



TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Germán Gómez-Herrero

Brain Connectivity Analysis with EEG



Julkaisu 877 • Publication 877

Tampere 2010

Tampereen teknillinen yliopisto. Julkaisu 877
Tampere University of Technology. Publication 877

Germán Gómez-Herrero

Brain Connectivity Analysis with EEG

Thesis for the degree of Doctor of Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB111, at Tampere University of Technology, on the 19th of March 2010, at 12 noon.

Supervisor

Karen Egiazarian, Dr.Tech., Professor
Department of Signal Processing
Tampere University of Technology
Tampere, Finland

Pre-examiners

Joachim Gross, Ph.D., Professor
Centre for Cognitive Neuroimaging
Department of Psychology
University of Glasgow
Glasgow, UK

Fa-Hsuan Lin, Ph.D., Assistant Professor
Athinoula A. Martinos Center for Biomedical Imaging
Massachusetts General Hospital,
Harvard Medical School, and
Massachusetts Institute of Technology
Charlestown, Massachusetts, USA

Opponent

Aapo Hyvärinen, Ph.D., Professor
Department of Mathematics and Statistics and
Department of Computer Science
University of Helsinki
Helsinki, Finland

ISBN 978-952-15-2342-7 (printed)
ISBN 978-952-15-2378-6 (PDF)
ISSN 1459-2045

Abstract

A problem when studying functional brain connectivity with EEG is that electromagnetic volume conduction introduces spurious correlations between any pair of EEG sensors. The traditional solution is to map scalp potentials to brain space before computing connectivity indices. The fundamental pitfall of this approach is that the EEG inverse solution becomes unreliable when more than a single compact brain area is actively involved in EEG generation. This thesis proposes an analysis methodology that partially overcomes this limitation. The underlying idea is that the inverse EEG problem is much easier to solve, if tackled separately for *functionally segregated* brain networks. The reason is that each of these *EEG sources* are likely to be spatially compact. In order to separate the contribution of each source to the scalp measurements, we use a blind source separation approach that takes into account that the sources, although functionally segregated, are not mutually independent but exchange information by means of *functional integration* mechanisms. Additionally, we also introduce a new set of information theoretic indices able to identify transient coupling between dynamical systems, and to accurately characterize coupling dynamics.

The analysis techniques developed in this thesis were used to study brain connectivity underlying the EEG-alpha rhythm, in a population of healthy elderly subjects, and in a group of patients suffering mild cognitive impairment (MCI). MCI is a condition at risk of developing to dementia and often a pre-clinical stage of Alzheimer's disease. The results of the analysis for the control population were in agreement with the previous literature on EEG-alpha, supporting the validity of the analysis approach. On the other hand, we found consistent connectivity differences between controls and MCIs, giving weight to the hypothesis that neurodegeneration mechanisms are active years before a patient is clinically diagnosed with dementia. Prospectively, these new analysis tools could provide a rational basis for evaluating how new drugs affect neural networks in early degeneration, which might have far-reaching implications for therapeutic drug development.

Preface

This thesis is based on part of the research carried out by the author between January 2004 and December 2009, at the Department of Signal Processing of Tampere University of Technology, Finland.

This work was accomplished under the supervision of Professor Karen Egiazarian to whom I am greatly indebted, not only for his guidance, but especially for the unconditional trust that he has put in me during all these years. I am also grateful to Professor José Luis Cantero from University Pablo de Olavide, Spain and to Petr Tichavský, from the Institute of Information Theory and Automation, Czech Republic, for sharing their knowledge and for inspiring many of the ideas presented in this thesis. Special thanks to Kalle Rutanen, who owns a share of whatever merit chapter 3 of this thesis may have. I would also like to express my sincere appreciation to Atanas Gotchev, Alpo Värri, and Professor Ulla Ruotsalainen for their valuable advice.

For the excellent research environment that I have enjoyed at the Department of Signal Processing, I have to thank Professor Moncef Gabbouj, Professor Ari Visa and Professor Jaakko Astola. Many thanks also to our secretaries, and especially to Virve Larmila, for helping me with so many practical issues.

I have been financially supported by the Graduate School of Tampere University of Technology, by Tampere Graduate School in Information Science in Engineering, by the Academy of Finland and by the European Commission. They are all gratefully acknowledged for making this research possible.

My warmest thanks go to Mirja, for her love and for making me so happy. I dedicate this thesis to my parents and to my brother. Their perseverance and dedication are examples that I will always aim to follow.

Tampere, March 2010

Germán Gómez-Herrero

Contents

Abstract	i
Preface	iii
Contents	v
1 Background and rationale	1
1.1 Brain connectivity analysis with EEG	2
1.2 Proposed approach	9
2 Blind Source Separation	13
2.1 Introduction	13
2.2 BSS of non-Gaussian i.i.d. sources	15
2.2.1 The mutual information contrast	16
2.2.2 The marginal entropy contrast	17
2.2.3 FastICA and EFICA	18
2.2.4 Optimization and reliability of ICA contrasts	21

2.3	BSS of spectrally diverse sources	25
2.4	Hybrid BSS algorithms	26
2.4.1	COMBI	27
2.4.2	M-COMBI	28
2.4.3	F-COMBI	32
2.5	BSS by entropy rate minimization	33
2.6	Experiments and results	36
2.6.1	Non-Gaussian and spectrally diverse sources	36
2.6.2	Non-linear sources with cross-dependencies	39
2.6.3	Real EEG data	46
2.7	Conclusions to the chapter	48
3	Measures of effective connectivity	51
3.1	Introduction	51
3.2	Information-theoretic indices	53
3.2.1	Estimation	57
3.3	Ensemble estimators	60
3.4	Experiments and results	61
3.4.1	Multivariate Gaussian distribution	63
3.4.2	Coupled Lorenz oscillators	64
3.4.3	Gaussian processes with time-varying coupling	68

3.4.4	Mackey-Glass electronic circuits	69
3.5	Conclusions to the chapter	70
4	Directional coupling between EEG sources	77
4.1	Introduction	77
4.2	Materials and methods	79
4.2.1	EEG model	79
4.2.2	Analysis procedure: VAR-ICA	80
4.2.3	Alternative approaches to VAR-ICA	85
4.2.4	Simulations	86
4.2.5	Assessing the accuracy of DTF estimates	89
4.2.6	EEG recordings and preprocessing	90
4.2.7	Reliability assessment	91
4.3	Results	94
4.3.1	Simulations	94
4.3.2	EEG Alpha	97
4.4	Conclusions to the chapter	98
5	Connectivity and neurodegeneration	103
5.1	Introduction	103
5.2	Methods	104
5.2.1	Subjects	104

5.2.2	EEG recordings and pre-processing	106
5.2.3	Connectivity analysis	106
5.2.4	Alpha peak frequency	107
5.2.5	Statistical analysis	107
5.3	Results	109
5.3.1	Alpha peak frequency	109
5.3.2	Connectivity between EEG-alpha sources	110
5.4	Discussion	111
6	Concluding remarks	115
A	Information theory	117
A.1	Basic definitions	117
A.2	Properties and relationships	119
B	The concept of state-space	121
	Bibliography	123

Introduction to the thesis

Outline of the thesis

This thesis is organized as follows. Chapter 1 gives the motivation of the thesis and reviews the most important methods that have been previously used to measure brain connectivity with EEG. Chapter 2 presents several algorithms for solving the linear and instantaneous blind source separation (BSS) problem, which plays a major role in the reconstruction of the neural sources underlying scalp EEG potentials. The proposed algorithms are extensively compared with other state-of-the-art BSS techniques using simulated sources and real EEG time-series. In chapter 3 we review the most important information theoretic indices that can be used to identify directional interactions between dynamical systems. Subsequently, we introduce the concept of partial transfer entropy and propose practical estimators that can be used to assess coupling dynamics in an ensemble of repeated measurements. Chapter 4 contains the most important contribution of the thesis and describes all the steps of the proposed connectivity analysis methodology. In chapter 5 we use the approach presented in chapter 4 to determine the differences in brain connectivity between a population of normal elderly controls and a group of patients suffering mild cognitive impairment. The concluding remarks and future research directions are given in chapter 6.

Publications and author's contribution

Most of the material presented in this monograph appears in the following publications by the author:

- [69] G. Gómez-Herrero, K. Rutanen, and K. Egiazarian. Blind source separa-

tion by entropy rate minimization. *IEEE Signal Processing Letters*, 17 (2): 153–156, February 2010.
DOI: 10.1109/LSP.2009.2035731

- [20] J. L. Cantero, M. Atienza, G. Gómez-Herrero, A. Cruz-Vadell, E. Gil-Neciga, R. Rodríguez-Romero, and D. García-Solis. Functional integrity of thalamocortical circuits differentiates normal aging from mild cognitive impairment. *Human Brain Mapping*, 30 (12): 3944–3957, December 2009.
DOI: 10.1002/hbm.20819
- [62] G. Gómez-Herrero, M. Atienza, K. Egiazarian, and J. L. Cantero. Measuring directional coupling between EEG sources. *Neuroimage*, 43(3): 497–508, November 2008.
DOI: 10.1016/j.neuroimage.2008.07.032
- [215] P. Tichavský, Z. Koldovský, A. Yeredor, G. Gómez-Herrero, and E. Doron. A hybrid technique for blind separation of non-Gaussian and time-correlated sources using a multicomponent approach. *IEEE Transactions on Neural Networks*, 19(3): 421–430, March 2008.
DOI: 10.1109/TNN.2007.908648
- [67] G. Gómez-Herrero, Z. Koldovský, P. Tichavský, and K. Egiazarian. A fast algorithm for blind separation of non-Gaussian and time-correlated signals. In *Proceedings of the 15th European Signal Processing Conference, EUSIPCO 2007*, pages 1731–1735, Poznan, Poland, September 2007.
- [66] G. Gómez-Herrero, E. Huupponen, A. Värri, K. Egiazarian, B. Vanrumste, A. Vergult, W. De Clercq, S. Van Huffel, and W. Van Paesschen. Independent component analysis of single trial evoked brain responses: Is it reliable? In *Proceedings of the 2nd International Conference on Computational Intelligence in Medicine and Healthcare, CIMED 2005*, Costa da Caparica, Portugal, June 2005

The contents of chapter 3 are still unpublished but parts of the chapter are included in a manuscript that is currently under review:

- [70] G. Gómez-Herrero, W. Wu, K. Rutanen, M. C. Soriano, G. Pipa, and R. Vicente. Assessing coupling dynamics from an ensemble of time-series. *Submitted*.

The contents of this thesis are also closely related to the following publications by the author:

- [63] G. Gómez-Herrero, W. De Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Van Huffel, and W. Van Paesschen. Automatic removal of ocular artifacts in the EEG without a reference EOG channel. In *Proceedings of the 7th Nordic Signal Processing Symposium NORSIG 2006*, Reykjavik, Iceland, June 2006.
DOI: 10.1109/NORSIG.2006.275210
- [32] I. Christov, G. Gómez-Herrero, V. Krasteva, I. Jekova, and A. Gotchev. Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification. *Medical Engineering & Physics*, 28(9):876–887, November 2006.
DOI: 10.1016/j.medengphy.2005.12.010
- [61] Gómez-Herrero, I. Jekova, V. Krasteva, I. Christov, A. Gotchev, and K. Egiazarian. Relative estimation of the Karhunen-Loève transform basis functions for detection of ventricular ectopic beats. In *Proceedings of Computers in Cardiology, CinC 2006*, pages 569 – 572, Valencia, Spain, September, 2006.
- [87] E. Huupponen, W. De Clercq, G. Gómez-Herrero, A. Saastamoinen, K. Egiazarian, A. Värri, A. Vanrumste, S. Van Huffel, W. Van Paesschen, J. Hasan, and S.-L. Himanen. Determination of dominant simulated spindle frequency with different methods. *Journal of Neuroscience Methods*, 156(1–2):275–283, September 2006.
DOI: 10.1016/j.jneumeth.2006.01.013
- [65] G. Gómez-Herrero, A. Gotchev, I. Christov, and K. Egiazarian. Feature extraction for heartbeat classification using matching pursuits and independent component analysis. In *Proceedings of the 30th International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2005*, Philadelphia, USA, pages 725–728, March 2005.
DOI: 10.1109/ICASSP.2005.1416111

The contribution of the author of this thesis to all the publications above has been crucial. In [20], the connectivity analysis was done using software provided by the author of this thesis, who also wrote the description of the method and had an active participation in writing the rest of the paper. However, the electrophysiological and genetic measurements, the statistical analysis, and the neurological interpretation of the results were done by the other co-authors. In [215], the author of this thesis proposed the original idea of combining complementary BSS algorithms based on the concept of multidimensional independent components¹. The author also had major contributions to the selection of the clustering strategy and to the realization of the numerical experiments. In [32], the author contributed the time-frequency descriptors and participated

¹This idea was first proposed in a research report by the author of this thesis [64].

in writing the paper. In [87], the author of this thesis contributed one of the compared methods (the best performing one) but most of the simulations and the writing were carried out by E. Huupponen. In all other publications above the author has been the main contributor.

Notation and conventions

Along the text of this thesis we try to explain any non-obvious notation whenever is first used. For convenience we summarize here the most important conventions that have been adopted.

Matrix operands are denoted with uppercase boldface fonts (e.g. \mathbf{A}), vectors are in lowercase boldface fonts (e.g. \mathbf{v}) and scalars are in italic fonts (e.g. A or a). However, \mathbf{A}_{ij} means the scalar element of matrix \mathbf{A} that is located in the i th row and in the j th column. Unless otherwise stated in the text, any vector is assumed to be a column vector so that $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_M]$ is a matrix that has vectors $\mathbf{a}_1, \dots, \mathbf{a}_M$ as columns. The transpose of matrix \mathbf{A} is denoted by \mathbf{A}^T , the Moore-Penrose pseudoinverse is \mathbf{A}^+ and the inverse \mathbf{A}^{-1} . In general, we assume continuous-valued discrete-time variables and time varying entities will be indexed by n (e.g. $\mathbf{A}(n)$, $\mathbf{a}(n)$, $a(n)$) with n a natural number that denotes the corresponding sampling instant. Continuous-time equations use the time index t .

The cardinality of a discrete set Γ is denoted by $|\Gamma|$. We denote the l^p norms as $\|\cdot\|_p$. The hat decoration $\hat{\cdot}$ denotes estimated values (e.g. \hat{x} is the estimate of x).

In chapter 2 we use the interference-to-signal ratio (ISR) [23] as a standard measure of source estimation accuracy in blind source separation (BSS) problems. The ISR between the k th and l th source estimate is given by $\Psi_{kl} = \mathbf{G}_{kl}^2 / \mathbf{G}_{kk}^2$ where $\mathbf{G} = \hat{\mathbf{B}}\mathbf{A}$, \mathbf{A} being the true mixing matrix and $\hat{\mathbf{B}}$ an estimate of \mathbf{A}^{-1} . The total ISR for the k th source estimate is given by $\Psi_k = \sum_i \Psi_{ki}$.

Random variables are denoted by uppercase italics (e.g. X) but in occasions we also use the notation \mathbf{x} to denote a random vector. For convenience we often use indistinctly \mathbf{x} or $p(\mathbf{x})$ to refer to the probability density function (pdf) of a random vector \mathbf{x} . Notice the difference between \mathbf{x} and $\mathbf{x}[n]$, where the latter denotes a realization of random vector \mathbf{x} at the sampling instant n . A discrete-time stochastic processes is denoted by $\{x(n)\}_n$ or $\{\mathbf{x}(n)\}_n$ if it is vector-valued. Then $\mathbf{x}(n)$ is used to refer to the distribution of stochastic

process $\{\mathbf{x}(n)\}_n$ at time-instant n , while a sample realization at time-instant n is denoted with square brackets, i.e. $\mathbf{x}[n]$. The M -dimensional Gaussian distribution is denoted by $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu} = [\mu_1, \dots, \mu_M]^T$ is the mean vector and $\boldsymbol{\Sigma}$ the covariance matrix of the distribution.

The differential entropy of a continuous random vector X (or \mathbf{x}) can be denoted by any of the following: $H(X) \equiv H(\mathbf{x}) \equiv H(p(\mathbf{x})) \equiv H_X$. Similarly, the Kullback-Leibler divergence between the distributions of two random vectors \mathbf{x} and \mathbf{y} can use any of the notations $K(X|Y) \equiv K(\mathbf{x}|\mathbf{y}) \equiv K(p(\mathbf{x})|p(\mathbf{y}))$ or even $K(p|q)$ if it has been previously specified that the pdf of \mathbf{x} is denoted by $p(\mathbf{x})$ and the pdf of \mathbf{y} is denoted by $q(\mathbf{y})$. Conditional probabilities are denoted by $p(X|Y)$ which means the pdf of random variable X given Y .

In chapter 3 we use the notation $I_{X \leftrightarrow Y}$ to refer to the mutual information between random variables X and Y . Similarly, $I_{X \leftrightarrow Y|Z}$ means partial mutual information between X and Y given Z and $T_{X \leftarrow Y|Z}$ denotes partial transfer entropy from Y towards X given Z . The operator $\langle \cdot \rangle_n$ stands for mean with respect to index n .

In chapter 4 the overall estimation accuracy for the directed transfer function (DTF) is assessed using the following index (in percentage):

$$\epsilon = 100 \cdot \frac{1}{M^2 N} \sum_{ijf} \sqrt{(\gamma_{ij}(f) - \hat{\gamma}_{ij}(f))^2} \quad (1)$$

where $\gamma_{ij}(f)$ denotes the true DTF from the j th source towards the i th source, $\hat{\gamma}_{ij}(f)$ stands for the corresponding DTF estimate, M is the total number of sources and N is the number of frequency bins of the DTF function. Since the DTF at a certain frequency is within the range $[0, 1]$, ϵ ranges from 0% (best case, no estimation error) to 100% (worst case, maximum possible estimation error).

Table 1 below contains the most common acronyms and abbreviations used in the thesis.

AD	Alzheimer's disease
ANOVA	Analysis of variance
AR	Autoregressive model
BSS	Blind source separation
CRLB	Cramér-Rao lower bound
DC	Directed coherence
DCM	Dynamic causal modeling
DTF	Directed transfer function
EEG	Electroencephalography
EFICA	Efficient FastICA [110]
ENRICA	Entropy rate-based independent component analysis [69]
ERP	Event-related potential
GC	Granger causality
ICA	Independent component analysis
i.i.d.	Independent and identically distributed
ISR	Interference-to-signal ratio
JADE	Joint approximate diagonalization of eigen-matrices [25]
JADE_{TD}	BSS using simultaneously JADE and TDSEP [154]
KL divergence	Kullback-Leibler divergence
KL estimator	Kozachenko-Leonenko estimator of differential entropy [112]
LORETA	Low resolution brain electromagnetic tomography [170]
LS	Least squares
MCI	Mild cognitive impairment
ME	Marginal entropy
MEG	Magnetoencephalography
MI	Mutual information
MILCA	Mutual information least-dependent component analysis [210]
ML	Maximum likelihood
NpICA	Non-parametric independent component analysis [15]
PCA	Principal component analysis
pdf	Probability density function
PMI	Partial mutual information [55]
PSI	Phase-slope index [159]
PTE	Partial transfer entropy [70]
RADICAL	Robust accurate direct independent component analysis [122]
SBNR	Signal-to-biological noise ratio
SNR	Signal-to-measurement noise ratio
SOBI	Second order blind identification [13]
TDSEP	Temporal decorrelation separation [233]
TE	Bi-variate transfer entropy [195]
VAR	Vector autoregressive model
WASOBI	Weights adjusted second order blind identification [228]

Table 1: List of frequently used acronyms and abbreviations.

Chapter 1

Background and rationale

The *electroencephalogram* (EEG) is a record of the temporal variations of brain electric potentials recorded from electrodes on the human scalp. The EEG is closely related to its magnetic counterpart, the *magnetoencephalogram* (MEG). EEG and MEG measure the same underlying electrical phenomena and their relative strenghts are still a matter of debate [121, 127, 139, 140]. In this thesis we focus on the EEG but most of the proposed techniques could be directly applied to MEG as well.

The human brain is an extremely complicated network that probably contains of the order of 10^{10} interconnected neurons. Each neuron consists of a central portion containing the nucleus, known as the cell body, and one or more structures referred to as axons and dendrites. The dendrites are rather short extensions of the cell body and are involved in the reception of stimuli. The axon, by contrast, is usually a single elongated extension.

Rapid signaling within nerves and neurons occurs by means of *action potentials*, which consist on a rapid swing (lasting around 1 ms) of the polarity of the neuron transmembrane voltage from negative to positive and back. These voltage changes result from changes in the permeability of the membrane to specific ions, the internal and external concentration of which are in imbalance. An action potential produces a current flow from the cell body to the axon terminal. However, the effect of these currents on scalp EEG potentials is probably negligible, since it is unlikely that enough spatially aligned neurons would fire simultaneously during the short duration of action potentials in order to form a measurable current dipole.

An action potential reaching a synapse triggers the release of neurotransmitters that bind to the receptors of a post-synaptic neuron. If the neurotransmitter is excitatory (resp. inhibitory), electrical current flows from the postsynaptic cell to the environment (resp. the opposite), therefore depolarizing (resp. polarizing) the cell membrane. These post-synaptic potentials produce a current flow (a dipole) that lasts tens or even hundreds of milliseconds. During this long time window, many spatially aligned dipoles (in the order of hundreds of millions [160]) may become simultaneously active, making such an event detectable at the scalp EEG sensors. Large pyramidal neurons in the neocortex are a major source of scalp EEG potentials, due to the spatial alignment of their dendritic trees perpendicular to the cortical surface [130] (see Fig. 1.1). Nevertheless, contributions from deep sources have also been reported [60, 62, 98, 217].

Due to its non-invasive nature and low cost, the EEG has become the method of choice for monitoring brain activity in many clinical and research applications. Moreover, EEG (together with MEG) is the only functional neuroimaging technique with enough temporal resolution to study fast cognitive processes. There are three basic neuroelectric examinations based on scalp brain potentials: (i) EEG studies that involve inspection of spontaneous brain activity in different experimental settings, (ii) event-related potential studies (ERPs) that use signal-averaging and other processing techniques to extract weak neural responses that are time-locked to specific sensory, motor or cognitive events and (iii) studies of event-induced modulation of ongoing brain activity. The methods developed in this thesis were mostly designed with sights set at the analysis of spontaneous EEG activity¹, but their generalization to induced oscillatory brain activity is rather straightforward, using for instance the same approach as in [211].

1.1 Brain connectivity analysis with EEG

The goal of cognitive neuroscience is to describe the neural mechanisms underlying cognition. Compelling evidence has firmly established that brain cells with common functional properties are grouped together into specialized (and possibly anatomically segregated) brain areas. Based on this principle of *functional segregation* [56], neuroimaging studies have traditionally aimed at identifying the brain areas that are dedicated to specific information processing tasks. However, high-level cognitive functions are likely to require the *functional integration* of many specialized brain networks [51] and neuroimaging studies investigating dependencies between remote neurophysiological events

¹An exception is chapter 3 where several connectivity indices specifically suited for ERPs are introduced.

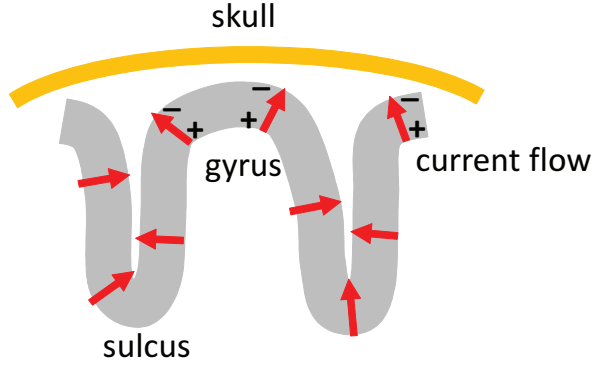


Figure 1.1: The EEG signal results mainly from the postsynaptic activity of the pyramidal neurons in the surface of the brain. Scalp potentials are especially sensitive to radially oriented dipoles generated in pyramidal neurons in the gyri.

(i.e. *functional connectivity*) have become increasingly prevalent. This interest has been further fostered by groundbreaking theories suggesting a major role of systems-level brain connectivity in neurodegeneration [19, 166], and in the emergence of consciousness [2, 143].

A major problem when studying connectivity between brain areas with EEG is that coupling between scalp EEG signals does not necessarily imply coupling between the underlying neural sources. The reason is that scalp EEG potentials do not exclusively reveal averaged postsynaptic activity from localized cortical regions beneath one electrode. On the contrary, they reflect the superposition of all active coherent neural sources located anywhere in the brain, due to conduction effects in the head volume [138, 160]. This superposition inevitably leads to misinterpretations of the connectivity results obtained between scalp EEG signals, especially when subcortical generators are actively involved (see Fig. 1.2 for an illustration of these effects).

An elementary vector dipole \mathbf{q} in the head volume is fully defined by its location vector (\mathbf{r}_q), its magnitude (a scalar, m) and its orientation (i.e. a pair of spherical coordinates $\Theta = \{\theta, \phi\}$). The electric potential generated by such elementary dipole at a scalp location \mathbf{r} is given by [8]:

$$v(\mathbf{r}) = a(\mathbf{r}, \mathbf{r}_q, \Theta)m \quad (1.1)$$

where $a(\mathbf{r}, \mathbf{r}_q, \Theta)$ is the solution to the quasi-static approximation of the forward electromagnetic problem [139], which, regardless of the head model con-

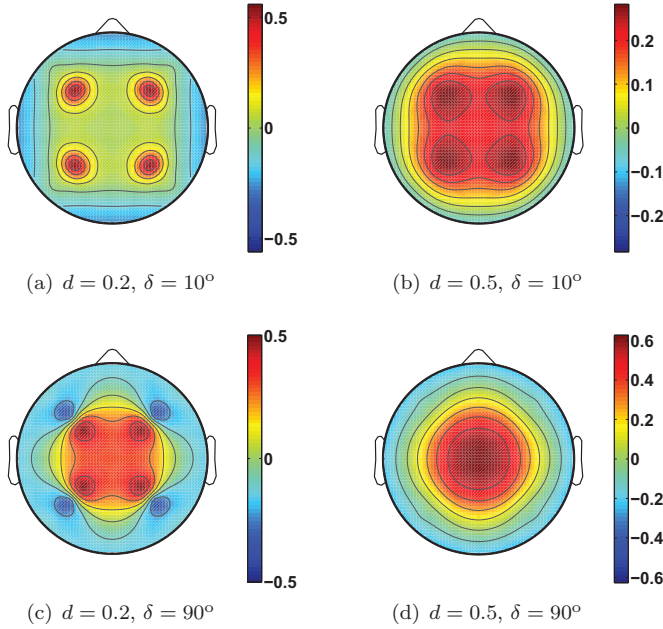


Figure 1.2: Distribution of scalp potentials generated on the surface of a single-layer spherical head when varying the radial angle (δ , degrees) and the depth (d , normalized with respect to the head radius) of four simulated dipoles. See Fig. 4.3 for an illustration of the locations of the dipoles. The figures clearly show that brain activity generated in the same brain locations (left column) can lead to completely different patterns of scalp potentials, depending on the orientation of the dipoles. At the same time, when the dipoles are very deep (Fig. 1.2(d)) their activity propagates across all scalp electrodes, making it difficult to identify at the scalp the number and location of the underlying generators. Only for the case of radially oriented and very shallow dipoles (Fig. 1.2(a)) one may study mutual interactions between the four EEG sensors located just above the dipoles.

sidered, always depends linearly on Θ and non-linearly on \mathbf{r}_q [153]. The scalp potential generated by R simultaneously active dipoles can be simply obtained by linear superposition:

$$v(\mathbf{r}) = \sum_i^R a(\mathbf{r}, \mathbf{r}_{qi}, \Theta_i) m_i \quad (1.2)$$

In the case of simultaneous EEG measurements at K scalp sensors we can write [8]:

$$\begin{aligned} \mathbf{v}(\{\mathbf{r}_i\}) &= \begin{bmatrix} v(\mathbf{r}_1) \\ \vdots \\ v(\mathbf{r}_K) \end{bmatrix} = \begin{bmatrix} a(\mathbf{r}_1, \mathbf{r}_{q1}, \Theta_1) & \cdots & a(\mathbf{r}_1, \mathbf{r}_{qR}, \Theta_R) \\ \vdots & \ddots & \vdots \\ a(\mathbf{r}_K, \mathbf{r}_{q1}, \Theta_1) & \cdots & a(\mathbf{r}_K, \mathbf{r}_{qR}, \Theta_R) \end{bmatrix} \begin{bmatrix} m_1 \\ \vdots \\ m_R \end{bmatrix} \\ &= \mathbf{A}(\{\mathbf{r}_{qi}, \Theta_i\}) \mathbf{m} \end{aligned} \quad (1.3)$$

where $\mathbf{A}(\{\mathbf{r}_{qi}, \Theta_i\})$ is the *leadfield matrix* mapping dipole magnitudes to scalp measurements. Each column of \mathbf{A} is commonly referred to as the *forward field* or *scalp topography* of the corresponding dipole. A discrete time component $n = 1, 2, \dots, L$ can be easily incorporated to (1.3) in order to account for time-evolving dipole strengths:

$$\mathbf{v}(\{\mathbf{r}_i\}, n) = \mathbf{A}(\{\mathbf{r}_{qi}, \Theta_i\}) \mathbf{m}(n) \quad (1.4)$$

Eq. (1.4) above defines a *fixed dipole model* because the orientations of the dipoles do not change with time. Although models with *rotating dipoles* are also possible [152, 190], we do not consider them in this thesis.

Based on the fixed dipole model, source connectivity analysis with EEG involves using the measured time-series of scalp potentials $v(\mathbf{r}_1, n), \dots, v(\mathbf{r}_K, n)$ to (i) assess mutual interactions between the underlying dipole activations $m_1(n), \dots, m_R(n)$ and (ii) to determine the cerebral localizations $\mathbf{r}_{q1}, \dots, \mathbf{r}_{qR}$ of those signal generators. Current approaches to this problem fall within two broad categories: parametric modeling and imaging methods. The former is based on the assumption that brain activity can be well represented by few equivalent current dipoles (ECDs) of unknown locations and orientations. On the other hand, imaging approaches consider distributed current sources

containing thousands of dipoles and impose only generic spatial constraints on the inverse solution.

The most straightforward parametric approach to the inverse problem is to find the set of R dipoles that minimizes the following least-squares contrast:

$$J_{LS}(\{\mathbf{r}_{qi}, \Theta_i\}, \mathbf{m}) = \|\mathbf{v}(\{\mathbf{r}_i\}) - \mathbf{A}(\{\mathbf{r}_{qi}, \Theta_i\})\mathbf{m}\|_2^2 \quad (1.5)$$

For any choice of $\{\mathbf{r}_{qi}, \Theta_i\}$, the optimal (in the least-squares sense) dipole magnitudes are:

$$\mathbf{m} = \mathbf{A}^+ \mathbf{v} \quad (1.6)$$

where $^+$ denotes pseudoinversion. Then, optimization of $J_{LS}(\{\mathbf{r}_{qi}, \Theta_i\}, \mathbf{m})$ can be more efficiently done in two steps. First, solve in $\{\mathbf{r}_{qi}, \Theta_i\}$ by minimizing the following cost function:

$$J_{LS}(\{\mathbf{r}_{qi}, \Theta_i\}) = \|\mathbf{v} - \mathbf{A}\mathbf{A}^+ \mathbf{v}\|_2^2 \quad (1.7)$$

and then obtain the dipole magnitudes with (1.6). By using an entire block of data in the least-squares fit, the temporal activations of the underlying dipoles can be reconstructed and source connectivity can be assessed using standard synchronization measures [81]. The drawbacks of the least-squares method are that the number of dipoles has to be decided a priori and that the nonconvexity of the cost function increases rapidly with the number of dipoles. This prevents using more than just few (e.g. 3 or 4) dipoles, if one wants to avoid getting systematically trapped in local minima.

Another popular parametric approach to source connectivity is linearly constrained minimum variance (LCMV) beamforming. LCMV beamformers retrieve the activity generated by a dipole at location \mathbf{r}_q with orientation Θ by means of a spatial filter (a $K \times 1$ vector of scalar coefficients) \mathbf{w} that solves the following linearly constrained optimization problem [74, 219, 220]:

$$\min_{\mathbf{w}} \{\mathbf{w}^T \Sigma_{\mathbf{v}} \mathbf{w}\} \quad (1.8)$$

subject to:

$$\mathbf{w}^T \mathbf{a}(\mathbf{r}_q, \Theta) = 1 \quad (1.9)$$

where $\Sigma_{\mathbf{v}}$ is the covariance matrix of the scalp EEG potentials. Using the method of Lagrange multipliers, the solution to (1.8) can be found be [220]:

$$\mathbf{w} = \Sigma_{\mathbf{v}}^{-1} \mathbf{a} [\mathbf{a}^T \Sigma_{\mathbf{v}}^{-1} \mathbf{a}]^{-1} \quad (1.10)$$

The LCMV beamformer tries to minimize the output variance of the spatial filter while leaving untouched the activity originating in the dipole of interest. Intuitively, this is equivalent to a spatial filter with a fixed passband and a data-adaptive stop-band. Eq. (1.8) may also incorporate a linear transformation of the scalp potentials in the time-domain. For instance, the Fourier transform allows defining a frequency-dependent spatial filter, which is especially suitable for the analysis of rhythmic brain activity. This is precisely the principle behind the so-called dynamic imaging of coherent sources [74, 117]. Other transforms (e.g. the wavelet transform) could also be used to define filters that let pass only activity located in certain area of the time-frequency plane [38, 120]. In practice, the ability of beamformers to remove interfering sources is limited, due to the reduced number of degrees of freedom, and due to the presence of cross-dependencies between brain sources.

Although beamformers can target brain areas selected a priori, this is a rather risky strategy because imprecise dipole locations can result in signal attenuation or even cancellation. A probably safer route is to define the beamformer target based on the ratio between the output variance of the beamformer at a given brain location and the output variance that would be obtained in the presence of noise only [230]:

$$\text{var}(\mathbf{r}_q) = \frac{\mathbf{a}^T \Sigma_{\epsilon}^{-1} \mathbf{a}}{\mathbf{a}^T \Sigma_{\mathbf{v}}^{-1} \mathbf{a}} \quad (1.11)$$

where Σ_{ϵ} is an estimate of the noise covariance. Clearly, localization of brain activity can then be done by finding the maxima of ratio (1.11). An alternative approach to identify the brain locations of interest is multiple signal classification (MUSIC) [152, 192]. MUSIC starts by performing a singular value decomposition (SVD) of the following $K \times L$ matrix of scalp potentials:

$$\mathbf{V} = \begin{bmatrix} v(\mathbf{r}_1, 1) & \cdots & v(\mathbf{r}_1, L) \\ \vdots & \ddots & \vdots \\ v(\mathbf{r}_K, 1) & \cdots & v(\mathbf{r}_K, L) \end{bmatrix} \quad (1.12)$$

where the columns correspond to different sampling instants and the rows to

EEG channels. The SVD decomposition yields the factorization $\mathbf{V} = \mathbf{U}\Sigma\mathbf{V}^T$. Assuming that $K > R$ and that the signal-to-noise ratio (SNR) is sufficiently large, the R first columns of \mathbf{U} , denoted by \mathbf{U}_S , form a basis for the signal subspace, while the noise subspace is spanned by the remaining columns. Then, MUSIC's brain activity function is defined as:

$$J(\mathbf{r}, \Theta) = \frac{\|\mathbf{P}_s^\perp \mathbf{a}(\mathbf{r}, \Theta)\|_2^2}{\mathbf{a}(\mathbf{r}, \Theta)} \quad (1.13)$$

where $\mathbf{P}_s^\perp = \mathbf{I} - \mathbf{U}_S \mathbf{U}_S^T$ is the orthogonal projector onto the noise subspace, and $\mathbf{a}(\mathbf{r}, \Theta)$ is the scalp topography for a dipole at location \mathbf{r} with orientation Θ . Function $J(\mathbf{r}, \Theta)$ is zero when $\mathbf{a}(\mathbf{r}, \Theta)$ corresponds to one of the true source locations and, therefore, the reciprocal of $J(\mathbf{r}, \Theta)$ has R peaks at or near the true locations of the R sources. Once the locations of the sources have been found, their time activations can be estimated using a beamformer like LCMV or simply using (1.6).

Imaging approaches to the inverse EEG problem avoid altogether the estimation of the location and orientation of the source dipoles. Instead, they build a dense grid of dipoles covering all brain regions where EEG activity could be plausibly generated. This grid is usually built upon an anatomical magnetic resonance (MR) image of the subject, so that dipoles are allowed to lie only within the cortex and few deep gray matter structures. Then, the imaging problem reduces to solving the linear system $\mathbf{v} = \mathbf{A}(\{\mathbf{r}_{qi}, \Theta_i\})\mathbf{m}$ for the dipole amplitudes \mathbf{m} . Since the grid of brain locations contains of the order of ten to one hundred thousand dipoles, the EEG imaging problem is hugely under-determined and constraints need to be imposed on the allowed current source distributions. Typically, this has been achieved through the use of regularization or Bayesian image restoration methods. A detailed review on this type of inverse solvers can be found elsewhere [8, 146]. We will just say here that the most common approaches enforce the sources to be smooth [165, 169, 170] and therefore suffer of poor spatial resolution. As a result, the number of active dipoles is typically very large and there is no obvious spatial separation between different EEG sources. Thus, dipoles need to be grouped into regions of interest (ROI) either manually using a priori knowledge [126, 211], or by means of heuristic and rather ad-hoc procedures [40, 116, 117, 124]. The lack of objective and theoretically funded approaches for the selection of ROIs is a major weakness of imaging techniques based on smoothness constraints.

Mapping scalp potentials to brain space is only half of the problem in brain connectivity studies with EEG. Another important issue is how to identify and characterize functional relationships between EEG signals in brain space. This is an especially difficult problem when we aim to study *effective*

connectivity [56], i.e. directed causal connections between cerebral systems. One approach is to use dynamical causal modeling (DCM) [58, 108] to model all data generation steps, from the neural signals to the transformation that these signals undertake before becoming the observed EEG measurements. By incorporating absolutely all the relevant parameters, DCM allows the inference of the specific neural mechanisms underlying a given effective connection. However, DCM requires a great deal of a priori information, which is often unavailable or inaccurate, especially when studying neurodegeneration. Moreover, the dynamical equations of EEG generators are still largely unknown which explains why DCM has been only rarely used with EEG². An alternative to DCM is based on the so-called Granger causality (GC) [73], which leans on the simple idea that the cause occurs before the effect and, therefore, knowledge of the cause helps forecasting the effect. Traditionally, GC-based methods have used linear vector autoregressive (VAR) models to quantify directed influences between EEG sources in the frequency domain [7, 49, 104, 119]. More recently, several GC indices based on information theory have been proposed, which are also sensitive to non-linear interactions [55, 195]. In this thesis, the analysis of real EEG data will be performed with linear GC indices, due to their well-known properties and proven robustness [105]. Nevertheless, in chapter 3 we use simulations to investigate the promising properties of information theoretic approaches, and propose novel measures for the characterization of time varying coupling patterns.

1.2 Proposed approach

In this thesis we use an extension of the fixed dipoles model (1.4). In our model, brain sources are represented by clusters of synchronous dipoles rather than by single dipoles. Let us consider M clusters and denote by Γ_i the set of dipoles belonging to the i th cluster. Then:

²Moreover, EEG studies using DCM have involved almost exclusively evoked and induced brain responses and not spontaneous EEG. See, however, recent works by Moran et al. [148, 149] for an approach to DCM of steady-state local field potentials.

$$\begin{aligned}
\mathbf{v}(\{\mathbf{r}_i\}, n) &= \\
&= \begin{bmatrix} \sum_{j \in \Gamma_1} a(\mathbf{r}_1, \mathbf{r}_{qj}, \Theta_j) m_j & \cdots & \sum_{j \in \Gamma_M} a(\mathbf{r}_1, \mathbf{r}_{qj}, \Theta_j) m_j \\ \vdots & \ddots & \vdots \\ \sum_{j \in \Gamma_1} a(\mathbf{r}_K, \mathbf{r}_{qj}, \Theta_j) m_j & \cdots & \sum_{j \in \Gamma_M} a(\mathbf{r}_K, \mathbf{r}_{qj}, \Theta_j) m_j \end{bmatrix} \begin{bmatrix} s_1(n) \\ \vdots \\ s_M(n) \end{bmatrix} \\
&= \mathbf{B}(\{\mathbf{r}_{qi}, \Theta_i, m_i\}) \mathbf{s}(n)
\end{aligned} \tag{1.14}$$

where each column of matrix $\mathbf{B}(\{\mathbf{r}_{qi}, \Theta_i, m_i\})$ now contains the scalp topography of a cluster of synchronous dipoles, or *brain source*. Notice that the temporal dynamics of all dipoles within a brain source are identical up to a scaling factor. This thesis rests upon the assumption that functionally segregated EEG sources can be approximately modeled by such dipole clusters, i.e. they can be characterized by a single temporal activation and a single scalp topography. If these scalp topographies are linearly independent then model (1.14) is an instance of the well-known linear and instantaneous blind source separation (BSS) problem³. Thus, the rationale of the proposed approach consists of the following steps:

- Use BSS techniques to estimate the temporal activation of the brain sources and their corresponding scalp topographies, i.e. obtain $\mathbf{s}(n)$ and $\hat{\mathbf{B}} = [\hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_M]$.
- Assess connectivity between the time courses of the brain sources.
- Using the scalp topography of a single brain source $\hat{\mathbf{b}}_i$ solve the inverse EEG problem, in order to obtain the magnitudes $m_1, \dots, m_{|\Gamma_i|}$ of the synchronous dipoles associated to that source.

The three steps above are depicted in Fig. 1.3. The first step is discussed in detail in chapter 2 where three novel BSS algorithms are also introduced. Chapter 3 reviews the most common indices used for assessing connectivity between dynamical systems and presents novel indices for the analysis of short-duration event-related EEG potentials. An integrated analysis framework is described and applied to real EEG in chapters 4 and 5.

³If $K > M$ the problem is sometimes called blind source extraction [33]. However, in this thesis we will enforce $K = M$ by linearly projecting the observed scalp potentials to their signal subspace using principal component analysis (PCA [100]).

