POLITECNICO DI MILANO
Corso di Laurea MAGISTRALE in Ingegneria Biomedica
Dipartimento di Elettronica e Informazione

Brain plasticity mechanisms underlying motor control reorganisation: a longitudinal study on post-stroke subjects

NEAR Lab
Neuroengineering and Medical Robotics

Relatore: Prof.ssa Alessandra Pedrocchi
Correlatore: Ing. Marta Gandolla

Tesi di Laurea di:
Lorenzo Niero, matricola 859230

Anno Accademico 2017-2018

Contents

List of Figures

List of Tables

Sommario

Introduzione

Questa tesi è incentrata sullo studio longitudinale dei pazienti post-ictus, per indagare il meccanismo di plasticità cerebrale alla base della riorganizzazione del controllo motorio in soggetti sottoposti ad un trattamento di riabilitazione motoria, supportato da stimolazione elettrica funzionale. Inoltre, uno degli obiettivi principali di questo lavoro è quello di ottenere maggiori informazioni sui fattori determinanti che portano al raggiungimento di un recupero motorio duraturo (chiamato effetto "carry-over") dopo la riabilitazione motoria basata su FES.
La FES (Functional Electrical Stimulation - stimolazione elettrica funzionale) dei muscoli degli arti superiori e inferiori ha ricevuto crescente attenzione poiché è stato dimostrato che la terapia riabilitativa basata su FES, combinata con il trattamento di terapia convenzionale, migliora in modo più efficace la capacità di deambulazione e migliora il recupero motorio rispetto alla sola terapia convenzionale, in pazienti sopravvisuti ad ictus [1].

FES ed effetto Carryover - Per quanto riguarda gli arti inferiori, la FES viene utilizzata principalmente per la correzione ortotica della sindrome da piede cadente, un problema piuttosto comune in seguito ad un ictus. Una percentuale di pazienti riapprende a dorsiflettere volontariamente la caviglia senza l'uso del dispositivo [2]. Il meccanismo di funzionamento di questo fenomeno, denominato "effetto carryover", non è stato ancora pienamente compreso e , infatti, l'effetto carryover è stato osservato solo in sottogruppi di pazienti neurologici e le caratteristiche di questi pazienti, con o senza effetto carryover, non sono chiare. Si ritiene che l'interazione tra sforzo volontario e stimolazione elettrica si traduca in un effetto neuroplastico, inducendo la formazione di nuove connessioni funzionali permanenti e, quindi, promuovendo il riapprendimento motorio [3][4][5][6]. Questo meccanismo segue i principi della teoria di Hebb, che afferma che una stimolazione ripetuta e persistente di una cellula postsinaptica da parte di una cellula presinaptica, favorisca l'aumento dell'efficacia sinaptica [7]. In effetti, la combinazione di sforzo volontario (segnale efferente) e percezione sensoriale di un movimento correttamente completato (segnale afferente) sembra facilitare la plasticità di Hebb.

Gli studi suggeriscono che la capacità di un paziente di pianificare il movimento e di percepire la stimolazione come parte del proprio ciclo di controllo motorio è molto importante perché l'effetto carryover abbia luogo [8].

Teoria dell'Active Inference - In questo studio, come ipotesi di lavoro, abbiamo impiegato la recente teoria che suggerisce che l'aggiornamento dei modelli interni del cervello segue i principi bayesiani, combinando distribuzioni di probabilità a priori e livelli noti di incertezza del feedback sensoriale con le conseguenze sensoriali del movimente [9]. Inoltre, la direzione interpretativa della tesi punta verso teoria dell'Active Inference come meccanismo di funzionamento del controllo motorio, che prevede che la corteccia motoria trasmetta previsioni discendenti delle conseguenze sensoriali del movimento piuttosto che i comandi cinematici di guida specificati dal controllo motorio ottimale. La possibilità di studiare i correlati neurali funzionalia i cambiamenti di connettività associati all'utilizzo della FES rappresenta un'utile condizione sperimentale per discriminare le due teorie alterantive "Optimal motor control" e "Active Inference", sulla base delle quali ci aspetteremmo che l'output del controllo motorio venga diversamente influenzato dalla presenza della FES (propriocezione alterata), cioè che i comandi cinematici rimangano inalterati nel caso del controllo ottimale, oppure che la predizione propiocettiva venga aggiornata nel caso dell'active inference.

Metodi

Partecipanti - I partecipanti coinvolti in questo studio consistono in 16 soggetti sani e 8 pazienti post-ictus. I pazienti sono stati sottoposti a una classificazione binaria, basata sull'esito del trattamento, a seconda che il soggetto abbia raggiunto o meno miglioramenti stabili (cioè l'effetto carryover) nel recupero motorio, dopo la riabilitazione [10].

In questo lavoro l'attenzione è stata focalizzata su soggetti che presentano la sindrome da piede cadente, detta anche paralisi flaccida, causata da debolezza alle caviglie e paralisi ai muscoli flessori dorsali del piede, e che impedisce ai soggetti che ne sono affetti di sollevare completamente la gamba o il piede mentre muovono un passo. Uno degli obiettivi di queste terapie è quindi l'uso della FES appropriatemente sincronizzato con la fase di oscillazione del passo per produrre una dorsiflessione pienamente funzionalecome, e anche come stimolo di propriocezione alterata. Questa procedura mira a stimolare i processi neuroplastici per la riorganizzazione del circuito di controllo motorio a livello corticale. L'esecuzione di un movimento volontario, infatti, richiede al cervello di integrare sia l'intenzione volontaria di eseguire il movimento, sia la conoscenza dello stato del corpo (cioè integrare il feedback sensoriale).

Il dataset e il portocollo sperimentale - Nella configurazione sperimentale, la FES viene utilizzata per fornire informazioni propriocettive pilotate dall'esterno durante l'esecuzione del movimento - in altre parole, per alterare sperimentalmente la re-afferenza (cioè l'input sensoriale che risulta dal movimento del soggetto). Il protocollo sperimentale è stato definito in modo tale da alterare solo la propriocezione in diverse condizioni, mantenendo costante la cinematica del movimento per rilevare quali aree e quali connessioni sono più sensibili all'alterazione della propriocezione. Il protocollo sperimentale originale [6][8], realizzato coninvolgendo la dorsiflessione della caviglia destra assistita da FES e a ritmo scandito da segnali acustici, consisteva in un design fattoriale 2×2, i cui fattori sperimentali erano rappresentati dall'intenione volontaria di eseguire il movimento (V) e la FES (F), ognuno dei quali aveva 2 livelli, cioè "volontario" / "passivo" e "presente" / "assente", rispettivamente. Questi comportano 4 condizioni: volontà del movimento con concomitante FES (FV), movimento passivo con concomitante FES (FP), volontà del movimento senza FES (V) e movimento passivo senza FES (P).
I dati di risonanza magnetica funzionale (fMRI) utilizzati in questo studio sono stati ottenuti durante sessioni di esercizio duranti le queli sia il movimento volontario che la propriocezione sono stati manipolati sperimentalmente e, così facendo, sono stati analizzati gli effetti sulle interazioni tra le aree corticali motorie e sensoriali, sia in soggetti sani che gruppi di pazienti post-ictus.
I dati utilizzati in questo lavoro sono relativi a 16 volumi cerebrali sani e 16 provenienti da pazienti post-ictus. In fatti, per ciascun paziente erano disponibili 2 sessioni di fMRI, acquisite una prima (PRE) e una dopo (POST) la riabilitazione.

DCM - Il lavoro in questa tesi si basa sulla struttura di lavoro del Dynamic Causal Modeling (DCM), implementato all'interno del software SPM di Matlab, come strumento per valutare la connettività effettiva tra regioni corticali del cervello. Oltre a fare inferenze sull'accoppiamento tra regioni cerebrali distinte, lo scopo del DCM è quello di esaminare come questo accoppiamento dipenda dal contesto sperimentale. Il DCM è usato per ricavare gli stati neuronali, dalla misura indiretta dell'attività cerebrale (cioè il segnale BOLD) all'interno di un imianto bayesiano [11]. Sono stati sviluppati diversi tipi di DCM, ma in questo studio è stato utilizzato un modello deterministico bilineare a stato singolo. DCM è un approccio per l'identificazione di sistemi di ingresso-stato-uscita non lineari; la richiesta di un input e la necessità di specificare le regioni del cervello di cui è composto il sistema, implicano che il DCM è tradizionalmente utilizzato per testare ipotesi specifiche che hanno motivato un particolare progetto sperimentale. Utilizzando un'approssimazione bilineare per le interazioni dynamiche tra gli stati,

$$\dot{z} = \left(A + \sum_j u_j B^{(j)} \right) z + Cu,$$

I paramteri del modello causale implicito si riducono a tre set: i parametri che mediano l'influenza degli input estrinsechi sugli stati del sistema (matrice C), i parametri che mediano le connessioni intrinseche tra gli stati (matrice A), e i paramtetri che permettono agli input di modulare queste connessioni (matrice B) L'identificazione procede seguendo una struttura bayesiana, dati degli input noti e deterministici, e la risposta del sistema misurata. Nello schema bayesiano di identificazione, le distribuzioni a posteriori dei parametri sono stimate attraverso un algoritmo di ottimizzazione, utilizzando i dati osservati e le distribuzioni a priori, sotto ipotesi di Gaussianità.

Gli Input - Per estrarre le serie temporali della risposta BOLD, relative a ingressi sperimentali opportunamente adattati per il nostro modello, abbiamo riformulato il modello lineare generale per specificare gli input diretti e modulatori; questi corrispondono a un'attività di stimolo che rappresenta l'effetto dei segnali volontari discendenti, V (che combina gli onset delle condizioni V e FV); un secondo input che codifica il contributo della stimolazione funzionale elettrica ascendente all input propriocettivo, E (che combina gli onset delle condizioni FV e FP) e un terzo input che rappresenta l'input propriocettivo dovuto ai movimenti, P (che combina gli onset di tutte le 4 condizioni). Per riassumere l'attività regionale di ciascun soggetto, per ciascun soggetto è stato eseguito un contrasto F tra tutte le covariate della nuova design matrix, ottenendo così un'immagine di contrasto relativa all'attivazione cerebrale per tutti e tre gli input sperimentali.

