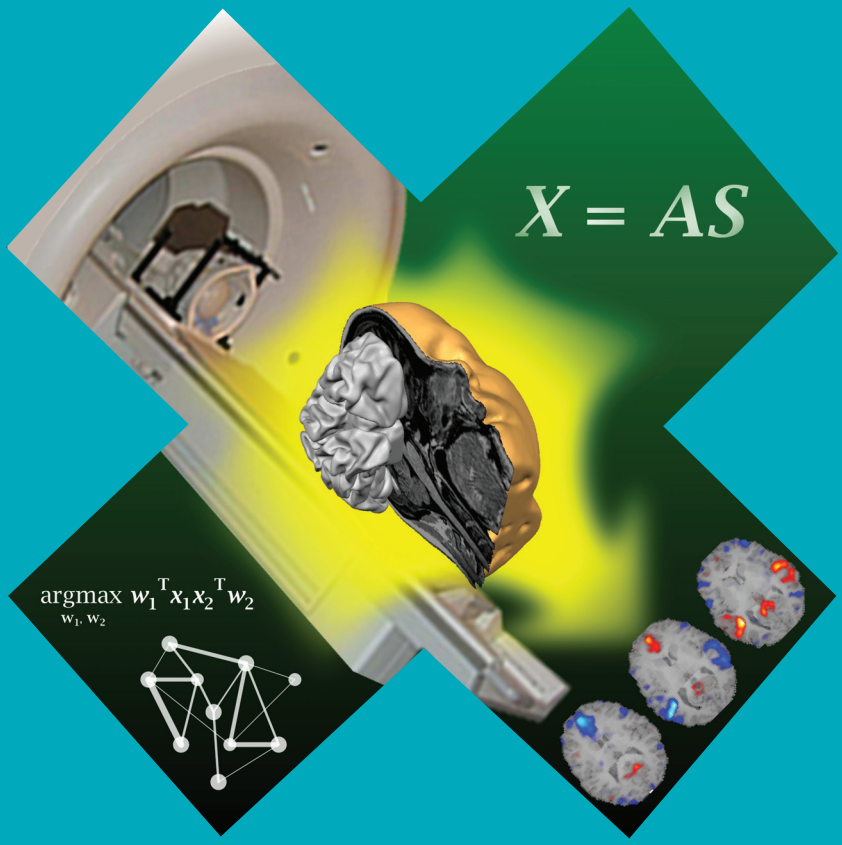
Department of Information and Computer Science

Data-driven Analysis for Natural Studies in Functional Brain Imaging

Jarkko Ylipaavalniemi



Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 70/2013

Data-driven Analysis for Natural Studies in Functional Brain Imaging

Jarkko Ylipaavalniemi

A doctoral dissertation completed for the degree of Doctor of Science in Technology to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall T2 of the school on 3 May 2013 at 12.

Aalto University
School of Science
Department of Information and Computer Science

Supervising professor
Prof. Erkki Oja

Thesis advisor
Doc. Ricardo Vigário

Preliminary examiners
Adj. Prof. Vesa Kiviniemi, University of Oulu, Finland
Assoc. Prof. Morten Mørup, Technical University of Denmark, Denmark

Opponent
Prof. Klaus-Robert Müller, Technical University of Berlin, Germany

Aalto University publication series
DOCTORAL DISSERTATIONS 70/2013

ISBN 978-952-60-5136-9 (printed)
ISBN 978-952-60-5137-6 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-5137-6

Unigrafia Oy
Helsinki 2013

Finland

441 697
Printed matter

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Jarkko Ylipaavalniemi

Name of the doctoral dissertation
Data-driven Analysis for Natural Studies in Functional Brain Imaging

Publisher School of Science

Unit Department of Information and Computer Science

Series Aalto University publication series DOCTORAL DISSERTATIONS 70/2013

Field of research Computer and Information Science

Manuscript submitted 5 November 2012 **Date of the defence** 3 May 2013

Permission to publish granted (date) 20 December 2012 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

Abstract

In neuroscience, functional magnetic resonance imaging (fMRI) has become a powerful tool in human brain mapping. Typically, fMRI is used with a rather simple stimulus sequence, aiming at improving signal-to-noise ratio for statistical hypothesis testing. When natural stimuli are used, the simple designs are no longer appropriate. The aim of this thesis is in developing data-driven approaches for reliable inference of brain correlates to natural stimuli.

Since the beginning of the nineteenth century, neuroscience has focused on the idea that distinct regions of the brain support particular mental processes. However, modern research recognizes that many functions rely on distributed networks, and that a single brain region may participate in more than one function. These rapid paradigm changes in neuroscience raise important methodological challenges. Purely hypothesis-driven methods have been used extensively in functional imaging studies. As the focus in brain research is shifting away from functional specialization towards interaction-based functional networks, those approaches are no longer appropriate. In contrast to the classic statistical hypothesis testing approaches, modern machine learning methods allow for a purely data-driven way to describe the data. They do not use the stimuli, and make no assumptions about whether the brain processes are stimulus related or not. The recordings for each brain region may contain a complicated mixture of activity, which is produced by many spatially distributed processes, and artifacts. Each process can be described as a component having a separate time series and spatial extent, and producing simultaneous changes in the fMRI signals of many regions.

The main contribution of the thesis is a reliable independent component analysis (ICA) approach, which is available in the Arabica toolbox. The usefulness of the approach was tested extensively with fMRI data, showing that the method is capable of providing insights into the data that would not be attainable otherwise. The new method was also theoretically analyzed and its asymptotic convergence was proven. The theory offers a thorough explanation of how the method works and justifies its use in practice. Then, the new method is further developed for analyzing networks of distributed brain activity, by combining it with canonical correlation analysis (CCA). The extension was shown to be particularly useful with fMRI studies that use natural stimuli. The approach is further extended to be applicable in cases where independent subspaces emerge, which often happens when using real measurement data that is not guaranteed to fit all the assumptions made in the development of the methods.

Keywords Canonical correlation analysis, independent component analysis, functional elements, functional networks, functional magnetic resonance imaging, natural stimulation, subspaces, reliability

ISBN (printed) 978-952-60-5136-9 **ISBN (pdf)** 978-952-60-5137-6

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Espoo **Location of printing** Helsinki **Year** 2013

Pages 173 **urn** http://urn.fi/URN:ISBN:978-952-60-5137-6

Tekijä
Jarkko Ylipaavalniemi

Väitöskirjan nimi
Aineistolähtöinen analysointi luonnollisille koeasetelmille toiminnallisessa aivokuvantamisessa

Julkaisija Perustieteiden korkeakoulu

Yksikkö Tietojenkäsittelytieteen laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 70/2013

Tutkimusala Informaatiotekniikka

Käsikirjoituksen pvm 05.11.2012 **Väitöspäivä** 03.05.2013

Julkaisuluvan myöntämispäivä 20.12.2012 **Kieli** Englanti

☐ **Monografia** ☒ **Yhdistelmäväitöskirja (yhteenveto-osa + erillisartikkelit)**

Tiivistelmä

Toiminnallisesta magneettiresonassikuvantamisesta (fMRI) on tullut tehokas työkalu neurotieteessä ihmisaivojen toiminnan kartoittamiseen. Yleensä fMRI-menetelmän yhteydessä käytetään melko yksinkertaista ärsykkeiden sarjaa, jolla pyritään hyvään signaali-kohinasuhteeseen tilastollista hypoteesitestausta varten. Kun siirrytään kohti luonnollisia ärsykkeitä, tällaiset yksinkertaiset koeasetelmat eivät ole enää päteviä. Tämän väitöstyön tavoite on ollut kehittää uusia aineistolähtöisiä menetelmiä luonnollisten ärsykkeiden aiheuttamien aivovasteiden luotettavaan tunnistamiseen.

Aina 1800-luvun alusta lähtien neurotieteessä on vallinnut käsitys siitä, että erilliset aivoalueet vastaavat tiettyjä neuropsykologisia toimintoja. Moderni tutkimus on kuitenkin osoittanut, että monet toiminnot perustuvat hajautettuihin aivoalueiden verkostoihin ja kukin aivoalue voi toimia osana useampaa verkostoa. Puhtaasti hypoteesitestaukseen perustuvat menetelmät ovat olleet hyvin suosittuja aivokuvantamisen yhteydessä. Kun tutkimus keskittyy yhä enemmän toiminnallisen erikoistumisen sijaan vuorovaikutuksiin perustuviin verkostoihin, tällaisia menetelmiä ei voida enää käyttää. Niiden sijaan modernit koneoppimismenetelmät mahdollistavat puhtaasti aineistolähtöisen tavan mittausten selittämiseen. Nämä menetelmät eivät tarvitse tietoa ärsykkeistä eivätkä tee oletuksia siitä ovatko mittaustulokset ärsykkeistä riippuvaisia. Mittausten oletetaan muodostuvan usean aktivaation sekoituksista, jotka voivat sisältää monia sijainniltaan hajautettuja prosesseja, sekä mittausvirheitä. Jokainen prosessi voidaan selittää komponenttina, jolla on muista eriävä aikasarja ja sijaintikartta.

Väitöskirjan pääasiallinen tulos on luotettava riippumattomien komponenttien menetelmä (ICA), joka on vapaasti saatavilla Arabica-ohjelmistossa. Uuden menetelmän toimivuutta on testattu usealla fMRI-aineistolla ja osoitettu, että menetelmä kykenee selittämään mittaukset uudella tavalla ja tuottamaan tietoa aineistosta, joka ei muuten olisi mahdollista. Menetelmän toimivuus on todistettu myös teoreettisen tilastotieteen keinoin, joka tarjoaa myös kattavat perusteet menetelmän käytännön hyödyntämiseen. Menetelmää myös jatkokehitettiin hajautettujen verkostojen analysointiin sopivaksi yhdistämällä se kanonisen korrelaatioanalyysin (CCA) kanssa. Tämän laajennuksen osoitettiin soveltuvan luonnollisia ärsykkeitä käyttävien fMRI-aineistojen analysointiin. Lopuksi menetelmää parannettiin myös sellaisia tilanteita varten, joissa havaitaan riippumattomia aliavaruuksia. Näin tapahtuu usein käytännössä, sillä aineiston ei voida taata täyttävän kaikkia menetelmissä tehtyjä oletuksia.

Avainsanat Aliavaruudet, toiminnallinen magneettiresonanssikuvantaminen, funktionaaliset verkostot, funktionaaliset yksiköt, kanoninen korrelaatioanalyysi, luonnolliset ärsykkeet, luotettavuus, riippumattomien komponenttien menetelmä

ISBN (painettu) 978-952-60-5136-9 **ISBN (pdf)** 978-952-60-5137-6

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Espoo **Painopaikka** Helsinki **Vuosi** 2013

Sivumäärä 173 **urn** http://urn.fi/URN:ISBN:978-952-60-5137-6

Preface

This work has been carried out in the Adaptive Informatics Research Centre at the Department of Information and Computer Science of the Aalto University School of Science, previously called the Helsinki University of Technology. The work has been mainly funded by the university and the Finnish Academy, with additional funding from the Helsinki Doctoral Programme in Computer Science (Hecse). I have also received grants from Tekniikan edistämissäätiö (TES) and Jenny ja Antti Wihurin rahasto.

I wish to thank my supervisor Professor Erkki Oja for providing the inspirational research conditions that have been created in the laboratory. I would also like to thank my instructor Docent Ricardo Vigário for the excellent and always fun guidance he has provided as the head of the neuroinformatics research group. Under his guidance I have also learned a lot about teaching. I am also grateful to the whole neuroinformatics group for maintaining a certain sense of humor, even during all the transitions in the university. I would also like to thank all my room mates during the years, particularly my co-author Nima Reyhani for the highly theoretical, as well as, all the otherwise amusing discussions.

The neuroscience data for this thesis has been provided by the Brain Research Unit at the O.V. Lounasmaa Laboratory, previously called the Low Temperature Laboratory, of the Aalto University. I would like to thank my colleagues there for the enlightening discussions concerning neuroscience, and especially Academician Riitta Hari for her invaluable support. Likewise, I am most thankful to my other co-authors, Samuel Kaski, Eerika Savia, Sanna Malinen, Seppo Mattila, Antti Tarkiainen, and Jyri Soppela, for their interdisciplinary expert knowledge and lively discussions.

I would like to express my gratitude to the pre-examiners of this thesis, Adjunct Professor Vesa Kiviniemi and Associate Professor Morten Mørup,

for their comments that helped improve the quality of this thesis.

I would also like to thank my parents, and my whole family, for raising me to be the curious creature I am. For keeping my life in balance between theory and nonsense, I am also grateful for all my friends outside of work.

Finally, I wish to thank my fiancée Johanna for everything. Without her, this project may have never ended.

Espoo, April 5, 2013,

Jarkko Ylipaavalniemi

Contents

Preface 7

Contents 9

List of Publications 11

List of Abbreviations 15

List of Symbols 17

1. Introduction 19

1.1 Current Trends in Neuroscience 19

1.2 Data-driven Analysis 21

1.3 Contributions and Organization of the Thesis 21

2. Functional Brain Imaging 23

2.1 Background 23

2.2 Magnetic Resonance Imaging 25

2.3 Functional Magnetic Resonance Imaging 29

2.3.1 Blood Oxygenation Level Dependent Signal 30

2.3.2 Experimental Settings and Stimulation 32

3. Preprocessing and Statistical Analysis 35

3.1 Typical Preprocessing 36

3.1.1 Slice Time Correction 36

3.1.2 Motion Correction 36

3.1.3 Coregistration 37

3.1.4 Functional-structural Coregistration 37

3.1.5 Spatial Normalization 37

3.1.6 Spatial Smoothing 38

3.1.7 Temporal Filtering 39

3.2 Hypothesis Testing 39
3.2.1 General Linear Model 40

4. Analysis Based on Dependence and Independence 41
4.1 Motivation 41
4.2 Principal Component Analysis 42
4.3 Canonical Correlation Analysis 44
4.4 Independent Component Analysis 45
4.4.1 Blind Source Separation 47
4.4.2 Independence and Non-Gaussianity 47
4.4.3 The FastICA Algorithm 48
4.4.4 Spatial ICA of fMRI Data 50

5. Uncertainty and Exploiting Variability 51
5.1 Errors in Assumptions and Estimation 52
5.2 Sparseness and Overfitting 53
5.3 Bootstrap and Combining Estimates 54
5.4 The Arabica Toolbox 56
5.5 Reliable Analysis using ICA 60

6. Networks of Related Independent Components 63
6.1 Shared Dependencies Between Datasets 64
6.2 Features of Natural Stimuli 65
6.3 Semi-blind Approach in Two Steps 65
6.3.1 Identifying Functional Elements with ICA 66
6.3.2 Linking Functional Networks with CCA 66
6.4 Networks of Brain Activation Related to Stimuli 67

7. Subspaces of Independent Components 69
7.1 Subspaces Emerging from ICA 70
7.2 Two Ways to Identify Subspaces 70
7.2.1 Estimates Linked Through Correlation 70
7.2.2 Estimates with Shared Variability 71
7.3 Further Analysis of Subspaces 71
7.3.1 Blind Refinement 72
7.3.2 Refinement Using Additional Data 72

8. Discussion 73

Bibliography 77

List of Publications

This thesis consists of an overview and of the following publications, which are referred to in the text by their Roman numerals.

I Jarkko Ylipaavalniemi and Ricardo Vigário. Analysis of Auditory fMRI Recordings via ICA: A Study on Consistency. In *Proceedings of IJCNN'04, International Joint Conference on Neural Networks*, pages 249–254. IEEE, Piscataway, NJ, 2004. DOI 10.1109/IJCNN.2004.1379908.

II Jarkko Ylipaavalniemi, Seppo Mattila, Antti Tarkiainen and Ricardo Vigário. Brains and Phantoms: An ICA Study of fMRI. In Justinian Rosca, Deniz Erdogmus, José C. Príncipe and Simon Haykin (Eds.): *Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation, ICA 2006*, pages 503–510. Springer, Berlin/Heidelberg, 2006. DOI 10.1007/11679363_63.

III Jarkko Ylipaavalniemi and Ricardo Vigário. Analyzing Consistency of Independent Components: An fMRI Illustration. *NeuroImage*, 39:1, pages 169–180, 2008. DOI 10.1016/j.neuroimage.2007.08.027.

IV Jarkko Ylipaavalniemi and Jyri Soppela. Arabica: Robust ICA in a Pipeline. In Tülay Adali, Christian Jutten, João Marcos Travassos Romano and Allan Kardec Barros (Eds.): *Proceedings of the 8th International Conference on Independent Component Analysis and Blind Signal Separation, ICA 2009*, pages 379–386. Springer, Berlin/Heidelberg, 2009. DOI 10.1007/978-3-642-00599-2_48.

V Nima Reyhani, Jarkko Ylipaavalniemi, Ricardo Vigário and Erkki Oja. Consistency and Asymptotic Normality of FastICA and Bootstrap FastICA. *Signal Processing*, 92:8, pages 1767—1778, 2012. DOI 10.1016/j.sigpro.2011.11.025.

VI Jarkko Ylipaavalniemi, Eerika Savia, Ricardo Vigário and Samuel Kaski. Functional Elements and Networks in fMRI. In Wei Zhang and Ilya Shmulevich (Eds.): *Proceedings of the 15th European Symposium on Artificial Neural Networks, ESANN 2007*, pages 561–566. D-side Publications, Bruxelles, Belgium, 2007.

VII Jarkko Ylipaavalniemi, Eerika Savia, Sanna Malinen, Riitta Hari, Ricardo Vigário and Samuel Kaski. Dependencies Between Stimuli and Spatially Independent fMRI Sources: Towards Brain Correlates of Natural Stimuli. *NeuroImage*, 48:1, pages 176–185, 2009. DOI 10.1016/j.neuroimage.2009.03.056.

VIII Jarkko Ylipaavalniemi and Ricardo Vigário. Subspaces of Spatially Varying Independent Components in fMRI. In Mike E. Davies, Christopher J. James, Samer A. Abdallah and Mark D. Plumbley (Eds.): *Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation, ICA 2007*, pages 665–672. Springer, Berlin/Heidelberg, 2007. DOI 10.1007/978-3-540-74494-8_83.

Author's Contribution in the Publications

Publication **I** is a feasibility study testing the consistency of independent components in a multiple run independent component analysis (ICA) approach. The author was responsible for implementing the method and performing all the experiments. The article was written together with the co-author.

It was shown in Publication **II** that the approach can be used to gain insight into the data that is not attainable with single run approaches. The author was responsible for the main idea, developing the methods, and performing all the experiments. The article was mostly written by the author.