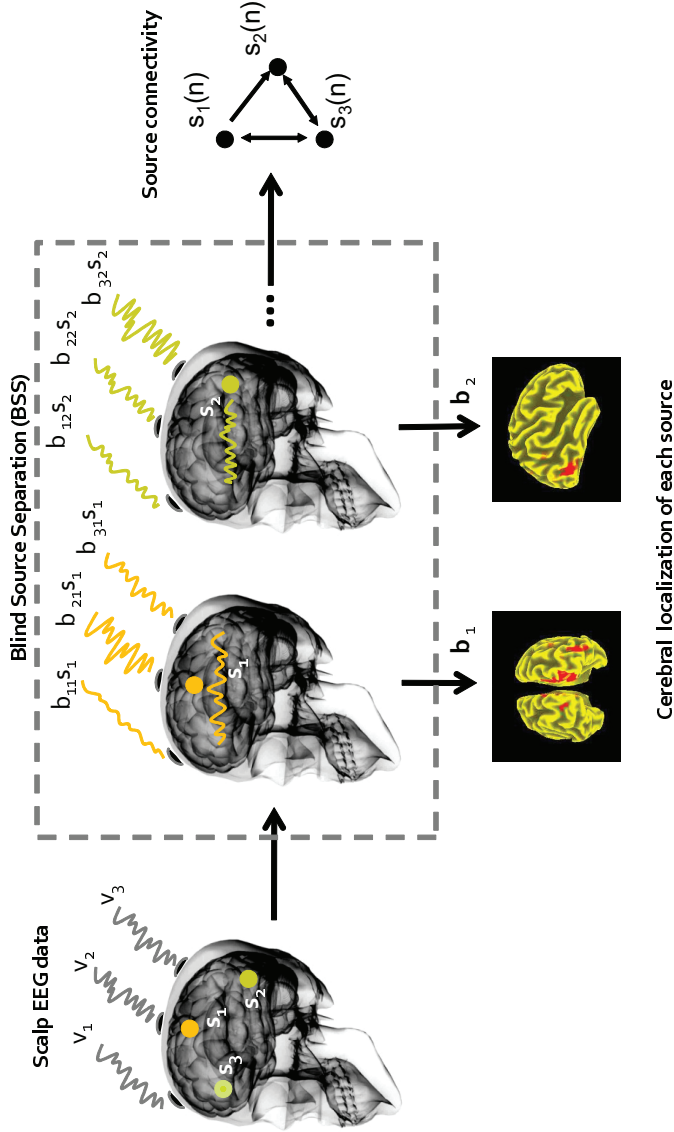


Figure 1.3: A schematic of the rationale behind the proposed methodology for assessing functional connectivity between neural EEG sources. Since the sources are functionally connected, the blind separation problem cannot be solved by enforcing mutual independence of the sources. At the same time, individual brain sources are likely to be anatomically compact and, therefore, they are probably easier to reconstruct individually, using inverse imaging methods based on smoothness constraints, or with ECD-based approaches [213]. A practical implementation of the whole methodology is described in chapter 4.

Chapter 2

Blind Source Separation

2.1 Introduction

Recall from chapter 1 that we assume that EEG potentials recorded at K scalp locations $\mathbf{v}(n) = [v_1(n), \dots, v_K(n)]^T$ can be approximately modeled as a *linear* and *instantaneous* superposition of $M \leq K$ underlying *brain sources* $\mathbf{s}(n) = [s_1(n), \dots, s_M(n)]^T$, i.e.:

$$\mathbf{v}(n) = \mathbf{\Omega}\mathbf{s}(n) + \boldsymbol{\eta} \quad (2.1)$$

where $\mathbf{\Omega}$ is an unknown $K \times M$ matrix having as columns the spatial distribution of scalp potentials generated by each source and $\boldsymbol{\eta} = [\eta_1, \dots, \eta_K]^T$ denotes additive *measurement noise*. We neglect for now the contribution of noise and we assume that $\mathbf{v}(n)$ has been linearly projected to its M -dimensional signal sub-space $\mathbf{x}(n) = [x_1(n), \dots, x_M(n)]^T$ so that:

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) = \sum_{j=1}^M \mathbf{a}_j s_j(n) \quad (2.2)$$

where $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_M]$ is an unknown $M \times M$ mixing matrix which is assumed to be of full column rank. The goal of blind source separation (BSS) is to estimate a separating matrix \mathbf{B} such that the source signals can be ap-

proximately recovered up to a permutation and scaling indeterminacy, i.e. $\mathbf{BA} \approx \mathbf{PA}$ where \mathbf{P} and \mathbf{A} are an arbitrary permutation matrix and an arbitrary diagonal matrix, respectively. This problem is found not only in the analysis of EEG data but also in a variety of applications ranging from wireless communications [177] and the geosciences [171] to image processing [11]. The term "blind" means that generic assumptions are made regarding the source signals but no a priori knowledge on the mixing coefficients is available. Most BSS algorithms are based on the common premise of mutually independent sources. Then, separation is achieved by optimizing a suitable BSS *contrast* that exploits either non-Gaussianity, spectral diversity or non-stationarity of the independent sources [26, 168]. For each of these three models, there exist algorithms which are asymptotically optimal under certain conditions:

- Efficient FastICA (EFICA) [110] for i.i.d. generalized-Gaussian distributed non-Gaussian sources.
- Weights-adjusted second-order blind identification (WASOBI) [228] for wide sense stationary parametric Gaussian sources with *spectral diversity*.
- Block Gaussian likelihood (BGL) [176] for Gaussian sources with *time-varying variances*.

Indeed, EEG sources are likely to fit approximately more than one of these models but probably none of them perfectly. Consequently, algorithms unifying two [36, 72, 83, 154] or even the three models [76, 91] have been proposed in the literature. In this chapter we present three BSS algorithms [67, 69, 215] that combine the first two models above and that offer different trade-offs between accuracy and computational complexity. The advantages of these novel algorithms over the state-of-the-art are highlighted using simulated source signals and real EEG data.

A fundamental pitfall of independence-based BSS contrasts is that they underperform in the presence of cross-dependencies between EEG sources. Such cross-dependencies are likely to occur due to (time-lagged) axonal flows of information across distributed brain areas. This problem has been largely overlooked in the literature but can seriously compromise the reliability of the estimated EEG sources. Algorithms assuming i.i.d. sources are also negatively affected by the characteristic $1/f$ spectrum of EEG signals making them prone to overlearning [189]. Special precautions must be taken in the analysis of single-trial event-related EEG potentials since the overlearned sources may have biologically plausible shapes [66]. These concerns are also briefly studied in this chapter and solutions are proposed.

2.2 BSS of non-Gaussian i.i.d. sources

Most BSS algorithms for non-Gaussian i.i.d. sources are ultimately based on the maximum likelihood (ML) principle. Due to the lack of temporal structure, the sources can be treated as a random vector \mathbf{s} which is fully characterized by its probability density function (pdf), denoted by $P_{\mathbf{s}}$. Then, given a set of N realizations of the mixed observations \mathbf{x} , the normalized log-likelihood is [25]:

$$L_N(\mathbf{A}|\mathbf{x}) \triangleq \frac{1}{N} \log \prod_{n=1}^N p(\mathbf{x}[n] | \mathbf{A}) = \frac{1}{N} \sum_{n=1}^N \log p(\mathbf{x}[n]|\mathbf{A}) \quad (2.3)$$

and, by the law of large numbers, we have that:

$$\lim_{N \rightarrow \infty} L_N(\mathbf{A}|\mathbf{x}) = L(\mathbf{A}|\mathbf{x}) \triangleq E[L_N(\mathbf{A}|\mathbf{x})] = \int p(\mathbf{x}) \log p(\mathbf{x} | \mathbf{A}) d\mathbf{x} \quad (2.4)$$

and setting $p(\mathbf{x}|\mathbf{A}) = \left[\frac{p(\mathbf{x}|\mathbf{A})}{p(\mathbf{x})} \right] p(\mathbf{x})$ in the equation above yields [22]:

$$L(\mathbf{A}|\mathbf{x}) = -K(p(\mathbf{x}|\mathbf{A}) || p(\mathbf{x})) - H(p(\mathbf{x})) \quad (2.5)$$

where K denotes Kullback-Leibler (KL) divergence and H means differential Shannon entropy¹. The ML estimate of the mixing matrix is then obtained by maximizing $L(\mathbf{A}|\mathbf{x})$:

$$\hat{\mathbf{A}}_{ML} = \arg \max_{\mathbf{A}} L(\hat{\mathbf{A}}|\mathbf{X}) = \arg \min_{\mathbf{A}} K(p(\mathbf{x}|\hat{\mathbf{A}}) || p(\mathbf{x})) \quad (2.6)$$

Note that the term $H(p(\mathbf{x}))$ in (2.5) was discarded because it does not depend on the parameter \mathbf{A} . So we finally obtain that the BSS contrast associated with the ML estimator is [22]:

$$\begin{aligned} \phi_{ML}(\hat{\mathbf{A}}) &= K(p(\mathbf{x}|\hat{\mathbf{A}}) || p(\mathbf{x})) = K(p(\hat{\mathbf{A}}^{-1}\mathbf{x}) || p(\mathbf{A}^{-1}\mathbf{x})) \\ &= K(p(\hat{\mathbf{s}}) || p(\mathbf{s})) \end{aligned} \quad (2.7)$$

¹See appendix A for a summary of the information theoretic concepts and properties used in this thesis.

where $p(\hat{\mathbf{s}})$ is the joint pdf of the estimated sources $\hat{\mathbf{s}} = \hat{\mathbf{A}}^{-1}\mathbf{x}$. The KL divergence is a (non-symmetric) measure of the difference between two probability distributions and, therefore, optimizing contrast ϕ_{ML} can be intuitively understood as finding the matrix $\hat{\mathbf{A}}$ that makes the pdf of the estimated sources as close as possible to the distribution of the true sources. A fundamental limitation is that, if \mathbf{s} is normally distributed, any rotation of the true sources minimizes the ML contrast:

$$\phi_{ML}(\mathbf{A}) = \phi_{ML}(\mathbf{A}\mathbf{R}) = 0 \quad \text{if} \quad \begin{cases} \mathbf{R}\mathbf{R}^T = \mathbf{R}^T\mathbf{R} = \mathbf{I} \\ \mathbf{s} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \end{cases} \quad (2.8)$$

which means that i.i.d. Gaussian sources can be recovered only up to an arbitrary unitary matrix \mathbf{R} . One such arbitrarily rotated version of the source estimates that minimizes the ML contrast is $\hat{\mathbf{s}} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{x}$ with $\boldsymbol{\Sigma}_{\mathbf{x}} = E[\mathbf{x}\mathbf{x}^T]$. Thus, for the sources to be uniquely determined, at most one of them can be Gaussian distributed. In the following, we will assume that this is the case.

2.2.1 The mutual information contrast

For i.i.d. sources, the basic premise of mutual independence means that $p(\mathbf{s}) = \prod_i p(s_i)$ and we can rewrite the ML contrast as:

$$\begin{aligned} \phi_{ML}(\hat{\mathbf{A}}) &= K(p(\hat{\mathbf{s}})||p(\mathbf{s})) \\ &= K(p(\hat{\mathbf{s}})||\prod_i p(\hat{s}_i)) + K(\prod_i p(\hat{s}_i)||\prod_i p(s_i)) \end{aligned} \quad (2.9)$$

A problem when trying to minimize $\phi_{ML}(\hat{\mathbf{A}})$ is that the second term in the right side of (2.9) depends on the true distribution of the sources, which is unknown. The technically simplest solution is to assume *a priori* a plausible distribution. This is the approach taken by *Infomax* [12], which is a BSS algorithm widely used among the neuroscientific community (see e.g. [46, 135–137]). A natural extension of Infomax consists in using a parametric pdf to model the distribution of the sources [123]. A more general approach does not assume any distribution for the sources but minimizes the ML contrast by optimizing not only over $\hat{\mathbf{A}}$ but also with respect to $p(\mathbf{s})$. For any given estimate of the mixing matrix, the distribution minimizing $\phi_{ML}(\hat{\mathbf{A}})$ is $p(\mathbf{s}) = \prod_i p(\hat{s}_i)$, which leads to the following BSS contrast [25]:

$$\phi_{MI}(\hat{\mathbf{A}}) \triangleq \min_{p(\mathbf{s})} \phi_{ML}(\hat{\mathbf{A}}) = K \left(p(\hat{\mathbf{s}}) \prod_i p(\hat{s}_i) \right) = I(\hat{\mathbf{s}}) \quad (2.10)$$

where I denotes *mutual information* (MI). Minimizing $\phi_{MI}(\hat{\mathbf{A}})$ is equivalent to finding the independent component analysis (ICA) [34, 92, 103] projection of the observed mixtures. This close connection between the linear and instantaneous BSS problem and ICA explains why some neuroscientists wrongly consider BSS and ICA to be equivalent terms. However, ICA is the solution to the BSS problem only when the sources are non-Gaussian, mutually independent and i.i.d., at least according to the definition of ICA given in [34]. As will be discussed later in this chapter, other source models lead to BSS solutions different from ICA. A remarkable algorithm based on contrast $\phi_{MI}(\hat{\mathbf{A}})$ is *MILCA* [210], which uses an MI estimator based on nearest-neighbors statistics [115].

2.2.2 The marginal entropy contrast

Unfortunately, estimating MI on the basis of a finite sample is difficult because it involves learning a multidimensional pdf. Thus, most ICA algorithms follow an indirect route to MI minimization, which is based on expressing MI as:

$$I(\hat{\mathbf{s}}) = \sum_{i=1}^M H(\hat{s}_i) - H(\hat{\mathbf{s}}) = \sum_{i=1}^M H(\hat{s}_i) - \log |\hat{\mathbf{A}}^{-1}| - H(\mathbf{x}) \quad (2.11)$$

Then, since $H(\mathbf{x})$ is constant with respect to $\hat{\mathbf{A}}$, the MI objective function is reduced to:

$$\phi_{MI}^*(\hat{\mathbf{A}}) = \sum_{i=1}^M H(\hat{s}_i) - \log |\hat{\mathbf{A}}^{-1}| \quad (2.12)$$

which involves only univariate densities that can be accurately and efficiently estimated using kernel methods and the fast Fourier transform [101, 203], as is done by algorithm *NpICA* [15]. Similarly, the algorithm RADICAL [122] uses another non-parametric estimator of entropy for univariate distributions due to Vasicek [221]. An alternative to non-parametric methods are approximations

of entropy based on the assumption that the pdf of the sources *is not very far from a Gaussian distribution*. Two well-known ICA algorithms that use different types of such approximations are FastICA [89] and JADE [25].

Contrast $\phi_{MI}^*(\hat{\mathbf{A}})$ can still be further simplified under the constraint that the estimated sources are spatially white. This constrained can be enforced by sphering the observed mixtures through the transformation $\mathbf{z} = \Sigma_{\mathbf{x}}^{-1/2}\mathbf{x}$. Then, the mixing matrix $\hat{\mathbf{A}}$ minimizing contrast (2.12) is $\hat{\mathbf{A}} = \Sigma_{\mathbf{x}}^{1/2}\hat{\mathbf{R}}^T$ where $\hat{\mathbf{R}} = [\hat{\mathbf{r}}_1, \dots, \hat{\mathbf{r}}_M]^T$ is the unitary matrix that minimizes the following contrast:

$$\phi_{ME}(\hat{\mathbf{R}}) = \sum_{i=1}^M H(\hat{\mathbf{r}}_i^T \mathbf{z}) \quad (2.13)$$

which qualitatively means that the ICA projection is that minimizing the marginal entropy of the estimated source signals. A major advantage of contrast $\phi_{ME}(\hat{\mathbf{R}})$ is that it involves optimization over the set of $M \times M$ orthogonal matrices, which is significantly easier than optimization over the set $\mathbb{R}^{M \times M}$. Indeed, most ICA algorithms discussed here are ultimately based on optimization of (2.13). Another appealing property of the marginal entropy contrast is that exhaustive search for the global minimum of $\phi_{ME}(\hat{\mathbf{R}})$ might be feasible, if the contrast can be evaluated efficiently. The reason is that the optimal M -dimensional rotation can be found by rotating only two dimensions at a time using what are known as Jacobi rotations (see Table 2.1). The only downside of using the orthogonal contrast (2.13) is that it imposes a lower bound on the asymptotic separation error. This is due to the blind trust put on the second order statistics that are used for whitening the source estimates [21, 23].

2.2.3 FastICA and EFICA

In this section we describe in more detail a variant of the popular FastICA algorithm, termed EFICA (efficient FastICA) [110], which is an essential building block of the BSS algorithms proposed in section 2.4.

FastICA is based on the marginal entropy contrast and approximates differential entropy by assuming that the pdf of the sources is not very far from the Gaussian distribution. This approximation takes the form [88]:

$$H(\hat{s}_i) \approx H(\nu) - \frac{1}{2} \sum_{k=1}^K (E[G_k(\hat{s}_i)])^2 \quad (2.14)$$

Algorithm:	Jacobi rotations
Input:	Sphered observations: $\mathbf{z}(n) \forall n = 1, \dots, N$. Initial source estimates: $\hat{\mathbf{s}} = [\hat{s}_1, \dots, \hat{s}_M]^T$ Initial estimation of rotation: $\hat{\mathbf{R}} = [\hat{\mathbf{r}}_1, \dots, \hat{\mathbf{r}}_M]^T$
Parameters:	J : number of rotation angles to evaluate. S: number of sweeps for Jacobi rotations
Procedure:	For each of S sweeps For each of $\frac{M(M-1)}{2}$ pairs of data dimensions (p, q) : i. Find rotation angle ϕ^* such that: $\phi^* = \arg \min_{\phi} (H(\hat{y}_p) + H(\hat{y}_q))$ with $\begin{bmatrix} \hat{y}_p \\ \hat{y}_q \end{bmatrix} = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \begin{bmatrix} \hat{s}_p \\ \hat{s}_q \end{bmatrix}$ ii. Update source estimates: $\hat{s}_p \leftarrow \cos(\phi^*)\hat{s}_p - \sin(\phi^*)\hat{s}_q$ $\hat{s}_q \leftarrow \sin(\phi^*)\hat{s}_p + \cos(\phi^*)\hat{s}_q$ iii. Update \mathbf{R} : $\hat{\mathbf{r}}_p \leftarrow \cos(\phi^*)\hat{\mathbf{r}}_p - \sin(\phi^*)\hat{\mathbf{r}}_q$ $\hat{\mathbf{r}}_q \leftarrow \sin(\phi^*)\hat{\mathbf{r}}_p + \cos(\phi^*)\hat{\mathbf{r}}_q$
Output:	$\hat{\mathbf{R}}$

Table 2.1: Exhaustive search of the optimum M-dimensional rotation through elementary 2-dimensional rotations

where $H(\nu) = \frac{1}{2}(1 + \log(2\pi))$ is the entropy of the normal density and $\{G_k\}$ is a set of non-linear functions. FastICA uses the simplest version of this entropy estimator, which involves only one non-linearity (i.e. $K = 1$). If the density of the sources $p(s_i)$ would be known, the optimal choice for this nonlinearity would be $-\log p(s_i)$. However, since the density of the sources is unknown, FastICA requires the user to choose among a set of nonlinearities that fit well some important and well-known densities [89]:

$$\begin{aligned} G_a(u) &= \frac{1}{a_1} \log \cosh(a_1 u) \\ G_b(u) &= -\frac{1}{a_2} \exp\left(-\frac{a_2 u^2}{2}\right) \\ G_c(u) &= \frac{1}{4} u^4 \end{aligned} \quad (2.15)$$

where $1 \leq a_1 \leq 2$ and $a_2 \approx 1$ are constants. G_a is suitable for distributions close to the exponential power family, G_b may be better for highly super-Gaussian distributions with outliers and G_c is recommended for estimating sub-Gaussian independent components.

EFICA generalizes FastICA by using different non-linearities for estimating the marginal entropy of each of the M sources. These M non-linearities are chosen in a data-adaptive fashion so that EFICA's residual error variance attains the Cramér-Rao lower bound (CRLB), when the distribution of the sources belongs to the class of generalized Gaussian distributions.

A common measure for evaluating separation accuracy is the interference-to-signal ratio (ISR) [23]. For a given estimate of the separating matrix $\hat{\mathbf{B}} = \hat{\mathbf{A}}^{-1}$, the corresponding ISR matrix is given (elementwise) by $\Psi_{kl} = \mathbf{G}_{kl}^2 / \mathbf{G}_{kk}^2$, where $\mathbf{G} = \hat{\mathbf{B}}\mathbf{A}$. Under the assumption that the sources are i.i.d. non-Gaussian random variables, the asymptotic ISR matrix for EFICA has the form [110]:

$$E[\Psi_{kl}] = \frac{1}{N} \frac{\gamma_k(\gamma_l + \tau_l^2)}{\tau_l^2 \gamma_k + \tau_k^2(\gamma_l + \tau_l^2)} \quad (2.16)$$

with

$$\begin{aligned} \gamma_k &= E[G_k^2(s_k)] - E^2[s_k G_k(s_k)] \\ \tau_k &= |E[s_k G'_k(s_k)] - E[G'_k(s_k)]| \end{aligned} \quad (2.17)$$

where $E[\cdot]$ denotes expectation, G_k is the non-linearity used with the k th source estimate and G'_k denotes the derivative of G_k . A useful property of

EFICA is that consistent estimates of the asymptotic ISR matrix can be obtained by replacing in (2.16) the true sources with the estimated sources and expectations with sample means.

2.2.4 Optimization and reliability of ICA contrasts

Most ICA algorithms use local optimization techniques that require computing the gradient of the contrast function [12, 15, 89, 93, 110, 175]. Local optimization is appealing for computational reasons and because, asymptotically, the ICA contrast has a unique minimum that corresponds to the BSS solution, if the sources are truly independent, i.i.d., and their distribution is *close to Gaussian* [16]. However, ICA contrasts are likely to have (potentially many) spurious minima if at least one of these conditions is fulfilled [66, 189, 225]:

1. The sample size is too small to consider asymptotic behavior.
2. The sources are not perfectly independent.
3. The sources are not i.i.d., i.e. they have temporal structure.

Spurious local minima can be avoided by global optimization techniques based on Jacobi rotations [25, 69, 122, 210]. However, the global minimum might also be spurious, invariably leading ICA to meaningless *overfitted* source estimates [66, 189]. This is a crucial issue in the analysis of EEG data since brain sources are likely to exchange information through time-delayed axonal pathways and perfect independence is probably unrealistic. One may argue that, if EEG sources were truly independent, functional connectivity studies using EEG would make no sense altogether.

Second-order temporal structure and time-lagged cross-dependencies can be accounted for by modeling the sources as a vector autoregressive (VAR) process:

$$\mathbf{s}(n) = \sum_{\tau=1}^p \mathbf{B}_s(\tau) \mathbf{s}(n - \tau) + \mathbf{z}(n) \quad (2.18)$$

where $\mathbf{B}_s(\tau) \forall \tau = 1, \dots, p$ are the coefficient matrices of the VAR process and $\mathbf{z}(n) = [z_1(n), \dots, z_M(n)]^T$ is the non-Gaussian i.i.d. *innovation* of the model².

²Note that we require the innovation process to be non-Gaussian, contrary to the standard definition of VAR models. This has implications on the estimation of the parameters of the

In this thesis we consider mutually independent brain sources in the sense that the 1-dimensional components of the innovation process in (2.18) behave like mutually independent random variables. Then, the observed mixtures can be written as:

$$\mathbf{x}(n) = \sum_{\tau=1}^p \underbrace{\mathbf{A}\mathbf{B}_s(\tau)\mathbf{A}^{-1}}_{\mathbf{B}_x(\tau)} \mathbf{x}(n-\tau) + \underbrace{\mathbf{A}\mathbf{z}(n)}_{\mathbf{w}(n)} \quad (2.19)$$

Obviously, if model (2.18) is a valid approximation of reality, a straightforward solution to minimize mutual dependencies and temporal structure is to pre-process the observed mixtures with the following VAR filter [62]:

$$\hat{\mathbf{w}}(n) = \mathbf{x}(n) - \sum_{\tau=1}^p \hat{\mathbf{B}}_x(\tau) \mathbf{x}(n-\tau) \approx \mathbf{A}\mathbf{z}(n) \quad (2.20)$$

where $\hat{\mathbf{B}}_x(\tau) \forall \tau = 1, \dots, p$ are the coefficients of a VAR model fitted to the observed mixtures using least squares (LS) [193] or any other VAR estimation algorithm [178]. VAR filtering is an effective approach for removing second order temporal structure (including cross-correlations), even when the VAR model is valid just approximately, as will be shown in section 2.6.

In order to illustrate finite-sample effects on ICA contrasts, we generate two mixed signals by rotating (by an angle $\phi_s = 60^\circ$) two i.i.d. and mutually independent Laplacian sources. Fig. 2.1 shows the shape of FastICA's and RADICAL's contrasts for different rotations of the observed mixtures. When the number of data samples is very large both contrasts are convex and have a single minimum at the right solution ($\phi_x = 30^\circ$). Notice that by rotating the original sources first by 60° and then by 30° , the source estimates are just 90° -rotated versions of the original sources:

$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \begin{bmatrix} \cos(90^\circ) & -\sin(90^\circ) \\ \sin(90^\circ) & \cos(90^\circ) \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} -s_2 \\ s_1 \end{bmatrix} \quad (2.21)$$

i.e. the estimates are just a permuted and scaled version of the true sources, which is the desirable result. However, if we now repeat the experiment for

model from its observations since most estimation methods assume Gaussian innovations. However, our numerical experiments in chapter 4 suggest that classical VAR estimators like ARFit [193] work well, even when the innovation process is non-Gaussian.

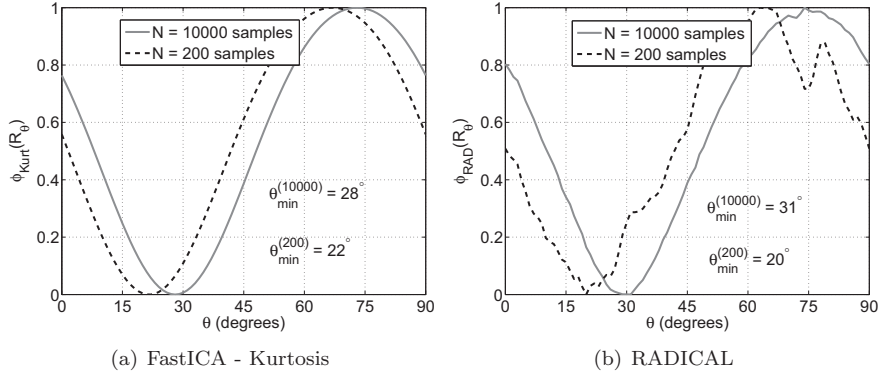


Figure 2.1: Finite-sample effects on the ICA contrasts of FastICA and RADICAL. Two mixtures were generated by rotating 60° two i.i.d. and mutually independent Laplacian sources ($N = 10000$). The figures show the values of the ICA contrast for different rotations of the observed mixtures. Perfect separation was achieved with $\phi = 30^\circ$. Note that with just 200 samples, the global minimum of the two contrasts is spurious.

just $N = 200$ samples, the global minimum of both RADICAL's and FastICA's contrast is spurious. Moreover, RADICAL's contrast is not convex anymore and contains several local minima (see Fig. 2.1). Similar experiments can be performed to illustrate the negative effects of temporal structure in the sources (Fig. 2.2) and of cross-dependencies between sources (Fig. 2.3). Although these toy experiments do not allow drawing general conclusions, they lead to the following intuitions:

- Limited data and cross-dependencies between sources are most problematic because they often lead to overfitting, i.e. to a spurious global minimum in ICA contrasts.
- Non-parametric estimators of marginal entropy are more affected by local minima than semi-parametric approaches. However, the global minimum of the latter might be less reliable, especially for sources with large scale temporal correlations.
- VAR estimation techniques that assume Gaussian innovations probably produce valid estimates, even when the innovations are not Gaussian.

Further numerical assessment of these preliminary findings and of the advantages of VAR pre-processing can be found in section 2.6.

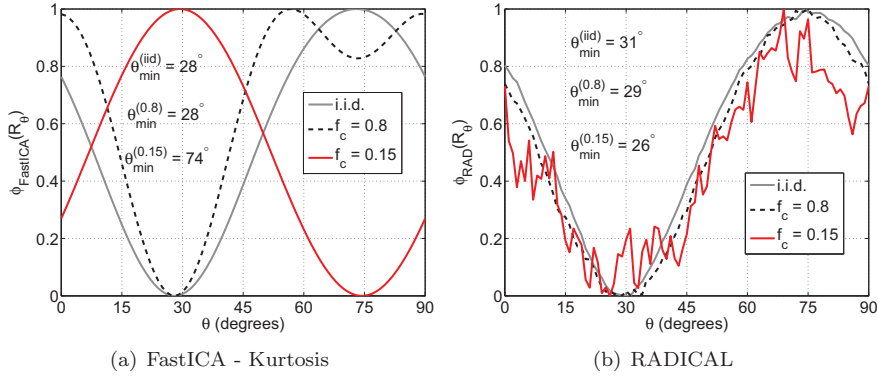


Figure 2.2: Values of the contrasts of FastICA and RADICAL for different rotations of two mixed non-white sources. The sources were obtained by filtering two mutually independent i.i.d. Laplacian noise processes ($N = 10000$) with a low-pass FIR filter with normalized cutoff $f_c = 0.8$ (dashed black line) and $f_c = 0.15$ (solid red line). The mixtures were obtained by rotating 60° the sources so that perfect separation was achieved with $\phi = 30^\circ$. Notice how low-pass filtering introduces many spurious local minima in RADICAL's contrast but global optimization would nevertheless lead to acceptable source estimates. On the other hand, FastICA is likely to be less robust to source autocorrelations, due to convergence to local minima or to a spurious global minimum when low frequencies dominate.

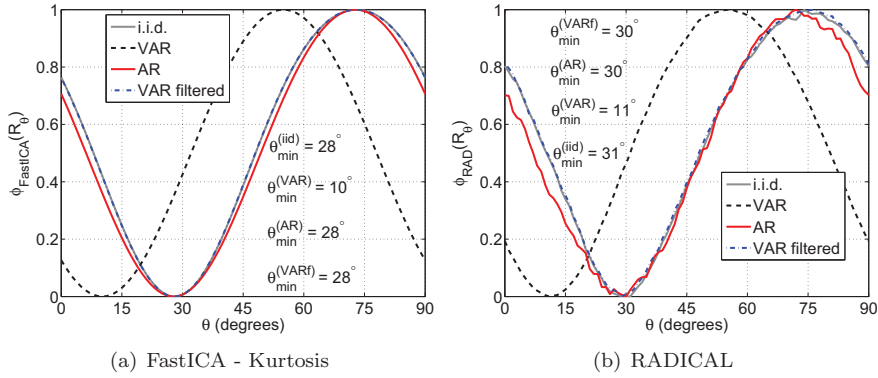


Figure 2.3: Effects of cross-dependencies between sources on the contrasts of FastICA and RADICAL. The mixtures were obtained by rotating 60° the sources so that perfect separation was achieved with $\phi = 30^\circ$. Three types of sources were used to generate the figures: (i) i.i.d. Laplacian sources (denoted as "i.i.d." in the legend), (ii) VAR sources with i.i.d. Laplacian innovations (denoted "VAR"), (iii) AR sources with i.i.d. Laplacian innovations (denoted "AR"). The blue dash-dotted line shows the values of the contrast when the mixtures of VAR sources are pre-processed with a VAR filter. The coefficients of the VAR filter were estimated using ARfit [193].

2.3 BSS of spectrally diverse sources

BSS can also be approached by considering mutually uncorrelated sources with second-order autocorrelations. This means that the time-lagged covariance of the sources is:

$$\mathbf{\Sigma}_s(\tau) = E [\mathbf{s}(n)\mathbf{s}^T(n - \tau)] = \mathbf{R}_s(\tau) \quad (2.22)$$

where $\mathbf{R}_s(\tau)$ is a diagonal matrix having as diagonal entries the autocorrelation of each source at time-lag τ . Thus, the time-lagged covariance matrices of the observed mixtures satisfy ³:

$$\mathbf{\Sigma}_x(\tau) = E [\mathbf{A}\mathbf{s}(n)\mathbf{s}^T(n - \tau)\mathbf{A}^T] = \mathbf{A}\mathbf{R}_s(\tau)\mathbf{A}^T \quad (2.23)$$

which clearly shows that \mathbf{A} alone explains cross-correlations of the mixtures at any lag. In this case, solving the BSS problem requires finding at least two lags for which the source signals have distinct autocorrelations so that the following system of linear equations has a unique solution [147, 168]:

$$\left. \begin{aligned} \mathbf{\Sigma}_x(\tau_1) &= \mathbf{A}\mathbf{R}_s(\tau_1)\mathbf{A}^T \\ \mathbf{\Sigma}_x(\tau_2) &= \mathbf{A}\mathbf{R}_s(\tau_2)\mathbf{A}^T \end{aligned} \right\} \implies \mathbf{\Sigma}_x(\tau_1)\mathbf{A}^{-1} = \mathbf{\Sigma}_x(\tau_2)\mathbf{A}^{-1}\mathbf{\Lambda} \quad (2.24)$$

where $\mathbf{\Lambda} = \mathbf{R}_s(\tau_1)\mathbf{R}_s^{-1}(\tau_2)$ is a diagonal matrix. Eq. (2.24) is a generalized eigenvalue problem which determines uniquely (up to permutation and scaling of its columns) the separating matrix \mathbf{A}^{-1} as long as the corresponding eigenvalues are all distinct. Note that this is equivalent to requiring that the ratios between the diagonal elements of the two time-lagged source covariance matrices are all different: $\mathbf{R}_s^{(ii)}(\tau_1)/\mathbf{R}_s^{(ii)}(\tau_2) \neq \mathbf{R}_s^{(jj)}(\tau_1)/\mathbf{R}_s^{(jj)}(\tau_2) \ \forall i \neq j$. As noted by Parra and Sajda [168], blind source separation of non-Gaussian and non-stationary sources can also be formulated as a generalized eigenvalue decomposition by replacing $\mathbf{R}_s(\tau_1)$ and $\mathbf{R}_s(\tau_2)$ with matrices that have the same properties upon those alternative source models.

In practice, blind separation using just two covariance matrices is not robust due to errors in the estimation of covariances from a limited amount

³In this thesis we consider only real-valued variables.

of noisy observations. A considerably more robust approach is based on *approximate*⁴ joint diagonalization (AJD) of many estimated time-lagged covariance matrices $\hat{\mathbf{R}}_{\mathbf{s}}(\tau_1), \hat{\mathbf{R}}_{\mathbf{s}}(\tau_2), \dots, \hat{\mathbf{R}}_{\mathbf{s}}(\tau_n)$. BSS algorithms SOBI [13] and TDSEP [233] diagonalize these matrices using successive Jacobi rotations [27]. WASOBI [228] formulates AJD as a weighted least squares problem [216, 229] with weights that are asymptotically optimal for the case of Gaussian sources. Namely, if all sources are Gaussian autoregressive (AR) processes of order p , then, under asymptotic conditions, the ISR matrix attained by WASOBI can be shown to reach the CRLB [47]:

$$E[\Psi_{kl}] = \frac{1}{N} \frac{\gamma_{kl}}{\gamma_{kl}\gamma_{lk} - 1} \frac{\sigma_k^2 R_l[0]}{\sigma_l R_k[0]} \quad (2.25)$$

where $R_k[m]$ denotes the autocorrelation of the k th source at lag m , σ_k^2 is the variance of the innovations of the k th source and γ_{kl} are given by:

$$\gamma_{kl} = \frac{1}{\sigma_k^2} \sum_{i,j=0}^p a_{il} a_{jl} R_k[i-j] \quad (2.26)$$

where $\{a_{il}\}_{i=0}^p$ are the AR coefficients of the l th source with $a_{0l} = 1$ for $k, l = 1, \dots, M$. As a final remark, we stress the fact that the autocorrelation sequence (i.e. the power spectrum) of the sources need to be different for separation to be feasible using only second-order statistics.

2.4 Hybrid BSS algorithms

In a mixture of non-Gaussian i.i.d. sources with Gaussian non-white sources, ICA and SOBI-like algorithms will be (at best) able to separate just some of the sources but never all of them. A compromise solution to this problem is to try to optimize a weighted average of both types of BSS contrasts. Several algorithms have been proposed in this direction, including JADE_{TD} [154], JCC [72], ThinICA [36] and the unifying model of [91]. An alternative approach is to successively use different contrasts to separate the sub-sets of sources for which they are more suitable [64]. The BSS algorithms EFWS [214], COMBI [214], M-COMBI [215] and F-COMBI [63] implement this idea by combining the strengths of EFICA and WASOBI.

⁴While the set of true covariance matrices admits exact diagonalization, it is almost surely impossible to jointly diagonalize the set of estimated matrices.

Recall that, given a realization of the sources, a common way of evaluating the accuracy of the separation produced by any BSS algorithm is the *realization*⁵ ISR matrix, which is defined (element-wise) as $\Psi_{kl} = \mathbf{G}_{kl}^2 / \mathbf{G}_{kk}^2$ where $\mathbf{G} = \hat{\mathbf{B}}\mathbf{A}$. $\hat{\mathbf{B}}$ is the estimated separating matrix and \mathbf{A} is the true mixing matrix. Ψ_{kl} measures the level of residual interference between the k th and l th estimated components. The total ISR of the k th estimated source is defined as $\psi_k = \sum_{l=1, l \neq k}^M \Psi_{kl}$. If the realization ISR matrix obtained by WASOBI and EFICA would be known, a straightforward combination of the separation abilities of the two algorithms would consist in selecting, for each source, the reconstructed version (either EFICA's or WASOBI's) with better ISR. This simple principle can then be extended into a successive scheme such that, in each interaction, only the best separated sources are accepted, and the remaining ones are subjected to an additional iteration of separation and selection.

Although the realization ISR matrices of WASOBI (Ψ^{WA}) and EFICA (Ψ^{EF}) are obviously unknown (nor can they be consistently estimated from the data), it is possible to approximate them with the (asymptotically) expected ISR matrices, thereby producing a selection strategy which performs well "on average". Consistent estimates of $\bar{\Psi}^{\text{EF}} = E[\Psi^{\text{EF}}]$ and $\bar{\Psi}^{\text{WA}} = E[\Psi^{\text{WA}}]$ can be obtained by replacing in (2.16) and (2.25) the true sources with the estimated sources and expectations with sample means. Then, the total ISR for each source obtained by each algorithm can be approximated by $\psi_l^{\text{WA}} \approx \bar{\psi}_l^{\text{WA}} = \sum_{l=1, l \neq k}^M \bar{\Psi}_{kl}^{\text{WA}}$ and $\psi^{\text{EF}} \approx \bar{\psi}^{\text{EF}} = \sum_{l=1, l \neq k}^M \bar{\Psi}_{kl}^{\text{EF}}$.

2.4.1 COMBI

Algorithm COMBI [214] uses consistent estimates of $\bar{\psi}^{\text{WA}}$ and $\bar{\psi}^{\text{EF}}$ in the following procedure:

1. Let \mathbf{x} be the observed mixtures and let $\mathbf{z} = \mathbf{x}$.
2. Apply both EFICA and WASOBI to \mathbf{z} and denote the estimated sources as $\hat{\mathbf{s}}^{\text{EF}}$ and $\hat{\mathbf{s}}^{\text{WA}}$, respectively. Approximate the corresponding ISR for all estimated sources as $\psi^{\text{EF}} \approx \bar{\psi}^{\text{EF}}$ and $\psi^{\text{WA}} \approx \bar{\psi}^{\text{WA}}$.
3. Let $E = \min_k \psi_k^{\text{EF}}$ and $W = \min_k \psi_k^{\text{WA}}$.
4. **If** $E < W$ **then:**

⁵We use the term *realization* ISR (denoted Ψ) to refer to the ISR matrix obtained by a single run of a BSS algorithm on a specific realization of the mixtures. This is to differentiate it from the (asymptotically) *expected* value of the ISR matrix (denoted $\bar{\Psi} = E[\Psi]$).

- Accept source estimates $\{\hat{s}_k^{EF} \mid \psi_k^{EF} < W\}$ and redefine \mathbf{z} as the set of rejected source estimates $\{\hat{s}_k^{EF} \mid \psi_k^{EF} \geq W\}$

Otherwise

- Accept source estimates $\{\hat{s}_k^{WA} \mid \psi_k^{WA} < E\}$ and redefine \mathbf{z} as the set of rejected source estimates $\{\hat{s}_k^{WA} \mid \psi_k^{WA} \geq E\}$
5. If \mathbf{z} contains more than one rejected source estimate, go to step (2). Otherwise, accept the rejected source estimate (if there is any) and terminate.

The asymptotic ISR expressions in (2.16) and (2.25) were derived under the assumption that all sources comply with their respective source models, which raises concerns on their reliability in hybrid scenarios of non-Gaussian and non-i.i.d. sources. However, it was empirically verified in [215] that those expressions remain reasonably accurate, even when their respective model assumptions are mildly violated. Moreover, exact ISR values are of little interest for COMBI, since only their comparative relations are used during the selection of the best source estimates. Such selection could also be performed using other empirical estimates of the realization ISR, such as those based on bootstrap resampling [64, 144]. However, bootstrap techniques are computationally expensive and their application to sources with time structure is far from trivial.

2.4.2 M-COMBI

A fundamental limitation of COMBI is that, at each selection step, it accepts or rejects *individual* source estimates. Consider the hypothetical case that EFICA (resp. WASOBI) would be able to attain a nearly block-diagonal ISR matrix, effectively separating the mixtures into *groups* of sources with poor intra-group separation but minimal inter-group interferences. A subsequent application of WASOBI (resp. EFICA) to each of these groups separately may be able to attain good separation of all sources, but COMBI is unaware of this possibility. In this section we present an extended version of COMBI, termed M-COMBI [215], that is able to account for *multidimensional* source estimates.

Standard BSS assumes that the one-dimensional unknown sources are mutually independent according to the independency contrast used (e.g. statistical independence or time-lagged uncorrelation). A straightforward generalization of this principle assumes that not all the M sources are mutually independent but they form D higher dimensional independent compo-

nents [24, 43]. Let d_l denote the dimensionality of the l th multidimensional component that groups together the one-dimensional source signals with indexes $\Gamma_l \equiv \{l_1, \dots, l_{d_l}\}$. Then, the l th multidimensional component is given by $\mathbf{s}_l(n) = [s_{l_1}(n), \dots, s_{l_{d_l}}(n)]^T$, where $l = 1, \dots, D$ and $d_1 + d_2 + \dots + d_D = M$. Therefore, we can rewrite the sources in (2.2) as $\mathbf{s}(n) = [\mathbf{s}_1^T(n), \dots, \mathbf{s}_D^T(n)]^T = \mathbf{P} [s_1(n), \dots, s_M(n)]^T$ where \mathbf{P} is a permutation matrix. Using the notation above and dropping matrix \mathbf{P} under the permutation indeterminacy of ICA, we can reformulate (2.2) as:

$$\mathbf{s}(n) = \mathbf{A}^{-1} \mathbf{x}(n) = [\mathbf{x}^T(n) \mathbf{B}_1^T, \dots, \mathbf{x}^T(n) \mathbf{B}_D^T]^T = [\mathbf{s}_1^T(n), \dots, \mathbf{s}_D^T(n)]^T \quad (2.27)$$

Then the goal of multidimensional BSS is to estimate the set of matrices $\{\mathbf{B}_l\}_{l=1, \dots, D}$, each of which is of dimension $d_l \times M$. Since the sub-components of a multidimensional independent component are arbitrarily mixed we can recover $\{\mathbf{B}_l\}_{l=1, \dots, D}$ only up to an invertible matrix factor [24].