Scelta delle regioni di interesse (ROI) - Per quanto riguarda i nodi del modello, le regioni di interesse (ROI) sono state scelte in base agli obiettivi dello studio e alle conoscenze pregresse sul loro ruolo nel meccanismo di controllo motorio [6][8][12][13][14][15] [16][17] [18]. Le regioni di interesse risultanti, da cui sono state estratte le serie temporali del segnale BOLD, sono le aree motorie controlaterali primaria e supplementare (M1 e SMA), l'area somato-sensoriale controlaterale primaria (S1) e il giro angolare controlaterale (AG). I volumi di interesse specifici per ogni soggetto, associati alle regioni sopra menzionate, sono stati selezionati utilizzando sfere del raggio di 4 mm, centrate nelle coordinate di massima attivazione rilevate all'interno delle mappe di attivazione mascherate. Mascherare l'immagine di contrasto significa prendere in considerazione un numero limitato di voxel contenuti in uno specifico spazio volumetrico, definito dalla corrispondente maschera della regione anatomica/funzionale del cervello. Le maschere per ogni ROI sono state ricavate dall'atlante SPM.

Modello gerarchico e Regressori - Nel contesto del nostro studio su gruppi, per verificare se la connettività effettiva fosse diversa tra sottogruppi di pazienti, abbiamo deciso di sfruttare lo schema del Parametric Empirical Bayes (PEB), costruendo un modello gerarchico rispetto ai parametri, come descritto in studi recenti [19][20]. Oltre al modello bilineare al primo livello, che descrive l'evoluzione temporale degli stati neurali, questo approccio introduce un secondo livello che modella ogni parametro al primo livello (cioè i parametri di accoppiamento tra regioni corticali) come una variabile casuale che oscilla attorno a un valor medio di gruppo. Il modello di secondo livello è rappresentato da un modello lineare generale (GLM), che include una combinazione lineare di uno o più regressori (o covariate) e un rumore additivo gaussiano. I regressori sono compattati nella cosiddetta matrice di progettazione (X):

$$\theta_i = X\beta + \varepsilon$$

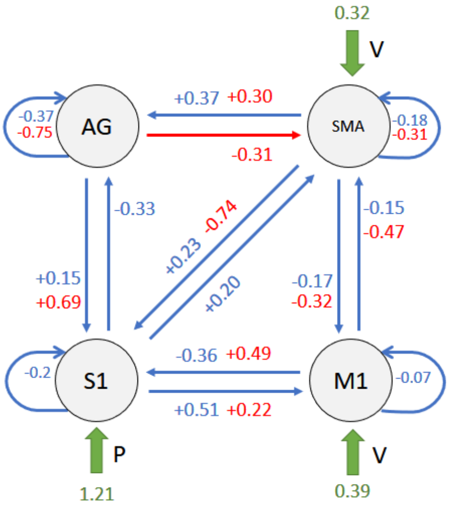
dove $\beta = \{\beta_1, \dots, \beta_n\}$ è il set di paramteri al seconod livello che esprime il contributo di ogni regressore inserito nella design matrix, e $\varepsilon \sim \mathcal{N}(0, \sigma)$.
Nell'analisi sui soggetti sani abbiamo utilizzato un singolo regressore che rappresenta la media di gruppo, poiché non sono stati ipotizzati altri effetti di gruppo all'interno del gruppo dei soggetti sani. Nella successiva analisi sui pazienti, eravamo interessati a osservare le differenze nella connettività effettiva tra 4 sottogruppi (CE-PRE, CE-POST, nCE-PRE, nCE-POST); questi sottogruppi derivano dal disegno fattoriale 2 × 2 ottenuto dalla combinazione del fattore di classificazione carry-over, con livelli "CE" (presente) e "nCE" (assente) e il fattore tempo, con livelli "PRE" (prima sessione fMRI) e "POST" (seconda sessione fMRI).

Al fine di evidenziare meglio le possibili differenze di connettività tra queste condizioni, abbiamo definito nella design matrix un regressore carry-over, che codifica l'appartenenza alla classe, e un regressore tempo, che codifica il punto temporale della sessione fMRI. Un ulteriore regressore che modella la media di gruppo è stato inserito nella matrice.

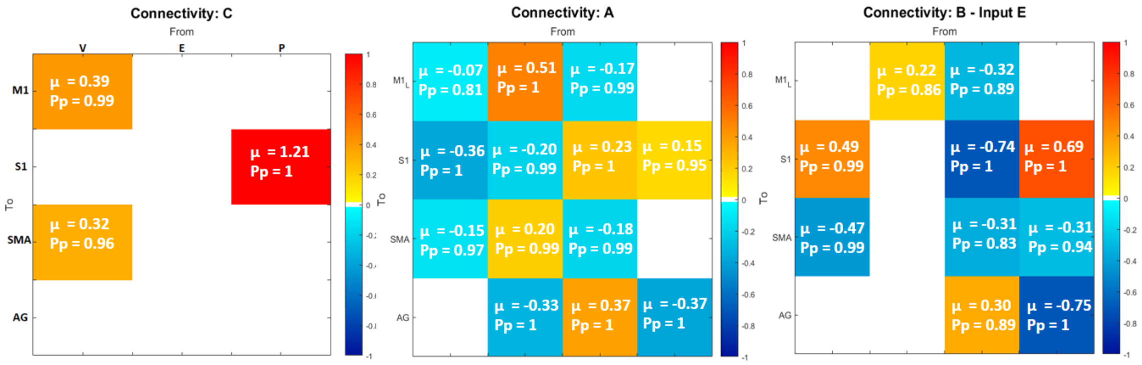
Procedure per l'identificazione della struttura del modello - Prima di identificare la struttura del modello, alcune semplificazioni sono state adottate, basate sulla connettività intrinseca e sugli input, in modo da ridurre lo spazio dei modelli. Questi vincoli sono basati su ipotesi derivate da conoscenze pregresse e informazioni trovate in letteratura, sulla connettività strutturale e funzionale. Abbiamo quindi scelto di adottare un approccio gerarchico, definendo prima uno spazio dei modelli realizzato facendo variare gli input (matrice C), poi utilizzando il modello vincente così selezionato per definire un nuovo spazio dei modelli ottenuto facendo variare le connessioni intrinseche (matrice A), ed infine utilizzando il modello così ottenuto per definire un ulteriore spazio dei modelli facendo variare gli effetti modulatori (amtrice B).
Questa procedura è stata applicata inizialmente al gruppo di soggetti sani. Nella successiva analisi sul gruppo di pazienti, si è ipotizzato che la matrice C avesse la stessa struttura di quella ottenuta per i soggetti sani; la matrice A è stata studiata a partire da dalla versione ridotta avente la stessa struttura di quella risultante dal gruppo dei soggetti sani. Nessun vincolo è stato imposto sulla matrice B e tutte le possibili combinazioni sono state testate. Questo approccio ci ha permesso di fare inferenza sulla struttura del modello, esplorando l'ampio spazio dei modelli in modo efficiente.

Risultati

I risultati riguardanti i oggetti sani sono riportati in **Figura 4.5**.



(a) Struttura di connettività del modello. Rappresentazione generale del modello risultante per soggetti sani; i valori blu sono relativi alle connessioni intrinseche nella matrice A, i valori rossi sono relativi agli effetti modulatori dell'input E (FES) nella matrice B, i valori verdi indicano gli input (V, E, P) nella matrice C.



(b) Rappresentazione matriciale. Medie a posteriori stimate μ e il loro livello di significatività P_p, riportate per ogni matrice di primo livello (da sinistra a destra: C - input; A - connettività intrinseca; B - effetti modulatori dell'input E).

Figura 1

L'inferenza sui parametri ha evidenziato una profonda interazione tra feedback sensoriale alterato artificialmente e movimento volontario, nei soggetti sani (connessioni $M1 \longrightarrow S1$ e $SMA \longrightarrow S1$ in **Figura 4**). Questa interazione è rispecchiata dagli effetti modulatori che l'input E ha sulle connessioni da M1 e SMA verso S1 (matrice B(2,1) e B(2,3) in **Figura 5**) e supporta la teoria dell'Active Inference per il controllo motorio, poiché questo effetto modulatorio indica un aggiornamento della previsione propriocettiva dovuta alla stimolazione elettrica applicata. In effetti, nel contesto del controllo ottimo, dato lo stesso output cinematico atteso in tutte le condizioni, i comandi cinematici di controllo non avrebbero bisogno di essere modulati dalla presenza di propriocezione alterata. Come si può vedere in **Figura 2**, nei pazienti i valori di connettività sono in contrasto con quanto osservato per i sani, e quindi questi valori evidenziano un malfunzionamento di questa interazione ($M1 \longrightarrow S1$, $SMA \longrightarrow S1$), che potrebbe indicare una propagazione difettosa della previsione propriocettiva inviata dalle aree motorie.
In **Figura 2** sono riportati la struttura del modello e le medie delle distribuzioni a posteriori stimate per i pazienti.
Si noti che ogni parametro di primo livello (cioè le connessioni) è espresso al secondo livello come la combinazione lineare di 4 componenti, associate ai quattro regressori inclusi nella matrice di progetto, ciascuno dei quali esprime un effetto di gruppo: medai di grupo, raggiungimento dell'effetto carryover (CE/nCE), tempo (PRE/POST). Ciò significa che i valori riportati in **Figure 2**, per ogni connessione, sono stati ottenuti come $\theta_i = \beta_{mean} \pm \beta_{carryover} \pm \beta_{time}$.