In Publication **III**, the motivation and application to functional magnetic resonance imaging (fMRI) data is thoroughly discussed. The author was responsible for all the experiments, and the article was mainly written by the author.

Publication **IV** introduces the reliable ICA toolbox, called Arabica. The author was responsible for the design of the toolbox and most of the implementation. The article was fully written by the author.

Publication **V** proves the asymptotic convergence of the bootstrap ICA method and thoroughly discusses the theory behind the method. The author was mainly responsible for the implementation and experiments. The article was written together with the co-authors.

Publication **VI** introduces a new two-step framework for studying activation networks. Publication **VII** improves the method and studies the framework with better controlled measurements. In Publications **VI** and **VII** the author was responsible for the main idea, together with the co-authors, and for performing all ICA experiments. Both articles were written in collaboration.

Finally, Publication **VIII** shows the importance of taking the emergence

of independent subspaces into account to be able to correctly interpret the components. The author was responsible for the main idea, together with the co-author, for implementing all the methods, and running all the experiments. The article was mostly written by the author.

List of Abbreviations

Methodology:

BSS	Blind Source Separation (problem)
CCA	Canonical Correlation Analysis
DCM	Dynamic Causal Modeling
GLM	General Linear Model
ICA	Independent Component Analysis
ISA	Independent Subspace Analysis
PCA	Principal Component Analysis

Neuroscience:

BOLD	Blood Oxygenation Level Dependent (signal changes in fMRI)
CNS	Central Nervous System
fMRI	Functional Magnetic Resonance Imaging
MRI	Magnetic Resonance Imaging

Organizations:

INCF	International Neuroinformatics Coordinating Facility
OECD	Organisation for Economic Co-operation and Development
WHO	World Health Organization

List of Symbols

N	dimensionality of a dataset
N_i	dimensionality of the i'th dataset
V	number of samples in a dataset, or number of voxels in fMRI
M	number of regressors in GLM
K	number of components in dimensionality reduction
z	a scalar random variable
μ_z	mean of random variable z
$\boldsymbol{x}$	a vectorial random variable $\in \mathbb{R}^N$
$\boldsymbol{z}$	a vectorial random variable $\in \mathbb{R}^K$
$\boldsymbol{w}$	a projection vector
$\boldsymbol{W}$	a weight matrix $\in \mathbb{R}^{N \times K}$
$\boldsymbol{X}$	a data matrix $\in \mathbb{R}^{N \times V}$
$\boldsymbol{X}_i$	the i'th data matrix $\in \mathbb{R}^{N_i \times V}$
$\boldsymbol{G}$	design matrix in GLM $\in \mathbb{R}^{N \times M}$
$\boldsymbol{\beta}$	parameter matrix in GLM $\in \mathbb{R}^{M \times V}$
$\boldsymbol{\epsilon}$	residual error matrix in GLM $\in \mathbb{R}^{N \times V}$
$\boldsymbol{A}$	mixing matrix in ICA $\in \mathbb{R}^{N \times K}$
$\boldsymbol{S}$	source matrix of independent components in ICA $\in \mathbb{R}^{K \times V}$
argmax	argument that maximizes subsequent expression

Matrices:

$\boldsymbol{x}^\top,\ \boldsymbol{X}^\top$	transpose of a vector or a matrix
$\boldsymbol{A}^{-1}$	inverse matrix of matrix $\boldsymbol{A}$
$\boldsymbol{I}$	identity matrix
$\| \cdot \|$	L_2-norm of a vector or a matrix
diag	vector of the diagonal elements of a matrix
λ	eigenvalue of a matrix
$\boldsymbol{\Lambda}$	diagonal matrix of eigenvalues in decreasing order

Probability:

Pr	a probability
$\xrightarrow{P}$	converges in probability
$\mathbb{E}[\,\cdot\,]$	expected value of a random variable
$p(s)$	probability density of random variable s
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$H(y)$	entropy of random variable y
$\mathcal{S}^d$	a spherical manifold with d dimensions

Dependency:

ρ	correlation
$\boldsymbol{C}$	covariance matrix
$\boldsymbol{C}_i$	covariance matrix of the i'th dataset in CCA
$\boldsymbol{C}_{ij}$	cross-covariance matrix between the i'th and j'th datasets in CCA

1. Introduction

In neuroscience, functional magnetic resonance imaging (fMRI) has become a powerful tool in human brain mapping. Typically, fMRI is used with a rather simple stimulus sequence, aimed at improving signal-to-noise ratio for statistical hypothesis testing. When natural stimuli are used, the simple designs are no longer appropriate. The aim of this thesis is in developing data-driven approaches for reliable inference of brain correlates to natural stimuli.

1.1 Current Trends in Neuroscience

The focus in brain research is shifting away from functional specialization towards interaction-based functional networks [see, *e.g.*, Tononi et al., 1994, McIntosh, 2000, Friston, 2002, Hari and Kujala, 2009]. At the same time, interest in the dynamics of large-scale neuronal populations has gained attraction, for a review see Deco et al. [2008]. Brain connectivity is studied with both healthy subjects and patients suffering from neurological diseases, such as Alzheimer, schizophrenia, or epilepsy. These rapid paradigm changes in neuroscience raise important methodological challenges. In such uncontrolled setups, it is extremely difficult to differentiate the stimulus-related processes from ongoing brain activity or, analogously, the stimulus features related to brain activity from all other aspects of the natural environment.

Since the beginning of the nineteenth century, neuroscience has focused on the idea that distinct regions of the brain support particular mental processes. However, modern research recognizes that many functions rely on distributed networks, and that a single brain region may participate in more than one function [Friston, 2002, Hari and Kujala, 2009].

Another emerging field of research, also interested in distributed net-

works of brain activity, deals with functional imaging of resting subjects, without a particular task to perform. Studies on resting state networks often focus on the so-called default-mode network [see, *e.g.*, Damoiseaux et al., 2006, Raichle and Snyder, 2007]. The default network has been hypothesized to generate spontaneous thoughts during mind-wandering and to be an essential component of creativity. Alternatively, default mode activity may represent underlying physiological processes going on in the brain that are unrelated to any particular thought.

Up to now, purely hypothesis-driven statistical methods, like general linear model (GLM), have been used extensively in functional imaging studies, which have focused on rather simple block designs aimed at optimizing stimulus control and signal-to-noise ratio. Currently, the focus is shifting from simple unimodal stimuli towards integration of multiple sensory stimuli to study cognitive processes and, generally, brain activation related to natural stimuli.

The recent interest in more natural setups is evident in experiments with real-life stimuli. Studies using movies as stimuli have been able to model signals in one brain with the activity from another brain [Hasson et al., 2004, Bartels and Zeki, 2005]. The results have revealed a surprising tendency of individual brains to operate coherently during natural viewing. Moreover, remarkable reconstructions of the viewed images were produced with a decoder model in another study using video clips [Nishimoto et al., 2011]. These results demonstrate that dynamic brain activity measured under naturalistic conditions can be decoded and provides a powerful environment to reveal connectivity in the brain. Furthermore, human interaction has been recently studied using a first-person video game as stimuli [Kätsyri et al., 2012], showing that winning activates the brain's reward circuitry differently depending on whether the opponent is another human player or a computer.

As natural stimuli are increasingly used in fMRI studies, challenges arise in analyzing the measurements. It is no longer feasible to assume that a single experimental variable could account for the brain activity, and it is extremely difficult to differentiate the stimulus-related processes from ongoing brain activity. Instead, relevant combinations of a rich set of stimulus features could explain the more complex activation patterns.

1.2 Data-driven Analysis

In contrast to the classic statistical hypothesis testing approaches, modern machine learning methods allow for a purely data-driven way to describe the data. They do not depend on the stimuli, and make no assumptions about whether the brain processes are stimulus related or not. One of the most widely used methods for data-driven signal decomposition is independent component analysis (ICA, Hyvärinen et al. [2001b]).

The recordings for each brain region may contain a complicated mixture of activity, which is produced by many spatially distributed processes, and may be corrupted by artifacts. Each process can be described as a component having a separate time series and spatial extent, and producing simultaneous changes in the fMRI signals of many regions.

However, some problems exist with the use of ICA. One concern is the tendency of the estimated independent components to change each time the analysis is performed. This behavior can be caused by many factors both in the algorithm and in the data, making it difficult to assess the reliability of the results.

1.3 Contributions and Organization of the Thesis

The main contribution of the thesis is a reliable ICA approach, made available in the Arabica toolbox. Unlike previously existing methods that rely on a single run, the new approach utilizes multiple runs of ICA with bootstrap. In addition to providing reliable signal estimates, the new method can provide additional information on the data, that is not possible with the previous methods. As a second contribution, the method is further developed for analyzing networks of brain activity, by combining it with another data-driven method called canonical correlation analysis (CCA, Hotelling [1936]), which looks for dependencies between two datasets. The third contribution extends the reliable ICA approach to be applicable in cases where independent subspaces emerge, which often occurs in practice, when using measurement data that does not completely fulfill the assumptions made by the algorithm.

The introductory part of the thesis is organized as follows: Chapter 2 gives a quick overview of the human brain and introduces functional brain imaging, with the focus on fMRI. Then, Chapter 3 describes the standard preprocessing and statistical analysis that is applied to nearly all

recorded fMRI data. In Chapter 4, modern machine learning methods for data-driven analysis are briefly reviewed. They form the basis on which all the new approaches, developed in the thesis, rely on. The reliable ICA approach is introduced in Chapter 5. Networked brain activity, and the extended framework for analyzing activation networks, are described in Chapter 6. Then, the emergence of independent subspaces is introduced in Chapter 7, together with the extensions making the reliable ICA method usable with subspaces. The thesis concludes with discussions in Chapter 8.

The presentation order of the publications reflects the research themes as follows. The first five publications developed the method for reliable analysis of fMRI data using ICA. Publication **I** is a feasibility study testing the consistency of independent components in a multiple run ICA approach. In Publication **II** the multiple run approach is further developed, and it is shown that the approach can be used to gain insight into the data that is not attainable with single run approaches. In Publication **III** the motivation and development of the approach is discussed in detail. The approach is also applied to multi-subject fMRI data. The method is discussed in Section 5.3. Publication **IV** introduces the Arabica toolbox for reliable ICA of fMRI data. The toolbox is discussed in Section 5.4. Finally, Publication **V** proves the asymptotic convergence of the Arabica method and thoroughly discusses the theory behind the method. This is discussed in Section 5.5.

The next two publications introduced a new two-step framework for analyzing activation networks. The framework is particularly useful for studies using natural stimulation. Publication **VI** introduces the extension and is discussed in Section 6.3. In Publication **VII** the framework is improved and studied with better controlled measurements, which is discussed in Section 6.4.

The last publication shows the importance of taking the emergence of independent subspaces into account, to be able to correctly interpret the components. Publication **VIII** develops extensions for the Arabica approach to cover the cases with independent subspaces. They are discussed in Section 7.2.

2. Functional Brain Imaging

Understanding how the human brain functions has always been one of the ultimate goals in science. The human brain is one of the most central topics in many research fields, including biomedicine, psychology, and information theory. Current knowledge of brain structure and function is substantial, and growing fast, due to modern imaging and analysis techniques. In this chapter, the key concepts needed to understand the thesis are briefly explained. The introduction is based on a great overview on the human brain by Kalat [2008].

In machine learning, the growing knowledge of the brain can also lead to new theories in, *e.g.*, signal processing, neural computation, pattern recognition, machine vision, and artificial intelligence. The theoretical models can, in turn, be used to describe or predict observed behavior in the real brain. The mutual benefits have led to the fusion of neuroscience and information technology into a rapidly growing research field called neuroinformatics.

The difficulty in understanding the brain has added to the excitement of the research, but brain research also has a huge economical and societal impact. In fact, the current WHO studies show that the cost of nervous system disorders far out-ranks other diseases. Following the successful footsteps of the Human Genome Project, the proposals of the OECD Mega Science Forum and Global Science Forum have recently led to the establishment of the International Neuroinformatics Coordinating Facility (INCF, `http://www.incf.org`).

2.1 Background

The principle of *localization of function* emerged at the beginning of the nineteenth century [Kalat, 2008], and it is still the basis for most neu-

roimaging studies. The idea is simply that distinct regions of the brain support particular mental processes. The early studies were crude and invasive, often leading to incorrect findings. Modern imaging methods are typically noninvasive and allow for detailed analysis of brain anatomy and function, including tracking changes during the lifetime of a subject, *e.g.*, studying the effects of aging, or the progress of an illness.

The basic anatomical structure of the human brain is depicted in Figure 2.1. The central nervous system (CNS) is formed by the brain and the spinal cord. The brain can be divided into neocortex, cerebellum, brain stem, and other subcortical areas. The brain stem and subcortical regions are mainly involved in lower level functions and signal processing. Higher level functions, such as conscious thought, are performed on the cortex, *i.e.*, the surface, of the brain. However, the higher level functions rely on the functions of the subcortical areas.

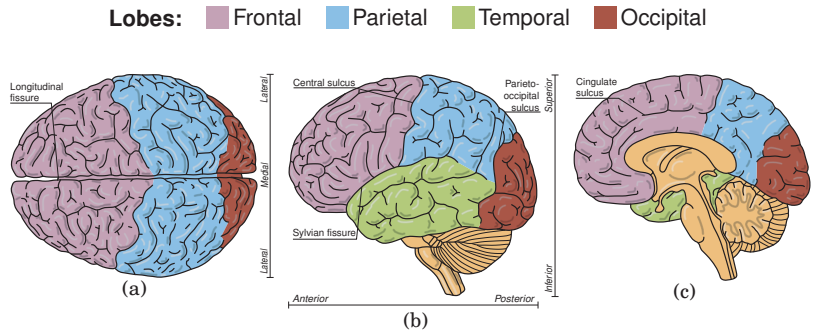


Figure 2.1. Anatomical structure of the human brain. (a) The horizontal view from above and the sagittal views from (b) the side and (c) middle of the brain show the basic structures and divisions, including the four lobes separated by sulci and fissures. Important names and directions are also shown.

The cortex of the brain essentially comprises of two kinds of tissues, the gray matter and the white matter. The gray matter contains the actual cell bodies of the neurons and most of it is concentrated on the surface of the brain. The connecting fibers between the neurons form the white matter. The inside of the brain is mostly white matter, but there are some *nuclei*, which are groups of neurons along signaling pathways. The surface is heavily folded to increase its area while keeping the volume of the brain fixed. The folds are called *sulci* and separate the surface into small sections, or *gyri*. Bigger folds that separate larger parts are sometimes called *fissures*.

The division of the cortex into the four lobes, as shown in Figure 2.1(b), is somewhat arbitrary. It is based on major sulci and fissures, visible on the

surface. Additionally, fine details, like the density of neurons and their size and shape, differ between the areas. Naturally, the boundaries are not so clear in a real brain, and can change slightly from one individual to another.

Functional brain imaging often deals with the surface of the brain, as the gray matter is mostly there. But the connections are also very important. The neuronal configuration is similar throughout the surface, but different inputs and outputs of the peripheral nervous system are connected to different areas of the brain. Thus, the different areas are specialized in processing different kind of information. Figure 2.2(a) shows the location of some of the well-known primary processing areas on the cortex. These areas are mainly connected *contralaterally*, meaning that areas on the left hemisphere are mainly responsible for signals from the right side of the body. The primary areas are then connected to additional areas nearby on the same hemisphere, or *ipsilaterally*. The additional areas usually perform more complex functions based on the processing done in the primary areas. The left and right hemispheres of the brain are functionally quite symmetric, but some tasks have a more dominant side. The brain is also very adaptive and, *e.g.* after an injury, nearby areas can take over some lost functionality.

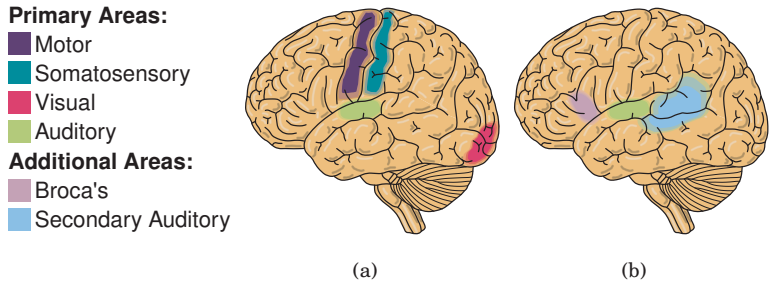


Figure 2.2. Functional areas of the human brain. (a) Primary areas involved in processing different sensory information and (b) areas involved in auditory and speech signal processing.

2.2 Magnetic Resonance Imaging

Magnetic resonance imaging uses strong magnetic fields to create images of tissues. The MRI scanner creates a *static magnetic field*, typically with field strengths of 3 Tesla or higher. For comparison, Earth's magnetic field is around $0.00005\,\mathrm{T}$. The scanner uses a *pulse sequence* of changing

magnetic gradients and oscillating electromagnetic fields. With a suitably tuned frequency of the electromagnetic field, energy is absorbed by atomic nuclei. Depending on the types of nuclei present, some amount of the energy is later emitted, and can be measured by the scanner. Figure 2.3 shows a typical MRI scanner. Different pulse sequences allow for the scanner to detect different tissue properties and distinguish between tissue types, making MRI a flexible and powerful tool. Figure 2.4 shows examples of MR-images. They are scans of a human head, using a setup that produces good contrast between different tissue types, thus revealing the anatomical structure in great detail. Since MRI is virtually *noninvasive* and is able to produce high quality images, it has quickly become very popular in structural imaging.

The complex physics of nuclear magnetic phenomena are very interesting, but certainly beyond the scope of this chapter. Therefore, only an overview of the main concepts is given in the following sections, for the purpose of understanding the signal generation in the MRI scanner. The overview is based on a more thorough introduction by Huettel et al. [2008].

Figure 2.3. An MRI scanner at the Advanced Magnetic Imaging Centre (AMI-Centre) of the Aalto University. The subject lies down on the table in front of the scanner, placing his or her head inside the volume coil. The table is then moved into the bore of the scanner, so that the head is positioned at the very center.

Due to the high water content, hydrogen atoms are by far the most abundant type of atoms in the human body. Therefore, hydrogen nuclei are the

most commonly targeted nuclei in MRI. Hydrogen nuclei have two key properties under normal conditions. Thermal energy causes the nucleus, consisting of a single proton, to spin about its axis. Because the proton carries a positive charge, the spinning motion generates an electrical current on its surface, that creates a small magnetic source. When placed within a magnetic field, the small magnetic source also generates torque. The strength of the maximum torque a magnetic source can generate is called the *magnetic moment*. The mass of the spinning proton also results in an *angular momentum*. A nucleus with these properties is often referred to as a *spin*, and only such nuclei can be studied using magnetic resonance. Different types of atoms have distinct numbers of protons and neutrons in their nuclei. If a nucleus has an even number of protons, the magnetic moment can be cancelled out by distributing the same amount of charges in opposite directions, and if a nucleus has an even atomic mass, the angular momentum can be cancelled by evenly distributed spin directions.