A multidimensional component according to certain independency contrast (e.g. non-Gaussianity) might be separable into one-dimensional components using an alternative independency measure (e.g. time-lagged cross-correlations). This suggests a procedure for combining complementary independency criteria:

1. Try BSS using certain independency-based contrast.
2. Detect the presence of multidimensional components in the source signals estimated in step (1).
3. Try BSS using an alternative BSS contrast in each multidimensional component found in step (2).

This is the idea underlying M-COMBI [215] and F-COMBI [67], which combine the complementary strengths of the non-Gaussianity contrast of EFICA and the criterion based on cross-correlations of WASOBI. For notational convenience, we identify a multidimensional source (or a *cluster* of sources) Γ either by the set of indices of the sources that it contains, denoted by $\Gamma \equiv \{l_1, l_2, \dots, l_d\} \subset \{1, 2, \dots, M\}$ or by the corresponding sources $\Gamma \equiv [s_{l_1}, \dots, s_{l_d}]^T \subset \mathbf{s}$. Given the same approximations of the ISR matrix used by COMBI, the total ISR of cluster Γ can be defined as:

$$\psi(\Gamma) = \alpha_\Gamma \sum_{k \in \Gamma, l \notin \Gamma} \bar{\Psi}_{kl} \quad (2.28)$$

where α_Γ is a normalization coefficient that depends on the cluster's cardinality (denoted as $|\Gamma|$) and on the total number of sources M . In [215] it was proposed to use:

$$\alpha_\Gamma = \frac{d-1}{|\Gamma|(d-|\Gamma|)} \quad (2.29)$$

so that $\psi(\Gamma)$ is $d-1$ times the average of the entries of matrix $\bar{\Psi}$ that are included in the sum in (2.28). Note that this definition of $\psi(\Gamma)$ is compatible with the original definition for one-dimensional sources: $\psi(\{k\}) = \psi_k = \sum_{l=1, l \neq k}^M \bar{\Psi}_{kl}$.

M-COMBI works iteratively with a stack of clusters \mathcal{C} . In each iteration, one of the clusters in the stack that is not a *singleton*⁶ is decomposed into two or more smaller clusters, until all clusters are singletons. The algorithm steps are summarized below:

1. Initialize the stack of clusters with just one cluster that contains all the observed mixtures, i.e. $\mathcal{S} = \{\Gamma_0^{(S)}\}$ with $\Gamma_0^{(S)} = \{1, 2, \dots, M\}$. Initialize the source estimates $\hat{\mathbf{s}} = \mathbf{x}$.
2. Pick any cluster in \mathcal{S} , say Γ^* , that is not a singleton and denote the corresponding source estimates as \mathbf{z} . Obviously $\mathbf{z} = \mathbf{x}$ in the first iteration.
3. Apply EFICA and WASOBI to \mathbf{z} in order to obtain the corresponding source estimates $\hat{\mathbf{s}}^{EF}$ and $\hat{\mathbf{s}}^{WA}$ as well as the estimated asymptotic ISR matrices $\bar{\Psi}^{EF}$ and $\bar{\Psi}^{WA}$.
4. Build the set \mathcal{C} of possible sub-clusters within Γ^* : $\mathcal{C} = \{\Gamma_1^{(C)}, \dots, \Gamma_K^{(C)}\}$ with $\Gamma_i^{(C)} \subset \Gamma^* \forall i = 1, \dots, K$. For example, if $\Gamma^* \equiv \{2, 5, 8\}$ then $\mathcal{C} = \{\{2\}, \{5\}, \{8\}, \{2, 5\}, \{2, 8\}, \{5, 8\}\}$.
5. Based on the estimated asymptotic ISR matrices $\bar{\Psi}^{EF}$ and $\bar{\Psi}^{WA}$ obtain the corresponding approximations for the cluster ISR: $\psi^{EF}(\Gamma_i^{(C)})$ and $\psi^{WA}(\Gamma_i^{(C)})$ for all $i = 1, \dots, K$.
6. Let $E = \min_i \psi^{EF}(\Gamma_i^{(C)})$ and $W = \min_i \psi^{WA}(\Gamma_i^{(C)})$.
7. **If** $E < W$, define the set of "best" separated clusters as follows:

$$\mathcal{C}^{BEST} = \mathcal{C}^{EF} = \{\Gamma_1^{EF}, \Gamma_2^{EF}, \dots, \Gamma_R^{EF}\} \quad (2.30)$$

⁶We denote by *singleton* a cluster that contains a single 1-dimensional source estimate.

with

$$\Gamma_1^{EF} = \arg \min_{\Gamma \in \mathcal{C}} \psi^{EF}(\Gamma) \quad (2.31)$$

and, for $k = 1, 2, \dots, R - 1$:

$$\Gamma_{k+1}^{EF} = \arg \min_{\Gamma \in \mathcal{T}(k)} \psi^{EF}(\Gamma) \quad (2.32)$$

with $\mathcal{T}(k) = \mathcal{C} - \{\Gamma \mid \exists l \in [1, k] \mid \Gamma_l \cap \Gamma \neq \emptyset\}$, and with R the value such that $\psi^{EF}(\Gamma_R) > W$ or $\mathcal{T}(R - 1)$ is empty. This procedure picks up the best clusters in a *greedy* scheme by selecting at each step the best remaining cluster in \mathcal{C} , among those disjoint with the clusters that have been already picked.

Else (for $E \geq W$) define the set of best separated clusters as $\mathcal{C}^{BEST} = \mathcal{C}^{WA}$, where \mathcal{C}^{WA} is obtained in an analogous manner to \mathcal{C}^{EF} .

8. Update \mathcal{S} by $\mathcal{S} = (\mathcal{S} - \Gamma^*) \cup \mathcal{C}^{BEST}$
9. If all clusters in \mathcal{S} are already singletons terminate. Otherwise, return to step (2).

The total number of possible sub-clusters in step (4) is $2^{|\Gamma^*|} - 2$, which might be prohibitively large when cluster Γ^* contains many sources. For such high-dimensional cases, say for $|\Gamma^*| \geq 20$, we take an alternative approach based on hierarchical clustering [206]⁷. A symmetric similarity matrix between clusters is defined as $\mathbf{D} \triangleq \bar{\Psi} + \bar{\Psi}^T$ and the set \mathcal{C} is built recursively so that initially \mathcal{C} contains all singletons. At each subsequent step, we look for the couple (k, l) for which \mathbf{D}_{kl} obtains its maximum value and add to \mathcal{C} the cluster formed by the union of the most recently created cluster containing signal k and the most recently created cluster containing source estimate l . Then we make $\mathbf{D}_{kl} = \mathbf{D}_{lk} = 0$ so that the same couple of source estimates are not reused in the next iterations. The update of \mathcal{C} terminates after $|\Gamma^*| - 1$ steps and contains $2^{|\Gamma^*|} - 2$ entries at the end. Note that this procedure will usually lead to set \mathcal{C} being significantly smaller than the number of all possible sub-clusters $2^{|\Gamma^*|} - 2$. This clustering scheme is an *ad hoc* approach that could be replaced by more sophisticated algorithms in the future. However, in our simulations, the proposed procedure worked well and usually outperformed the spectral clustering method advocated in [144] in a similar context.

⁷If \mathbf{z} was obtained with EFICA, no clustering is actually required, as \mathcal{C} can be simply defined as the set of all singletons. This is because, in the case of EFICA, at most one cluster of unresolved (Gaussian) components may exist. By contrary, many clusters of spectrally identical sources can be found within the estimates produced by WASOBI.

2.4.3 F-COMBI

A practical limitation of M-COMBI [215] is its iterative nature, which may lead to relatively lengthy computations for high-dimensional problems. In this section we propose a simplified variant of M-COMBI with similar separation performance but significantly smaller computational load.

As M-COMBI, F-COMBI is also based on detecting the presence of clusters of sources with high mutual interference. A symmetric similarity matrix between clusters, defined as $\mathbf{D} \triangleq \bar{\Psi} + \bar{\Psi}^T$ is used to define a set of M partition levels of the estimated sources into different clusters by means of agglomerative hierarchical clustering with single linkage. By single linkage we mean that the similarity between clusters of sources is defined as the similarity between the closest pair of source estimates. The output of this clustering algorithm is a set of $i = 1, 2, \dots, M$ possible partition levels of the estimated sources. At each particular level the method joins together the two clusters from the previous level which are most similar. Therefore, at level $i = 1$ each source forms a cluster whereas at level $i = d$ all the sources belong to the same cluster. For assessing the goodness-of-fit of the $i = 2, \dots, M - 1$ partition levels we propose using the validity index $I_i = D_i^{intra} / D_i^{inter}$ where D_i^{intra} and D_i^{inter} roughly measure the average intra-cluster and inter-cluster similarities, respectively. They are defined, for $1 < i < M$, as follows:

$$D_i^{intra} = \frac{\sum_{j=1, |\Gamma_j^{(i)}| > 1}^{M-i+1} \sum_{k \in \Gamma_j^{(i)}, l \in \Gamma_j^{(i)}} \mathbf{D}_{kl}}{\sum_{j=1, |\Gamma_j^{(i)}| > 1}^{M-i+1} |\Gamma_j^{(i)}| \frac{(|\Gamma_j^{(i)}| - 1)}{2}} \quad (2.33)$$

$$D_i^{inter} = \frac{\sum_{j=1}^{M-i+1} \sum_{k \in \Gamma_{i,j}, l \notin \Gamma_j^{(i)}} \mathbf{D}_{kl}}{\sum_{j=1}^{M-i+1} |\Gamma_j^{(i)}| (M - |\Gamma_j^{(i)}|)} \quad (2.34)$$

where $\Gamma_j^{(i)}$ is the set of indices of the sources belonging to the j^{th} cluster at the i^{th} partition level. We also define $I_1 = 1 / (\max_{ij} \mathbf{D}_{ij})$ and $I_M = 10$. Finally we choose the best cluster partition to be that one corresponding to the maximum of the validity index I . By setting $I_M = 10$ we consider that the separation failed completely (there is just one M -dimensional cluster) if $D_i^{intra} < 10 \cdot D_i^{inter} \forall i = 2, \dots, M - 1$. The definition of I_1 implies that all estimated sources will be considered to be 1-dimensional (perfect separation) if $(\max_{ij} \mathbf{D}_{ij}) < \min_{i > 2} (1/I_i)$. Since we have set $I_M = 10$, a necessary condition for perfect separation is that the maximum ISR between any pair of 1-dimensional source

estimates is below -10 dB. Based on this heuristic index of cluster quality, F-COMBI uses a simple combination strategy:

1. Apply WASOBI to the observed mixtures.
2. If the maximum entry in the estimated ISR matrix is below a very small threshold (e.g. -30 dB) then all sources were accurately estimated and we are finished.
3. If the minimum value in the ISR matrix is above a high threshold (-10 dB) then WASOBI was not able to produce any separation of the observed data, i.e. $\mathcal{C}^{WA} = \{\Gamma_1\}$ with $\Gamma_1 = \{1, 2, \dots, M\}$. In this case we skip step (4) and go directly to step five.
4. Use hierarchical clustering and the quality index I defined above to determine the optimal partition of the sources estimated by WASOBI into $1 < R < M$ clusters, i.e. $\mathcal{C}^{WA} = \{\Gamma_1, \Gamma_2, \dots, \Gamma_R\}$.
5. Apply EFICA to each cluster in \mathcal{C}^{WA} .

The reason for using WASOBI first, instead of EFICA, is that the former is considerably faster than the latter for high dimensional mixtures, which is the target application of F-COMBI. Moreover, WASOBI is often able to break the original high-dimensional problem into more multidimensional clusters than EFICA (which, asymptotically, can produce at most one cluster of Gaussian components).

2.5 BSS by entropy rate minimization

As we saw in section 2.2, for sources that behave like i.i.d. non-Gaussian random variables, the linear and instantaneous BSS problem can be solved using ICA [34]. Recall that, if the observed mixtures have a covariance matrix $\Sigma_{\mathbf{x}} = E[\mathbf{x}\mathbf{x}^T]$, the ICA-based separating matrix is $\mathbf{B}_{opt} = \mathbf{R}_{opt}\Sigma_{\mathbf{x}}^{-1/2}$ where \mathbf{R}_{opt} is the $M \times M$ unitary matrix that minimizes:

$$\phi_{ME}(\mathbf{R}) = \sum_{i=1}^M H(\mathbf{r}_i^T \Sigma_{\mathbf{x}}^{-1/2} \mathbf{x}) = \sum_{i=1}^M H(\hat{s}_i) \quad (2.35)$$

where $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_M]^T$. But in most practical applications, and especially in the case of EEG, the sources are not i.i.d. and are better modeled as stochastic processes with (second and higher-order) temporal correlations. Let us

consider sources that behave like mutually independent stationary Markov processes of order d so that their temporal structure is confined within the vector $\mathbf{s}_i^{(d)} = [s_i(n), s_i(n-1), \dots, s_i(n-(d-1))]$. Note that, due to the stationarity of s_i , the joint distribution of $\mathbf{s}_i^{(d)}$ is invariant with respect to shifts in the time index. The amount of temporal structure in s_i (i.e. its temporal predictability) can then be assessed by its *entropy rate* [35]:

$$H_r(s_i) = H(\mathbf{s}_i^{(d)}) - H(\mathbf{s}_i^{(d-1)}) \quad (2.36)$$

Indeed, in the case of Markovian sources of order d , a maximum likelihood estimate of the separating matrix is obtained by minimizing the entropy rate of the estimated sources [83]. However, estimating the entropy rate requires the combination of two different estimates of joint entropy, whose estimation errors might not cancel each other. Instead, we propose minimizing the following BSS contrast [69]:

$$\phi_{ER}(\mathbf{R}) = \sum_{i=1}^M H(\hat{\mathbf{s}}_i^{(d)}) \quad (2.37)$$

Intuitively, minimizing (2.37) will lead to source estimates that are maximally non-Gaussian (i.e. spatially independent) and that have maximum temporal structure. This intuition stems from the basic information-theoretic equality:

$$H(X_1, X_2, \dots, X_L) = \sum_{i=1}^L H(X_i) - I(X_1, X_2, \dots, X_L) \quad (2.38)$$

where X_1, X_2, \dots, X_L is any set of random variables. If X_1, X_2, \dots, X_L denote the state probabilities of a source estimate at L correlative time instants, then minimizing (2.38) is equivalent to making source estimates as non-Gaussian as possible while maximizing information sharing between correlative time-instants.

Contrast (2.37) is a valid BSS contrast because, if $\mathbf{s}_i^{(d)}$ and $\mathbf{s}_j^{(d)}$ are mutually independent and at least one of them is not Gaussian distributed, then the following inequality holds [35, 225]:

$$\phi_{ER}(\alpha \mathbf{s}_i^{(d)} + \beta \mathbf{s}_j^{(d)}) \geq \min(\phi_{ER}(\mathbf{s}_i^{(d)}), \phi_{ER}(\mathbf{s}_j^{(d)}))$$

with equality if and only if $\alpha = 0$ and $\phi_{ER}(\mathbf{s}_j^{(d)}) < \phi_{ER}(\mathbf{s}_i^{(d)})$, $\beta = 0$

and $\phi_{ER}(\mathbf{s}_i^{(d)}) < \phi_{ER}(\mathbf{s}_j^{(d)})$, or $\alpha\beta = 0$ and $\phi_{ER}(\mathbf{s}_j^{(d)}) = \phi_{ER}(\mathbf{s}_i^{(d)})$. A generalization of this inequality to more than two sources can be found in [37]. However, if the sources are not perfectly independent, local minima will appear and even the global minimum of the proposed contrast might be spurious [225]. Moreover, broad autocorrelations of the sources also contribute to the presence of local minima by effectively increasing the memory of the Markov sources beyond the assumed model order. These two problems can be significantly minimized by pre-processing the observed mixtures with a VAR filter, as was discussed in section 2.2. One may argue that second order autocorrelations are part of the temporal structure that contrast ϕ_{ER} is based on. Indeed, there is a trade-off between too broad temporal structure (beyond the assumed Markov order) and too narrow structure (insufficient for ϕ_{ER} to be able to exploit it). Nevertheless, contrast ϕ_{ER} is also sensitive to temporal dependencies of higher order than those removed by the VAR filter, which explains why VAR pre-processing is beneficial in most cases (see section 2.6).

Evaluation of the BSS contrast (2.37) requires the estimation of multivariate Shannon entropies. We use an estimator based on k-nearest-neighbor distances [112]:

$$\hat{H}(\hat{\mathbf{s}}_i^{(d)}) = h_{k-1} - h_{N-1} + dE[\log 2\epsilon] \quad (2.39)$$

where $h_t = -\sum_{r=1}^t r^{-1}$ and ϵ is the maximum-norm distance from a sample realization of vector $\hat{\mathbf{s}}_i^{(d)}$ (corresponding to a time instant $n = n^*$) to its k:th nearest neighbor realization (corresponding to a time instant $n_k \neq n^*$). The expectation operator $E[\cdot]$ can be approximated by the sample mean. In the numerical experiments performed in section 2.6 below we always used $k = 20$ and $d = 4$. We preferred this estimator instead of fixed-bandwidth kernel-based approaches due to its higher sensitivity to finer (high-order) details of the distributions [210]. The optimum rotation minimizing (2.37) is found using Jacobi rotations [34]. The resulting ENRICA (Entropy Rate-based ICA) algorithm can be summarized in the following steps:

1. Whiten the data through the transformation $\mathbf{z} = \hat{\Sigma}_{\mathbf{x}}^{-1/2} \mathbf{x}$ where $\hat{\Sigma}_{\mathbf{x}} = \frac{1}{N-1} \sum_{i=1}^N \mathbf{x}(n)\mathbf{x}^T(n)$.
2. Perform a temporal whitening by fitting a VAR model to \mathbf{z} using ARfit [193] and computing the residuals of the model. Denote the residuals by $\mathbf{q}(n) \forall n = 1, \dots, N$.
3. Using Jacobi rotations (see Table 2.1), find the unitary matrix $\hat{\mathbf{R}} = [\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \dots, \hat{\mathbf{r}}_M]^T$ minimizing $\sum_{i=1}^M \hat{H}(\hat{\mathbf{s}}_i^d)$, where $\hat{\mathbf{s}}_i(n) = \hat{\mathbf{r}}_i^T \mathbf{q}(n)$.

4. Estimate the separating matrix as $\hat{\mathbf{B}} = \hat{\mathbf{R}}\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}$.

2.6 Experiments and results

All the numerical experiments and the figures shown below can be reproduced using MATLAB code and datasets available online at the URL:

<http://www.cs.tut.fi/~gomezher/bss.htm>.

2.6.1 Non-Gaussian and spectrally diverse sources

In this section we use simulations to show the ability of F-COMBI and M-COMBI to separate both non-Gaussian and time-correlated sources. Due to their computational simplicity, the target application of these algorithms are high-dimensional mixtures and, therefore, we used mixtures of up to 30 sources. This prevented us to include in the comparison non-parametric ICA algorithms that, due to their computational demands, are hardly applicable to these high-dimensional problems. The performance of non-parametric algorithms such as ENRICA [69], RADICAL [122] or MILCA [210] is assessed in sections 2.6.2 and 2.6.3, where low-dimensional mixtures are used. The following algorithms were included in this first set of experiments:

- BSS algorithms for time-correlated sources. Namely, SOBI [13] and WASOBI [228].
- Popular ICA algorithms based on approximations of differential entropy: EFICA [110] and JADE [25].
- A hybrid algorithm that is able to separate both non-Gaussian and time-correlated sources: JADE_{TD} [154]. We did not include in this first set of experiments other algorithms of this type like JCC [72] and ThinICA [36], due to their extremely high computational demands, and due to the fact that, in a set of similar experiments [67, 215], they did not outperform JADE_{TD}. Nevertheless, JCC and ThinICA were included in the performance analysis for the low-dimensional problem in section 2.6.2.

The implementations of these algorithms were obtained from their respective authors' public web-pages or provided directly to us by the authors. The only

exception was JADE_{TD} which we implemented using the publicly available implementations of JADE and TDSEP. The unifying model of [91] was not included in the comparison because the implementation kindly provided by Prof. Hyvärinen did not allow the separation of AR sources of order greater than 1. JADE_{TD} is based on joint diagonalization of quadricovariance eigen matrices and time-delayed correlation matrices. We selected the time lags of the cross-correlations to be $0, 1, \dots, K$ where K denotes the maximum AR order of the source signals (see Table 2.2).

The source signals are generated by feeding AR filters with random i.i.d. samples with different distributions. The mixtures are obtained using randomly generated well-conditioned mixing matrices. The characteristics of the sources are summarized in Table 2.2 where we can see that the simulated dataset consists of $d = 3 \cdot K$ sources. The first K sources have a Gaussian distribution and therefore cannot be separated by means of typical ICA contrasts. On the contrary, sources $K + 1$ to $3 \cdot K$ are all easily separated by exploiting their non-Gaussianity. It can also be observed that for a fixed value of m , the sources with indexes $n \cdot K + m$ for $n = 0, 1, 2$ all have the same spectrum and therefore cannot be separated by means of SOBI, WASOBI or other algorithms exploiting different spectra of the source signals. The multidimensional structure of the simulated dataset for $K = 3$ can be observed in Fig. 2.4. Separation accuracy of the i th source was assessed using the interference-to-signal ratio (ISR) [23]:

$$\text{ISR}_i = 10 \log \left(\sum_{k \neq i} \mathbf{G}_{ik}^2 / \mathbf{G}_{ii}^2 \right) \quad (2.40)$$

where $\mathbf{G}_{ik} = (\hat{\mathbf{B}}\mathbf{A})_{ik}$ with \mathbf{A} the true mixing matrix and $\hat{\mathbf{B}}$ the estimated separating matrix. In general, ISR values above -10 dBs are probably unacceptable in most applications.

Overall separation accuracy is assessed with the average (across sources) ISR. Fig. 2.5 depicts the 90th percentile⁸ of the average ISR for different number of data samples when the dimensionality of the dataset is fixed to 9 (i.e. $K = 3$). The results indicate that M-COMBI and F-COMBI perform almost identically well for this dataset, and that both of them significantly outperform the benchmark algorithms, in terms of estimation accuracy. As shown in Fig. 2.6, computation times⁹ for mixtures of 9 sources are very small in all cases and M-COMBI is only slightly slower than the other algorithms. Fig. 2.7 shows the quality of individual source estimates in terms of the 90th

⁸Percentiles were obtained by generating 100 random surrogates of the sources and of the mixing matrices.

⁹All computations were performed under MATLAB 7.8.0 (R2009a) for Windows XP, running on a Dell Optiplex 960 (Intel Core2 Quad CPU 2.83 GHz, 3.21 GB of RAM).

Source #	distribution	AR filter coefficients
1	Gaussian	$[1, \rho]$
2	Gaussian	$[1, 0, \rho]$
\vdots	\vdots	\vdots
K	Gaussian	$[1, 0, \dots, 0, \rho]$
$K + 1$	Laplacian	$[1, \rho]$
\vdots	\vdots	\vdots
$2 \cdot K$	Laplacian	$[1, 0, \dots, 0, \rho]$
$2 \cdot K + 1$	Uniform	$[1, \rho]$
\vdots	\vdots	\vdots
$3 \cdot K$	Uniform	$[1, 0, \dots, 0, \rho]$

Table 2.2: Characteristics of the source signals used in the experiments shown in Section 2.6.1. Note that by modifying K we can set the dimensionality of the dataset to any multiple of three.

percentile value of the corresponding ISR. Clearly, F-COMBI and M-COMBI are the most robust algorithms and they are the only ones able to separate the Gaussian sources with different spectra.

As was shown in section 2.2.4, temporal structure in the source signals has a negative impact on ICA contrasts that assume i.i.d. sources. By applying ICA on the residuals of a VAR model fitted to the observed mixtures, we can minimize this negative impact and increase overall separation accuracy [69]. Fig. 2.8 shows the benefits of using VAR pre-processing for i.i.d. based algorithms like EFICA and JADE. These benefits are especially evident in the case of EFICA, boosting the average ISR to values close to those obtained by M-COMBI and F-COMBI. However, Fig. 2.9 shows that VAR pre-processing improves only the estimates of the non-Gaussian sources and that F-COMBI and M-COMBI remain as the only alternatives to separate the Gaussian components.

A major advantage of M-COMBI and F-COMBI is that they offer excellent performance for high-dimensional mixtures (See Fig. 2.10). F-COMBI is faster (see Fig. 2.12) but M-COMBI is more reliable and stable, due to its exhaustive search for multidimensional components. The differences in terms of reliability are evident in Fig. 2.13 where M-COMBI outperforms F-COMBI when the AR sources have a very weak temporal structure. This is due to the fixed order that F-COMBI uses to apply its two complementary BSS contrasts: first temporal structure and, subsequently, non-Gaussianity. Thus, a poor (but acceptable) separation in the first stage will necessarily have a negative impact on the final source estimates. However, notice from Fig. 2.13 that this does not mean that

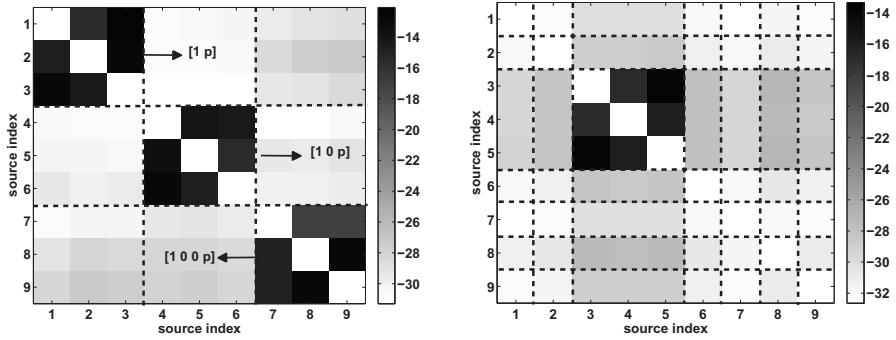


Figure 2.4: The results of clustering the ISR matrices estimated by WASOBI (left) and EFICA (right) for the simulated dataset of section 2.6.1 when $K = 3$, $N = 3000$, and $\rho = 0.6$. The scale is in dBs. The dashed lines mark the clusters of unresolvable components, which in the case of WASOBI correspond to components having the same spectrum. In the case of EFICA there is only one multidimensional cluster of unresolvable components that corresponds to the Gaussian sources.

F-COMBI fails to separate some of the sources, but just that M-COMBI is able to separate much more accurately the non-Gaussian components.

2.6.2 Non-linear sources with cross-dependencies

In this set of experiments we generate the hidden sources using three identical Lorenz oscillators $\Phi_i : (\dot{X}_i(t), \dot{Y}_i(t), \dot{Z}_i(t))$, $\forall i = 1, 2, 3$, described by the differential equations:

$$\begin{aligned}\dot{X}_i(t) &= 10(Y_i(t) - X_i(t)) \\ \dot{Y}_i(t) &= 28X_i(t) - Y_i(t) - X_i(t)Z_i(t) \\ &\quad + \sum_{j \neq i} K_{ij} (Y_j(t - \tau_{ji}))^2 \\ \dot{Z}_i(t) &= X_i(t)Y_i(t) - \frac{8}{3}Z_i(t)\end{aligned}$$

We integrate these equations using a fourth order Runge-Kutta method with a time step of 0.003 but we record only every 100th point leading to an effective time step of 0.3. The sources are obtained from the Y components of the oscillators, i.e. $s_i(t) = Y_i(t) \forall i = 1, 2, 3$. In total 3000 samples are generated for each source. We consider the case of uncoupled ($K_{ij} = 0 \forall i, j$) and unidirectionally coupled oscillators ($K_{21} = 1$, $K_{32} = 1$, $K_{ij} = 0$ otherwise). In the latter case the coupling delays are $\tau_{21} = 10$ and $\tau_{32} = 15$. In order to obtain

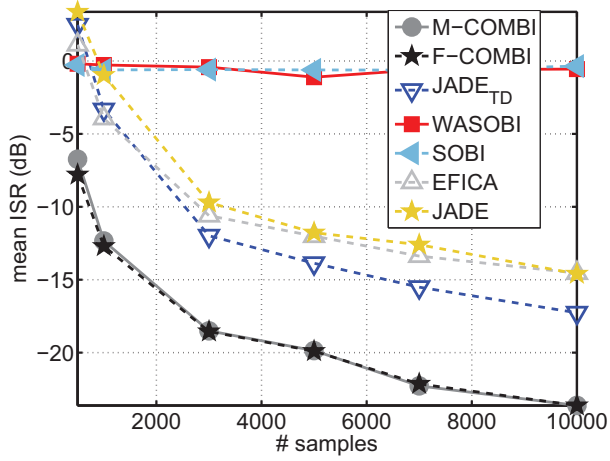


Figure 2.5: 90th percentile value (across 100 random repetitions of the mixtures) of the mean ISR for random mixtures of the sources described in Table 2.2. Separation accuracy is displayed for different sample sizes.

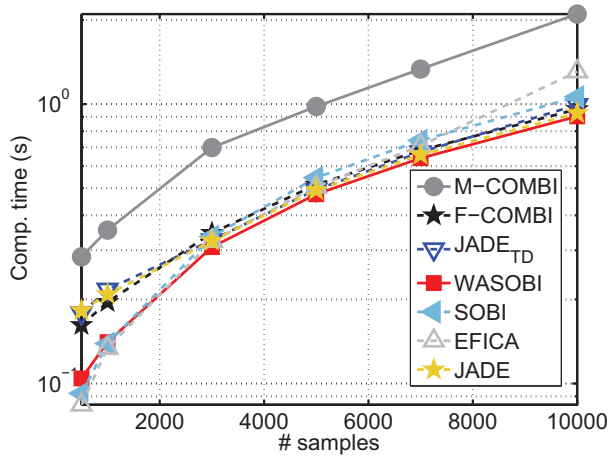


Figure 2.6: Average computation time versus the number of observations for the dataset described in Table 2.2 with $K = 3$.

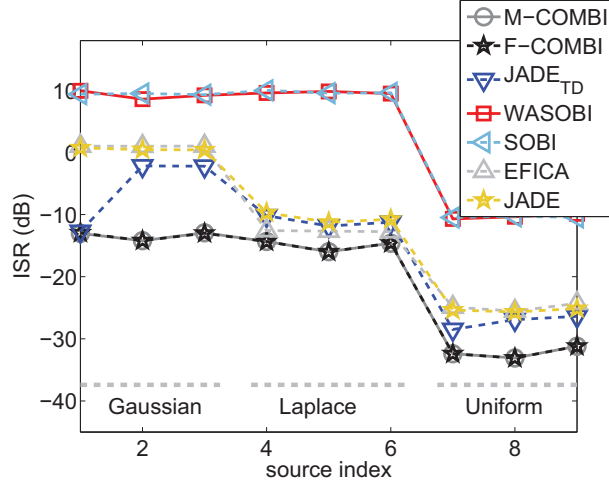


Figure 2.7: 90th percentile value of the ISR obtained for each source. The sample size was fixed to $N = 10000$ data samples.

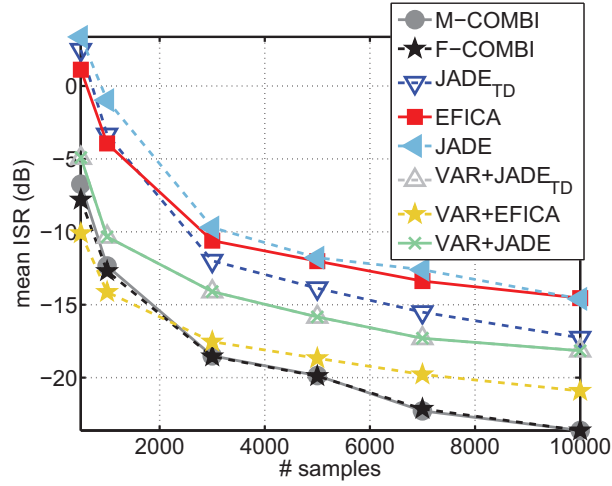


Figure 2.8: 90th percentile value of the average ISR with and without VAR pre-processing.

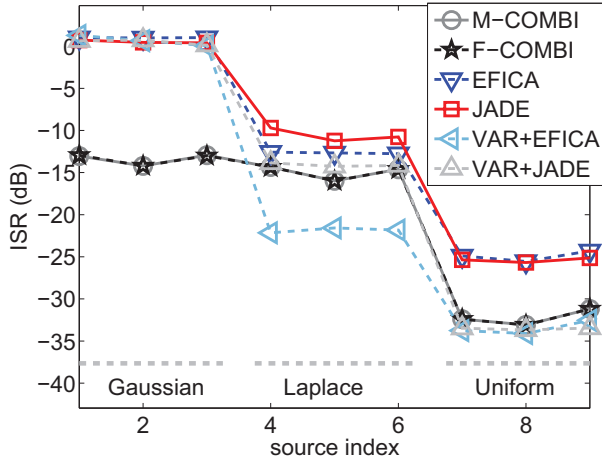


Figure 2.9: 90th percentile value of the ISR obtained for each source with and without VAR pre-processing. The sources contained 10000 data samples.

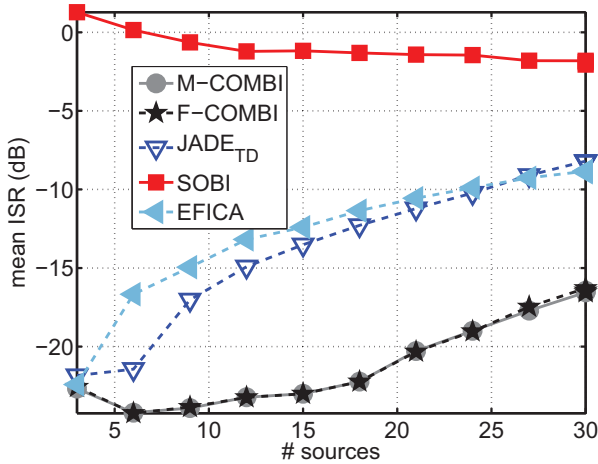


Figure 2.10: 90th percentile of the average ISR (in dBs) across 100 random surrogates of the mixtures of 3 to 30 sources.

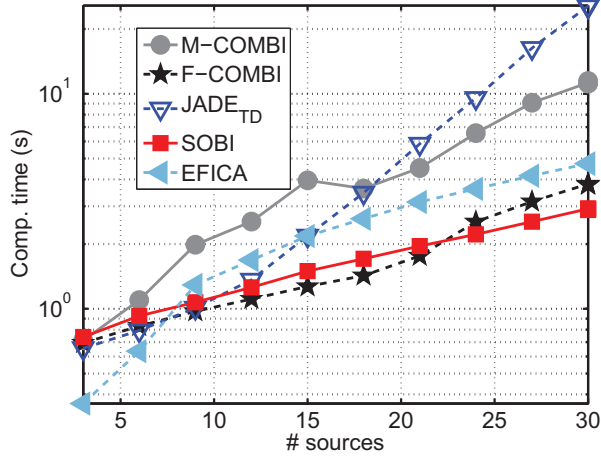
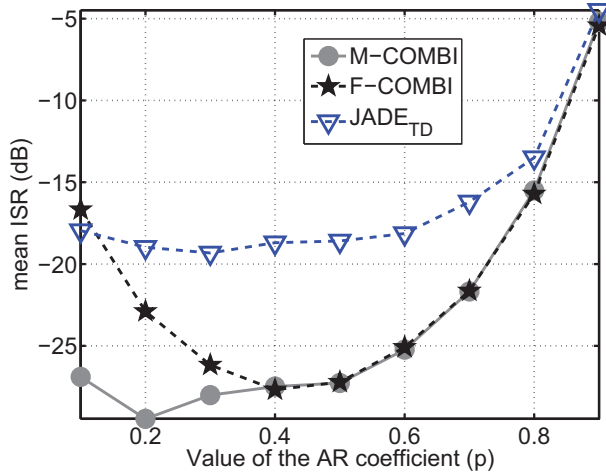


Figure 2.11: Computation times for mixtures of 3 to 30 sources.

Figure 2.12: 90th percentile of the average ISR (in dBs) across 100 random surrogates of the mixtures for different values of the AR coefficient ρ (See table 2.2 for the meaning of this simulation parameter).

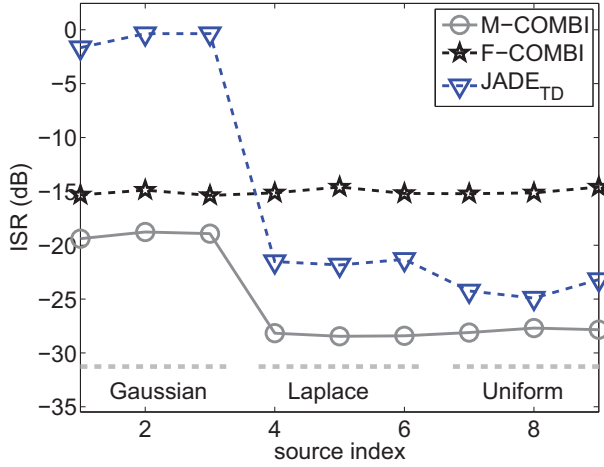


Figure 2.13: 90th percentile value of the ISR obtained for each source when $N = 10000$ samples and $p = 0.1$. Notice that, contrary to JADE_{TD} , F-COMBI successfully separated all sources with acceptable accuracy (with an ISR well below -10 dBs in all cases).

average results, we generate 200 realizations of the mixtures by using different initial conditions for the Lorenz systems and random well-conditioned mixing matrices.

The results in table 2.3 show that, both for the coupled and uncoupled cases, the optimal separating matrix corresponds to a robust global minimum of the contrast function used by the algorithm ENRICA [69]. For coupled Lorenz oscillators, VAR filtering is quite effective in removing spurious global minima in ICA-based contrasts, but it is not able to remove the numerous local minima, which explains the poor results of ICA algorithms based on local optimization (e.g. Infomax, EFICA and NpICA). Moreover, the fact that the distribution of the Lorenz sources is far from Gaussian and multimodal explains the poor performance of algorithms based on parametric or semi-parametric approximations of entropy (e.g. Infomax, JADE, and EFICA). Indeed, VAR pre-processing is especially useful for non-parametric ICA algorithms that use brute-force optimization to find the global minimum of the contrast function (RADICAL [122] and MILCA [210]).

BSS Algorithm		ISR (dB)					
		uncoupled			coupled		
		2.5%	50%	97.5%	2.5%	50%	97.5%
Infomax [12]		-32	-13	0	-1	0	0
	+VAR	-17	-4	1	-5	-2	-1
Ext. Infomax [123]		-3	-1	1	-18	0	1
	+VAR	-7	-1	0	-26	-3	1
JADE [25]		-21	-5	0	-3	0	1
	+VAR	-25	-16	0	-13	-4	-1
EfICA [110]		-32	-7	0	-1	0	0
	+VAR	-9	-2	0	-22	-2	-1
SOBI [13]		-9	-1	1	-7	-6	-5
	+VAR	-15	-3	0	-13	-3	1
WASOBI [228]		-14	-2	1	-21	-17	-14
	+VAR	-18	-3	1	-26	-12	-1
RADICAL [122]		-38	-32	-26	-28	-1	0
	+VAR	-37	-28	-21	-33	-21	-1
MILCA [210]		-40	-32	-26	-26	-8	0
	+VAR	-35	-26	-20	-30	-18	-6
NpICA [15]		-46	-33	0	-33	-1	0
	+VAR	-36	-29	0	-33	-4	-2
MCOMBI [215]		-31	-6	0	-34	-24	-1
	+VAR	-13	-3	0	-28	-2	-1
FCOMBI [67]		-31	-6	0	-21	-17	-14
	+VAR	-12	-2	0	-27	-2	-1
ThinICA [36]		-41	-30	-24	-19	-1	0
	+VAR	-38	-28	-23	-28	-18	-8
JADE _{TD} [154]		-17	-4	0	-3	-1	0
	+VAR	-25	-15	-1	-13	-4	-1
JCC [72]		-5	2	5	-6	1	6
	+VAR	-21	-5	0	-14	-3	0
ENRICA [69]		-42	-34	-29	-42	-34	-27
	+VAR	-41	-34	-29	-40	-33	-26

Table 2.3: Accuracy of the tested algorithms in the blind separation of three Lorenz oscillators. The different columns denote the 2.5%, the 50% and the 97.5% percentiles of the average ISR. Percentile values were obtained by generating 200 random surrogates of the Lorenz mixtures. The rows marked with the term “+VAR” indicate that VAR filtering was used to pre-process the observed mixtures. ISR values above -10 dB are probably unacceptable.

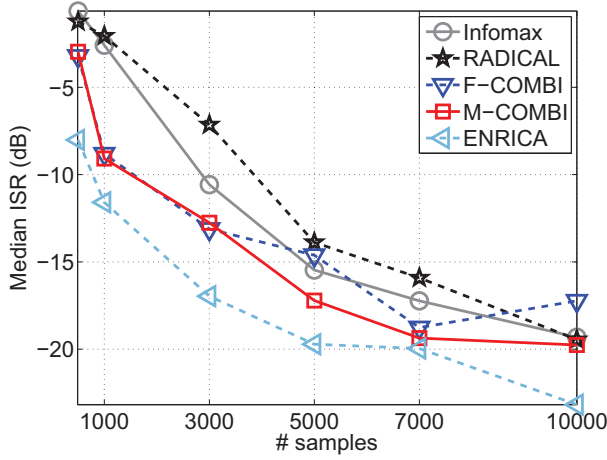


Figure 2.14: 90th percentile of the median ISR (in dBs) across 200 random surrogates of the three EEG sources. For clarity we show only the results of some of the best performing algorithms. Full results of this experiment are available online [68].