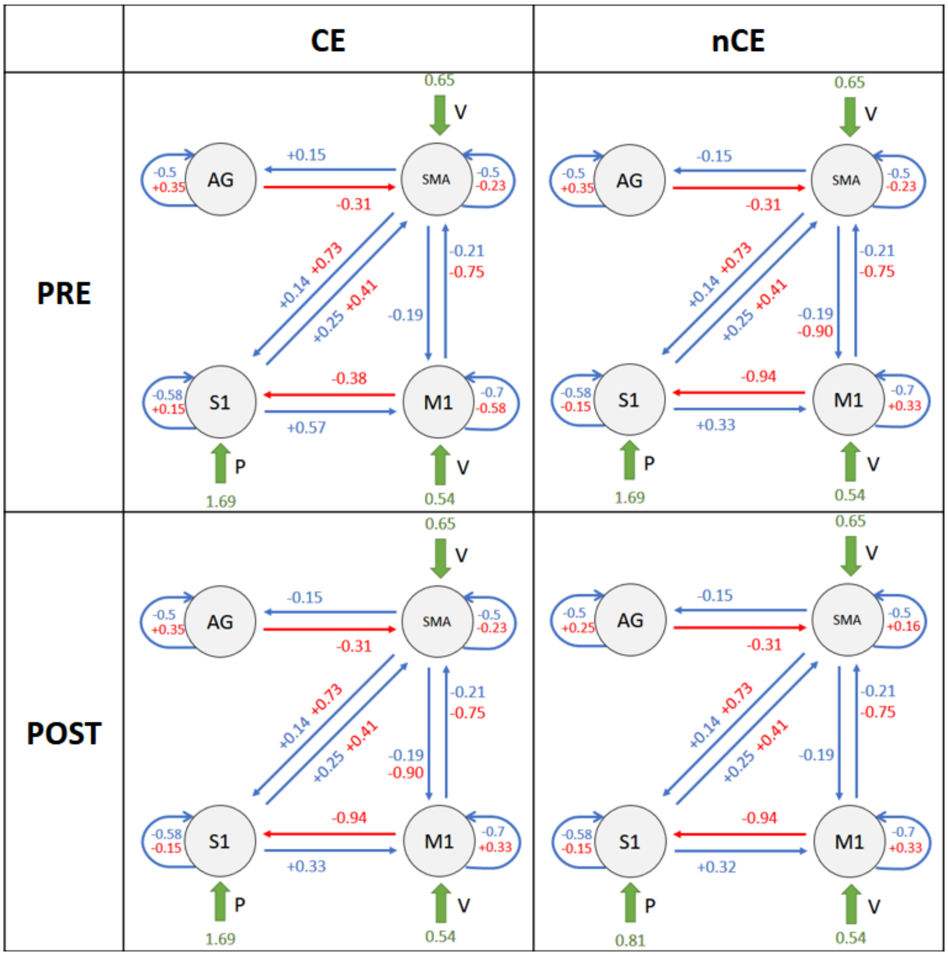
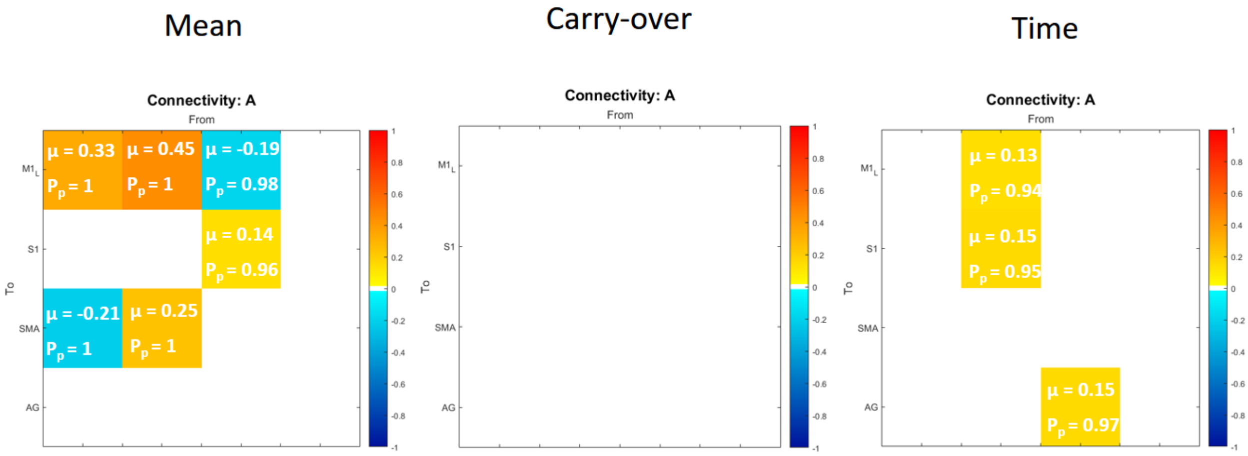


Figura 2: 2×2 factorial representation of patients model structure, along with estimated parameter posterior means. Values are obtined as $\theta_i = \beta_{mean} \pm \beta_{carryover} \pm \beta_{time}$. *Blue values are relative to intrinsic connections in matrix A, red values are relative to modulatory effects of input E (FES) in matrix B, green values indicate driving inputs (V, E, P) contribution in matrix C. Values are reported for each subgroup, localised as follows: in the upper left panel is the CE-PRE subgroup; in the upper right panel is the nCE-PRE subgroup; in the lower left panel is the CE-poST subgroup; in the lower right panel is the nCE-POST subgroup*

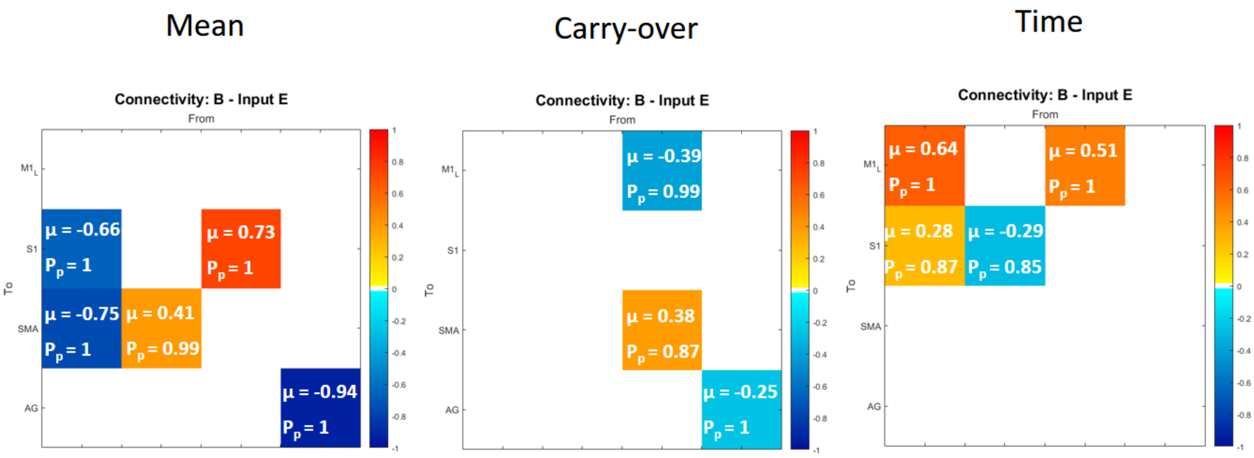
In **Figure 7**, i valori medi stimati (μ) dei parametri a secondo livello sono riportate per ogni parametro, insieme al loro livello di significatività (P_p), in rappresentazione matriciale; i livelli di colore indicano l'intensità del valore riortato.
Anche la natura gerarchica della elaborazione motoria è stata evidenziata dai risultati, principalmente sottolineata dal ruolo di SMA, che agisce sia come supervisore dell'area motoria primaria, sopprimendo i programmi motori

standard di M1 ogni qualvolta le condizioni esterne cambiano, sia come unità di elaborazione motoria di ordine superiore, che interviene qualora l'attività desiderata comporti movimenti più complessi. In questo contesto, suggeriamo che la dorsiflessione della caviglia, che sembra essere un movimento piuttosto semplice per un soggetto sano, potrebbe essere un compito impegnativo per i soggetti con disabilità, che quindi richiederebbe un maggiore contributo della SMA. A proposito di questa connessione ($SMA \longrightarrow M1$), abbiamo osservato differenze sia nel carry-over che nel tempo (matrice B(1,3) in **Figura 7b - pannelli *carryover* e *time***), indicando che potrebbe essere soggetta a cambiamenti dovuti a processi neuroplastici e, inoltre, che questa interazione potrebbe rappresentare un aspetto fondamentale e caratterizzante per l'insorgenza dell'effetto carry-over. Quest'ultimo risultato mostra come l'attività soppressiva di SMA nei confonti di M1, meccanismo che è proprio di un circuito di controllo motorio sano, sia preservata nei pazienti che esibiscono l'effetto carry-over ma venga a mancare nei pazienti senza l'effeto carry-over.

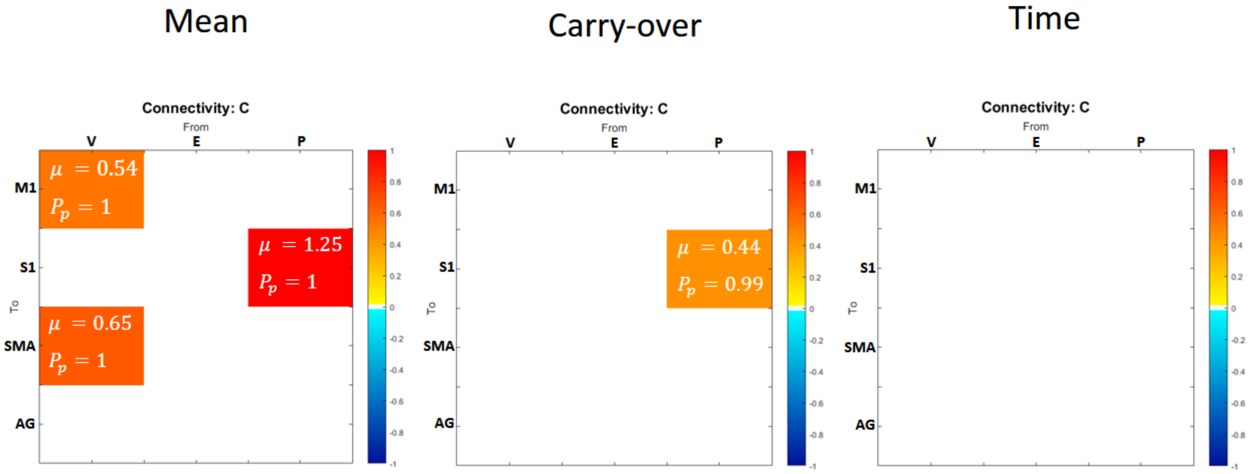
Un altro risultato interessante, che potrebbe rappresentare un potenziale marker dell'effetto carryover, è relativo all'input propriocettivo (P). Come mostrato in **Figura 7c - pannello *carry-over***, la componente relative all'effetto carry-over dell'input P entrante in S1 ha un valore stimato di 0.44. Questo vuol drie che i pazienti appartenenti al gruppo nCE hanno dimostrato di essere meno capaci di interiorizzare il segnale afferente sensoriale e di sfruttare queste informazioni propriocettive per integrarle nel circuito di controllo motorio. Ciò potrebbe rappresentare un potenziale punto di partenza per studi futuri su fattori chiave determinanti per il verificarsi dell'effetto carryover.



