Alexandre Nuno de Morais Sayal Abreu Campos

BRAIN CONNECTIVITY ANALYSIS FOR REAL-TIME FMRI NEUROFEEDBACK EXPERIMENTS

Thesis submitted to the Faculty of Sciences and Technology of the University of Coimbra for the degree of Master in Biomedical Engineering with specialisation in Clinical Informatics and Bioinformatics, supervised by Prof. Dr. Miguel Castelo-Branco and Dr. Bruno Direito.

September 2016

UNIVERSIDADE DE COIMBRA

FCTUC

FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Alexandre Nuno de Morais Sayal Abreu Campos

Brain Connectivity Analysis for real-time fMRI Neurofeedback Experiments

Thesis submitted to the
University of Coimbra for the degree of
Master in Biomedical Engineering

Supervisors:

Prof. Dr. Miguel Castelo-Branco (IBILI, Faculty of Medicine, University of Coimbra)
Dr. Bruno Direito (IBILI, Faculty of Medicine, University of Coimbra)

Coimbra, 2016

This work was developed in collaboration with:

IBILI - Institute for Biomedical Imaging and Life Sciences

Faculty of Medicine - University of Coimbra

Esta cópia da tese é fornecida na condição de que quem a consulta reconhece que os direitos de autor são pertença do autor da tese e que nenhuma citação ou informação obtida a partir dela pode ser publicada sem a referência apropriada.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognize that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without proper acknowledgement.

Acknowledgments

The completion of this work would not have been possible without the collaboration and guidance of a number of people over these five years.

First of all, I would like to thank professor Miguel Castelo-Branco for the opportunity of studying at IBILI and participating in a number of projects currently in progress, as well as for the attentive and inspiring perspectives and supervision that led to the conclusion of this work.

Bruno Direito, as supervisor of this work and as a friend, collaborated in a infinite number of ways for the development and completion of this work. I cannot possibly thank him enough.

Proceeding without mentioning the fellow colleagues of "Sala de Cálculo" would be a serious mistake, and so I thank João Andrade, João Lima, Carlos Amaral, Susana Mouga, Marco Simões and Miguel Bernardes for the friendship and incredible work atmosphere.

Tiago Fernandes came from Lisbon (the only city in Portugal) and provided a fundamental contribution regarding the implementation and validation of causality measures, to whom I am truly thankful. He was a student of professor Alexandre Andrade, from the University of Lisbon, which also contributed greatly for the valid application of connectivity methods and to whom I am thankful for the availability and interest on this work.

A special thank you to Daniela Pereira, neuroradiologist that implemented the N-back experiment that was included in this work, and provided valuable feedback during the course of analysis.

The input and advice of Miguel Patrício was absolutely fundamental to the systematic review performed for the state of the art of this work. My sincere greetings.

To enumerate all the investigators at IBILI that contributed, one way or another, to the success of this work is a quite difficult task, and so I leave here a big greeting to all of them.

A major greeting to Michael Luehrs from BrainInnovation, The Netherlands, for the technical help provided on the company's software, neurofeedback techniques and implementation.

No work matters, of course, if not shared with our closest friends. To my colleagues and friends of "XIS" André Pedrosa, Fábio Reis, Sara Oliveira, André Matos, Rute Marto, Raquel Monteiro, Carolina Travassos, Mafalda Lima, João Marques and Mariana Barros, a huge thank you for all those moments that matter the most and cannot be forgotten.

My four months experience in Helsinki, Finland was one of the most personally and professionally changing. My special greeting to Aalto University, its professors, and to all my international colleagues which I know I can still call friends in an upcoming day. Kiitos!

For the unconditional support, joy and love over all these years, I do need to thank my one and only Dori, who gave a true meaning to all of this effort. I hope we can keep on travelling "na estrada para ser feliz".

Family matters the most, and so I certainly love to thank my mother and father for the biggest pillar of all: the one of daily support, sacrifice, friendship and love. The dedication needed to give the world's most kind meaning to "Para o infinito e mais além" cannot be written in this page. Obrigado.

To my sister, the biggest bear hug of all, specially for all those "boring" moments that make part of me, today and always.

Agora sem ser em "língua de crocodilo", um agradecimento especial ao meu avô materno, que me ajudou a ver o mundo desde que nasci. Também a minha avó materna merece estas palavras, pelo apoio mais feliz e mimado do mundo. Aos meus avós paternos, distantes mas presentes, o meu sincero obrigado.

To my uncles, a special thank you for the craziest support of all, that mattered so much.

A final greeting to the University of Coimbra and the Physics Department for hosting this incredible journey.

Resumo

Imagiologia por Ressonância Magnética Funcional (IRMf) é uma das técnicas mais inovadoras para estudar o cérebro humano. Com a utilização de um sistema de IRMf em tempo real, torna-se possível o desenvolvimento de experiências de "neurofeedback", em que o estímulo é adaptado em tempo real à actividade medida do cérebro do participante. Os processos cognitivos dependem da comunicação entre regiões no cérebro, e por isso, é de grande interesse a análise de conectividade antes, durante e depois da experiência de Neurofeedback. Este trabalho analisa e compara alguns dos métodos actualmente usados em dados de IRMf: Correlação, Coerência, Causalidade de Granger e Modelação Causal Dinâmica. Testes executados em dados simulados mostram que a métrica de Coerência consegue devolver resultados válidos mesmo com baixas taxas de amostragem dos sinais, ao contrário das métricas de Correlação e Causalidade de Granger. A análise preliminar de conectividade em dados experimentais apresenta resultados interessantes relativamente ao funcionamento da rede de memória de trabalho. Os estudos efectuados levantam algumas questões relacionadas com a aplicação de métodos de conectividade a dados de IRMf, validação de resultados e optimização de protocolos experimentais, assim como um número de desafios a serem ultrapassados em trabalho futuro.

Abstract

Functional Magnetic Resonance Imaging (fMRI) has evolved to become one of the most innovative methods to study the human brain. With a real time fMRI system, it is possible to perform neurofeedback experiments, where stimuli are adapted in real time to the participants' measured brain activity. Cognitive processes depend on the communication between regions of the brain, and so the analysis of connectivity before, during and after a neurofeedback experiment is of great interest. This work analyses and benchmarks some of the methods currently used on fMRI data: Correlation, Coherence, Granger Causality and Dynamic Causal Modelling. Tests performed on simulated datasets showed that Coherence presents robustness to downsampling, unlike Correlation and Granger Causality. Preliminary analysis of connectivity on an experimental dataset allowed for interesting insights on the working memory network functioning. The work developed has raised a number of questions related to the application of connectivity methods to Functional Magnetic Resonance Imaging (fMRI) data, results validation and optimisation of fMRI experimental setups, as well as a number of challenges to be surpassed in future work.

List of Acronyms

AAL Automated Anatomical Labelling.

AC Anterior Commissure.

ACC Anterior Cingulate Cortex.

AG Angular Gyrus.

AIC Akaike Information Criteria.

AMG Amygdala.

ANOVA Analysis of Variance.

BA Brodmann Area.

BIC Bayesian Information Criteria.

BOLD Blood Oxygenation Level-Dependent.

CBF Cerebral Blood Flow.

CD Causal Density.

CEN Central Executive Network.

CF Causal Flow.

CG Cingulate Gyrus.

DACC Dorsal Anterior Cingulate Cortex.

DACG Dorsal Anterior Cingulate Gyrus.

DCM Dynamic Causal Modelling.

deoxyHb Deoxygenated Haemoglobin.

DLPFC Dorsolateral Prefrontal Cortex.

DMN Default Mode Network.

DMPFC Dorsal Medial Prefrontal Cortex.

DTI Diffusion Tensor Imaging.

EEG Electroencephalography.

EPI Echo-Planar Imaging.

FDR False Discovery Rate.

FEF Frontal Eye Fields.

fMRI Functional Magnetic Resonance Imaging.

FWE Family-wise Error.

FWHM Full Width at Half Maximum.

GCM Granger Causality Mapping.

GLM General Linear Model.

HIPS Horizontal segment of Intraparietal Sulcus.

HRF Haemodynamic Response Function.

IC Independent Component.

ICA Independent Component Analysis.

IFG Inferior Frontal Gyrus.

IPL Inferior Parietal Lobe.

IPS IntraParietal Sulcus.

LPFC Lateral Prefrontal Cortex.

LWR Levinson-Wiggins-Robinson.

M1 Primary Motor Cortex.

ME Motor Execution.

MFG Medial Frontal Gyrus.

MFPC Medial Frontopolar Cortex.

MI Motor Imagery.

MPFC Medial Prefrontal Cortex.

MPRAGE Magnetisation-Prepared Rapid Gradient-Echo.

MRI Magnetic Resonance Imaging.

MTG Medial Temporal Gyrus.

NF Neurofeedback.

NMR Nuclear Magnetic Resonance.

OFC Orbito-Frontal Cortex.

oxyHb Oxygenated Haemoglobin.

PC Posterior Commissure.

PCG Posterior Cingulate Gyrus.

PeC Pearson's Correlation.

PFC Prefrontal Cortex.

pgACC Pregenuel Anterior Cingulate Cortex.

PMA Premotor Cortex.

PPC Posterior Parietal Cortex.

PPI Psychophysiological Interaction.

PPL Posterior Parietal Lobe.

PrecentralG Precentral Gyrus.

pSTS Posterior Superior Temporal Sulcus.

rACC Rostral Anterior Cingulate Cortex.

RFX Random Effects.

ROI Region of Interest.

rt-fMRI Real-time Functional Magnetic Resonance Imaging.

rt-fMRI-NF Real-time Functional Magnetic Resonance Imaging Neurofeedback.

SEF Supplementary Eye Fields.

SEM Structural Equation Modelling.

SFG Superior Frontal Gyrus.

SMA Supplementary Motor Area.

SN Salience Network.

SNR Signal-to-Noise Ratio.

SPL Superior Parietal Lobe.

STS Superior Temporal Sulcus.

SVAR Structural Vector Autoregression.

TAL Talairach Space.

TE Echo Time.

TR Repetition Time.

VAR Vector Autoregression.

VMPFC Ventral Medial Prefrontal Cortex.

Contents

List of Acronyms	xiii
-------------------------	-------------

Contents	xv
-----------------	-----------

1 Introduction	1
-----------------------	----------

1.1 Motivation	2
--------------------------	---

1.2 Main Contributions	2
----------------------------------	---

2 Background	3
---------------------	----------

2.1 fMRI Principles	3
-------------------------------	---

2.1.1 History	3
-------------------------	---

2.1.2 BOLD Imaging	4
------------------------------	---

2.2 fMRI Experiment Design	5
--------------------------------------	---

2.2.1 Experimental Design	6
-------------------------------------	---

2.2.2 Practical Considerations	7
--	---

2.3 fMRI Data Processing and Analysis	8
---	---

2.3.1 Preprocessing	9
-------------------------------	---

2.3.2 Transformation to Normalised Space	10
--	----

2.3.2.1 Anatomical-Functional Coregistration	10
--	----

2.3.2.2 Spatial Normalisation	11
---	----

2.3.3 General Linear Model (GLM)	12
--	----

2.3.4 Region of Interest Analysis	14
---	----

2.3.5 Multi-Subject Analysis	15
--	----

2.4 fMRI Neurofeedback Experiments	16
--	----

2.4.1	Definition	16
2.4.2	Experimental Setup	16
2.4.3	Neurofeedback Types	18
2.5	Brain Connectivity	19
2.5.1	Functional Connectivity	19
2.5.1.1	Pearson's and Partial Correlation	19
2.5.1.2	Coherence and Partial Coherence	20
2.5.2	Effective Connectivity	21
2.5.2.1	Granger Causality Mapping (GCM)	22
2.5.2.2	Dynamic Causal Modelling (DCM)	23
2.5.2.3	Structural Equation Modelling (SEM)	25
2.5.2.4	Structural Vector Autoregression (SVAR)	26
2.6	Case Study - Working Memory Network	27
3	State of the Art	29
3.1	Review Structure	29
3.2	Functional Connectivity	31
3.2.1	Pearson's Correlation	31
3.3	Effective Connectivity	38
3.3.1	Granger Causality Mapping (GCM)	38
3.3.2	Dynamic Causal Modelling (DCM)	40
3.3.3	Structural Equation Modelling (SEM)	41
3.3.4	Structural Vector Autoregression (SVAR)	42
4	Datasets	45
4.1	Simulated fMRI Dataset	45
4.1.1	Processes in the order of millisecond	47
4.1.2	Processes in the order of second	47
4.2	N-back Task Dataset - Working Memory	48
4.2.1	Experimental Setup	48
4.2.1.1	Participants	48

4.2.1.2	MRI acquisition	48
4.2.1.3	Functional Localiser	48
4.2.1.4	Neurofeedback Runs	49
4.2.2	Neurofeedback Calculation	50
4.2.3	Offline Analysis	51
4.2.4	ROI Selection and Time Course Export	51
5	Methods	53
5.1	Pearson's and Partial Correlation	53
5.2	Coherence and Partial Coherence	54
5.3	Granger Causality Mapping	55
5.4	Dynamic Causal Modelling	56
6	Results	59
6.1	Simulated fMRI Dataset	59
6.1.1	Processes in the order of millisecond	59
6.1.1.1	Pearson's Correlation	59
6.1.1.2	Coherence	60
6.1.1.3	Granger Causality Mapping	61
6.1.2	Processes in the order of second	62
6.1.2.1	Pearson's Correlation	62
6.1.2.2	Coherence	62
6.1.2.3	Granger Causality Mapping	63
6.2	N-back Task Dataset - Working Memory	65
6.2.1	Group GLM and ROI Selection	65
6.2.1.1	Brief overview on the function of the ROIs found in the Localiser	66
6.2.2	Connectivity Analysis for the Localiser Run	67
6.2.2.1	Pearson's and Partial Correlation	67
6.2.2.2	Coherence and Partial Coherence	68
6.2.2.3	Granger Causality Mapping	71

6.2.3	Connectivity Analysis for the Neurofeedback Runs	72
6.2.3.1	Pearson's and Partial Correlation	72
6.2.3.2	Coherence and Partial Coherence	73
6.2.3.3	Granger Causality Mapping	75
6.2.4	Connectivity changes between Train and Transfer Runs	77
6.2.4.1	Pearson's and Partial Correlation	77
6.2.4.2	Coherence and Partial Coherence	78
6.2.4.3	Granger Causality Mapping	79
6.2.5	Dynamic Causal Modelling	81
6.2.5.1	DCM for the Localiser Run	81
6.2.5.2	DCM for the Neurofeedback Runs	82
6.2.5.3	DCM for Train and Transfer Runs	82
7	Discussion	85
7.1	Results for the Simulated fMRI Dataset	85
7.2	Results for the N-back Task Dataset	87
7.3	DCM Analysis	91
7.4	Comparison of Connectivity Methods	92
7.5	Experimental and Data Processing Notes	93
8	Conclusions	95
8.1	Future Work	96
	References	99
	Online References	105

1

Introduction

fMRI has evolved to become one of the most widely used methods to study the human brain. It allows to non-invasively record high-resolution images of Blood Oxygenation Level-Dependent (BOLD) responses as an indirect measure of neuroelectric activity.

By implementing a Real-time Functional Magnetic Resonance Imaging (rt-fMRI) system, it is possible to perform a new type of experiment where stimuli are adapted in real time to the participants' measured brain activity. The real-time biofeedback of neuronal responses is called Neurofeedback (NF). There is increasing evidence of higher levels of activation in regions of interest after sessions of NF, with participants successfully managing to modulate their neuronal responses to a wide variety of stimuli [Coben and Evans, 2011].

Based on the assumption that cognitive processes depend not only on the activation of a single region in the human brain, but also (or even mainly) on the communication between regions related to those processes, the evaluation of connectivity before, during and after a neurofeedback session reveals itself fundamental for a more complete understanding of the functioning human brain.

There are a number of methods to assess neural connectivity in fMRI data, and an even larger number of different implementations of those methods. Regarding Functional connectivity, Pearson's and Partial Correlation and Coherence are the best known, and regarding Effective connectivity, Granger Causality Mapping (GCM), Structural Equation Modelling (SEM) and Dynamic Causal Modelling (DCM) are the most common. Each of these has got advantages and disadvantages, including some methodological drawbacks that will be discussed.

This thesis aims to provide a detailed explanation on the analysis of fMRI data from a brain connectivity perspective, together with all the challenges encountered: starting with raw scanner data, through preprocessing and methods implementation, and ending with visualisation of connectivity maps.

1.1 Motivation

Analysing connectivity between areas can provide us a more complete characterisation of the functioning brain than mere activation patterns: it is proven and expected that function relies not on a single, isolated region, but on a network of regions working together. Because of this, it is expected that neurofeedback training may affect the functional patterns of the network. The ability to analyse these connectivity patterns is of major importance and constitutes the main motivation for this work.

Additionally, after an overall review of the published articles on connectivity for NF experiments, it is clear that the search for the best method and implementation is far from over. The multiple implementations of each method, together with its application validity and reliability, should be carefully evaluated and systematised.

1.2 Main Contributions

The main contributions of this work are:

- Creation of a pipeline of analysis that can provide the researcher information about the connectivity patterns between regions, based on a systematic analysis and comparison of the methods and understanding of the strengths and caveats of each one.
- Connectivity methods validation for fMRI data, using a simulated fMRI dataset with known connectivity patterns.
- Study and optimisation of a series of parameters related to fMRI data acquisition, experimental design, statistical frameworks, among others.
- Preliminary analysis of connectivity metrics on real functional data from a neurofeedback fMRI experiment.

2

Background

This chapter introduces a number of topics explored throughout the work. Basic theoretical and practical aspects of each topic will be provided, together with some more detailed mathematical aspects.

2.1 fMRI Principles

The ability to watch a functioning human brain is one of the most striking scientific developments.

2.1.1 History

The first attempt of mapping the human brain function takes us back to the early nineteenth century, when Phrenologists stated that the amount of brain devoted to a cognitive function determined its influence on behaviour. The brain maps were based on questionable descriptions of individuals with extreme characteristics, and as such, the idea collapsed on scientific grounds in the late 1830s. Even so, Phrenology introduced the idea of **brain function localisation** - different areas of the brain are responsible for different functions - and in the next decades, scientists began the search for changes in brain physiology.

In 1924, the physicist Wolfgang Pauli made a very important discovery, that is still one of the fundamental pillars of Magnetic Resonance Imaging (MRI). He postulated that atomic nuclei had properties, namely spin and magnetic moment, that could only take discrete values (*quanta*). After this, some experiments were performed which confirmed Pauli's statement. In 1937, Isidor Rabi performed an experiment using strong oscillating magnetic fields, a detector plate and a gaseous beam of a single element, allowing him to realise that if the frequency of

the oscillating magnetic field matched the spin frequency of the atomic nucleus, the nucleus would absorb energy from the field. This effect was called **magnetic resonance**.

In 1945, Edward Purcell attempted to demonstrate magnetic resonance in a block of paraffin wax: if he matched the resonance frequency of the wax to the oscillating magnetic field, the wax would absorb energy, and thus change its electrical conductance. With a strong magnetic field, he was able to verify a clear resonance effect. Using a different system to prove the same effect on bulk matter, Felix Bloch was able to measure changes in the energy absorbed by water. He named this **Nuclear Induction**, or as it is known today, Nuclear Magnetic Resonance (NMR) - the pillar of all current MRI techniques.

In the following years, several major discoveries were made in this field. Perhaps one of the most important is the invention of Echo-Planar Imaging (EPI), purposed by Peter Mansfield in 1976: NMR data was collected from an entire image slice at one time, by sending one electromagnetic pulse from a transmitter coil and then rapidly changing magnetic field gradients while recording the MR signal. This signal could be converted into an image using Fourier analysis.