2.6.3 Real EEG data

In order to assess the expected performance on real EEG data, we test the BSS algorithms in table 2.3 using mixtures of three time-series extracted from a real EEG dataset [42, 223]. Mutually independent EEG sources are emulated by selecting, from different electrodes, EEG epochs that did not overlap in time. This approach ensures that the time-courses of the sources mimic the dynamics of the underlying true sources. However, the lack of cross-dependencies between sources is probably unrealistic and, therefore, the numerical results of this experiment should be taken as positively biased estimates. From Fig. 2.14 it is obvious that BSS based on temporal structure clearly outperforms i.i.d.-based approaches, at least for realistic sample sizes. The poor convergence of Infomax raises concerns on common practices among the EEG research community. For instance, [45] recommends to use about $30M^2$ samples to estimate M sources but, in our experiments, Infomax needed at least 10 times more samples to produce reliable source estimates. Fig. 2.16 shows the 90th percentile of the median ISR for mixtures of more than three EEG sources. Even in high-dimensional problems, ENRICA consistently outperforms the benchmark algorithms. The major disadvantage of ENRICA with respect to its closer competitor (M-COMBI) is computation time. Separation of 10 sources with 5000 samples, requires about 14 minutes for ENRICA compared to less than 1 second for M-COMBI. However, lengthy computations are often acceptable for offline analysis of EEG data.

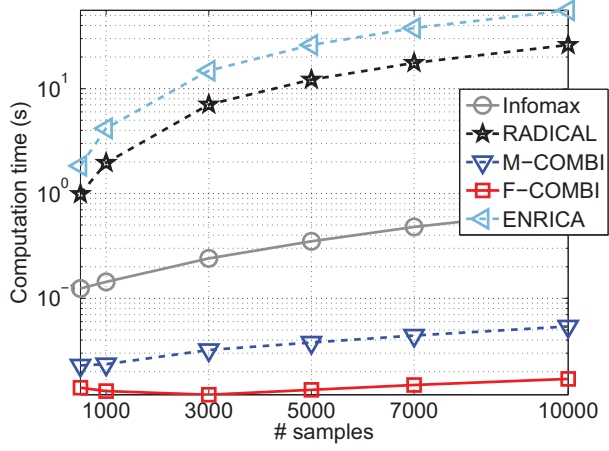


Figure 2.15: Computation times for different sample sizes of mixtures of 3 EEG sources.

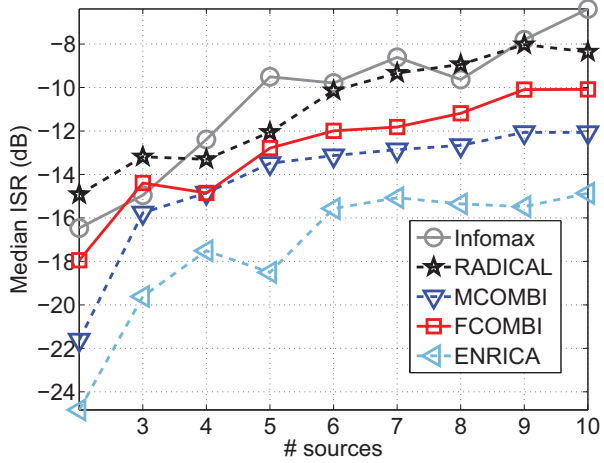


Figure 2.16: 90th percentile of median ISR (in dBs) across 100 random surrogates of mixtures of 2 to 10 EEG sources. Each source contained 5000 data samples.

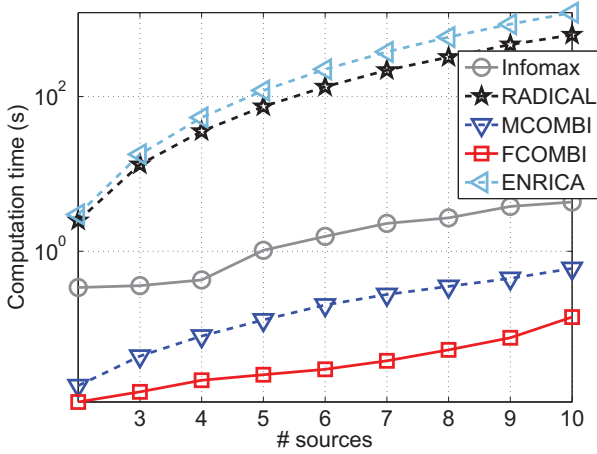


Figure 2.17: Computation times for mixtures of 2 to 10 EEG sources. Each source contained 5000 data samples.

2.7 Conclusions to the chapter

In this chapter we have reviewed the most common approaches to the blind separation of neural sources underlying scalp EEG measurements. Furthermore, we have proposed three novel algorithms (M-COMBI, F-COMBI and ENRICA) with attractive features for the analysis of EEG data. M-COMBI and F-COMBI combine the strengths of two complementary BSS methods (WASOBI and EFICA) in order to improve separation accuracy in hybrid mixtures of non-Gaussian and time-correlated sources. Apart from their robustness, the main advantage of these two algorithms is their computational simplicity, in contrast to the computational burden imposed by non-parametric approaches like ENRICA. This is especially the case of F-COMBI that allows for almost real-time separation (e.g. in less than 1 second) of mixtures of up to 10 sources in a powerful personal computer. Moreover, parallel implementations of F-COMBI are straightforward, and could easily boost separation speed in modern multicore machines. Although M-COMBI is more reliable for offline applications, F-COMBI may be more appropriate for applications in which speed is critical, like brain computer interfaces [199] or neuroprosthetics [80].

F-COMBI and M-COMBI are useful mainly for high-dimensional mixtures for two reasons. First, because multidimensional clusters of components that cannot be separated by WASOBI and EFICA are more easily found in high-dimensional mixtures. Second, because their computational burden is affordable. With respect to similar algorithms aiming at the separation of hy-

brid mixtures of non-Gaussian and spectrally diverse Gaussian sources (e.g. ThinICA, JCC or JADE_{TD}), M-COMBI and F-COMBI are clearly faster. The results shown in [215] and [67] with mixtures of 40 sources show that M-COMBI and F-COMBI can be up to 10 and 100 times faster, respectively.

The brain sources underlying EEG measurements are likely to be connected by neuroelectric pathways, especially in experimental paradigms that aim to study functional brain connectivity. Time-lagged information exchange between the hidden sources can have a very negative impact on BSS algorithms that assume mutual independence, as shown in this chapter with coupled Lorenz oscillators, and as will be confirmed in chapter 4 with VAR sources. In our experience, BSS algorithms based on temporal structure (e.g. WASOBI) are less affected by this problem than ICA methods. For the latter, pre-processing the mixtures using a VAR filter can be helpful, especially in the case of ICA algorithms based on brute-force global optimization of the contrast function (e.g. RADICAL and MILCA). The poor performance of parametric and semiparametric ICA algorithms with the Lorenz dataset is mainly explained by the distribution of the data, which is multimodal and very far from Gaussian. We have to admit that this distribution is rather uncommon and these results do not imply poor performance in real-life applications, including the analysis of EEG data. In fact, in chapter 4 we found that the combination of VAR filtering and EfICA obtains excellent results in the separation of sources with characteristics similar to those of real EEG.

Our experiments with real EEG time-series have raised concerns on whether it is possible to accurately separate many EEG sources (e.g. more than 10) from relatively short EEG epochs (e.g. in the order of 10 to 20 seconds). We found that the common rule-of-thumb of requiring $30M^2$ samples [45] to separate M sources is overoptimistic. We would recommend using of the order of ten times as many samples, assuming a standard sampling rate of 250 Hz.

Chapter 3

Measures of effective connectivity

3.1 Introduction

The intriguing ability of neuronal populations to establish oscillatory *coupling* at large-scale levels has been potentially regarded as one of the brain mechanisms underlying cognition (e.g. [218]). These large-scale interactions are due to a complex pattern of underlying brain connectivity. In this macroscale context, *anatomical* connectivity refers to the specific arrangement of macroscopic fiber pathways linking different brain regions. In contrast, *functional* connectivity is a fundamentally statistical concept, which refers to the existence of synchronized patterns between the temporal activations of often spatially remote neural systems. By this definition, functional connectivity can be measured e.g. by cross correlation, cross-spectra [164] or mutual information [201].

In highly interconnected cerebral systems it is relevant not only to identify anatomical and functional links, but also to measure to what extent the individual brain networks contribute to information production, and at what rate they exchange information among each other. These directional interactions form a pattern of *effective* brain connectivity [56] that carries important information on the *functional integration* mechanisms of the brain.

There are several distinct approaches to understanding and measuring effective connectivity. One approach - dynamic causal modeling (DCM) [57, 58]

- uses state-space continuous-time models for each and every step of the physiological and biophysical chain of events leading to the observed functional measurements. These detailed models are then used to characterize dynamical and structural perturbations in the system due to known deterministic inputs. DCM is especially suitable for functional magnetic resonance imaging (fMRI) studies, due to the relatively simple models involved, and due to the fact that fMRI records responses to deterministic experimental manipulations. However, electrophysiological DCM models (see e.g. [40, 41]) are complex and still largely speculative (even more in the case of neurological disorders), making DCM much less suitable for EEG studies. In addition, application of DCM methods to spontaneous EEG signals is not straightforward [148, 149].

An alternative to DCM is to consider that an effective connection between two neural networks exists whenever the prediction of the neural states of one of the networks can be improved by incorporating information from the other. This so-called *Granger causality* (GC) was originally proposed by Wiener [227], and later formalized by Granger [73], in the context of linear regression models of stochastic processes. Since then, linear GC has also been extended to non-linear models [30, 54, 191] and to non-parametric GC indices derived from information theory [29, 55, 195, 222].

Closely related to linear GC is the concept of *directed coherence* (DC) introduced by Saito and Harashima [186] in the early 1980s. Like linear GC, effective connectivity indices based on DC rely heavily on vector autoregressive (VAR) models to infer causal relationships between temporally structured time-series. The most prominent DC-based connectivity measure is the directed transfer function (DTF) [104], which roughly measures to what extent a spectral component in a neural signal induces the generation of the same spectral component in another neural signal. The DTF has been used in previous EEG studies (see e.g. [6, 52, 111, 119]) and we will use it in chapters 4 and 5 to characterize flows of oscillatory activity underlying the generation of the human alpha rhythm.

Lastly, Nolte et al. have recently proposed the so-called phase-slope index (PSI) [159] as a new method to estimate the direction of causal interactions between time-series. The idea behind the PSI is that a time-lagged interaction between a pair of systems induces a slope in the phase of the cross-spectra of those systems. The existence of this slope allows detecting the interaction. The PSI is a promising tool for assessing effective connectivity, and has been tentatively applied to EEG data [159]. However, a major limitation of the PSI is that it is a bivariate measure and, therefore, might be confounded by the presence of common drivers in systems consisting of more than two variates.

In this chapter, we review the most important information theoretic ap-

proaches that have been proposed in the literature to assess effective connectivity between dynamical systems. We then present *partial transfer entropy* (PTE) as a natural extension of the well-known bivariate transfer entropy [195] for assessing causality in multivariate systems. Finally, we propose several practical estimators of PTE that (i) effectively integrate the information available in an ensemble of repeated measurements and (ii) are able to identify time-varying coupling patterns.

3.2 Information-theoretic indices

The average number of information units¹ needed to encode a discrete random variable Z with states $\{z_1, \dots, z_M\}$ and a probability mass function $p(z)$ is given by the Shannon entropy [201]:

$$H_Z \triangleq - \sum_{i=1}^M p(z_i) \log p(z_i) \quad (3.1)$$

Shannon entropy can be understood as the *average information* carried by Z . For a continuous d -dimensional random vector X with pdf $f(\mathbf{x})$, a similar information measure can be defined by discretizing X using some bins in \mathbb{R}^d . If we denote by Y the binned variable then we have that [114]:

$$H_Y \approx H_X - d \log \Delta \quad (3.2)$$

where Δ is the Lebesgue measure of the bins and H_X is the *differential entropy* of the continuous random vector X :

$$H_X \triangleq - \int_{\mathcal{S}_f} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x} \quad (3.3)$$

where \mathcal{S}_f is the support of $f(\mathbf{x})$. Notice that H_X is not a true information measure and that it can be negative. Moreover, H_X is not invariant under homeomorphisms of X .

¹The units of information depend on the base of the logarithm in (3.1). In this thesis we use natural logarithms so that the units of information are *nats*.

The *Kullback-Leibler (KL) divergence* [118] measures the average excess number of information units that would have been used to encode the discrete variable Z , if one would have wrongly assumed that the probability mass function of Z was $q(z)$ instead of $p(z)$:

$$K(p||q) \triangleq \sum_i p(z_i) \log \frac{p(z_i)}{q(z_i)} \quad (3.4)$$

The KL divergence can also be defined for continuous random variables with pdfs f and g :

$$K(f||g) \triangleq \int_{\mathcal{S}_f \cap \mathcal{S}_g} f(\mathbf{x}) \log \frac{f(\mathbf{x})}{g(\mathbf{x})} d\mathbf{x} \quad (3.5)$$

Although the KL divergence is not a distance metric, it can be intuitively understood as a measure of how different the pdfs f and g are. The reason is that the KL divergence between two pdfs is zero, if and only if the two distributions are identical.

Since we aim to measure information transfer in multivariate dynamical systems, we will consider in the following three stochastic processes X , Y and Z , that can be approximately modeled with a stationary Markov model of finite order. Thus, the *phase-space*² of process X can be reconstructed using the embedded vector $\mathbf{x}(n) = (x(n), \dots, x(n - K_x + 1))$, where K_x is the corresponding Markov order. Similarly, we could construct $\mathbf{y}(n)$ and $\mathbf{z}(n)$ for processes Y and Z , respectively. A useful notational convention is that $p(\mathbf{x}(n))$ denotes the probability to find X in the state $\mathbf{x}(n)$ at time n , i.e. the joint probability of finding X at states $x(n), \dots, x(n - K_x + 1)$ during time instants $n, n-1, \dots, n-K_x+1$. Due to stationarity, $p(\mathbf{x}(n))$ does not depend on the time index and we can simply write $p(\mathbf{x})$. In the absence of any direct or indirect transfer of information between X and Y , one can argue that the joint pdf of their corresponding state-vectors can be factorized as follows:

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y}) \quad (3.6)$$

indicating that knowledge of the outcome of one of the systems does not provide any information on the outcome of the other. By measuring the KL divergence

²See appendix B for information regarding the concept of phase-space and phase-space reconstruction through delay embedding.

between the pdfs at both sides of the equality (3.6) we can assess how far from the truth our null hypothesis that X and Y are not mutually related is. This leads to the definition of *mutual information* (MI) [35]:

$$I_{X \leftrightarrow Y} \equiv I(\mathbf{x}, \mathbf{y}) \triangleq \int p(\mathbf{x}, \mathbf{y}) \log \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} d\mathbf{x}d\mathbf{y} \quad (3.7)$$

MI can be used for detecting synchronization between two time-series. However, since MI does not incorporate any dynamical information, it cannot identify directional interactions. A somewhat ad-hoc solution to this limitation is to introduce a delay in one of the two signals:

$$I_{X \leftrightarrow Y}(\tau) \triangleq I(\mathbf{x}(n), \mathbf{y}(n - \tau)) \quad (3.8)$$

which is one of the few information-theoretic indices commonly used in neuroscience (e.g. [96, 97]). Even if $I_{X \leftrightarrow Y}(\tau)$ can identify directionality, it cannot determine whether there is a true information transfer between X and Y , or if their time-lagged synchronization is due to an indirect connection mediated by Z . This is a limitation inherent to the bivariate nature of MI.

A natural extension of MI-based connectivity analysis is to consider that, in the absence of any direct interaction between X and Y , the following factorization must hold:

$$p(\mathbf{y}, \mathbf{x} \mid \mathbf{z}) = p(\mathbf{y} \mid \mathbf{z}) p(\mathbf{x} \mid \mathbf{z}) \quad (3.9)$$

which conceptually means that \mathbf{x} (resp. \mathbf{y}) does not contain any additional information regarding \mathbf{y} (resp. \mathbf{x}) apart from the information that is also shared with \mathbf{z} . Again, the validity of this null-hypothesis can be assessed using the KL divergence, which leads to the definition of *partial mutual information* (PMI) [55]:

$$I_{X \leftrightarrow Y \mid Z} \triangleq I(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) = \int p(\mathbf{x}, \mathbf{y}, \mathbf{z}) \log \frac{p(\mathbf{x}, \mathbf{y}, \mathbf{z}) p(\mathbf{z})}{p(\mathbf{x}, \mathbf{z}) p(\mathbf{y}, \mathbf{z})} d\mathbf{x}d\mathbf{y}d\mathbf{z} \quad (3.10)$$

PMI has the appealing property of being able to discriminate between a direct link between X and Y and an indirect relationship mediated by Z . As

with MI, we can introduce a delay in the PMI definition, in order to identify directionality in the information exchange:

$$I_{X \leftrightarrow Y|Z}(\tau) = I(\mathbf{x}(n), \mathbf{y}(n - \tau) \mid \mathbf{z}(n)) \quad (3.11)$$

An alternative to the static probabilities used by MI and PMI is to study transition probabilities and consider that, in the absence of a direct information transfer in the direction $X \leftarrow Y$, any forecast of X based on knowledge of the present and past of X and Z is not affected by the additional knowledge of the present and past of Y , that is:

$$p(x(n+1) \mid \mathbf{x}(n), \mathbf{z}(n)) = p(x(n+1) \mid \mathbf{x}(n), \mathbf{y}(n - \tau), \mathbf{z}(n)) \quad (3.12)$$

with $\tau \geq 0$. The term $p(x(n+1) \mid \mathbf{x}(n), \mathbf{z}(n))$ denotes the probability of finding the Markov process X in state $x(n+1)$ at time-instant $n+1$, given that X and Y are in states $\mathbf{x}(n)$ and $\mathbf{y}(n)$ at time-instant n . The right side of the equality can be understood in an analogous manner. Then, the KL divergence between the distributions at both sides of (3.12) leads to the definition of *partial transfer entropy* (PTE)³:

$$T_{X \leftarrow Y|Z}(\tau) = \int p(\mathbf{v}) \log \frac{p(x(n+1) \mid \mathbf{x}(n), \mathbf{y}(n - \tau), \mathbf{z}(n))}{p(x(n+1) \mid \mathbf{x}(n), \mathbf{z}(n))} d\mathbf{v} \quad (3.13)$$

with $\mathbf{v} \equiv (x(n+1), \mathbf{x}(n), \mathbf{y}(n - \tau), \mathbf{z}(n))$. PTE is a multivariate extension of bivariate transfer entropy [195] and, like PMI, it is a directed measure of information transfer. But, contrary to PMI, PTE is not confounded by the presence of static correlations due to the common history or common input signals. Nevertheless, PMI and PTE are indeed very closely related:

$$T_{X \leftarrow Y|Z}(\tau) \equiv I_{X \leftrightarrow Y|X, Z}(\tau) \quad (3.14)$$

with $\tau \geq 0$ and $X^+ \equiv x(n+1)$. Thus, PTE simply extends PMI by including in the condition the history of X .

³Notice that in the formulas of PTE and PMI we have omitted the lag of the condition for notational simplicity. Ideally, the state-space $\mathbf{z}(n)$ should be shifted by a lag τ_z such that $T_{X \leftarrow Y|Z}$ (resp. $I_{X \leftrightarrow Y|Z}$) is minimized. For the PTE, the lag τ_z must be positive, i.e. only delays are considered.

3.2.1 Estimation

Let $V = (V_1, \dots, V_m)$ denote a random variable. Then, an *entropy combination* is defined by [70, 185]⁴:

$$C(V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}) = \sum_{i=1}^p s_i H(V_{\mathcal{L}_i}) - H(V) \quad (3.15)$$

where $\forall i \in [1, p]: \mathcal{L}_i \subset [1, m]$ and $s_i \in \{-1, 1\}$ such that $\sum_{i=1}^p s_i \chi_{\mathcal{L}_i} = \chi_{[1, m]}$ where $\chi_{\mathcal{S}}$ is the characteristic function of a set \mathcal{S} . It can be easily checked that MI, PMI and PTE are all entropy combinations (see Table 3.1):

$$I_{X \leftrightarrow Y} \equiv -H_{XY} + H_X + H_Y \quad (3.16)$$

$$I_{X \leftrightarrow Y|Z} \equiv -H_{XZY} + H_{XZ} + H_{ZY} - H_Z \quad (3.17)$$

$$T_{X \leftarrow Y|Z} \equiv -H_{WXY} + H_{WX} + H_{YZ} - H_{XZ} \quad (3.18)$$

where $W \equiv X^+ \equiv x(n+1)$.

A straightforward approach to estimate entropy combinations would be to combine separate estimates of each of the involved multidimensional entropies. The differential entropy of a random vector can be estimated using *plug-in* estimators that perform numerical integration of an estimate of the random vector's pdf. However, estimation of the density of a (probably high-dimensional) random vector is a complex problem, which involves the selection of crucial parameters such as bin width for histogram methods, or kernel type and number of components for methods based on mixtures models. An alternative that we have found to be generally more robust and accurate is to use the nearest-neighbor entropy estimator by Kozachenko and Leonenko [112]. For N realizations $\mathbf{x}[1], \mathbf{x}[2], \dots, \mathbf{x}[N]$ of a d -dimensional random vector X , this estimator takes the form:

$$\hat{H}_X = -\psi(k) + \psi(N) + \log(v_d) + \frac{d}{N} \sum_{i=1}^N \log(\epsilon(i)) \quad (3.19)$$

⁴The term *entropy combination* and its definition in (3.15) were proposed by my colleague K. Rutenen [185].

Index	Entropy combination parameters		
$I_{X \leftrightarrow Y}$	V	=	$(V_1, V_2) = (X, Y)$
	p	=	2
	\mathcal{L}_1	=	$\{1\}$
	\mathcal{L}_2	=	$\{2\}$
	s_1	=	1
	s_2	=	1
	$C(X, Y)$	=	$-H_{XY} + H_X + H_Y$
$I_{X \leftrightarrow Y Z}$	V	=	$(V_1, V_2, V_3) = (X, Z, Y)$
	p	=	3
	\mathcal{L}_1	=	$\{1, 2\}$
	\mathcal{L}_2	=	$\{2, 3\}$
	\mathcal{L}_3	=	$\{2\}$
	s_1	=	1
	s_2	=	1
	s_3	=	-1
	$C(X, Z, Y)$	=	$-H_{XZY} + H_{XZ} + H_{ZY} - H_Z$
$T_{X \leftrightarrow Y Z}$	V	=	$(V_1, V_2, V_3, V_4) = (W, X, Z, Y)$
	p	=	3
	\mathcal{L}_1	=	$\{1, 2, 3\}$
	\mathcal{L}_2	=	$\{2, 3, 4\}$
	\mathcal{L}_3	=	$\{2, 3\}$
	s_1	=	1
	s_2	=	1
	s_3	=	-1
	$C(W, X, Z, Y)$	=	$-H_{W X Z Y} + H_{W X Z} + H_{X Z Y} - H_{X Z}$

Table 3.1: Equivalence between the definition of entropy combination (3.15) and the definitions of MI, PMI, and PTE. This table has been adapted with permission from [185].

where ψ is the digamma function⁵, v_d is the volume of the d -dimensional unit ball⁶, and $\epsilon(i)$ is the distance from $\mathbf{x}[i]$ to its k th nearest neighbor in the set $\{\mathbf{x}[j]\}_{\forall j \neq i}$. The Kozachenko-Leonenko estimator is based on the assumption that the density $p(X)$ is constant within an ϵ -ball. The validity of this assumption determines the bias of the final entropy estimate. Therefore, the number of neighbors k and the number of available data samples N control the trade-off between bias and variance of the estimator. In general, the larger the sample size the smaller one should choose the parameter k . Below we will use numerical experiments to assess the bias and variance of MI and PTE estimators based on the Kozachenko-Leonenko entropy estimator.

The size of the ϵ -balls used by the Kozachenko-Leonenko estimator depends directly on the dimensionality of the random vector. Thus, the bias of estimates of the differential entropies in (3.15) will, in general, not cancel, leading to a poor estimator of entropy combinations. This problem can be partially overcome by noticing that (3.19) holds for any value of k so that we do not need to have a fixed k . Therefore, we fix the number of neighbors in the joint space but use a pointwise adaptive value of k in the marginal spaces, so that the corresponding ϵ -balls approximately match the size of the balls in the joint space. By using the maximum norm to compute distances between points we guarantee that this approximation holds at least for one of the marginal spaces. This idea was originally proposed in [114] for estimating mutual information and was generalized in [70, 185] to the following estimator of entropy combinations:

$$\hat{C}(V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}) = F(k) - \sum_{i=1}^p s_i \langle F(k_i(n)) \rangle_n \quad (3.20)$$

where $F(k) = \psi(k) - \psi(N)$ and $\langle \dots \rangle_n = \frac{1}{N} \sum_{n=1}^N (\dots)$ denotes averaging with respect to the time index. The term $k_i(n)$ accounts for the number of neighbors of the n th realization of the marginal vector $V_{\mathcal{L}_i}$ located at a distance strictly less than $\epsilon(n)$, where $\epsilon(n)$ denotes the radius of the ϵ -ball in the joint space. The point itself is included in this counting. From (3.20) and using Table 3.1 we can write the following estimator for MI [114]:

⁵ $\psi(x) = \Gamma(x)^{-1} \frac{d\Gamma(x)}{dx}$ with $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$. The digamma function satisfies the recursion $\psi(x+1) = \psi(x) + \frac{1}{x}$ and $\psi(1) = -0.5772156\dots$ is the Euler-Mascheroni constant.

⁶The volume of the unit ball depends on the used norm. In this thesis we assume an l^∞ norm so that $v_d = 2^d$ for any dimensionality d . If we would have used the Euclidean norm then the volume of the unit ball in d dimensions would be $c_d = \frac{\pi^{\frac{d}{2}}}{\Gamma(1+d/2)}$ with Γ denoting the Gamma function.

$$\hat{I}_{X \leftrightarrow Y} = \psi(k) + \psi(N) - \langle \psi(k_x(n)) + \psi(k_y(n)) \rangle_n \quad (3.21)$$

Similarly for PMI [55]:

$$\hat{I}_{X \leftrightarrow Y|Z} = \psi(k) - \langle \psi(k_{xz}(n)) + \psi(k_{zy}(n)) - \psi(k_z(n)) \rangle_n \quad (3.22)$$

and for PTE [70]:

$$\hat{T}_{X \leftarrow Y|Z} = \psi(k) - \langle \psi(k_{wxz}(n)) + \psi(k_{xzy}(n)) - \psi(k_{xz}(n)) \rangle_n \quad (3.23)$$

where $W \equiv X^+ \equiv x(n+1)$ and $\langle \dots \rangle_n = \sum_{n=1}^N (\dots)$.

3.3 Ensemble estimators

A practical pitfall common to all the non-parametric indices described in the previous section is that their estimation require many stationary data samples. This limits significantly their use in the analysis of experimental datasets involving time-varying or short-duration causal links. For instance, fast brain responses to different types of external stimuli – or event-related potentials (ERPs) – are likely to cause rapid topological reorganization of information processing pathways. Typically, an ensemble of repeated measurements is obtained in ERP experiments by presenting the stimuli many times. However, to date, there is not an efficient strategy for integrating the coupling information across data trials, in order to produce a more accurate estimation of information-theoretic GC indices. In this section we extend the estimators of MI, PMI and PTE for the characterization of coupling dynamics in an ensemble of repeated measurements. These extensions are based on the following time-adaptive version of the entropy combination estimator in (3.20):

$$\hat{C}(\{V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}\}, n) = F(k) - \sum_{i=1}^p s_i F(k_i(n)) \quad (3.24)$$

This naive time-adaptive estimator is not useful in practice due to its large variance, which stems from the fact that a single data point is used for producing the estimate at each time instant. However, let us consider the case

of an ensemble of r' repeated measurements (trials) from the dynamics of V . Let us also denote by $\{\mathbf{v}^{(r)}[n]\}_r$ the measured dynamics for those trials ($r = 1, 2, \dots, r'$). Similarly, we denote by $\{\mathbf{v}_i^{(r)}[n]\}_r$ the measured dynamics for the marginal vector $V_{\mathcal{L}_i}$. A straightforward approach for integrating the information from different trials is to average together estimates obtained from individual trials:

$$\hat{C}^{\text{avg}}(\{V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}\}, n) = \frac{1}{r'} \sum_{r=1}^{r'} \hat{C}^{(r)}(\{V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}\}, n) \quad (3.25)$$

where $\hat{C}^{(r)}(\{V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}\}, n)$ is the estimate obtained from the r th trial. However, this approach makes a poor use of the available data and will typically produce useless estimates, as will be shown in the experimental section of this chapter. A more effective procedure takes into account the multi-trial nature of our data by searching for neighbors across ensemble members, rather than from within each individual trial. This *nearest ensemble neighbors* [113] approach is illustrated in Fig. 3.1 and leads to the following ensemble estimator of entropy combinations:

$$\hat{C}^{\text{en}}(\{V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}\}, n) = F(k) - \frac{1}{r'} \sum_{r=1}^{r'} \sum_{i=1}^p s_i F\left(k_i^{(r)}(n)\right) \quad (3.26)$$

where the counts of marginal neighbors $\{k_i^{(r)}(n)\}_{\forall i=1, \dots, p}^{\forall r=1, \dots, r'}$ are computed as shown in Fig. 3.1.

3.4 Experiments and results

In this section we use simulations and experimental data from electronic circuits to assess the validity of estimators (3.20) and (3.26). All the experiments and figures shown in this section have been produced and can be replicated using MATLAB code and datasets publicly available online at:

<http://www.cs.tut.fi/~gomezher/causality/>

This MATLAB code relies on efficient C++ implementations of estimators (3.20) and (3.26) that are included in the TIM library [184], a software package developed by my colleague K. Rutanen.

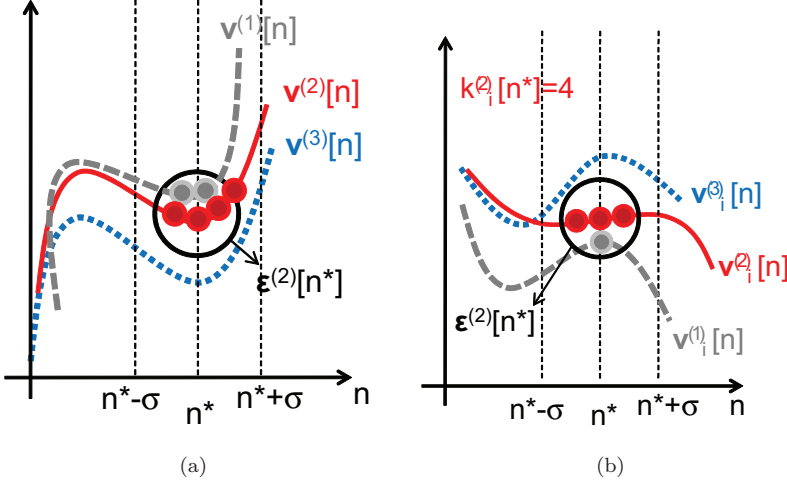


Figure 3.1: For each time instant $n = n^*$ and for each member of the ensemble of joint phase-space trajectories $\{\mathbf{v}^{(r)}[n]\}_{1 \leq n \leq N}^{1 \leq r \leq r'}$ we define the neighborhood size $\epsilon^{(r)}[n^*]$ as the maximum norm distance between $\mathbf{v}^{(r)}[n^*]$ and its k th nearest neighbor in the set $\{\mathbf{v}^{(j)}[n] \mid 1 \leq j \leq r'; |n - n^*| \leq \sigma\}$. This is illustrated in Fig. 3.1(a) for $k = 5$ neighbors. Then we denote by $k_i^{(r)}[n^*]$ the number of neighbors of $\mathbf{v}_i^{(r)}[n^*]$ in the set $\{\mathbf{v}_i^{(j)}[n] \mid 1 \leq j \leq r'; |n - n^*| \leq \sigma\}$ which are within the radius $\epsilon^{(r)}[n^*]$, as depicted in Fig. 3.1(b). These neighbor counts are obtained for all $i = 1, \dots, p$ marginal trajectories. For rapidly changing connectivity patterns, small values of σ might be needed to track the coupling dynamics. On the other hand, larger values of σ will generally lead to lower variance of the estimates.

3.4.1 Multivariate Gaussian distribution

In this first set of numerical experiments we assess the bias and variance of entropy combination estimates for the case of Gaussian-distributed data. We consider an m -dimensional Gaussian random variable $V \sim N(0, \Sigma)$ and the following entropy combination of its marginal components:

$$C(V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}) = \sum_{i=1}^p s_i H(V_{\mathcal{L}_i}) - H(V) \quad (3.27)$$

where $V_{\mathcal{L}_i}$ is the random vector formed by selecting from V the dimensions specified by the set of indices $\mathcal{L}_i \subset [1, m]$ and $s \in \{-1, 1\}^p$ subject to $\sum_{i=1}^p s_i \chi_{\mathcal{L}_i} = \chi_{[1, m]}$. Then, the analytical value of such an entropy combination can be easily found to be:

$$C(V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}) = \frac{1}{2} \log \left(\frac{\prod_{i=1}^p |\Sigma_{\mathcal{L}_i}|^{\text{sign}(s_i)}}{|\Sigma|} \right) \quad (3.28)$$

where $\Sigma_{\mathcal{L}_i}$ denotes the covariance matrix between the marginal components of V corresponding to the set of indices \mathcal{L}_i . Using the expression above and Table 3.1 we can easily obtain the analytical values of the MI, PMI and PTE for Gaussian distributed variables. Fig. 3.2 shows the bias and variance of the MI estimator (3.21) for different sample sizes and different cross-correlation coefficient between the marginal components of a 2-dimensional Gaussian variable. It can be clearly observed that the bias of the MI estimator is mainly affected by the cross-correlation coefficient and that very accurate estimates can be obtained, even with very small sample sizes, as long as the cross-correlation is small. As shown by Fig. 3.3, the null-distribution of the MI estimator (i.e. the bias and variance for mutually independent marginal components) is very tightly located around zero, which suggests that the estimator will lead to very few type I errors, when assessing information flow between a pair of random variables. In contrast, Fig. 3.4 shows that classical estimators based on fixed-size binning of the data⁷ require extremely large sample sizes to produce estimates with low bias and variance, even for perfectly uncorrelated variables. Due to the typically non-stationary nature of brain activity, it is often impossible to use long data epochs to characterize functional connectivity between EEG time series, suggesting that estimators based on nearest neighbors are much more appropriate for EEG applications than binning-based approaches.

⁷Namely, Fig. 3.4 was obtained using the mutual information estimator included in the MATLAB toolbox EEGLAB [46], which is a popular software package among EEG researchers.

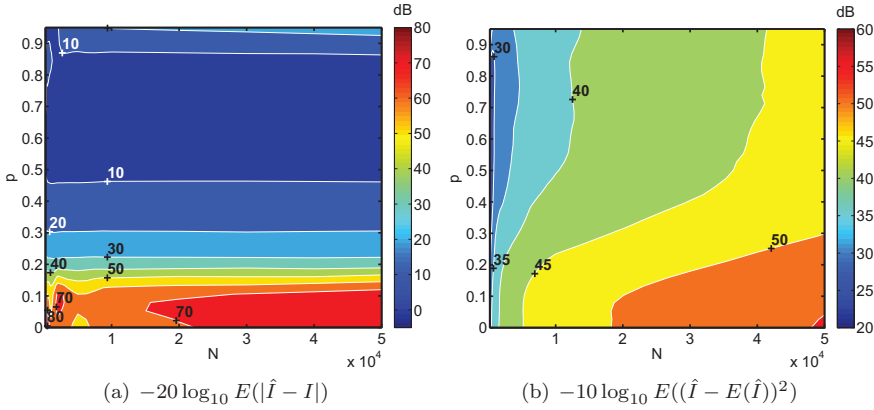


Figure 3.2: Bias (left) and variance (right) of the MI estimator (with $k = 10$ nearest neighbors) for different sample sizes (N) and for different levels of cross-correlation between two Gaussian random variables. The results are the average of 100 independent realizations. Figures are in inverted dB, i.e. the larger the plotted values, the smaller the bias (resp. variance) of the estimator.

In the case of the PTE we have that $V = (X^+, X, Z, Y) \equiv (W, X, Z, Y)$ and the null-hypothesis that there is no flow from Y towards X is the same as saying that the covariance between Y and X^+ is zero. Therefore, we obtained the null-distribution of the PTE estimator (see Fig. 3.5) by using a random positive definite covariance matrix Σ with $\sigma_{wx} = 0$. Fig. 3.6 shows the bias and variance of the PTE estimator for different cross-correlation coefficient between Y and X^+ , i.e. for different levels of information sharing between Y and X^+ . The amount of information transfer has an important impact on the accuracy of the estimates and, therefore, comparisons between large PTE values must be made with caution. However, the good localization around zero of the null distribution confirms that large PTE values can be used with confidence to reject the null hypothesis that there is no transfer of information.

3.4.2 Coupled Lorenz oscillators

In this set of experiments we use, as in section 2.6.2, three identical Lorenz oscillators $\Phi_i : (\dot{X}_i(t), \dot{Y}_i(t), \dot{Z}_i(t))$, $\forall i = 1, 2, 3$, described by the differential equations:

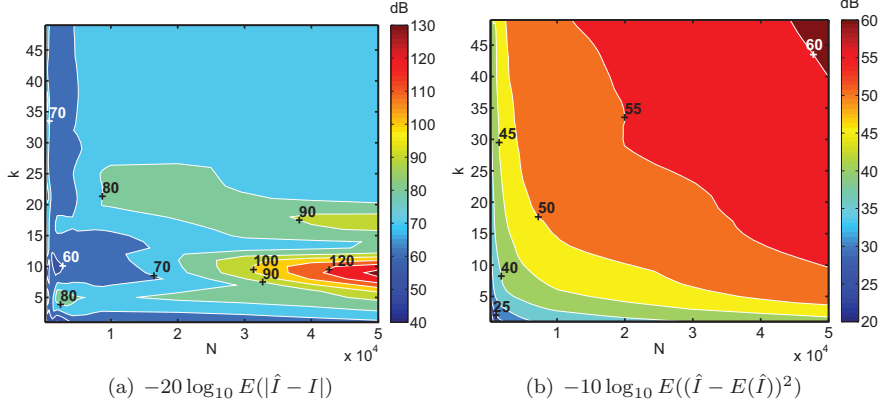


Figure 3.3: Bias (left) and variance (right) of the null distribution of the MI estimator for different number of nearest neighbors (k) and for different sample sizes (N). The results were obtained from 100 independent realizations. Figures are in inverted dB, i.e. the larger the plotted values, the smaller the bias (resp. variance) of the estimator.

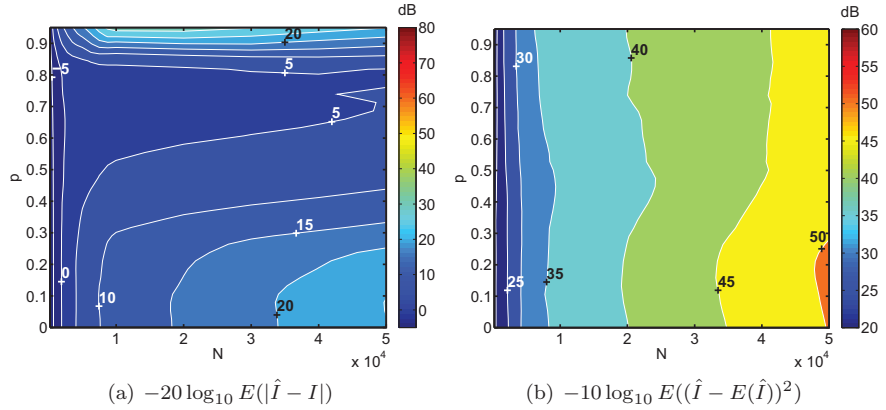


Figure 3.4: Bias (left) and variance (right) of the MI estimator based on fixed-size binning for different sample sizes (N) and for different levels of cross-correlation between two Gaussian random variables. The results were obtained from 100 independent realizations. Figures are in inverted dB, i.e. the larger the plotted values, the smaller the bias (resp. variance) of the estimator.

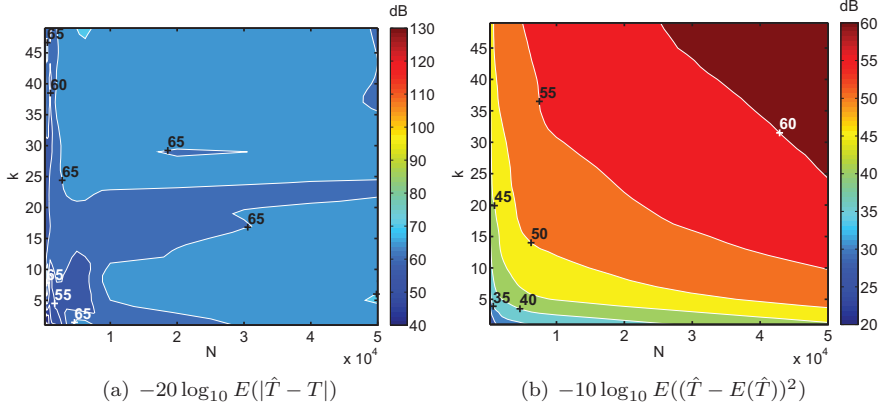


Figure 3.5: Bias (left) and variance (right) of the PTE estimator for mutually independent Gaussian-distributed random variables. The figures were obtained from 100 independent realizations. The figures are in inverted dB, i.e. the larger the plotted values, the smaller the bias (resp. variance) of the estimator.