(a) Componenti matrice A



(b) Componenti matrice B



(c) Componenti matrice C

Figura 3: Valori stimati a secondo livllo, suddivisi per ognuna delle matrici a primo livello (A, B, C), relativi ai 3 regressori. Da sinistra a destra i pannelli riportano: media di gruppo (β_{mean}); effetto carryover ($\beta_{carryover}$); tempo (β_{time}).

Conclusioni

Per concludere, abbiamo presentato un approccio gerarchico personalizzato per l'identificazione della struttura di rete del modello. Abbiamo quindi analizzato e riportato i risultati dell'inferenza sui parametri del modello, per il nostro studio longitudinale di gruppo. Questi risultati supportano la teoria alternativa dell'Active Inference per il controllo motorio, forniscono alcune informazioni sui meccanismi di attenuazione sensoriale, hanno evidenziato la natura gerarchica del controllo motorio e infine potrebbero indicare potenziali meccanismi di azione della FES nel promuovere l'apprendimento motorio nella neuroriabilitazione.

Struttura della tesi

La tesi è strutturata come segue.
Il Capitolo 1 introduce il problema della menomazione post-ictus, fa' una rapida revisione della letteratura sullo stato dell'arte nel campo della riabilitazione motoria, introduce brevemente le procedure utilizzate durante il lavoro e dà un'idea degli obiettivi dello studio.
Il capitolo 2 offre una panoramica completa della teoria alla base della struttura del Dynamic Causal Modeling (DCM).
Il capitolo 3 presenta materiali e metodi utilizzati in questo lavoro.
Nel capitolo 4 sono presentati e discussi i risultati.
Il capitolo 5 riporta conclusioni sui risultati ottenuti, limiti dello studio e direzioni di ricerca future.

Summary

Introduction

Objective of the work and State of the art - This thesis is focused on the longitudinal study of post-stroke patients, to investigate the mechanism of brain plasticity underlying motor control reorganisation in subjects attending a motor rehabilitation treatment, supported by functional electrical stimulation (FES). Moreover, one of the major objectives of this work is to obtain more information about the determinant factors leading to the achievement of stable improvements (i.e. the "carry-over" effect) after FES-based motor rehabilitation.
FES of upper and lower limb muscles has been receiving increasing attention as it has been demonstrated that FES-based therapy combined with conventional therapy treatment more effectively improves the walking ability and enhances the motor recovery when compared with conventional therapy alone, in stroke survivors [1].

FES and Carryover - As for lower limbs, FES is primarily used for the orthotic correction of foot drop, a common issue following a stroke. A proportion of patients relearn the ability to voluntarily dorsiflex the ankle without the device [2]. The underlying functioning mechanism of this phenomenon, referred to as the "carryover effect", has not yet been fully understood and, in fact, the carryover effect has been observed only in subgroups of neurological patients and the characteristics of those with and without FES carryover are not clear. It is thought that this interaction between volitional effort and the electrical stimulation results in a neuroplastic effect, inducing long-lasting formation of new functional connections and, therefore, promoting motor re-learning [3] [4] [5] [6]. This mechanism follows the principles of Hebbian theory, which claims that an increase in synaptic efficacy arises from a presynaptic cell's repeated and persistent stimulation of a postsynaptic cell [7]. Indeed, the combination of volitional effort (efferent signal) and the sensory perception of a properly completed movement

(afferent signal) seems to facilitate Hebbian-like plasticity.
Studies suggest that the ability of a patient to plan the movement and to perceive the stimulation as a part of his/her own control loop is important for the FES carryover effect to take place [8].

Active Inference account - In this study we embraced as work hypothesis the recent theory suggesting that the updating of the internal models of the brain follows Bayesian principles, combining a priori probability distributions and known levels of uncertainty of sensory feedback with sensorial consequences [9]. Moreover, the interpretative direction of the thesis pointed toward the Active Inference account of motor control, which envisages that the motor cortex sends descending predictions of the sensory consequences of movement rather than the driving kinematics commands specified by optimal motor control. The possibility to study functional brain correlates and connectivity changes associated with the use of FES represent a useful experimental condition to discriminate between the two alternative theories of optimal motor control and Active Inference, based on which we would expect the motor control output to be differently influenced by the presence of FES (altered proprioception), i.e. unaltered kinematics commands in the case of optimal motor control, or updated propriocptive prediciton in case of Active Inference account.

Methods

Participants - Participants recruited for this study consisted of 16 healthy subjects and 8 post-stroke patients. Patients were subjected to a binary classification, based on the outcome of the treatment, depending whether or not the subject has achieved long lasting improvements (i.e. carry-over effect) in motor recovery after the rehabilitation [10].
In this work the attention has been focused on subjects exhibiting foot drop, or what is occasionally referred to as drop foot, caused by deficiency of control of the anterior muscles of the lower leg and preventing affected subjects from being able to fully lift their leg or foot while taking a step. One of the aim of these therapies is therefore the use of FES properly synchronised with the swing phase of gait in order to produce a full and functional dorsiflexion, as well as an altered proprioception stimulus. This procedure aims at stimulating the neuroplastic processes of motor control loop reorganization at cortical level. The execution of a voluntary movement, in fact, requires the brain to integrate both the volitional intention to execute the movement and the knowledge about the state of the body (i.e. integrate sensory feedback).

Dataset and experimental protocol - In the experimental setup, FES is used to provide externally driven proprioceptive information during movement execution - in other words, to experimentally alter reafference (i.e. the sensory input that results from the agent's movement). The experimental protocol was defined so that only proprioception in different conditions was altered, while maintaining constant movement kinematics to reveal which areas and which connections are sensible to proprioception alteration. The original experimental protocol [6][8], performed using FES assisted ankle dorsiflexion (right leg for healthy subjects, paretic leg for patients) paced with auditory cues, consisted in a 2×2 factorial design, which experimental factors included volitional intention to perform the movement (V) and FES (F), each of which had 2 levels, i.e. "voluntary"/"passive" and "present"/"absent", respectively. These translated into 4 conditions: will of movement with concurrent FES (FV), passive movement with concurrent FES (FP), will of movement without FES (V) and passive movement without FES (P).

Functional Magnetic Resonance Imaging (fMRI) data were acquired during these training sessions in which both volitional movement and proprioception were experimentally manipulated and, by doing so, the effects on the interactions between cortical motor and sensory areas were examined, in both healthy subjects and post-stroke patients groups. Data used in this work consisted in 16 healthy brain volumes and 16 brain volumes coming from 8 post-stroke patients. In fact, for each patient 2 sessions of fMRI were available, acquired one before (PRE) and one after (POST) the rehabilitation.

DCM - The work in this thesis is based on the Dynamic Causal Modelling (DCM) framework, implemented within the SPM software in Matlab, as a tool to assess effective connectivity among brain cortical regions. Besides making inferences about the coupling between distinct brain regions, the purpose of DCM is to examine how this coupling is dependent upon the experimental context. DCM is used to infer hidden neuronal states from the indirect measure of brain activity (i.e. the BOLD signal) within a Bayesian framework [11]. Different types of DCM have been developed, but in this study a deterministic single-state bilinear model has been employed. DCM is an approach to the identification of nonlinear input-state-output systems; the requirement of an input, and the need to specify the brain regions that the system is composed of, mean that DCM is traditionally used to test a specific hypothesis that motivated a particular experimental design.

By using a bilinear approximation to the dynamics of interactions among states

$$\dot{z} = \left(A + \sum_{j} u_j B^{(j)} \right) z + Cu$$

the parameters of the implicit causal model reduce to three sets: parameters that mediate the influence of extrinsic inputs on the states (matrix C), parameters that mediate intrinsic coupling among the states (matrix A), and parameters that allow the inputs to modulate that coupling (matrix B). Identification proceeds in a Bayesian framework given known inputs and the observed responses of the system.
In the Bayesian identification scheme, parameters posterior distributions are estimated through an optimisation algorithm, using observed data and prior distributions, under Gaussian assumptions.

Inputs - In order to extract BOLD response time-series related to experimental inputs conveniently suited for our model, we reformulated the general linear model to specify the driving and modulatory inputs; these are a stimulus function representing the effect of descending voluntary signals, V (combining onsets from V and FV conditions); a second input encoding the contribution of ascending functional electrical stimulation to proprioceptive input, E (combining onsets from FV and FP conditions), and a third input representing underlying proprioceptive input from all movements, P (combining onsets from all conditions). To summarise, the regional activity of each subject, an F-contrast was performed across all covariates of the new design matrix for each subject, obtaining a contrast image of brain activation for all three experimental inputs.

ROIs selection - Concerning the nodes of the model, regions of interest (ROIs) were chosen based on the objectives of the study and prior knowledge about their role in motor control [6][8][12][13][14][15] [16][17] [18]. The resulting regions of interest, from which BOLD time-series were extracted, were the contralateral primary and supplementary motor areas (M1 and SMA), the contralateral primary somato-sensory area (S1) and the contralateral angular gyrus (AG). Subject-specific volumes of interest, associated to the regions mentioned above, were selected using 4mm radius spheres, centred in the coordinates of maximum activation detected within the masked activation maps. Masking the contrast image means taking into consideration a limited number of voxels contained in a specific volumetric space, defined by the corresponding labeled mask of the anatomical/functional brain region.

Labeled masks for each ROI were taken from the SPM atlas of maximum probability tissue labels.