The strong static magnetic field is necessary for MRI, since it causes the atomic nuclei to align with the magnetic field lines. Two critical properties of the magnetic field are *field homogeneity* and *field strength*. The field needs to be uniform both in space and time so that the measured signals do not change unexpectedly, depending on when the imaging is done, or how the body is positioned in the field. Modern MRI scanners use liquid helium cooled superconducting electromagnets to create the strong and stable magnetic fields. Since maintaining the field using superconductivity requires little electricity, the static fields are always active.

The process is called magnetic resonance imaging, because the signal is actually produced by using *radiofrequency coils* that generate and receive electromagnetic fields at the resonant frequencies of the atomic nuclei within the static magnetic field. At typical field strengths, most atomic nuclei of interest for MRI happen to have resonant frequencies in the radiofrequency portion of the electromagnetic spectrum. Unlike the static magnetic field, the radiofrequency fields are turned on, according to the pulse sequence, only during small portions of the image acquisition process. After a body is placed into the strong magnetic field, an equilibrium state is reached in which the spins become aligned with the magnetic field. Due to the magnetic moments of the spins, they will actually precess around an axis parallel or antiparallel to the magnetic field. The radiofrequency waves that resonate at a suitable frequency perturb the

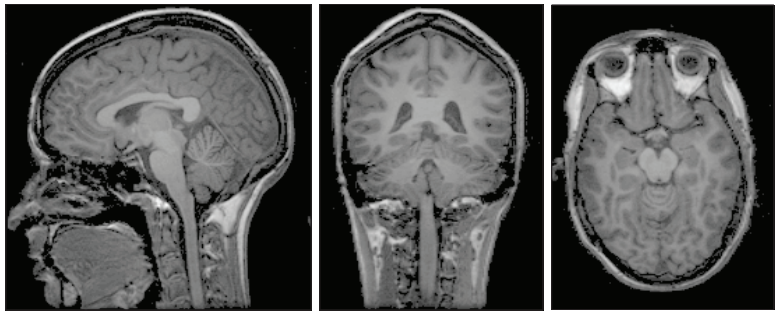
equilibrium state. During this *excitation*, the atomic nuclei absorb the energy of the radiofrequency pulse. When the pulse ends, the atomic nuclei return to the equilibrium state, releasing the previously absorbed energy. The release of energy during *relaxation* can be detected by the radiofrequency coils as the raw MR-signal. The amount of energy that can be transmitted or received by a radiofrequency coil depends on its distance from the tissue being measured. Therefore, the radiofrequency coils are typically placed immediately around the head, as *volume coils*.

The raw MR-signal does not contain any spatial information. To create an image, *gradient coils* are used to generate magnetic gradients superimposed on the strong static magnetic field. The gradient magnetic fields cause the MR-signal to depend on the spatial location in a controlled fashion. Like the radiofrequency coils, the gradient coils are only used briefly during the image acquisition. Ideally, the main magnetic field would be perfectly homogeneous and the gradients would be perfectly linear. This is hardly the case in reality, and modern scanners use additional *shimming coils* that generate compensatory magnetic fields that correct for the inhomogeneities. Unlike the other magnetic fields, the shimming coils are typically adjusted for each subject, since each person's head distorts the magnetic field slightly differently. The shimming coils are also left on for the whole duration of the imaging session. The magnitudes of the magnetic fields generated by the gradient and shimming coils are orders of magnitude smaller than that of the static magnetic field.

Most MRI scanners construct three-dimensional images from sets of two-dimensional slices. Restricting the MR-signal to one two-dimensional slice at a time is called *slice selection*, and the key element is to ensure that there is a match only between the precession frequency of the spins within the targeted slice and the radiofrequency pulse. All excited spins within a selected slice contribute to the MR-signal, but a pattern of multiple excitations is used with different combinations of weak gradients, applied immediately after each excitation. The goal is a *frequency encoding*, where the frequencies of the spins vary along one direction, and a *phase encoding*, where the phases of the spins vary along the other direction of the two-dimensional slice. The image of the slice is then recovered from the recorded MR-signals using the inverse Fourier transform. Depending on the pulse sequence, the resulting image depicts the spatial distribution of some property of the atomic nuclei within the sample. Typical properties are spin density, spin mobility, and relaxation times of the tissues in

which the spins reside.

The scanning speed is determined by the amount of time it takes for the spin relaxation processes and adjusting of the magnetic fields during the pulse sequence. Producing high resolution images, such as the ones in Figure 2.4, can take several minutes. Naturally, the quality of the images is strongly affected by the inhomogeneities in the magnetic fields, internal magnetic interactions within the tissues, and electromagnetic interference from the environment.



(a) (b) (c)

Figure 2.4. Examples of structural magnetic resonance images. The images show slices of a human head viewed from (a) sagittal, (b) frontal and (c) horizontal directions.

2.3 Functional Magnetic Resonance Imaging

Much can be learned about the brain from studying its structure. However, structural studies cannot reveal short-term physiological changes associated with the active functioning of the brain. Functional neuroimaging studies aim to identify the different parts of the brain where particular mental processes occur, and to characterize the associated patterns of brain activation. Typically, *functional neuroimaging* studies build maps that link brain activation to mental function.

The idea in functional MRI is to record a sequence of MR images at different time points. Figure 2.5 shows an example of an image sequence during an onset of neuronal activity. The images produced in fMRI are not a direct measure of neuronal activity. Instead, fMRI creates images of physiological changes that are *correlated* with neuronal activity. The measure is based on the differing magnetic properties of oxygenated and deoxygenated hemoglobin molecules, as further explained in Section 2.3.1.

The measured signal change is small compared with the total image intensity, and more importantly, the change related to neuronal activity is very small compared with other sources of spatial and temporal variability across and within images. Careful analysis of the whole sequence is required to detect the activation patterns.

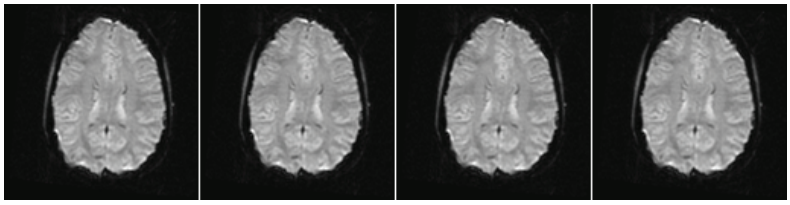


Figure 2.5. Examples of functional magnetic resonance images. A sequence of scanned slices without processing. The first two images are scanned during rest and the last two during a task. Note that the images do not show a direct measure of neuronal activity.

Additional problems arise from the long duration of the scanning. Movement of the head and other physiological changes during the measurements will distort the images. Therefore, the setup used in fMRI is a careful compromise between fast scanning and high resolution images. Current fMRI scanners are able to produce full head volumes with a time interval of a few seconds, but the spatial resolution is much smaller than that used in structural imaging. The spatial resolution of fMRI determines the ability to separate adjacent brain regions with different functional behavior. Smaller voxel size leads to reduced signal-to-noise ratio and the organization of the vascular system introduces additional spatial constraints. The temporal resolution refers to the ability to estimate the timing of neuronal activity. Using a short repetition time can improve temporal resolution and increase statistical power, but the slowness and variability of the vascular response limits the ability to make precise temporal measurements.

2.3.1 Blood Oxygenation Level Dependent Signal

As the information processing activity of neurons increases their metabolic requirements, the vascular system provides energy in the form of glucose and oxygen to meet these demands. Since the oxygen in the blood is bound to hemoglobin molecules, the increased oxygen consumption leads to a higher concentration of deoxygenated hemoglobin. The *hemodynamic* changes are measurable with blood oxygenation level dependent (BOLD) contrast [Ogawa et al., 1992], visible in certain types of MR images. BOLD

is actually based on changes in the magnetic properties of water molecules, which reflect the changes in the concentration of paramagnetic deoxyhemoglobin. Thus, fMRI is a very indirect measure of neuronal activity.

The changes in BOLD after a brief neuronal activity, *i.e.*, *hemodynamic response*, consists of a short onset delay, a rise to a peak after a few seconds, a return to baseline, and a prolonged undershoot. Sustained neuronal activity results in an additional plateau following the onset peak. Figure 2.6 shows an illustration of a typical hemodynamic response model. Early studies [*e.g.*, Mandeville et al., 1999, Logothetis et al., 2001] showed that the BOLD response is actually shaped by local changes in blood flow and blood volume, not the extraction of oxygen by active neurons. Oxygen-rich blood from the lungs is pumped through the heart to the aorta, leading to several large arteries. Each artery branches into smaller arteries and to even smaller arterioles that eventually terminate in *capillaries*. The extraction of oxygen and glucose from the blood, and the removal of waste carbon dioxide, occurs at the surface of the capillaries, which are thin-walled vessels 5 to $10\,\mu$m in diameter. The red blood cells, with a width of about $7.5\,\mu$m, actually need to deform as they move through the narrowest capillaries.

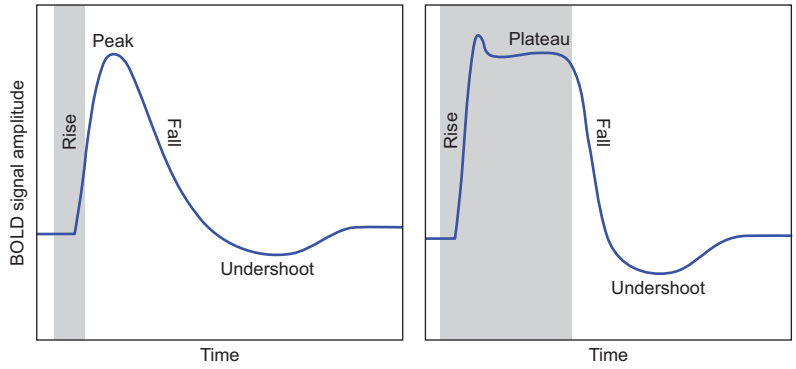


Figure 2.6. Illustration of BOLD signal responses to neuronal activity. The shaded area depicts the time of neuronal activity. On the left, response to a short transient activation, and on the right, response to a longer sustained activation.

The increased metabolic demands caused by neuronal activation results in an increased inflow of oxygenated blood, after a short latency. The vascular system overcompensates by supplying more oxygen to the area than is extracted, resulting in a decrease of deoxyhemoglobin concentration. This is visible as the sharp peak in the BOLD response after onset. According to most models, such as the *balloon model* [Buxton and Frank, 1997], the initially greater blood inflow than outflow leads to the

expansion of the capillary veins, and an increased blood volume. After the neuronal activation returns to baseline, the blood flow also decreases rapidly. However, the blood volume remains elevated for an extended period, and the relatively low blood flow compared to the blood volume results in higher deoxyhemoglobin concentration. This would explain the undershoot in the BOLD response. Recent studies [*e.g.*, Harshbarger and Song, 2008] also support models, where continuously elevated regional metabolism can explain the undershoot.

Early experiments have shown that the hemodynamic responses behave roughly linearly [*c.f.*, Boynton et al., 1996, Dale and Buckner, 1997]. However, the hemodynamic responses can be nonlinear at intervals of less than 6 s, and taking the nonlinearities into account in analysis models can dramatically improve their sensitivity [Wager et al., 2005]. Small nonlinearities during this *refractory period* can reduce the power of experimental analyses, but large nonlinearities could preclude the study of short intervals.

2.3.2 Experimental Settings and Stimulation

Traditionally, any experimental scientific study starts with a carefully chosen research hypothesis. The aim of the experimental measurements is to provide evidence for or against the hypothesis. In neuroscience, the research questions can vary from very generic to focused details of known brain processes. More specific hypotheses can be falsified more easily and are considered more informative. The way in which the measurements are set up to test the hypothesis is referred to as the *experimental design*.

A well-designed experiment has one or many *independent variables* that are intentionally manipulated during the measurements using different *experimental conditions*. The different conditions can occur between subjects, such that, different groups of subjects correspond to different values of the variables. This is typical in, *e.g.*, studies comparing patients with control subjects. More commonly though, the conditions are varied within subjects using a planned sequence of stimulation. During the experiment, one or many *dependent variables* are measured, and the values should reflect the effects of the independent variables. Typically, BOLD signal change is the primary dependent variable.

Usually, fMRI studies use a controlled stimulus, like visual patterns or audible sounds, designed to test a specific hypothesis. The most common way of building the stimulus is to use a *blocked design* that alternates

between two conditions in blocks. Although each block contains multiple individual stimuli, it is assumed that the cognitive processes of interest are constant during a block. Typically, one of the conditions is a *control condition* during which the subject is not engaged in task performance, *e.g.* watching a blank screen. There can also be more than two types of blocks in a single study. The timing of the blocks and number of repetitions must be planned carefully.

Another popular type of experimental design is the *event-related design*, where the stimuli comprise of discrete short-duration events. To analyze event-related measurements, it is often required that signals from many repeated events are averaged to improve the signal-to-noise ratio. Non-linearities in the hemodynamic response reduce the amplitudes of rapidly recurring events and present an additional challenge for most studies. In addition, the refractory effects can also be used to study functional adaptation within a brain region.

The challenge in designing the experimental conditions is to minimize multiple explanations for the measured effects. Any property that co-varies with the independent variables is known as a *confounding factor*. A well-designed experiment has enough conditions to distinguish the confounding factors from the independent variables, but this is not always possible and additional experiments are typically needed to determine the true cause of an observed effect. Two systematic approaches are often used to minimize confounding factors. If the order of experimental conditions can be chosen arbitrarily, *randomization* of the order of repeating conditions can make the factors uncorrelated with the independent variables. When the factors cannot be made completely random, *counterbalancing* the order of conditions over several trials can be used to ensure that the factors would be equally present during all conditions. In either case, the goal is to make confounding factors influence all conditions similarly.

3. Preprocessing and Statistical Analysis

The signal-to-noise ratio in fMRI is very low and depends on the field strength of the scanner. A higher field strength does improve the signal-to-noise ratio, but also makes the measurements susceptible to contamination by other types of noise. In addition to the intrinsic *thermal noise* within both the subject and the scanner, there are also strong sources of system noise. One particularly important type is *scanner drift*, which among other causes is produced by subtle changes in the strength of the static field that slowly alters the resonant frequency of hydrogen protons. There are many sources of signal fluctuations in the imaging hardware, and naturally, the scanner room must be shielded from any extraneous radiofrequency signals.

Even more important types of noise are *motion artifacts* and *physiological noise*, especially under high field strength. Subject motion, such as moving the head or swallowing, is very common and disruptive for fMRI studies. In the worst case, it can render the whole recording useless. There is also oscillatory muscle activity due to breathing and heartbeat, while inside the brain, blood pulses through arteries and veins. Many physiological factors cause changes in blood flow, blood volume and oxygen metabolism. The signal effects are usually not produced by motion during image acquisition, but rather motion causes variability across the time-series of images. Moreover, movement is often correlated with the experimental task and motion also introduces both spatial and temporal correlations in the measured images.

For the hypothesis testing, any task-unrelated neural activity can also be considered as noise. The task-related responses occur within an active brain, where other neural processes are altering the BOLD contrast at every moment. On top of everything, the sampling rate is not fast enough to fully resolve all the noise effects, leading to *aliasing* that causes arti-

factual low frequency signals to appear. Thus, the detection and analysis of interesting brain processes is very difficult.

3.1 Typical Preprocessing

Essentially, any statistical analysis of the fMRI measurements assumes that each voxel represents a unique and fixed location in the brain, sampled at regular intervals. Considering the scanning process and the noise sources described above, these assumptions seem invalid. Therefore, before analysis the fMRI data is almost always processed using methods that try to overcome these problems [Worsley and Friston, 1995]. The aim of preprocessing is to improve the functional resolution of an experiment by removing uninteresting variability from the data and preparing it for further statistical analysis.

3.1.1 Slice Time Correction

The fMRI scanner acquires images one slice at a time and covers the full imaging volume by collecting several adjacent slices in rapid succession. Most fMRI studies use interleaved slice acquisition, where the scanner first collects every other slice and then the remaining interleaved slices in between. This helps to avoid cross-slice excitation during the imaging. However, this poses a problem since adjacent parts of the brain are acquired at non-adjacent time points. The typical approach for correcting the timing discrepancies is to temporally interpolate the values in every slice to a global offset. This leads to a loss of some information, especially in event-related experimental designs. Sometimes analysis is performed for each slice separately to avoid slice time correction, but such approaches have their own limitations.

3.1.2 Motion Correction

Head motion has a drastic effect on the data. Even very small movements, the size of a single voxel, can cause signals from very different tissue types to mix up. Particularly, around the borders of tissues, motion leads to spurious activations that form a pattern around the edges of the tissues. It is typical to observe a distinctive ring pattern around the edges of the brain. Moreover, motion effects that co-occur with stimulus presentation are difficult to separate from real brain activation. Even though motion is

mainly a spatial problem, movement that is perpendicular to the imaging slices can effect the timing of the activation.

A carefully executed study will aim to prevent head motion, *e.g.*, by using head restraints or by training of the subjects. Nevertheless, the chances of acquiring a complete dataset that is not corrupted by motion are very small. The goal of motion correction is to adjust the images so that the brain is always in the same position. Generally, the process of aligning two image volumes is called *coregistration*. In the case of motion correction, the time series of image volumes are coregistered to a single reference volume.

3.1.3 Coregistration

Since the size and shape of the brain does not change during scanning, it is often assumed that a *rigid-body transformation* is sufficient, meaning that two images can be superimposed exactly upon each other by translations and rotations. In practice, inhomogeneities in the magnetic field can cause distortions in the images. There are several methods for estimating the optimal translation and rotation parameters, *e.g.*, based on minimizing the *mutual information* between volumes. Once the realignment parameters are determined, the original image is resampled using spatial interpolation to estimate the values without head motion. In some studies, the motion parameters are used in the analysis model to isolate motion effects.

3.1.4 Functional-structural Coregistration

In most studies, it is important to understand how the activation corresponds to the underlying neuroanatomy. Unfortunately, functional images have lower resolution and very small anatomical contrast compared to structural images. Because of these limitations, also high resolution structural images are typically acquired. Using similar coregistration approaches as above, the functional images can then be mapped to the structural ones. However, additional parameters are needed to account for scaling differences between the images.