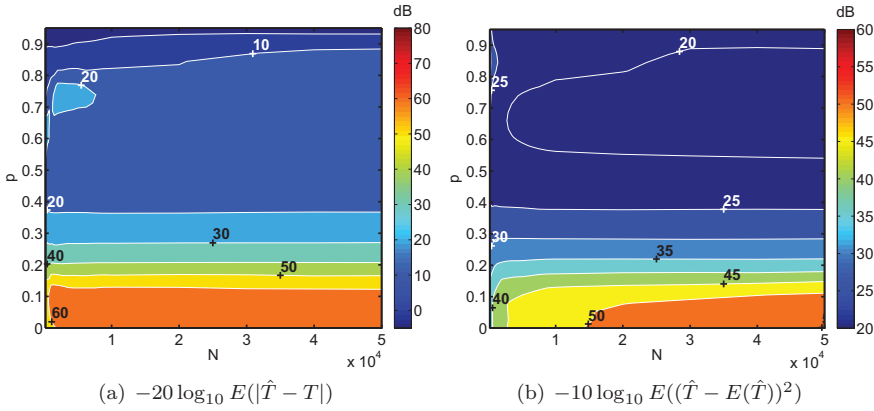


Figure 3.6: Bias (left) and variance (right) of the PTE estimator for a 4-dimensional random vector $V \equiv (X^+, X, Z, Y)$ and different values of the cross-correlation coefficient between X^+ and Y . The other entries in the covariance matrix of C were randomized but we ensured that the diagonal entries were fixed to one and that the matrix was positive definite. The values shown were obtained from 100 independent realizations of $\mathbf{v}[1], \dots, \mathbf{v}[N]$. The figures are in inverted dB, i.e. the larger the plotted values, the smaller the bias (resp. variance) of the estimator.

$$\begin{aligned}
\dot{X}_i(t) &= 10(Y_i(t) - X_i(t)) \\
\dot{Y}_i(t) &= 28X_i(t) - Y_i(t) - X_i(t)Z_i(t) \\
&\quad + \sum_{j \neq i} K_{ij} (Y_j(t - \tau_{ij}))^2 \\
\dot{Z}_i(t) &= X_i(t)Y_i(t) - \frac{8}{3}Z_i(t)
\end{aligned}$$

We integrate these equations using a fourth order Runge-Kutta method with a time step of 0.003, but we keep only every 100th point leading to an actual time step of 0.3. We cut away transient dynamics at the beginning and consider only the time-series of the Y component of the oscillators as representative of each dynamical node. We generate $N = 3000$ samples for each time-series and consider the case of a chain of unidirectionally coupled oscillators ($K_{21} = 1$, $K_{32} = 1$, $K_{ij} = 0$ otherwise) with delays $\tau_{21} = 10$ and $\tau_{32} = 15$.

The goal of our connectivity analysis is to reveal the pattern of effective connectivity underlying this network of Lorenz systems. Fig. 3.7 (left column) shows that time-lagged MI shows two peaks roughly at the correct lags, therefore revealing the direction of information flow between nodes. However, the MI analysis alone is not able to tell whether the information shared between Φ_1 and Φ_3 is due to a direct interaction between the two systems, or due to an indirect connection mediated by Φ_2 . On the other hand, a PTE analysis (Fig. 3.7, right column) clearly shows that Φ_1 and Φ_3 are not directly connected and that the information shared by Φ_1 and Φ_3 is actually mediated by Φ_2 . Similar results can be obtained with the PMI [55].

An inherent limitation of PTE, PMI and, in general, of GC indices is that temporal precedence does not necessarily imply *effective* information exchange but simply time-lagged synchronized dynamics [57]. Such synchronization can be due to unidirectional coupling (effective connectivity) but can also be due to other synchronization mechanisms. To prove this point, we simulated the case of the second Lorenz system driving the other two systems, i.e. $K_{12} = 1$, $K_{32} = 1$, and $K_{ij} = 0$ otherwise. The delays are set to $\tau_{12} = 10$ and $\tau_{32} = 15$. The coupling in this case is strong enough so that the driving system (Φ_2) enforces *lag synchronization* between the two driven systems (Φ_1 and Φ_3) with a lag $\tilde{\tau}_{31} = \tau_{32} - \tau_{12} = 5$ time steps. This synchronization, which occurs without any direct effective connection between the driven systems, is robust to small differences in the parameters of the Lorenz systems and in the coupling factors (as long as the coupling is not too weak [141]). From a PMI perspective (Fig. 3.8, left column) the underlying coupling structure is $\Phi_1 \rightarrow \Phi_3$ with $\tilde{\tau}_{31} = 5$. On the other hand, a PTE analysis (Fig. 3.8, right column) reveals information transfer in the directions $\Phi_2 \rightarrow \Phi_1$ and $\Phi_1 \rightarrow \Phi_3$ with $\tilde{\tau}_{12} = 10$ and $\tilde{\tau}_{31} = 5$. Hence, neither the PMI nor the PTE alone can reveal the true underlying effective connectivity. However, a pairwise MI analysis (not shown here) reveals that the three systems share information between each other, therefore rejecting the possibility that Φ_2 is truly disconnected from the other

two oscillators, as suggested by the PMI analysis. Indeed, the PTE found a flow of information from Φ_2 towards Φ_1 . The fact that the PMI did not detect this flow can only be explained by the presence of an identical flow from Φ_2 towards Φ_3 . Ultimately, without making assumptions about the motion equations of the three systems, we cannot unequivocally tell whether the coupling $\Phi_1 \rightarrow \Phi_3$ is real or if it is just due to synchronization via a common source. Nevertheless, the principle of parsimony would tell that the most likely explanation for the observed PMI and PTE is a coupling pattern consisting of the links $\Phi_2 \rightarrow \Phi_1$ with $\tilde{\tau}_{12} = 10$ and $\Phi_2 \rightarrow \Phi_3$ with $\tilde{\tau}_{32} = 15$. This toy example supports the idea that MI, PMI and PTE offer complementary information about the underlying coupling structure and that a simultaneous MI/PMI/PTE analysis increases the chances of revealing the true effective connectivity pattern.

3.4.3 Gaussian processes with time-varying coupling

In order to demonstrate that (3.26) can be used to characterize time-varying coupling patterns, we simulated three non-linearly coupled Gaussian processes with a time-varying coupling factor:

$$\begin{aligned} x^{(r)}[n] &= 0.4x^{(r)}[n-1] + \eta_x \\ y^{(r)}[n] &= 0.5y^{(r)}[n-1] + \kappa_{yx}[n] \sin(x^{(r)}[n - \tau_{yx}]) + \eta_y \\ z^{(r)}[n] &= 0.5z^{(r)}[n-1] + \kappa_{zy}[n] \sin(y^{(r)}[n - \tau_{zy}]) + \eta_z \end{aligned}$$

where the index $r = 1, \dots, 50$ corresponds to the trial number and $n = 1, \dots, 1500$ denote time instants. The terms η_x , η_y and η_z are normally distributed noise processes, which are mutually independent across trials and time instants. The coupling delays were $\tau_{yx} = 10$, $\tau_{zy} = 15$ and the dynamics of the coupling were described by:

$$\kappa_{yx}[n] = \begin{cases} \sin(0.004\pi n) & \text{for } 250 \leq n < 750 \\ 0 & \text{otherwise} \end{cases}$$

$$\kappa_{zy}[n] = \begin{cases} \cos(0.004\pi n) & \text{for } 750 \leq n < 1250 \\ 0 & \text{otherwise} \end{cases}$$

Fig. 3.9 shows the time-varying PTEs obtained with the ensemble estimator of entropy combinations given in (3.26), and Fig 3.9 shows the corresponding results for the naive average estimator (3.25). The significance of PTE values (at each time-instant) was assessed using a non-parametric permutation

test [172]. Before PTE estimation, each time-series was time-delayed so that they had maximum mutual information with the destination of the flow. That is, before computing some $T_{a \leftarrow b|c}(n)$, the time-series b and c were delayed so that they shared as much information as possible with the future of time-series a . In addition, the PTE series shown in Fig. 3.9 and Fig. 3.10 have been smoothed with a moving average filter of order 20, as a simple way of reducing estimator variance. As long as the order of the filter is not too large, this post-processing does not affect crucially the temporal resolution of the PTE estimates. The latter is mainly determined by the size of the time window used in the nearest neighbor search (parameter σ in Fig. 3.1). Notice from Fig. 3.9 that the proposed ensemble estimator accurately characterized the coupling dynamics. On the other hand, the naive average estimator (Fig. 3.10) did not detect any significant information flow between the three Gaussian processes.

3.4.4 Mackey-Glass electronic circuits

In this last experiment we assess the performance of estimator (3.26) using voltage measurements obtained from two one-way coupled Mackey-Glass (MG) electronic circuits [134]. The drive system generated a chaotic signal thanks to a time-delayed feedback line [187]. A time-varying coupling factor was simulated by modulating the drive signal with a sinusoid of period 100 samples. The second MG circuit did not have any feedback and, therefore, its output voltage was determined by the input signal, by errors in the A/D and D/C blocks, by the presence of electrical noise and by its non-linear input-output transfer function. The experimental setup is illustrated in Fig. 3.11. This dataset was provided by M. C. Soriano from the Institute for Cross-Disciplinary Physics and Complex Systems at Palma de Mallorca (Spain).

Fig. 3.12 shows the results of the TE analysis for different lags between the outputs of the two MG circuits and for different time-instants. The TE reveals that, indeed, there is a strong transfer of information in the direction $\text{MG1} \rightarrow \text{MG2}$ around lag $\tau = 20$ samples. A zoomed view of the coupling dynamics for that lag (Fig. 3.12, bottom) shows a clear sinusoidal pattern with an approximate period of 100 samples. TE estimates for the opposite direction ($\text{MG1} \leftarrow \text{MG2}$) were non-significant for the vast majority of lags and time-instants. A linear analysis using lagged correlations in sliding time-windows did not reveal significant coupling at any lag. These results are encouraging and suggest that the time-adaptive estimator (3.26) is able to accurately characterize information transfer between real-life complex physical systems. Nevertheless, we have to admit that the experimental setup used here is rather artificial and still too simple to be able to draw conclusions on the expected performance on EEG time-series.

3.5 Conclusions to the chapter

In this chapter we have reviewed the most important information theoretic indices that can be used for characterizing information transfer between time-series and have introduced *partial transfer entropy* as a natural extension of bivariate transfer entropy (TE). Moreover, we have presented practical estimators of entropy combinations (e.g. MI, PMI or PTE) that can be used to identify time-varying coupling patterns from an ensemble of repeated measurements. Using simulated data we showed the benefits of these estimators with respect to simple averaging of individual trial estimates. We are currently investigating the application of these ensemble estimators to study the generation mechanisms of auditory event-related potentials in humans. Another topic for future research is the definition of automatic rules for adaptive selection of the temporal window size in the time-varying estimator of entropy combinations (i.e. the parameter σ in Fig. 3.1).

The two major advantages of non-parametric information-transfer measures over more traditional parametric approaches like those based on VAR models is that they are less affected by model misspecification and that are sensitive also to non-linear interactions. The downside is that estimation of information theoretic quantities is a delicate issue and large number of stationary data samples might be needed to obtain accurate estimates. This problem is partially overcome by the ensemble estimators proposed in this chapter, at the expense of requiring stationarity across a set of repeated measurements. In practice, perfect stationarity across time or across trials is unlikely, which raises concerns regarding the applicability of these indices to real electrophysiological datasets. A more accurate description of the relative strengths and weaknesses of these measures will require the analysis of both realistically simulated data and real ERPs. We are currently working on these two issues and our preliminary results suggest that the key to success lies in the pre-processing before information-theoretic indices are estimated. In particular, pervasive long-term correlations in EEG time-series can render impossible the reconstruction of their phase-space using simple delay embedding, or may require high-dimensional embedded vectors, which complicates the estimation from short duration EEG epochs. In agreement with the results in chapter 2, we have found that calculation of information-theoretic quantities is typically improved by diminishing long-range second-order temporal structure using VAR filters, provided of course that the interactions between time-series are not purely linear.

A practical pitfall of the PTE (and also of the PMI) and in general of multivariate indices is that partialization over a large number of variables is unfeasible, due to the difficulty of estimating high-dimensional pdfs. This means that PTE and PMI can only be applied to networks with few nodes (e.g. less

than 10). In the case of stationary EEG rhythms such small networks might be enough to accurately describe the underlying brain dynamics. However, this is obviously a limiting factor, if one aims to apply these techniques to fMRI data, or if one wants to compute information flow at the sensor level. A partial solution consists on using first the TE or MI to assess pairwise connectivity between all network nodes. Then PTE (or PMI) is used iteratively in subsets of network nodes (e.g. at most 10 nodes), in order to remove indirect connections. Consider for instance a network with K nodes (X_1, X_2, \dots, X_K) and that using pairwise TE we have obtained that the j th node receives inflows (direct or indirect) from the subset of nodes $\{X_l\}_{l \in \Gamma}$ with $\Gamma \subset [1, K]$. In order to determine the underlying effective connectivity pattern, one would like to assess whether the flow $X_j \leftarrow X_i$ with $i \in \Gamma$ is due to a true direct connection between the two nodes or the influence has been mediated by another intermediate node. When K is too large to directly apply the PTE, we suggest the following approximation:

$$\hat{T}_{X_j \leftarrow X_i | \{X_l\}_{l \in [1, K] - \{i, j\}}} \sim \min_m \left\{ \hat{T}_{X_j \leftarrow X_i | \{X_l\}_{l \in \Gamma_m \subset \Gamma - \{i, j\}}} \right\} \quad (3.29)$$

where $m = 1, \dots, \binom{|\Gamma|}{\alpha}$ and $\{\Gamma_m\}$ are all unordered subsets of Γ containing some $\alpha < |\Gamma|$ elements. That is, we try to determine whether the flow from X_i towards X_j can be explained by some of the other nodes sending inflows to X_j but partializing at a time only with respect to subsets of $\alpha < |\Gamma|$ nodes. By using a small value of α we make feasible the estimation of the PTEs at the right side of Eq. 3.29. There is no guarantee, however, that this approach will remove all indirect flows, but one would expect to remove at least some of them.

In summary, we can conclude that information theory provides a sound and flexible framework for assessing information exchange in dynamical systems. However, the theoretical advantages of indices like MI, PMI or PTE might be counterbalanced by the difficulties involved in their estimation from real EEG data, which is an issue currently under investigation. These difficulties might be partially overcome by careful pre-processing and the use of accurate estimators but, in any case, information-theoretic indices are likely to offer real advantages only in those cases where information exchange cannot be described by simpler parametric models of Granger causality.

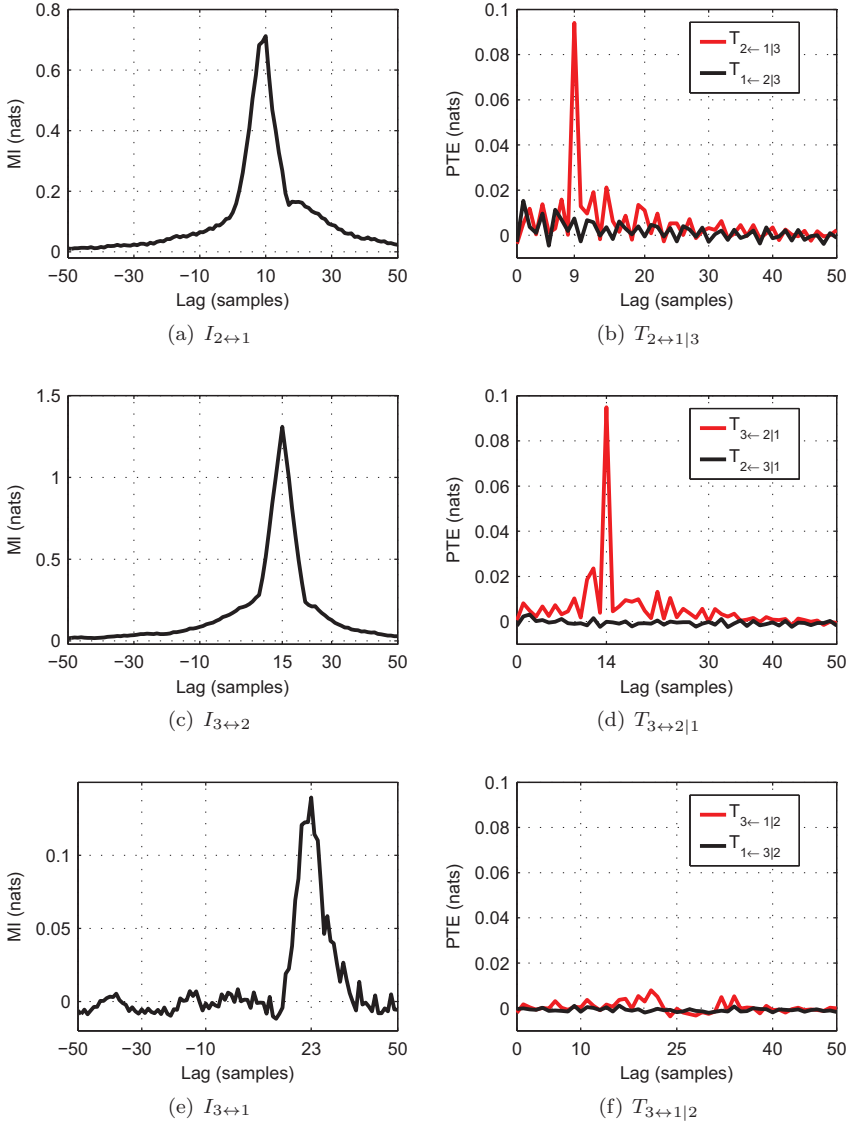


Figure 3.7: MI (left column) and PTE (right column) estimates for the simulation using a chain $\Phi_1 \rightarrow \Phi_2 \rightarrow \Phi_3$ of unidirectionally coupled Lorenz oscillators. In all cases, the number of nearest neighbors was $k = 20$. The figures show the average MI and PTE values obtained in 100 repetitions of the time-series, which were obtained using different random initializations of the Lorenz oscillators.

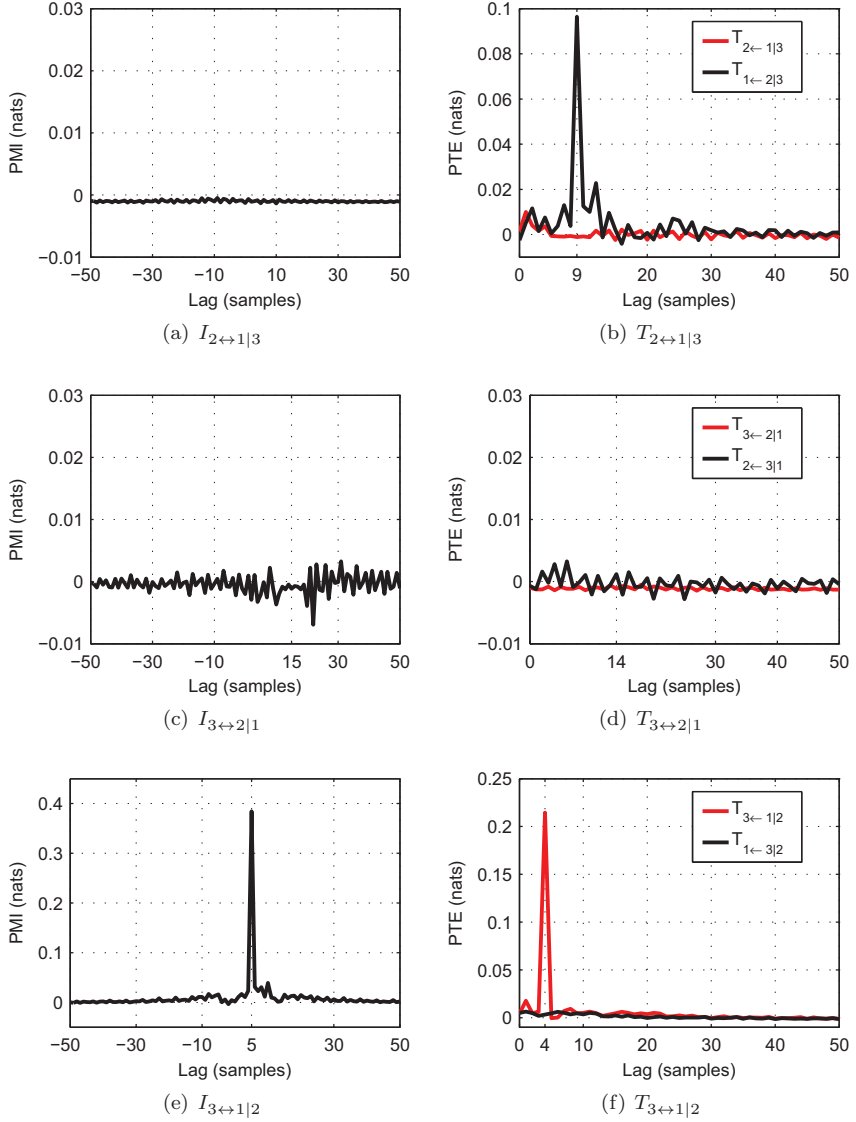


Figure 3.8: PMI (left column) and PTE (right column) estimates for the simulation of two identical Lorenz systems having a common drive (Φ_2). In all cases, the number of nearest neighbors was $k = 20$. The figures show the average MI and PTE values obtained in 100 repetitions of the time-series, which were obtained using different random initializations of the Lorenz oscillators.

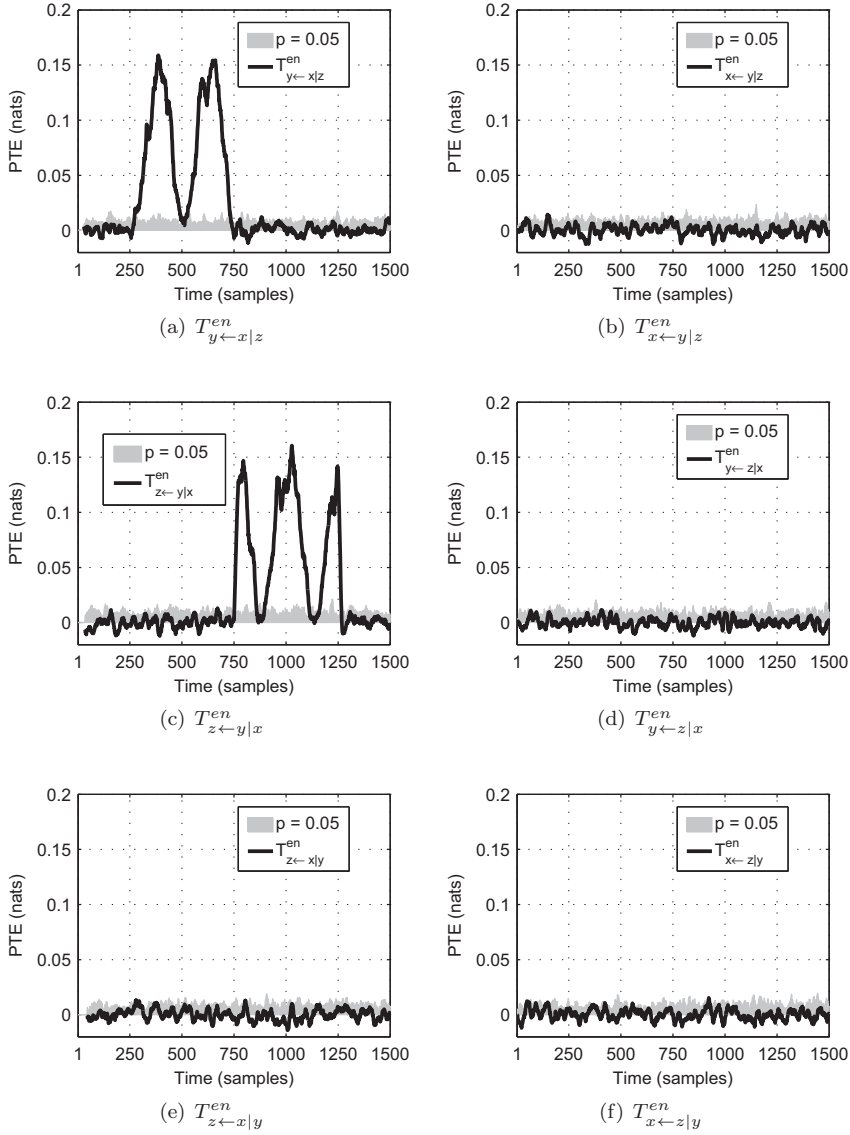


Figure 3.9: Time-varying PTE estimates obtained with the ensemble estimator (3.26) for the Gaussian processes with time-varying underlying connectivity. In all cases, the number of nearest neighbors was $k = 20$ and the window size was $\sigma = 5$ samples. The time-varying PTE estimates shown in the figures have been temporally smoothed using a moving average filter of order 20. The significance thresholds were obtained with a permutation test [172].

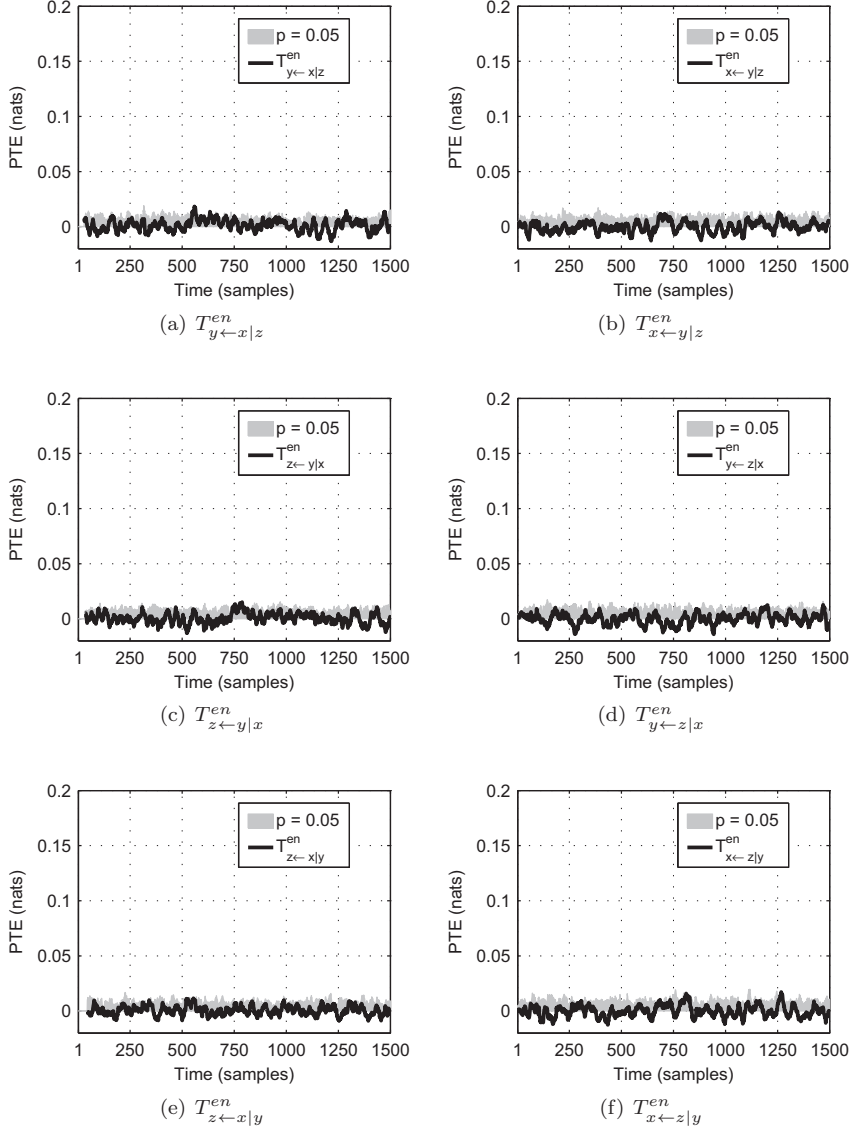


Figure 3.10: Time-varying PTE estimates obtained with the average estimator (3.25) for the Gaussian processes with time-varying underlying connectivity. In all cases, the number of nearest neighbors was $k = 20$. The time-varying PTE estimates shown in the figures have been temporally smoothed using a moving average filter of order 20. The significance thresholds were obtained with a permutation test [172].

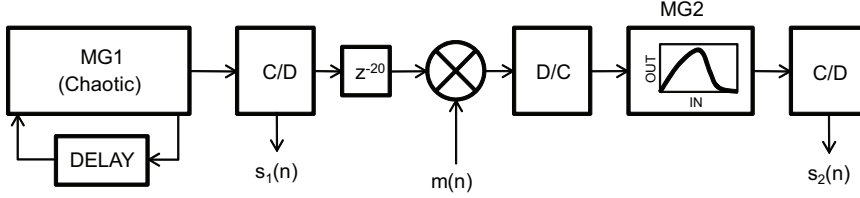


Figure 3.11: Experimental setup with two Mackey-Glass (MG) circuits unidirectionally coupled. The modulation signal was $m(n) = \frac{1}{2} + \frac{1}{4} \sin(\frac{2\pi}{T}n)$ with $T = 100$ samples. Fig. 3.12 below depicts the amount of information transfer from $s_1(n)$ to $s_2(n)$, as measured by the TE.

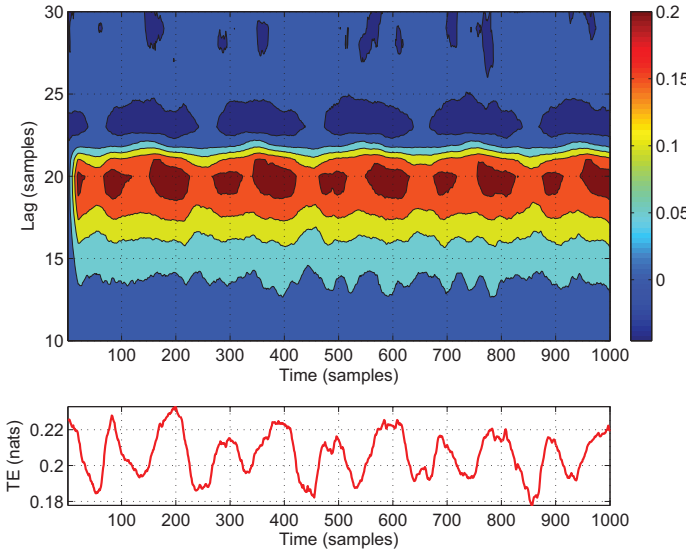


Figure 3.12: TE from the first electronic circuit towards the second. The upper figure shows time-varying TE versus the lag introduced in the temporal activation of the first circuit. Clearly, there is a directional flow of information time-locked at lag $\tau = 20$ samples, which is significant for all time-instants ($p < 0.01$). On the other hand, the flow of information in the opposite direction was much smaller ($T_{1 \leftarrow 2} < 0.0795$ nats $\forall(t, \tau)$) and only reached significance ($p < 0.01$) for about 1% of the (n, τ) tuples. Moreover, the location of the significant tuples were discontinuous and displayed a random pattern in the (n, τ) plane. The lower figure shows the temporal pattern of information flow for $\tau = 20$, i.e. $T_{2 \leftarrow 1}(n, \tau = 20)$, which resembles a sinusoid with a period of roughly 100 data samples.

Chapter 4

Measuring directional coupling between EEG sources

4.1 Introduction

Oscillatory synchronization between distant brain areas is a key mechanism underlying information processing in the human brain [218]. These systems-level interactions can be studied by means of (symmetric) synchronization measures but, in profusely interconnected cerebral systems, it is of major importance to characterize also the directionality of these interactions. That is, to be able to infer causal (drive-response) relationships between the studied subsystems. Analysis techniques based on Granger causality criteria identify causality by measuring how the history of a neural signal predicts the future of another [7, 73, 195]. A closely related concept consists in measuring to what extent a spectral component in a neural signal induces the generation of the same spectral component in another neural signal [49, 104]. To date, all these information-flow measures have been typically computed between scalp EEG signals recorded in humans [7, 82, 104, 119] and from intracranial recordings both in humans and animals [52, 54, 105]. Only recently they have started to be applied in brain source space [4, 9, 211].

EEG connectivity analysis at the sensor level is seriously compromised by the field spread caused by volume conduction effects. As was discussed in

chapter 1, a straightforward solution to these problems is to assess connectivity between a small number of equivalent intracranial current dipoles [81]. However, estimating the number, locations and orientations of multiple discrete neuroelectric dipoles is not easy, due to the high number of possible model configurations that fit well the measured scalp potentials. A different approach is to compute coherence between regions of interest (ROI) of a brain activation map, obtained either with a linear imaging method [124, 211], or with a beamformer [75]. All these approaches to brain connectivity analysis find the location of the cerebral EEG generators first and, only afterwards, estimate the corresponding temporal activations. As a result, their performance is very dependent on how accurate the EEG inverse solution is, since errors in this first analysis stage will propagate and probably ruin the whole analysis. At the same time, solving the EEG inverse problem is extremely difficult, if the underlying current source distribution does not exhibit some desirable properties, such as being spatially smooth and compact.

In this chapter we propose an alternative strategy to brain connectivity analysis with EEG. Instead of attempting to solve the EEG inverse problem directly, we use blind source separation (BSS) methods to separate the relevant brain sources into separate spatio-temporal components. Then, we separately map each of these components to brain space using a state-of-the-art inverse solver. BSS algorithms make mild assumptions about the mixing process but impose strong requirements on the statistical properties of the source signals. A prototypical example of BSS techniques is independent component analysis (ICA), which assumes that the sources are mutually independent non-Gaussian random variables (see [92] for a review). Multiple studies have shown the usefulness of ICA techniques to remove artifacts (e.g. [63, 102]), to separate physiological sources (e.g. [137]), and even to study directional connectivity between cortical areas [3, 9]. However, if we aim at determining causality relationships between different spatiotemporal EEG sources, we are implicitly assuming that those sources are functionally interrelated, which violates the assumption of independence made by ICA. Several methods have been previously proposed to overcome this pitfall [145, 157–159], but they allow only pairwise connectivity analysis and, with the exception of [159], they are not able to identify directionality.

4.2 Materials and methods

4.2.1 EEG model

The time-varying neural current density responsible for the scalp EEG potentials can be modeled by a discrete set of M signal generators with distinct spatial properties (see chapter 1 for further details). Let us denote by $\mathbf{s}(n) = [s_1(n), \dots, s_M(n)]^T$ the multivariate activation pattern of those generators at time instant n . We assume that source dynamics can be approximately described by a vector autoregressive (VAR) model of order p , at least for a certain time window $n = 1, \dots, L$. Then, we have that:

$$\mathbf{s}(n) = \sum_{\tau=1}^p \mathbf{B}_s(\tau) \mathbf{s}(n - \tau) + \mathbf{h}(n) \quad (4.1)$$

where $\mathbf{B}_s(\tau) \forall \tau = 1, \dots, p$ are the coefficient matrices of the VAR model and $\mathbf{h}(n) = [h_1(n), \dots, h_M(n)]^T$ represents the corresponding multivariate *innovation process*. We assume that each EEG generator is a source of *independent* activity in the sense that the elements of the residual vector $\mathbf{h}(n)$ behave like mutually independent random variables. Functional relationships between EEG sources are therefore exclusively caused by time-lagged axonal propagation of macroscopic neural behavior among distant regions of the brain (modeled by the coefficient matrices $\mathbf{B}_s(\tau) \forall \tau = 1, \dots, p$).

Several Granger-causality measures can be directly computed from estimates of $\mathbf{B}_s(\tau)$, as explained in the following section. However, scalp EEG sensors do not record the multivariate vector $\mathbf{s}(n)$ but a linear mixture of its components. Therefore, the EEG recorded at time instant n using K electrodes is a multivariate signal $\mathbf{x}(n)$ satisfying:

$$\mathbf{x}(n) = \mathbf{\Omega} \mathbf{s}(n) + \boldsymbol{\eta} + \mathbf{\Phi}_{noise} \boldsymbol{\epsilon}(n) \quad (4.2)$$

where $\mathbf{\Omega}$ is an unknown $K \times M$ matrix whose columns contain the spatial distribution of scalp potentials generated by each EEG source, $\boldsymbol{\eta} = [\eta_1, \dots, \eta_K]^T$ denotes the additive noise at each electrode (*measurement noise*), which is assumed to be white and Gaussian, and the term $\mathbf{\Phi}_{noise} \boldsymbol{\epsilon}(n)$ represents the contribution of noisy EEG sources (*biological noise*). By noisy sources we refer to temporally structured sources unrelated to the EEG features under study (e.g. unrelated to the generation of the EEG-alpha if this rhythm is

the target of the study). We further assume that $K \geq M$. Especially for high-density EEG recordings and stationary brain rhythms, it is relatively safe to assume that the number of electrodes is greater than the number of spatially distinct sources contributing to the scalp EEG. Furthermore, for the sake of simplicity, we neglect the contribution of measurement and biological noise in the derivations below. However, we will study the effects of noise in section 4.2.4 using simulations. By combining (4.1) and the noiseless version of (4.2), we obtain that the EEG measurements follow the VAR model:

$$\mathbf{x}(n) = \sum_{\tau=1}^p \underbrace{\mathbf{\Omega} \mathbf{B}_s(\tau) \mathbf{\Omega}^+}_{\mathbf{B}_x(\tau)} \mathbf{x}(n - \tau) + \underbrace{\mathbf{\Omega} \mathbf{h}(n)}_{\mathbf{v}(n)} \quad (4.3)$$

where $^+$ denotes Moore-Penrose pseudoinversion¹. Granger causality studies applied to human EEG signals have typically used estimates of $\mathbf{B}_x(\tau)$ to measure directional flows of macroscopic synaptic activity between *scalp EEG electrodes*, with the implicit assumption that causal relationships between electrodes imply functional connectivity between their respective underlying cortical regions. This is equivalent to assuming that $\mathbf{B}_x(\tau) \approx \mathbf{B}_s(\tau)$. Indeed, this assumption is not valid in general because the VAR model that best fits the observed EEG data might be strongly affected by volume conduction effects (matrix $\mathbf{\Omega}$ in Eq. 4.3).

4.2.2 Analysis procedure: VAR-ICA

In this section we describe the components of the proposed methodology for measuring source connectivity with EEG. Our approach, called *VAR-ICA*, is depicted in Fig. 4.1 and consists of the following steps:

Principal Component Analysis (PCA)

We first apply PCA on the measured scalp potentials because previous studies have demonstrated its ability to integrate brain activity spread across EEG leads, reducing the effects of measurement noise, and removing second-order instantaneous cross-correlations caused by volume conduction (see e.g. [179]). Even more important, PCA reduces the dimensionality of the data and avoids ill-conditioned covariance matrices, which results in a faster and more robust estimation of the VAR model in the next analysis step.

¹If \mathbf{M} is a real matrix of full column rank then $\mathbf{M}^+ = \mathbf{M}^T (\mathbf{M} \mathbf{M}^T)^{-1}$

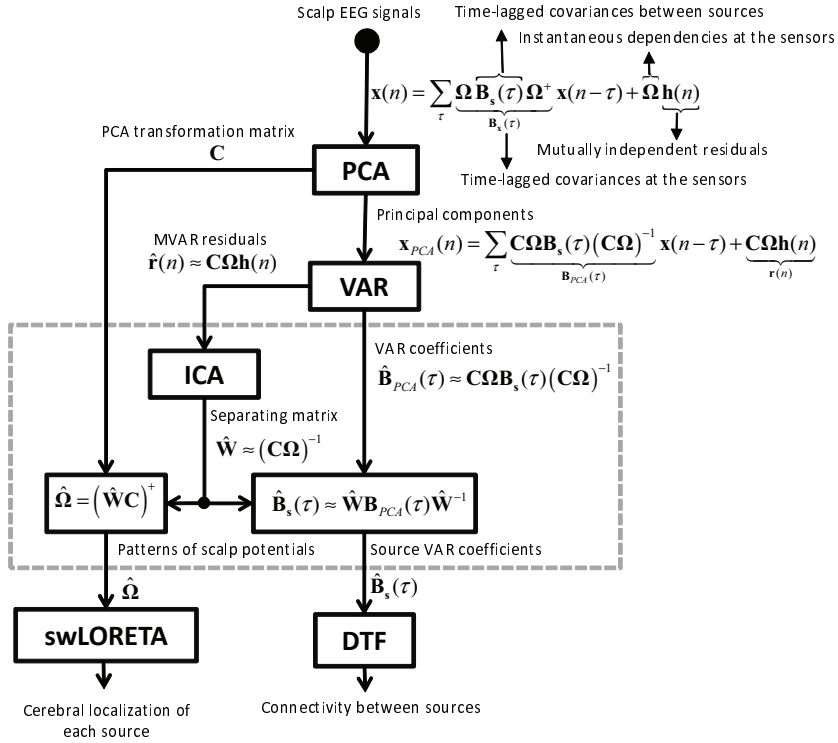


Figure 4.1: Block diagram of the proposed VAR-ICA methodology. The PCA and ICA blocks model instantaneous cross-dependencies between scalp EEG signals, which are exclusively caused by field spread due to volume conduction (matrix $\boldsymbol{\Omega}$ in the diagram). The PCA block takes care of second order correlations whereas the ICA block models dependencies of greater orders. On the other hand, the VAR block models time-lagged covariances between scalp EEG signals (matrices $\mathbf{B}_x(\tau)$ in the diagram), which are partially caused by time-lagged flows of activity between EEG sources (matrices $\mathbf{B}_s(\tau)$ in the diagram). By combining the PCA, VAR and ICA models it is possible to obtain an estimate of the leadfield matrix and of the coefficients of the VAR model driving the EEG sources. From the columns of matrix $\hat{\boldsymbol{\Omega}}$, swLORETA obtains the cerebral activation map for each EEG source separately. From the coefficients of the source VAR model it is straightforward to compute the DTF and other VAR-based information flow indices. The dashed rectangle identifies the blocks that are modified when ICASSO is introduced (see section 4.2.7 for more details).

PCA linearly transforms the scalp EEG signals into a set of mutually uncorrelated principal components. From (4.3) it obviously follows that the PCA-transformed data is also a VAR process:

$$\mathbf{x}_{PCA}(n) = \sum_{\tau=1}^p \underbrace{\mathbf{C}\boldsymbol{\Omega}\mathbf{B}_s(\tau)}_{\mathbf{B}_{PCA}(\tau)} (\mathbf{C}\boldsymbol{\Omega})^{-1} \mathbf{x}_{PCA}(n - \tau) + \underbrace{\mathbf{C}\boldsymbol{\Omega}\mathbf{h}(n)}_{\mathbf{r}(n)} \quad (4.4)$$

where \mathbf{C} is an $M \times K$ matrix implementing the PCA transformation. We suggest using as many principal components as necessary to reconstruct most (e.g. 99%) of the variance contained in the EEG signals.

Vector Autoregressive (VAR) modeling

Time-delayed cross-dependencies between the estimated principal components cannot be explained by instantaneous volume conduction effects. Consequently, they are likely to be of neural origin, i.e. caused by time-delayed axonal propagation of information between brain generators. To disclose the dynamical characteristics of these time-delayed cross-dependencies, a VAR model is fitted to $\mathbf{x}_{PCA}(t)$ using the algorithm ARfit [193]. We denote by $\hat{\mathbf{B}}_{PCA}(\tau) \forall \tau = 1, \dots, p$ the estimated model coefficients. The model order is automatically selected using Swartz's Bayesian Criterion (SBC) [198]. Traditionally, the order is determined by locating the minimum of the SBC as a function of model order (p values varied between 2 and 30 in our case). Instead, we select the order that guarantees that greater model orders do not significantly reduce the SBC, i.e. the order for which the reduction in the SBC has reached 90% of the maximum reduction achievable within the tested range of model orders. This modification is motivated by the fact that, for our EEG data, both SBC and Akaike's information criterion [1] dropped monotonically with increasing model order, which is in agreement with previous EEG studies [18, 211].