In 1977, the first human NMR scanner was built by Raymond Damadian's FONAR Corporation. The field strength was just 0.05 T and the complete image took nearly 4 hours to acquire, but in the end the experiment was successful. As a result of the "negative" connotation attributed to the word nuclear, which was associated with ionising radiation, in the early 1980s NMR became MRI - Magnetic Resonance Imaging. General Electric created the first commercial human-body 1.5 T scanner in 1982.

2.1.2 BOLD Imaging

The transition from MRI to functional MRI relies on two fundamental aspects: firstly, the identification of a biophysical property that is altered by information processing in the brain. Secondly, an MRI contrast that can detect this property.

Back in 1936, Linus Pauling discovered that Haemoglobin, the protein responsible for carrying oxygen in the blood, has different magnetic properties depending on whether it is or not bound to oxygen. Oxygenated Haemoglobin (oxyHb) is **diamagnetic** (no unpaired electrons and zero magnetic moment) while Deoxygenated Haemoglobin (deoxyHb) is **paramagnetic** (with unpaired electrons and a significant magnetic moment). As a consequence, completely deoxygenated blood has a magnetic susceptibility about 20% larger than fully oxygenated blood.

[Ogawa et al., 1990] demonstrated that the presence of deoxygenated blood, compared with the presence of oxygenated blood, decreases the MR signal. This finding, later called Blood Oxygenation Level-Dependent (BOLD), is the contrast that enables the measurement of functional changes in the brain.

When a certain region of the brain increases its activity in response to a stimulus, the extraction of oxygen from the local capillaries increases, leading to an initial drop in the concentration of oxyHb and an increase in deoxyHb. It would be expected that, as the level of deoxyHb increases, the MR signal would decrease. But in fact, 3 to 6 seconds after the stimulus onset, an increase in the MR signal is observed. This is due to a large increase in Cerebral Blood Flow (CBF), which delivers a surplus of oxyHb and clears deoxyHb.

The Haemodynamic Response is the change in the MR signal triggered by neuronal activity. This is a smooth signal, representing the changes in blood flow that MRI detects, with a small dip in signal before and after the larger increase in signal, possibly reflecting a transient imbalance between the metabolic activity and blood flow. A typical Haemodynamic Response Function (HRF) is shown in figure 2.1. The shape of the response can greatly vary, depending on the stimulus, the brain region and the local vasculature.

Figure 2.1: Typical Haemodynamic response function (adapted from [Siero et al., 2013]).

2.2 fMRI Experiment Design

All scientific research begins with a question.

When designing an fMRI experiment, several parameters must be chosen and decisions must be taken in order to answer the questions of interest in the most efficient way. A few of these will be explored in the following sections.

To test an hypothesis like "Manipulating these independent variables will cause changes in the dependent variables", the experiment must contain at least two values of the independent

variables and must be able to measure changes in the dependent variables. The simplest way to set up an experiment is to have two conditions that occur at different times. These are usually separated into an experimental condition and a control condition, which differ only in the effect of interest [Huettel et al., 2004]. The experimental condition is often called the task condition and the control condition is often called the baseline condition.

2.2.1 Experimental Design

The first decision is on the type of experimental design. A block design scheme is shown on figure 2.2. Multiple events of a condition are clustered in blocks and presented alternately with baseline (rest) blocks.

Figure 2.2: Block design fMRI.

The analysis is usually performed by contrasting between conditions (for instance, Task vs Baseline or Task 1 vs Task 2). A few considerations about this design are:

1. Choosing a Baseline: it could be as simple as a fixation cross or a blank screen, but it might be of interest to use the baseline blocks to subtract activation that occurs concurrently with the process of interest.
2. Order of conditions: the order of conditions should be randomised across runs in order to avoid order effects (for instance, condition 1 could influence the processing in condition 2).
3. Block duration: due to haemodynamic lag, the BOLD signal can take some time to increase/decrease. As so, the block has to be long enough to assure that the signal has sufficient time to increase in task blocks and to return to its default value in baseline blocks. This duration is always a multiple of the Repetition Time.
4. Run duration: To avoid distortion of the BOLD response and participant's fatigue, runs should be kept around 5 to 10 minutes in length.

In an event-related design, events of a condition are presented individually, rather than clustered in blocks (figure 2.3). Each event is usually separated by 12-14 seconds, as to allow the BOLD signal to return to its baseline value. There is also the case of rapid event-related designs, which use variable inter-stimuli intervals.

Figure 2.3: Event-related design fMRI.

Since this type of design is not suitable for Real-time Functional Magnetic Resonance Imaging Neurofeedback (rt-fMRI-NF) experiments, as it will be made clear in section 2.4, further considerations about this design will not be given, but can be explored in Chapter 11 of [Huettel et al., 2004].

2.2.2 Pratical Considerations

One of the most important parameters for connectivity analysis in fMRI experiments is the Repetition Time (TR). This is the time between the acquisition of volumes, and it determines the sampling frequency of the BOLD signals. Intuitively, the lower its value, the higher the temporal resolution of the signal. However, one should find an optimal value regarding the goal of the experiment, since the number of slices (single 2D MRI images) that is possible to acquire decreases with the decrease of the TR value, and so, a smaller section of the brain will be covered. This trade-off is shown in figure 2.4: on the left side with a TR = 2 s, the maximum number of slices is 32; on the right side with a TR = 1 s, the maximum number of slices is 16. All other acquisition parameters were kept constant.

Figure 2.4: Relation between the TR and the number of slices. Left side image with a TR of 2 s, right side image with a TR of 1 s.

Connectivity methods greatly rely on the temporal resolution of the signal to accurately retrieve the functional connections between regions. A slow acquisition rate of 0.5 Hz (corresponding to a typical TR of 2 s) is an obstacle for this analysis. Assuming that the goal of an fMRI experiment is to assess neural connectivity, the TR should be set as small as possible. Nevertheless, in section 4.1 is presented a dataset which intends to prove if even with low acquisition rates the signal contains enough information for the methods to work.

Other MRI parameters that have to be set:

1. Slice orientation - adjusted to the brain coverage intended.
2. Slice thickness - smaller the value, higher the spatial resolution in the direction perpendicular to the slice orientation.
3. In-plane resolution - 2D functional image resolution. The higher its value, the lower the number of slices it is possible to acquire.

2.3 fMRI Data Processing and Analysis

The steps to take from scanner data to activation maps.

To accomplish an interpretable functional map of the brain from raw scanner data, some processing and analysis steps must be taken. The pipeline of this process is shown in figure 2.5.

The software used for data processing was BrainVoyager QX 2.8 from BrainInnovation [1] and pipeline automation was performed using MATLAB [6] routines. In the following sections, each of the processing steps will be explored.

Figure 2.5: Pipeline of fMRI data analysis.

2.3.1 Preprocessing

fMRI data suffers from spatial and temporal inaccuracy caused by subject's head motion, physiological oscillations like heartbeats and respiration, inhomogeneities in the static field and differences in the timing of image acquisition [Huettel et al., 2004]. If uncorrected, this variability may reduce or even eliminate the detection power of an experiment. As such, preprocessing is used to remove variability of no interest from the data and to prepare the data for statistical analysis. Some of the most important preprocessing steps are presented below, in the most usual (but not consensual) order of application:

Slice Scan Time Correction Shifts the data of a slice in time to the same time point as when the reference slice was scanned [?]. This changes the data in a way as if the whole volume would have been measured at the same moment in time. The problem of different slice scanning times stems from the fact that a functional volume is usually not covered at once but with a series of successively measured 2D slices. For instance, for a functional volume of 30 slices and a TR of 3 seconds, the data of the last slice is measured almost 3 seconds later than the data of the first slice. Despite the sluggishness of the haemodynamic response, an imprecise specification of time in the order of 3 seconds will lead to a suboptimal statistical analysis, specially in event-related designs.

3D Motion Correction The most damaging problem of fMRI is head motion: assuming that the voxel size is $3 \times 3 \times 3$ mm, if the subject moves his head only by 3 mm, each voxel's time series will contain data from two different parts of the brain that are separated by exactly 3 mm. If those two parts have very different properties, even this small motion can cause very large changes in raw signal over time. Motion correction operates by selecting one functional volume of a run (or a volume from another run of

the same scanning session) as a reference to which all other functional volumes are aligned [?]. Most head motion algorithms describe head movements by six parameters, three translation (displacement) parameters and three rotation parameters.

Temporal filtering Voxel time courses of fMRI data show low-frequency drifts, which are thought to be caused by physiological and physical scanner-related noise. If not considered, these signal drifts can substantially reduce the power of statistical data analysis, as well as invalidate event-related averaging, since this analysis assumes stationary time courses (time courses with a constant signal level). As such, one of the most important preprocessing steps is the removal of low-frequency drifts, which should be performed always, with the use of a high-pass filter. Unfortunately, this preprocessing step is also one of the most dangerous, since condition-related signal changes may also be removed if this filtering is not applied properly.

Spatial Smoothing Low-pass spatial filters are applied to reduce the high-frequency spatial components and blur the images. The most common blurring technique is the application of a Gaussian filter (with the shape of a normal distribution) with a certain Full Width at Half Maximum (FWHM). When a Gaussian spatial filter is applied, it effectively spreads the intensity at each voxel over nearby voxels. If there was no spatial correlation in fMRI data (impossible to predict whether a voxel would be active by whether its neighbours were active) then spatial filtering would reduce Signal-to-Noise Ratio (SNR). But fMRI data has spatial correlation, due both to functional similarity of adjacent brain regions and to blurring introduced by the vascular system and so, with the application of a filter that matches the expected spatial correlation of the data, it is possible to considerably increase SNR with small loss of spatial resolution.

2.3.2 Transformation to Normalised Space

2.3.2.1 Anatomical-Functional Coregistration

The differences in image resolution between a structural (anatomical) and a functional image are clearly evident (figure 2.6). It is of great advantage to use information from structural images to guide normalisation of functional images. This computational process which maps functional images to structural images is called coregistration.

Figure 2.6: Comparison between a functional (left) and structural (right) image.

2.3.2.2 Spatial Normalisation

With coregistration, one subject's brain activity can be anatomically localised. However, the problem of comparing activity across individuals, whether in the same study or across studies remains: there is wide variation in shape, size, orientation, and gyral anatomy of the brain. To make inter-subject comparison possible, each subject's brain must be transformed to have the same size and shape - spatial normalisation.

The transformation for Talairach Space (TAL), originally defined in [Talairach and Tournoux, 1988], is performed in two steps:

AC-PC Transformation The midpoint of the Anterior Commissure (AC) is located first, acting as the origin of the Talairach space. The brain is then rotated around this point so that the Posterior Commissure (PC) appears in the same axial plane as the AC. The connection of AC and PC in the middle of the brain forms the y -axis of the Talairach coordinate system. The x -axis runs from the left to the right hemisphere through the AC. The z -axis runs from the inferior part of the brain to the superior part through the AC. In order to further specify the x and z -axes, a y - z plane is rotated around the y (AC-PC) axis until it separates the left and right hemisphere (mid-sagittal plane). This AC-PC space keeps the original size of the brain intact while providing a common orientation, being an interesting space for individual clinical applications.

TAL Transformation A cubic bounding box in AC-PC space is defined: it should run parallel to the three axes defined in the first step, enclosing the cortex precisely (figure 2.7). This box requires the specification of additional points on the borders of the cerebrum. The mid-sagittal y - z plane separates two sub-cuboids containing the left and right hemisphere. An axial (x - y) plane through the origin separates two sub-cuboids containing the space below and above the AC-PC plane. Two coronal (x - z) planes, one running through AC and one running through PC, separate three sub-cuboids; the first con-

2. Background

tains the anterior portion of the brain anterior to the AC, the second contains the space between AC and PC and the third contains the space posterior to PC. These planes separate 12 sub-cuboids. In a final Talairach transformation step, each of the 12 sub-cuboids is expanded or shrunk linearly to match the size of the corresponding sub-cuboid of the standard Talairach brain.

To reference any point in the brain, x , y , z coordinates are specified in millimetres of Talairach space.

Figure 2.7: Talairach transformation Bounding Box. The cursor marks the origin of the Talairach Space - AC.

In figure 2.8 it is shown the effect of Talairach transformation. The first image on the left is in the Native Space, the middle one is in AC-PC Space and the right one is on Talairach Space.

Figure 2.8: Talairach transformation (Sagittal View). From left to right, the anatomical image is displayed in the Native, AC-PC and Talairach Space.

2.3.3 General Linear Model (GLM)

Hypothesis-driven statistical data analysis is a fundamental tool to identify brain regions that show increased or decreased responses in specific experimental conditions as compared to other conditions. Due to the presence of physiological and physical noise fluctuations, differences observed between conditions might simply due to chance. Because of this, statistical

data analysis is required to explicitly assess the effect of measurement variability (noise fluctuations) on estimated condition differences: if it is very unlikely that an observed effect is solely the result of noise fluctuations, it is accepted that the observed effect reflects a true difference between conditions.

A linear model is given by the following equation:

$$y = a_0 + a_1x_1 + \dots + a_nx_n + e \quad (2.1)$$

where the observed data y is equal to a weighted combination of several model factors x_n plus an additive error term e . The parameter weights a_n indicate how much of each factor contributes to the overall data.

Given the data and a specified set of model factors, the objective is to calculate what combination of weights serves to minimise the error term. Extending equation 2.1 to include a large number of dependent variables, the equation for a General Linear Model (GLM) is obtained:

$$X = G\beta + \epsilon \quad (2.2)$$

where X , the experimental data, is represented as a 2D matrix with n time points and v voxels; G is the design matrix, which specifies the linear model to be evaluated, with m model factors of length n ; β is the parameter matrix containing m rows and v columns, indicating the amplitude of each model factor for each voxel; ϵ is the error term ($n \times v$ matrix).

The GLM uses a cost function (least-squares error) to determine the smallest amount of residual error and so determine the best model for the data.

In figure 2.9 an example design matrix plot with three experimental factors and one constant factor is shown. Black indicates that the factor is predicted to have no effect at those time points, and white indicates that it should have its largest effect.

Each of the factors (columns) should be a prediction on how the haemodynamic activity would change if it is associated with that factor. It is possible to define three types of factor [Huettel et al., 2004]:

1. Experimental factors - model factors associated with specific hypothesis.
 - (a) Covariates - with a continuous range of values, representing the amount of some known quantity (e.g. predicted haemodynamic responses, response time and physiological changes like the respiratory cycle).
 - (b) Indicators - with integral values, representing a qualitative level (e.g. experimental conditions, subject treatment, demographic conditions)

2. Nuisance factors - model factors associated with known non-experimental sources of variability (e.g. artifacts associated with breathing, motion parameters, white-matter signal). These can reduce the amount of residual variation in the error term and improve the validity of the GLM.

Figure 2.9: Example of a Design Matrix Plot, in dark-light colormap. Dark blocks indicate that the factor is predicted to have no effect at those time points, and white blocks indicate that it should have its largest effect.

2.3.4 Region of Interest Analysis

Voxel-wise analysis are valid for research hypothesis related to a particular cognitive process, however, a more targeted approach - a Region of Interest (ROI) analysis - which evaluates statistical tests on a predetermined collection of voxels chosen to reflect *a priori* anatomical distinctions within the brain, is usually preferred.

Intuitively, a ROI approach has several advantages, namely:

1. The number of ROIs is a lot smaller than the number of voxels, reducing the complexity of the statistical analysis.
2. Each ROI combines data from several voxels, and so there will be an increase of the SNR, assuming that the ROI is functionally homogeneous.
3. Solves problems regarding the comparison of data across subjects, since these ROIs can be drawn on a subject-specific basis.

ROIs can be defined based on two major criteria:

Anatomical ROIs Based upon *a priori* expectations about the involvement of certain brain areas in a task. These are drawn on structural images collected at the beginning of an fMRI session, without consultation of functional activity maps, and therefore, they can provide an unbiased estimate of activity within a given brain area.

Functional ROIs Based on the statistical activation map of a certain functional run: the ROIs will include the voxels that were activated by a particular stimulus.

A parameter of major importance is the size of the ROI (the number of voxels it contains). With a smaller number of voxels, the signal comes from a more select set of voxels, but the risk of capturing noise is much higher. With a larger size, the effect of averaging is higher and helps cancel the noise, but it might cancel the dynamic of interest as well.

2.3.5 Multi-Subject Analysis

All fMRI experiments should acquire data from several subjects, in a search for greater validity and robustness of the results. There are two major statistical approaches for multi-subject fMRI studies.

The first analysis involves combining all data points from all subjects into a single analysis, or averaging the time courses across subjects, resulting in a single time course. This approach is known as **fixed-effects analysis**, since it assumes that the experimental effect is constant (fixed) across subjects (fixed-effects models assume that the experimental manipulation has the same effect upon BOLD signal in every subject). The disadvantage of this approach is intuitive: it restricts the statistical inferences to the particular sample of subjects of the study and restricts the sensitivity to extreme results from individual subjects. For instance, if 70% of the subjects were unable to perform the task (no measurable effects), a fixed-effect analysis might return that the difference between conditions of interest is still significant.

The second approach allows to make inferences about the population from which the subjects were drawn (each subject is considered one of many possible subjects who could have participated in the experiment) and assumes that the experimental manipulation could have a different effect on each subject. This approach is called **random-effects analysis** and is divided in two steps: the first one calculates summary statistics for each subject independently in the comparison of interest; the second one tests the distribution of individual subject's statistics for significance. If this second step returns a significant result, then it is possible to conclude that the experimental manipulation would have an effect on the population from which the subjects were drawn. These are, in general, strongly recommended.

2.4 fMRI Neurofeedback Experiments

Is it possible to train participants to control the functioning of specific regions within their brain using fMRI?

2.4.1 Definition

Neurofeedback is a specific form of biofeedback, a process by which a system feeds back information to an individual about its own brain activity, to allow for training of voluntary regulation of brain activity [Weiskopf, 2012].

The clinical goal of this technique is to normalise the abnormal functional activation pattern associated with a clinical condition. One of the main hypothesis is that the functional normalisation may have a therapeutical potential by improving the clinical symptoms. This therapeutic potential is based on the concept of self-directed neuroplasticity, *i.e.* the brain's ability to change itself.

A rt-fMRI system, which processes images as fast as they are acquired, enables this new type of experiment where stimuli adapt online to the participants' measured brain activity (rt-fMRI-NF).

2.4.2 Experimental Setup

A rt-fMRI-NF session usually consists on four types of run: the first is a simple anatomical run, where a structural image of the brain is acquired. This will allow for coregistration and easier visualisation of activated areas. The second is called a localiser, where the participant is performing a specific task and/or seeing a specific stimulus, known to activate the regions of interest. This allows to select the seed region for the feedback. The third is a set of neurofeedback runs, where the participant is performing a specific task and trying to improve its own performance. Immediately before and after these neurofeedback runs, two runs with the same task but without feedback to the participant are performed (called Train and Transfer runs, respectively). These allow to compare activations before and after the session, therefore evaluating the effects of NF.

A typical setup for a rt-fMRI-NF experiment is shown in figure 2.10. During the acquisition of fMRI data, a region of interest (seed ROI) is selected and the time course from this seed is

the basis for the neurofeedback value calculation. After some processing and analysis of this signal, the feedback is presented to the participant at each volume (or time point). The loop is closed when the participant interprets this feedback. The effect should be reflected in the following time points.

Figure 2.10: Real-time fMRI Neurofeedback setup.

Regarding practical implementation, three computers are required for this type of experiment. Figure 2.11 illustrates this setup (adapted from [5]). The MRI console computer is responsible for controlling the MRI scanner and exporting the images to a shared folder in the network. Then, Computer B, running a rt-fMRI software like Turbo-BrainVoyager [4] processes the images in real time and allows the selection of a seed ROI. The signal values are then sent to Computer C, which calculates and presents the feedback to the participant, through a projector mounted inside the MRI room.