3.1.5 Spatial Normalization

Two individuals in the same fMRI experiment may differ in overall brain size by more than $30\,\%$. There is substantial variation in the shape of the

brain, and the organization of sulci and gyri can vary so much that even major landmarks have different positions and orientations in different individuals. This remarkably variable morphology makes combining and comparing data from different individuals or studies difficult.

Normalization is like coregistration, except that the compared images are assumed to originally differ in shape rather than as a result of image distortion. The goal is to compensate for the shape differences by stretching, squeezing, and warping the images of each brain so that they correspond to those of every other brain. The most widely used normalized space is Talairach space [Talairach and Tournoux, 1988], which was derived from landmark measurements on a single brain of an elderly woman. However, since the single brain used was unrepresentative of the population at large, more recent works have created probabilistic templates based on hundreds of individuals. A commonly used space is MNI space [Chau and McIntosh, 2005] that has been scaled to match the landmarks within the well-established Talairach space. Most automated normalization algorithms are based on such probabilistic templates.

Moreover, nearly all normalization templates are based on samples from a population of young and healthy adults. Other groups, such as elderly individuals, young children, and patients with brain lesions, systematically differ from the template population. By definition, normalization emphasizes that which is common among individuals and de-emphasizes that which is unique. Small, but meaningful, variations may be lost in the process. An alternative is to use individual region-of-interest based analyses.

3.1.6 Spatial Smoothing

In most fMRI analyses, a low-pass spatial filter, typically a Gaussian filter, is used to reduce high-frequency components and smooth the images. Spatial smoothing improves signal-to-noise ratio by preserving signals of interest, since there are functional similarities between adjacent brain regions and blurring caused by the vascular system, while removing uninteresting variation in the data due to noise sources. Moreover, when using standard statistical approaches, reducing the dimensionality of the data by smoothing can reduce the problem of many false-positives with multiple statistical comparisons. Additionally, the smoothing can compensate for small errors made in the coregistration and normalization steps.

3.1.7 Temporal Filtering

In addition to all the spatial preprocessing, also the time series of the data can be processed. Temporal filters are typically constructed to reduce the influence of physiological oscillations, such as breathing and heart beating. High-pass filtering can also remove slow drift-like trends. When the experimental setup allows it, filters can also be designed to retain important frequency ranges based on repeating stimuli. Using temporal filters can substantially improve functional signal-to-noise ratio. However, temporal filtering is limited by the slow sampling rate.

3.2 Hypothesis Testing

The standard way of analyzing an fMRI sequence is to perform statistical significance testing, based on a hypothesis model. Essentially, the analysis reveals the areas of the brain that most probably fit a given hypothesis. There are many different hypothesis-driven approaches that differ in their model assumptions. However, they all express significance as the probability that the results could occur by chance, meaning that a very low value indicates a reliable finding. More importantly, all hypothesis-driven methods assume that each time point can be assigned to a particular experimental condition, or that the whole time series of the hemodynamic response can be predicted. Furthermore, when using such methods, it can be challenging to appropriately control for the rates of false positive and false negative findings.

The most commonly used approach is statistical parametric mapping (SPM, `http://www.fil.ion.ucl.ac.uk/spm`), which is a regression analysis based on the general linear model (GLM, *c.f.*, Friston et al. [2006]). The assumption in GLM is that the data are composed of the linear combinations of several model factors, along with uncorrelated noise. The model factors are simply given as multiple regressor time series. The following section describes GLM in more detail.

There are many technical problems with such analysis, but perhaps most importantly, the stimulation setup has to be simple enough to allow for predicting the responses, and forming the reference time series, in the first place. Therefore, detecting previously unknown phenomena is extremely difficult, and an expensive dataset may only be suitable for testing a single hypothesis.

3.2.1 General Linear Model

The general linear model (GLM, *c.f.*, Worsley and Friston [1995]) is a statistical multiple regression model. Based on the timing and duration of events in the stimuli, researchers generate a predicted hemodynamic response. These model factors contain the predicted time series for the entire measurement session. The relative contribution of each of these regressors to each voxel within the data is then statistically evaluated. Given the data and a chosen set of regressors the model is, in matrix form:

$$\boldsymbol{X} = \boldsymbol{G}\boldsymbol{\beta} + \boldsymbol{\epsilon} \, . \tag{3.1}$$

The fMRI data are represented as a two-dimensional data matrix $\boldsymbol{X}$ consisting of N time points by V voxels. The design matrix $\boldsymbol{G}$, of size $N \times M$, contains the M regressors as columns. The parameter matrix $\boldsymbol{\beta}$ with dimensions $M \times V$ contains the parameter weights of each regressor for every voxel. Finally, the error matrix $\boldsymbol{\epsilon}$ has the same size as $\boldsymbol{X}$.

The goal of regression is to find the parameter weights that minimize the error term, typically in the least-squares sense. Note that the spatial structure of the fMRI data is lost as all voxels in the imaging volume are rearranged as one row in the matrix. The parameter values and error term are calculated independently for each voxel.

The regressors represent hypothesized factors that may or may not contribute to the data. To evaluate the statistical significance of a regressor, the amount of variability it explains is compared with the amount of variability explained by the error term. All the variability in the data that cannot be explained with the set of chosen regressors is considered as additive noise by the model. Therefore, in some studies, it is crucial to add so called nuisance regressors that model known artifacts or uninteresting signals. Adding regressors makes the statistical testing more conservative, but can improve the validity of the model.

4. Analysis Based on Dependence and Independence

The previously discussed analysis methods assume the ability to reliably estimate the different factors affecting the fMRI time series in advance. Arguably, under many conditions and for many brain areas, such predictions are very difficult, if not impossible. Even in the simplest experiment, there could be changes in strategy by the subject, changes in task performance associated with learning, habituation, fatigue, or with other processes whose temporal behavior cannot be predicted in advance.

The hypothesis-driven methods also have major technical drawbacks. Typically, voxels in which the signal exceeds a predefined level of significance are identified as active. However, even in areas of activation, the task-related hemodynamic signal changes are typically very small, suggesting that other time varying phenomena must produce the bulk of the measured signals. Typical methods also rely on univariate techniques that ignore relationships between voxels, hindering the detection of brain regions acting as functional units during the experiment.

4.1 Motivation

The recordings for each voxel may contain a complicated mixture of activity, which is produced by many spatially distributed processes, including task-related and task-unrelated brain activations, as well as motion or scanner artifacts. Each process can be described as a component having a separate time series and spatial extent, and producing simultaneous changes in the fMRI signals of many voxels. The component processes may or may not be task-related, or they could be only transiently task-related.

This chapter briefly introduces three data-driven analysis methods that make minimal assumptions about the data. They are based on identify-

ing uncorrelated, mutually dependent, or statistically independent signal components, respectively.

4.2 Principal Component Analysis

Principal component analysis (PCA, *c.f.*, Bishop [2007]) is a classic technique in statistical data analysis, feature extraction, and data compression. In some literature, the method is also called the Karhunen-Loève transform or the Hotelling transform. It has two equivalent formulations by Pearson [1901], which defines it as the projection that minimizes the mean squared distance between data points and their projections, and by Hotelling [1933], which is based on orthogonal projections that maximize the variance of the projected data. The following uses the maximum variance perspective in matrix form, since it is easier to relate to the other methods.

As before, assume a data matrix $\boldsymbol{X}$ consisting of V samples of N dimensional random vectors $\boldsymbol{x}$ as columns, such as, the recorded time series of each voxel. Without any loss of generality, the vectors $\boldsymbol{x}$ can be assumed to have zero mean $\mathbb{E}[\boldsymbol{x}] = 0$, or to have been centered as a preprocessing step, to simplify the formulas. The goal is to find a projection vector $\boldsymbol{w}$ that maximizes the variance of the projected scalar variable z,

$$z = \boldsymbol{w}^\top \boldsymbol{x} \ , \tag{4.1}$$

where the norm of the projection vector is constrained to be $\|\boldsymbol{w}\| = 1$, so that the variance cannot be maximized by simply increasing the length of the projection vector. By definition, the variance of z is,

$$\mathrm{var}[z] = \mathbb{E}[(z - \mu_z)^2] = \mathbb{E}[\boldsymbol{w}^\top \boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{w}] = \boldsymbol{w}^\top \boldsymbol{C}\boldsymbol{w} \ , \tag{4.2}$$

where $\mu_z = 0$ is the mean of z and the matrix $\boldsymbol{C}$ is the estimated covariance matrix $\boldsymbol{C} = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top] \approx \frac{1}{V}\boldsymbol{X}\boldsymbol{X}^\top$. Maximizing Equation (4.2) with the constraint $\|\boldsymbol{w}\| = 1$, using a Lagrange multiplier λ, leads to

$$\boldsymbol{C}\boldsymbol{w} = \lambda \boldsymbol{w} \ , \tag{4.3}$$

which is the eigenvalue equation for matrix $\boldsymbol{C}$. It is also easy to see from Equations (4.2) and (4.3) that the variance of the projection is equal to the eigenvalue. Therefore, the largest variance is attained when the largest eigenvalue and its corresponding eigenvector are chosen as the principal component. Since the PCA problem reduces to solving an eigenvalue

equation, it is possible to look for the components either all at once or one at a time, by deflation.

Using this formulation, the data variance can be fully described with a decomposition of N orthogonal principal components by using the eigenvalue decomposition of the covariance matrix $\boldsymbol{C}$,

$$\boldsymbol{C}\boldsymbol{W} = \boldsymbol{W}\boldsymbol{\Lambda} \,, \tag{4.4}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues in descending order, and $\boldsymbol{W}$ is a matrix with the corresponding orthonormal eigenvectors as columns. The data can be projected onto the principal components using the projection matrix $\boldsymbol{W}$ without loss of information, in matrix form applied to the whole data matrix, as

$$\boldsymbol{Z} = \boldsymbol{W}^\top \boldsymbol{X} \,. \tag{4.5}$$

The projected decomposition $\boldsymbol{Z}$ found with PCA is the set of uncorrelated components that contain most of the variance and can reconstruct the whole data with minimal residual error. The reconstruction model is actually analogous to the one in GLM, where the matrix $\boldsymbol{W}$ is used instead of a predefined design matrix.

The properties of PCA make it widely used in dimensionality reduction, since to reduce the dimensionality of data $\boldsymbol{X}$ to K, while preserving as much of the variance as possible, the projection matrix $\boldsymbol{W}$ is simply built using only the eigenvectors corresponding to the K largest eigenvalues. Usually, the number of principal components required to adequately represent the data to a specified level of accuracy is much smaller than the original dimensionality of the data.

Another popular use of PCA is as a preprocessing step for data whitening, or sphering. In whitening, the projected data is further normalized by making its variance along all directions equal to one. This is accomplished simply by dividing each principal component with its standard deviation, that is, with the square root of the corresponding eigenvalue.

Principal component analysis can be used to isolate uncorrelated activation patterns in functional imaging data [see, *e.g.* Moeller and Strother, 1991]. However, as task-related changes in fMRI are only a small part of the total signal variance, capturing the greatest variance in the data may reveal little information about task-related activations or other processes of interest.

4.3 Canonical Correlation Analysis

Another classic approach is called canonical correlation analysis (CCA, Hotelling [1936]). Whereas PCA searches for uncorrelated components within one dataset, CCA looks for dependencies between two datasets. The method projects both datasets into a common space in such a way that the dependencies between the two sets are maximized. Just as PCA is good at dimensionality reduction, CCA can be used to lower the dimensionality of the datasets, while preserving variation that is shared by both sets and discarding variation present in only one dataset.

As before, denote centered random vectors $\boldsymbol{x}_1$ from dataset $\boldsymbol{X}_1$, with dimensionality N_1. Similarly, assume another related dataset $\boldsymbol{X}_2$, such as from another trial or subject in the same study, which is represented by random vectors $\boldsymbol{x}_2$ with dimensionality N_2. The goal in CCA is to look for projection vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$, such that the Pearson correlation between the projections is maximized

$$\underset{\boldsymbol{w}_1,\boldsymbol{w}_2}{\operatorname{argmax}}\ \rho = \underset{\boldsymbol{w}_1,\boldsymbol{w}_2}{\operatorname{argmax}}\ \boldsymbol{w}_1^\top \boldsymbol{x}_1 \boldsymbol{x}_2^\top \boldsymbol{w}_2 = \underset{\boldsymbol{w}_1,\boldsymbol{w}_2}{\operatorname{argmax}}\ \boldsymbol{w}_1^\top \boldsymbol{C}_{12} \boldsymbol{w}_2 \tag{4.6}$$

with normalization constraints $\boldsymbol{w}_1^\top \boldsymbol{C}_1 \boldsymbol{w}_1 = \boldsymbol{w}_2^\top \boldsymbol{C}_2 \boldsymbol{w}_2 = 1$, where the matrices $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are the estimated covariance matrices of the individual datasets, and the matrix $\boldsymbol{C}_{12}$ is the estimated cross-covariance matrix $\boldsymbol{C}_{12} = \mathbb{E}[\boldsymbol{x}_1 \boldsymbol{x}_2^\top]$ between the datasets.

To obtain the desired formulation of CCA, the two data vectors can be concatenated into one $(N_1 + N_2)$ dimensional vector, whose covariance matrix $\boldsymbol{C}$ is

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_2 \end{pmatrix}, \tag{4.7}$$

where the blocks on the diagonal are the covariances, and the off-diagonal blocks are the cross-covariances, from above. While PCA leads to an eigenvalue problem, maximizing Equation (4.6) leads to a so-called generalized eigenvalue problem [Timm, 2002], where $\lambda = 1 + \rho$,

$$\begin{pmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \end{pmatrix} = \lambda \begin{pmatrix} \boldsymbol{C}_1 & 0 \\ 0 & \boldsymbol{C}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \end{pmatrix}. \tag{4.8}$$

When denoting the block diagonal of $\boldsymbol{C}$ by $\boldsymbol{D}$,

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{C}_1 & 0 \\ 0 & \boldsymbol{C}_2 \end{pmatrix} , \tag{4.9}$$

Equation (4.8) can be written simply as

$$\boldsymbol{C}\boldsymbol{W} = \lambda \boldsymbol{D}\boldsymbol{W} , \tag{4.10}$$

where the columns of the matrix $\boldsymbol{W}$ contain the concatenated projection vectors. There are several ways to generalize correlation to more than two random vectors [Kettenring, 1971], leading to several possibilities to generalize CCA for multiple datasets. This formulation allows for a straightforward generalization to more than two datasets [Bach and Jordan, 2002], where the vectors and matrices are simply concatenations of more than two block elements. Solving the generalized eigenvalue problem is based on the Cholesky decomposition and the symmetric QR-algorithm [Golub and van Loan, 1996].

In recent years, recording multiple related datasets has become relatively easy in many research fields, making CCA and its generalizations quite popular. In functional brain imaging, CCA has been used to search for dependencies between subjects or between stimuli and the measurements [*c.f.*, Friman et al., 2001, Calhoun et al., 2009, Hardoon et al., 2007]. Even dependencies between fMRI and electroencephalography (EEG) measurements, recorded using the same stimulation, have been studied [Mantini et al., 2007].

4.4 Independent Component Analysis

Unlike the methods above, independent component analysis (ICA, Hyvärinen et al. [2001b]) is a modern machine learning approach, based only on the assumption that the component signals are statistically independent. This seems reasonable in many applications and, in fact, does not have to hold exactly for ICA to be usable in practice. Just like before in GLM, the generative model used in ICA is an instantaneous linear mixture of random variables. Again, denoting the observed mixed signals as data matrix $\boldsymbol{X}$, the mixture model in matrix form is, neglecting the error term for simplicity,

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} . \tag{4.11}$$

The mixing matrix $\boldsymbol{A}$ holds the mixing weights and the matrix $\boldsymbol{S}$ contains the corresponding original sources. Unlike before, both $\boldsymbol{A}$ and $\boldsymbol{S}$

are unknown in ICA. Typically, the rank of matrix $\boldsymbol{A}$ is lower than the dimensionality of the data matrix $\boldsymbol{X}$, so that there are less sources than observed signals.

The problem of jointly solving both the mixing and the original sources is not only considerably difficult, but also ambiguous. Since there are no constraints on the variances of the sources, it immediately follows that the signs and scaling of the sources cannot be uniquely defined. A source signal can be multiplied with any scalar, as long as the corresponding mixing weights are divided by the same value. Additionally, the order of the sources is not fixed. Fortunately, these ambiguities in the model are not so crucial in practice. For example, the signs and scaling of the independent components can often be fixed with simple normalization, after performing ICA.

Many algorithms have been implemented for ICA, such as, FastICA [Hyvärinen and Oja, 1997], Infomax [Bell and Sejnowski, 1995, Amari et al., 1995], and JADE [Cardoso, 1989]. They all work in a clear data-driven manner, based on the assumption that the original sources are statistically independent. In recent years, ICA has become very popular in many research fields, including functional brain imaging [see, *e.g.*, Makeig et al., 1995, Olshausen and Field, 1996, Jahn et al., 1998, Vigário et al., 2000, Tang et al., 2002, Calhoun et al., 2003, McKeown et al., 2003, Calhoun et al., 2009].

Before estimating the independent components, the observed mixtures can be whitened, that is, transformed to be uncorrelated and have unit variances. This does not constrain the estimation in any way due to the inherent scale ambiguity, and since independence implies uncorrelatedness. Whitening simplifies the component estimation by restricting the structure of the mixing, and is a required step in many ICA algorithms.

Additionally, if the goal is to find K independent components and the whitening is done using PCA, the complexity of the problem can be lowered by reducing the data by retaining only the K strongest principal components. Assuming that the weakest principal components contain mainly noise, the dimensionality of the data is reduced in an optimal manner to improve the signal-to-noise ratio. However, when this does not hold for the data, as in the case of fMRI, the dimensionality must not be reduced too much to remove interesting components.

4.4.1 Blind Source Separation

The idea of solving the underlying source signals in Equation (4.11) using only the observed signals with unknown mixing and minimal, if any, information on the sources is called blind source separation (BSS, *c.f.*, Cardoso [1990], Jutten and Herault [1991]). To make the solution for both the mixing and sources at the same time identifiable, all approaches need to make some constraining assumptions, either on the mixing process or the source signals.

For example, when assuming that the sources contain significant autocorrelations, the problem can be solved by using temporal decorrelation algorithms, such as, SOBI [Belouchrani et al., 1993] or TDSEP [Ziehe and Müller, 1998]. The assumptions in ICA [Jutten and Herault, 1991, Comon, 1994] are the most widely used approach for solving the BSS problem. The solutions have many useful applications, such as, identifying signals of interest, removing artifacts, and reducing noise.