Independent Component Analysis (ICA)

The instantaneous higher-order cross-dependencies remaining in the residuals of the VAR model that were estimated in the previous step are likely to be of non-neural origin, i.e. caused by volume conduction effects². Based on this

²Assuming that there is not any instantaneous flow of activity between brain generators. This is a plausible assumption since instantaneous flows between brain EEG sources located more than few mm apart are unlikely, because of axonal propagations delays [53]. Despite

assumption, we use ICA to estimate an $M \times M$ matrix $\hat{\mathbf{W}} \approx (\mathbf{C}\mathbf{\Omega})^{-1}$ that minimizes the mutual dependencies between the components of the multivariate residual $\mathbf{r}(n)$ in (4.4)³. Namely, we use the ICA algorithm EfICA [110]. This choice is motivated by EfICA's robustness and computational efficiency and by the fact that, for the large sample sizes used in this study, EfICA's performance is excellent (see the numerical results in chapter 2). However, for studies that require shorter analysis windows, ENRICA [69] might be a more suitable choice. BSS algorithms that exploit lagged cross-correlations (e.g. M-COMBI [215] and F-COMBI [67]) are a bad choice since such second-order temporal structure is absent, or at least very diminished, in the residuals of the VAR model.

The idea of applying ICA to the residuals of a linear predictor has been proposed elsewhere [90,189], although in an univariate context. A combination of VAR modeling and ICA was previously used in the field of audio signal processing [31]. After VAR-ICA was introduced in [62] we have become aware of similar combinations of VAR models and ICA [94].

DTF computation between EEG sources

We use $\hat{\mathbf{W}} \approx (\mathbf{C}\mathbf{\Omega})^{-1}$ and model (4.4) to estimate the VAR parameters of the underlying EEG sources: $\hat{\mathbf{B}}_{\mathbf{s}}(\tau) = \hat{\mathbf{W}}\hat{\mathbf{B}}_{PCA}(\tau)\hat{\mathbf{W}}^{-1} \approx \mathbf{B}_{\mathbf{s}}(\tau)$. Then, the spatiotemporal spectral properties of the EEG sources can be uncovered by transforming (4.1) to the frequency domain:

$$\mathbf{S}(f) = \underbrace{\left(\mathbf{I} - \sum_{\tau=1}^p \mathbf{B}_{\mathbf{s}}(\tau) e^{-j2\pi\tau \frac{f}{f_s}} \right)^{-1}}_{\mathbf{T}(f)} \cdot \mathbf{H}(f) \quad (4.5)$$

where $\mathbf{H}(f)$ is the discrete Fourier transform of the residual process $\mathbf{h}(n)$, f_s is the sampling frequency and \mathbf{I} denotes the identity matrix. The transfer matrix $\mathbf{T}(f)$ describes transfer of spectral properties between EEG sources.

The directed transfer function (DTF) [104] from the j^{th} EEG source to the i^{th} source at certain frequency f is denoted by $\gamma_{ij}(f)$ and is defined as the

reports of zero-lag synchronization between distant brain areas (e.g. [180,181]), such instantaneous dependencies must necessarily involve time-lagged interactions with intermediate neural systems [224].

³Note that the permutation indeterminacy of ICA becomes irrelevant once the brain sources are localized in section 4.2.2.

ratio of influence of $s_j(t)$ on $s_i(t)$, with respect to the combined influence of $s_1(t), \dots, s_M(t)$ on $s_i(t)$. The DTF can be compactly expressed in terms of elements of the transfer matrix as:

$$\gamma_{ij}(f) = \frac{|T_{ij}(f)|^2}{\sum_{m=1}^M |T_{im}(f)|^2} \quad (4.6)$$

DTF values are in the interval $[0, 1]$. The DTF can also be defined for a band of frequencies, e.g. for the alpha-rhythm frequency band (7.5 – 12.5 Hz) by integrating the DTF across alpha frequencies: $\gamma_{ij}(7.5 \text{ Hz} < f < 12.5 \text{ Hz}) = \int_{7.5 \text{ Hz}}^{12.5 \text{ Hz}} \gamma_{ij}(f) df$.

Although the DTF is a good indicator of the total spectral influence from one electromagnetic source to another, it is noteworthy mentioning that knowledge of the VAR model underlying the source signals is all we need to compute other alternative measures of information flow in the frequency domain. An example is the partial directed coherence (PDC) [7] which, contrary to the DTF, allows to discriminate between direct and indirect flows between EEG sources. Both DTF and PDC have somewhat complementary advantages and limitations [49, 119]. Probably, the best approach would be to compute several such complementary measures simultaneously [49], which is a topic for further research. The DTF was chosen here only for illustration purposes and any other VAR-based information flow measure could be used in combination with the proposed VAR-ICA methodology.

Cerebral activation maps of EEG sources

Each column of matrix $\hat{\Omega} = (\hat{\mathbf{W}}\mathbf{C})^+ \approx \Omega$ contains the distribution of scalp potentials generated by a single EEG source and, therefore, it is everything that is needed to estimate its corresponding current source distribution. Practically any inverse method in the literature could be used to estimate the corresponding brain activation maps. Given that we are mainly interested on the EEG-alpha rhythm, we use swLORETA, which has shown a better performance when compared with other similar techniques for the reconstruction of deep EEG sources [165].

We used a realistic head model based on the 3D magnetic resonance (MR) images of the Collin's brain provided by the Montreal Neurological Institute (MNI). This model was built using the boundary element method and consisted of three tissue types (scalp, skull and brain) with conductivities 0.33, 0.0042

and 0.33. The location of the electrodes were obtained from the standard electrode coordinates included in the ASA analysis package (ANT software BV, Enschede, The Netherlands) [231] for the Collin’s brain template. The brain activation maps produced by the ASA software displayed probabilities of source activation based on Fisher’s F-test.

4.2.3 Alternative approaches to VAR-ICA

Most previous studies of effective brain connectivity with EEG have assumed that Granger causality between scalp EEG signals is equivalent to connectivity between the cortical regions underlying the corresponding electrodes, completely disregarding volume conduction effects and the presence of deep or non-normally oriented sources [7, 14, 105, 119]. If the locations of the EEG sources are known a priori, a conceptually equivalent approach is to fit a VAR model only to the signals recorded from *the electrodes closest to the underlying EEG sources*. In the following we refer to this traditional strategy as the *low-dimensional VAR* approach. Actually, low-dimensional VAR can be regarded as the optimal case of sensor-level DTF estimators since the locations of the underlying neuroelectrical sources are typically unknown in real experiments.

Astolfi et al. [3] used ThinICA [36] to remove volume conduction effects before estimating the PDC. The DTF and the PDC are both based on VAR modeling and therefore, the same approach can be directly extrapolated to the estimation of the DTF. The approach of Astolfi and colleagues differs from ours in two important factors. First, it applies ThinICA on the principal components directly, rather than on the residuals of the VAR model that fits best the principal components. Second, both ThinICA and EfICA are able to separate non-Gaussian i.i.d. sources but only ThinICA is also able to separate Gaussian sources with temporal structure and zero time-lagged cross-correlations. As will be shown later with simulations, the presence of time-delayed cross-correlations between EEG sources has an important negative effect on the accuracy of ThinICA and, therefore, on the estimation of the DTF. In the simulations we refer to the method of Astolfi and colleagues simply as ThinICA-VAR in order to stress the fact that VAR modeling is applied after ThinICA. Since the number of time-delayed covariance matrices used by ThinICA was not specified in [3], we tested two possibilities: ThinICA1-VAR, which uses only a single covariance matrix for time-lag 0 and ThinICA5-VAR, which uses covariances for time-lags 0, 1, 2, 3, 4.

4.2.4 Simulations

The proposed methodology was evaluated by studying the effects of the following simulation parameters: L (number of temporal data samples), d (depth of the source dipoles, using the head radius as reference unit), δ (orientation angle of the source dipoles, in degrees), SNR (signal-to-measurement-noise ratio, in dBs), SBNR (signal-to-biological noise ratio, in dBs) and α (Gaussianity of the residuals). Each tested set of values for those parameters resulted in a simulation instance. In order to assess average performance, each simulation instance was repeated 200 times with random spatial and temporal characteristics of the noise. In addition, any parameter that was not under evaluation during certain simulation instance was either fixed to some favorable values ($L = 38400$, SNR=15 dB, SBNR= 15 dB) or varied randomly within certain plausible ranges ($0.2 \leq d \leq 0.8$, $0^\circ \leq \delta \leq 90^\circ$, $1 \leq \alpha \leq 3$). This simulation set-up allowed us to assess the average effects of each tested parameter for an arbitrary combination of the remaining parameters.

We used a one-shell spherical head of normalized radius ($r = 1$). Inside this simulated head we located four *source dipoles* $\vec{\mathbf{q}}_j(t) = s_j(t) \cdot \vec{\mathbf{e}}_j \quad \forall j = 1, 2, 3, 4 \quad \forall t = 1, \dots, L$ with orientations $\vec{\mathbf{e}}_j = [e_{j,x}, e_{j,y}, e_{j,z}]^T$ and time-varying activations $s_j(t)$ following a VAR model as in (4.1). Both the residuals and the coefficients of the VAR model (the coupling strengths between sources) were random in each simulation surrogate, with the only requirement that the resulting VAR model had to be stable. In order to assess the amount of data needed by different techniques to reach their asymptotic performance, we ran simulations using $1280 < L < 38400$ samples. Fig. 4.2 (left) and Fig. 4.3 show the positions of the simulated electrodes ($K = 16$) and the positions and orientations of the simulated source dipoles ($M = 4$). As shown in Fig. 4.3, the spatial characteristics of the source dipoles could be varied using two parameters: the distance to the scalp (d) and the angle that they formed with their radial component (δ). The rotation angles of the dipoles around their radial components also affected the performance of the different methods but the effect of the angle δ was much more pronounced. Because of this, only δ was used to vary the orientation of the dipoles and the rotation angles of the dipoles were random in each simulation surrogate.

The residual processes driving the *source dipoles*, which are denoted by $h_1(n), \dots, h_M(n)$ in model (4.1), followed a generalized Gaussian distribution with probability density function (pdf): $p(h_i) \propto e^{-|h_i|^\alpha}$. By varying α between 1 and 3 we varied the Gaussianity of the residuals. If $\alpha < 2$, they showed a super-Gaussian distribution (acute peak around the mean and "fat" tails). If $\alpha = 2$ they were exactly Gaussian, and if $\alpha > 2$ they were sub-Gaussian (smaller peak around the mean and "thin" tails).

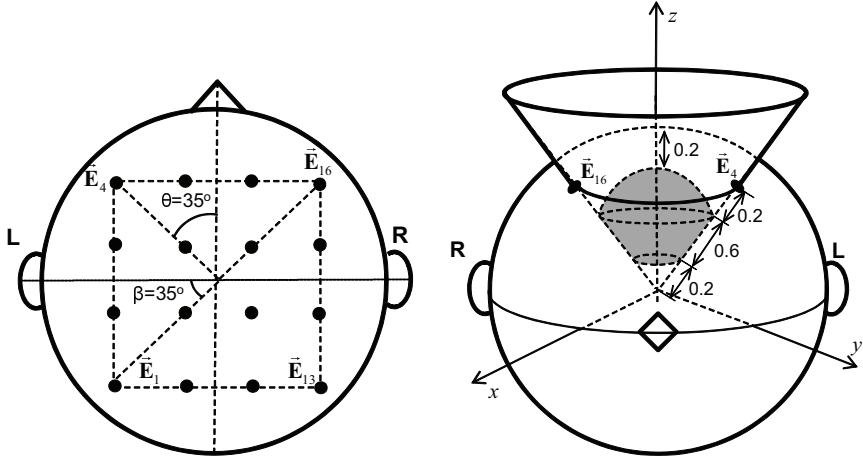


Figure 4.2: (Left) Location of the electrodes in the simulated head with normalized radius $r = 1$. The black-filled circles illustrate the positions of the electrodes. The position of the corner electrodes E_1, E_4, E_{13}, E_{16} were fixed. The other electrodes were placed uniformly in θ and β in such a way that the angular differences between two neighboring electrodes were $\Delta\theta = \Delta\beta = 70^\circ/3$. This electrode grid is similar to that used in [109]. (Right) Possible locations of the noisy dipoles. The valid locations (in gray) were within the volume generated by the intersection of the head sphere with a cone whose vertex is located in the center of the head and whose lateral surface contains the four corner electrodes E_1, E_4, E_{13}, E_{16} . Additionally, the noisy dipoles were required to be located at a minimum distance of 0.2 from the scalp and at the same minimum distance from the center of the head.

The signal-to-measurement-noise ratio (SNR) and the signal-to-biological-noise ratio (SBNR) were defined as the mean standard deviation of the signal across EEG channels, divided by the standard deviation of the Gaussian thermal noise introduced by the EEG sensors (*measurement noise*) and by the mean standard deviation of the *biological noise* across EEG channels, respectively. The biological noise was simulated using four *noisy dipoles* which were randomly located in each simulation surrogate within the volume depicted in Fig. 4.2 (right). The orientations of the noisy dipoles also varied randomly across simulation surrogates and their individual temporal activations were obtained with an autoregressive (AR) model of order 5 driven by white Gaussian residuals. Both the residuals and the coefficients of these AR models were randomized in each simulation surrogate. The simulation setup can be replicated using MATLAB code publicly available online at:

<http://www.cs.tut.fi/~gomezher/varica.htm>

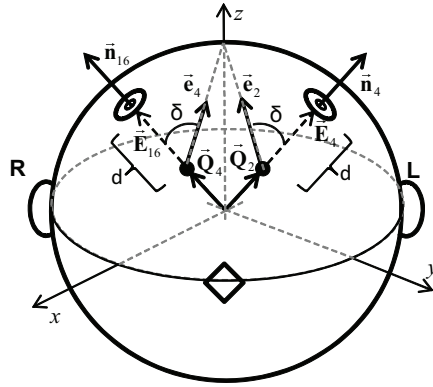


Figure 4.3: Anterior frontal view of the simulated head with the positions of the two frontal source dipoles (\vec{Q}_2 , \vec{Q}_4) marked with filled black circles and with the position of the two frontal corner electrodes (E_4 , E_{16}) marked with empty circles. The source dipoles were located in the radii that connected the center of the head with the corner electrodes. The distance d between the dipoles and the scalp was a free parameter of the simulations and its value was normalized according to the head radius ($r = 1$). The dipoles orientation vectors \vec{e}_2 , \vec{e}_4 formed an angle δ with their corresponding location vectors \vec{Q}_2 , \vec{Q}_4 . Their rotation angle around their radial component was random in each simulation surrogate. If $\delta = 0^\circ$ the dipoles were radial whereas if $\delta = 90^\circ$ they were tangential to the scalp surface. The positions and orientations of dipoles 1 and 3 were exactly symmetrical to those of dipoles 2 and 4 in respect of the midline coronal plane.

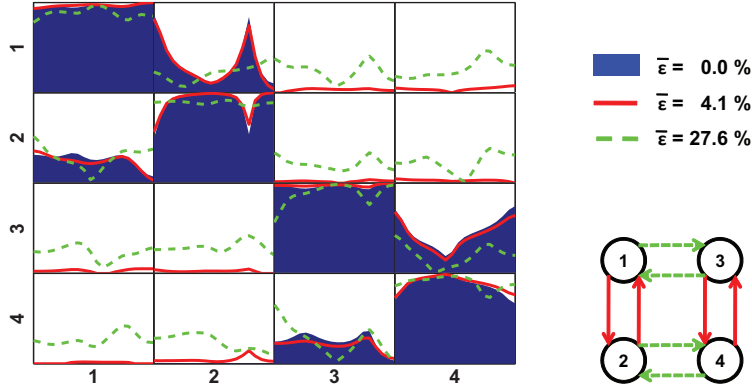


Figure 4.4: An illustration of the true absolute value of the DTF (blue filled areas), an accurate estimate (red solid line) and a poor estimate (green dashed line). The horizontal axis of each subplot is normalized frequencies and the vertical axis is absolute DTF values. The numbers indicate the source index. The connectivity graph in the lower right corner shows the true activity flows (red solid arrows) as well as the spurious flows corresponding to the low quality DTF estimate (green dashed arrows).

4.2.5 Assessing the accuracy of DTF estimates

The overall estimation accuracy for the source DTF was measured using the following index (in percentage):

$$\epsilon = 100 \cdot \frac{1}{M^2 N} \sum_{ijf} \sqrt{(\gamma_{ij}(f) - \hat{\gamma}_{ij}(f))^2} \quad (4.7)$$

where $\gamma_{ij}(f)$ denotes the true DTF between the i^{th} and j^{th} simulated dipoles, $\hat{\gamma}_{ij}(f)$ stands for the DTF estimate obtained with any of the tested methods, M is the total number of source dipoles (in our case, $M = 4$) and N is the number of frequency bins employed in the DTF analysis (in our simulations, $N = 10$ uniformly distributed bins). Since the DTF at a certain frequency is within the range $[0 \ 1]$, ϵ ranges from 0% (best case, no estimation error) to 100% (worst case, maximum possible estimation error). Fig. 4.4 shows two sample DTF estimates and their corresponding estimation error.

As was said in the previous section, each simulation instance (corresponding

to a set of values of the simulation parameters) was repeated 200 times. The results obtained from these simulation surrogates were summarized by means of two indexes:

1. The mean DTF estimation error, defined as $\bar{\epsilon} = \frac{1}{200} \sum_{i=1}^{200} \epsilon_i$, where ϵ_i is the DTF error obtained in the i^{th} simulation surrogate. Clearly, $\bar{\epsilon}$ varies between 0% (perfect DTF estimation in every surrogate) and 100% (worst possible estimation in every surrogate).
2. The p-value of the paired t-test for the null hypothesis $\bar{\epsilon}_0 < \bar{\epsilon}$, where $\bar{\epsilon}_0$ is the mean DTF estimation error obtained with the traditional low-dimensional VAR methodology and $\bar{\epsilon}$ is the mean DTF error for an alternative method. We denoted this performance index by $\bar{\epsilon}_{p-val}$. A very low value of this index allows to confidently reject the null hypothesis that the corresponding method does not perform better than the traditional approach.

A simulation framework such as the one described above is necessary to demonstrate the validity of VAR-ICA, that is, to demonstrate that if the assumed model is approximately fulfilled the connectivity analysis results will be valid. However, there is a large gap between simulations and real-life data and simulations alone can hardly be enough to assess the full potential of VAR-ICA. Thus, we will further validate VAR-ICA by analyzing real recordings of human EEG-alpha rhythm. Although little can be said for sure about the EEG generation mechanisms, EEG-alpha is one of the best studied EEG phenomena, allowing us to compare our findings with those of the previous literature.

4.2.6 EEG recordings and preprocessing

EEG recordings were obtained from 20 cognitively normal volunteers (10 females, mean age: 68.4 ± 6.1 yr). Inclusion criteria for this study were (i) no cognitive dysfunctions corroborated by neuropsychological exploration, (ii) clinical dementia rating global score of 0 (no dementia), and (iii) normal independent function both judged clinically and by means of a standardized scale for the activities of daily living. None of them had a history of neurological, psychiatric disorders and/or major medical illness.

EEG recordings were obtained between 9-10 AM in all participants in a relaxed wakefulness state with eyes closed. Vigilance level was constantly controlled based on EEG patterns characteristic of drowsiness [188]. EEG was

continuously acquired and referenced to linked mastoids from 59 scalp locations according to the International 10-20 system. Vertical ocular movements were recorded with a pair of electrodes placed above and below the left eye. The horizontal electrooculogram was acquired with another pair 1 cm apart from the outer canthus of each eye. Electrophysiological measurements were recorded with 10 mm diameter gold disk electrodes (Grass, Inc.). Electrode-scalp impedance was always kept below $5\text{ K}\Omega$. All electrophysiological variables were amplified (BrainAmp MR, Brain Vision), filtered (0.1 – 100 Hz bandpass), digitized (250 Hz, 16-bit resolution), and stored in digital format for off-line analysis.

EEG epochs containing prominent ocular, muscular and/or any other type of artifacts were manually identified and eliminated by an experienced researcher with expertise in human neurophysiology⁴. Residual ocular artifacts present in the remaining EEG epochs were corrected by adaptively regressing out the signals recorded at the peri-ocular electrodes [77]. A total of 150 s (37500 samples) of artifact-free EEG containing alpha rhythm were then available for each participant. The selected epochs were filtered within 6 – 13 Hz using a finite impulse response (FIR) filter of order 100. The average DTF for alpha-EEG frequencies was obtained as the average of the DTF values obtained in 10 equally spaced frequency bins within the alpha band (7.5 Hz - 12.5 Hz).

4.2.7 Reliability assessment

In the analysis of real EEG data, we used ICASSO [78,79] to assess the significance of the ICA estimates obtained in the VAR-ICA procedure. EfICA, FastICA and many other ICA algorithms involve stochastic optimization, which raises concerns about the repeatability and reliability of the analysis with real data [79,189]. ICASSO overcomes these concerns by identifying clusters of ICA estimates that are consistently found across random initializations of the ICA algorithm and across random bootstrap surrogates of the input data.

We introduced three modifications in the standard ICASSO procedure originally proposed in [79]. First, we used EfICA as ICA algorithm instead of FastICA, which was motivated by the generally better performance of the former, as reported in [110] and confirmed by our own numerical experiments. Second, ICA estimates were clustered according to the cross-correlation coefficients between their corresponding spatial patterns of scalp potentials. Third,

⁴This was done by Prof. Cantero from the laboratory of functional neuroscience at University Pablo de Olavide (Seville, Spain). All data acquisition and pre-processing described in section 4.2.6 was performed at Prof. Cantero's group.

we clustered together ICA estimates obtained from all available subjects. This latter modification allowed the detection of stable clusters of spatial components across the group of studied subjects. ICASSO was set-up to run EfICA 75 times on each subject's EEG with random initial conditions and with random bootstrap re-sampling. We used ICASSO's default agglomerative clustering with average-linkage criterion. The number of clusters in the data was automatically selected with the R-index [79]. Each cluster was uniquely represented by a single *centrotype* ICA estimate, which is just the estimate in the cluster that has the maximum sum of similarities to other points in the cluster. Only centrotypes of significant clusters were considered as valid ICA-estimates. A cluster was considered significant if it contained ICA estimates from at least 50% of the subjects (high inter-subject repeatability) and from at least 50% of the ICA-runs corresponding to those subjects (high intra-subject reliability).

Since the significant centrotype ICA estimates might have been obtained from different subjects and/or different ICA runs, we had to redefine the computation of the DTF between EEG sources and we had to find a unique matrix $\mathbf{\Omega}$ representative of the whole population of subjects. This was done by noting that the 1-dimensional temporal activation of the i th centrotype EEG source, denoted by $s_i^*(t)$, can be retrieved from the EEG measurements of the corresponding subject by:

$$s_i^*(t) = \mathbf{v}_i^T \mathbf{x}_{(i)}(t) \quad \forall i = 1, \dots, K \quad (4.8)$$

where \mathbf{v}_i^T is a spatial filter that corresponds to certain row of the matrix $\mathbf{V} = \mathbf{W}_{(i)} \mathbf{C}_{(i)}$ where $\mathbf{W}_{(i)}$ and $\mathbf{C}_{(i)}$ denote the ICA separating matrix and the PCA transformation matrix corresponding to the analysis run in which the i th centrotype was found. Similarly, $\mathbf{x}_{(i)}(n)$ denotes the EEG measurements of the subject from which the i th centrotype ICA-estimate was obtained. Then, the whole population of subjects can be characterized with R centrotype EEG sources that are defined by the set of spatial topographies $\hat{\mathbf{\Omega}}^* = [\boldsymbol{\omega}_1^*, \dots, \boldsymbol{\omega}_R^*]$ and the set of spatial filters $\mathbf{V}^* = [\mathbf{v}_1^*, \dots, \mathbf{v}_R^*]$. The former can be used to obtain the brain activity maps characterizing the whole population and the latter to estimate the temporal courses of the EEG sources in a single-subject basis. Using the temporal activations of the sources in each subject, we re-estimate the coefficients of the VAR model that best fit their mutual dynamics and compute the DTF for individual subjects. Lilliefors' test [125] was used to reject the null hypothesis that the time courses of the residuals of the EEG sources were normally distributed.

Since the distribution of DTF estimates obtained from a VAR model is not well established [105, 119], we used bootstrap surrogates to determine whether the obtained DTF estimates were statistically significant [105]. Recall that the DTF measures the ratio between the outflow from the EEG source j to the EEG source i , in respect to all the inflows to the destination EEG source.

Therefore, if $\gamma_{ij}(f) > \gamma_{ik}(f) \forall k \neq \{i, j\}$ then we can infer that the i th EEG source is *mainly driven by* the j th EEG source. However, for these inferences to be valid we need to define statistical tests able to reject the null hypothesis that $\gamma_{ij}(f) \leq \gamma_{ik}(f)$ for some $k \neq \{i, j\}$. Approaching this task analytically is complex due to (i) the highly non-linear relationship between data samples and DTF estimates and (ii) the interdependence between DTF values that were obtained from the same subject at the same frequency. Thus, we employed a numerical approach consisting of the following steps:

1. After the VAR-ICA analysis, we have estimates of all the unknowns in model (4.1), that is the order and the coefficients of the VAR model that best fit the centrotpe EEG sources: $\hat{\mathbf{B}}_s(\tau) \approx \mathbf{B}_s(\tau) \forall \tau = 1, \dots, p$ and $\hat{\mathbf{h}}(n) \approx \mathbf{h}(n) \forall n = 1, \dots, L$. By randomly shuffling time instants of the estimated multivariate residual process $\hat{\mathbf{h}}(n)$ we generated J surrogates of the EEG sources:

$$\hat{\mathbf{s}}^{(m)}(n) = \sum_{\tau=1}^p \hat{\mathbf{B}}_s(\tau) \hat{\mathbf{s}}^{(m)}(n - \tau) + \hat{\mathbf{h}}^{(m)}(t) \quad \forall m = 1, \dots, J \quad (4.9)$$

where $^{(m)}$ indexes the surrogates.

2. We re-estimated the coefficients of the VAR model for each surrogate: $\hat{\mathbf{B}}_s^{(m)}(\tau) ; \forall \tau = 1, \dots, p ; \forall m = 1, \dots, J$.
3. Using $\hat{\mathbf{B}}_s^{(m)}$, we re-estimated the DTF in the alpha band for each surrogate: $\gamma_{ij}^{(m)} ; \forall m = 1, \dots, J \forall i, j = 1, \dots, M$.
4. For each EEG source, we computed the differences between inflows of different origin:

$$\Delta\gamma_{ij,ik}^{(m)} = \gamma_{ij}^{(m)} - \gamma_{ik}^{(m)} ; \forall i ; \forall k \neq \{i, j\} \quad (4.10)$$

5. Finally we computed the maximum and minimum inflow differences:

$$\begin{aligned} \Delta\gamma_{ij,ik}^{MIN} &= \min_m \left(\Delta\gamma_{ij,ik}^{(m)} \right) ; \forall i ; \forall k \neq \{i, j\} \\ \Delta\gamma_{ij,ik}^{MAX} &= \max_m \left(\Delta\gamma_{ij,ik}^{(m)} \right) ; \forall i ; \forall k \neq \{i, j\} \end{aligned} \quad (4.11)$$

Only if both $\Delta\gamma_{ij,ik}^{MIN}$ and $\Delta\gamma_{ij,ik}^{MAX}$ were positive we were confident that, for the i th EEG source, the inflow from the j th source was proportionally larger than the inflow from the k th source since this was true for every surrogate.

4.3 Results

4.3.1 Simulations

Fig. 4.5 shows that VAR-ICA is reliable even for small data lengths, outperforming the traditional low-dimensional VAR approach with as few as 6400 data samples ($\epsilon_{p-val} < 0.001$). The poor performance of ThinICA-VAR is caused by the presence of time-delayed covariances between the source signals, in violation of the source model used by ThinICA. This explanation is supported by the fact that including more time-lagged covariance matrices did not improve the asymptotic performance of ThinICA-VAR but, actually, made it worse.

A major pitfall of the traditional low-dimensional VAR approach is its sensitivity to volume conduction effects. Namely, Fig. 4.6 shows that DTF estimation accuracy is seriously compromised even for very shallow cortical dipoles, if they were not radially oriented. On the other hand, VAR-ICA is considerably more robust to the presence of non-radially oriented dipoles. Fig. 4.6 also shows that we can safely reject the null hypothesis that VAR-ICA does not outperform the traditional approach.

From Fig. 4.7 we can conclude that VAR-ICA is largely undisturbed by the presence of measurement noise and that it compares favorably with respect to the traditional approach for low levels of biological noise. As expected, the latter type of noise was more problematic and all methods produced very poor DTF estimates for SBR values under 10 dBs. This result shows the relevant role of structured noise on the performance of brain connectivity analyses, suggesting that this type of noise should always be considered when evaluating new analysis methodologies.

VAR-ICA is quite robust to mild violations of the assumption of non-Gaussian residuals and VAR-ICA fails only when the residuals are almost perfectly Gaussian (see the sharp peak around $\alpha = 2$ in Fig. 4.8). On the other hand, ThinICA-VAR is much more affected by the distribution of the residuals and the hypothesis that the traditional approach performs better could only be rejected with confidence ($\epsilon_{p-val} < 0.001$) when the distribution was clearly super-Gaussian (see the fat tails around $\alpha = 2$ in Fig. 4.8).

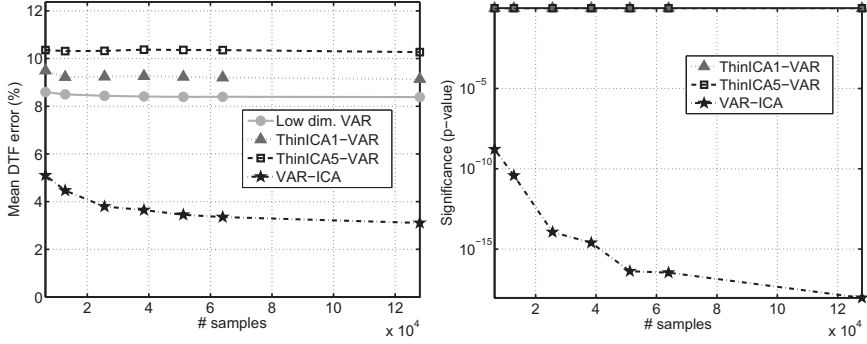


Figure 4.5: (Left) Mean DTF estimation error for the low-dimensional VAR approach, the two variants of ThinICA-VAR, and VAR-ICA, versus the number of simulated data samples. (Right) Probability (p-value) that low-dimensional VAR performed better (achieved a lower mean DTF error) than VAR-ICA and/or ThinICA-VAR.

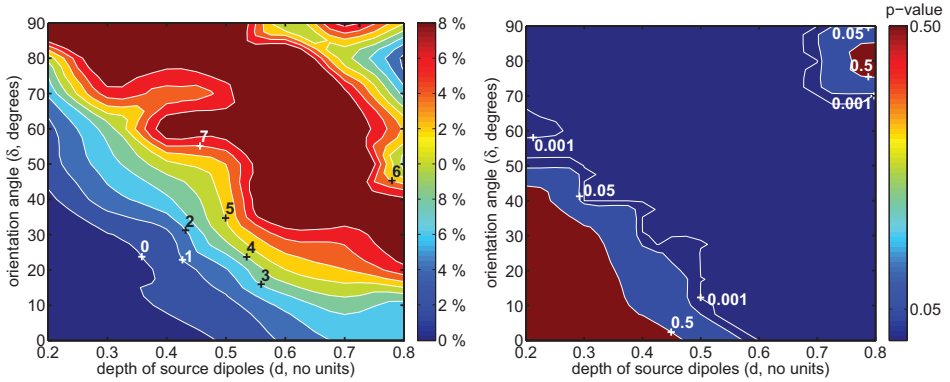


Figure 4.6: Volume conduction effects. (Left) The color scale shows the value of the difference $\bar{\epsilon}_0 - \bar{\epsilon}_{VAR-ICA}$, where $\bar{\epsilon}_0$ denotes the mean DTF error obtained with the traditional low-dimensional VAR approach and $\bar{\epsilon}_{VAR-ICA}$ denotes the error of the proposed VAR-ICA approach. The more positive this value, the greater the average improvement obtained with VAR-ICA. (Right) Probability (p-value) that low-dimensional VAR is more accurate than VAR-ICA, i.e. the probability of the event $\bar{\epsilon}_0 < \bar{\epsilon}_{VAR-ICA}$.

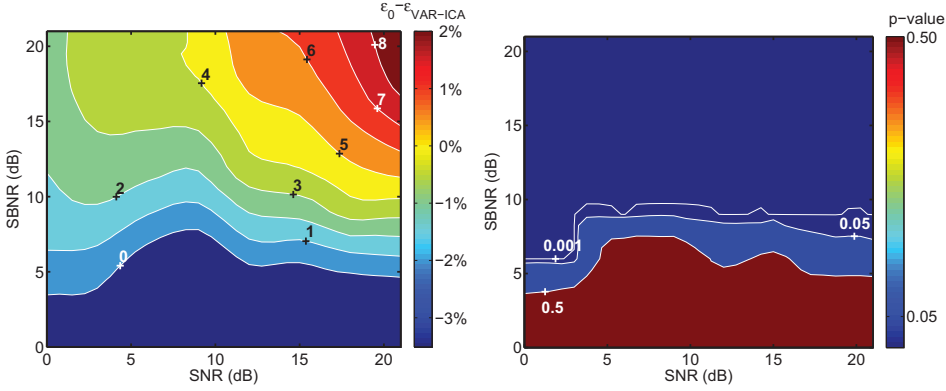


Figure 4.7: Effects of Gaussian thermal noise at the EEG sensors (SNR) and biological noise ($SBNR$). The interpretation of this figure is analogous to that of Fig. 4.6.

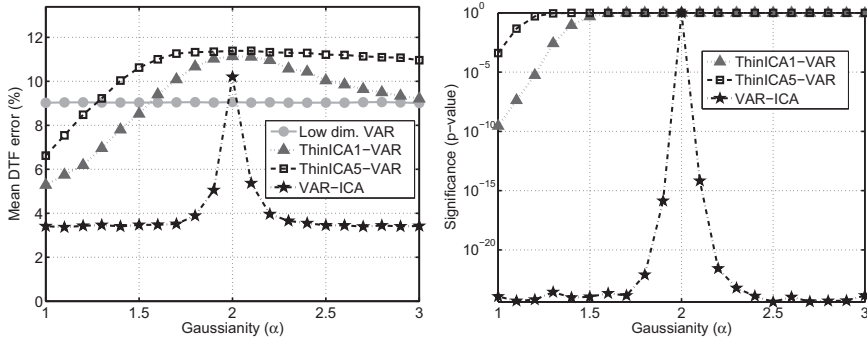


Figure 4.8: (Left) Effects of the Gaussianity of the VAR model residuals on the average DTF error. If $\alpha < 2$, the residuals showed a super-Gaussian distribution, if $\alpha = 2$ they were exactly Gaussian, and if $\alpha > 2$ they were sub-Gaussian. (Right) Probability (p-value) that the traditional low-dimensional VAR approach outperforms the alternative techniques.

4.3.2 EEG Alpha

Fig. 4.9 depicts the results of the ICASSO analysis applied to alpha EEG recordings from 20 healthy elderly subjects. The R-index suggested an optimal partition of 12 clusters in the data. Clusters 10, 11 and 12 were the only significant ones according to the criteria given in section 4.2.7 and contained 1125, 1050 and 825 ICA estimates, respectively (out of a total of 4425 estimates). Those three clusters exhibited a high repeatability within the same subject and across subjects. In particular, ICA-estimates within cluster 10 were obtained from 15 different subjects while 14 and 11 subjects contributed to clusters 11 and 12, respectively. All remaining clusters were much smaller and had high cross-similarities with one or more of the three major clusters. Clusters 10, 11 and 12 were selected for further analysis since they conveyed the most reproducible and stable features of the EEG-alpha rhythm.

The normalized scalp topographies corresponding to the representative centropypes of clusters 10, 11 and 12 are shown in Fig. 4.9 (right panel). According to the brain activation probability map obtained with swLORETA (see Fig. 4.10), the single electrical dipoles most likely to be generating each of those topographies were located in caudal regions of the thalamus (cluster 10), in the precuneus (cluster 11), and in the middle occipital gyrus, within the limits of the cuneus (cluster 12). Lillieford's test rejected the Gaussianity hypothesis for the centropype ICA estimates of clusters 10, 11 and 12 ($p < 0.01$).

Fig. 4.11 summarizes the results regarding directed flows between the alpha generators in the pre-cuneus (P), the cuneus (C) and the thalamus (T). There was a clear bidirectional link between the generation of EEG-alpha in the thalamus and the precuneus. EEG-alpha oscillations originating in the thalamus were mainly driven by the cuneus generator in 12 subjects ($p < 0.01$). By contrary, the thalamic source was mainly driven by EEG-alpha generation in the precuneus only in 1 subject ($p < 0.01$). Similarly, the thalamic source had a main effect on the generation of EEG-alpha in the cuneus in 12 of the subjects ($p < 0.01$) whereas the hypothesis that the inflow to the cuneus was larger from the precuneus than from the thalamus did not reach significance ($p > 0.01$) in any of the subjects. Moreover, the participation of the precuneus in the EEG alpha generation did not exert a major effect on either the thalamus or the cuneus, which rules out the possibility that the bidirectional flow between thalamus and cuneus might be due to indirect flows through the precuneus. Only in 1 subject the flow of EEG-alpha activity from precuneus to thalamus was significantly larger ($p < 0.01$) than the flow from the cuneus to the thalamus. Overall, the precuneus seemed to behave like a sink of EEG-alpha activity generated in the thalamus and/or the cuneus, whereas the major mechanism regulating EEG-alpha generation was a strong bidirectional causal feedback between thalamus and cuneus. The origin of the EEG-alpha activity inflow

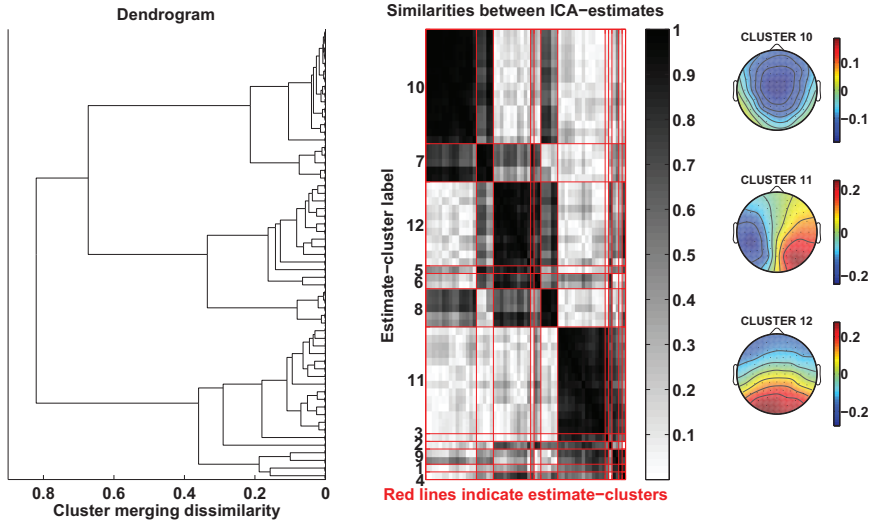


Figure 4.9: (Left) Dendrogram illustrating the arrangement in 12 clusters (as suggested by the R-index) of the ICA-estimates obtained with ICASSO. The horizontal axis represents the dissimilarity values at which clusters are merged at each possible partition level. The vertical axis indexes ICA-estimates. (Middle) Similarity matrix. The color scale indicates the cross-correlation coefficient between the scalp topographies of individual ICA-estimates. Clusters of ICA-estimates are indicated with red lines and their corresponding labels are depicted in the left vertical axis. (Right) Normalized distributions of scalp potentials corresponding to the centropotypes of clusters 10, 11 and 12, which are, by far, the largest and most compact.

to the precuneus is uncertain because the strong bidirectional link between thalamus and cuneus does not allow us to discard the possibility that the flow from thalamus to precuneus (from cuneus to precuneus) is actually caused by an indirect flow through the cuneus (thalamus). This issue could be clarified by incorporating to the analysis additional connectivity indices like the PDC or the partial transfer entropy proposed in chapter 3. This is a topic for future research.

4.4 Conclusions to the chapter

In this chapter we have presented a new methodology for estimating directional flows of activity between EEG sources. The major features of the proposed VAR-ICA approach are that (i) it removes spurious flows between scalp EEG signals due to volume conduction effects and (ii) it does not make a-priori

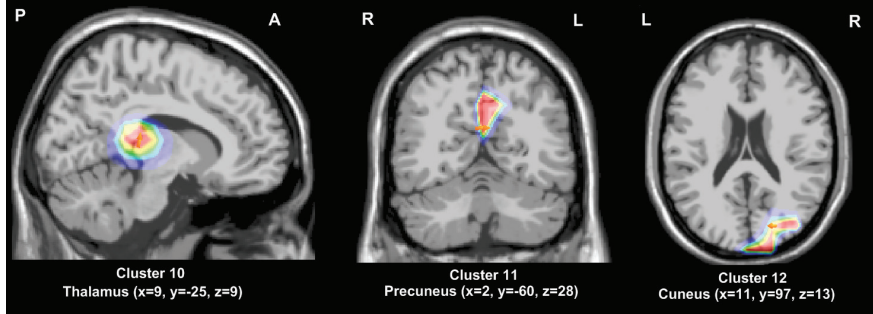


Figure 4.10: Localization obtained with swLORETA of electric dipole sources for the scalp distribution of alpha oscillations associated with clusters 10, 11 and 12 (thalamus, precuneus and cuneus, respectively).

assumptions about the cerebral localization of the underlying EEG generators.

An advantage of VAR-ICA is that it allows solving the inverse EEG problem separately for each source. It has been previously reported that, by localizing each independent source of activity separately, the localization error can be significantly reduced [213]. In this context, we have to admit that, although the conductivity values of scalp, skull and brain that we used can be considered as a de-facto standard in head modeling, they might be far from the true values [71]. This problem could be overcome by measuring tissue connectivity *in vivo* for each subject. One possibility is to relate tissue conductivity with water diffusion maps obtained with diffusion-tensor imaging (DT-MRI) [183]. However, we discarded this idea in the present study because it would have meant a qualitative increase in the complexity of the analysis.

The poor performance of ThinICA-VAR is mainly explained by the fact that high-order ICA contrasts are very disturbed by the presence of time-lagged cross-correlations between the sources. Such correlations violate not only the premise of independent sources but also the assumption that the sources lack temporal structure. VAR-ICA gets around this problem by applying ICA on the residuals of the VAR model, which are (ideally) free of second-order temporal structure and still contain the same instantaneous spatial dependencies as the original source signals.