Figure 2.11: Practical real-time fMRI Neurofeedback setup.

2.4.3 Neurofeedback Types

There are several types of NF presentation. The most commonly used is a thermometer: the higher the activation or the activation variation in the region selected, the higher the temperature the thermometer displays. An example thermometer is shown in figure 2.12, together with the corresponding time course (baseline blocks in grey, task blocks in green). The current condition can be indicated to the participant using text (as shown) or by an auditory cue.

Figure 2.12: Neurofeedback using a graphical thermometer.

There are also other types of NF presentation that may be better suited, depending on the experiment itself, such as auditory cues (positive and negative-eliciting sounds) and visual cues (colour gradients, avatars).

The formula for calculating the neurofeedback values is also a matter of great variability. The simplest is to feedback directly the BOLD value. This has the intuitive drawback of not showing clearly the difference between baseline and task values.

The most commonly used method relies on the relation between the BOLD values during the task block and the values during the baseline blocks, resulting in a BOLD signal variation. An example formula would be:

$$\text{SignalVar}(\%) = \frac{x(t) - b(t)}{b(t)} \times \frac{100}{\text{maxPSC}} \quad (2.3)$$

where $x(t)$ is the BOLD value at time point t , $b(t)$ is the baseline value for the same time point t and maxPSC is a threshold that defines the maximum percent signal value.

What varies the most in the formula above is the definition of the baseline value $b(t)$. This can be, for instance, the value from the last point in the baseline block preceding the current task block, or the average of all baseline values until the current time point. Additionally, to account for the haemodynamic lag, the classification of time points (in baseline/task) is shifted by a value related to the TR of the acquisition.

Another technique uses the definition of an additional control ROI to feedback the difference of BOLD signal changes between the two ROIs, as such:

$$\text{SignalVar} = (x_{\text{seed}}(t) - b_{\text{seed}}(t)) - (x_{\text{control}}(t) - b_{\text{control}}(t)) \quad (2.4)$$

where $x_{\text{seed}}(t)$ is the BOLD value from the seed ROI and $x_{\text{control}}(t)$ is the BOLD value from the control ROI.

2.5 Brain Connectivity

Can we hear the conversation between regions in the brain?

Imaging neuroscience has firmly established functional segregation as a principle of brain organisation in humans. As cognitive processes depend not only on the activation of a single region in the human brain, but also (or even mainly) on the communication between regions related to those processes, the investigation about the integration of segregated areas has proven to be relevant, although more difficult to assess.

2.5.1 Functional Connectivity

One approach to characterise integration is in terms of functional connectivity, which is usually inferred on the basis of correlations among measurements of neuronal activity [Friston, 2011]. Functional connectivity is defined as statistical dependencies among remote neurophysiological events.

2.5.1.1 Pearson's and Partial Correlation

The Pearson product-moment correlation, commonly referred as Pearson's Correlation (PeC), is a measure of the strength of the linear relationship between two variables. In practice, it

measures how close the observations are to the straight line that best describes the linear relationship between the variables [Petrie and Sabin, 2000].

The PeC coefficient between variable x and y (with n values) is defined as

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.5)$$

where $\bar{x}$ and $\bar{y}$ are the mean values of variables x and y .

Its values lie between -1 and +1, where +1 indicates a perfect positive linear correlation, -1 a perfect negative linear correlation and 0 a null linear correlation. Intuitively, x and y can be interchanged without affecting its value ($r_{xy} = r_{yx}$).

It is used on fMRI data as a measure of functional connectivity between two areas. If two areas show strong correlation when performing a specific task, it is an indication that they might be connected and working together to perform the referred task.

A typical correlation analysis involves the following steps:

1. Definition of seed ROIs.
2. Extraction of seed's time course.
3. Correlate seed's time course with each other and/or with all other voxels in the brain.
4. Convert correlation values to z values using Fisher's r -to- z transformation.

$$z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) = \text{arctanh}(r) \quad (2.6)$$

5. Apply statistical test to retrieve significant connections.

Partial correlation measures the degree of association between two variables, taking into account the effect of third variable. It can be simply derived as a function of simple correlation coefficients:

$$r_{xy \cdot z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \quad (2.7)$$

2.5.1.2 Coherence and Partial Coherence

Coherence measures the linear time-invariant relationship between two time series x and y at frequency λ :

$$\text{Coh}_{xy}(\lambda) = |R_{xy}(\lambda)|^2 = \frac{|f_{xy}(\lambda)|^2}{f_{xx}(\lambda)f_{yy}(\lambda)} \quad (2.8)$$

where $f_{xy}(\lambda)$ is the cross-spectrum of x and y , and $f_{xx}(\lambda)$ is the power spectrum of x [Sun et al., 2004].

Its values lie between 0 and 1, where 1 indicates that x can predict y in a linear fashion perfectly and 0 indicates that x and y have no linear relationship. Intuitively, x and y can be interchanged without affecting its value ($\text{Coh}_{xy}(\lambda) = \text{Coh}_{yx}(\lambda)$).

The assumption is that two regions in the brain which have the same underlying neural activity will be highly coherent, even if they present very different haemodynamic response shapes. However, as the HRF could act as a band-pass filter, coherence may be relevant only within the bandwidth of the haemodynamic response (typically 0 - 0.15 Hz).

A typical coherence analysis involves:

1. ROI selection and specific time course extraction.
2. (Optional) Splitting of the time courses based on the experimental conditions.
3. Estimation of coherence between pairs of ROIs and conditions.
4. Contrasting maps for each ROI/condition.
5. Conversion of coherence values to z values by Fisher's r -to- z transformation (applying the arc-hyperbolic tangent transform).
6. Statistical analysis.

Partial coherence can be used to determine if the relation between two variables x and y is consequence of a common input r (reference function), *i.e.* it estimates the amount of additional improvement in predicting y from x given r :

$$\text{Coh}_{xy \cdot r}(\lambda) = \frac{|R_{xy}(\lambda) - R_{xr}(\lambda)R_{ry}(\lambda)|^2}{(1 - |R_{xr}(\lambda)|^2)(1 - |R_{ry}(\lambda)|^2)} \quad (2.9)$$

A partial coherence value of 1 indicates that x predicts y perfectly, without the contribution of r . An interesting choice for the reference function r is the stimulus presentation paradigm convolved with the HRF, as it allows to assess the connectivity due entirely to the neural response of interest, excluding any direct stimulus-induced response.

2.5.2 Effective Connectivity

Integration within a distributed system is usually better understood in terms of effective connectivity: it refers explicitly to the influence that one neural system exerts over another, either at a synaptic or population level [Friston, 2011]. Effective connectivity is dynamic (activity-dependent) and depends on a model of interactions or coupling.

2.5.2.1 Granger Causality Mapping (GCM)

Granger Causality Mapping (GCM) [Roebroeck et al., 2005] is a method that aims to explore the influences between neuronal populations and their directions. The method does not require prior assumptions of a model, nodes or connections among them. The mathematical formulation relies on Granger causality to establish the existence and direction of influence from information data. In other words, the GCM maps are established based on the temporal precedence of the information, and identify the sets of voxels, ROIs or nodes that are connected (source or target).

Assuming that $x_1(t)$ and $x_2(t)$ are the time courses of two sets of voxels, and according to Granger causality, if x_1 Granger-causes x_2 , $x_1(k)$, $k < t$ contains useful information to predict $x_2(t)$. The direction of the interaction is identified based on the temporal precedence of one time series over the other.

The most common operationalisation of Granger causality utilises Vector Autoregression (VAR) modelling of time series data, which captures the temporal and cross region interactions among multiple regions, with delayed effects of one or multiple lags in a network [Chen et al., 2011]. In a network with two variables, the framework is described by

$$X_1(t) = \sum_{i=1}^p A_{11i} X_1(t-i) + \sum_{j=1}^p A_{12j} X_2(t-j) + \epsilon_1(t) \quad (2.10)$$

$$X_2(t) = \sum_{j=1}^p A_{21j} X_1(t-j) + \sum_{i=1}^p A_{22i} X_2(t-i) + \epsilon_2(t) \quad (2.11)$$

where it is possible to see the dependence of X_1 on the past of X_2 .

The Granger causality from X_2 to X_1 , written $\mathcal{F}_{X_2 \rightarrow X_1}$, stands to quantify the degree to which the past of X_2 helps predict X_1 , over and above the degree to which X_1 is already predicted by its own past. Omitting the dependency on the past of X_2 , X_1 becomes predicted by its own past only:

$$X'_1(t) = \sum_{i=1}^p A'_{11i} X_1(t-i) + \epsilon'_1(t) \quad (2.12)$$

The unconditional Granger causality from X_2 to X_1 is defined as

$$\mathcal{F}_{X_2 \rightarrow X_1} = \ln \frac{|\text{cov}(\epsilon_1)|}{|\text{cov}(\epsilon'_1)|} \quad (2.13)$$

quantifying the reduction in prediction error when the past of the process X_2 is included in the explanatory variables of a VAR model for X_1 .

This unconditional definition ignores the possibility of a joint dependency between X_1 and X_2 and a third set of variables X_3 , which can lead to spurious causalities. These can be eliminated by taking into account the common dependencies. Assuming that the full (equation 2.10) and reduced (equation 2.12) are now

$$X_1(t) = \sum_{i=1}^p A_{11i} X_1(t-i) + \sum_{j=1}^p A_{12j} X_2(t-j) + \sum_{k=1}^p A_{13k} X_3(t-k) + \epsilon_1(t) \quad (2.14)$$

$$X'_1(t) = \sum_{i=1}^p A'_{11i} X_1(t-i) + \sum_{k=1}^p A'_{13k} X_3(t-k) + \epsilon'_1(t) \quad (2.15)$$

Granger causality from X_2 to X_1 knowing X_3 , $\mathcal{F}_{X_2 \rightarrow X_1 | X_3}$, is defined as in equation 2.13 and quantifies the degree to which the past of X_2 helps predict X_1 over and above the degree to which X_1 is already predicted by its own past and the past of X_3 .

Two measures based on Granger causality, Causal Density (CD) and Causal Flow (CF), are commonly used to assess connectivity between brain areas.

Causal Density is a measure of the overall causal structure of a system. Considering a set of time series ($n > 2$), it is possible to determine the pairwise Granger causality between time series and determine the causal density as the average of all pairwise Granger causalities between elements. Mathematically, the CD of a time series X is

$$CD(X) = \frac{1}{n(n-1)} \sum_{i \neq j} \mathcal{F}_{X_i \rightarrow X_j | X_{ij}} \quad (2.16)$$

where X_{ij} is the subsystem of X with the variables i and j omitted and n is the total number of connections. Nodes completely disconnected from the network will present low scores of CD, while high values will only be achieved by nodes that present potential predictive information.

Causal Flow represents a measure of the interaction of a specific node with the rest of the network. A node in a causal network can be characterised by the difference between its inputs and outputs. One can interpret a node with a highly positive CF as a node that exerts a strong causal influence on the network.

2.5.2.2 Dynamic Causal Modelling (DCM)

Dynamic Causal Modelling is a model-driven mathematical framework which aims to estimate and make inferences about the coupling between brain areas, based on the variations imposed by the experimental changes (*i.e.* experimental stimuli) [Friston et al., 2003].

DCM constructs a causal model of interacting cortical regions or nodes. Neuronal activity in a given region causes changes in neuronal activity in interconnected regions. The activity of each node (and of the connections between nodes) is also directly modulated by input variables, *i.e.* changes in the experimental conditions. The model is supplemented by a forward model that provides posterior estimates of neurobiologically interpretable quantities (*e.g.* effective connectivity) from the measurements of brain activity.

The brain activity responses are evoked by known experimental stimuli that elicits changes in sensory or cognitive sets. This can be modelled by using a dynamic input-state-output model: the inputs are functions that encode experimental manipulations; the state variables correspond to both the neuronal activities and other neurophysiological or biophysical variables; the outputs are measured haemodynamic responses of the ROIs.

DCM assumes that the responses are driven by changes in external inputs, which may interact with the model in one of two ways. Firstly, inputs can trigger changes in the neuronal activity of the nodes directly (*e.g.* sensory inputs may cause direct responses from sensory areas). Secondly, they can affect the system through the modulation of the coupling between nodes (*i.e.* effective connectivity).

The mathematical formulation is based on ordinary differential equations. Considering f a linear non function, u the external stimuli/input, x the activity/neuronal states of the brain regions and θ the model parameters, the formulation is

$$\dot{x} = f(x, u, \theta) = \left(A + \sum_{j=1}^m u_j B^{(j)} \right) x + Cu \quad (2.17)$$

where

$$\dot{x} = \frac{dx}{dt} \quad A = \frac{\partial f}{\partial x} \Big|_{u=0} \quad B = \frac{\partial^2 f}{\partial x \partial u} \quad C = \frac{\partial f}{\partial u} \Big|_{x=0}$$

These three parameter matrices describe the nature of the three causal components which underlie the modelled neural dynamics: Matrix A represents the context-independent connectivity among brain regions, mediated by anatomical connections; Matrix B represents the context-dependent changes in effective connectivity induced by the j th input u_j ; Matrix C represents the direct inputs into the system that drive regional activity. The posterior distributions of these parameters can provide information about the impact of different mechanisms on determining the dynamics of the model.

Parameter estimation uses standard approaches to system identification based on Bayesian inference. A more detailed mathematical explanation of DCM models, namely about its estimation, can be consulted in [Friston et al., 2003].

2.5.2.3 Structural Equation Modelling (SEM)

Structural Equation Modelling (SEM) is a widely used technique in behavioural sciences and a standard for analysing effective connectivity in fMRI. McIntosh and Gonzalez-Lima were the first to introduce this method in a neuroimaging study in 1994 [McIntosh and Gonzalez-Lima, 1998], and despite the increasing application of dynamic methods such as DCM, it is still in common use for path analysis and modelling of interactions among covariant brain areas.

In a simplistic way, the analysis of effective connectivity involves looking at possible causal influences between brain regions, given measurements of the activity of those regions. SEM is a model-driven algorithm that can use hypothesis derived from *a priori* knowledge (anatomical or neurophysiological models) to draw the path strength of a direction, either directly between two regions or indirectly from one region to another through other regions.

A generic SEM model, which assumes no lagged effects (*i.e.* $X(t)$ does not depend on previous time points $X(t - a)$) can be represented by:

$$X(t) = c + A_0 X(t) + \epsilon(t) \quad (2.18)$$

where $\epsilon(t)$ is the residuals vector, assumed to follow a multivariate Gaussian distribution, c is a constant for the baseline and A_0 is the instantaneous path matrix

$$A_0 = \begin{pmatrix} 0 & \alpha_{120} & \dots & \alpha_{1n0} \\ \alpha_{210} & 0 & \dots & \alpha_{2n0} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n10} & \alpha_{n20} & \dots & 0 \end{pmatrix} \quad (2.19)$$

with components α_{ij0} specifying all the possible contemporaneous effects among the n regions [Chen et al., 2011].

Knowing that the baseline satisfies $c = (I - A_0)E[X(t)]$ and replacing $E[X(t)]$ by the vector of sample means, the model can be reduced to:

$$X(t) - \bar{X} = A_0(X(t) - \bar{X}) + \epsilon(t) \quad (2.20)$$

The covariance structure of this model Σ among n regions can be predicted by

$$\Sigma = (I - A_0)^{-1} \epsilon \epsilon^T (I - A_0)^{-T} \quad (2.21)$$

where $\Psi = \epsilon \epsilon^T$ is the covariance matrix of the residuals that model the unaccounted vari-

ability in the latent variables. By comparing Σ with the covariance matrix estimated from the data, one can use the Maximum Likelihood approach to minimise the discrepancy between the two covariance structures and obtain the estimate of the path matrix A_0 .

The application of SEM to fMRI data usually follows these steps [Storkey et al., 2007]:

1. ROI selection based on *a priori* knowledge of the brain regions of importance.
2. Definition of a connectivity model between these regions: the model should state which regions are connected and the direction of this connectivity.
3. Definition of a structural equation model: its parameters are estimated using maximum likelihood methods.
4. Comparison of connection parameters across subject classes.

2.5.2.4 Structural Vector Autoregression (SVAR)

A very important difference between VAR models (Granger causality) and SEM is that, while the former captures the effects of the past of the regions in the network on the current state, the latter focuses on the instantaneous effects of neural regions. Each of them can identify some important features of the network, but they fail to identify others. For instance, Granger causality ignores the instantaneous correlation among regions of the network, while SEM assumes that all neural interactions are instantaneous and no lagged correlation is considered within and across regions.

A more realistic approach would include both instantaneous and lagged interactions: Structural Vector Autoregression (SVAR) can model both instantaneous (contemporaneous) and lagged effects among network regions using a unified analytical framework. While no interactions within the brain are truly instantaneous, the inclusion of the instantaneous effect terms makes possible to model interactions with delay times much shorter than the lag time set by the temporal resolution of fMRI [Zotey et al., 2011].

A multivariate SVAR model of order p for a network of n ROIs is defined as:

$$X(t) = A_0X(t) + A_1X(t-1) + \dots + A_pX(t-p) + c_1z_1(t) + \dots + c_mz_m(t) + \epsilon(t) \quad (2.22)$$

where $X(t)$ is the vector containing fMRI values for n ROIs at time point t and $X(t-1)$ is the vector for the same ROIs at the preceding time $t-1$. Matrices A_0 and A_1 contain path coefficients for different pairs of ROIs. A_0 describes the contemporaneous effects as in the SEM equation while $A_1, \dots, A_p$ are the lagged effects as in a VAR model. While the general structure of A_0 needs to be defined *a priori*, based on an hypothesis about the instantaneous

effects of the network, the definition of A_1 is data-driven. Functions z_m are exogenous variables which are independent of the interactions within the network, modelled by the vectors c_m . $\epsilon(t)$ is the vector of residuals, assumed to be Gaussian distributed.

2.6 Case Study - Working Memory Network

Self-modulation of executive functions with real-time fMRI.

Working memory involves the temporary storage and manipulation of information that is assumed to be necessary for a wide range of complex cognitive activities such as reasoning, comprehension and learning [Baddeley, 2003].

According to a multi-component model proposed by [Baddeley and Hitch, 1974], working memory could be divided into an attention control system - the central executive - and two modality-based temporary storage systems - the phonological loop and the visuospatial sketch pad. Among the three systems, the central executive system controls attention and information flow to and from the phonological loop and visuospatial sketch pad, which separately buffer the verbal-acoustic and visual material. With respect to the central executive function, accumulating evidence showed that the Dorsolateral Prefrontal Cortex (DLPFC) played a crucial role in a variety of executive control processes [Zhang et al., 2013].

This case study main proposal is to develop a rt-fMRI-NF protocol for executive functions training. The hypothesis is that rt-fMRI-NF has therapeutic potential in several neuropsychiatric disorders with executive functions deficits, recognised by having high impact in daily-life.

3

State of the Art

This chapter presents an overview of current literature on the topics studied in this work. Firstly, the format of the review is explained, and then a number of papers measuring functional and effective connectivity on fMRI are presented.

3.1 Review Structure

The state of the art was performed as a Systematic Review, following the PRISMA Statement guidelines [Moher et al., 2009]. The objective of this review is to analyse papers which study connectivity on fMRI experiments and perform neurofeedback based on the activity of a certain region of the brain. The search criteria and methodology are presented below.

Searches were performed on Pubmed [7] with the keywords "fMRI", "neurofeedback" and "connectivity", finding 37 papers published from September 2011 up to October 2015. The last search update was made on November 20th, 2015. The title, objective, subject of study and eligibility verdict of these 37 papers was recorded on a spreadsheet.

Published papers in English-only were initially filtered by reading the abstract to identify which papers presented original data (so to exclude reviews) and which actually computed connectivity by any method (excluding references of connectivity in future work).