Hierarchical Model and Regressors - In the context of our group study, as to test whether effective connectivity is different between subgroups of patients, we decided to exploit the Parametric Empirical Bayes (PEB) scheme, constructing a hierarchical model over the parameters, as described in recent studies [19][20]. In addition to the bilinear model at the first level, which describes the neural states temporal evolution, this approach introduces a second level that models each parameter at the first level (i.e. coupling parameters between cortical regions) as a random variable oscillating around a group mean. Second level model is represented by a general linear model (GLM), including a linear combination of one or more regressors (or covariates) and an additive Gaussian noise. Regressors are compacted in the so called design matrix (X):

$$\theta_i = X\beta + \varepsilon$$

where $\beta = \{\beta_1, \ldots, \beta_n\}$ is the set of second level parameters that express the contribution of each regressor (or group effect) included in the design matrix, and $\varepsilon \sim \mathcal{N}(0, \sigma)$.

In healthy subjects analysis, we used a single regressor representing the group mean, as no group effects were assumed within the healthy group. In the subsequent analysis on patients, we were interested in observing differences in effective connectivity among subgroups, derived from the 2×2 factorial design given by the combination of carry-over classification factor, with levels ”CE” (present) and ”nCE” (absent), and time factor, with levels ”PRE” (first session) and ”POST” (second session). This way, 4 subgroups were identified: CE-PRE, CE-POST, nCE-PRE, nCE-POST.

In order to better highlight possible connectivity differences among these subgroups, we defined in the design matrix a carry-over regressor, encoding the class membership, and a time regressor, encoding the session time point. A regressor modelling the overall group mean was also included in the design matrix.

Model structure identification procedure - Before identifying the model structure some simplifications have been imposed based on intrinsic connectivity and inputs, to reduce the model space. These constraints are based on assumptions derived from prior knowledge and literature information about structural and functional connectivity.

Then, we have chosen to adopt a hierarchical approach, by first defining a model space of varying inputs (matrix C), then using the winning model to define a new model space of varying intrinsic connectivity (matrix A) and finally using the model thus obtained to define a further model space of varying modulatory effects (matrix B).

This was applied to healthy subjects group at first. In the subsequent analysis on patients group, C matrix was assumed to have the same structure as that obtained for healthy subjects; A matrix was investigated starting from a reduced matrix having the same structure as that resulting from healthy group. No constraints were imposed on the B matrix, and all possible combinations were tested. This approach allowed us to make inference on network structure, by exploring the wide model space in an efficient way.

Results

Results concerning healthy subjects are reported in **Figure 4** and **Figure 5**.

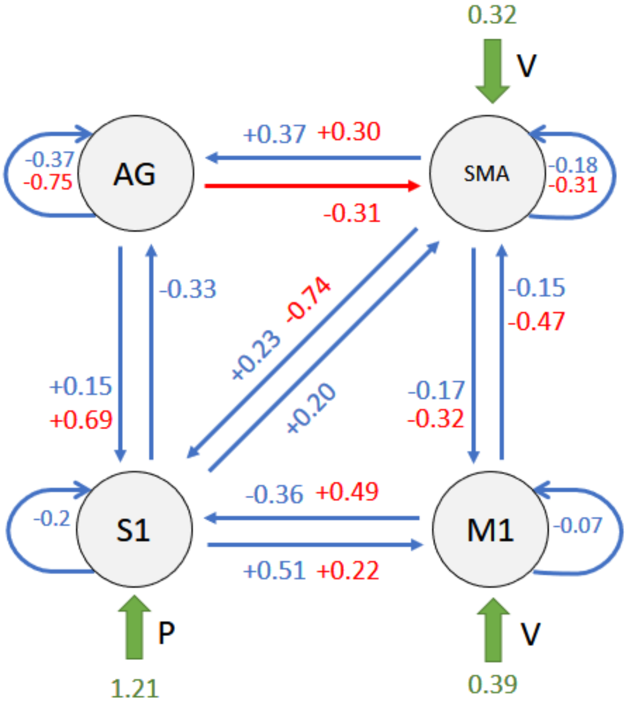


Figure 4: Model connectivity structure. Overall representation of final resulting model for healthy subjects; blue values are relative to intrinsic connections in matrix A, red values are relative to modulatory effects of input E (FES) in matrix B, green values indicate driving inputs (V, E, P) contribution in matrix C.

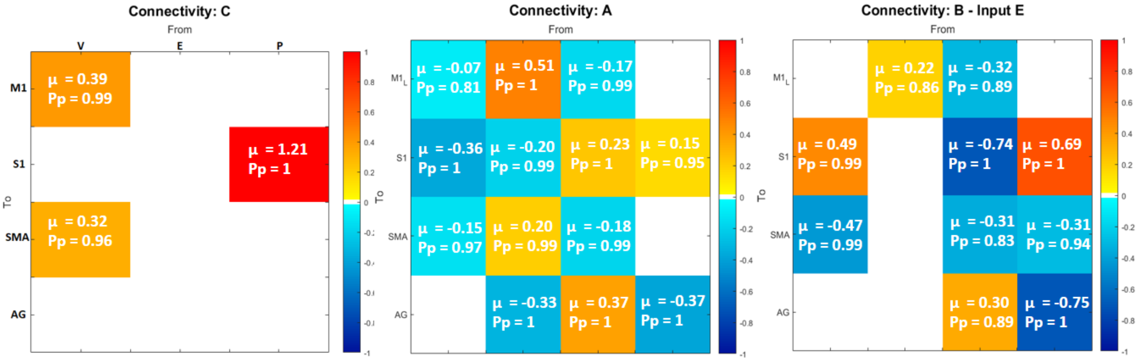


Figure 5: Matrix representation. Estimated posterior means μ and their significance level P_p, expressed for each first level matrix (from left to right: C - driving inputs; A - intrinsic connectivity; B - modulatory effects of input E).

Inference about parameters highlighted a profound interaction between artificially altered sensory feedback and volitional movement in healthy subjects ($M1 \longrightarrow S1$ and $SMA \longrightarrow S1$ in **Figure 4**). This interaction is reflected by modulatory effects of input E over connections from M1 and SMA toward S1 (matrix B(2,1) and B(2,3) in **Figure 5**) and supports the Active Inference account for motor control, as this modulation suggests an update of the proprioceptive prediction due to the applied electrical stimulation. Indeed, in the optimal control account, given the same expected kinematic output in all conditions, driving kinematics commands would not need to be modulated by the presence of altered proprioception.

As can be seen in **Figure 6**, patients connectivity values are in contrast with healthy ones, and thus they underlined a malfunctioning of this interaction ($M1 \longrightarrow S1$, $SMA \longrightarrow S1$), which could indicate a defective propagation of proprioceptive prediction sent by motor areas.

In **Figure 6** model structure and estimated posterior means for patients are reported. Note that, each first level parameter (i.e. connections) is expressed, at the second level, as the linear combination of 4 components, each associated with one of four regressors included in the design matrix, each of which expresses a group effect: group mean, carryover effect achievement (CE/nCE), time (PRE/POST). That means, values reported in **Figure 6**, for each connection, have been obtained as $\theta_i = \beta_{mean} \pm \beta_{carryover} \pm \beta_{time}$. In **Figure 7**, second level estimated mean values (μ) are reported for each parameter, along with the their significance level (P_p), in matrix representation; colour levels indicate the magnitude of the value.

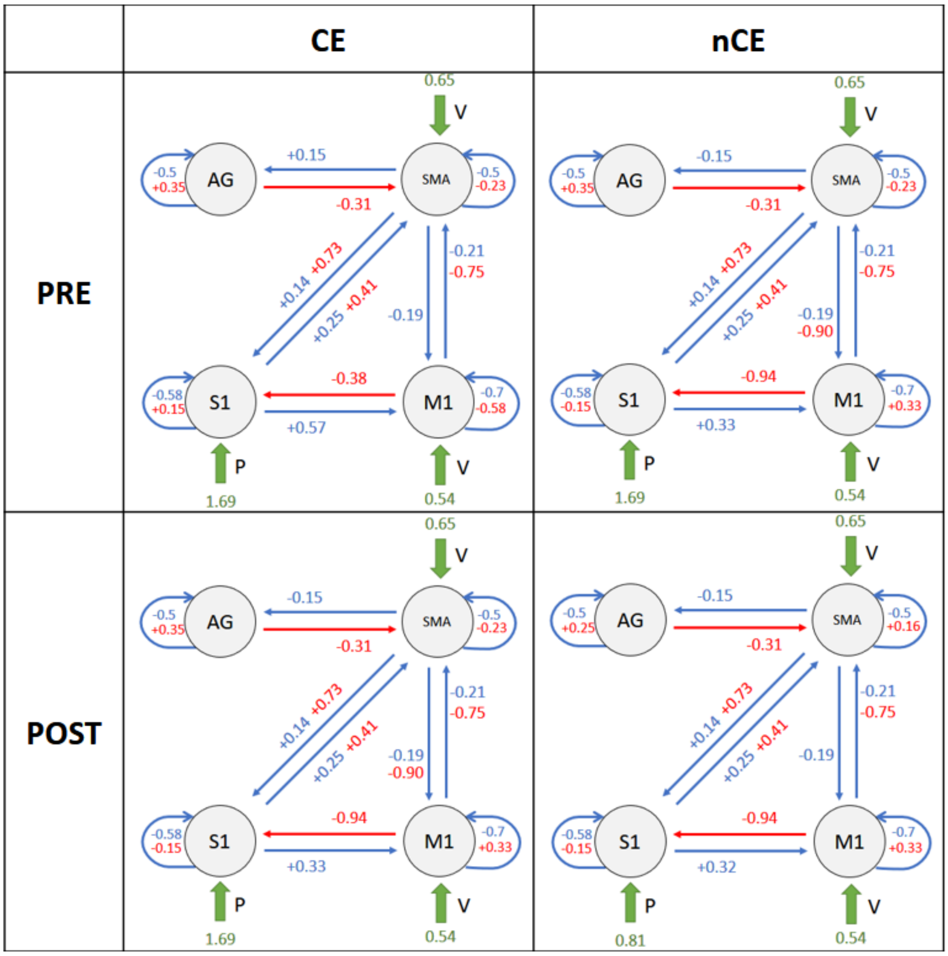
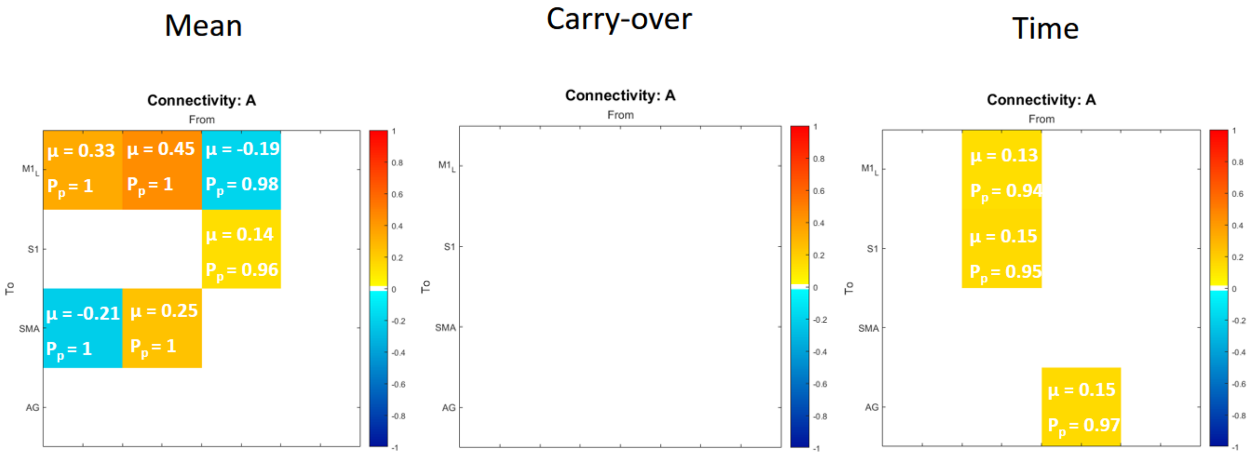