Only few source space studies have attempted to provide a global pattern of directional connectivity across a population of subjects [3,211]. In this chapter, ICASSO was used to integrate the information obtained from several subjects, in order to provide a concise and simple description of the whole population. This comes at the cost of requiring that the number of active EEG sources and their cerebral activation maps are similar across subjects. This is probably the

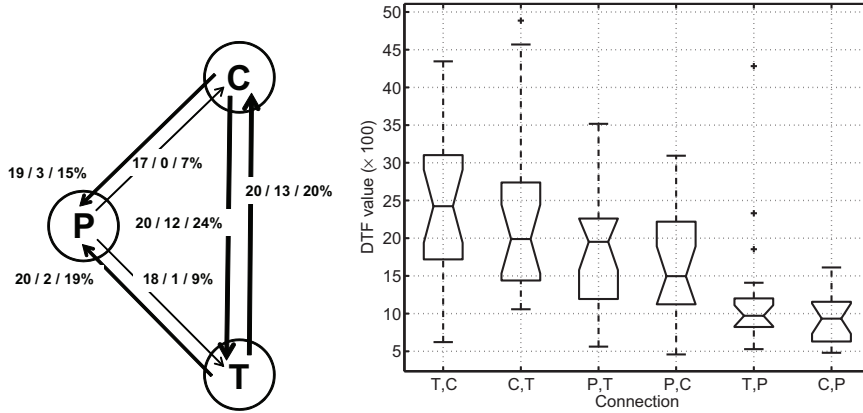


Figure 4.11: (Left) Causal flows between EEG-alpha generators in the precuneus (P), cuneus (C) and thalamus (T). Each flow is characterized by three numbers. The first number corresponds to the number of subjects for which the directed flow was significant ($p < 0.01$). The second number is the number of subjects for which the flow was identified as the most important inflow to the destination EEG source ($p < 0.01$). The third number is the median DTF value across subjects. Based on the results, the directional flows were ranked according to their qualitative significance into three groups (identified by different line widths in the diagram): (i) the bidirectional flow between thalamus and cuneus, (ii) the inflows to the precuneus originated in cuneus and thalamus, (iii) the outflows from precuneus to cuneus and thalamus. (Right) Spread across subjects of DTF values corresponding to directional flows of activity between EEG-alpha generators in the precuneus (P), the cuneus (C) and the thalamus (T). The horizontal axis depicts the six possible directional flows, with X,Y meaning an inflow to generator X originated in generator Y. The vertical axis represents corresponding DTF values (in percentages). The lines within the boxes indicate the lower quartile, median and upper quartile values. The lines extending from each end of the boxes show the extent of the spread of DTF values across subjects. The notches in each box represent a robust estimate of the uncertainty about the medians of each box.

case of EEG-alpha, given the positive results obtained here. However, the same might not hold for more complex EEG patterns and/or more heterogeneous subject populations.

The proposed analysis methodology disregards the possibility of instantaneous flows of information between cerebral EEG generators. Additionally, the innovation processes driving each EEG source are assumed to be non-Gaussian. Instantaneous synaptic communication between neuronal populations located more than few mm apart are unlikely, due to axonal propagation delays [53]. Although there is no fundamental reason to believe that the innovation process is non-Gaussian, several previous studies have also found meaningful non-Gaussian sources of brain activity (e.g. [3, 85, 137]), suggesting the existence of non-Gaussian generators in the brain. A limitation of VAR-ICA is the use of VAR models to describe brain dynamics. If the model does not fit well the data or if the model parameters cannot be accurately estimated, VAR-ICA will fail to produce reliable results. Indeed, this pitfall is shared with all connectivity analyses that are based on VAR models.

Brain oscillations in the range of alpha activity are one of the fundamental electrophysiological phenomena of the human EEG. This brain activity can be easily identified by its topographic distribution (maximum amplitude over parieto-occipital regions), frequency range (8–13 Hz), and reactivity (it suffers a dramatic amplitude attenuation with the opening of the eyes [95]). The study of alpha oscillations has generated a vast amount of literature related with physiological, maturational, clinical, and cognitive aspects [156, 174, 196].

Highly coherent alpha oscillations with significant phase shifts have been observed in both the visual cortex and the lateral geniculate nucleus in non-human mammals [28, 131, 133], supporting the involvement role of thalamo-cortical circuits in the generation of waking-alpha rhythm. Neocortical neurons located in the layer V of the occipital cortex seem to be intrinsic alpha generators, as revealed by results from *in vitro* preparations [202] and *in vivo* recordings [132]. They may receive thalamic inputs in order to maintain activation of cortical columns at an optimal level depending on the brain activation state. The number and exact location of alpha generators remain, however, unclear.

From the analysis of the EEG-alpha rhythm recorded from 20 volunteers under resting conditions, we found that the bidirectional feedback between thalamus and cuneus was crucial in the EEG-alpha generation. The precuneus seemed to play a secondary (or independent) role and was not the source of any causal inflow neither to the thalamus nor to the cuneus. This finding is consistent with [194] reporting a positive correlation between EEG-alpha power and metabolism of the lateral thalamus, as well as occipital cortex (cuneus)

and adjacent parts of the parietal cortex (precuneus) in humans. Our results also revealed that thalamocortical synaptic transmission remained alike from thalamus to cortex and vice-versa, which is in agreement with neural simulations showing that bidirectional coupling between distant brain areas engenders strong oscillatory activity [39]. These findings, together with results from human studies employing 3D equivalent dipole modeling [10, 98, 197], support the notion that complex interactions between local and non-local EEG sources, instead of a single or multiple isolated neural generators, are responsible for the genesis of the human alpha rhythm [161].

In conclusion, in this chapter we presented a novel methodology for measuring directed interactions between EEG sources. The proposed approach is based on well-established techniques such as VAR modeling, ICA, clustering and swLORETA. Simulated experiments showed improved robustness and accuracy with respect to more traditional approaches. We further evaluated the validity of our method using EEG recordings of alpha waves from a set of 20 control subjects. The proposed technique estimated current source distributions and directed flows of brain activity in agreement with the most recent findings about the generation mechanisms of the alpha rhythm in humans.

Chapter 5

The connectivity profile of neurodegeneration

5.1 Introduction

Cognitive impairment leading to dementia is among the most important hazards associated with population aging in developed countries. The European Community Concerted Action on the Epidemiology and Prevention of Dementia group (EURODEM) has estimated that 5.3 million people are affected by dementia in Europe, being Alzheimer's disease (AD) its most prevalent form. The annual cost attributable to dementia care across Europe are probably no less than 90 billion euros [232]. These overwhelming financial costs and the dramatic effects on the well-being of dementia patients and their families call for recognizing dementia as a major health scourge.

Functional integration of spatially distributed brain regions is a well-known mechanism underlying various cognitive and perceptual tasks. Indeed, mounting evidence suggests that impairment of such mechanisms might be the first step of a chain of events triggering neurodegeneration in AD and other neurological disorders. Namely, it has been recently confirmed that synaptic dysfunction rather than cell death is the pivotal event in AD initiation [166,200], which has raised novel theories about the key mechanisms underlying cognitive deterioration. For instance, it has been postulated that accumulation of certain neurotoxins (Beta-amyloid protein) in the brain may trigger a process of synaptic reorganization which in turn would increase the vulnerability of

healthy neurons to those neurotoxins [204]. New evidence obtained in a rodent model of AD confirms such a redistribution of synaptic drive in early stages of AD [19]. Thus, in-vivo assessment of systems-level connectivity in the brain might be the key to new breakthroughs in our understanding of neurodegeneration and specifically of AD.

Cell death and synaptic loss caused by AD affects mainly association areas and the pyramidal cells that supply long projections among distant neocortical regions [151]. This neuropathological pattern results in a global disruption of long-range neural circuits that is revealed by changes in brain oscillations recorded with the electroencephalogram (EEG). Namely, abnormal EEG activity in the alpha band has been repeatedly reported in AD [5, 84, 162]. In the generation of EEG alpha rhythm, the thalamus plays a crucial role by means of intrinsic mechanisms [86] and dynamical interactions of thalamocortical networks [129]. Postmortem studies in AD patients have revealed amyloid deposits and neurofibrillary tangles in the thalamus [17, 142] as well as significant loss of its gray matter [107], postulating a mild involvement of this subcortical structure in neurodegeneration. Moreover, it has been reported [44] that human alpha response of the thalamus precedes that of the cortex, predicating the existence of causal flows of information (in the Granger sense) between the neural generators of alpha rhythm. Damage of cortical pyramidal cells and thalamic neurons might affect the pattern of communication between cortex and thalamus during the generation of alpha rhythm even in preclinical stages of AD. The analysis methodology presented in chapter 4 is used in this chapter to test this hypothesis by assessing functional connections between thalamic and cortical EEG-alpha sources in patients suffering mild cognitive impairment (MCI), a condition considered to be a transitional stage between normal aging and AD [173].

5.2 Methods

5.2.1 Subjects

The data consisted of EEG recordings from twenty MCI patients (nine females, mean age: 66.8 ± 4.7 years) and from twenty cognitively normal volunteers (ten females, mean age: 68.4 ± 6.1 years). The latter were the same subjects that were analyzed in chapter 4. Patients with MCI and controls were matched in educational years and handedness. The demographics and neuropsychological profile of the participants of this study are summarized in Table 5.1. The recruitment and selection of subjects as well as all neuropsychological tests were performed by researchers from the Laboratory of Functional Neuroscience

	Controls (n=20)	MCI (n=20)	P
Age, years (mean \pm SD)	66.8 \pm 4.7	68.4 \pm 6.1	< 0.3
Gender (F/M)	10/10	9/11	N/A
Education, years (mean \pm SD)	13.5 \pm 5.6	11.8 \pm 6.5	< 0.4
MMSE (mean \pm SD)	28.4 \pm 1.4	26.6 \pm 2.7	0.01
CDR (sum of boxes)	0	0.5	N/A
Immediate recall (mean \pm SD)	14.4 \pm 2.9	9.9 \pm 2.2	< 10 ⁻⁵
Delayed recall (mean \pm SD)	13.3 \pm 2.4	6.4 \pm 3.5	< 10 ⁻⁷

Table 5.1: Subject demographics. F, females; M, males; MMSE, mini-mental state exam (where the range from best to worst performance is 30 – 0); CDR, clinical dementia rating (where CDR=0 no dementia, CDR=0.5, questionable or very mild dementia); N/A, not applicable.

at the University Pablo de Olavide (Seville, Spain) and by the Dementia Unit of the Neurology Service at the University Hospital Virgen del Rocío (Seville, Spain).

The diagnosis of MCI was based on consensus criteria [173]: (i) subjective memory complaints confirmed by the informant, (ii) objective memory decline on neuropsychological tests evidenced by scores greater or equal than 1.5 standard deviations below the age-appropriate mean, (iii) clinical dementia rating (CDR) global score of 0.5 (questionable dementia), (iv) normal independent function both judged clinically and by means of a standardized scale for the activities of daily living, and (v) not meeting DSM-IV criteria for dementia. Depression was excluded by clinical interview and the Geriatric Depression Scale (GDS) of Yesavage (shorter form). The cutoff to be included in the study was 0 – 5. The diagnosis of MCI was finally based on a clinical consensus following evaluation in the Dementia Unit by a senior neurologist and a clinical neuropsychologist.

Inclusion criteria for the healthy elderly group were: (i) no subjective memory complaints corroborated by neuropsychological exploration, (ii) CDR global score of 0 (no dementia), and (iii) normal independent function both judged clinically and by means of standardized scale for the activities of daily living. None of them had history of neurological, psychiatric disorders and/or major medical illness.

To avoid interference on EEG recordings and neuropsychological performance, the uptake of pharmacological compounds known to significantly affect any cognitive domain was considered cause of exclusion, both in healthy controls and patients with MCI. Individuals with medical conditions that may

affect brain structure or function were also excluded.

5.2.2 EEG recordings and pre-processing

Continuous EEG recordings were obtained at the facilities of the Laboratory of Functional Neuroscience of University Pablo de Olavide with all participants in relaxed wakefulness and eyes closed between 9 and 10 AM. Vigilance level was constantly controlled to avoid intermittent alpha oscillatory behavior caused by the drowsiness state. EEG was referenced to linked mastoids from 59 scalp locations according to the International 10-20 system¹. Vertical and horizontal ocular movements were also simultaneously recorded. Electrode-scalp impedance was always kept below 5 k Ω . All electrophysiological variables were amplified (BrainAmp MR, Brain VisionVR), filtered (0.1-100 Hz bandpass), digitized (250 Hz, 16-bit resolution), and stored in digital format for off-line analysis. EEG epochs containing prominent ocular, muscular, and/or any other type of artifacts were manually identified and eliminated. A total of 150 s of artifact-free EEG containing alpha rhythm were then available for each participant. The selected epochs were filtered within 6-13 Hz using an FIR bandpass filter.

5.2.3 Connectivity analysis

In order to allow for comparisons between two different populations of subjects, we had to assume that the spatial characteristics of EEG sources were approximately the same in both populations. This is plausible since neuronal loss in MCI patients is mild and a drastic spatial reorganization of the current source distribution underlying EEG-alpha is unlikely. Based on this premise, we identified stable clusters of EEG-alpha sources in the control population and characterized each cluster using a single *centrotype* source (see chapter 4 for details). The final outcomes of VAR-ICA were:

- The distribution of scalp potentials for the centrotypes of three significant clusters of EEG-sources: $\omega_1, \dots, \omega_3$. These distributions were used by swLORETA to obtain a brain activation probability map for each source.
- Three spatial filters retrieving the time-course of the centrotype EEG sources from their corresponding subject and ICA run: $\mathbf{v}_1, \dots, \mathbf{v}_3$.

¹The locations were Fp1, Fp2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, and O2.

For an arbitrary subject (control or MCI), an estimate of the activation of the i th centrotpe source was obtained as $\mathbf{s}_i(n) = \mathbf{v}_i^T \mathbf{x}(n)$, where $\mathbf{x}(n) = [x_1(n), \dots, x_{59}(n)]^T$ are the time-courses of the potentials at each scalp sensor. Subsequently, a VAR model was fitted to these sources using ARfit [195] and, based on the VAR model coefficients, the normalized directed transfer function (DTF, [105]) between sources was computed for both healthy elderly subjects and patients with MCI. The steps of the connectivity analysis are illustrated in Fig. 5.1.

5.2.4 Alpha peak frequency

The alpha peak frequency was computed for the temporal activation of each EEG-alpha source. Spectral power was estimated using the Welch's periodogram (4-s segments, 1 Hz resolution, 50% overlapping and Hanning windowing) as implemented in MATLAB v.7.4.0. Alpha peak frequency was identified as the maximum spectral power value within the defined alpha band (7.5-12.5 Hz).

5.2.5 Statistical analysis

The statistical significance of the normalized DTF values was assessed using data surrogates as described in section 4.2.7. DTF values that did not reach the $p < 0.01$ threshold were considered not significant and removed from the analysis.

Alpha peak frequency, probability of source activation, and directionality of the synaptic flow were evaluated with two types of mixed ANOVAs: (i) without including covariates; and (ii) with scores provided by the Mini Mental State Examination (MMSE) and age as covariates (if they were significant; otherwise this ANOVA was not performed). Each covariate was tested independently for significance and only significant covariates ($p < 0.05$) were included in the final ANOVA. Mauchly's W was computed to check for violations of the spherity assumption. When Mauchly's test was significant, the Greenhouse-Geisser correction was applied to all repeated-measures analysis. Homogeneity of variance was evaluated with Levene's test. The Student-Newman-Keuls or the Games-Howell procedures were applied for multiple post hoc comparisons depending on population variances.

To confirm previous reports on alpha slowing in patients with MCI, alpha peak frequency was evaluated with a mixed ANOVA including, as the within-

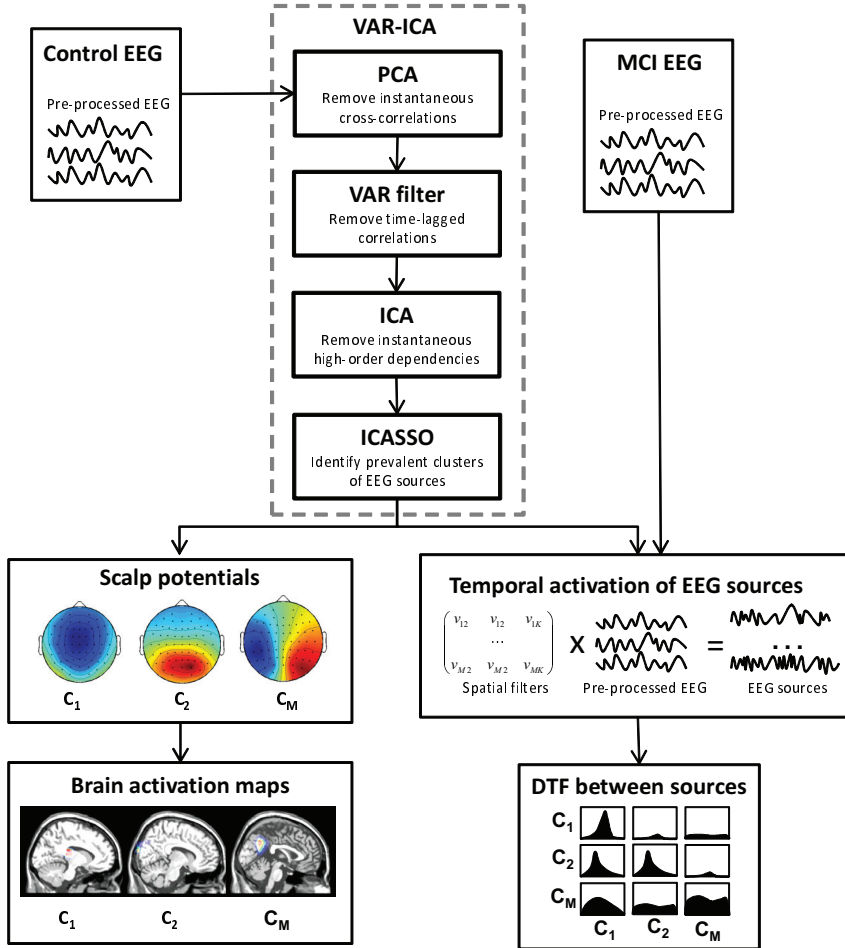


Figure 5.1: Flow chart displaying the main steps of the connectivity analysis. Note that VAR-ICA was applied only to the control population in order to obtain (i) the patterns of scalp potentials of prevalent EEG-alpha sources and (ii) a set of spatial filters retrieving the temporal activation of the those sources. The former were used by swLORETA to compute the brain activation maps while the latter were used to estimate the temporal activation of the EEG sources both in the control population and in the MCI population.

	Thalamus	Cuneus	Precuneus
Controls	9.47 ± 0.19	9.62 ± 0.18	9.57 ± 0.22
MCI	8.72 ± 0.21	8.97 ± 0.25	9.10 ± 0.30

Table 5.2: Means and standard deviations of alpha peak frequency in healthy elderly subjects and MCI patients for the centrotypes of the three clusters (thalamus, cuneus, and precuneus). Adapted with permission from [20].

subject factor, the electrodes showing the maximum amplitude within the alpha band for each EEG-alpha centrotype, and group (healthy elderly and MCI) as the between-subject factor.

Directionality of the synaptic flow between EEG-alpha sources was analyzed for the lower and upper alpha band separately in controls and patients with MCI. The ANOVA included EEG-alpha source as the within-subject factor and group (healthy elderly and MCI) as the between-subject factor.

5.3 Results

The spatial patterns of the EEG-alpha sources are shown in Fig. 4.9. The brain activation maps of the centrotypes of the three significant clusters of EEG sources are shown in Fig. 4.10. The maximum activation of each of these sources was located in caudal regions of the thalamus ($x = 9$, $y = -25$, $z = 9$), in the precuneus ($x = 2$, $y = -60$, $z = 28$), and in the middle occipital gyrus, within the limits of the cuneus ($x = 11$, $y = -97$, $z = 13$).

5.3.1 Alpha peak frequency

The alpha peak frequency corresponding to the thalamic, precuneus, and cuneus EEG-alpha sources was obtained for controls and patients with MCI. As there were no significant main effects of the MMSE scores and age on the alpha peak frequency, these two covariates were not included in the ANOVA. As expected, patients with MCI showed a significant decrease in the alpha peak frequency as compared to controls [$F(1, 38) = 7.02$; $p < 0.02$]. On the average, the alpha peak frequency in patients with MCI was 0.6 Hz (SE = 0.17) lower than in controls. This abnormal slowing was evident for the three EEG-alpha sources (see Table 5.2).

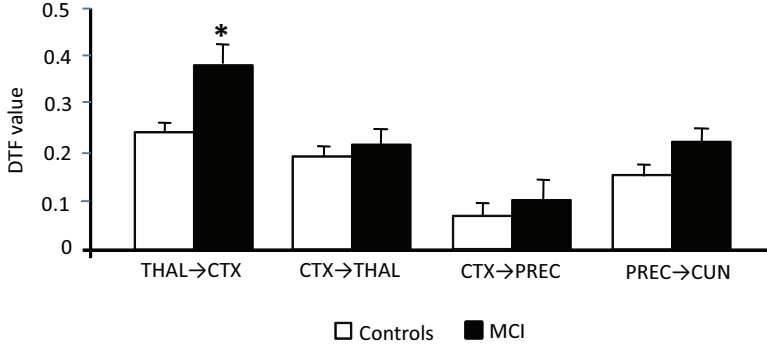


Figure 5.2: Effects of normal and pathological aging on synaptic flow between EEG-alpha sources. Mean strength of the directionality (absolute value of the DTF) of the synaptic flow between thalamus and cortex (cuneus and precuneus) and between cortico-cortical structures involved in the generation of the lower alpha band in healthy elderly controls and patients with MCI. Error bars denote mean standard errors. CTX = cortex, THAL = thalamus, CUN = cuneus, PREC = precuneus. $*p < 0.008$. This figure has been adapted with permission from [20].

5.3.2 Connectivity between EEG-alpha sources

The two-way ANOVA (with MMSE as covariate) performed to evaluate the directionality of the synaptic flow between thalamus, cuneus, and precuneus within the lower alpha band showed a significant group effect and a significant directionality \times group interaction ($p < 0.04$). Namely, the synaptic transfer was abnormally facilitated in these patients (~ 24 % higher) relative to controls, particularly from thalamus to cortex (~ 38 % higher in MCI, $p < 0.008$). No differences between controls and patients with MCI were obtained for cortico-cortical interactions in the lower alpha band. Analysis only revealed a higher flow of synaptic activity from cuneus to precuneus than in the opposite direction [$F(1, 38) = 22.61$, $p < 0.00003$]. Fig. 5.2 shows comparisons of DTF values obtained between brain sources participating in the generation of lower alpha oscillations.

For the upper alpha subdivision, the three-way ANOVA (with age as covariate) revealed a significant group effect [$F(1, 37) = 9.32$, $p < 0.005$]. In particular, patients with MCI showed a decreased bidirectional synaptic flow of about 32% between thalamic and cortical sources when compared to healthy elderly subjects. Analysis also showed a significant interaction between the factors group and age [$F(2, 37) = 8.17$, $p < 0.002$]. In agreement with this finding, we found that the bidirectional synaptic flow between thalamocortical sources of the upper alpha band significantly decreased with age in healthy elderly sub-

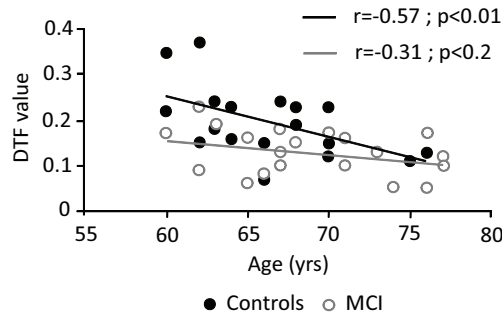


Figure 5.3: Effects of age on thalamocortical synaptic flow in the upper alpha band. Scatter plot showing the degree of relationship between thalamocortical DTF values within the upper alpha band and age in healthy elderly controls (black circles) and patients with MCI (gray empty circles). The value represented for each circle results from averaging the DTF values obtained between thalamus-cuneus, thalamus-precuneus, cuneus-thalamus, and precuneus-thalamus for the upper alpha band in one specific subject. The black and gray lines reflect the best fit for controls and patients with MCI, respectively. Note that only healthy elderly controls showed a significant decrease of the synaptic transmission from thalamus to cortex with age. This figure has been adapted with permission from [20].

jects but not in patients with MCI. The scatter plot shown in Fig. 5.3 shows the degree of linear relationship between these two factors in both groups as derived from Pearson's correlation. Regression fitness reached statistical significance only for healthy elderly subjects ($p < 0.01$) suggesting that the strength of the thalamocortical synaptic flow within the upper alpha decreases with age as a result of normal aging processes. No significant effects were found for cortico-cortical interactions in the upper alpha band in any of the groups.

5.4 Discussion

In this chapter we have investigated synaptic flow in normal elderly subjects and amnesic MCI patients. The VAR-ICA methodology presented in chapter 4 was applied to the control population in order to obtain a set of spatial filters able to selectively enhance scalp activity due to different cerebral sources. These filters were then used to approximate the temporal activation of the neural sources underlying EEG-alpha in controls and MCI patients. Finally, information flow at brain level was quantified in both controls and MCI patients using the DTF. In addition to the typical EEG-alpha slowing associated with neurodegeneration, patients with MCI showed a profile of synaptic

transmission between neural sources significantly distinct from that of normal elderly. In particular, the MCI population was characterized by an abnormal enhancement of the thalamic source together with a facilitation of information flow from thalamus to cortex in the lower alpha range (7.5 – 10 Hz).

It has been repeatedly reported in animals that synchronizing pacemakers within thalamocortical networks play a major role in the generation of alpha rhythm [129, 207]. In humans, a PET study by Schreckenberger et al. [194] found a positive correlation between EEG-alpha power and metabolism of the lateral thalamus, the occipital cortex (cuneus) and adjacent parts of the parietal cortex (precuneus). These results are in agreement with the localization of EEG-alpha sources reported here.

The contribution of deep brain structures to EEG and MEG measurements is usually negligible in comparison with the field generated by post-synaptic currents in pyramidal neurons of the cortical mantle. Moreover, the three-dimensional symmetry of dendritic organization in thalamic neurons significantly diminishes the contribution of the thalamus to scalp-recorded electrophysiological measurements [48, 130, 160]. Nevertheless, unambiguous thalamic sources of EEG and MEG activity have been reported in a number of occasions [60, 98, 128, 217] and the possibility that highly synchronous thalamic oscillations can be captured under specific experimental manipulations cannot be ruled out. Further evidence has shown that the highest degree of cross-correlation in the alpha range between the lateral pulvinar nucleus and different cortical regions was observed in the precentral area [59], where the thalamic centrotypic source that we found was best represented.

Our analyses did not show any preferential direction in the flow of information between thalamus and cortex, in agreement with the notion that the human alpha rhythm is not the result of one or more isolated neural generators but arises from complex interactions between spatially distributed brain sources [161].

Alpha slowing is typically considered as a definitive indication of an underlying neurological pathology [155] and has been consistently reported in AD patients (e.g. [150]). However, only Fernandez et al. [50] has reported such a slowing in MCI patients and most other studies have either found no differences between MCI and controls [84, 99] or identified only a non-significant tendency toward a reduction of alpha peak frequency [163]. The mechanisms by which neurodegeneration slows down EEG-alpha are still unknown. One hypothesis is that β -amyloid plaques in deep gray matter may slow down high-threshold bursting neurons by disrupting the glutamatergic system [20, 86, 167]. However, it is not possible to determine at this point whether the alpha slowing that we found in the MCI population is specific to AD or is just a consequence

of neurological deterioration. To discard this possibility will require a follow-up analysis of the same MCI subjects after some of them have converted to AD.

Animal studies have confirmed that the posterior parietal gyri and the occipital lobes are densely connected with caudal portions of the thalamus [209]. Therefore, aberrant behavior of large ensembles of thalamocortical relay cells are reflected in EEG oscillations [208], which constitute an *in-vivo* global index of thalamocortical integrity. This idea is supported by our finding that normal information flow between thalamic and cortical EEG-alpha sources is disrupted in MCI patients. A bidirectional decrease in information exchange in high-alpha is expected in normal aging (see Fig. 5.3) but the key to cognitive deterioration in MCI might be in the increased influence of the thalamus on the generation of cortical alpha activity. This hypothesis is in agreement with the abnormal enhancement of functional connectivity that has been detected in AD patients [226] as well as with the results of Rombouts et al. [182] reporting that MCI patients show less deactivation of the default mode network when compared to controls but higher deactivation than AD patients.

In summary, in this chapter we used VAR-ICA to characterize information flow in normal aging and MCI. Based on the results, we postulate that deterioration of communication pathways between cortex and thalamus accounts for the more rapid slowing of EEG-alpha in MCI patients than in healthy elderly subjects. In a future follow-up study we will confirm whether these connectivity changes correlate with the prognosis of MCI patients, thereby providing an *in-vivo* functional marker for early diagnosis of AD.

Chapter 6

Concluding remarks

In this thesis we have presented a novel methodology for measuring functional source connectivity with scalp EEG. The proposed approach is based on estimating the temporal activation and spatial characteristics of each relevant neural source using BSS techniques. Then, finding the cerebral localization of each of these sources at a time is much easier than solving the EEG inverse problem directly from raw scalp potentials. In this way, we avoid imposing very restrictive prior requirements on the location of the EEG sources. This localization "blindness" comes at the expense of assuming that the dynamics of the EEG sources can be approximately described with a VAR model driven by non-Gaussian innovations. The plausibility of this assumption was assessed by studying the human EEG-alpha rhythm in a population of healthy elderly subjects. Our analysis found cerebral current source distributions in agreement with the most recent findings on the generators of EEG-alpha, supporting the notion that the novel approach is valid, at least for the analysis of stationary EEG rhythms.

This thesis has made technical contributions to the linear and instantaneous BSS problem in the form of three novel algorithms. The first two algorithms (M-COMBI and F-COMBI) allow for the separation of non-Gaussian and (possibly Gaussian) spectrally diverse sources. Their computational simplicity makes these algorithms suitable for the separation of high-dimensional mixtures. This is especially the case of F-COMBI, which can be several orders of magnitude faster than some state-of-the-art competitors. The third algorithm, called ENRICA, is based on the minimization of the entropy rate of the source estimates. ENRICA is computationally demanding but not much more than other well-known non-parametric BSS algorithms. Computation time is

rarely the crucial factor in offline analysis of EEG data and ENRICA is an attractive choice in these cases, due to the superior accuracy and reliability that has shown in our experiments with real EEG.

Several time-adaptive estimators of information transfer were presented in chapter 3. Using simulated data it was shown that these estimators can accurately describe time-varying coupling dynamics in an ensemble of repeated measurements. We anticipate that these new connectivity indices might provide useful information on the generation mechanisms of event related brain potentials, which is a topic for further research.

In the last chapter of the thesis we identified aberrant connectivity patterns that differentiate normal aging from mild cognitive impairment (MCI), a clinical condition at risk of developing Alzheimer's disease (AD). The fact that these connectivity differences can be detected years before diagnosis of AD are encouraging, and suggest that they could be used as a non-invasive tool to improve early diagnosis in clinical settings. Prospectively, the analysis methods developed in this thesis could provide a rational basis for evaluating how new drugs affect neural networks in early degeneration, which might have far-reaching implications for therapeutic drug development.

Appendix A

Information theory

In this appendix we summarize the most important concepts and relationships from information theory that are used in this thesis. For a comprehensive reference on information theory see [35].

A.1 Basic definitions

Definition: The *differential entropy* of a continuous random vector X with a probability density function (pdf) p is defined as:

$$H(p) = - \int_{\mathcal{S}_p} p(\mathbf{x}) \log(p(\mathbf{x})) d\mathbf{x} \quad (\text{A.1})$$

where \mathcal{S}_p is the support of p .

Definition: The *Kullback-Leibler divergence* between two pdfs p and q is defined as:

$$K(p||q) = \int_{\mathcal{S}_p \cap \mathcal{S}_q} p(\mathbf{y}) \log \frac{p(\mathbf{y})}{q(\mathbf{y})} d\mathbf{y} \quad (\text{A.2})$$

where \mathcal{S}_p and \mathcal{S}_q are the supports of p and q , respectively.

For convenience, we use in this thesis the notational equivalences $H(p) \equiv H(\mathbf{x}) \equiv H(p(\mathbf{x})) \equiv H_X$ and $K(p||q) \equiv K(\mathbf{x}||\mathbf{y}) \equiv K(p(\mathbf{x})||p(\mathbf{y})) \equiv K_{X||Y}$ if p and q are the densities of the random vectors \mathbf{x} and \mathbf{y} , respectively.

Definition: The *mutual information* between the marginal components of a continuous random vector \mathbf{x} is defined as:

$$I(\mathbf{x}) = K\left(p(\mathbf{x}) \parallel \prod_{i=1}^M p_i(x_i)\right) = \int_{\bigcap_{i=1}^M \mathcal{S}_i} p(\mathbf{x}) \log\left(\frac{\mathbf{x}}{\prod_{i=1}^M p_i(x_i)}\right) d\mathbf{x} \quad (\text{A.3})$$

where \mathcal{S}_i is the support of p_i .

Definition: A stochastic process $\{x(n)\}_n$ is said to be *stationary* if the joint distribution of any subset of the sequence of random variables is invariant with respect to shifts in the time index, i.e.,

$$\begin{aligned} p(x(1), x(2), \dots, x(n)) \\ = p(x(1+l), x(2+l), \dots, x(n+l)) \end{aligned} \quad (\text{A.4})$$

Definition: A discrete stochastic process $x(1), x(2), \dots$ is said to be a *Markov process of order d* if, for $n = 1, 2, \dots$,

$$\begin{aligned} p(x(n+1) | x(n), \dots, x(n-d+1)) \\ = p(x(n+1) | x(n), \dots, x(n-d+1), x(n-d)) \end{aligned} \quad (\text{A.5})$$

Definition: The *entropy rate* of a stochastic process $\{x(n)\}_n$ is defined by:

$$H_{rate}(\{x(n)\}_n) = \lim_{n \rightarrow \infty} \frac{1}{n} H(x(1), x(2), \dots, x(n)) \quad (\text{A.6})$$

For a stationary Markov process of order d , the entropy rate is given by:

$$H_{rate}(\{x(n)\}) = H(\mathbf{x}^{(d)}) - H(\mathbf{x}^{(d-1)}) \quad (\text{A.7})$$

with $\mathbf{x}^{(d)} = [x(d), x(d-1), \dots, x(1)]$.

A.2 Properties and relationships

- $H(\mathbf{x}, \mathbf{y}) = H(\mathbf{y}|\mathbf{x}) + H(\mathbf{x}) = H(\mathbf{x}|\mathbf{y}) + H(\mathbf{y})$
- $H(\mathbf{x}, \mathbf{y}) = H(\mathbf{x}) + H(\mathbf{y}) - I(\mathbf{x}, \mathbf{y})$
- $H(\mathbf{A}\mathbf{x}) = H(\mathbf{x}) + \log |\mathbf{A}|$ for any invertible matrix \mathbf{A}
- $H(\mathbf{x}) \leq \frac{1}{2} \log ((2\pi e)^m |\Sigma_{\mathbf{x}}|)$ with equality iff \mathbf{x} is jointly Gaussian with mean zero and covariance matrix $\Sigma_{\mathbf{x}}$.
- $K(p||q) \geq 0$ with equality iff p and q agree p -almost everywhere.
- $K(\mathbf{x}|\mathbf{y}) = K(f(\mathbf{x})||f(\mathbf{y})) = K(f^{-1}(\mathbf{x})||f^{-1}(\mathbf{y}))$ for any invertible transformation f of the sample space.
- If the pdf of random vector \mathbf{z} has support $(0, 1)^n$ and \mathbf{u} is a random vector uniformly distributed on $(0, 1)^n$ then $H(\mathbf{z}) = -K(\mathbf{z}||\mathbf{u})$.
- $K(p(\mathbf{y})||p(\mathbf{s})) = K(p(\mathbf{y})||\prod_i p(y_i)) + K(\prod_i p(y_i)||p(\mathbf{s}))$ for any distribution $p(\mathbf{s})$ with independent components, i.e. $p_{\mathbf{s}}(\mathbf{s}) = \prod_i p(s_i)$.

Appendix B

The concept of state-space

Dynamical systems are often described using a set of state variables so that the state of the system at a given time n corresponds to a point $\mathbf{z}(n) \in \mathcal{M} \subseteq \mathbb{R}^d$ where \mathbb{R}^d is the *state-space* of the system and d is the state-space dimension. The state-space is often called a *phase-space* when \mathcal{M} forms a smooth finite-dimensional manifold that consists of an infinite number of points.

A discrete-time dynamical system is said to be *deterministic* when there is a deterministic *evolution operator* $\Phi : \mathcal{M} \times \mathbb{Z} \mapsto \mathcal{M}$ such that $\Phi(\mathbf{z}(n), \tau) = \mathbf{z}(n + \tau)$. In other words, when one can write some mathematical rule by which future states of the system can be precisely determined from the current state. Moreover, the dynamical system is said to be *stationary* if the evolution operator does not change with time. On the contrary, a non-stationary system is one whose temporal evolution is subject to some outside influence that is not accounted for in the evolution operator Φ . Note that for any non-stationary evolution operator, a new evolution operator Φ' can be constructed which is stationary by including in the system definition all outside influences, i.e. by simply increasing the state-space dimension.

Typically, we cannot observe directly the true states $\mathbf{z}(n)$ of a dynamical system. Instead, we usually have an *observation function* $g : \mathcal{M} \mapsto \mathbb{R}$ that provides us with a scalar time-series of system measurements. Obviously $g(\cdot)$ cannot offer a complete description of the system but, according to Taken's theorem [212], if d_e is sufficiently large, the evolution of the *embedded vector*:

$$x(n) \mapsto \mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-d_e)] \quad (\text{B.1})$$

will be the same as $\mathbf{z}(n)$. In practice, such reconstruction of the state-space is typically not possible due to the presence of noise and finite quantisation errors but it can be presumed that the data embeded by (B.1) approximates the topology of the underlying manifold attractor.

The most common approach to reconstruct the state-space of a dynamical system is to use *time-delayed embedding*:

$$\mathbf{x}(n) = [x(n), x(n - \tau), x(n - 2\tau), \dots, x(n - (m - 1)\tau)] \quad (\text{B.2})$$

where τ is the difference in samples between successive components of the embedded vector or *embedding delay* and m is the *embedding dimension*. Choosing the right values for τ and m can be a tricky issue but the most commonly employed method is to take $\tau = 1$ and successively embed in higher dimensions until the results of the analysis are consistent. See however [106, 205] for comprehensive overviews on how to find good embedding parameters.

Bibliography

- [1] H. Akaike. Autoregressive model fitting for control. *Ann. Inst. Stat. Math.*, 23:163–180, 1971.
- [2] M. T. Alkire, A. G. Hudetz, and G. Tononi. Consciousness and anesthesia. *Science*, 322:876–880, 2008.
- [3] L. Astolfi, H. Bakardjian, F. Cincotti, D. Mattia, M. Grazia Marciani, F. De Vico Fallani, A. Colosimo, S. Salinari, F. Miwakeichi, Y. Yamaguchi, P. Martinez, A. Cichocki, A. Tocci, and F. Babiloni. Estimate of causality between independent cortical spatial patterns during movement volition in spinal cord injured patients. *Brain Topograph.*, 19:107–123, 2007.
- [4] L. Astolfi, F. Cincotti, D. Mattia, M. Grazia Marciani, L. A. Baccala, F. de Vico Fallani, S. Salinari, M. Ursino, M. Zavaglia, L. Ding, J. Christopher Edgar, G. A. Miller, B. He, and F. Babiloni. Comparison of different cortical connectivity estimators for high-resolution EEG recordings. *Hum. Brain Mapp.*, 28(2):143–157, 2007.
- [5] C. Babiloni, G. Binetti, E. Cassetta, G. Dal Forno, C. Del Percio, R. Ferreri, G. Frisoni, K. Hirata, B. Lanuzza, C. Miniussi, D. V. Moretti, F. Nobili, G. Rodriguez, G. L. Romani, S. Salinari, and P. M. Rossini. Sources of cortical rhythms changes as a function of cognitive impairment in pathological aging: a multicenter study. *Clin. Neurophysiol.*, 117:252–268, 2006.
- [6] F. Babiloni, F. Cincotti, C. Babiloni, F. Carducci, D. Mattia, L. Astolfi, A. Basilisco, P. M. Rossini, L. Ding, Y. Ni, J. Cheng, J. Christine, J. Sweeney, and B. He. Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function. *Neuroimage*, 24(11):118–131, 2005.

- [7] L. A. Baccalá and K. Sameshima. Partial directed coherence: A new concept in neural structure determination. *Biol. Cybern.*, 84:463–474, 2001.
- [8] S. Baillet, J. C. Mosher, and R. M. Leahy. Electromagnetic brain mapping. *IEEE Signal Process. Mag.*, 18:14–30, 2001.
- [9] H. Bakardjian, A. Cichocki, F. Cincotti, D. Mattia, F. Babiloni, M. Grazia Marciari, F. De Vico Fallani, F. Miwakeichi, Y. Yamaguchi, P. Martinez, S. Salinari, A. Tocci, and L. Astolfi. Estimate of causality between cortical spatial patterns during voluntary movements in normal subjects. *International Journal of Bioelectromagnetism*, 8(1):II/1–II/18, 2006.
- [10] E. Başar, M. Schürmann, C. Başar-Eroglu, and S. Karakaş. Alpha oscillations in brain functioning: an integrative theory. *Int. J. Psychophysiol.*, 26(1–3):5–29, 1997.
- [11] E. Be’ery and A. Yeredor. Blind separation of superimposed shifted images using parametrized joint diagonalization. *IEEE T. Image Process.*, 17(3), 2008.
- [12] A. J. Bell and T. J. Sejnowski. A non-linear information maximization algorithm that performs blind separation. In *Advances in Neural Information Processing Systems 7*, pages 467–474. The MIT Press, Cambridge, MA, 1995.
- [13] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE T. on Signal Processing*, 45(2):434–444, 1997.
- [14] C. Bernasconi and König. On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings. *Biol. Cybern.*, 81:199–210, 1999.
- [15] R. Boscolo, H. P. Pan, and V. P. Roychowdhury. Independent component analysis based on nonparametric density estimation. *IEEE T. Neural Networ.*, 15:55–65, 2004.
- [16] R. Boscolo and Roychowdhury. On the uniqueness of the minimum of the information-theoretic cost function for the separation of mixtures of nearly Gaussian signals. In *Proc. ICA’03*, pages 137–141, 2003.
- [17] H. Braak and E. Braak. Demonstration of amyloid deposits and neurofibrillary tangles in whole brain sections. *Brain Pathol.*, 1:742–746, 1991.