Next, a full-text reading excluded papers that performed neurofeedback of the connectivity instead of the BOLD signal, and also those that did not compare brain connectivity (either functional or effective) before and after a NF session.

The final selection includes 18 studies that fulfil all the requirements. For these, the fields Title, Subject of Study, Objective, ROIs, NF-Target ROI, Methods for Connectivity, Software, Metrics, Results, Number of Subjects, TR, Authors and Year were added to the spreadsheet.

Based on the field Methods for Connectivity, the studies were divided into 5 groups (figure 3.1).

The summary of the review process is presented in figure 3.2.

Figure 3.1: Distribution of the selected articles per method.

Figure 3.2: Flow of information throughout the different phases of the systematic review.

3.2 Functional Connectivity

3.2.1 Pearson's Correlation

In the following section, ten research papers that use Pearson's Correlation (PeC) to assess functional connectivity in NF experiments will be presented, separated by topic of research.

Resting State

Four studies [Harmelech et al., 2013, Scheinost et al., 2013, Yuan et al., 2014, Karch et al., 2015] evaluated post-neurofeedback changes in brain connectivity during rest.

[Harmelech et al., 2013] states that during rest, the cerebral cortex displays rich and coordinated patterns of spontaneous activity. The mechanism that shapes these patterns is largely unknown, but the authors demonstrate that an Hebbian-like sustained process plays a role in focusing these coherent patterns. To localise the region of interest, the Dorsal Anterior Cingulate Cortex (DACC), subjects performed two localiser tasks involving self-initiation and volitional behaviour. Both tasks have proven to activate this region. In the experiment, the left DACC was defined as the region of overlap ($r=15$ mm) between the self-initiated action and the self-initiated thought task activations, for each subject. It was used as seed for the NF sessions and for connectivity analysis. [Scheinost et al., 2013] and [Karch et al., 2015] used a similar localiser procedure, showing alternately neutral and interest images. In the former, the images were contamination-related and were used to target the Orbito-Frontal Cortex (OFC), which was defined as the 30 OFC voxels, bilaterally, that were maximally differentially activated by contamination-related stimuli. In the latter, the authors showed alcohol-relevant pictures (triggers for alcohol consumption). The objective was to define regions of interest for each patient that had the most extensive addiction-related BOLD response. Note that for connectivity purposes, the region defined by [Scheinost et al., 2013] is not used (the authors perform an unbiased analysis without *a priori* assumptions regarding ROIs). [Yuan et al., 2014] selected three spherical ROIs (7 mm radius) based on the stereotaxic array of Talairach and Tournoux [Talairach and Tournoux, 1988]: for the control group, on the Horizontal segment of Intraparietal Sulcus (HIPS), and for the active group, on the left Amygdala (AMG) and on the Pregenua Anterior Cingulate Cortex (pgACC).

Rest scans were included in all experiments, but with different timings. [Harmelech et al., 2013] performed three rest scans: one before, one immediately after and one 24h after the NF session. [Yuan et al., 2014] performed two rest scans, one before and one at least 2 days after

the NF session. [Karch et al., 2015] performed two rest scans, one before and one after the NF training. [Scheinost et al., 2013] uses a more innovative and extensive training, consisting on four MRI scanning sessions, each spaced by a few days. In the first and the last session, two resting state runs are performed.

In table 3.1, a comparison of the preprocessing steps performed by each paper is presented.

Table 3.1: Comparison of preprocessing steps on Resting State papers.

	[Harmelech et al., 2013]	[Scheinost et al., 2013]	[Yuan et al., 2014]	[Karch et al., 2015]
Exclude first volumes	0	0	5	5
Slice scan time correction	yes	yes	yes	yes
3D motion correction	yes	yes	yes	yes
Linear trend removal	yes	no	yes	yes
Temporal filtering	Bandpass 0.009 - 0.08 Hz	Lowpass 0.12 Hz	Bandpass 0.005 - 0.1 Hz	Bandpass 0.01 - 0.1 Hz
Spatial smoothing	Gaussian FWHM=4 mm	no	Gaussian FWHM=6 mm	Gaussian FWHM=8 mm

Regarding correlation analysis, [Scheinost et al., 2013] calculates a measure called *degree* for each voxel in two steps: firstly, the time course of a voxel is correlated with every other voxel in the grey matter; secondly, the degree for that voxel is defined as the number of connections with a correlation greater than 0.4. This process is repeated for every voxel. [Harmelech et al., 2013] also performed voxel-wise analysis: for each voxel, the time course was extracted, transformed to percent signal change and correlated with the time courses from all other voxels in the cortex. [Yuan et al., 2014] and [Karch et al., 2015] extract the time course from the regions of interest, average the signal within them, and then compute the correlation coefficient between these signals and the time course from all other brain voxels.

The four papers compare connectivity measures before and after NF sessions: the first step is to convert correlation values to z values using Fisher's r -to- z transformation (equation 2.6). An additional step, clearly stated in [Yuan et al., 2014], is to convert these values to z scores, by subtracting the mean and dividing by the standard deviation (equation ??). Next, for statistical analysis, t-tests are used to compare the first and last rest run: [Yuan et al., 2014] and [Karch et al., 2015] perform a paired one-sided t-test; [Scheinost et al., 2013] performs a Wilcoxon's signed-rank test; [Harmelech et al., 2013] performs a two-tailed paired t-test.

All the papers observe statistically significant differences between rest runs: In [Scheinost et al., 2013] the group map of changes in degree of connectivity showed significant decreases in brain regions associated with emotion processing (Insula, Hippocampus, among others) and significant increases in prefrontal areas associated with emotion regulation and cognitive control (right Lateral Prefrontal Cortex (LPFC) and bilateral portions of Brodmann Area (BA)

8). In [Yuan et al., 2014] it is shown that the abnormal hypo-connectivity in connections with the left AMG in depression patients was reversed, suggesting a therapeutic effect of rt-fMRI-NF using recall of happy autobiographical memories. [Harmelech et al., 2013] see a decrease in functional connectivity in regions that were inactivated during the NF sessions as opposed to an increase in the activated regions. Also, a significant relationship between the level of co-activation during the NF training and connectivity increment (or decrement) in the resting-state connectivity was found, specially on the last session (1 day after). In [Karch et al., 2015] patients demonstrated increased connections between the left middle frontal area (BA 46) and the DLPFC, lentiform nucleus, thalamus and the parahippocampal gyrus. In the other hand, connections between the right Anterior Cingulate Cortex (ACC), the left and the right DLPFC (BA 9) did not change in consequence of the rt-fMRI paradigm. The authors also suggest that the changes in connectivity were due specifically to the feedback and not to non-specific aspects of the task, based on the comparison with a control group of subjects (this group did not receive feedback from the NF target ROI - "sham feedback").

Emotion Processing

[Zotev et al., 2011] and [Sarkheil et al., 2015] investigate the possibility of training subjects to self-regulate the activity of key areas for emotion processing, as the AMG and the LPFC.

[Sarkheil et al., 2015] used a functional localiser to identify a region in the left LPFC, using a sequence of emotional eliciting images. The region was made cubic, centred in the peak of activation in the left LPFC. Additionally, a region with 5 mm in the AMG was defined, based on Talairach coordinates reported in literature. [Zotev et al., 2011] defined three 5 mm spherical ROIs based on quantitative meta-analyses of functional neuroimaging studies which investigate the role of the left and right AMG in emotion processing and the role of HIPS in number processing. The neurofeedback signal was based on the activation in the left AMG for the participants in the experimental group and on the activation in the left HIPS for the subjects in the control group.

Preprocessing steps performed by both papers are shown in table 3.2.

Functional connectivity in [Sarkheil et al., 2015] was calculated based on the correlation between the individual's LPFC signal and all other brain voxels. The time course of the signal from the LPFC served as a predictor in the GLM to estimate the beta values in each brain voxel. The individually calculated maps of beta estimates were then submitted to a two-sample t-test to reveal group differences.

Table 3.2: Comparison of preprocessing steps on Emotion Processing papers.

	[Zotev et al., 2011]	[Sarkheil et al., 2015]
Exclude first volumes	0	0
Slice scan time correction	yes	yes
3D motion correction	yes	yes
Linear trend removal	no	yes
Temporal filtering	Lowpass 0.08 Hz	Highpass 5 sin/cos cycles
Spatial smoothing	Gaussian FWHM=5 mm	Gaussian FWHM=6 mm

This test revealed clusters with increased connectivity level with the left LPFC in the feed-back group including the right LPFC and Posterior Cingulate Gyrus (PCG). The authors state that this implies an effect of feedback modulation. Analysing the PeC between beta values from the left LPFC and right AMG revealed a significant negative correlation between them. The correlation with the left AMG, however, did not reach statistical significance.

In [Zotev et al., 2011] a GLM-based functional connectivity analysis was also performed, using as seed the left AMG: the time course of the mean fMRI signal from the ROI was used as a stimulus regressor. The GLM-based R-squared statistics were converted to correlation coefficient values r and the resulting correlation maps for each run were transformed to the Talairach space. These correlation coefficient values were converted to z values using Fisher's r -to- z transformation. Group t-test with respect to zero level was employed to determine the functional connectivity pattern for each run. Correction for multiple comparisons was based on False Discovery Rate (FDR).

To identify the regions for which the functional connectivity with the left AMG increased during the experiment, mean values of the correlation coefficients for several spherical ROIs (of 5 mm radius) were determined for each run. The ROIs were centred at locations that were characterised by peak t-values in the statistical group connectivity analysis of the experimental group for the Transfer run. For each of these ROIs, a one-way Analysis of Variance (ANOVA) for repeated measures on mean correlation coefficients was applied.

For the experimental group, six regions within the AMG network demonstrated a significant increase in functional connectivity over the course of the neurofeedback training. No significant changes in functional connectivity were observed for the same six regions for the control group.

Working Memory

[Zhang et al., 2015] aimed to investigate the dynamic functional connectivity between three networks engaged in working memory training during rt-fMRI neurofeedback. These networks are the Central Executive Network (CEN), the Default Mode Network (DMN) and the Salience Network (SN). The study starts with a localiser run in which the subject's target ROI in the left DLPFC is selected, and is followed by four feedback runs. Preprocessing steps included slice time scan correction, 3D motion correction and spatial smoothing with a Gaussian filter of FWHM = 8 mm.

Independent Component Analysis (ICA) was performed on the rt-fMRI data of all feedback runs of all subjects in the experimental and control group to identify the spatially independent networks referred. For each selected Independent Component (IC), a two-way repeated-measure ANOVA was conducted and corrected using FDR at an overall $\alpha < 0.001$. Based on this analysis at the group level, the ROIs of the three networks were defined as spherical regions (radius of 6 mm) centred on the local maximum peak. This procedure returned eleven key brain areas, including the bilateral DLPFC, bilateral Posterior Parietal Cortex (PPC), bilateral Insula, ACC, Ventral Medial Prefrontal Cortex (VMPFC) and bilateral Medial Temporal Gyrus (MTG)/Angular Gyrus (AG).

The mean time course of each ROI of every subject was extracted, and the PeC coefficient between each ROI pair within and between the three networks was calculated and transformed to z values using Fisher's r -to- z transformation. Then, a one-way repeated-measure ANOVA was performed on these z values.

Using a paired t-test, results showed that in the experimental group the functional connectivity of the ROI pairs increased significantly in the last training run compared to the first training run. Specifically, the functional connectivity between the SN and CEN, between the SN and DMN, and within the CEN changed significantly as the training progressed.

Motor Execution and Imagery

[Hui et al., 2014] investigated changes in the motor network engaged in both Motor Execution (ME) and Motor Imagery (MI) task induced by real-time neurofeedback training of the right Premotor Cortex (PMA).

The ROIs were manually selected: one on the right dorsal PMA and another on the background, *i.e.* a region that was not being activated by the task. Unusually, the feedback presen-

ted to the subject was calculated based on the signal from these two ROIs. For offline analysis, the dorsal PMA ROI was defined as a 6 mm radius sphere around the *maxima* on the GLM t-contrast map.

Preprocessing included slice scan time correction, 3D motion correction, re-slicing and spatial smoothing with a Gaussian kernel of FWHM = 8 mm.

Correlation analysis was performed on nodes from the task-related network, identified by ICA. A conjunction analysis revealed eight brain regions engaged in the motor network. The mean time course was extracted from each ROI for each pre and post-training run of each subject. PeC coefficients between the time courses of each pair of ROIs were calculated and converted into z values using Fisher's r -to- z transformation. Statistical analysis was performed with a two-way repeated measures ANOVA, using training (pre-training and post-training; within-subject) and group (experimental and control; between-subjects) as main factors. For functional connections that showed significant interaction effects, tests of simple effect were further carried out and corrected by Family-wise Error (FWE) method.

The results show that the participation of brain areas in the motor network as well as the functional connections among nodes within the motor network were changed by the real-time neurofeedback training. Specifically, the activity of the target ROI during neurofeedback training showed significant relation with the functional connectivity between the target ROI and Posterior Parietal Lobe (PPL) during the pre-/post-training MI task.

Auditory Perception

[Haller et al., 2013] trained participants to voluntarily regulate activity in the auditory cortex. The authors used a functional localiser to identify each subject's primary auditory cortices, which were used as target regions for the NF runs. The acquired images were corrected for slice scan time differences, spatially realigned and smoothed with a Gaussian kernel of FWHM = 4 mm and high pass filtered at 1/128 Hz. Data for functional connectivity analysis was obtained by ICA: PeC coefficients were calculated for all pairwise combination of ICs' associated time courses. For the time courses of selected ICs, condition-specific time courses were generated (by concatenating blocks of the same condition). For the five functional network connections that showed learning effects, the correlation between the ICs' condition-specific time courses was computed, which resulted in 24 correlation coefficients for training and 36 correlation coefficients for transfer. Fisher's r -to- z transformation was applied, followed by a) paired t-test to confirm the changes in functional network connectivity between baseline and regulation during training; b) paired t-test between change in connectivity baseline-regulation

during training versus transfer, and c) paired t-tests between the different conditions during transfer. The confidence level of these tests was set to $p < 0.05$ (Bonferroni corrected for the number of functional network connections).

Increases in functional network connectivity were found during the neurofeedback training phase between the auditory target region and the auditory pathway, visual regions (related to visual feedback processing), Insula (related to introspection and self-regulation) and working memory and high-level visual attention areas (related to cognitive effort). The auditory target region was identified as the hub of the reconfigured functional networks without *a priori* assumptions. During the transfer phase, specific functional connectivity reconfiguration between auditory and attention network was found, confirming the specific effect of self-regulation on functional connectivity.

Neuropsychiatric disorders

[Berman et al., 2013] investigated the ability of healthy volunteers to modulate brain activity within their right Anterior Insular Cortex, during a rt-fMRI session with a blink suppression task. The goal was to develop and establish a rt-fMRI-based neurofeedback training methodology that could be used for future investigation as a therapeutic intervention in neuropsychiatric conditions associated with disordered suppression, where a role for the insular cortex has been implicated.

During a functional localiser run, a 25 voxel square region of interest in the axial plane was initially defined, using anatomical landmarks, in the area of the Right Anterior Insular Cortex and then adjusted to lead to a maximum amount of BOLD signal during the blink suppression blocks. A reference ROI encompassing the entire brain volume in an axial slice distant to the interest ROI was defined and included in the feedback calculation.

fMRI data was preprocessed by 3D motion correction, excluding the first 10 volumes of each run and spatial smoothing with Gaussian kernel with FWHM = 8 mm.

For each run and each subject, the time series of the functionally localised ROI was used as seed and correlated with each voxel in the brain. Correlation values were converted to z values and group-level connectivity maps for each of the neurofeedback runs were generated. The maps were clustered and statistics of each cluster were extracted and tested for changes using a repeated measures one-way ANOVA and paired t-tests.

At the group-level, increasing brain activity was achieved within the right Insular Cortex during feedback training. Neurofeedback training resulted in a significant increase in functional connectivity detected between the target ROI and the Medial Frontal Gyrus (MFG), by

comparing the first and last functional run: it increases across successive training runs, but disappears for the transfer run.

3.3 Effective Connectivity

3.3.1 Granger Causality Mapping (GCM)

Four studies [Lee et al., 2012, Ruiz et al., 2013, Sitaram et al., 2014, Kadosh et al., 2015] investigated the modulation of effective connectivity derived from rt-fMRI-NF training using GCM. Each study proposed a specific approach to analyse effective connectivity in terms of measures (Granger causality index, CD and CF), ROI selection (based on atlases, GLM and/or literature), functional runs used in the analysis (first vs. last or weakest vs. strongest).

[Lee et al., 2012] defined an experiment to explore the efficacy of rt-fMRI-NF on the modulation of auditory- and attention-related networks. Additionally, the authors investigated if rt-fMRI-NF training reinforced the causality within attention networks. The study comprised two groups, a control group (with "sham feedback") and an experimental group (the feedback was computed based on the the BOLD activity from the ROIs). First, a pre rt-fMRI session (three scans that focused the primary and secondary auditory cortex) was used to determine the NF target associated to auditory attention-related networks. In the rt-fMRI training run, the participants were instructed to self-regulate the activity in the pre-selected ROIs. Post training runs were performed to evaluate the level of neuronal activation in these brain areas. The ROIs used for connectivity analysis were defined according to the changes in the activation patterns in the pre rt-fMRI sessions, *i.e.* based on statistically different spatial patterns activations comparing pre and post rt-fMRI-NF training.

In the experimental group, nine regions were selected (six that showed increased activation and three displaying decrease in the BOLD activity) for effective connectivity analysis. The results indicate that the connectivity strengthened for some connections (for instance, the causality from the right Cingulate Gyrus (CG) to the bilateral MFG, left CG, and left Precentral Gyrus (PrecentralG) was significantly reinforced) and others weakened after the neurofeedback training. Additionally, the authors performed a whole-brain causality analysis (to this end, the authors considered the Automated Anatomical Labelling (AAL) to determine 90 regions). The results show enhanced causality levels in the experimental group in the last scan of the rt-fMRI-NF session, in agreement with the results based on the ROIs networks analysed.

[Ruiz et al., 2013] evaluated the ability of schizophrenia patients to self-control the BOLD activity of the Anterior Insula Cortex and investigated whether the self regulation ability is correlated with clinical measures and symptom severity. Additionally, the authors studied changes in connectivity associated with neurofeedback training. The feedback was determined based on the difference between two ROIs. ROI₁ was based on a functional localiser and focused on the right and left Anterior Insular Cortex (the task included imagery of emotionally relevant experiences). A reference region, ROI₂, was selected from a background area (e.g. motor cortex and parietal areas). Feedback was presented during up-regulation blocks in the form of a graphical thermometer. During baseline blocks, no feedback was presented. Effective connectivity was assessed based on CD of the elements considered in the network involved in self regulation. The authors considered two different sources to create the network. On one hand, brain areas that have been consistently reported in the literature as part of the emotional network. On the other hand, ROIs extracted from the group GLM of the neurofeedback runs. A total of eight ROIs were chosen to complete the network. To determine variations in the connectivity, the CD was compared between the weakest regulation session and the strongest regulation session, of the first and last day of the training respectively.

Generally, the authors found a larger CD in the strongest session of the regulation compared to the weakest. An increase of CD among the Insula Cortex, AMG and Medial Prefrontal Cortex (MPFC) is also reported. Additionally, new connections appeared among the ACC, Pre-cuneus, Supramarginal Gyrus and the rest of the ROIs.