4.4.2 Independence and Non-Gaussianity

Theoretically, statistical independence means that the sources do not contain any information on each other. In other words, the joint probability density function of the sources is factorisable to its marginal probability densities $p(s_1, ..., s_K) = \prod_i p(s_i)$. Since a direct optimization of independence would require exact determination of the density functions, which are generally not available, the sources have to be estimated by approximating independence with other objective functions. These can be based on statistical concepts such as mutual information and negentropy [Hyvärinen et al., 2001b].

Essentially, all objective functions measure how non-Gaussian the estimated sources are. An intuitive explanation for this is offered by the central limit theorem, which states that the distribution of a mixture of independent and identically distributed (*i.i.d.*) random variables tends to be Gaussian, regardless of what the original distributions are. Therefore, as the sources are made more non-Gaussian they must also become more unmixed. Commonly used measures of non-Gaussianity are skewness and kurtosis, or the third and fourth order cumulants.

Negentropy is a basic concept in information theory, measuring how much smaller the differential entropy of a random variable is when compared to the entropy of a Gaussian variable. Since the entropy of a Gaus-

sian variable is the largest among all random variables of equal variance, negentropy is a very natural measure of non-Gaussianity. Negentropy is defined, using differential entropy H, as

$$J(y) = H(\tilde{y}) - H(y) \ , \tag{4.12}$$

where $\tilde{y}$ is a Gaussian random variable having the same covariance as the measured variable y. For the purpose of ICA, negentropy also has the benefit of being invariant for invertible linear transformations. Again, since using negentropy is computationally difficult, requiring a known or estimated density function, simpler approximations of negentropy are needed. The aforementioned higher order cumulants are one way of estimating negentropy, but a more general and robust estimator has the form

$$J(y) \propto (\mathbb{E}[G(y)] - \mathbb{E}[G(\tilde{y})])^2 \ , \tag{4.13}$$

where constant terms have been removed and G is any nonlinear non-quadratic function. Choosing a G that does not grow too fast leads to a robust estimator, and using $G(y) = y^4$ is identical to the kurtosis based estimation.

By definition, in the special case of a Gaussian random variable, uncorrelatedness is equal to independence. It is also easy to see from the intuitive explanation, and commonly used measures of non-Gaussianity, that theoretically ICA cannot identify Gaussian components. This is due to the fact that any attempt at unmixing Gaussian components cannot make them measurably more non-Gaussian. In practice, ICA can identify one Gaussian component among many non-Gaussian ones, and typically the interesting signals are not Gaussian.

4.4.3 The FastICA Algorithm

The remainder of this thesis focuses on the use of the FastICA algorithm (`http://research.ics.aalto.fi/ica/fastica`, Hyvärinen [1999]), though most considerations apply also to other ICA algorithms. FastICA is one of the fastest and most robust ICA algorithms, even with large datasets and under noisy conditions. FastICA uses a fixed-point optimization of negentropy, based on an approximative Newton method. FastICA also uses whitening with PCA as the first step.

Assuming the data has been whitened using the projection in Equation (4.5) and normalizing the variance along all directions equal to one, so that the random vector z denotes the samples of the whitened data,

the goal is to find an unmixing projection vector $\boldsymbol{w}$ that produces an independent component $\boldsymbol{s} = \boldsymbol{w}^\top \boldsymbol{z}$. The gradient of the approximation of negentropy in Equation (4.13) with respect to $\boldsymbol{w}$ becomes

$$\gamma \, \mathbb{E}[\boldsymbol{z} g(\boldsymbol{w}^\top \boldsymbol{z})] \, , \tag{4.14}$$

where g is the derivative of G and the scaling term is $\gamma = \mathbb{E}[G(\boldsymbol{w}^\top \boldsymbol{z})] - \mathbb{E}[G(\tilde{y})])^2$. Since z is assumed to be whitened, $\tilde{y}$ is a standardized Gaussian random variable, and the optimization can be performed under the constraint $\mathbb{E}[(\boldsymbol{w}^\top \boldsymbol{z})^2] = \|\boldsymbol{w}\| = 1$. Therefore, the whole scaling coefficient γ becomes a constant and can be removed.

The gradient in Equation (4.14) could be used directly in a fixed-point iteration. However, the convergence of the algorithm can be significantly improved with the Newton method. Typically, this would increase the computational complexity due to a matrix inversion needed in the Newton iteration. In FastICA, the matrix inversion is avoided by further approximating that matrix with its diagonal. This is a reasonable approximation, since the data is assumed to be whitened, and the resulting fixed-point update rule in FastICA is

$$\boldsymbol{w} \leftarrow \mathbb{E}\left[\boldsymbol{z} g(\boldsymbol{w}^\top \boldsymbol{z}) - \mathbb{E}[g'(\boldsymbol{w}^\top \boldsymbol{z})]\boldsymbol{w}\right] \, , \tag{4.15}$$

where g' is the derivative of g. The whole algorithm for finding one maximally non-Gaussian component is outlined in Table 4.1. As in PCA, it is possible to look for many components either all at once or one at a time, by deflation.

1. Whiten the data to get $\boldsymbol{z}$.
2. Choose an initial (*e.g.*, random) vector $\boldsymbol{w}$ of unit norm.
3. Let $\boldsymbol{w} \leftarrow \mathbb{E}\left[\boldsymbol{z} g(\boldsymbol{w}^\top \boldsymbol{z}) - \mathbb{E}[g'(\boldsymbol{w}^\top \boldsymbol{z})]\boldsymbol{w}\right]$, with a suitable g.
4. Let $\boldsymbol{w} \leftarrow \boldsymbol{w}/\|\boldsymbol{w}\|$.
5. Repeat from step 3, if not converged.

Table 4.1. The FastICA algorithm. The steps for estimating one independent component.

In practice, step 3 of the algorithm is estimated from a finite sample $\{\boldsymbol{z}_i\}_{i=1}^n$ by

$$\hat{\boldsymbol{w}}_n \leftarrow \frac{1}{n} \sum_{i=1}^{n} \left(\boldsymbol{z}_i g(\hat{\boldsymbol{w}}_n^\top \boldsymbol{z}_i) - g'(\hat{\boldsymbol{w}}_n^\top \boldsymbol{z}_i) \hat{\boldsymbol{w}}_n \right) \, . \tag{4.16}$$

4.4.4 Spatial ICA of fMRI Data

The first application of independent component analysis to fMRI data was done by McKeown et al. [1998], McKeown and Sejnowski [1998]. Since then, ICA has become a very popular tool for fMRI studies. For recent reviews, consider reading Calhoun et al. [2003], McKeown et al. [2003]. The ICA model used in fMRI analysis is illustrated in Figure 4.1.

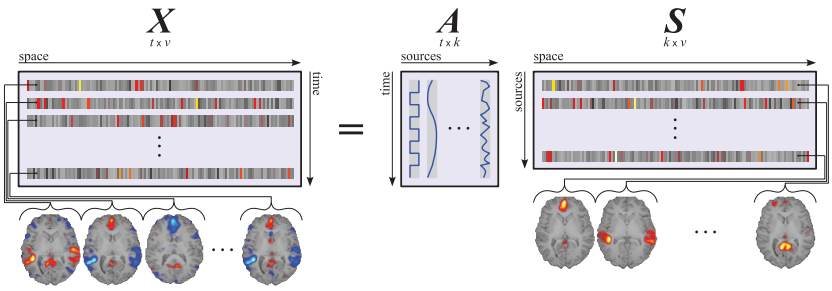


Figure 4.1. Spatial ICA of fMRI data. The rows of the data matrix $\boldsymbol{X}$ and source matrix $\boldsymbol{S}$ are the vectorized volumes. The columns of the mixing matrix $\boldsymbol{A}$ are the time series. Note, that the statistical independence applies to the spatial patterns.

As with GLM, to use fMRI data in the ICA model, all voxels in the imaging volume are rearranged as one row in the data matrix $\boldsymbol{X}$. Since ICA is only concerned with the statistics of the observations, the order of the voxels can be freely chosen, as long as every volume is reordered in the same way. Each row of $\boldsymbol{S}$ contains one of the K independent spatial patterns and the corresponding column of $\boldsymbol{A}$ holds its activation time series. In contrast to GLM, no assumptions are made on the shape of the individual time series.

5. Uncertainty and Exploiting Variability

Independent component analysis has become widely adopted in recent years for performing blind source separation. However, some methodological problems have remained with the use of ICA. One concern is the tendency of the estimated independent components to change slightly each time the analysis is performed. Such variability can be caused by many factors. For example, the theoretical assumption of statistical independence may not hold for the data. The particular ICA algorithm used may also be inherently stochastic. Furthermore, significant noise or other properties of the data can cause variations in the solutions.

Under expert supervision, such behavior has usually been overcome by comparing the estimated components with known sources, or with references obtained by other analysis methods [see, *e.g.*, Calhoun et al., 2001]. Another used approach is to evaluate the results based on expert knowledge of the potential sources. Such approaches may not be possible in practice, and effectively cancel the true benefits of data-driven analysis. Recently, bootstrap methods [Efron and Tibshirani, 1994] have been used successfully to identify consistently reproducible components [see, *e.g.*, Duann et al., 2003]. Additionally, some methods [*c.f.*, Himberg et al., 2004] have proposed grouping similar estimated solutions and visualizing the compactness of the estimates. However, the full potential of analyzing the consistency has not been exploited before. Moreover, the convergence of FastICA, especially under bootstrap, has not been properly studied or theoretically analyzed before.

This chapter presents a novel approach, based on observations made by the author in Publications **I** and **II**, offering a better insight to both ICA and the data. The approach runs the FastICA algorithm multiple times with bootstrap sampling, followed by a suitable clustering of the estimated solutions. The method easily reveals the consistent independent

components, but, perhaps more importantly, also helps to interpret the less consistent phenomena.

All the algorithmic optimization is performed in an optimization landscape that is partly defined by the data. Therefore, the data samples will also affect the convergence of the algorithm. Restricting structure in the data can also bias the estimated solutions. Perhaps the simplest example of such an effect is the increased number of local optima when the signal-to-noise ratio of the data is low. In addition to solving the reliability problems, consistency analysis of ICA can reveal the properties of the variability that may actually tell more about the underlying data. For example, components that are clearly separated from other components, but contain a high degree of variability could identify an independent subspace in the data. On the other hand, consistent components with very weak signal-to-noise ratios could be reliably discriminated from overfitting.

As explained in Publication **III** with greater detail, the variability can be considered on a surface defined by the objective function and the given data. For illustrative purposes, Figure 5.1 shows an example of a 3-dimensional optimization landscape, where in addition to the global minimum, the surface contains several local minima. Different starting points and directions, or the tendency to bias towards a certain solution, cause the optimization to converge along different paths, illustrated as thick curves. Ideally, a robust algorithm should always reach the same optimum, but in practice it can get stuck on a local optimum. Even the estimates of the same optimum, reached along different paths, can vary slightly, depending on the implementation of the algorithm. The estimates should still form consistent groups with high similarity. Thus, the most reliable solution can be found as the mean of a consistent group and the spread of the group can be used to analyze the properties of the variability.

5.1 Errors in Assumptions and Estimation

The ICA algorithms are usually derived under strict assumptions, such as statistical independence, stationarity and noiseless data. With real measurement data, the assumptions often fail and the behavior of the estimation cannot be guaranteed, even if the estimation would seem robust.

The algorithm itself may also be a source of variability. The selected op-

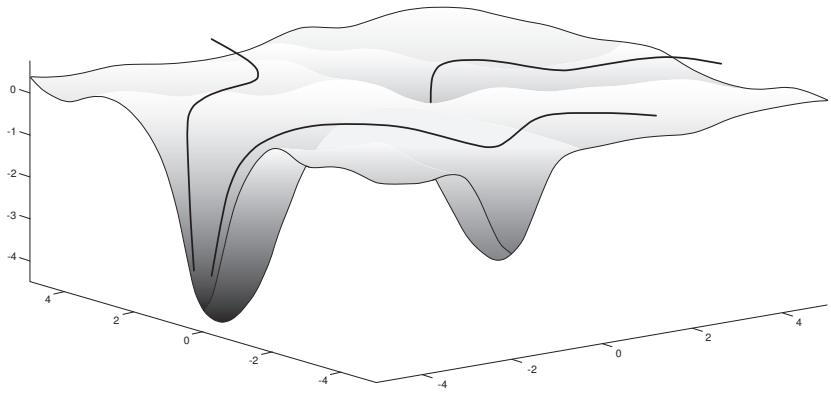


Figure 5.1. Optimization landscape. Different initial conditions and data properties cause the algorithm to converge along different paths (illustrated as curves) on the 3-dimensional optimization surface. The algorithm may get stuck on a local minimum and even the optimal estimates may be slightly different.

timization schemes and implemented iteration methods are usually tuned for fast convergence, which can affect the estimation accuracy of the algorithm. The performance of the algorithm in the presence of noise also depends on the implementation. Particularly, FastICA can search for the independent components one at a time, by deflation, or all at once. Methods that solve all the components at once tend to spread the errors evenly among all components, whereas in deflation, the errors accumulate to the last components. In FastICA, the sensitivity to outliers in the data can also be tuned with the choice of the nonlinearity in the update rule. Additionally, the initial conditions of the algorithm are usually selected randomly.

5.2 Sparseness and Overfitting

The measures commonly used in the optimization, like kurtosis, give relatively high scores for sparse sources, whether or not they are truly independent. Therefore, independence and sparseness are strongly connected [see, *e.g.*, Li et al., 2004], which results in a natural tendency of any ICA algorithm to bias towards sparse, rather than strictly independent, solutions. Indeed, when the degrees of freedom in ICA are too high, the model is likely to overfit to the data, which is known to result in strong sparse contamination of the estimated components [Särelä and Vigário, 2003]. Additionally, the overfitting contamination can change freely on each application of the algorithm. On the other hand, a too low model

order can result in poor separation altogether. Whitening, and a suitable dimensionality reduction, is important in controlling the complexity and the size of the decomposition.

Arguably, the good performance of ICA in practice, even under ill conditions, can be in part attributed to the fact that the sparse components often form a very natural and easily interpreted decomposition of the data [see, *e.g.*, McKeown and Sejnowski, 1998, Vigário et al., 2000]. In some cases, the best goal may even be to search for sparse, rather than strictly independent components. The problem with the natural bias towards sparse components is that it can easily be confused with the sparseness produced by overfitting. Thus, it is crucial to fully analyze the properties of the variability.

5.3 Bootstrap and Combining Estimates

The main idea in bootstrap [Efron and Tibshirani, 1994] is to get many values for an estimator, to be able to characterize its distribution. This is achieved by performing the estimation multiple times using a differently resampled version of the data each time. The goal is to create many possible versions of the data from a single set of actual measurements. The resampling is done with replacement, that is, using independently drawn samples of the data. Perhaps the most common example of bootstrap is the calculation of the standard deviation of an estimated parameter to assess its reliability. Resampling the available data to discover the distribution of an estimator, while retaining the structure within the data, has strong connections to Bayesian sampling theory [see, *e.g.*, Rubin, 1981, Clyde and Lee, 2001]. Essentially the same approach is used also in boosting and bagging [see, *e.g.*, Bauer and Kohavi, 1999].

The majority of previous work on bootstrap in machine learning has been done with supervised algorithms, where the performance is easy to verify with the given ground truth [Strother et al., 2002]. The behavior of unsupervised algorithms, like ICA, is more difficult to quantify, since there is no ground truth and it is difficult to know which estimates represent the same component [Baumgartner et al., 2000]. Most commonly the goal has been to analyze the stability of an algorithm or to select the best performing model [see, *e.g.*, Rao and Tibshirani, 1997, Breakspear et al., 2004]. The effects of artificially increasing the level of noise in the data, on the consistency of the independent components, have also been stud-

ied [Harmeling et al., 2003, 2004]. However, fMRI data, like many real datasets, has a poor signal-to-noise ratio to begin with.

A recent approach, proposed by Meinecke et al. [2002], relies on a single initial ICA decomposition, and aims at validating the consistency of those independent components. The consecutive runs are applied on resampled versions of the estimated independent components from the first run, with the assumption that the consecutive runs should result in an identity mixing matrix. Therefore, the method is straightforward to implement and sidesteps any issues arising from the sign and permutation ambiguities. However, since the estimation takes place only in the local vicinity of the initial solution, the optimization landscape may not be covered sufficiently, and the assumptions made in bootstrap could be violated. Additionally, the initial run essentially defines the set of independent components to be analyzed, and a poor solution on the initial run will render the consecutive runs practically useless.

On the other hand, Himberg et al. [2004] have proposed a visualization method, based on clustering components from multiple runs of ICA without heavy restrictions. However, their method is computationally heavy and leaves the user with a nonlinear two-dimensional cluster representation. Moreover, the visualization itself is unstable and changes every time the method is applied, making it difficult, or even misleading, to interpret.

A proper combination strategy for ICA has to take into account the sign, scale and order ambiguities in the model. Even the optimal number of components is usually not known. Some simple combination strategies, or validation of manual combinations, have been studied in connection with the reproducibility of ICA decompositions [see, *e.g.*, Duann et al., 2003, Ito et al., 2003, Stögbauer et al., 2004]. The combination approach in the new method is rather simple and efficient, allowing for easy interpretation of the results. It has strong connections to recently developed clustering methods based on spectral properties of a similarity matrix [see, *e.g.*, Meilă and Shi, 2000, 2001]. There are also some studies analyzing the stability of clustering solutions [see, *e.g.*, Lange et al., 2004].

As mentioned before, randomizing the initial conditions changes the starting point and direction of optimization. This allows converging to the optimum by approaching it from different directions, and actually, some of the points that are surrounded by local optima, as defined by properties of the data or noise, may only be reachable from certain directions. Therefore, starting with different initial conditions is important for com-

plete analysis, and helps in detecting even the weakest local optima. This can be crucial for data with poor signal-to-noise ratio, or when the data is such that it causes a strong bias or overfitting tendency.

5.4 The Arabica Toolbox

Performing FastICA many times under bootstrap resampling is actually rather difficult in practice, due to such problems as the sign, scaling, and permutation ambiguities. A toolbox based on all the previous considerations is introduced in Publication **IV**.