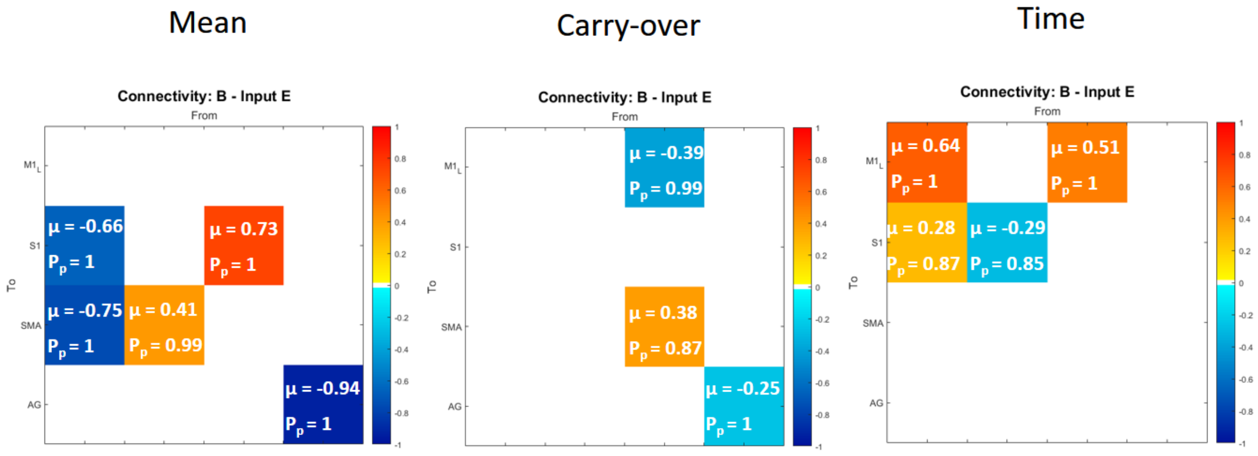
Figure 6: 2×2 factorial representation of patients model structure, along with estimated parameter posterior means. Values are obtined as $\theta_i = \beta_{mean} \pm \beta_{carryover} \pm \beta_{time}$. Blue values are relative to intrinsic connections in matrix A, red values are relative to modulatory effects of input E (FES) in matrix B, green values indicate driving inputs (V, E, P) contribution in matrix C. Values are reported for each subgroup, localised as follows: in the upper left panel is the CE-PRE subgroup; in the upper right panel is the nCE-PRE subgroup; in the lower left panel is the CE-POST subgroup; in the lower right panel is the nCE-POST subgroup.

The hierarchical nature of motor processing has also been highlighted, mostly by the role of SMA. This region, in fact, seems to act as higher order motor processing unit, intervening whenever the desired motor task involves more complex movements. In this context, we suggest that ankle dorsiflexion, that seems to be a fairly simple movement for a healthy subject, may be quite a challenging task for impaired subjects, requiring an increased contribution of SMA. Moreover, SMA showed a suppressive activity over primary motor area M1, in agreement with what is reported in literature [14][21][22], repressing standard motor programs in M1 whenever external condition change. About this connection ($SMA \longrightarrow M1$), we observed differences in both carry-over and time conditions (matrix B(1,3) in **Figure 7b - *carryover* and *time* panels**), indicating that it is subjected to changes due to neuroplastic processes and, furthermore, that this interaction could represent a fundamental and characterising aspect for the onset of carry-over effect. This latter result shows that the suppressive activity of SMA over M1, mechanism that is proper to a healthy motor control circuit, is preserved in patients exhibiting the carry-over effect but not in those without the carry-over effect.

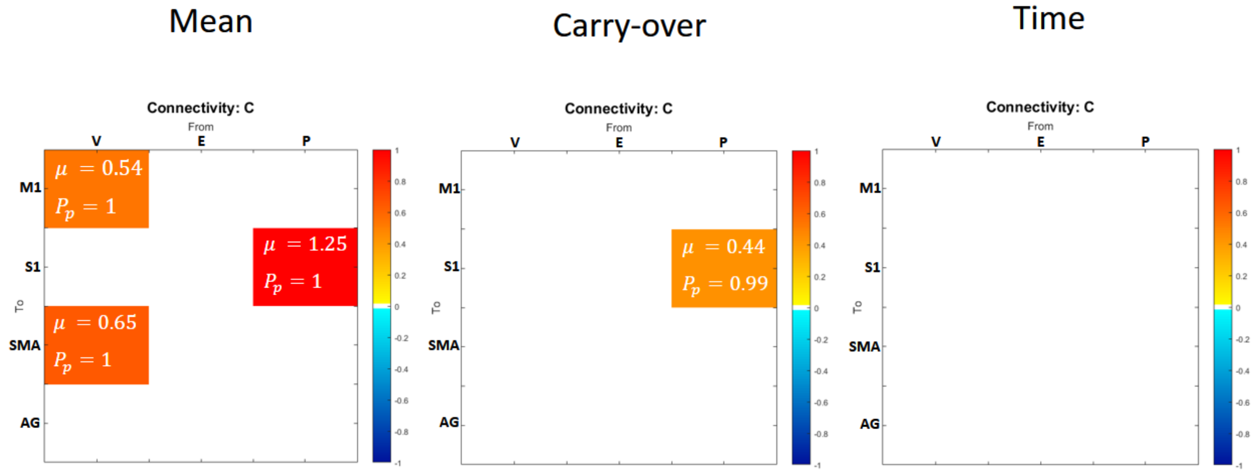
Another interesting result, which could represent a potential carry-over marker, is related to the proprioceptive driving input (P). As shown in **Figure 7c - element C(2,3) in *carry-over* panel**, the carry-over related component of input P entering in S1 has an estimated value of 0.44, meaning that patients belonging to the nCE group have shown to be less able to internalise the sensory afferent signal and exploit the proprioceptive information to integrate it into the motor control loop. This represents a potential starting point for future studies on key factors that determine the occurrence of the carry-over effect.



(a) A matrix components



(b) B matrix components



(c) C matrix components

Figure 7: Second level estimates (μ) and their significance level (P_p), subdivided for each of the first level matrices (A, B, C), relative to 3 regressors. From left to right panels represent: group mean (β_{mean}); carryover effect ($\beta_{carryover}$); time (β_{time}). Colour levels express the magnitude of the value, i.e., red tones are for positive values, blue tones are for negative values

Conclusions

To conclude, we presented a customised hierarchical approach for the identification of network structure. We then analysed and reported results about inference on model parameters, for our longitudinal group study. These results support the alternative theory of Active Inference account for motor control, provide some insight into the mechanisms of sensory attenuation and may speak to potential mechanisms of action of FES in promoting motor learning in neurorehabilitation.

Structure of the Thesis

The thesis is structured as follows.
Chapter 1 introduces the problem of post-stroke impairment, outlines a quick review of literature about the state of the art in the field of motor rehabilitation, briefly introduces the procedures used during the work and outlines the objectives of the study.
Chapter 2 gives a comprehensive overview of the theory behind Dynamic Causal Modelling (DCM) framework.
Chapter 3 presents materials and methods used in this work.
In Chapter 4 results are presented and discussed.
Chapter 5 reports conclusions about obtained results, limitations of the study and future research directions.

Chapter 1

Introduction

Stroke is one of the leading causes of long-term disability in adults worldwide [23]. In fact, as most patients survive the initial injury, the biggest effect on patients and families is usually through long-term impairment, limitation of activities and reduced participation. The most common clinical deficit and widely recognised impairment caused by stroke is motor impairment, which can be caused by ischaemic or haemorrhagic injury to the motor cortex, pre-motor cortex, motor tracts, or associated pathways in the cerebrum or cerebellum, and can be regarded as a loss or limitation of function in muscle control or movement or a limitation in mobility [24]. After-stroke motor impairment may include hemiparesis, incoordination and spasticity and typically affects the control of movement of the face, arm, and leg of one side of the body [25], contralateral to the stroke lesion side. Therefore, much of the focus of stroke rehabilitation, and in particular the work of physiotherapists and occupational therapists, is aimed to the recovery of impaired movement and the associated functions, throughout the recovery of self-performed daily life activities.