[Sitaram et al., 2014] presented a pilot study that aimed to explore the ability to learn to regulate the brain activity of the left Anterior Insula using rt-fMRI-NF. The changes in functional connectivity of the network involved were also studied and reported. The authors had previously found that healthy volunteers successfully learnt to control BOLD signal in the target area. The experimental group consisted of four criminal psychopaths that were trained to regulate the Anterior Insula by employing negative emotional feelings from previous episodes. The protocol considered four neurofeedback runs per day for one to three days depending on the availability of the subjects. The NF target ROI₁ (in the left Anterior Insula), was functionally localised based on a localiser run during which the subjects were instructed to recall emotionally relevant personal experiences (a counting backwards strategy was used as contrast condition). The region consisted of a single-slice, rectangular area, of 5×5 voxels. A second region ROI₂ was selected from a background area (e.g. motor cortical surface) and used to cancel global effects and to minimise unspecific activations. The analysis of effective connectivity was performed on one subject only, based on Granger causality framework. To this end, the authors evaluated the network dynamics during three different runs (one of "weak regulation", a second of "intermediate", and a third of "strongest regulation"). CD and CF

were compared among runs. The nodes of the network were defined according to the literature describing the networks involved in self-induced and stimulus-induced emotions. Eight to ten regions were selected based on this principle and on the statistical map produced during the neurofeedback runs (clusters threshold of 50 voxels).

The CD increased in the brain network involved in up-regulation of the BOLD signal on the left Anterior Insula. The authors also mention the high correlation between the percent BOLD increase and the CD from the early stage to the final stage of the training. The CF results show that there is a significant modification of the structure of the connectivity model - in the early training stage, the up-regulation is driven from the Posterior Cingulate (causal source). With training, the sources shifted to the anterior part of the brain finally setting in the left Anterior Insula during the "strongest regulation" run.

[Kadosh et al., 2015] analysed rt-fMRI-NF training of the emotion regulation network in adolescence. The study included 17 adolescent subjects that were instructed to up-regulate the bilateral Insula, one of the key regions in the emotion regulation network. The NF target region was selected based on a localiser run that consisted of a stimulus set of nine coloured photographs of female faces (three emotional expressions - neutral, fearful and happy). The region was delineated in the right Anterior Insula. The authors also analysed effective connectivity, considering the Granger causality framework. The Granger Causality was estimated for each ROI pair individually and for each block and then averaged across blocks and runs. To this end, eight 10 mm ROIs were selected according to group local *maxima* in the localiser run, as well as two 10 mm personally-tailored ROIs based on the peak in the bilateral Insula. The authors aimed at identifying the differences between the two conditions, up- and down-regulation, in terms of the Granger information flow.

The results show significant variations between conditions, increased during the up-regulation condition: the direction from the left AMG node to the right Insula node and from the left Pre-centralG to the left MFG. The results sustain the hypothesis that NF training had a differential effect on the Granger information flow within the emotion regulation network.

3.3.2 Dynamic Causal Modelling (DCM)

Two studies [Scharnowski et al., 2014, Paret et al., 2015] have analysed the changes in connectivity across different areas of the brain before and after rt-fMRI-NF training by means of DCM. As a preliminary step, both performed a Psychophysiological Interaction (PPI) analysis, and identified appropriate ROIs (or nodes) to construct the DCM model. Both studies report an increase in connectivity as a result of the rt-fMRI-NF training.

[Scharnowski et al., 2014] demonstrated that training participants to up-regulate their activity in the visual cortex is associated with an increase of effective connectivity between the visual areas and the Superior Parietal Lobe (SPL). Moreover, the participants that learnt this skill showed increased top-down control of the SPL over the visual cortex (cognition influences sensory data). To determine the nodes of the DCM model, the authors used previous exploratory PPI analysis. According to this rule, two nodes were selected: the trained visual ROI, *i.e.* neurofeedback target, and the contra lateral SPL to the visual ROI. The comparison between the DCM analysis of the last neurofeedback run and the first showed that the fully connected model is the architecture that best fits the data. Subsequently, the model parameters were analysed and a systematic increase in top-down connectivity strength from contralateral SPL onto the visual ROI was found (only found in learners, *i.e.* participants able to up-regulate activity in the NF target region). The authors also report a decrease in the bottom-up connectivity from the visual ROI onto the contralateral SPL.

[Paret et al., 2015] studied the effect of rt-fMRI-NF on the connectivity of the down-regulation of the activity in the AMG. The first part of the study focused on a PPI analysis to identify prefrontal regions interconnected with the AMG. To assess the causal information flow and further evaluate the functional structure, the authors used a DCM model. Two nodes were selected, the AMG neurofeedback target and the VMPFC. The fully connected model was used in accordance with the literature, *i.e.* the authors assumed bidirectional intrinsic connectivity between the nodes, while the input enters the system at the AMG, VMPFC or both. The condition was assumed to modulate the connectivity in the top-down, bottom-up or both directions. The results highlighted a model characterised by network input through VMPFC and modulation of connectivity from the AMG to the VMPFC by the regulation condition.

3.3.3 Structural Equation Modelling (SEM)

[Shen et al., 2015] use SEM to calculate the interactions between regions of the working memory network during rt-fMRI-NF training.

To select ROIs in the working memory network, a functional localiser run was performed, consisting in a 3-back task. Online, it was used to locate the target region left DLPFC for the subsequent neurofeedback runs. Offline, with a group GLM, it was possible to define six ROIs in the local activation maximum of the bilateral DLPFC, bilateral Inferior Parietal Lobe (IPL), ACC and left Inferior Frontal Gyrus (IFG). Preprocessing steps included slice scan time correction, 3D motion correction, normalisation with resampling and spatial smoothing with a Gaussian kernel of FWHM = 8 mm.

The method used to build the SEM model was based on [McIntosh and Gonzalez-Lima, 1994]: the initial model is derived from *a priori* knowledge of the functional connections of the Baddeley model [Baddeley and Hitch, 1974] that corresponds to the central executive system (the ACC and bilateral DLPFC) and the phonological loop (the bilateral IPL and left IFG).

For each run in each group, the extracted time courses of each ROI were averaged across subjects, preprocessed by removing the signal outliers (that exceeded two standard deviations) and normalised. The path coefficients were estimated using a maximum likelihood algorithm that minimised the difference between the observed and expected variance-covariance matrices. The goodness-of-fit index (GFI) and the normalised fit index (NFI) were calculated to assess the fit of the data to the SEM model [Zhuang et al., 2005]. Based on the constructed initial model, the path coefficients of each subject were measured. The path differences within each group were compared between the first run of the first session and the last run in the second session (seven days after) with a stacked models approach on the same SEM model [McIntosh and Gonzalez-Lima, 1994]. This approach included the comparison of a free model, in which all connections were allowed to vary between the dataset, with a constrained model, in which a given connection was forced to be equal for the dataset.

The results of this work revealed that the path from the prefrontal to parietal cortices exhibited a significantly increased strength. Additionally, the alteration of this connection was significantly correlated with improved performance, suggesting an important role of the frontoparietal connection in working memory training and behavioural improvement.

3.3.4 Structural Vector Autoregression (SVAR)

[Zotev et al., 2013] used SVAR to analyse effective connectivity changes in a network responsible for emotion regulation during rt-fMRI induced training. In a previous study [Zotev et al., 2011], already referred in section 3.2.1, the authors have found an enhancement in functional connectivity between the left AMG and six regions on the Prefrontal Cortex (PFC).

Participants were split in experimental and control group (with "sham feedback "). For the subjects in the experimental group, the NF target ROI was selected on the left AMG, based on a meta-analysis of functional neuroimaging studies regarding the role of the AMG in emotion processing. For the subjects in the control group, the feedback signal was based on the signal from the HIPS, defined also based on a meta-analysis, this time regarding the role of HIPS in number processing. During the experiment, the participants were asked to perform a positive emotion induction task, based on retrieval of happy autobiographical memories. The remaining ROIs for connectivity analysis were selected based on the results of the previous study

[Zotev et al., 2011], as a 10 mm sphere on the regions that showed a significant enhancement in functional connectivity strength with the left AMG. Because of this, the actual method used should be referred to as PeC-informed SVAR.

The application of SVAR followed [Chen et al., 2011], with a first order ($p = 1$) SVAR model. Preprocessing of fMRI data included slice scan timing correction, volume registration, correction for cardiorespiratory artifacts and spatial smoothing with a Gaussian filter of FWHM = 5 mm. The multivariate analysis was performed according to equation 2.22 for the network of seven ROIs for each of the six experimental runs. The exogenous variables z_m included six motion parameters, time courses of the four ROIs within white matter and ventricle and six Legendre polynomials for modelling the baseline. A process for optimising the best starting model for instantaneous effects was performed, followed by a statistical analysis of the resulting path coefficients.

The SVAR analysis demonstrated that the left Rostral Anterior Cingulate Cortex (rACC) plays an important role during rt-fMRI-NF training, modulating the left AMG and the other network regions. According to the analysis, rt-fMRI-NF training leads to a significant enhancement in the time-lagged effect of the left rACC on the left AMG, potentially consistent with the ipsilateral distribution of the mono synaptic projections between these regions. The training is also accompanied by significant increases in the contemporaneous effects of the left rACC on four other regions – the bilateral Dorsal Medial Prefrontal Cortex (DMPFC), the right Medial Frontopolar Cortex (MFPC), and the left Superior Frontal Gyrus (SFG). The instantaneous effects of the left AMG on the bilateral DMPFC are also significantly enhanced.

These exploratory results show insights into the causal relationships within the network of regions engaged during a rt-fMRI-NF procedure targeting the AMG, besides suggesting that the rACC may constitute a promising target for rt-fMRI training along with the AMG in patients with affective disorders.

4

Datasets

Functional and effective connectivity methods were applied on a simulated and an experimental dataset. These datasets will be explored in this chapter.

4.1 Simulated fMRI Dataset

This simulated dataset is based on a VAR model originally defined by [Baccalá and Sameshima, 2001], and explored in detail in [Seth et al., 2013] with the objective of testing Granger Causality against haemodynamic convolution and downsampling. The model is described by the following equations:

$$\begin{aligned}x_1(t) &= 0.95\sqrt{2}x_1(t-l) - 0.9025x_1(t-2l) + w_1(t) \\x_2(t) &= 0.5x_1(t-2l) + w_2(t) \\x_3(t) &= -0.4x_1(t-3l) + w_3(t) \\x_4(t) &= -0.5x_1(t-2l) + 0.25\sqrt{2}x_4(t-l) + 0.25\sqrt{2}x_5(t-l) + w_4(t) \\x_5(t) &= -0.25\sqrt{2}x_4(t-l) + 0.25\sqrt{2}x_5(t-l) + w_5(t)\end{aligned}\tag{4.1}$$

where l determines the neural delay (relative to the model update rate) and $w_i(t)$ are zero-mean uncorrelated white processes with variance of 1.

The model connectivity structure is shown in figure 4.1, alongside with the corresponding connectivity matrix on figure 4.2. These matrices should be read starting with the x axis (from region x) to the y axis (to region y). For instance, the black square shown on position [5,4] of the matrix should be read as "influence from region 4 to region 5".

To simulate BOLD signals, each output of the VAR model is convolved with one of five different HRF kernels (figure 4.3). The difference is visible on the response time of each HRF, purposely chosen to confound the analysis: notice, for instance, that variable x_1 is the model

Figure 4.1: Model Connectivity Scheme.

Figure 4.2: Model Connectivity Matrix.

source signal, but the correspondent HRF (in blue) is the last one to peak. This is a method to simulate the existing variability in real data.

Figure 4.3: HRF kernels.

The HRF kernels were generated based on the SPM12 [8] function `spm_hrf(TR, p)`, which receives as input the repetition time in seconds `TR` and the vector `p` with a set of parameters that shape the HRF. Following the HRFs parameters shown in [Seth et al., 2013], the code below was implemented to create them. Notice the vector `delay_rsp`, which contains five different values for the parameter "delay of response (relative to onset)" of vector `p`. The HRFs created with each of these values `hrf(:, i)` is convolved with each model variable stored in `x`.

```
1 delay_rsp = [3.85 2.3 3 2.4 2.7]*2.5; % delay of response
2 for i = 1:5
3     p = [delay_rsp(i) 16 1 1 6 0 32];
4     hrf(:,i) = spm_hrf(1/fs,p);
5     BOLD(:,i) = conv(X(:,i) , hrf(:,i));
6 end
```

The main purpose of this dataset is method validation: depending on the ability to replicate the results in [Seth et al., 2013] or simply to retrieve the true connectivity matrix, one can assess if the implementation of each connectivity method is correct.

4.1.1 Processes in the order of millisecond

350 s of data were generated with a time step of 1 ms ($f_s = 1000$ Hz) and a neural delay l of 20 ms, based on the assumption that typical neural delays are in the range [20,60] ms [Seth et al., 2013]. The first 50 s of data was discarded to allow for any numerical “burn-in” effects.

For analysis, three signals were created:

1. **Raw 250 Hz** - VAR data was downsampled to 250 Hz.
2. **BOLD 250 Hz** - VAR data was downsampled to 250 Hz and then convolved with the HRF kernels.
3. **BOLD 0.5 Hz** - VAR data was downsampled to 0.5 Hz and then convolved with the HRF kernels, representing typical fMRI BOLD signals with TR = 2 s.

4.1.2 Processes in the order of second

To assess the ability of the methods to recover relations in the order of the second, which will be the requirement for real fMRI data, 350 s of data were generated with a time step of 1 s ($f_s = 1$ Hz) and a neural delay l of 1 s. Once again, the first 50 s of data were discarded.

For analysis, four signals were created:

1. **Raw 1 Hz** - direct output of the VAR model.
2. **BOLD 1 Hz** - VAR data was convolved with the HRF kernels.
3. **Raw 0.5 Hz** - VAR data was downsampled to 0.5 Hz.
4. **BOLD 0.5 Hz** - VAR data was downsampled to 0.5 Hz and then convolved with the HRF kernels, representing typical fMRI BOLD signals with TR = 2 s.

4.2 N-back Task Dataset - Working Memory

4.2.1 Experimental Setup

4.2.1.1 Participants

Seventeen adult volunteers, with no history of neurological or psychiatric diseases, were included in the study. Each volunteer had one scanning session consisting in an anatomical acquisition, neurofeedback training and Diffusion Tensor Imaging (DTI) (figure 4.4).

Figure 4.4: N-back Task - Experimental protocol scheme.

4.2.1.2 MRI acquisition

The rt-fMRI-NF system is based on a 3.0 T Siemens Magnetom Trio scanner with a standard head coil, acquired at Institute of Nuclear Sciences Applied to Health (ICNAS), Coimbra, Portugal. For each subject, a T_1 -weighted Magnetisation-Prepared Rapid Gradient-Echo (MPRAGE) sequence was used to obtain the anatomical images (matrix = 256×256 , 176 partitions, 1 mm^3 isotropic voxels, TR = 2530 ms, Echo Time (TE) = 3.4 ms). A single-shot T_2^* -weighted gradient-echo, EPI sequence was used for the functional imaging acquisition (TR = 2000 ms, TE = 30 ms, matrix = 64×64 , in-plane resolution = $3 \times 3 \text{ mm}^2$, slices = 32, slice thickness = 2.5 mm, slice gap = 0.5 mm, flip angle = 90°). The online analysis of whole-brain fMRI data was performed with Turbo-BrainVoyager software (Brain Innovation, Maastricht, The Netherlands) [4]. The data analysis included online incremental 3D motion correction, drift removal, spatial smoothing (FWHM = 5 mm) and incremental statistical analysis based on the GLM. The threshold for statistical significance was set at $t > 3.0$.

4.2.1.3 Functional Localiser

Before NF training, participants performed a classical N-back task, that reliably promotes cortical BOLD activation in the DLPFC, IntraParietal Sulcus (IPS) and Dorsal Anterior Cingu-

late Gyrus (DACG). The task consists of three conditions: a 1-back and a 2-back condition distributed randomly in 10 blocks (5 blocks per condition), alternating with baseline blocks. Each block has 15 volumes and 15 digits, with 5 of them targets and the remaining 10 non-targets. Before each block, the instruction to remember 1 or 2 preceding numbers appears. The participant is instructed to press a button when the number displayed matches the one from 1 step earlier in the sequence (1-back condition) or 2 steps earlier (2-back condition). A schematic example of the 2-back task is displayed on figure 4.5. The total length of the run is 315 volumes.

Figure 4.5: Localiser example sequence for the 2-back task.

This run allowed to select a seed ROI in the left DLPFC, based on the statistical activation map contrasting 1-back and 2-back conditions (2-back > 1-back).

4.2.1.4 Neurofeedback Runs

During the five neurofeedback runs, the participant is instructed to imagine a sequence of numbers and then recall the sequence in inverse order. The number, length and variability of the sequences per block was left to the participant's discretion, while stating that the sequence should not be based on any mathematical series (e.g. 1-3-5-7-9). During all the runs the participant is seeing the thermometer and the written instruction. On the first and last run (train and transfer) the thermometer value does not change (no feedback), while on the three NF runs the objective is to increase the thermometer value during the imagery blocks and maintain its close to zero value during the baseline blocks.

Each run consists of six blocks of imagery task, each with 15 volumes, alternated with baseline blocks of the same length. The total length of each run is 195 volumes.

4.2.2 Neurofeedback Calculation

Neurofeedback values were calculated using equation 2.3. $b(t)$ is calculated based on the "baseline" condition prior to the active modulation condition.

Due to the BOLD response delay, the values defined by a baseline interval should take the haemodynamic delay into account: to allow the signal to decay from the previous modulation condition, values at the beginning of the baseline condition are excluded, using the `shiftBegin` property. The rule-of-thumb used for this value is $6/TR$ which is 3 in this case. For instance, if the baseline condition is defined from volume 16 to 31, the actual points considered for baseline calculation are 19 to 31. Since it takes about 2-3 seconds until the BOLD signal rises in the modulation block, a `shiftEnd` value should also be defined as $1/3$ of the `shiftBegin` value. In this case, it is 1. Retrieving the previous example, the volumes that are used to calculate the baseline are 19 to 32.

The example code for calculating the baseline values $b(t)$ is shown below:

```
1 if time > ignore
2     if BaselineIndexesUpdate(time)
3         Baseline(time) = mean( TC(last-shiftBegin+1 : time) );
4     else
5         last = time;
6         if time == ignore + 1
7             Baseline(time) = TC(time-1);
8         else
9             Baseline(time) = Baseline(time-1);
10        end
11    end
12 end
```

where

- `time` is the current time point.
- `ignore = 5` - The first five points of the run are ignored to allow for BOLD signal stabilisation.
- `shiftBegin = 3`
- `TC` is a vector with the BOLD signal.
- `Baseline` contains the baseline values $b(t)$.
- `BaselineIndexesUpdate` is a vector with the same length of the run that contains ones in the indexes for which the baseline value should be updated, taking into account the shifts referred.

4.2.3 Offline Analysis

BrainVoyager QX 2.8 software package (Brain Innovation, Maastricht, The Netherlands) [1] was used for fMRI offline image analysis. Preprocessing of functional data included slice scan time correction, 3D motion correction with intrasession alignment and temporal high-pass filtering. Functional data was coregistered to anatomical data and subsequently normalised to Talairach coordinate space.

Statistical maps were computed using a GLM. The design matrix of the GLM contained predictors encoding the experimental conditions, six detrended motion parameters and spikes (confounds were created based on the algorithms of [3]). Single-subject maps were created for all runs and subjects. Multi-subject analysis was performed using Random Effects (RFX) Analysis.

All these steps were automated using MATLAB and the COM functionality of BrainVoyager QX, which allowed for relatively autonomous analysis of all subjects' data, reducing the probability of repetition errors and speeding processing times.

4.2.4 ROI Selection and Time Course Export

The ROIs used for connectivity analysis were selected with the following procedure:

1. A group activation map for the localiser run was generated using RFX Analysis. The map was corrected using FDR with $\alpha < 0.01$ and the contrast used was 2-back > 1-back.
2. From this map, a number of regions were selected with a large number of voxels.
3. For each subject, the voxel with the highest t-value in the subject localiser map, within the corresponding ROI selected in the previous step, was chosen to be the centre of a spherical ROI with a specific radius.