In the Arabica approach, a separate initial run is not used, allowing for any combination of independent components to be found on each run. One important benefit is that ICA is free to find the optimal solution on each run, which allows for the results to include weaker local mimima that are very difficult to find on all runs. Clearly, this leads to a harder combination problem. The FastICA algorithm is well suited for such use, since it is very robust and converges quickly even for large datasets. The approach is outlined in Table 5.1.

Consistent components are easy to identify, since the number of the grouped estimates is close to the number of bootstrap rounds performed. The mean, or centroid, of each cluster can be considered as the best estimate of the true decomposition. The mean representatives can also depart somewhat from a strictly independent solution, which can be beneficial when the strict independence cannot always be guaranteed. In cases where the strict independence is more crucial, the estimates from the single run closest to the mean solution could be picked instead.

To further characterize the differences among the components, measures of compactness and discrimination are calculated. Some of the weaker components appearing only a few times, since they are difficult for ICA to separate, can still be considered reliable due to a very compact and well discriminated nature. Ranking based on these measures allows for the significant components to be shown first, and similarly behaving ones close to each other. The variability can be fully interpreted by showing it in all of the similarity, mixing, and signal spaces simultaneously. The additional information can be used to correctly identify the underlying source signals much better than in a single run ICA approach.

The toolbox name was inspired by the finest species of coffee beans, but can also serve as the acronym for ”Adaptive Reliable Algorithms for

- Perform the following estimation steps many times.
    1. Draw a bootstrap sample of the given data matrix.
    2. Whiten the bootstrap sample using PCA, possibly reducing the dimensionality.
    3. Randomize the initial conditions for FastICA.
    4. Estimate a reasonable number of independent components.
- Form groups with the estimates corresponding to the same independent component.
    1. Collect all estimates from the multiple runs.
    2. Calculate a similarity matrix of the estimates, taking into account the ambiguities in ICA.
    3. Cluster the estimates, using given similarity threshold and linkage path length.
- Visualize the independent component with their variability information.
    1. Calculate the statistical properties of the variability in each cluster.
    2. Rank the clusters based on the number of estimates and their compactness.
    3. Show the components, highlighting their spatio-temporal variations.

Table 5.1. Bootstrap FastICA in the Arabica toolbox. The steps for reliable estimation of independent components.

Bootstrap ICA". The toolbox is open source and available at `http://launchpad.net/arabica`. It is implemented using Matlab, and supports execution in the LONI Pipeline environment (`http://pipeline.loni.ucla.edu`, Rex et al. [2003]). The design was motivated by the recent shift towards cluster and grid computing, to allow for ever larger datasets to be collected and analyzed in a feasible manner. The environment allows for graphical definition of processing flows consisting of connected modules, that can be executed in one or many computers. Such an environment is very useful, since it offers the ability to reuse, replace, and modify only certain parts of a larger processing pipeline. The environment also allows for easy sharing of parts, or even complete, processing pipelines between researchers. As an example, Figure 5.2 illustrates typical usage of the toolbox.

As mentioned before, other considerations include the amount of dimensionality reduction performed during the whitening stage, as well as the number of independent components estimated on each run. It is common practice to estimate as many independent components as the number of retained whitened signals. This causes the demixing matrix to be square and makes the computations somewhat easier. In the presented method, it can be beneficial for the number of components to be smaller. The reason can be considered from a deflation point of view. In the deflation approach, the estimated components are affected by the accumulated errors of all previously estimated components and the first components have a higher overfitting tendency. Similarly, in each bootstrap run, all the simultaneously estimated components affect each other. Overfitting, or otherwise easy to estimate components, can make matters worse by appearing in every run. Therefore, the choice of estimating a smaller set of independent components than the total number of whitened signals allows for more freedom to better account for the estimation errors. In each run, different sets of independent components can be estimated, resulting in a total number of reliable components that is typically still larger than the number of whitened signals.

There is also a theoretically justified way of making the approach even faster, while allowing for improved identification of weaker phenomena. The amount of data used on each run can be further reduced in the bootstrap sampling, by resampling the original data into smaller subsets. The easiest way of achieving this is by randomly picking less than the total number of samples available. However, a too strong reduction can even-

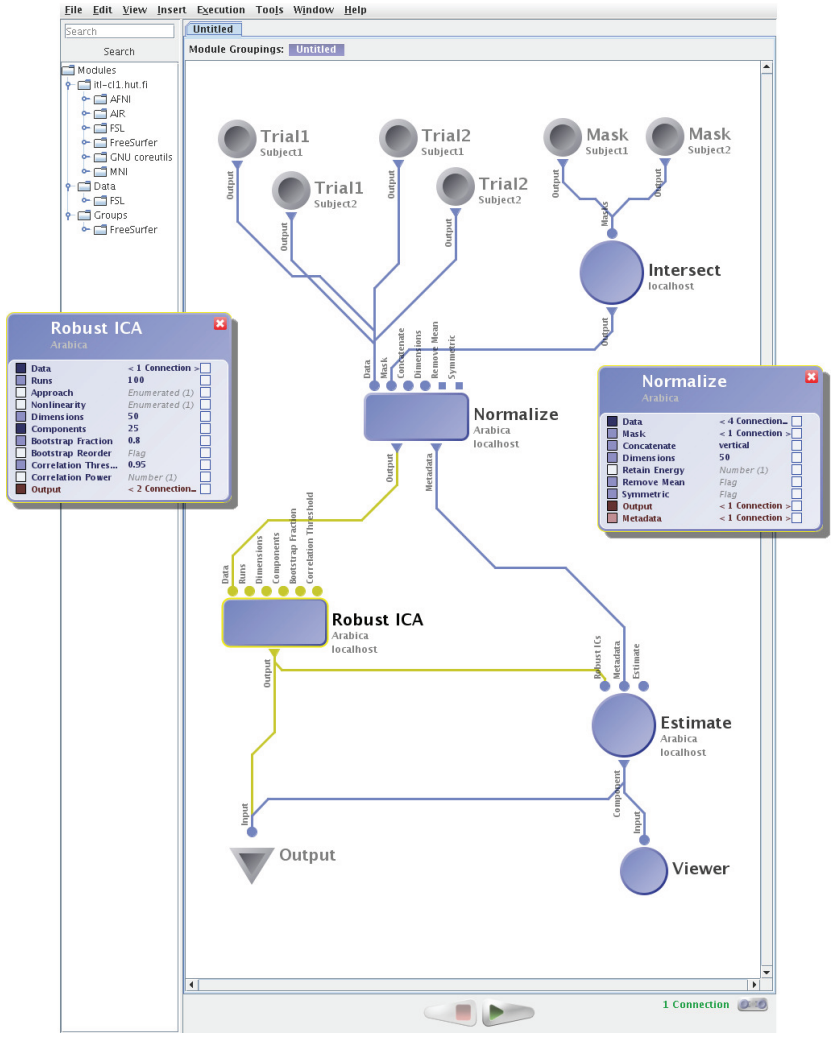


Figure 5.2. An illustrative example of a typical Arabica process in the LONI Pipeline environment. The interface shows an editable flow chart of a processing pipeline for multi-subject fMRI data. The popups show selected parameters for two of the Arabica modules. The illustrated process starts from the top with the dataset consisting of functional measurements of two subjects in two trials, followed by the individual skull stripping volume masks. In this simplified illustration, the datasets are already assumed to be suitably preprocessed. The datasets are fed to the normalize module, which allows for a group analysis of multiple datasets. The module also outputs metadata that is needed in the later stages of the pipeline. The normalized group data is the input to the robust ICA module that produces the clustered component estimates. The last steps in the example output the final results and visualize the reliable ICs of the individual subjects.

tually cause an increase in the variability of the estimates. Although the behavior is easy to confirm in practice, the amount of reduction should be controlled carefully since suitable values are most likely dependent on the data.

5.5 Reliable Analysis using ICA

The basic theoretical foundations of ICA, as well as its various implementations, are rather well understood [*c.f.*, Hyvärinen et al., 2001b, Comon, 1994, Cichocki and Amari, 2002]. In particular, for the FastICA algorithm, some theoretical convergence bounds have been presented earlier [see, *e.g.*, Hyvärinen, 1999, Oja and Yuan, 2006, Tichavský et al., 2006]. However, a complete proof for asymptotic convergence had not been shown before, and critically, the bootstrap setup had not been studied before. The following is a short summary of the new results presented in Publication **V**, giving formal theoretical grounds for the practical success of the bootstrap FastICA approach.

The main result of Publication **V** is a proof of asymptotic normality of FastICA and bootstrap FastICA, using the method of empirical process theory and Z-estimators [van der Vaart and Wellner, 1996]. This means that the bootstrap FastICA is a consistent estimator whose distribution around the true parameters approaches a normal distribution, with standard deviation shrinking in proportion to the sample size. Also, a probabilistic convergence rate is derived. In addition to the theoretical importance, the results of Publication **V** offer an elegant way to check the convergence of the bootstrap algorithm in practice, using a multivariate normality test [Anderson, 2003].

In the following, $\mathbb{E}[\cdot]$ is the expectation, Pr denotes a probability, and $\xrightarrow{P}$ means converges in probability. Let $\boldsymbol{z}$ denote the whitened data and $\boldsymbol{w}$ the demixing vectors in the whitened space from n runs of FastICA. More specifically $\boldsymbol{w}_\circ$ is the true solution, $\hat{\boldsymbol{w}}$ the sample estimator and $\hat{\boldsymbol{w}}^*$ the bootstrap estimator.

Theorem 1 (Consistency and Asymptotic Normality of FastICA)**.** *Let us assume* $\mathbb{E}[\boldsymbol{z}] = 0$ *and* $\boldsymbol{z}$ *has all moments up to the fourth;* $\mathbb{E}[\boldsymbol{z}\boldsymbol{z}^\top] = \boldsymbol{I}_d$*; and function* $G : \mathbb{R} \to \mathbb{R}$ *has bounded and continuous derivatives, and that* G *and its first derivative* g *are Lipschitz. Further, we assume that the quantities* $\mathbb{E}[g'(s_i)] \neq 0$*,* $\mathbb{E}[s_i^2 g'(s_i)] \neq 0$*,* $\mathbb{E}[g^2(s_i)] \neq 0$*,* $\mathbb{E}[s_i^2 g^2(s_i)] \neq 0$*,* $\mathbb{E}[G(\boldsymbol{w}^\top \boldsymbol{z})], \forall \boldsymbol{w} \in \mathcal{S}^{p-1}$*, and* $i = 1, \ldots, p$*, exist and are bounded. Then*

the sequence of estimates $\hat{\boldsymbol{w}}_n$ that is produced by the FastICA iteration, in Equation (4.16), is consistent and asymptotically normal, as

$$\hat{\boldsymbol{w}}_n \xrightarrow{P} \boldsymbol{w}_\circ$$
$$\sqrt{n}(\hat{\boldsymbol{w}}_n - \boldsymbol{w}_\circ) \rightsquigarrow \mathcal{N}(0, \boldsymbol{\Sigma}) \ ,$$

where the covariance of the Gaussian distribution is

$$\boldsymbol{\Sigma} = \boldsymbol{A}\text{diag}\left[\frac{\mathbb{E}[g^2(s_\circ)]}{\mathbb{E}[g'(s_\circ)^2]}, \ldots, \frac{\mathbb{E}[s_\circ^2 g^2(s_\circ)]}{(\mathbb{E}[s_\circ^2 g'(s_\circ))^2]}, \ldots, \frac{\mathbb{E}[g^2(s_\circ)]}{\mathbb{E}[g'(s_\circ)^2]}\right]\boldsymbol{A}^\top$$

and $\boldsymbol{A}$ is the true mixing matrix. In addition, assuming that $\mathbb{E}[g'^2(s_i)]$, $\mathbb{E}[s_i g'^2(s_i)]$, $\mathbb{E}[s_i^2 g'^2(s_i)]$, $\mathbb{E}[s_i^4 g'^2(s_i)]$, and the source signals, are bounded. Then the bootstrap FastICA is also asymptotically normal, as

$$\sqrt{n}(\hat{\boldsymbol{w}}_n^* - \hat{\boldsymbol{w}}_n) \rightsquigarrow \mathcal{N}(0, c^2 \boldsymbol{V}_{\boldsymbol{w}_\circ}^{-1} \boldsymbol{U}_{\boldsymbol{w}_\circ} \boldsymbol{V}_{\boldsymbol{w}_\circ}^{-1}) \ ,$$

where

$$\boldsymbol{U}_{\boldsymbol{w}_\circ} = \mathbb{E}[\boldsymbol{z}\boldsymbol{z}^\top g^2(\boldsymbol{w}_\circ^\top \boldsymbol{z})]$$
$$\boldsymbol{V}_{\boldsymbol{w}_\circ} = \mathbb{E}[\boldsymbol{z}\boldsymbol{z}^\top g'(\boldsymbol{w}_\circ^\top \boldsymbol{z})] \ ,$$

and c is a positive constant.

The theorem justifies the use of FastICA in a bootstrap and randomly initialized manner. However, on each run, a different set of independent components may be estimated and the total number of estimates for each component will vary. Moreover, the whitening step can change the signs of individual dimensions of the whitened space on each run, depending on the bootstrap sampled data. It is, therefore, crucial to correctly identify and group similar components from the various runs in practical implementations. Only then, a statistical analysis of each group can then be performed.

6. Networks of Related Independent Components

The purely hypothesis-driven methods, like GLM, have been used extensively in functional imaging studies, which have focused on rather simple block designs aimed at optimizing stimulus control and signal-to-noise ratio. Currently, the focus is shifting from simple unimodal stimuli towards integration of multiple sensory stimuli to study cognitive processes and, generally, brain activation related to natural stimuli. The recent interest in more natural setups is evident in experiments with real-life stimuli, such as movies [Hasson et al., 2004, Bartels and Zeki, 2005, Damoiseaux et al., 2006], video clips [Nishimoto et al., 2011], or human interactions [Kätsyri et al., 2012].

Modern research recognizes that brain functions can rely on distributed processing networks, and that a single brain region may participate in more than one function. As natural stimuli are increasingly used in fMRI studies, challenges are created in analyzing the measurements. It is no longer feasible to assume that a single feature of the experimental design could account for the brain activity. Instead, relevant combinations of a rich set of stimulus features could explain the more complex activation patterns.

The simplest approaches to model interactions among different brain regions are based on pair-wise correlations, often called *functional connectivity* [Friston, 1994]. Such methods suffer from spurious correlations due to confounding effects. Recently developed methods, such as dynamic causal modeling (DCM, Friston et al. [2003], Penny et al. [2004]) and Granger causality analysis [Granger, 1969, Seth, 2005, Sato et al., 2007], aim at *effective connectivity* analysis. They are based on auto-regressive or non-linear modeling of neuronal interactions. However, these approaches require a priori hypothesis of the interaction network using a small number of regions of interest.

This chapter presents the method first suggested in Publication **VI**, and then further refined in Publication **VII**. The aim of the approach is to infer brain correlates of natural stimuli using both ICA and CCA. In such uncontrolled setups, it is extremely difficult to differentiate the stimulus-related processes from ongoing brain activity or, analogously, the stimulus features related to brain activity from all other aspects of the natural environment. Statistical hypotheses can no longer be self-evidently derived from the experimental setup. Instead, identifying the testable hypotheses can be considered as one goal of the analysis.

6.1 Shared Dependencies Between Datasets

As a data-driven method, ICA looks for the spatially independent patterns of activity without any prior knowledge on the location or temporal dynamics of the activity. It has quickly become a common analysis tool in fMRI studies using natural stimuli [see, *e.g.*, Bartels and Zeki, 2005, Damoiseaux et al., 2006, Malinen et al., 2007]. However, in most studies the majority of the found independent patterns are left without an explanation in terms of the stimulus features. Some of the unexplained components can actually be unrelated to the stimulation, but it is possible to extend the analysis to identify some of the components as stimulus-related, by considering temporal correlations between them [see, *e.g.*, Calhoun et al., 2002].

In addition to ICA, other modern statistical machine learning methods, such as support vector machines [Vapnik, 1998], and Gaussian-process classifiers [Rasmussen and Williams, 2005], have recently been tested for finding brain correlates of individual stimulus features [see, *e.g.*, Kamitani and Tong, 2005, Haynes and Rees, 2005]. It has even been possible to predict the stimulus features from the brain activity with a well-trained predictive model [see, *e.g.*, Schneider et al., 2006, Sona et al., 2007, Nishimoto et al., 2011]. After training, the methods can be quite accurate, but they require training data with the correct target classes. Thus, they are only suited for studies where the experimental design readily defines suitable classes of interest, such as visual object categories.

All the above approaches have a practical problem with fMRI data, because of the huge dimensionality and number of samples in the recordings. Without regularization, interpreting the results can be very difficult and the methods can find strong correlations among the noise in the

datasets. In many studies, the problem has been solved by actually constraining the methods to a manually selected set of regions of interest. In the presented approach, ICA can be seen as a kind of regularization that imposes a purely data-driven, and theoretically justified, constraint on the data.

6.2 Features of Natural Stimuli

Under strict laboratory control, it can be possible to design an experiment so that all relevant states of the experimental variables are presented in a well balanced and adequately repeated fashion. However, with natural stimuli the analysis has to cope with imbalances in the data. Since CCA is invariant to linear transformations of the stimulus features, it is able to compensate for such imbalances, as long as the stimuli are expressive enough to include all relevant combinations of the experimental variables.

Naturally, when the experimental design forces certain stimuli to always co-occur, it is impossible to distinguish the corresponding brain correlates from each other. Somewhat surprisingly, similar confusion may happen with stimuli that never co-occur. When the intention is to design uncorrelated variables by creating a sequence without any overlap, the regularity of the setup can lead to unexpected negative correlations between the variables. Such emerging correlations cannot be distinguished from true negative correlations between the stimuli. Thus, any analysis could find dependencies that reflect the characteristics of the experimental design, rather than of the observed brain responses. Due to the importance of trying to avoid the effects of spurious correlations, the topic has recently acquired more scientific interest [see, *e.g.*, Aguirre et al., 1998, Fox et al., 2009, Murphy et al., 2009].

6.3 Semi-blind Approach in Two Steps

The new two-step approach uses ICA, as a first step, to identify spatially independent brain processes, which are considered to be the functional elements of brain activity within the measured data. As a second step, temporal dependencies between stimuli and the functional elements are identified using CCA. The proposed method is illustrated in Figure 6.1. Essentially, it looks for combinations of the stimulus features that max-

imally correlate with combinations of the functional elements. It is particularly attractive that the results are invariant to aspects of the stimuli that are not reflected in brain activity and, similarly, to the brain activity that is unrelated to the stimuli.