Stroke patients show varying degrees and types of neurological deficits, that depend on size and location of the brain lesion [26], and whose recovery strongly depends upon the type of rehabilitation program chosen and perforomed and the degree of commitment of the patients and family members. Most of them recover at least some of their lost motor functions over time, though the degree of this recovery is highly variable, depending on the severity of the damage and the type and intensity of rehabilitation therapy. There is evidence that increasing the intensity and repetition of post-stroke therapy can enhance motor recovery [27]. Brain mapping studies in patients have revealed that the brain reorganizes after stroke in relation to recovery of motor function. In recent years, convincing evidence has been produced in neuroscience that have led to the knowledge that the brain can change or reorganize itself in response to sensory input, experience and learning [28]. This ability of the brain and other parts of the central nervous system to reorganize itself is referred to as Neuroplasticity [29] and exclusively of cortex as Cortical plasticity [30]. Neuroplasticity occurs in both a healthy and injured brain [31]: it is the basic mechanism of learning processes, in which the nervous system change its structure and its function over a lifetime, in reaction to environmental diversity, experience and repetition, triggering the dynamic destruction and reformation of new synaptic connections between neurons. Basically, brain plasticity is the ability for the brain to recover and restructure itself. This adaptive potential of the nervous system also allows the brain to recover after disorders or injuries. Therefore, clinical improvements of motor function after stroke are accompanied by profound functional reorganization within motor areas of both hemispheres. It is believed that appropriately boosting these neuroplastic processes may restore function via recruitment of spared areas and pathways.

Traditional neurorehabilitation approaches for post stroke patients are oriented towards motor and functional recovery [32]; motor recovery refers to the ability of the patient to execute movements under voluntary control achieving eventually the same performances as before the stroke [33], while functional recovery refers to the improvement of the individual's ability to autonomously perform activities such as self-care and mobility [34]. Post stroke recovery also occurs at brain level, referred to as neurological recovery, which is generally associated with the structural and functional reorganization of brain. Both motor and functional recoveries are conditioned by neurological recovery. Accordingly, neural reorganization after stroke is thought to be an important mechanism to facilitate motor recovery.

Various novel stroke rehabilitative methods for motor recovery have been developed based on basic science and clinical studies characterizing brain remodeling due to neural plasticity [35] [36]. In addition to the more classic rehabilitation techniques, other methods include Selective Serotonin Reuptake Inhibitor Medications (SSRI), Constraint-Induced Movement Therapy (CIMT), Noninvasive Brain Stimulation such as Trans-cranial Magnetic Stimulation (TMS), Mirror Therapy, Robot-assisted Therapy [37] [38].

One commonly adopted therapy in the rehabilitation of stroke is Functional Electrical Stimulation (FES) [39]. FES of upper and lower limb muscles has been receiving increasing attention as a therapeutic modality in post-stroke rehabilitation. A meta-analysis of controlled studies supported the conclusion that FES promotes the recovery of muscle strength after stroke, with a reasonable likelihood of clinically significant results [1].

In this work the attention has been focused on subjects exhibiting foot drop, or what is occasionally referred to as drop foot, a common issue following a stroke or other neurological injury. Foot drop is caused by interruptions in the common fibular and sciatic nerves and aggravated by the paralysis of the anterior (front) muscles of the lower leg. Those experiencing foot drop are often unable to fully lift their leg or foot while taking a step forward. Due to this weakness, the knee compensates, bending more deeply and lifting the leg higher off the ground to avoid dragging or "dropping". This compensation creates what is referred to as a steppage gait, that can then cause exaggerated flexion at the hip and knee. While this prevents the foot from dropping, it also creates an awkward ground reaction vector on the joints and subsequent slap of the foot as the pressure of impact is diffused by the body. There could be a problem of maladaptive plasticity as well, i.e. the motor re-learning mechanism produce a neural reorganization compensating for this emi-paresis, but leading to lasting defective movements. Men and women with foot drop complications are often unable to perform complete dorsiflexion (the flexing of the foot toward the body).

It has been demonstarted that FES therapy combined with conventional therapy treatment more effectively improves the walking ability and enhances the motor recovery when compared with conventional therapy alone in stroke survivors [40]. Externally induced dorsiflexion using FES was first introduced by Liberson et al. in 1961 as an alternative treatment for foot-drop [2]. The application of the peroneal nerve stimulation (PNS) has both a positive therapeutic and a positive orthotic effect in improving speed and reduced effort while walking [41] [42].

In the experimental setup, functional electrical stimulation (FES) is used to provide externally driven proprioceptive information during movement execution – in other words, to experimentally alter reafference (i.e. the sensory input that results from the agent's movement). FES delivered to a nerve tract containing both efferent motor and afferent sensory fibers, will synchronously depolarize motor and sensory axons that are bundled together, eliciting muscle contraction through two pathways. The first (direct descending pathway) conveys signals along the efferent motor fibers that generate muscle contraction by direct motoneuron depolarization. The second (indirect ascending pathway) communicates signals via the afferent sensory fibers [43] that code proprioceptive signals from muscle spindles, Golgi tendon organs and cutaneous receptors [44], but in particular Ia fibers responsible for muscle spindle information [45]. This second pathway produces muscle contractions through a central mechanism, providing excitatory synaptic input to spinal neurons that recruit motor units in the natural order [46]. Therefore, the proprioceptive signal elicited by the sensory fiber stimulation creates the impression that the muscle is extended (i.e. muscle spindles discharge), and leads to firing of the motor neurons in order to produce a contraction. During FES, it has been demonstrated that this information can be useful at the level of the spinal cord, inducing a reinforcement of the muscle contraction through the myotatic reflex circuit, however few notions about altered proprioceptive information sent up to the cortex are available in the literature.

FES is primarily used for the orthotic correction of foot drop, but a proportion of patients relearn the ability to voluntarily dorsiflex the ankle without the device [2]. This phenomenon, referred to as the "carryover effect", has been observed in a number of subsequent studies [47] [48]. The functioning mechanism of this effect has not yet been fully understood, although it has been hypothesised that an interaction between volitional effort and the electrical stimulation of FES results in a neuroplastic effect on the central nervous system [3] [4] [5] [6].

Indeed a recent study [8] suggests that the mechanism through which FES carryover take place is based on movement prediction together with sense of body ownership. In other words, the ability of a patient to plan the movement and to perceive the stimulation as a part of his/her own control loop is important for the FES carryover effect to take place. However, the carryover effect has been observed only in subgroups of neurological patients and the characteristics of those with and without FES carryover are not clear. In patients showing FES carryover, the execution of the movement with concurrent volitional intention and FES allow them to correctly plan the movement and to perceive it as self-generated. By doing so, the motor control loop correctly updates itself [49], showing a long-lasting formation/strengthening of new functional connections, following Hebbian principles.

Hebbian theory claims that an increase in synaptic efficacy arises from a presynaptic cell's repeated and persistent stimulation of a postsynaptic cell. It is an attempt to explain synaptic plasticity, the adaptation of brain neurons during the learning process [7]. Indeed, the combination of volitional effort and the sensorial perception of a properly completed movement provides somatosensory feedback that facilitates Hebbian-like plasticity [50]. The aim of these therapies is therefore the use FES as stimulus of altered proprioception, during task-oriented rehabilitation procedures, in order to stimulate neuroplastic processes for motor control loop reorganization at cortical level.

The execution of a voluntary movement in fact requires the brain to integrate both the volitional intention to execute the movement and the knowledge about the state of the body (i.e. integrate sensory feedback). In humans, changing proprioceptive input influences motor cortex excitability [51] [52]. Conversely, the response of somatosensory cortex neurons to proprioception is modified by the nature of the motor task [53]. Currently, motor control theory proposes that internal models generate motor commands that are sent to the periphery to produce the desired movement. In this account, internal models combine sensory inputs, prior knowledge and volitional intention to produce motor commands [54]. Forward models are thought to be responsible for predicting the sensory consequences of action, given the motor commands [55] [56]. It has been recently suggested that the updating of the internal model follows Bayesian principles [54], combining a priori probability distributions and known levels of uncertainty of sensory feedback with sensorial consequences [57].

However, an alternative account of motor control has been proposed, drawing on the hierarchical generative models used in perceptual and active inference [58]. In this account, motor cortex sends descending predictions of the sensory consequences of movement rather than the driving commands specified by optimal motor control. Here, proprioceptive prediction errors are generated at the level of the spinal cord and result in activation of motor neurons through classical reflex arcs. Although there are commonalities between the two accounts, the key difference is that under optimal control, given a same task and a same state of the system, motor signals are context independent commands, whereas under active inference they are context-dependent predictions [59].

In order to disambiguate between these two theoretical accounts of motor control, a recent work, performed at Nearlab with the collaboration of Villa Beretta Rehabilitation Center and the Institute of Neurology (UCL, London) [6], investigated the effect of altered signals from the primary motor cortex, by experimentally manipulating both volitional movement and proprioception (i.e. context) and examining the effects on the interactions between cortical motor and sensory areas. It has been defined an experimental protocol that only alters proprioception in different conditions while maintaining constant movement kinematics to reveal which areas and which connections are sensible to proprioception alteration. This study has highlighted that M1 and S1 exhibit a profound interaction between artificially altered sensory feedback and volitional movement. Changes in coupling between these regions support an active inference account of motor control, in which sensorimotor integration rests upon the context-sensitive assimilation of descending motor predictions.

In the context of the same aforementioned study, using the same experimental protocol and considering further cortical regions in addition to M1 and S1, the purpose of this thesis was therefore to investigate the overall functioning mechanisms of the motor control loop within the brain of healthy subjects and, conversely, of neurological patients, in order to corroborate the alternative theory of the active inference account of motor control with particular attention to the investigation of the carryover effect. Specifically, through the analysis of time series extracted from functional magnetic resonance images (fMRI) taken during FES therapy sessions, has been tried to assess how and where altered proprioception information interact with the intentional movement, how cortical brain regions interact and influence each other and how this coupling/effective connectivity is influenced by altered proprioception.