Based on these subject-specific ROIs, the time course for each was extracted to MATLAB using the NeuroElf v1.1 package [9]. The labelling of the regions was performed based on BrainVoyager Brain Tutor v2.5 [2] and on a Talairach Daemon [Lancaster et al., 2000], alongside with the consulting of a trained neuroradiologist.

5

Methods

In the following chapter are explored detailed aspects regarding the practical implementation of each connectivity method.

5.1 Pearson's and Partial Correlation

Pearson's correlation (PeC) was calculated based on the MATLAB function `corr`. The time courses of all ROIs are the input of the function, which then returns a symmetrical correlation matrix containing the correlation values r between all pairs of ROIs.

Regarding the simulated dataset, only PeC was calculated: the function was applied to the complete signals from the five nodes.

For the Localiser and Neurofeedback runs of the N-back dataset, correlation values for each ROI pair are calculated for each subject. The resulting correlation matrices are then averaged across all subjects, resulting in a single matrix. The three Neurofeedback runs were treated as one single run, *i.e.* the time courses were concatenated.

To compare the Train and Transfer runs, condition-specific time courses are generated by concatenating all demeaned blocks of each condition, and the correlation between all possible ROI pairs is calculated for each (each block is demeaned to minimise the effects of discontinuities between them). The first result shown is the averaging, across all subjects, of the difference between the correlation on the Transfer and the Train runs, for each condition. The second result shown is a paired-sample t-test: firstly, the correlation values are converted to z values using Fisher's r -to- z transformation - this assures the normality of the values; then, the MATLAB function `ttest` is used to perform the t-test: correlation matrices of each subject are concatenated in a single 3-dimensional matrix to input this function. Unless stated otherwise, the p value threshold was set to $p < 0.01$ (uncorrected).

Partial Correlation is calculated using MATLAB function `partialcorr`, which tests partial correlation between pairs of ROI time courses while controlling for the effects of a third variable z . In this case, variable z was set to be an impulse train corresponding to the beginning of each condition, convolved with a canonical HRF. This procedure is used in [Sun et al., 2004], where the authors state that by using the stimulus presentation paradigm as the reference function, the method determines whether the relationship between two brain regions is entirely due to a stimulus-locked response, or whether there is an additional task-induced association. In figure 5.1, the impulse train for a condition and the result of the convolution of that impulse with the HRF are shown.

The analysis of each run of the N-back dataset is then identical to the above explained for PeC.

Figure 5.1: Reference function for Partial Correlation and Partial Coherence analysis.

5.2 Coherence and Partial Coherence

Coherence is calculated using MATLAB function `mscohere`, which finds the magnitude-squared coherence estimate of the input signals using Welch's averaged modified periodogram method.

Following the procedure in [Sun et al., 2004], coherence was calculated for the frequencies within the bandwidth of the haemodynamic response: 0 - 0.15 Hz. For the simulated dataset, the complete signals are used. For the N-back dataset, the condition-specific time courses of each ROI were created by concatenating all demeaned blocks of each condition. The function estimates the condition-specific coherence between the time courses of two regions of the brain, using a 30-point (double of the block size) discrete Fourier transform and a Hamming window of length 15 (block size). This choice of parameters was mainly inferred from [Sun

et al., 2004]. The coherence maps show the average coherence on the frequency band 0 - 0.15 Hz. Regarding the simulated dataset, all these parameters were kept at default values.

Partial Coherence was calculated using equation 2.9 and the same reference function as for correlation analysis (figure 5.1).

For the N-back dataset, two tests were analysed: the first is the averaging of the coherence matrices of all subjects for the contrast of interest; the second starts by transforming coherence values to z values, using Fisher's r -to- z transformation, and then applying a t-test on these values. Unless stated otherwise, the p value threshold was set to $p < 0.01$ (uncorrected).

5.3 Granger Causality Mapping

Granger Causality Mapping was implemented using the MVGC multivariate Granger causality toolbox [Barnett and Seth, 2014].

A MATLAB function `mvgc_demo_function` was created to perform Granger Causality analysis, based on the one in the `mvgc_demo` script that is provided with the toolbox. This function receives the prepared fMRI data and a set of parameters that tune the methods used. The data can be input in several ways. Specifically, the calculation can be performed for each subject individually, or each subject's data can be concatenated, either in tensor ($n_{ROIs} \times l_{run} \times n_{subjects}$) or by creating a single time course with length $l_{run} \cdot n_{subjects}$. Additionally, the calculation can be performed for the complete time course, or this can be split by experimental condition. Each condition's block can be either concatenated in a single time course or distributed in tensor ($n_{ROIs} \times l_{block} \times n_{blocks}$).

Regarding the simulated dataset, the complete signal was input. Regarding the N-back task dataset, and after exhaustive testing, the calculation was performed for each subject and for each condition individually, with blocks as trials, and the final set of parameters chosen was:

- VAR model estimation regression mode - Levinson-Wiggins-Robinson (LWR) algorithm.
- Information criteria regression mode - LWR algorithm.
- Model order estimation method - Akaike Information Criteria (AIC) for the N-back dataset and Bayesian Information Criteria (BIC) for the simulated dataset.
- Statistical test for MVGC - Granger's F -test.

The function returns the time-domain pairwise-conditional causalities in a square matrix of Granger F values and performs a significance test using a theoretical null distribution, which identifies which causalities are statistically significant on a matrix with zeros and ones. The results of the `mvgc_demo` script provided with the toolbox are shown in figure 5.2, where the referred result matrices are displayed.

Figure 5.2: MVGC Toolbox output matrices.

For both datasets, two types of results are shown. The first is the average of the Granger F values across all subjects. In the case of the N-back dataset, these values are contrasted between conditions. The second is the average of the significance tests performed for each subject. Each of the tests was performed at $p < 0.01$ (uncorrected), unless otherwise stated.

5.4 Dynamic Causal Modelling

Dynamic Causal Modelling was applied using the DCM10 module from SPM12 [8]. This method was only applied to the N-back task dataset. The MATLAB function `spm_dcm_estimate`, used for estimating the parameters of the DCM for fMRI data, requires the following inputs:

- $\mathbf{a}$ - Endogenous (structural) connections ($n \times n$).
- $\mathbf{b}$ - Bilinear modulations ($n \times n \times n_{cond}$).
- $\mathbf{c}$ - Exogenous connections ($n \times n_{cond}$).
- $\mathbf{d}$ - Nonlinear modulations ($n \times n$).
- $\mathbf{U}$ - Exogenous inputs - experimental protocol ($n_{cond} \times v$).
- $\mathbf{Y.y}$ - BOLD signals ($n \times v$).
- $\mathbf{Y.x0}$ - Confounds ($n_{conf} \times v$).
- $\mathbf{n}$ - Number of regions.
- $\mathbf{v}$ - Number of scans.

Figure 5.3 shows all possible models tested with 2 nodes, assuming a structural connection between all the nodes and none nonlinear modulations. The orange circle represents a

direct input into the system that drives that region's activity. The arrows represent a modulatory action of a certain condition of interest on that connection. The confounds provided were the same as for the GLM calculation.

For instance, for model A1 and the localiser run, the parameter matrices are:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}; B_{baseline} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}; B_{1back} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}; B_{2back} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}; C = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Bayes theory is used to identify which model is the best compromise between model fit (accuracy) and model complexity. The method calculates each model's exceedance probability and then compares between them.

Additionally, hypothetical models with 3 nodes were created and tested, and are shown in figure 5.4. It was computationally not feasible to perform DCM with the 512 possible models, so a set of 7 models was chosen based on theoretical expectations.

Figure 5.3: Theoretical DCM Models with 2 nodes. Orange circles represent a direct input into the system that drives that region's activity and the arrows represent a modulatory action of a certain condition of interest on that connection.

Figure 5.4: Theoretical DCM Models with 3 nodes. Orange circles represent a direct input into the system that drives that region's activity and the arrows represent a modulatory action of a certain condition of interest on that connection.

6

Results

In this chapter are presented the results of this work, both for the simulated and experimental datasets.

6.1 Simulated fMRI Dataset

The following results show the application of connectivity methods on the simulated datasets presented in chapter 4. All results in this section are averaged over 100 tests (100 datasets were created and the methods applied to each one and averaged). Take note that a "Raw" dataset contains signals that were not convolved with an HRF, unlike a "BOLD" dataset, which contains signals that were convolved with an HRF.

6.1.1 Processes in the order of millisecond

6.1.1.1 Pearson's Correlation

The correlation results for the three datasets (Raw 250 Hz, BOLD 250 Hz and BOLD 0.5 Hz) are shown in figure 6.1. It is easily observed that the method failed to retrieve the true model relations. However, for the BOLD datasets, an unexpected stronger negative correlation appears for $1 \leftrightarrow 3$, $1 \leftrightarrow 4$ and $4 \leftrightarrow 5$.

Figure 6.1: Pearson's Correlation Results for the Simulated fMRI Dataset (processes in the order of millisecond).

6.1.1.2 Coherence

Coherence results for the frequency band 0 - 0.15 Hz are shown in figure 6.2. Unlike correlation, this metric shows the true model relations $1 \leftrightarrow 2$, $1 \leftrightarrow 3$, $1 \leftrightarrow 4$ and $4 \leftrightarrow 5$ presenting higher coherence scores.

Figure 6.2: Coherence Results for the Simulated fMRI Dataset (processes in the order of millisecond).

6.1.1.3 Granger Causality Mapping

As clearly stated in [Seth et al., 2013], Granger Causality is invariant to HRF convolution but not to downsampling. These results were replicated and are presented in figures 6.3 and 6.4. Regarding figure 6.3, it is possible to observe that at 250 Hz both the Raw and BOLD signals provide enough information for Granger Causality to successfully retrieve the model connectivity relations. Notice however that connection $5 \rightarrow 4$ was lost. On the other hand, at 0.5 Hz, an inversion of directionality is observed and connections $5 \rightarrow 4$ and $4 \rightarrow 5$ are lost.

Figure 6.3: MVGC Results for the Simulated fMRI Dataset (processes in the order of millisecond).

In figure 6.4 it is shown the average matrix of significance: for every Granger Causality calculation, an FDR test was applied ($\alpha < 0.05$), retrieving the significant Granger Causality values. For both the Raw 250 Hz and BOLD 0.5 Hz signals, this test returns the expected results, as they match the F values in figure 6.3. However, for the BOLD 250 Hz signal, the significant connections are symmetrical.

Figure 6.4: MVGC Significant Results for the Simulated fMRI Dataset (processes in the order of millisecond).

6.1.2 Processes in the order of second

6.1.2.1 Pearson's Correlation

The correlation results for the four datasets (Raw 1 Hz, Raw 0.5 Hz, BOLD 1 Hz, BOLD 0.5 Hz) are shown in figure 6.5. Again, the method fails to retrieve the true model relations for all of them. For the BOLD datasets, an unexpected stronger negative correlation appears for $1 \leftrightarrow 3$, $1 \leftrightarrow 4$ and $4 \leftrightarrow 5$, as in figure 6.1.

Figure 6.5: Pearson's Correlation Results for the Simulated fMRI Dataset (processes in the order of second).

6.1.2.2 Coherence

Coherence results for the band 0 - 0.15 Hz are shown in figure 6.6. The metric manages to retrieve the true model connections for all the signals.

Figure 6.6: Coherence Results for the Simulated fMRI Dataset (processes in the order of second).

6.1.2.3 Granger Causality Mapping

The results for the application of Granger Causality to the four simulated signals are shown in figures 6.7 and 6.8. Surprisingly, the results for the BOLD 1 Hz signal are the best in this analysis: both in figure 6.7 and in figure 6.8 it is possible to distinguish the true model relations (except for $5 \rightarrow 4$).

Figure 6.7: MVGC Results for the Simulated fMRI Dataset (processes in the order of second).

Figure 6.8: MVGC Significant Results for the Simulated fMRI Dataset (processes in the order of second).

6.2 N-back Task Dataset - Working Memory

6.2.1 Group GLM and ROI Selection

The group GLM map for the Localiser run contains the regions that were shown to be activated during an N-back task, such as the IPS, DLPFC and Supplementary Motor Area (SMA), as well as other regions like the Frontal Eye Fields (FEF), Insula and Premotor cortex. In figure 6.9 the map is shown on a brain mesh for the left and right hemispheres, contrasted between the activation during the 2-back and the 1-back conditions (FDR at $\alpha < 0.01$).

Figure 6.9: Group GLM Map for the Localiser run.

Using the method for ROI selection described in section 4.2.4, eleven ROIs were drawn on the group Localiser GLM map. An additional region, on the right Posterior Superior Temporal Sulcus (pSTS), was selected based on anatomical landmarks to act as a control region (outside of the working memory network). The regions are described and labeled on table 6.1.

The two main regions of interest - the left DLPFC (NF target) and the left IPS - are highlighted on figure 6.10 in green and blue, respectively.

Figure 6.10: Group GLM Map for the Localiser run highlighting the two main regions of interest: the left DLPFC (in green) and the left IPS (in blue).

Table 6.1: Regions of Interest selected on the group GLM of the Localiser, indicating Talairach coordinates of the centre of gravity and the number of voxels.

Index	Name	X	Y	Z	Number of Voxels
1	Left IPS	-32	-58	41	10020
2	Right IPS	35	-55	41	10296
3	Left Insula	-32	18	11	2807
4	Right Insula	29	13	13	3746
5	Left DLPFC	-40	24	36	4646
6	Right DLPFC	37	28	36	3956
7	Left FEF	-26	-5	57	5443
8	Right FEF	25	-4	59	5636
9	Left Premotor C	-43	0	37	4575
10	Right Premotor C	47	1	36	2171
11	SMA	0	4	53	5587
12	Right pSTS	50	-45	13	-

The regions were kept big enough to encompass the subject-specific peaks of activation for each ROI. From this point, all results are shown for three ROI sizes: spheres with 19, 81 and 257 voxels around this peak of activation.

6.2.1.1 Brief overview on the function of the ROIs found in the Localiser

The IPS main functions are related to perceptual-motor coordination and visual attention, which allows for visually-guided pointing, object manipulation and focus on specific visual targets. It receives widespread connections from visual and premotor areas.

The Insula is a brain structure implicated in cognitive, affective, and regulatory functions, including interoceptive awareness, emotional responses and empathic processes. While classically considered a limbic region, recent evidence from network analysis suggests a critical role for the Insula (particularly the anterior division) in high-level cognitive control and attentional processes [Menon and Uddin, 2010].

The DLPFC is located in the MFG of the human brain, encompassing BA 46 and BA 6. It is a functionally defined region, which has a major (but not exclusive) role in executive functions, such as working memory, planning, inhibition, and reasoning. The DLPFC is also the highest cortical area involved in motor planning, organisation and regulation.

The FEF play an important role in visual attention and eye movements: the region is activated during the initiation of eye movements, such as voluntary saccades and pursuit eye movements. Together with the Supplementary Eye Fields (SEF), the IPS and the Superior Col-

liculus, the region constitutes one of the most important networks involved in the generation and control of eye movements.

The Premotor Cortex receives input from the cerebellum and is involved in planning movements based on external visual cues. It projects mainly to Primary Motor Cortex (M1) and is involved mostly in control of postural and proximal limb muscles. Together with SMA, its function can be generally described as to transform the intention to perform a complex motor act into the specific sequence of movements necessary to accomplish the act.

The SMA is defined as an area in the frontal agranular cortex, corresponding to the medial part of BA 6, involved with complex motor and motor planning tasks [Chung et al., 2005]. However, recent studies show that the SMA is not a functionally homogenous area, but rather composed of subregions with distinct functional roles: during complex motor and heat sensory tasks, activation tends to occur in the contralateral posterior portion of the region, whereas word generation and working memory tasks tend to produce activation in the anterior portion of SMA, particularly on the left side.

The Superior Temporal Sulcus (STS) is considered a hub for social perception and cognition, including the perception of faces and human motion, as well as understanding others' actions, mental states, and language [Deen et al., 2015]. Evidence suggests that the posterior part of this sulcus is more specific for the perception of human expressions and biological motion.

6.2.2 Connectivity Analysis for the Localiser Run

6.2.2.1 Pearson's and Partial Correlation

Figures 6.11, 6.12 and 6.13 show correlation results for each pair of ROIs, averaged across all subjects. Denote that the matrices are symmetrical (correlation is a metric of functional connectivity) and that the correlation between a region and itself was set to zero for visualisation purposes (the true value is one). The connection of interest $1 \leftrightarrow 5$ and others such as $5 \leftrightarrow 9$, $1 \leftrightarrow 2$, $9 \leftrightarrow 10$ display higher correlation values, common to the analysis with the three ROI sizes. The left DLPFC is correlated with a significant number of regions. The correlation values for the pairs that include region 12 (pSTS) are among the lowest.

Figure 6.11: Mean Pearson's and Partial Correlation values of all subjects for the Localiser run (19 voxels ROIs).

Figure 6.12: Mean Pearson's and Partial Correlation values of all subjects for the Localiser run (81 voxels ROIs).

Figure 6.13: Mean Pearson's and Partial Correlation values of all subjects for the Localiser run (257 voxels ROIs).

6.2.2.2 Coherence and Partial Coherence

In figures 6.14, 6.15 and 6.16 the average coherence results for all subjects are shown, for the three ROI sizes. Denote that the matrices are symmetrical (coherence is a metric of functional

connectivity) and that the coherence between a region and itself was set to zero for visualisation purposes (the true value is one). Some of the connections of interest are visible in the three figures, as $1 \leftrightarrow 5$.

Figure 6.14: Mean Coherence and Partial Coherence values of all subjects, contrasting 2-back and 1-back blocks with baseline blocks (19 voxels ROIs, Localiser run).

Figure 6.15: Mean Coherence and Partial Coherence values of all subjects, contrasting 2-back and 1-back blocks with baseline blocks (81 voxels ROIs, Localiser run).

Figure 6.16: Mean Coherence and Partial Coherence values of all subjects, contrasting 2-back and 1-back blocks with baseline blocks (257 voxels ROIs, Localiser run).

In figures 6.17, 6.18 and 6.19 the t-test results for a confidence level of 99% are shown for the three ROI sizes.

Figure 6.17: t-test results for Coherence and Partial Coherence differences between the 2-back + 1-back conditions and the baseline, across all subjects (19 voxels ROIs, Localiser run).

Figure 6.18: t-test results for Coherence and Partial Coherence differences between the 2-back + 1-back conditions and the baseline, across all subjects (81 voxels ROIs, Localiser run).

Figure 6.19: t-test results for Coherence and Partial Coherence differences between the 2-back + 1-back conditions and the baseline, across all subjects (257 voxels ROIs, Localiser run).

6.2.2.3 Granger Causality Mapping

In figures 6.20, 6.21 and 6.22 the results for the application of the MVGC toolbox to Localiser data are shown. On the left side of the figures, the average of the Granger F matrices across all subjects is shown. On the right side is shown the average of the significant maps at a 99% confidence level.

Figure 6.20: Average MVGC Results for the Localiser runs contrasting the 2-back + 1-back conditions and the baseline (19 voxels ROIs).

Figure 6.21: Average MVGC Results for the Localiser runs contrasting the 2-back + 1-back conditions and the baseline (81 voxels ROIs).

Figure 6.22: Average MVGC Results for the Localiser runs contrasting the 2-back + 1-back conditions and the baseline (257 voxels ROIs).

6.2.3 Connectivity Analysis for the Neurofeedback Runs

6.2.3.1 Pearson's and Partial Correlation

Figures 6.23, 6.24 and 6.25 show correlation results for each pair of ROIs, averaged across all subjects, on the data from the three Neurofeedback runs.

Figure 6.23: Mean Pearson's and Partial Correlation values of all subjects for the Neurofeedback runs (19 voxels ROIs).