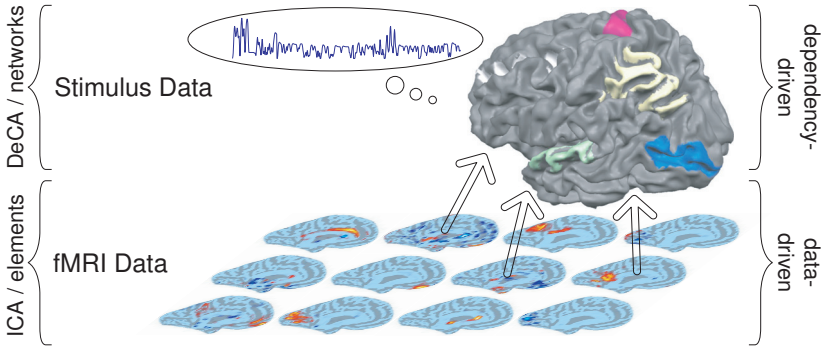


Figure 6.1. Illustration of the analysis framework. As the first step, ICA is applied to the fMRI measurements to find spatially independent patterns of brain activity. The second step uses CCA to identify functional combinations based on the temporal dependencies between the stimuli and the ICA components.

6.3.1 Identifying Functional Elements with ICA

The first step uses the reliable ICA method, discussed before, to find the functional elements of brain activity in the measurements. The ICA decomposition forms a natural data-driven basis to further describe the data, since it does not use the stimuli and does not make any assumptions about whether the components should be stimulus related or not.

Unfortunately, some literature has adopted the habit of calling independent components with distributed activation patterns as functional networks. The naming originates from simple region-based correlation analysis, but can be rather misleading when applied to ICA. The patterns could be networks, as a special case, when all the regions in the component would share the identical activation time series, without any delays. However, it is more likely that such cases are evidences of the forced co-occurance under the used stimuli or, even worse, of the spurious correlations, mentioned before. The proposed method makes a clear distinction between the functional elements and their combination into networks.

6.3.2 Linking Functional Networks with CCA

The second step uses CCA to look for combinations of the functional elements that are temporally correlated with combinations of the stimulus

features. This actually serves two purposes, since a typical problem with ICA is that the majority of the identified components are difficult to explain in terms of the stimuli. Firstly, the CCA results offer the best description of each component using the stimulus features. Secondly, when the stimulus features are better correlated with a combination of the functional elements, instead of just a single one, CCA identifies the best possible combination.

6.4 Networks of Brain Activation Related to Stimuli

The combination of many functional elements is a clear evidence for complex stimulus-related processes involving several brain regions. Such processes can be characterized as networks of brain regions, which dynamically function together in response to particular stimuli, such as networks responsible for multi-modal integration of sensory inputs.

7. Subspaces of Independent Components

The strict assumption of statistical independence used in ICA may not be completely valid for a given dataset. This is most likely the case for any fMRI measurement, since it is quite hard to believe that brain processes would be truly independent of each other. The dependency structure in the data, whether due to connections between the brain processes or spurious correlations, can lead to the formation of independent subspaces in addition to the easily separable individual directions. Subspaces can also form when there are no evidences of dependency among many brain processes during the measurement of the data.

Subspace ICA models generalize the assumption of component independence to independence between groups of components. Independent subspace analysis (ISA) was first introduced by Cardoso [1998] using geometric motivations. The model is quite general, and algorithmic enhancements in this setting have been studied, although identifiability and applicability to arbitrary random vectors have only been proven for equal group sizes. Moreover, if the observations contain additional spatial or temporal structures, these have been used for the multidimensional separation [Vollgraf and Obermayer, 2001].

In the special case of equal group sizes, Hyvärinen and Hoyer [2000] have proposed combining ICA with invariant feature subspace analysis, leading to more efficient algorithms. A related relaxation of the ICA assumption is given by topographic ICA [Hyvärinen et al., 2001a], where dependencies between all components are modeled along a topographic structure. On the other hand, Bach and Jordan [2003] formulate ISA as a component clustering problem using a graphical tree model. However, all the methods above are limited by the assumption of equal group sizes or less general semi-parametric models, and hence are not fully blind. Some recent development has been made towards a non-parametric ISA method

[see, *e.g.*, Theis, 2006], mainly suitable for dimensionality reduction.

The potential for identifying independent subspaces using the reliable ICA approach was hinted in Publication **II** and further studied in Publication **VIII**. Not unlike the network analysis discussed previously, the subspaces also reveal connections between the individual functional elements, and in a purely data-driven manner.

7.1 Subspaces Emerging from ICA

As explained before, the optimization landscape in ICA is defined by the structure of the data, additive noise, as well as the objective function used. When the data cannot be decomposed into purely independent directions, the landscape has areas with large valleys, or sets of many nearby local optima, instead of a clearly identifiable optimal point. Each bootstrap run, of the reliable ICA approach, can result in estimates from different parts of the valley, since according to the independence criterion used they are all equally optimal.

The emergence of a subspace in ICA also means that the basis directions within the subspace cannot be uniquely identified. In fact, independent components belonging to the subspace can appear with any rotation inside the subspace. Therefore, even if there is a strong relation between the subspace and the stimuli, the relation may not be readily visible as a high correlation between a component and the stimuli.

7.2 Two Ways to Identify Subspaces

The subspace detection is based on analyzing the variability of the estimated independent components. This can, and should, be done using both the mixing and the spatial patterns. In some cases, there are enough estimates of the mixing vectors that cover the whole subspace, so that the clustering of the mixing vectors can identify a very large group of related estimates. Sometimes the mixing vectors can form several distict groups, but the spatial patterns that they produce are still very related.

7.2.1 Estimates Linked Through Correlation

Having the bootstrap results of the reliable ICA approach, the most direct way of identifying subspaces is by looking for clusters, where the number

of estimates is much larger than the total number of bootstrap runs performed. This can only happen when many estimates, that produce very similar spatial patterns, were found during each run of the ICA.

For example, it may happen that the whole subspace is independent from all the rest of the components and easily separated, but within the subspace any direction results in roughly the same spatial pattern. Naturally the dimensionality of the subspace needs to be large enough, and the model order big enough, that more than one estimated component per run falls within the subspace. In such a case, it is questionable whether the directions favored by ICA are the best possible basis for the subspace.

7.2.2 Estimates with Shared Variability

It is also possible to identify a subspace by comparing the spatial variances of the components. When two or more components share the same spatial variance, that is only present in a few voxels, the estimation of those components must be dependent on each other. Additionally, the point of the variance often coincides with a strong point of activation in one of the components. This happens, for example, when one or more of the directions within the subspace are so distinct that the clustering does not directly group them all together. However, when the whole subspace appears rotated in a certain run of ICA, all the directions within the subspace tend to rotate together.

7.3 Further Analysis of Subspaces

Even though any direction within the independent subspace should look equally good for ICA, it is possible that the algorithm will prefer certain directions over others. For example, this can be due to the restrictions imposed by the other components found during the same run. On the other hand, there may be additional data available that would allow for more meaningful directions to be identified within a subspace.

The further refinement of the decompositions can be done either blindly by clustering them into their most consistent constituents, or by using CCA with some additional data.

7.3.1 Blind Refinement

A purely data-driven refinement can be done by clustering all the estimates belonging to the same subspace again, but using a much higher similarity threshold, and perhaps even a different similarity measure. This identifies the most distinct and consistent directions within the subspace. This may not help in interpreting the function of the whole subspace, but it could reveal details about the internal structure and dependencies within the subspace.

7.3.2 Refinement Using Additional Data

Another method is based on finding better basis directions inside the subspace using CCA. This approach needs additional data for the CCA to find the linear combinations that maximally correlate among the datasets.

If available, the additional data can be the stimulus features. This leads to a situation that is similar to the previous network analysis approach, with the difference that CCA is used to find the rotation of the whole subspace that best corresponds with the stimuli, regardless of any other components that may exist outside the subspace. The identified rotation should make the whole subspace, as well as each component, easier to interpret and describe in terms of the stimulus.

In a study with multiple trials or subjects, the additional data could also be a similar subspace identified from another dataset. Then the aim would be to find the rotations for the two subspaces that best correspond with each other. This would make identifying commonalities, or differences, between the trials or subject much more accurate. Such analysis would be easy to extend to more than two datasets with generalized CCA methods.

8. Discussion

In contrast to the classic statistical hypothesis testing approaches, modern machine learning methods allow for a purely data-driven analysis. One of the most widely used methods for data-driven signal decomposition is independent component analysis (ICA, Hyvärinen et al. [2001b]). However, some problems have remained with the use of ICA. One concern is the tendency of the estimated independent components to change each time the analysis is performed.

The main contribution of the thesis was a new reliable method for ICA, which is available in the Arabica toolbox. It is based on multiple runs of the FastICA algorithm [Hyvärinen and Oja, 1997] using bootstrap. The new method was also theoretically analyzed and its asymptotic convergence was proven. Secondly, the reliable ICA method was further extended by combining it with canonical correlation analysis (CCA, Hotelling [1936]) to identify networks of distributed brain activity. Thirdly, another improvement was made by adding ways to cope with independent subspaces, which often emerge when using real measurement data that is not guaranteed to fit all the assumptions made in the development of the methods.

Several possibilities for future improvements were also identified, both in the reliable ICA method itself and in the ways it could be applied to fMRI datasets, and potentially other kinds of data. The different topics and potential future research questions are discussed in detail below.

Reliable ICA

The new reliable ICA approach is very fast due to efficient bootstrap sampling and fast clustering. Surprisingly, the amount of samples drawn at each bootstrap iteration can be reduced quite heavily, making the overall

performance significantly better. This is mainly possible due to the robust estimation of the distribution, or more specifically, the non-Gaussianity, of the sources. The reduced sampling also has another benefit, as it makes the ICA less susceptible of errors caused by non-stationarity of the data. However, the minimum number of samples that still leads to good estimation is likely heavily dependent on the data, and should be studied more.

Since ICA has become an important tool for the multivariate analysis in functional magnetic resonance imaging (fMRI), the usefulness of the new approach was tested extensively with fMRI data. The tests showed that the method produces very robust results in practice. Furthermore, it was shown that the method is capable of providing insights into the data that would not be attainable otherwise. The method allows reliable interpretation of even the weaker components, which may be difficult for ICA to identify.

The phantom study also demonstrated that some components in real fMRI may stem from scanner artifacts, even though they appear reliable. Therefore, great care has to be taken when interpreting ICA components. Better approaches for quality control and functional calibration of the scanner could be studied in the future. Although other kinds of data were not studied in this thesis, the benefits should straightforwardly carry over to any kind of data.

The new method was also theoretically analyzed and its asymptotic convergence was proven. The theory offers a thorough explanation of how the method works and justifies its use in practice. The new theoretical results also offer ways to further develop the method in the future. Firstly, the current approach for clustering the estimated components could be improved by considering the fact that the intra-cluster distances of correctly grouped estimates should form a Gaussian distribution. The theory can even provide an estimate for the covariance of the distribution.

Secondly, an online clustering during the bootstrap iterations would allow the convergence of each component to be monitored using a normality test. This could be used as an overall stopping criterion for the algorithm, but also individually mark directions as sufficiently covered and force ICA to look in other directions.

Thirdly, if a good enough converge can be guaranteed by monitoring the bootstrap iterations, an early stopping criterion could be used to make the multiple runs even faster. Even though this would make any individual estimate weaker, it should not reduce the accuracy of the method, since

the final estimate is constructed as the mean of all the corresponding estimates.

Network Analysis

The reliable ICA method was further extended by combining it with CCA to identify networks of distributed brain activity. The extension was shown to be particularly useful with fMRI studies that use complex or natural stimuli. This is due to the fact that, unlike any other network analysis method for fMRI studies, the new approach can identify relevant combinations of both activation patterns and stimuli without requiring a predefined structure for the network. Therefore, the method should be capable of finding previously unknown network behavior in the brain.

However, the current implementation has no constraints on how the different activation patterns can be combined. This may not be the best approach, since in practice there is a difference between combining the activity of overlapping versus non-overlapping regions. Overlapping patterns could be due to increases in neuronal activation, changes in the neuronal population, or even inhibitory processes. Moreover, even when the combination of overlapping areas would correctly reflect changes in the activation, it is unlikely that the changes would be linear.

Developing a CCA method that would be based on a neuronal interaction model, instead of generic linear summing, could significantly improve the results in the future.

Subspace Analysis

Another improvement was made to the reliable ICA approach by studying independent subspaces, which often emerge when using real measurement data that is not guaranteed to fit all the assumptions made in the development of the methods. A novel approach was introduced for identifying the subspaces using the uncertainty of the estimates from the bootstrap ICA.

Furthermore, it was shown that it is still possible to improve the interpretability of the subspaces by refining the directions within the subspace or the rotation of the whole subspace. The refinement can be done both blindly, without using any additional information, or semi-blindly, by uti-

lizing the stimulus time series to identify more meaningful directions.

Correct comparisons over multiple datasets, such as, trials or subjects, are only possible by matching the subspaces in each dataset. The subspace analysis would benefit hugely from the ability to identify and cluster matching subspaces or components from several datasets. This could be possible by recent multi-way clustering approaches, and should definitely be studied further.

Bibliography

G. K. Aguirre, E. Zarahn, and M. D'Esposito. The inferential impact of global signal covariates in functional neuroimaging analyses. *Neuroimage*, 8(3):302, Oct. 1998. ISSN 1053-8119. doi: 10.1006/nimg.1998.0367.

S.-i. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In *Proceedings of the 1995 Conference on Advances in Neural Information Processing Systems (NIPS 1995)*, volume 8, pages 757–763, Denver, CO, Nov. 1995.

T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience, 3rd edition, 2003. ISBN 978-0471360919.

F. R. Bach and M. I. Jordan. Kernel Independent Component Analysis. *Journal of Machine Learning Research*, 3:1–48, 2002. doi: 10.1162/153244303768966085.

F. R. Bach and M. I. Jordan. Finding Clusters in Independent Component Analysis. In *Proc. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 891–896, 2003.

A. Bartels and S. Zeki. Brain dynamics during natural viewing conditions — A new guide for mapping connectivity in vivo. *NeuroImage*, 24(2):339–349, 2005. doi: 10.1016/j.neuroimage.2004.08.044.

E. Bauer and R. Kohavi. An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning*, 36(1-2):105–139, July 1999. doi: 10.1023/A:1007515423169.

R. Baumgartner, R. L. Somorjai, R. Summers, W. Richter, L. Ryner, and M. Jarmasz. Resampling as a Cluster Validation Technique in fMRI. *Journal of Magnetic Resonance Imaging*, 11(2):228–231, Feb. 2000. doi: 10.1002/(SICI)1522-2586(200002)11:2<228::AID-JMRI23>3.0.CO;2-Z.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159, Nov. 1995. doi: 10.1162/neco.1995.7.6.1129.

A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. Second-Order Blind Separation of Correlated Sources. In *Proceedings of the 1993 International Conference on Digital Signal Processing (DSP 1993)*, pages 346–351, Nicosia, Cyprus, July 1993. doi: 10.1.1.54.1662.

C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2nd edition, 2007. ISBN 978-0387310732.

G. M. Boynton, S. A. Engel, G. H. Glover, and D. J. Heeger. Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1. *Journal of Neuroscience*, 16(13):4207–4221, July 1996.

M. Breakspear, M. J. Brammer, E. T. Bullmore, P. Das, and L. M. Williams. Spatiotemporal Wavelet Resampling for Functional Neuroimaging Data. *Human Brain Mapping*, 23(1):1–25, Sept. 2004. doi: 10.1002/hbm.20045.

R. B. Buxton and L. R. Frank. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *Journal of Cerebral Blood Flow and Metabolism*, 17(1):64–72, Jan. 1997. ISSN 0271-678X. doi: 10.1097/00004647-199701000-00009.

V. D. Calhoun, T. Adali, V. B. McGinty, J. J. Pekar, T. D. Watson, and G. D. Pearlson. fMRI Activation in a Visual-Perception Task: Network of Areas Detected Using the General Linear Model and Independent Components Analysis. *NeuroImage*, 14(5):1080–1088, Nov. 2001. doi: 10.1006/nimg.2001.0921.

V. D. Calhoun, J. J. Pekar, V. B. McGinty, T. Adali, T. D. Watson, and G. D. Pearlson. Different activation dynamics in multiple neural systems during simulated driving. *Human Brain Mapping*, 16(3):158–167, 2002. doi: 10.1002/hbm.10032.

V. D. Calhoun, T. Adali, L. K. Hansen, J. Larsen, and J. J. Pekar. ICA of functional MRI Data: An Overview. In *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003)*, pages 281–288, Nara, Japan, Apr. 2003.

V. D. Calhoun, J. Liu, and T. Adali. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *NeuroImage*, 45(1): 163–172, 2009. doi: 10.1016/j.neuroimage.2008.10.057.

J.-F. Cardoso. Blind Identification Of Independent Components With Higher-order Statistics. In *Proceedings of the Workshop on Higher-Order Spectral Analysis (WHOSA 1989)*, pages 157–162, Vail, CO, June 1989.

J.-F. Cardoso. Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem. In *Proceedings of the 1990 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1990)*, volume 5, pages 2655–2658, Albuquerque, NM, Apr. 1990.

J.-F. Cardoso. Multidimensional Independent Component Analysis. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1998)*, pages 1941–1944, Seattle, Washington, USA, 1998.

W. Chau and A. R. McIntosh. The Talairach coordinate of a point in the MNI space: How to interpret it. *NeuroImage*, 25(2):408–16, Apr. 2005. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2004.12.007.

A. Cichocki and S.-i. Amari. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. Wiley-Interscience, New York, NY, 1st edition, June 2002. ISBN 978-0471607915.

M. A. Clyde and H. K. H. Lee. Bagging and the Bayesian Bootstrap. In *Proceedings of the 8th International Workshop on Artificial Intelligence and Statistics (AISTATS 2001)*, Key West, Florida, Jan. 2001.

P. Comon. Independent component analysis, A new concept? *Signal Processing*, 36(3):287–314, Apr. 1994. doi: 10.1016/0165-1684(94)90029-9.

A. M. Dale and R. L. Buckner. Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, 5(5):329–340, 1997.

J. S. Damoiseaux, S. A. R. B. Rombouts, F. Barkhof, P. Scheltens, C. J. Stam, S. M. Smith, and C. F. Beckmann. Consistent Resting-State Networks Across Healthy Subjects. *Proceedings of the National Academy of Sciences*, 103(37): 13848–13853, 2006. doi: 10.1073/pnas.0601417103.

G. Deco, V. K. Jirsa, P. A. Robinson, M. Breakspear, and K. J. Friston. The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS computational biology*, 4(8):1–35, Jan. 2008. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000092.