- [18] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. Bressler. Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. *P. Natl. Acad. Sci. USA*, 101(26):9849–9854, 2004.
- [19] M. A. Bushe, G. Eichhoff, H. Adelsberger, D. Abramowski, K.-H. Wiederhold, C. Haass, M. Staunfenbiel, A. Konnerth, and O. Garaschuk. Clusters of hyperactive neurons near amyloid plaques in a mouse model of Alzheimer’s disease. *Science*, 321:1686–1689, 2008.
- [20] J. L. Cantero, M. Atienza, G. Gómez-Herrero, A. Cruz-Vadell, E. Gil-Neciga, R. Rodríguez-Romero, and D. García-Solis. Functional integrity of thalamocortical circuits differentiates normal aging from mild cognitive impairment. *Hum. Brain Mapp.*, 30(12):3944–3957, 2009. doi: 10.1002/hbm.20819.
- [21] J.-F. Cardoso. On the performance of orthogonal source separation algorithms. In *Proc. EUSIPCO*, pages 776–779, Edinburgh, Ireland, 1994.
- [22] J.-F. Cardoso. Infomax and maximum likelihood for source separation. *IEEE Signal Proc. Lett.*, 4:112–114, 1997.
- [23] J.-F. Cardoso. Blind signal separation: Statistical principles. *P. IEEE*, 86(10):2009–2025, 1998.
- [24] J.-F. Cardoso. Multidimensional independent component analysis. In *Proc. ICASSP’98*, pages 12–15, Seattle, WA, 1998.
- [25] J.-F. Cardoso. High-order contrasts for Independent Component Analysis. *Neural Comput.*, pages 157–192, 1999.
- [26] J.-F. Cardoso. The three easy routes to independent component analysis: contrasts and geometry. In *Proc. ICA’01, San Diego, USA*, pages 463–468, December 2001.
- [27] J.-F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J. Matrix Anal. Appl.*, 17(1):161–164, 1996.
- [28] M. Chatila, C. Milleret, A. Rougeul, and P. Buser. Alpha rhythm in the cat thalamus. *C. R. Acad. Sci. III*, 316:51–58, 1993.
- [29] M. Chávez, J. Martinerie, and M. Le Van Quyen. Statistical assessment of nonlinear causality: application to epileptic EEG signals. *J. Neurosci. Methods*, 124:113–128, 2003.
- [30] Y. Chen, G. Rangarajan, J. Feng, and M. Ding. Analyzing multiple nonlinear time series with extended Granger causality. *Phys. Lett. A*, 324(1):26–35, 2004.

- [31] Y. Cheung and L. Xu. Dual multivariate auto-regressive modeling in state space for temporal signal separation. *IEEE T. Syst. Man Cyb.*, 33:386–398, 2003.
- [32] I. Christov, G. Gómez-Herrero, V. Krasteva, I. Jekova, and A. Gotchev. Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification. *Med. Eng. Phys.*, 28(9):876–887, 2006.
- [33] A. Cichocki and S. Amari. *Adaptive blind signal and image processing: learning algorithms and applications*. John Wiley & Sons, New York, 2002.
- [34] P. Comon. Independent component analysis - a new concept? *Signal Proces.*, 36:287–314, 1994.
- [35] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.
- [36] S. A. Cruces-Álvarez and A. Cichocki. Combining blind source extraction with joint approximate diagonalization: Thin algorithms for ICA. In *Proc. ICA'03, Nara, Japan*, pages 463–469, 2003.
- [37] S. A. Cruces-Alvarez, A. Cichocki, and S. Amari. From blind signal extraction to blind instantaneous signal separation: criteria, algorithms, and stability. *IEEE T. Neural Networ.*, 15(4):859–873, 2004.
- [38] S. S. Dalal, A. G. Guggisberg, E. Edwards, K. Sekihara, A. M. Findlay, R. T. Canolty, M. S. Berger, R. T. Knight, N. M. Barbaro, H. E. Kirsch, and S. S. Nagarajan. Five-dimensional neuroimaging: localization of the time-frequency dynamics of cortical activity. *Neuroimage*, 40(4):1686–1700, 2008.
- [39] O. David and K. J. Friston. A neural mass model for MEG/EEG: coupling and neuronal dynamics. *Neuroimage*, 20(3):1743–1755, 2003.
- [40] O. David, L. Garnero, D. Cosmelli, and F. J. Varela. Estimation of neural dynamics from MEG/EEG cortical current density maps: application to the reconstruction of large-scale cortical synchrony. *IEEE T. Biomed. Eng.*, 49(9):975–987, 2002.
- [41] O. David, S. J. Kiebel, L. M. Harrison, J. Mattout, J. M. Kilner, and K. J. Friston. Dynamic causal modeling of evoked responses in EEG and MEG. *Neuroimage*, 30(4):1255–1272, 2006.
- [42] W. De Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel. Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE T. Biomed. Eng.*, 53(12):2583–2587, 2006.

- [43] L. De Lathauwer, D. Callaerts, B. De Moor, and J. Vandewalle. Fetal electrocardiogram extraction by source subspace separation. In *Proc. IEEE Signal Processing workshop on higher-order statistics, Girona, Spain, June 12–14*, pages 134–138, 1995.
- [44] J. C. De Munck, S. I. Gonçalves, L. Huijboom, J. P. Kuijer, P. J. Pouwels, R. M. Heethaar, and F. H. Lopes da Silva. The hemodynamic response of the alpha rhythm: an EEG/fMRI study. *Neuroimage*, 35:1142–1151, 2007.
- [45] A. Delorme and S. Makeig. EEGLAB user manual. Available online: <http://sccn.ucsd.edu/eeglab/maintut>.
- [46] A. Delorme and S. Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Meth.*, 134:9–21, 2004. Available online: <http://sccn.ucsd.edu/eeglab/>.
- [47] E. Doron, A. Yeredor, and P. Tichavský. Cramér-rao-induced bound for blind separation of stationary parametric Gaussian sources. *IEEE Signal Process. Lett.*, 14(6):417–420, 2007.
- [48] D. Druckman and J. I. Lacey. *Brain and cognition. Some new technologies*. National Academy press, Washington, DC, 1989.
- [49] M. Eichler. On the evaluation of information flow in multivariate systems by the directed transfer function. *Biol. Cybern.*, 94:469–482, 2006.
- [50] A. Fernandez, R. Hornero, A. Mayo, J. Poza, P. Gil-Gregorio, and T. Ortiz. MEG spectral profile in Alzheimer’s disease and mild cognitive impairment. *Clin. Neurophysiol.*, 117:306–314, 2006.
- [51] R. S. J. Frackowiak, J. T. Ashburner, W. D. Penny, S. Zeki, K. J. Friston, C. D. Frith, R. J. Dolan, and C. J. Price, editors. *Human Brain Function*. Academic Press, 2 edition, 2005.
- [52] P. J. Franaszczuk and G. K. Bergey. Application of the directed transfer function method to mesial and lateral onset temporal lobe seizures. *Brain Topogr.*, 11:13–21, 1998.
- [53] W. J. Freeman. Characteristics of the synchronization of brain activity imposed by finite conduction velocities of axons. *Int. J. Bifurcation Chaos*, 10(10):2307–2322, 2000.
- [54] W. A. Freiwald, P. Valdes, J. Bosch, R. Biscay, J. C. Jimenez, L. M. Rodriguez, V. Rodriguez, A. K. Kreiter, and W. Singer. Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. *J. Neurosci. Meth.*, 94:105–119, 1999.

- [55] S. Frenzel and B. Pompe. Partial mutual information for coupling analysis of multivariate time series. *Phys. Rev. Lett.*, 99(20):204101, 2007.
- [56] K. J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Hum. Brain Mapp.*, 2:56–78, 1994.
- [57] K. J. Friston. Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLOS Biology*, 7(2):220–225, 2009.
- [58] K. J. Friston, L. Harrison, and W. Penny. Dynamical causal modelling. *Neuroimage*, 19:1273–1302, 2003.
- [59] S. Fujita, K. Fujita, and S. Matsumoto. Cross-correlation analysis of the lateral pulvinar and scalp EEG in man. *Appl. Neurophysiol.*, 42:294–301, 1979.
- [60] R. Gobbele, T. D. Waberski, H. Simon, E. Peters, F. Klostermann, G. Curio, and H. Buchner. Different origins of low- and high-frequency components (600 Hz) of human somatosensory evoked potentials. *Clin. Neurophysiol.*, 115:927–937, 2004.
- [61] Gómez-Herrero, I. Jekova, V. Krasteva, I. Christov, A. Gotchev, and K. Egiazarian. Relative estimation of the Karhunen-Loève transform basis functions for detection of ventricular ectopic beats. In *Proc. Computers in Cardiology, CinC 2006*, pages 569 – 572, Valencia, Spain, September, 2006.
- [62] G. Gómez-Herrero, M. Atienza, K. Egiazarian, and J. L. Cantero. Measuring directional coupling between EEG sources. *Neuroimage*, 43(3):497–508, November 2008.
doi: 10.1016/j.neuroimage.2008.07.032.
- [63] G. Gómez-Herrero, W. De Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Van Huffel, and W. Van Paesschen. Automatic removal of ocular artifacts in the EEG without a reference EOG channel. In *Proc. 7th Nordic Signal Process. Sym. NORSIG 2006*, pages 130–133, Reykjavik, Iceland, 2006.
- [64] G. Gómez-Herrero and K. Egiazarian. Independent component analysis by a resampling strategy. Technical report, Tampere University of Technology, Department of Signal Processing, 2005. Available online: <http://www.cs.tut.fi/~gomezher/rica.htm>.
- [65] G. Gómez-Herrero, A. Gotchev, I. Christov, and K. Egiazarian. Feature extraction for heartbeat classification using matching pursuits and independent component analysis. In *Proc. of the 30th Int. Conf. on Acoustics, Speech, and Signal Process., ICASSP 2005*, pages 725–728, 2005.

- [66] G. Gómez-Herrero, E. Huupponen, A. Värri, K. Egiazarian, B. Vanrumste, A. Vergult, W. De Clercq, S. Van Huffel, and W. Van Paesschen. Independent component analysis of single trial evoked brain responses: Is it reliable? In *Proc. 2nd Int. Conf. on Comput. Int. in Medicine and Healthcare, CIMED 2005*, pages 69–76, Costa da Caparica, Portugal, 2005.
- [67] G. Gómez-Herrero, Z. Koldovský, P. Tichavský, and K. Egiazarian. A fast algorithm for blind separation of non-Gaussian and time-correlated signals. In *Proc. 15th Eur. Signal Process. Conf., EUSIPCO 2007*, pages 1731–1735, Poznan, Poland, 2007.
- [68] G. Gómez-Herrero and K. Rutanen. MATLAB/C++ implementation of ENRICA, 2009. Available online: <http://www.cs.tut.fi/~gomezher/enrica.htm>.
- [69] G. Gómez-Herrero, K. Rutanen, and K. Egiazarian. Blind source separation by entropy rate minimization. *IEEE Signal Proc. Let.*, 17(2):153–156, 2010.
- [70] G. Gómez-Herrero, W. Wu, K. Rutanen, M. C. Soriano, G. Pipa, and R. Vicente. Assessing coupling dynamics from an ensemble of multivariate time-series. Submitted.
- [71] S. Gonçalves, J. de Munck, J. Verbunt, R. Heetharr, and F. da Silva. *In vivo* measurement of brain and skull resistivities using an EIT-based method and the combined analysis of SEF/SEP data. *IEEE T. Biomed. Eng.*, 50:1124–1128, 2003.
- [72] I. F. Gorodnitsky and A. Belouchrani. Joint cumulant and correlation based signal separation with application to EEG data. In *Proc. ICA '01, San Diego, CA, USA*, pages 475–480, 2001.
- [73] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.
- [74] J. Gross, J. Kujala, M. Hämäläinen, L. Timmermann, A. Schnitzler, and R. Salmelin. Dynamic imaging of coherent sources: studying neural interactions in the human brain. *Proc. Nat. Acad. Sci.*, 98(2):646–699, 2001.
- [75] J. Gross, F. Schmitz, I. Schnitzler, K. Kessler, K. Shapiro, and B. Hommel. Modulation of long-range neuronal synchrony reflects temporal limitations of visual attention in humans. *Proc. Natl. Acad. Sci.*, 101:13050–13055, 2004.
- [76] R. Guidara, S. Hosseini, and S. Deville. Blind separation of nonstationary Markovian sources using an equivariant Newton-Raphson algorithm. *IEEE Signal Proc. Let.*, 16(5), 2009.

- [77] P. He, G. Wilson, and C. Russell. Removal of ocular artifacts from electro-encephalogram by adaptive filtering. *Med. Biol. Eng. Comp.*, 42:407–412, 2004.
- [78] J. Himberg and A. Hyvärinen. ICASSO toolbox, 2005. Available online: <http://www.cis.hut.fi/projects/ica/icasso/>.
- [79] J. Himberg, A. Hyvärinen, and F. Esposito. Validating the independent components of neuroimaging time-series via clustering and visualization. *Neuroimage*, 22:1214–1222, 2004.
- [80] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442:164–171, 2006.
- [81] K. Hoechstetter, H. Bornfleth, D. Weckesser, N. Ille, P. Berg, and M. Scherg. BESA source coherence: A new method to study cortical oscillatory coupling. *Brain Topogr.*, 16(4):233–238, 2004.
- [82] C. J. Honey, R. Kötter, M. Breakspear, and O. Sporns. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *P. Natl. Acad. Sci. USA*, 104(24):10240–10245, 2007.
- [83] S. Hosseini and C. Jutten. Markovian source separation. *IEEE. T. Signal Proces.*, 51(12):3009–3019, 2003.
- [84] C. Huang, L. Wahlund, T. Dierks, P. Julin, B. Winblad, and V. Jelic. Discrimination of Alzheimer’s disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study. *Clin. Neurophysiol.*, 11:1961–1967, 2000.
- [85] R.-S. Huang, T.-P. Jung, A. Delorme, and S. Makeig. Tonic and phasic electroencephalographic dynamics during continuous compensatory tracking. *Neuroimage*, 34(4):1896–1909, 2008.
- [86] S. W. Hughes, M. Lörincz, D. W. Cope, K. L. Blethyn, K. A. Kekesi, H. R. Parri, G. Juhasz, and V. Crunelli. Synchronized oscillations at alpha and theta frequencies in the lateral geniculate nucleus. *Neuron*, 42:253–268, 2004.
- [87] E. Huupponen, W. De Clercq, G. Gómez-Herrero, A. Saastamoinen, K. Egiazarian, A. Värri, A. Vanrumste, S. Van Huffel, W. Van Paesschen, J. Hasan, and S.-L. Himanen. Determination of dominant simulated spindle frequency with different methods. *J. Neurosci. Meth.*, 156(1–2):275–283, 2006.
- [88] A. Hyvärinen. New approximations of differential entropy for independent component analysis and projection pursuit. In *Advances Neural Inform. Processing Syst.* 10, pages 273–279. MIT Press, 1998.

- [89] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE T. on Neural Networks*, 10(3):626–634, 1999.
- [90] A. Hyvärinen. Complexity pursuit: Separating interesting components from time series. *Neural Comput.*, 13(4):883–898, 2001.
- [91] A. Hyvärinen. A unifying model for blind separation of independent sources. *Signal Process.*, 85(7):1419–1427, 2005.
- [92] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.
- [93] A. Hyvärinen and E. Oja. Independent component analysis by general nonlinear Hebbian-like learning rules. *Signal Process.*, 64(3):301–313, 1998.
- [94] A. Hyvärinen, S. Shimizu, and P. O. Hoyer. Causal modelling combining instantaneous and lagged effects: an identifiable model based on non-gaussianity. In *Proc. Int. Conf. on Machine Learning (ICML2008)*, pages 424–431, Helsinki, Finland, 2008.
- [95] International Federation of Societies for Electroencephalography and Clinical Neurophysiology (IFSECN). A glossary of terms commonly used by clinical electroencephalographers. *Electroencephalogr. Clin. Neurophysiol.*, 376:538–548, 1974.
- [96] A. A. Ioannides, G. K. Kostopoulos, N. A. Laskaris, L. Liu, T. Shibata, M. Schellens, V. Poghosyan, and A. Khurshudyan. Timing and connectivity in the human somatosensory cortex from single trial mass electrical activity. *Human Brain Mapp.*, 15(4):231–246, 2002.
- [97] A. A. Ioannides, L. C. Liu, J. Kwapien, S. Drozd, and M. Streit. Coupling of regional activations in a human brain during an object and face affect recognition task. *Human Brain Mapp.*, 11(2):77–92, 2000.
- [98] S. A. Isaichev, V. T. Derevyankin, Yu. M. Koptelov, and E. N. Sokolov. Rhythmic alpha-activity generators in the human EEG. *Neurosci. Behav. Physiol.*, 31(1):49–53, 2001.
- [99] V. Jelic, M. Shigeta, P. Julin, O. Almkvist, B. Winblad, and L. O. Wahlund. Quantitative electroencephalography power and coherence in Alzheimer’s disease and mild cognitive impairment. *Dementia*, 7:314–323, 1996.
- [100] I. T. Jolliffe. *Principal Component Analysis*. Springer, 2 edition, 2002.
- [101] M. C. Jones and H. W. Lotwick. Remark as r50: A remark on algorithm as 176. Kernel density estimation using the fast Fourier transform. *J. Roy. Stat. Soc. C-App.*, 33(1):120–122, 1984.

- [102] T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37:163–178, 2000.
- [103] J. Jutten, C. Hérault. Blind separation of sources, Part I: an adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.
- [104] M. Kamiński and K. J. Blinowska. A new method of the description of the information flow in the brain structures. *Biol. Cybern.*, 65:203–210, 1991.
- [105] M. Kamiński, M. Ding, W. Truccolo, and S. Bressler. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assesment of significance. *Biol. Cybern.*, 85:145–157, 2001.
- [106] H. Kantz and T. Schreiber. *Nonlinear time series analysis*. Cambridge University Press, 1997.
- [107] G. B. Karas, P. Scheltens, S. A. Rombouts, P. J. Visser, R. A. Van Schijndel, N. C. Fox, and F. Barkhof. Global and local gray matter loss in mild cognitive impairment and Alzheimer’s disease. *Neuroimage*, 23:708–716, 2004.
- [108] S. J. Kiebel, M. I. Garrido, R. Moran, C.-C. Chen, and K. J. Friston. Dynamic causal modeling for EEG and MEG. *Human Brain Mapp.*, 30(6):1866–1876, 2009.
- [109] S. A. Klein and T. Carney. The usefulness of the laplacian in principal component analysis and dipole source localization. *Brain Topogr.*, 8(2):91–108, 1995.
- [110] Z. Koldovský, P. Tichavský, and E. Oja. Efficient variant of algorithm fastica for independent component analysis attaining the Cramer-Rao lower bound. *IEEE T. Neural Networks*, 17(5):1265–1277, 2006.
- [111] A. Korzeniewska, S. Kasicki, M. Kamiński, and K. J. Blinowska. Information flow between hippocampus and related structures during various types of rat’s behavior. *J. Neurosci. Meth.*, 73(1):49–60, 1997.
- [112] L. F. Kozachenko and N. N. Leonenko. A statistical estimate for the entropy of a random vector. *Probl. Infor. Transm.*, 23(95):916, 1987.
- [113] M. A. Kramer, E. Edwards, M. Soltani, M. S. Berger, R. T. Knight, and A. J. Szeri. Synchronization measures of bursting data: application to the electrocorticogram of an auditory event-related experiment. *Phys. Rev. E.*, 70:011914, 2004.

- [114] A. Kraskov. *Synchronization and interdependence measures and their applications to the electroencephalogram of epilepsy patients and clustering of data*. PhD thesis, Department of Physics, University of Wuppertal, 2004.
- [115] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Phys. Rev. E*, 69(6):066138, 2004.
- [116] J. Kujala, J. Gross, and R. Salmelin. Localization of correlated network activity at the cortical level with MEG. *Neuroimage*, 39:1706–1720, 2008.
- [117] J. Kujala, K. Pammer, P. Cornelissen, A. Roebroek, E. Formisano, and R. Salmelin. Phase coupling in a cerebro-cerebellar network at 8-13 hz during reading. *Cereb. Cortex*, 17:1476–1485, 2007.
- [118] S. Kullback. *Information Theory and statistics*. Wiley, New York, 1959.
- [119] R. Kús, M. Kamiński, and K. Blinowska. Determination of EEG activity propagation: pair-wise versus multichannel estimate. *IEEE T. Biomed. Eng.*, 51(9):1501–1510, 2004.
- [120] H. Laaksonen, J. Kujala, and R. Salmelin. A method for spatiotemporal mapping of event-related modulation of cortical rhythmic activity. *Neuroimage*, 42(1):207–217, 2008.
- [121] R. M. Leahy, J. C. Mosher, M. E. Spencer, M. X. Huang, and J. D. Lewine. A study of dipole localization accuracy for MEG and EEG using a human skull phantom. *Electroencephalogr. Clin. Neurophysiol.*, 107:159–173, 1998.
- [122] E. G. Learned-Miller and J. W. Fisher III. Independent component analysis using spacings estimates of entropy. *J. Mach. Learn. Res.*, 4:1271–1295, 2003.
- [123] T.-W. Lee, M. Girolami, and T. J. Sejnowski. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.*, 11(2):417–441, 1999.
- [124] D. Lehmann, P. L. Faber, L. R. R. Gianotti, K. Kochi, and R. D. Pasqual-Marqui. Coherence and phase locking in the scalp EEG and between LORETA model sources, and microstates as putative mechanisms of brain temporo-spatial functional organization. *J. Physiology-Paris*, 99:29–36, 2006.
- [125] H. W. Lilliefors. On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *Journal of the American Statistical Association*, 64:387–389, 1969.

- [126] F.-H. Lin, K. Hara, V. Solo, M. Vangel, J. W. Belliveau, S. M. Stufflebeam, and M. Hämäläinen. Dynamic Granger-Geweke causality modeling with application to interictal spike propagation. *Hum. Brain Mapp.*, 30:1877–1886, 2009.
- [127] A. K. Liu, A. M. Dale, and J. W. Belliveau. Monte Carlo simulation studies of EEG and MEG localization accuracy. *Hum. Brain Mapp.*, 16:47–62, 2002.
- [128] R. R. Llinas, U. Ribary, D. Jeanmonod, E. Kronberg, and P. P. Mitra. Thalamocortical dysrhythmia: a neurological and neuropsychiatric syndrome characterized by magnetoencephalography. *P. Natl. Acad. Sci. USA*, 96(26):15222–15227, 1999.
- [129] F. Lopes da Silva. Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalogr. Clin. Neurophysiol.*, 79:81–93, 1991.
- [130] F. Lopes da Silva and A. Van Rotterdam. Biophysical aspects of EEG and magnetoencephalogram generation. In E. Niedermeyer and F. Lopes da Silva, editors, *Electroencephalography. Basic principles, Clinical applications and related fields*, pages 78–91. Williams & Wilkins, Baltimore, 1993.
- [131] F. H. Lopes da Silva, T. H. M. T. Lierop, C. F. Schrijer, and W. Storm van Leeuwen. Organization of thalamic and cortical alpha rhythms: spectra and coherences. *Electroencephalogr. Clin. Neurophysiol.*, 35:627–639, 1973.
- [132] F. H. Lopes da Silva and W. Storm van Leeuwen. The cortical source of the alpha rhythm. *Neurosci. Lett.*, 6:237–241, 1977.
- [133] F. H. Lopes da Silva, J. E. Vos, J. Mooibroek, and A. van Rotterdam. Relative contributions of the intracortical and thalamo-cortical processes in the generation of alpha rhythms, revealed by partial coherence analysis. *Electroencephalogr. Clin. Neurophysiol.*, 50:449–456, 1980.
- [134] M. C. Mackey and L. Glass. Oscillation and chaos in physiological control systems. *Science*, 197:287–289, 1977.
- [135] S. Makeig, T.-P. Jung, Bell. A. J., D. Ghahremani, and T. Sejnowski. Blind separation of auditory event-related brain responses into independent components. *P. Natl. Acad. Sci. USA*, 94:10979–10984, 1997.
- [136] S. Makeig and J. Onton. *Oxford handbook of event-related potential components*, chapter ERP features and EEG dynamics: An ICA perspective. Oxford University Press, New York, 2009.

- [137] S. Makeig, M. Westerfield, T. Jung, S. Enghoff, J. Townsend, E. Courchesne, and T. Sejnowski. Dynamic brain sources of visual evoked responses. *Science*, 295(5555):690–694, 2002.
- [138] J. Malmivuo and R. Plonsey. *Bioelectromagnetism: Principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, New York, 1995. Available online: <http://butler.cc.tut.fi/~malmivuo/bem/book/index.htm>.
- [139] J. Malmivuo and V. Suihko. Effect of skull resistivity on the spatial resolutions of EEG and MEG. *IEEE T. Biomed. Eng.*, 51(7):1276–1280, 2004.
- [140] J. Malmivuo, V. Suihko, and H. Eskola. Sensitivity distributions of EEG and MEG measurements. *IEEE T. Biomed. Eng.*, 44(3):196–208, 1997.
- [141] A. Maritan and J. R. Banavar. Chaos, noise, and synchronization. *Phys. Rev. Lett.*, 72(10):1451–1454, 1994.
- [142] E. Masliah, R. Terry, and G. Buzsaki. Thalamic nuclei in Alzheimer disease: evidence against the cholinergic hypothesis of plaque formation. *Brain Res.*, 493:241–246, 1989.
- [143] M. Massimini, F. Ferrarelli, R. Huber, S. K. Esser, H. Singh, and G. Tononi. Breakdown of cortical effective connectivity during sleep. *Science*, 309:2228–2232, 2005.
- [144] F. Meinecke, A. Ziehe, M. Kawanabe, and K.-R. Müller. A resampling approach to estimate the stability of one-dimensional or multidimensional independent components. *IEEE T. Biomed. Eng.*, 49:1514–1525, 2002.
- [145] F. C. Meinecke, A. Ziehe, J. Kurths, and K.-R. Müller. Measuring phase synchronization of superimposed signals. *Phys. Rev. Lett.*, 94:084102, 2005.
- [146] C. M. Michel, M. M. Murray, G. Lantz, S. Gonzalez, L. Spinelli, and T. Grave de Peralta. EEG source imaging. *Clin. Neurophysiol.*, 115:2195–2222, 2004.
- [147] L. Moldegey and H. G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.*, 72(23).
- [148] R. J. Moran, K. E. Stephan, S. J. Kiebel, N. Rombach, W. T. O’Connor, K. J. Murphy, R. B. Reilly, and K. J. Friston. Bayesian estimation of synaptic physiology from the spectral responses of neural masses. *Neuroimage*, 42:272–284, 2008.

- [149] R. J. Moran, K. E. Stephan, T. Seidenbecher, H.-C. Pape, R. J. Dolan, and K. J. Friston. Dynamic causal models of steady-state responses. *Neuroimage*, 44:796–811, 2009.
- [150] D. V. Moretti, C. Babiloni, G. Binetti, E. Cassetta, G. Dal Forno, F. Ferrer, R. Ferri, B. Lanuzza, C. Miniussi, F. Nobili, G. Rodriguez, S. Salinari, and P. M. Rossini. Individual analysis of EEG frequency and band power in mild Alzheimer’s disease. *Clin. Neurophysiol.*, 115:299–308, 2004.
- [151] J. H. Morrison and P. R. Hof. Life and death of neurons in the aging brain. *Science*, 278:412–419, 1997.
- [152] J. Mosher, P. Lewis, and R. Leahy. Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE T. Biomed. Eng.*, 39(6):541–557, 1992.
- [153] J. C. Mosher, D. W. Leahy, and P. S. Lewis. EEG and MEG: Forward solutions for inverse methods. *IEEE T. Biomed. Eng.*, 46(3):245–259, 1999.
- [154] K.-R. Müller, P. Philips, and A. Ziehe. JADE_{TD}: Combining higher-order statistics and temporal information for blind source separation (with noise). In *Proc. ICA ’99, Aussois, France*, pages 87–92, 1999.
- [155] E. Niedermeyer. Alpha rhythms as physiological and abnormal phenomena. *Int. J. Psychophysiol.*, 26:31–49, 1997.
- [156] E. Niedermeyer and F. H. Lopes da Silva, editors. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Williams & Wilkins, Baltimore, 3rd edition, 1993.
- [157] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett. Identifying true brain interaction from EEG data using the imaginary part of the coherency. *Clin. Neurophysiol.*, 115:2292–2307, 2004.
- [158] G. Nolte, C. Meinecke, A. Ziehe, and K.-R. Müller. Identifying interactions in mixed and noisy complex systems. *Phys. Rev. E*, 73:051913, 2006.
- [159] G. Nolte, A. Ziehe, V. V. Nikulin, Schlogl A., N. Kramer, T. Brismar, and K. R. Muller. Robustly estimating the flow direction of information in complex physical systems. *Phys. Rev. Lett.*, 100:234101, 2008.
- [160] P. L. Nunez and R. Srinivasan. *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, New York, 2nd edition, 2006.
- [161] P. L. Nunez, B. M. Wingeier, and R. B. Silberstein. Spatial-temporal structures of human alpha rhythms: theory, microcurrent sources, multiscale measurements, and global binding of local networks. *Hum. Brain Mapp.*, 13(3):125–164, 2003.

- [162] D. Osipova, J. Ahveninen, O. Jensen, A. Ylikoski, and E. Pekkonen. Altered generation of spontaneous oscillations in Alzheimer's disease. *Neuroimage*, 27:8835–841, 2005.
- [163] D. Osipova, K. Rantanen, J. Ahveninen, R. Ylikoski, O. Häppölä, T. Strandberg, and E. Pekkonen. Source estimation of spontaneous MEG oscillations in mild cognitive impairment. *Neurosci. Lett.*, 405:57–61, 2006.
- [164] R. K. Otnes and L. Enochson. *Digital time series analysis*. Wiley, New York, 1972.
- [165] E. Palmero Soler, K. Dolan, V. Hadamschek, and P. A. Tass. swLORETA: a novel approach to robust source localization and synchronization tomography. *Phys. Med. Biol.*, 52:1783–1800, 2007.
- [166] J. J. Palop, J. Chin, and L. Mucke. A network dysfunction perspective on neurodegenerative diseases. *Nature*, 443:768–773, 2006.
- [167] K. Parameshwaran, M. Dhanasekaran, and V. Suppiramaniam. Amyloid beta peptides and glutamatergic synaptic dysregulation. *Exp. Neurol.*, 210:7–13, 2008.
- [168] L. Parra and P. Sajda. Blind source separation via generalized eigenvalue decomposition. *J. Mach. Learn. Res.*, 4:1261–1269, 2003.
- [169] R. D. Pascual-Marqui. Standardized low resolution brain electromagnetic tomography (sLORETA). *Meth. Find. Exp. Clin.*, 24D:5–12, 2002.
- [170] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann. Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.*, 18:49–65, 1994.
- [171] C. Paulus and J. I. Mars. New multicomponent filters for geophysical data processing. *IEEE. T. Geosci. Remote*, 44(8):2260–2270, 2006.
- [172] F. Pesarin. *Multivariate permutation tests: with applications in biostatistics*. John Wiley & sons, New York, 2001.
- [173] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen. Mild cognitive impairment. clinical characterization of the outcome. *Arch. Neurol.*, 56:303–308, 1999.
- [174] G. Pfurtscheller, A. Stancak, and C. Neuper. Event-related synchronization (ERS) in the alpha band—an electrophysiological correlate of cortical idling: a review. *Int. J. Psychophysiol.*, 24:39–46, 1996.
- [175] D.-T. Pham. Blind separation of instantaneous mixture of sources via an independent component analysis. *IEEE T. Signal Proces.*, 44(11):2768–2779, 1996.

- [176] D.-T. Pham and J.-F. Cardoso. Blind separation of instantaneous mixtures of non stationary sources. *IEEE T. Signal Proces.*, 49(9):1837–1848, 2001.
- [177] K. Raju, T. Ristaniemi, J. Karhunen, and E. Oja. Jammer suppression in DS-CDMA arrays using independent component analysis. *IEEE T. Wirel. Commun.*, 5(1):77–82, 2006.
- [178] G. C. Reinsel. *Elements of Multivariate Time Series Analysis*. Springer, 2003.
- [179] J. E. Richards. Recovering dipole sources from scalp-recorded event-related-potentials using component analysis: principal component analysis and independent component analysis. *Int. J. Psychophysiol.*, 54:201–220, 2004.
- [180] E. Rodriguez, N. George, J.-P. Lachaux, J. Martinerie, B. Renault, and F. J. Varela. Perception’s shadow: long-distance synchronization of human brain activity. *Nature*, 397:430–433, 1999.
- [181] P. R. Roelfsema, A. K. Engel, P. König, and W. Singer. Visuomotor integration is associated with zero time lag synchronization among cortical areas. *Nature*, 385:157–161, 1997.
- [182] S. A. Rombouts, F. Barkhof, R. Goekoop, C. J. Stam, and P. Scheltens. Altered resting state networks in mild cognitive impairment and mild Alzheimer’s disease: an fMRI study. *Hum. Brain Mapp.*, 26:231–239, 2005.
- [183] H. A. Rowley, P. E. Grant, and T. P. Roberts. Diffusion MR imaging. Theory and applications. *Neuroimag. Clin. N. Ann.*, 9:343–361, 1999.
- [184] K. Rutanen. TIM C++ library. Available online: <http://www.tut.fi/tim>.
- [185] K. Rutanen. Estimation of entropy combinations. Technical report, Department of Signal Processing, Tampere University of Technology, 2009.
- [186] Y. Saito and H. Harashima. Tracking of information within multichannel record: causal analysys in EEG. In N. Yamaguchi and K. Fujisawa, editors, *Recent advances in EEG and EMG data processing*, pages 133–146. Elsevier, Amsterdam, The Netherlands, 1981.
- [187] S. Sano, A. Uchida, S. Yoshimori, and R. Roy. Dual synchronization of chaos in Mackey-Glass electronic circuits with time-delayed feedback. *Phys. Rev. E*, 75:016207, 2007.
- [188] J. Santamaria and K. H. Chiappa. *The EEG of drowsiness*. Demos Medical Publishing, New York, 1987.

- [189] J. Särelä and R. Vigário. Overlearning in marginal distribution-based ICA: Analysis and solutions. *J. Mach. Learn. Res.*, 4:1447–1469, 2003.
- [190] M. Scherg and D. Von Cramon. Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. *Electroencephalogr. Clin. Neurophysiol.*, 62(1):32–44, 1985.
- [191] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer. Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Phys. Rev. E*, 54(6):6708, 1996.
- [192] R. O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE T. Antenn. Propag.*, 34(3):276–280, 1986.
- [193] T. Schneider and A. Neumaier. Algorithm 808: ARFIT - a matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM T. Math. Soft.*, 27(1):58–65, 2001.
- [194] M. Schreckberger, C. Lange-Asschenfeld, M. Lochmann, K. Mann, T. Siessmeier, H.-G. Buchholz, P. Bartenstein, and Gründer. The thalamus as the generator and modulator of EEG alpha rhythm: a combined PET/EEG study with lorazepam challenge in humans. *NeuroImage*, 22:637–644, 2004.
- [195] T. Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85(2):461–464, 2000.
- [196] M. Schurmann and E. Başar. Functional aspects of alpha oscillations in the EEG. *Int. J. Psychophysiol.*, 39:151–158, 2001.
- [197] M. Schurmann, T. Demiralp, E. Başar, and C. Başar-Eroglu. Electroencephalogram alpha (8–15 Hz) responses to visual stimuli in cat cortex, thalamus, and hippocampus: a distributed alpha network? *Neurosci. Lett.*, 292:175–178, 2000.
- [198] T. Schwarz. Estimating the dimension of a model. *Ann. Stat.*, 6:461–464, 1978.
- [199] S. H. Scott. Converting thoughts into action. *Nature*, 442:141–142, 2006.
- [200] D. J. Selkoe. Alzheimer’s disease is synaptic failure. *Science*, 298:789–791, 2002.
- [201] C. E. Shannon and W. Weaver. *The mathematical theory of information*. University of Illinois Press, Urbana, IL, 1949.
- [202] L. R. Silva, Amitai Y., and Connors B. W. Intrinsic oscillations of neocortex generated by layer 5 pyramidal neurons. *Science*, 251:432–435, 1991.

- [203] B. W. Silverman. Algorithm as 176: Kernel density estimation using the fast fourier transform. *J. Roy. Stat. Soc. C-App.*, 31(1):93–99, 1982.
- [204] D. H. Small. Network dysfunction in Alzheimer’s disease: does synaptic scaling drive disease progression? *Trends Molec. Med.*, 14:103–108, 2008.
- [205] M. Small. *Applied nonlinear time series analysis: applications in physics, physiology and finance*. World Scientific, 2005.
- [206] H. Späth. *Cluster analysis algorithms for data reduction and classification of objects*. Ellis Horwood, Chichester, U.K., 1980.
- [207] M. Steriade, P. Gloor, R. R. Llinas, F. H. Lopes da Silva, and M. M. Mesulam. Report of IFCN committee on basic mechanisms. basic mechanisms of cerebral rhythmic activities. *Electroencephalogr. Clin. Neurophysiol.*, 76:481–508, 1990.
- [208] M. Steriade, E. G. Jones, and R. R. Llinas. *Thalamic oscillations and signaling*. Wiley Interscience, New York, 1990.
- [209] M. Steriade, E. G. Jones, and D. A. McCormick. *Thalamus*. Elsevier, Amsterdam, 1997.
- [210] H. Stögbauer, A. Kraskov, S. A. Astakhov, and P. Grassberger. Least dependent component analysis based on mutual information. *Phys. Rev. E* 70, 6, 2004.
- [211] G. G. Supp, A. Schlögl, N. Trujillo-Barreto, M. M. Müller, and T. Gruber. Directed cortical information flow during human object recognition: analyzing induced EEG gamma-band responses in brain’s source space. *PLoS One*, 2(8):e684, 2007. doi:10.1371/journal.pone.0000684.
- [212] F. Takens. Detecting strange attractors in turbulence. *Lect. Notes in Math.*, 898:366–381, 1981.
- [213] A. C. Tang, B. A. Pearlmutter, N. A. Malaszenko, D. Phung, and B. C. Reeb. Independent components of magnetoencephalography: Localization. *Neural Computation*, 14(8):1827–1858, 2002.
- [214] P. Tichavský, Z. Koldovský, E. Doron, A. Yeredor, and G. Gómez-Herrero. Blind signal separation by combining two ICA algorithms: HOS-based EFICA and time structure-based WASOBI. In *Proc. EU-SIPCO 2006, Florence, Italy*, pages 1–5, 2006.
- [215] P. Tichavský, Z. Koldovský, A. Yeredor, G. Gómez-Herrero, and E. Doron. A hybrid technique for blind separation of non-Gaussian and time-correlated sources using a multicomponent approach. *IEEE T. Neural Networks*, 19(3):421–430, March 2008.

- [216] P. Tichavský, A. Yeredor, and J. Nielsen. A fast approximate joint diagonalization algorithm using a criterion with a block diagonal weight matrix. In *Proc. ICASSP 2008*, pages 3321–3324, 2008.
- [217] N. J. Trujillo-Barreto, E. Aubert-Vazquez, and P. A. Valdes-Sosa. Bayesian model averaging in EEG/MEG imaging. *Neuroimage*, 21:1300–1319, 2004.
- [218] P. J. Uhlhass and W. Singer. Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. *Neuron.*, 52:155–168, 2006.
- [219] B. D. Van Veen and K. M. Buckley. Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Mag.*, 5:4–24, 1988.
- [220] B. D. Van Veen, W. Van Drongelen, M. Yuchtman, and A. Suzuki. Localization of brain electrical activity via linear constrained minimum variance spatial filtering. *IEEE T. Biomed. Eng.*, 44(9):867–880, 1997.
- [221] O. Vasicek. A test for normality based on sample entropy. *J. Roy. Stat. Soc. B*, 38(1):54–59, 1976.
- [222] P. F. Verdes. Assessing causality from multivariate time series. *Phys. Rev. E*, 72:026222, 2005.
- [223] A. Vergult, W. De Clerq, A. Palmmini, B. Vanrumste, P. Dupont, S. Van Huffel, and W. Van Paesschen. Improving the interpretation of ictal scalp EEG: BSS-CCA algorithm for muscle artifact removal. *Epilepsia*, 48(5):950–958, 2007.
- [224] R. Vicente, L. L. Gollo, C. R. Mirasso, I. Fischer, and G. Pipa. Dynamical relaying can yield zero time lag neuronal synchrony despite long conduction delays. *Proc. Nat. Acad. Sci. USA*, 105(44):17157–17162, 2008.
- [225] F. Vrins and M. Versleynsen. On the entropy minimization of a linear mixture of variables for source separation. *Signal Process.*, 85:1029–1044, 2005.
- [226] K. Wang, M. Liang, L. Wang, L. Lian, X. Zhang, K. Li, and T. Jiang. Altered functional connectivity in early Alzheimer’s disease: a resting-state fMRI study. *Hum. Brain Mapp.*, 28:967–978, 2007.
- [227] N. Wiener. In E. F. Beckenbach, editor, *Modern mathematics for engineers*. McGraw-Hill, New York, 1956.
- [228] A. Yeredor. Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting. *IEEE Signal Proc. Lett.*, 7(7):197–200, 2000.

- [229] A. Yeredor. Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. *IEEE T. Signal Process.*, 50(7):1545–1553, 2002.
- [230] T. Yoshimoto, M. Kotani, S. Kuriki, H. Karibe, and N. Nakasato, editors. *Functional neuroimaging by synthetic aperture magnetometry (SAM)*, pages 302–305. Tohoku Univ. Press, Sendai, Japan, 1999.
- [231] F. Zanow and T. R. Knösche. ASA - advanced source analysis of continuous and event-related EEG/MEG signals. *Brain Topography*, 16(4):287–290, 2004.
- [232] A. Zbrozek. Costs of formal dementia care in Europe. In *Program and abstracts of the 9th International Conference on Alzheimer’s Disease and Related Disorders*, Philadelphia, USA, July 2004.
- [233] A. Ziehe and K.-R. Müller. TDSEP—an efficient algorithm for blind separation using time structure. In *Proc. ICANN’98*, pages 675–680, 1998.