Figure 6.24: Mean Pearson's and Partial Correlation values of all subjects for the Neurofeedback runs (81 voxels ROIs).

Figure 6.25: Mean Pearson's and Partial Correlation values of all subjects for the Neurofeedback runs (257 voxels ROIs).

6.2.3.2 Coherence and Partial Coherence

In figures 6.26, 6.27 and 6.28 the results for the average coherence of all subjects are shown, for the three ROI sizes.

Figure 6.26: Mean Coherence and Partial Coherence values of all subjects, contrasting imagery blocks with baseline blocks (19 voxels ROIs, Neurofeedback runs).

Figure 6.27: Mean Coherence and Partial Coherence values of all subjects, contrasting imagery blocks with baseline blocks (81 voxels ROIs, Neurofeedback runs).

Figure 6.28: Mean Coherence and Partial Coherence values of all subjects, contrasting imagery blocks with baseline blocks (257 voxels ROIs, Neurofeedback runs).

In figures 6.29, 6.30 and 6.31 the t-test results for a confidence level of 99% are shown for the three ROI sizes.

Figure 6.29: t-test results for Coherence and Partial Coherence differences between the imagery condition and the baseline, across all subjects (19 voxels ROIs, Neurofeedback runs).

Figure 6.30: t-test results for Coherence and Partial Coherence differences between the imagery condition and the baseline, across all subjects (81 voxels ROIs, Neurofeedback runs).

Figure 6.31: t-test results for Coherence and Partial Coherence differences between the imagery condition and the baseline, across all subjects (257 voxels ROIs, Neurofeedback runs).

6.2.3.3 Granger Causality Mapping

In figures 6.32, 6.33 and 6.34 the results for the application of the MVGC toolbox to Neurofeedback data are shown. On the left side of the figures the average of F values across subjects is shown. On the right side is shown the average of the significant maps at a 99% confidence level.

Figure 6.32: Average MVGC Results for the Neurofeedback runs contrasting the imagery condition and the baseline (19 voxels ROIs).

Figure 6.33: Average MVGC Results for the Neurofeedback runs contrasting the imagery condition and the baseline (81 voxels ROIs).

Figure 6.34: Average MVGC Results for the Neurofeedback runs contrasting the imagery condition and the baseline (257 voxels ROIs).

6.2.4 Connectivity changes between Train and Transfer Runs

6.2.4.1 Pearson's and Partial Correlation

Figures 6.35, 6.36 and 6.37 show the results for each pair of ROIs and averaged across all subjects, of the comparison of correlation values between the Train and Transfer runs.

Figure 6.35: Mean Pearson's and Partial Correlation values, contrasting Transfer and Train runs, of all subjects (19 voxels ROIs).

Figure 6.36: Mean Pearson's and Partial Correlation values, contrasting Transfer and Train runs, of all subjects (81 voxels ROIs).

Figure 6.37: Mean Pearson's and Partial Correlation values, contrasting Transfer and Train runs, of all subjects (257 voxels ROIs).

6.2.4.2 Coherence and Partial Coherence

The results of coherence analysis are separated by experimental conditions in this case. Figures 6.38, 6.40 and 6.42 show the results for the three ROI sizes of the comparison between the imagery blocks of the Train and Transfer runs. Figures 6.39, 6.41 and 6.43 show the results for the baseline blocks.

Figure 6.38: Results for Coherence and Partial Coherence, contrasting Transfer and Train runs, for the imagery blocks (19 voxels ROIs).

Figure 6.39: Results for Coherence and Partial Coherence, contrasting Transfer and Train runs, for the baseline blocks (19 voxels ROIs).

Figure 6.40: Results for Coherence and Partial Coherence, contrasting Transfer and Train runs, for the imagery blocks (81 voxels ROIs).

Figure 6.41: Results for Coherence and Partial Coherence, contrasting Transfer and Train runs, for the baseline blocks (81 voxels ROIs).

Figure 6.42: Results for Coherence and Partial Coherence, contrasting Transfer and Train runs, for the imagery blocks (257 voxels ROIs).

Figure 6.43: Results for Coherence and Partial Coherence, contrasting Transfer and Train runs, for the baseline blocks (257 voxels ROIs).

6.2.4.3 Granger Causality Mapping

Shown in figures 6.44, 6.45 and 6.46 are the results of the t-tests performed on the Granger F values contrasting the Train and Transfer runs.

Figure 6.44: MVGC Results: t-test across all subjects, contrasting the Transfer and Train runs, for imagery and baseline blocks (19 voxels ROIs).

Figure 6.45: MVGC Results: t-test across all subjects, contrasting the Transfer and Train runs, for imagery and baseline blocks (81 voxels ROIs).

Figure 6.46: MVGC Results: t-test across all subjects, contrasting the Transfer and Train runs, for imagery and baseline blocks (257 voxels ROIs).

6.2.5 Dynamic Causal Modelling

6.2.5.1 DCM for the Localiser Run

In figure 6.47 the results for the DCM analysis of the models with 2 nodes are shown. It is clear that model A3, that assumes a direct influence on the left IPS and a modulation of the interaction in both directions, is the winner for the 19 and 81 voxel ROIs. For the 257 voxels ROIs, model B3 shows greater probability: the difference is that B3 assumes a direct influence on the left DLPFC.

Figure 6.47: Model Exceedance Probability Results for RFX analysis of the models with 2 nodes for the Localiser run.

The results for the DCM analysis with the models with 3 nodes are shown in figure 6.48. A considerable decrease in robustness is shown, since the three ROI sizes return different best models: for instance, the best model for the ROIs with 19 voxels - A2 - is one of the worst for the ROIs with 81 voxels.

Figure 6.48: Model Exceedance Probability Results for RFX analysis of the models with 3 nodes for the Localiser run.

6.2.5.2 DCM for the Neurofeedback Runs

The DCM results for the three neurofeedback runs are shown in figure 6.49 for the 2 node models and in figure 6.50 for the 3 node models. Regarding the former, model C2 is considered the best with 19 voxel ROIs for the three runs, but this result is considerably less evident for the other two ROI sizes.

Figure 6.49: Model Exceedance Probability Results for RFX analysis of the models with 2 nodes for the three NF runs.

Regarding the results for the 3 node models, B2 is the best model in NF1 for the 19 and 257 voxel ROIs, but loses significantly to model A3 for the 81 voxel ROIs. The remaining results oscillate between these two models.

Figure 6.50: Model Exceedance Probability Results for RFX analysis of the models with 3 nodes for the three NF runs.

6.2.5.3 DCM for Train and Transfer Runs

The comparison between the DCM results obtained for the Train and Transfer runs are shown in figures 6.51 (2 node models) and 6.52 (3 node models). For the 2 node models, C2 is the clear winner for the Train run with the 19 and 257 voxels ROIs. With the 81 voxel ROIs, this

model is considered worse than model A3. C2 is also the best model for the Transfer run with 19 and 81 voxels ROIs.

Figure 6.51: Model Exceedance Probability Results for RFX analysis of the models with 2 nodes for the Train and Transfer runs.

Regarding the 3 node models, it is shown that model A3 is the clear winner for the three ROI sizes. This model assumes a direct influence on the left IPS and modulation of the connections $IPS \rightarrow Insula$, $Insula \rightarrow DLPFC$ and $DLPFC \rightarrow IPS$. No evident change from the train to the transfer run is shown.

Figure 6.52: Model Exceedance Probability Results for RFX analysis of the models with 3 nodes for the Train and Transfer runs.

7

Discussion

The results obtained are discussed in this chapter, starting with the analysis of the results for the simulated dataset.

7.1 Results for the Simulated fMRI Dataset

Processes in the order of millisecond

The main goal of this dataset was to replicate and extend the study by [Seth et al., 2013], in order to validate the application of the MVGC Toolbox. Additionally, functional connectivity methods - correlation and coherence - were implemented to assess their capability of recovering the causal relations and therefore validating their application on fMRI data.

As referred in the results section, Pearson's Correlation was unable to distinguish the model relations. The appearance of a relatively strong negative correlation for the relations $1 \leftrightarrow 3$, $1 \leftrightarrow 4$ and $4 \leftrightarrow 5$ has no immediate explanation. On the other hand, the distinction between coherence values for the true model relations and the remaining ones is very clear (differences around 0.1 and higher), in any of the three signals.

When evaluating Granger Causality results, the implementation of the method seems valid, since after 100 tests the method is able to recover 4 out of the 5 model relations for the Raw 250 Hz and BOLD 250 Hz signals, and a causality inversion is observed for the BOLD 0.5 Hz signal, just like the authors in [Seth et al., 2013] observed. After several tests, while varying some parameters, it was not possible to recover the causal relation from node 5 to 4, unlike in the mentioned paper. The reason for this absence was not confirmed. In figure 6.4, for the BOLD 250 Hz signal, the method shows already some indecision about the direction of the causality, as well as some connections that do not exist in the theoretical model with strong statistical significance.

Processes in the order of second

By creating a dataset where the causality dynamics was in the order of the second, the objective was to prove that the application of the three connectivity methods is valid for real fMRI data, assuming that in these the causality dynamics is not only present in the order of neural communications, but also in the order of fMRI acquisition rate.

Pearson's correlation fails again to provide any information about the connectivity of the data, for any of the four signals. Coherence manages to retrieve all true model relations in the signal of interest BOLD 0.5 Hz, revealing some independence to the sampling rate of the signal and therefore validating its application to real fMRI data.

Regarding Granger Causality results for the Raw 1 Hz and Raw 0.5 Hz, the lower than expected accuracy value is clear: for an fMRI signal with these characteristics, the results would be certainly unpredictable. For the BOLD 1 Hz signal, the method manages to recover 4 of the 5 model relations, although showing some indecision about its causal direction. For the signal of interest BOLD 0.5 Hz, the method manages to recover 4 of the 5 model relations as well, but also fails by recovering other 4 relations that are not from the model. These results tend to show that even with a clear causal dynamics in the order of the second it is not immediate for the method to recover the data connectivity.

Brief conclusions

The method that shows greater potential to be applied on real fMRI data is Coherence. Both for the datasets with dynamics in the order of the millisecond and the second, the method managed to retrieve the true model relations. Serious doubts are placed for Pearson's Correlation and Granger Causality, based on these results. The studies of [Seth et al., 2013] and [Rodrigues and Andrade, 2014] state specific limitations of Granger Causality, one of them clearly identified - downsampling.

7.2 Results for the N-back Task Dataset

Localiser Run

The success of the task seems clear in the Localiser RFX analysis GLM map, which shows the expected regions of the working memory network being activated during the conditions of interest with high statistical significance (FDR with $\alpha < 0.01$, figures 6.9 and 6.10). The contrast used (2-back $>$ 1-back) is quite specific. The ROI selection process was not straightforward, though. Even in the Talairach Space, the labelling/identification of the regions is not immediate, specially when one wants functional labels for the regions, instead of pure anatomical ones. Both softwares used mainly return anatomical labels, sometimes divergent, and so further research had to be performed to define each of the regions functionally. After consulting with a trained neuroradiologist, the final set of regions proved to be a sound theoretical model for the working memory network.

The results of Pearson's Correlation analysis are surprisingly informative. In figure 6.11, most of the theoretically expected connections show higher correlation values: between the left IPS and left DLPFC (1 $\leftrightarrow$ 5) and interhemispheric IPS and DLPFC, among others such as the motor areas FEF, SMA and premotor cortex, that are fully explainable. Note the substantially lower correlation values for the connections with the pSTS (control node purposely outside the network), and the equivalent results for the analysis with the three ROI sizes. The higher number of connections with the left DLPFC seems to suggest that this region is an important integrating hub.

Partial Correlation removes magnitude of the correlation values by eliminating the direct stimulus dependencies. Even so, the theoretically expected connections are still very distinguishable.

Regarding Coherence results, the expectations were that the regions involved in the 1-back and 2-back tasks, such as the IPS and DLPFC, would show higher coherence values between them, in the average of all subjects. As shown in figures 6.14, 6.15 and 6.16, this is exactly the case for the 1 $\leftrightarrow$ 5 (left IPS $\leftrightarrow$ left DLPFC) connection. Additionally, connections such as right IPS $\leftrightarrow$ left premotor cortex and left DLPFC $\leftrightarrow$ right FEF also show considerable differences in coherence but are less expected, specially because the regions are in different hemispheres. The control ROI appears to work as expected, but in a much more discrete way when compared to correlation results.

Note that the contrast used for coherence analysis is different from the one used for the GLM. As it was shown by the contrast 2-back $>$ 1-back, on the GLM, the activation during the

2-back condition is higher, but it was assumed that the neural mechanisms involved in the 1-back and 2-back tasks are the same. As such, the simplest approach was to join both tasks and contrast them against the baseline.

Partial Coherence removes some magnitude of the coherence values by eliminating the direct stimulus dependencies. Even so, the referred connections also appear, but become more evident for the motor areas and less to the main connection of interest $1 \leftrightarrow 5$.

The t-tests of figures 6.17, 6.18 and 6.19 confirm the results found with the coherence means, showing clear evidence for the higher coherence between the motor areas 7, 8, 9 and 10, between the IPS and DLPFC, among others. One unexpected connection with the pSTS is displayed ($7 \leftrightarrow 12$).

The size of the ROI shows to have little influence on the coherence values, since no major differences are noticeable between the analysis with the three ROI sizes.

The results of Granger Causality analysis are considerably more disperse. The matrices showing the average of the significant maps display, for the three ROI sizes, a causal influence of the left IPS on the left DLPFC ($1 \rightarrow 5$). Additionally, interhemispheric connections from the left to the right DLPFC ($5 \rightarrow 6$) and from the left to the right Insula are also significant. Other less expected relations appear, including three with the control region, the pSTS.

By superimposing the results for the three ROI sizes (figure 7.1) it is clear that a minority of the connections is common to all ROI sizes - analysing the matrix on the right of figure 7.1, in the total of 15 different connections, 6 are common to the three sizes, 4 are common to two sizes and the remaining only appear for one ROI size. This reveals an important dependence of Granger Causality on the ROI number of voxels.

Figure 7.1: Overlap of Granger Causality results for the Localiser run, for the three ROI sizes. On the left side is shown the sum of the Granger F values and on the right side is shown the sum of the significant connections.

It is important to notice that the connections of interest referred appear only for a minority

of the subjects. With such a robust task as the N-back performed in the localiser of this experiment, the disperse results obtained are disappointing for Granger Causality.

Neurofeedback Runs

The neural mechanisms that are involved in the N-back task are expected to be identical to the ones used during the recall of number sequences in the Neurofeedback runs.

Coherence results do not verify this assumption, since the difference in the coherence values between the imagery and baseline conditions is very small, for any of the ROI sizes (figures 6.26, 6.27 and 6.28). The link $9 \leftrightarrow 10$ shows the greater distinguishable difference revealing coherence between the premotor cortices of the two hemispheres. Partial coherence reveals to be even less informative, by not showing a single significant difference between imagery and baseline blocks for any ROI size.

The t-tests performed between the two conditions reveal one single significant connection for the 81 voxel ROIs - $5 \leftrightarrow 10$, at a 99% confidence level. This connection between the left DLPFC and the right premotor displays a higher value on the average matrices as well, but it was not one of the most expected, mainly because the regions are on different hemispheres.

Correlation tells a very different story: the interest connection $1 \leftrightarrow 5$ has one of the lowest values of the analysis, and the higher correlation values are for the connections involving motor areas, as $2 \leftrightarrow 8$, $3 \leftrightarrow 9$, $9 \leftrightarrow 10$, $9 \leftrightarrow 11$ and $10 \leftrightarrow 11$. These results are consistent for the three ROI sizes and for partial correlation. The results also indicate that the left Insula and IPS are communicating.

Regarding Granger Causality analysis, the results are again a lot more disperse. Summing the results for the three ROI sizes - figure 7.2 - it is possible to see that a number of connections include the control region pSTS. However, some relevant interhemispheric connections such as $3 \rightarrow 4$ (Insula) and $6 \rightarrow 5$ (DLPFC) appear.

Figure 7.2: Overlap of Granger Causality results for the Neurofeedback runs, for the three ROI sizes. On the left side is shown the sum of the Granger F values and on the right side is shown the sum of the significant connections.

Train vs Transfer Runs

A number of studies illustrate that the effect of NF is more evident after a significant number of sessions [Coben and Evans, 2011]. The subjects of this experiment only performed one NF session, and as such, large connectivity changes are not expected when comparing the metrics before and after the single NF session.

Regarding correlation results, the variations are in fact low: the highest increase in correlation - around 0.2 - is shown for connection $3 \leftrightarrow 7$ in figures 6.35, 6.36 and 6.37. The highest decrease is shown for connection $7 \leftrightarrow 9$. Denote also that the correlation for the connection of interest $1 \leftrightarrow 5$ decreases slightly, on average, from the Train to the Transfer run, which is rather unexpected. The analysis is equivalent for Partial Correlation results.

Coherence results tell a slightly different story. Although the variations are small between the runs, there are significant differences between the imagery and baseline blocks for the three ROI sizes. For instance, in figure 6.38a, it is clear that the interhemispheric coherence between the Insula, as well as the coherence between the left IPS and DLPFC has increased in the Transfer run for the imagery blocks. These changes are not visible for the analysis on the baseline blocks (figure 6.39a), as expected. Unfortunately, the t-test results suggest that none of these changes is statistically significant for any of the ROI sizes, at a 99% confidence level.

Considering Granger Causality results, it is seen that only the causal relation from the left Insula to the left FEF suffered a significant change for the imagery blocks. The results are not consistent across ROI sizes.

7.3 DCM Analysis

The objective of applying DCM was to better assess the causality directions of expected and found connections in the working memory network studied by this experiment. With that in view, models with the two main regions of interest - left IPS and DLPFC - were created, varying the direction of connectivity between them and the stimulus influence.

Regarding the results for these models with 2 nodes, it is possible to see that during the Localiser run, DCM predicts that the task influences the communication between the regions on both directions and stimulates directly the IPS. This result is seen for two of the ROI sizes in figure 6.47. Physiologically, it is a valid result: the IPS integrates visual information (the sequence of numbers) and initiates the task, communicating with the DLPFC that interprets this information.

During the NF runs, the results are not so clarifying. However, there is a tendency for model C2, which assumes a direct influence of the stimulus on both regions and on the communication from the DLPFC to the IPS. Physiologically, this is also a valid result: the IPS loses its generating functions, since it is an imagery task, and keeps its executive functions. Unfortunately, this result is not common to the analysis with the three ROI sizes or to the three NF runs.

When comparing connectivity patterns during Train and Transfer runs, a major change is not expected - the subjects performed only one NF session. Then, DCM should return the same model for both runs. This is exactly the case for the analysis with 19 voxel ROIs: model C2 is the best fit for both runs, although being more evident for the Train run. Notice that model C2 is also the best model for the NF runs, as expected since the imagery task is the same. However, when analysing the results for the 81 voxel ROIs, model A3 wins for both the Train and Transfer run (the same model chosen for the Localiser), and for the 257 voxel ROIs, model B1 wins for the Transfer run.

Keeping in mind the integrating functions of the Insula, models with 3 nodes were additionally tested. Due to the high computational requirements of performing DCM with all possible models, a set of seven models was selected, mainly to assess the effect of the presence of this third integrating region on the information flow, during all the runs.

The results are in general less robust, as they do not tend so evidently to one model. For the Localiser run, it is difficult to choose a best model, since the analysis with the three ROI sizes returns a different result. Even so, the model with the most different probability is A2 for the 19 voxel ROIs. This model assumes a direct stimulus influence on the IPS, just like the

2-node model A3. It also assumes that the IPS communicates with the DLPFC directly and through the Insula.