J.-R. Duann, T.-P. Jung, S. Makeig, and T. J. Sejnowski. Consistency of Infomax ICA Decomposition of Functional Brain Imaging Data. In *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003)*, pages 289–294, Nara, Japan, Apr. 2003.

B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability Series*. Chapman and Hall/CRC, 1st edition, 1994. ISBN 978-0412042317.

M. D. Fox, D. Zhang, A. Z. Snyder, and M. E. Raichle. The global signal and observed anticorrelated resting state brain networks. *Journal of neurophysiology*, 101(6):3270–83, June 2009. ISSN 0022-3077. doi: 10.1152/jn.90777.2008.

O. Friman, J. Cedefamn, P. Lundberg, M. Borga, and H. Knutsson. Detection of neural activity in functional MRI using canonical correlation analysis. *Magnetic Resonance in Medicine*, 45(2):323–330, Feb. 2001. doi: 10.1002/1522-2594(200102)45:2<323::AID-MRM1041>3.0.CO;2-\#.

K. J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1-2):56–78, Oct. 1994. ISSN 10659471. doi: 10.1002/hbm.460020107.

K. J. Friston. Functional integration and inference in the brain. *Progress in Neurobiology*, 68(2):113–143, Oct. 2002. ISSN 03010082. doi: 10.1016/S0301-0082(02)00076-X.

K. J. Friston, L. M. Harrison, and W. D. Penny. Dynamic causal modelling. *NeuroImage*, 19(4):1273–1302, Aug. 2003. ISSN 10538119. doi: 10.1016/S1053-8119(03)00202-7.

K. J. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. D. Penny. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, 2006. ISBN 978-0123725608.

G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins University Press, 3rd edition, 1996. ISBN 978-0801854149.

C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424–238, Nov. 1969. doi: 10.2307/1912791.

D. Hardoon, J. Mourão Miranda, M. Brammer, and J. Shawe-Taylor. Unsupervised analysis of fMRI data using kernel canonical correlation. *NeuroImage*, 37(4):1250–1259, 2007. doi: 10.1016/j.neuroimage.2007.06.017.

R. Hari and M. V. Kujala. Brain basis of human social interaction: from concepts to brain imaging. *Physiological Reviews*, 89(2):453–479, 2009. doi: 10.1152/physrev.00041.2007.

S. Harmeling, F. Meinecke, and K.-R. Müller. Analysing ICA Components by Injecting Noise. In *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003)*, pages 149–154, Nara, Japan, Apr. 2003.

S. Harmeling, F. Meinecke, and K.-R. Müller. Injecting noise for analysing the stability of ICA components. *Signal Processing*, 84(2):255–266, Feb. 2004. doi: 10.1016/j.sigpro.2003.10.009.

T. B. Harshbarger and A. W. Song. Differentiating sensitivity of post-stimulus undershoot under diffusion weighting: implication of vascular and neuronal hierarchy. *PloS one*, 3(8):e2914, Jan. 2008. ISSN 1932-6203. doi: 10.1371/journal.pone.0002914.

U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, and R. Malach. Intersubject synchronization of cortical activity during natural vision. *Science*, 303(5664):1634–1640, 2004. doi: 10.1126/science.1089506.

J.-D. Haynes and G. Rees. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, 8(5):686–691, 2005. doi: 10.1038/nn1445.

J. Himberg, A. Hyvärinen, and F. Esposito. Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage*, 22(3):1214–1222, July 2004. doi: 10.1016/j.neuroimage.2004.03.027.

H. Hotelling. Analysis of a Complex of Statistical Variables into Principal Components. *The Journal of Educational Psychology*, 24:417–441, 1933.

H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–377, 1936.

S. A. Huettel, A. W. Song, and G. McCarthy. *Functional Magnetic Resonance Imaging*. Sinauer Associates, Sunderland, MA, 2nd edition, Apr. 2008. ISBN 978-0878932863.

A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, May 1999. doi: 10.1109/72.761722.

A. Hyvärinen and P. Hoyer. Emergence of Phase- and Shift-Invariant Features by Decomposition of Natural Images into Independent Feature Subspaces. *Neural Computation*, 12(7):1705–1720, July 2000. ISSN 0899-7667. doi: 10.1162/089976600300015312.

A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural computation*, 9(7):1483–1492, Oct. 1997. doi: 10.1162/neco.1997.9.7.1483.

A. Hyvärinen, P. O. Hoyer, and M. Inki. Topographic independent component analysis. *Neural computation*, 13(7):1527–58, July 2001a. ISSN 0899-7667. doi: 10.1162/089976601750264992.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley-Interscience, New York, NY, 1st edition, May 2001b. ISBN 978-0471405405.

D. Ito, T. Mukai, and N. Murata. An Approach of Grouping Decomposed Components. In *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003)*, pages 745–750, Nara, Japan, Apr. 2003.

O. Jahn, A. Cichocki, A. A. Ioannides, and S.-i. Amari. Identification and elimination of artifacts from MEG Signals using efficient Independent Components Analysis. In *Proceedings of the 11th International Conference on Biomagnetism (BIOMAG 1998)*, volume 11, Sendai, Japan, Aug. 1998.

C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1):1–10, July 1991. doi: 10.1016/0165-1684(91)90079-X.

J. W. Kalat. *Biological Psychology*. Wadsworth Publishing, New York, NY, 10th edition, July 2008. ISBN 978-0495603009.

Y. Kamitani and F. Tong. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8(5):679–685, 2005. doi: 10.1038/nn1444.

J. Kätsyri, R. Hari, N. Ravaja, and L. Nummenmaa. The Opponent Matters: Elevated fMRI Reward Responses to Winning Against a Human Versus a Computer Opponent During Interactive Video Game Playing. *Cerebral Cortex*, page In press, Sept. 2012. ISSN 1460-2199. doi: 10.1093/cercor/bhs259.

J. R. Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58 (3):433–451, 1971. doi: 10.2307/2334380.

T. Lange, V. Roth, M. L. Braun, and J. M. Buhmann. Stability-Based Validation of Clustering Solutions. *Neural Computation*, 16(6):1299–1323, June 2004. doi: 10.1162/089976604773717621.

Y. Li, A. Cichocki, and S.-i. Amari. Analysis of Sparse Representation and Blind Source Separation. *Neural Computation*, 16(6):1193–1204, June 2004. doi: 10.1162/089976604773717586.

N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843): 150–157, July 2001. ISSN 0028-0836. doi: 10.1038/35084005.

S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski. Independent Component Analysis of Electroencephalographic Data. In *Proceedings of the 1995 Conference on Advances in Neural Information Processing Systems (NIPS 1995)*, volume 8, pages 145–151, Denver, CO, Nov. 1995.

S. Malinen, Y. Hlushchuk, and R. Hari. Towards natural stimulation in fMRI–issues of data analysis. *NeuroImage*, 35(1):131–139, 2007. doi: 10.1016/j.neuroimage.2006.11.015.

J. B. Mandeville, J. J. A. Marota, C. Ayata, M. A. Moskowitz, R. M. Weisskoff, and B. R. Rosen. MRI measurement of the temporal evolution of relative CMRO(2) during rat forepaw stimulation. *Magnetic Resonance in Medicine*, 42(5):944–951, Nov. 1999. ISSN 0740-3194. doi: 0.1002/(SICI)1522-2594(199911)42:5<944::AID-MRM15>3.0.CO;2-W.

D. Mantini, M. G. Perrucci, C. Del Gratta, G. L. Romani, and M. Corbetta. Electrophysiological Signatures of Resting State Networks in the Human Brain. *Proceedings of the National Academy of Sciences*, 104(32):13170–13175, 2007. doi: 10.1073/pnas.0700668104.

A. R. McIntosh. Towards a network theory of cognition. *Neural Networks*, 13(8-9):861–870, Nov. 2000. ISSN 08936080. doi: 10.1016/S0893-6080(00)00059-9.

M. J. McKeown and T. J. Sejnowski. Independent component analysis of fMRI data: Examining the assumptions. *Human Brain Mapping*, 6(5-6):368–372, Dec. 1998. doi: 10.1002/(SICI)1097-0193(1998)6:5/6<368::AID-HBM7>3.0.CO;2-E.

M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski. Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*, 6(3):160–188, Aug. 1998. doi: 10.1002/(SICI)1097-0193(1998)6:3<160::AID-HBM5>3.0.CO;2-1.

M. J. McKeown, L. K. Hansen, and T. J. Sejnowski. Independent component analysis of functional MRI: What is signal and what is noise? *Current Opinion in Neurobiology*, 13(5):620–629, Oct. 2003. doi: 10.1016/j.conb.2003.09.012.

M. Meilă and J. Shi. Learning Segmentation by Random Walks. In *Proceedings of the Neural Information Processing Systems Conference 2000 (NIPS 2000)*, pages 1–7, Denver, CO, Dec. 2000.

M. Meilă and J. Shi. A Random Walks View of Spectral Segmentation. In *Proceedings of the 8th International Workshop on Artificial Intelligence and Statistics (AISTATS 2001)*, pages 1–6, Key West, Florida, Jan. 2001.

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 Transactions on Biomedical Engineering*, 49(12):1514–1525, Dec. 2002. doi: 10.1109/TBME.2002.805480.

J. R. Moeller and S. C. Strother. A regional covariance approach to the analysis of functional patterns in positron emission tomographic data. *Journal of Cerebral Blood Flow and Metabolism*, 11(2):121–135, 1991. doi: 10.1038/jcbfm.1991.47.

K. Murphy, R. M. Birn, D. A. Handwerker, T. B. Jones, and P. A. Bandettini. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *NeuroImage*, 44(3):893–905, Mar. 2009. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2008.09.036.

S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J. L. Gallant. Reconstructing visual experiences from brain activity evoked by natural movies. *Current biology : CB*, 21(19):1641–6, Oct. 2011. ISSN 1879-0445. doi: 10.1016/j.cub.2011.08.031.

S. Ogawa, D. W. Tank, R. Menon, J. M. Ellermann, S.-G. Kim, H. Merkle, and K. Ugurbil. Intrinsic Signal Changes Accompanying Sensory Stimulation: Functional Brain Mapping with Magnetic Resonance Imaging. *Proceedings of the National Academy of Sciences*, 89(13):5951–5955, July 1992.

E. Oja and Z. Yuan. The FastICA Algorithm Revisited: Convergence Analysis. *IEEE Transactions on Neural Networks*, 17(6):1370–1381, 2006. doi: 10.1109/TNN.2006.880980.

B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, June 1996. doi: 10.1038/381607a0.

K. Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 2(6):559–572, 1901.

W. D. Penny, K. E. Stephan, A. Mechelli, and K. J. Friston. Comparing dynamic causal models. *NeuroImage*, 22(3):1157–1172, July 2004. ISSN 10538119. doi: 10.1016/j.neuroimage.2004.03.026.

M. E. Raichle and A. Z. Snyder. A default mode of brain function: a brief history of an evolving idea. *NeuroImage*, 37(4):1083–1090, Oct. 2007. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2007.02.041.

J. S. Rao and R. J. Tibshirani. The out-of-bootstrap method for model averaging and selection. *Technical Report*, pages 1–23, May 1997.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 1st edition, 2005. ISBN 978-0262182539.

D. E. Rex, J. Q. Ma, and A. W. Toga. The LONI Pipeline Processing Environment. *NeuroImage*, 19(3):1033–1048, 2003. doi: 10.1016/S1053-8119(03)00185-X.

D. B. Rubin. The Bayesian Bootstrap. *Annals of Statistics*, 9(1):130–134, Jan. 1981. doi: 10.1214/aos/1176345338.

J. Särelä and R. Vigário. Overlearning in Marginal Distribution-Based ICA: Analysis and Solutions. *Journal of Machine Learning Research*, 4:1447–1469, Dec. 2003.

J. a. R. Sato, P. A. Morettin, P. R. Arantes, and E. Amaro. Wavelet based time-varying vector autoregressive modelling. *Computational Statistics & Data Analysis*, 51(12):5847–5866, Aug. 2007. ISSN 01679473. doi: 10.1016/j.csda.2006.10.027.

W. Schneider, A. Bartels, E. Formisano, J. V. Haxby, R. Goebel, T. Mitchell, T. Nichols, and G. J. Siegle. Competition: Inferring Experience Based Cognition from fMRI. In *Proceedings of the 12th Meeting of the Organization for Human Brain Mapping (OHBM 2006)*, Florence, Italy, June 2006.

A. K. Seth. Causal connectivity of evolved neural networks during behavior. *Network: Computation in Neural Systems*, 16(1):35–54, Jan. 2005. ISSN 0954-898X. doi: 10.1080/09548980500238756.

D. Sona, S. Veeramachaneni, and E. Olivetti. Inferring Cognition from fMRI Brain Images. In *17th International Conference on Artificial Neural Networks (ICANN 2007)*, volume 4669, pages 869–878, Porto, Portugal, 2007. doi: 10.1007/978-3-540-74695-9_89.

H. Stögbauer, R. G. Andrzejak, A. Kraskov, and P. Grassberger. Reliability of ICA Estimates with Mutual Information. In *Proceedings of the 5th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2004)*, pages 209–216, Granada, Spain, Sept. 2004. doi: 10.1007/978-3-540-30110-3_27.

S. C. Strother, J. Anderson, L. K. Hansen, U. Kjems, R. Kustra, J. Sidtis, S. Frutiger, S. Muley, S. LaConte, and D. Rottenberg. The Quantitative Evaluation of Functional Neuroimaging Experiments: The NPAIRS Data Analysis Framework. *NeuroImage*, 15(4):747–771, Apr. 2002. doi: 10.1006/nimg.2001.1034.

J. Talairach and P. Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System - An Approach to Cerebral Imaging*. Thieme Medical Publishers, New York, NY, 1st edition, Jan. 1988. ISBN 978-0865772939.

A. C. Tang, B. A. Pearlmutter, N. A. Malaszenko, D. B. Phung, and B. C. Reeb. Independent Components of Magnetoencephalography: Localization. *Neural Computation*, 14(8):1827–1858, Aug. 2002. doi: 10.1162/089976602760128036.

F. J. Theis. Towards a general independent subspace analysis. In *Proceedings of the 2006 Conference on Advances in Neural Information Processing Systems (NIPS 2006)*, pages 1361–1368, Denver, CO, 2006.

P. Tichavský, Z. Koldovský, and E. Oja. Performance Analysis of the FastICA Algorithm and Cramér–Rao Bounds for Linear Independent Component Analysis. *IEEE Transactions on Signal Processing*, 54(4):1189–1203, 2006. doi: 10.1109/TSP.2006.870561.

N. H. Timm. *Applied Multivariate Analysis*. Springer, New York, NY, 1st edition, 2002. ISBN 978-0387953472.

G. Tononi, O. Sporns, and G. M. Edelman. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences*, 91(11):5033–5037, May 1994. ISSN 0027-8424. doi: 10.1073/pnas.91.11.5033.

A. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With applications to Statistics*. Springer, 1st edition, 1996. ISBN 978-0387946405.

V. N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, New York, NY, 1st edition, 1998. ISBN 978-0471030034.

R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja. Independent Component Approach to the Analysis of EEG and MEG Recordings. *IEEE Transactions on Biomedical Engineering*, 47(5):589–593, May 2000. doi: 10.1109/10.841330.

R. Vollgraf and K. Obermayer. Multi-dimensional ICA to separate correlated sources. In *Proceedings of the 2001 Conference on Advances in Neural Information Processing Systems (NIPS 2001)*, pages 993–1000, Denver, CO, 2001.

T. D. Wager, A. Vazquez, L. Hernandez, and D. C. Noll. Accounting for nonlinear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies. *NeuroImage*, 25(1):206–18, 2005. ISSN 10538119. doi: 10.1016/j.neuroimage.2004.11.008.

K. J. Worsley and K. J. Friston. Analysis of fMRI Time-Series Revisited — Again. *Neuroimage*, 2(3):173–181, 1995. doi: 10.1006/nimg.1995.1023.

A. Ziehe and K.-R. Müller. TDSEP - An Effective Algorithm for Blind Separation Using Time Structure. In *Proceedings of the 8th International Conference on Artificial Neural Networks (ICANN 1998)*, volume 8, pages 675–680, Skövde, Sweden, Sept. 1998.

DISSERTATIONS IN INFORMATION AND COMPUTER SCIENCE

Aalto-DD97/2012 Turunen, Ville T.
Morph-Based Speech Retrieval: Indexing Methods and Evaluations of Unsupervised Morphological Analysis. 2012.

Aalto-DD115/2012 Vierinen, Juha
On statistical theory of radar measurements. 2012.

Aalto-DD117/2012 Huopaniemi, Ilkka
Multivariate Multi-Way Modelling of Multiple High-Dimensional Data Sources. 2012.

Aalto-DD137/2012 Paukkeri, Mari-Sanna
Language- and domain-independent text mining. 2012.

Aalto-DD133/2012 Ahlroth, Lauri
Online Algorithms in Resource Management and Constraint Satisfaction. 2012.

Aalto-DD158/2012 Virpioja, Sami
Learning Constructions of Natural Language: Statistical Models and Evaluations. 2012.

Aalto-DD20/2013 Pajarinen, Joni
Planning under uncertainty for large-scale problems with applications to wireless networking. 2013.

Aalto-DD29/2013 Hakala, Risto
Results on Linear Models in Cryptography. 2013.

Aalto-DD44/2013 Pylkkönen, Janne
Towards Efficient and Robust Automatic Speech Recognition: Decoding Techniques and Discriminative Training. 2013.

Aalto-DD47/2013 Reyhani, Nima
Studies on Kernel Learning and Independent Component Analysis. 2013.

Since the beginning of the nineteenth century, neuroscience has focused on the idea that distinct regions of the brain support particular mental processes. However, modern research recognizes that many functions rely on distributed networks, and that a single brain region may participate in more than one function. The focus in brain research is rapidly shifting towards interaction-based functional networks, studied with both healthy subjects and patients suffering from neurological diseases. Another emerging field of research, also interested in distributed networks of brain activity, deals with resting subjects, without a particular task to perform. Functional magnetic resonance imaging has become a powerful tool in human brain mapping, typically used with a rather simple stimulus sequence tailored for statistical hypothesis testing. When natural stimuli are used, the simple designs are no longer appropriate. The aim of this thesis is in developing data-driven approaches for reliable inference of brain correlates to natural stimuli.

Aalto-DD 70/2013

ISBN 978-952-60-5136-9
ISBN 978-952-60-5137-6 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Information and Computer Science
www.aalto.fi