Analysing the results for the NF runs, it is shown that the best model oscillates between B2 and A3. The only difference between the two is the direction of connectivity between the IPS and DLPFC. There is not enough physiological information to assess which one is more correct, but comparing with the results for the 2-node models, A3 is the one that matches the directionality. In fact, in the comparison between Train and Transfer runs, the winning model is common to the three ROI sizes and both runs - model A3.

7.4 Comparison of Connectivity Methods

A general concordance between the results given by the different connectivity methods is expected, even with the metrics of functional and effective connectivity looking at different properties of the signal.

Coherence proved to be the only method to retrieve the model connectivity of the simulated dataset, regardless of the sampling rate. However, on real fMRI data, Correlation also returns surprisingly interesting results, in relative accordance to the Coherence analysis and the theoretical model. This discrepancy could indicate that the simulated dataset does not capture a realistic data correlation structure.

Instead of choosing one of them for analysing functional connectivity, it is important to state that the metrics assess "similarity" by looking at different spaces: Coherence analyses in the frequency space, while Correlation analyses in the time space. Because of this, each of them is more robust to certain signal properties. A complete comparison is performed in [Guevara and Corsi-Cabrera, 1996] for Electroencephalography (EEG) signals.

Partial metrics correct for direct stimuli-related responses, but the explanation behind the strong reduction in magnitude of Correlation and Coherence values is unclear. Additionally, the risk of reducing a connectivity value incorrectly seems high (Type II error).

Regarding effective connectivity, Granger Causality is the most widely used method, but on this study, it showed to be highly influenceable by a number of parameters, while failing to return expected results. Additionally, the relations found by this metric have no meaningful similarity with the ones found by Correlation and Coherence. Mainly because of this fact, it was not possible to define a clear network of communication for the working memory task in study. Supporting the results of [Seth et al., 2013], this metric shows no evident promise

of application to data with this temporal resolution, which precludes naive application of the metric to fMRI data in general.

DCM evaluates fMRI data in a different perspective, simply because it requires *a priori* models to work. Nevertheless, the metric showed great potential to validate Granger Causality results, for instance. A very clear drawback of this method is its computational requirements: it was only possible to perform DCM on simple models with 2 nodes and on a selection of models with 3 nodes, mainly because the method requires considerably more computational time than any other method tested.

A final note regarding the absence of corrections for multiple comparisons, due mainly to the exploratory/preliminary character of this study. To perform these corrections, it would be necessary, at first, to have more data.

7.5 Experimental and Data Processing Notes

The first and main conclusion regarding experimental decisions is related to the TR. The drastic effect of downsampling shown for the simulated dataset, as well as the statistically weak results found for Granger Causality for the N-back task dataset, indicate clearly that the TR should be chosen as small as possible. If the regions of interest are well defined *a priori*, as they could be for a NF experiment, a whole-brain analysis is not required and MRI coverage should be compromised as much as possible to acquire BOLD signals with a higher sampling frequency.

The decision about block and run duration is also important: too short and the number of points is insufficient, too long and the subject gets tired very easily. An average block duration of 30 s and run duration of 10 min were found to be advisable values.

Data processing proved to have a very important role: spatial normalisation is fundamental for a successful ROI selection process and should be performed as equal as possible for all subjects; preprocessing should avoid at all costs to remove important information for connectivity analysis, as for instance temporal smoothing would; head motion confounds proved also to have great influence on the GLM results, and because of that they should always be considered. In the literature, however, the processing steps are far from consensual.

Regarding NF, the number of sessions and its temporal spacing is a matter of variability. Due to the small number of studies and the incoherence between those, it is not possible to define an optimal number for these parameters. Even in EEG literature, where the number of studies is much larger, the numbers vary greatly.

8

Conclusions

The main objective of this work was the creation of a pipeline of analysis that could provide the researcher information about the connectivity patterns between regions, based on a systematised analysis and comparison of the methods and understanding of the strengths and caveats of each one.

Regarding the results found for the simulated dataset, Coherence showed robustness to downsampling, unlike Correlation and Granger Causality. The creation of a dataset with causal dynamics in the order of fMRI data was successful.

The preliminary analysis of connectivity on the experimental dataset allowed for interesting insights on the working memory network functioning, although the validation of the results is incomplete, due mainly to the heterogeneity of the results, ROI size differences and uncertain *a priori* expectations.

Connectivity methods application and validation is not consensual in the literature studied, concluding that the variability of the analysis starts early on the processing stages and remains until the statistical analysis of the results obtained.

8.1 Future Work

Diffusion Tensor Imaging A great contribution to connectivity analysis is the knowledge about anatomical connections between regions of interest. If it is possible to identify, using Diffusion Tensor Imaging (DTI), that specific regions are anatomically connected, then there is a greater chance they will communicate during a specific task. This would allow to better target the methods that require an *a priori* model of connectivity.

Neurofeedback Experiment The development of a fMRI experiment that includes a number of NF sessions is fundamental to be able to compare the effects of NF on brain connectivity. Additionally, this experiment should use the best set of parameters for connectivity analysis: for this matter, a parameter optimisation study should be performed before on available datasets, so as to design the experiment to potentiate the validity of the analysis.

Granger Causality Mapping Further investigation could be performed on GCM for analysing fMRI data, specially regarding its validity. Additionally, with the development of faster MRI scanners, the sampling rate could be increased and the results compared between datasets with different TRs.

State-Space GC A new growing approach in the literature on Granger Causality applications is the coupling of this method with State-spaces [Barnett and Seth, 2015, Solo, 2015, Valdes-Sosa et al., 2011]. A strong moving average component is often present in neural data, and it is poorly represented by an autoregressive model of finite order. This component can be intrinsic to the signal or can be induced by data acquisition, sampling or preprocessing procedures. Ideally, an autoregressive moving average (ARMA) can consistently model this data. To build such a model, a finite order linear state-space is required: [Barnett and Seth, 2015] shows a rigorous demonstration of state-spaces potential when coupled with Granger Causality, suggesting a powerful approach that can go beyond linear, stationary and homoscedastic assumptions and should become the default approach for Granger Causality analysis. [Solo, 2015] achieves similar results, but states that the method is strongly affected by downsampling, questioning its application to fMRI data and advising further research in this field.

Surrogates The application of surrogates is a widely used procedure to assess the validity of results: the methods in study are applied to signals that were transformed in a specific way, as to confound the analysis, and the results are then compared with the original. For instance, a time reversed Granger Causality analysis involves applying the method

to time reversed time courses of ROIs - the frequency spectrum of the data is maintained, but the temporal relations are inverted. In a multivariate analysis as the one performed by the MVGC Toolbox, this procedure can help eliminate false results.

Visualisation of Connectivity Maps Connectivity matrices provide a simple way of showing results of functional and effective connectivity, giving information based on the colour and position of each square. However, other graphics could be of more interest, depending on the analysis and the target audience. For instance, for effective connectivity metrics, it is possible to count the number of connections that a region receives and initiates and create a bar plot that clearly allows to identify the role of each region in the network. If a more clear anatomical view is advisable, for instance to present the results to people with no technical background, 3D graphics of the brain can be created, in which the regions of interest and the connections found by the methods are displayed.

SEM and SVAR Implementation of SEM and SVAR methods for connectivity analysis, and consequent comparison with the studied methods.

References

- [Baccalá and Sameshima, 2001] Baccalá, L. A. and Sameshima, K. (2001). Partial Directed Coherence: a new Concept in Neural Structure Determination. *Biological Cybernetics*, 84:463–474.
- [Baddeley, 2003] Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36(3):189–208.
- [Baddeley and Hitch, 1974] Baddeley, A. and Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, 8:47–89.
- [Barnett and Seth, 2014] Barnett, L. and Seth, A. K. (2014). The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods*, 223:50–68.
- [Barnett and Seth, 2015] Barnett, L. and Seth, A. K. (2015). Granger causality for state space models.
- [Berman et al., 2013] Berman, B. D., Horovitz, S. G., and Hallett, M. (2013). Modulation of functionally localized right insular cortex activity using real-time fMRI-based neurofeedback. *Frontiers in Human Neuroscience*, 7(October):1–11.
- [Chen et al., 2011] Chen, G., Glen, D. R., Saad, Z. S., Paul Hamilton, J., Thomason, M. E., Gotlib, I. H., and Cox, R. W. (2011). Vector autoregression, structural equation modeling, and their synthesis in neuroimaging data analysis. *Computers in Biology and Medicine*, 41(12):1142–1155.
- [Chung et al., 2005] Chung, G. H., Han, Y. M., Jeong, S. H., and Jack, C. R. (2005). Functional heterogeneity of the supplementary motor area. *AJNR. American journal of neuroradiology*, 26(August):1819–1823.
- [Coben and Evans, 2011] Coben, R. and Evans, R. J. (2011). *Neurofeedback and Neuromodulation Techniques and Applications*. Elsevier, first edition.

- [Deen et al., 2015] Deen, B., Koldewyn, K., Kanwisher, N., and Saxe, R. (2015). Functional Organization of Social Perception and Cognition in the Superior Temporal Sulcus. *Cerebral cortex (New York, N.Y. : 1991)*, pages 1–14.
- [Friston et al., 2003] Friston, K., Harrison, L., and Penny, W. (2003). Dynamic causal modeling. *NeuroImage*, 19(4):1273–1302.
- [Friston, 2011] Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connectivity*, 1(1):13–36.
- [Guevara and Corsi-Cabrera, 1996] Guevara, M. A. and Corsi-Cabrera, M. (1996). EEG coherence or EEG correlation? *International Journal of Psychophysiology*, 23(3):145–153.
- [Haller et al., 2013] Haller, S., Kopel, R., Jhooti, P., Haas, T., Scharnowski, F., Lovblad, K.-O., Scheffler, K., and Van De Ville, D. (2013). Dynamic reconfiguration of human brain functional networks through neurofeedback. *NeuroImage*, 81:243–252.
- [Harmelech et al., 2013] Harmelech, T., Preminger, S., Wertman, E., and Malach, R. (2013). The Day-After Effect: Long Term, Hebbian-Like Restructuring of Resting-State fMRI Patterns Induced by a Single Epoch of Cortical Activation. *Journal of Neuroscience*, 33(22):9488–9497.
- [Huettel et al., 2004] Huettel, S. A., Song, A. W., and McCarthy, G. (2004). *Functional Magnetic Resonance Imaging*. Sinauer Associates, Inc, Sunderland, Massachusetts, USA.
- [Hui et al., 2014] Hui, M., Zhang, H., Ge, R., Yao, L., and Long, Z. (2014). Modulation of functional network with real-time fMRI feedback training of right premotor cortex activity. *Neuropsychologia*, 62(1):111–123.
- [Kadosh et al., 2015] Kadosh, K. C., Luo, Q., Burca, C. D., Sokunbi, M. O., Feng, J., Linden, D. E. J., and Lau, J. Y. F. (2015). Using real-time fMRI to influence differential effective connectivity in the developing emotion regulation network. *NeuroImage*, 125:616–626.
- [Karch et al., 2015] Karch, S., Keeser, D., Hümmer, S., Paolini, M., Kirsch, V., Karali, T., Kupka, M., Rauchmann, B.-S., Chrobok, A., Blautzik, J., Koller, G., Ertl-Wagner, B., and Pogarell, O. (2015). Modulation of Craving Related Brain Responses Using Real-Time fMRI in Patients with Alcohol Use Disorder. *Plos One*, 10(7):e0133034.
- [Lancaster et al., 2000] Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., Kochunov, P. V., Nickerson, D., Mikiten, S. A., and Fox, P. T. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping*, 10(3):120–131.

- [Lee et al., 2012] Lee, J.-H., Kim, J., and Yoo, S.-S. (2012). Real-time fMRI-based neurofeedback reinforces causality of attention networks. *Neuroscience Research*, 72(4):347–354.
- [McIntosh and Gonzalez-Lima, 1994] McIntosh, A. and Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Human Brain Mapping*, 2(March):2–22.
- [McIntosh and Gonzalez-Lima, 1998] McIntosh, A. R. and Gonzalez-Lima, F. (1998). Large-scale functional connectivity in associative learning: interrelations of the rat auditory, visual, and limbic systems. *Journal of Neurophysiology*, 80(6):3148–3162.
- [Menon and Uddin, 2010] Menon, V. and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, pages 1–13.
- [Moher et al., 2009] Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Grp, P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Reprinted from Annals of Internal Medicine). *Physical Therapy*, 89(9):873–880.
- [Ogawa et al., 1990] Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24):9868–72.
- [Paret et al., 2015] Paret, C., Ruf, M., Fungisai Gerchen, M., Kluetsch, R., Demirakca, T., Jungkunz, M., Bertsch, K., Schmahl, C., and Ende, G. (2015). fMRI neurofeedback of amygdala response to aversive stimuli enhances prefrontal-limbic brain connectivity. *NeuroImage*, 125:182–188.
- [Petrie and Sabin, 2000] Petrie, A. and Sabin, C. (2000). *Medical Statistics at a Glance*. Number 9. Blackwell Science Ltd, Oxford, first edition.
- [Rodrigues and Andrade, 2014] Rodrigues, J. and Andrade, A. (2014). Lag-based effective connectivity applied to fMRI: A simulation study highlighting dependence on experimental parameters and formulation. *NeuroImage*, 89:358–377.
- [Roebroeck et al., 2005] Roebroeck, A., Formisano, E., and Goebel, R. (2005). Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage*, 25(1):230–242.
- [Ruiz et al., 2013] Ruiz, S., Lee, S., Soekadar, S. R., Caria, A., Veit, R., Kircher, T., Birbaumer, N., and Sitaram, R. (2013). Acquired self-control of insula cortex modulates emotion recognition and brain network connectivity in schizophrenia. *Human Brain Mapping*, 34(1):200–212.

- [Sarkheil et al., 2015] Sarkheil, P., Zilverstand, A., Kilian-Hütten, N., Schneider, F., Goebel, R., and Mathiak, K. (2015). fMRI feedback enhances emotion regulation as evidenced by a reduced amygdala response. *Behavioural Brain Research*, 281:326–332.
- [Scharnowski et al., 2014] Scharnowski, F., Rosa, M. J., Golestani, N., Hutton, C., Josephs, O., Weiskopf, N., and Rees, G. (2014). Connectivity Changes Underlying Neurofeedback Training of Visual Cortex Activity. *PLoS ONE*, 9(3):e91090.
- [Scheinost et al., 2013] Scheinost, D., Stoica, T., Saksa, J., Papademetris, X., Constable, R. T., Pittenger, C., and Hampson, M. (2013). Orbitofrontal cortex neurofeedback produces lasting changes in contamination anxiety and resting-state connectivity. *Translational Psychiatry*, 3(4):e250.
- [Seth et al., 2013] Seth, A. K., Chorley, P., and Barnett, L. C. (2013). Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling. *NeuroImage*, 65:540–55.
- [Shen et al., 2015] Shen, J., Zhang, G., Yao, L., and Zhao, X. (2015). Real-time fMRI training-induced changes in regional connectivity mediating verbal working memory behavioral performance. *Neuroscience*, 289:144–152.
- [Siero et al., 2013] Siero, J. C. W., Bhogal, A., and Jansma, J. M. (2013). Blood oxygenation level-dependent/functional magnetic resonance imaging: Underpinnings, practice, and perspectives. *PET Clinics*, 8(3):329–344.
- [Sitaram et al., 2014] Sitaram, R., Caria, A., Veit, R., Gaber, T., Ruiz, S., and Birbaumer, N. (2014). Volitional control of the anterior insula in criminal psychopaths using real-time fMRI neurofeedback: a pilot study. *Frontiers in Behavioral Neuroscience*, 8(October):1–13.
- [Solo, 2015] Solo, V. (2015). State Space Methods for Granger-Geweke Causality. *arXiv*.
- [Storkey et al., 2007] Storkey, A. J., Simonotto, E., Whalley, H., Lawrie, S., Murray, L., and McGonigle, D. (2007). Learning structural equation models for fMRI. *Advances in Neural Information Processing Systems*, 19:1329–1336.
- [Sun et al., 2004] Sun, F. T., Miller, L. M., Esposito, M. D. ., and Wheeler, H. H. (2004). Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. *NeuroImage*, 21:647–658.
- [Talairach and Tournoux, 1988] Talairach, J. and Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain. *Theime, Stuttgart, Germany*, 270:132.

- [Valdes-Sosa et al., 2011] Valdes-Sosa, P. A., Roebroeck, A., Daunizeau, J., and Friston, K. (2011). Effective connectivity: Influence, causality and biophysical modeling. *NeuroImage*, 58(2):339–361.
- [Weiskopf, 2012] Weiskopf, N. (2012). Real-time fMRI and its application to neurofeedback. *NeuroImage*, 62(2):682–692.
- [Yuan et al., 2014] Yuan, H., Young, K. D., Phillips, R., Zotev, V., Misaki, M., and Bodurka, J. (2014). Resting-state functional connectivity modulation and sustained changes after real-time functional magnetic resonance imaging neurofeedback training in depression. *Brain connectivity*, 4(9):690–701.
- [Zhang et al., 2013] Zhang, G., Yao, L., Zhang, H., Long, Z., and Zhao, X. (2013). Improved Working Memory Performance through Self-Regulation of Dorsal Lateral Prefrontal Cortex Activation Using Real-Time fMRI. *PLoS ONE*, 8(8):1–9.
- [Zhang et al., 2015] Zhang, Q., Zhang, G., Yao, L., and Zhao, X. (2015). Impact of real-time fMRI working memory feedback training on the interactions between three core brain networks. *Frontiers in Behavioral Neuroscience*, 9(September):1–9.
- [Zhuang et al., 2005] Zhuang, J., LaConte, S., Peltier, S., Zhang, K., and Hu, X. (2005). Connectivity exploration with structural equation modeling: An fMRI study of bimanual motor coordination. *NeuroImage*, 25(2):462–470.
- [Zotev et al., 2011] Zotev, V., Krueger, F., Phillips, R., Alvarez, R. P., Simmons, W. K., Bellgowan, P., Drevets, W. C., and Bodurka, J. (2011). Self-Regulation of Amygdala Activation Using Real-Time fMRI Neurofeedback. *PLoS ONE*, 6(9):e24522.
- [Zotev et al., 2013] Zotev, V., Phillips, R., Young, K. D., Drevets, W. C., and Bodurka, J. (2013). Prefrontal Control of the Amygdala during Real-Time fMRI Neurofeedback Training of Emotion Regulation. *PLoS ONE*, 8(11):e79184.

Online References

- [1] BrainInnovation. BrainVoyager. <http://www.brainvoyager.com/products/brainvoyager.html>, 2016. [Last access 19-July-2016].
- [2] BrainInnovation. BrainVoyager Brain Tutor. <http://www.brainvoyager.com/products/braintutor.html>, 2016. [Last access 19-July-2016].
- [3] BrainInnovation. Motion Correction Processor Plugin for BrainVoyager. <http://support.brainvoyager.com/available-tools/49-available-plugins/491-motion-correction-processor.html>, 2016. [Last access 4-May-2016].
- [4] BrainInnovation. Turbo-BrainVoyager. <http://www.brainvoyager.com/products/turbobrainvoyager.html>, 2016. [Last access 19-July-2016].
- [5] BrainInnovation. Turbo-BrainVoyager User's Guide. <http://download.brainvoyager.com/tbv/TBVUsersGuide/TurboBrainVoyagerUsersGuide.html>, 2016. [Last access 19-July-2016].
- [6] MathWorks. MATLAB. <https://www.mathworks.com/products/matlab/>, 2016. [Last access 19-July-2016].
- [7] PubMed. PubMed. <https://www.ncbi.nlm.nih.gov/pubmed>, 2015. [Last access 19-Oct-2015].
- [8] SPM. SPM - Statistical Parametric Mapping. <http://www.fil.ion.ucl.ac.uk/spm/>, 2016. [Last access 19-July-2016].
- [9] Jochen Weber. NeuroElf v1.1. <http://neuroelf.net/>, 2016. [Last access 20-Jun-2016].