Brain connectivity is a concept that can be defined at different spatial scales, but procedures and models used in this study deals with the macro scale aspects that are studied in neuroimaging, as opposed to micro scale connections between individual neurons. Connectivity, when discussing the whole brain, can refer to three different concepts regarding brain organisation and function, that are anatomical, functional, and effective connectivity. *Structural connectivity* refers to networks in the brain, formed by physical connections between neurons, neural populations or anatomically segregated brain regions. The physical connections can be formed by synapses between neurons, or white matter fibre pathways between neural populations. Physical pathways are relatively stable over short time periods, but due to neural plasticity significant morphological changes can occur over longer time periods. In the field of neuroimaging, *functional connectivity* is defined as temporal correlations between neurological events and it is an observable phenomenon, i.e. correlations in BOLD (Blood Oxygantion Level Dependent) signal between spatially separated brain regions, and so measuring it does not require a model.
Effective connectivity, instead, is related to causality and directionality, and it attempts to explain these correlations by way of some model explaining how they arise. The parameters of such a model are said to be the effective connections. Therefore, effective connectivity methods, that attempt to make inferences about causation and directed influences between regions, usually require a model that attempts to describe the causal influence between regions exhibiting some degree of functional interaction.

Crucially this work was aimed to better understand the underlying mechanism and key features of the carryover effect, by evaluating differences between groups of patients showing or not a FES carryover. The presence/absence of a long-term recovery of motor functions (i.e. carryover effect), in patients following the rehabilitation program, was determined using a quantitative comprehensive method proposed in [10] that evaluates functional improvement on a single-subject yes/no base combining multiple measures (e.g., kinematic, muscular) in order to assess a stable functional improvement and it is validated against clinical evaluation.

In this context the Dynamic Causal Modelling (DCM) methods have been used to model interactions among neuronal populations at cortical level and to infer directed connectivity among brain regions and the modulatory effect of externally driven electrical input over these couplings. DCM is a technique for the characterization of effective connectivity within networks of distributed neuronal responses [60] [61], as measured with fMRI [11] or electromagnetic responses as detectd by EEG or MEG [62]. This technique is based on the Bayesian inference framework which integrate prior knowledge and observations to estimate parameters of the posterior distributions.
DCM is an approach aimed at the identification of nonlinear input-state-output systems. By using a bilinear approximation to the dynamics of interactions among states, the parameters of the implicit causal model reduce to three sets. These comprise (i) parameters that mediate the influence of extrinsic inputs on the states, (ii) parameters that mediate intrinsic coupling among the states, and (iii) parameters that allow the inputs to modulate that coupling. Identification proceeds in a Bayesian framework given known deterministic inputs and the observed responses of the system. In the context of the analysis of effective connectivity using experimentally designed inputs and fMRI responses, the coupling parameters correspond to effective connectivity and the bilinear parameters reflect the changes in connectivity induced by inputs (in this case FES).
A number of previous studies used the framework of Dynamic Causal Modeling to model effective connectivity between brain areas involved in the motor control mechanism [6] [63] [64] [65] [14]. Similarly, in this work functional magnetic resonance images have been used, of subjects undergoing experimentally designed task-oriented activities. Following the Statistical Parametric Mapping (SPM) paradigm, after undergoing a pre-processing step and a statistical analysis, the Blood-Oxygen-Level-Dependent (BOLD) images resulted in activation maps relative to defined inputs, from which the time series of activation of neuronal clusters have been extracted to characterize the desired Regions Of Interest (ROI) and used in the subsequent modeling stage.

From the analysis of estimated parameters of DCM models we would be able to discriminate the behaviour of patients showing or not the carryover effect and to highlight any changes in effective connectivity throughout the rehabilitation program. In other words the scientific question to which we tried to answer is: how does the motor control loop update itself in a successful motor recovery and what goes wrong when the rehabilitation treatment does not show significant motor improvements ?

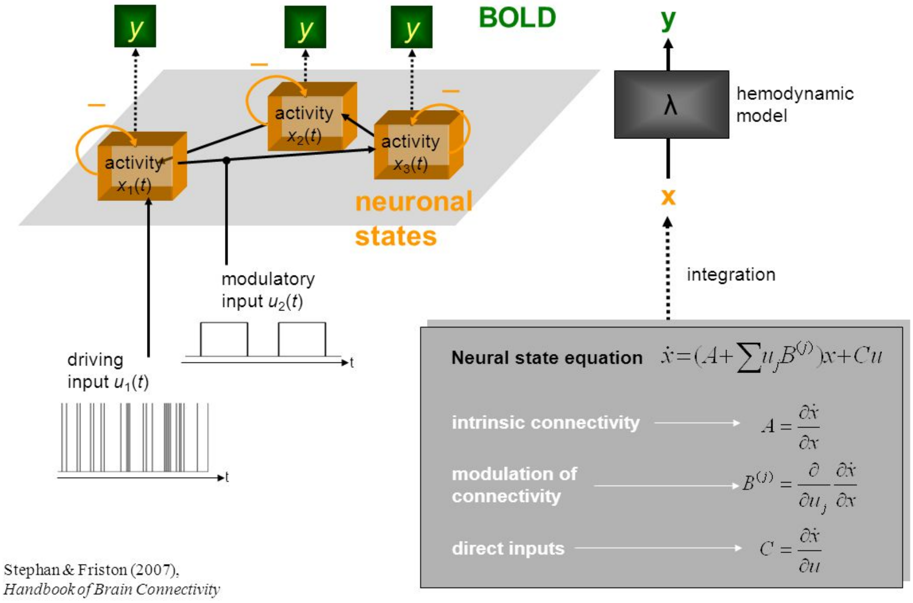
The thesis is structured as follows.
In Chapter 2 a brief background about stroke symptoms and causes is given, along with a discussion about the principles behind neural motor control and the rehabilitation after stroke.
Chapter 3 provides the theoretical background to functional magnetic resonance imaging based on the underlying principles of nuclear magnetic resonance, the localised change in blood flow that accompanies neural activity, and the corresponding signal that is measured; then data processing and the statistical analysis procedures for the creation of activation maps are described.
Chapter 4 gives a comprehensive overview of the theory behind Dynamic Causal Modelling (DCM) framework.
Chapter 5 presents materials and methods used in this work.
In Chapter 6 results are presented and discussed.
Chapter 7 outlines conclusions about obtained results, limitations of the study and future research directions.

Chapter 2

Dynamic Causal Modeling

2.1 Introduction – DCM for fMRI

Dynamic Causal Modelling (DCM) was first introduced for fMRI data by Friston and colleagues in 2003 [11], and integrated into the open-source Statistical Parametric Mapping (SPM) software running in Matlab. DCM was introduced as a way of inferring effective connectivity and it is fundamentally different from previously employed methods in that it was in-



Stephan & Friston (2007),
Handbook of Brain Connectivity

Figure 2.1: A schematic representation summarising the meaning of parameters in the bilinear equation and the forward mapping of neural states into BOLD response

vented specifically for analysing functional brain imaging data. DCM uses an input-state-output model, a concept defined well before the advent of neuroimaging, that however, in this case has been adpated to this particular purpose. DCM is used for electroencephalography (EEG) and magnetoencephalography (MEG) as well as fMRI, but this thesis focuses on the use of DCM for fMRI. The purpose of DCM is to make inferences about the coupling between distinct brain regions, and to examine how this coupling is dependent upon the experimental context. This means it requires a biologically plausible model of measured brain responses, which is both dynamic and non-linear in nature. DCM is used to infer hidden neuronal states from measured brain activity, in this case from the BOLD signal, within a Bayesian framework. A schematic representation of the overall principles of DCM is showen in **Figure 2.1**.

Numerous types of DCM have been developed, but all of them are based on the following characteristics [66, 67]:

- The idea of DCM is to construct realistic model of interacting cortical regions, with a system of differential equations.
- This neural model is then supplemented with a forward model of how the synaptic activity within these cortical regions translates to the measured response (BOLD in the case of fMRI).
- Inversion of the model based on Bayesian statistics, allows the parameters of the neuronal model of interacting cortical regions to be estimated from the data to give a measure of the effective connectivity.

Traditional DCM treats the brain as a deterministic dynamical system of interacting brain regions which can have several inputs, and treats an experiment as a designed perturbation of the system's dynamics. The inputs to the system are the usual stimulus functions that reflect the experimental design which are used in basic general linear model (GLM) methods. In this original format, bilinear differential equations are used to model the system, with the bilinear term representing context dependent modulation of effective connectivity. Since then DCM has been extended to allow for neuro-physiological phenomena that are considered important. Three major extensions to DCM are listed:

- Non-linear DCM [68] attempts to model how connectivity between two regions may be dependent on connectivity in another region, a process that is caused by synaptic interactions and that has been established though invasive electrophysiological experiments.

- Two-state model DCM [69] allows regions to have more than one state, i.e. modelling within region connectivity between excitatory and inhibitory neuronal populations.
- Stochastic DCM (sDCM) [70, 71], which allows stochastic inputs or error terms, and thus can be applied to data in the absence of experimental manipulation such as resting-state data.

Because in this work a deterministic bilinear single-state DCM has been used, the different types of DCM mentioned above will not be further discussed, and all references to DCM should be assumed to refer to this. Bilinear DCM requires direct inputs as it treats the brain as a dynamical system of coupled neuronal regions, in which the experiment is a designed perturbation of this system. In this respect, it is different from established methods of connectivity such as SEM and other multivariate autoregressive processes, in which there is no designed perturbation, where the inputs are treated as stochastic and unknown. The requirement of an input, and the need to specify the brain regions that the system is composed of, mean that DCM is traditionally used to test a specific hypothesis that motivated a particular experimental design, and therefore is not used as an exploratory technique as are other analyses of effective connectivity. The experimental inputs to a DCM and how they enter the model are an important aspect of the technique and form the basis of its ability to infer direct causal interactions between regions, i.e. effective connectivity. Since its inception, DCM has been widely adopted by the fMRI neuroimaging community and has been used to probe a variety of cognitive and neurophysiological questions [67].

2.2 Neuronal state equations

Neuronal state equations are the basis of all variants of DCM, and are known as "generative models", in that they provide a model of interaction of neuronal regions, from which the observed data were generated [60].

The original variant of DCM is based on a bilinear model of neural activity and it is the one used exclusively in this thesis. Given any number of brain regions with neuronal states $z = [z_1, \dots, z_N]$, an arbitrary model can be created, with inputs u and parameters θ, of the effective connectivity between these regions:

$$\dot{z} = f(z, u, \theta) \tag{2.1}$$

A simple truncated Taylor expansion around the system's resting state ($z_0 = 0$, $u_0 = 0$) provides an approximation to the function that is the bilinear

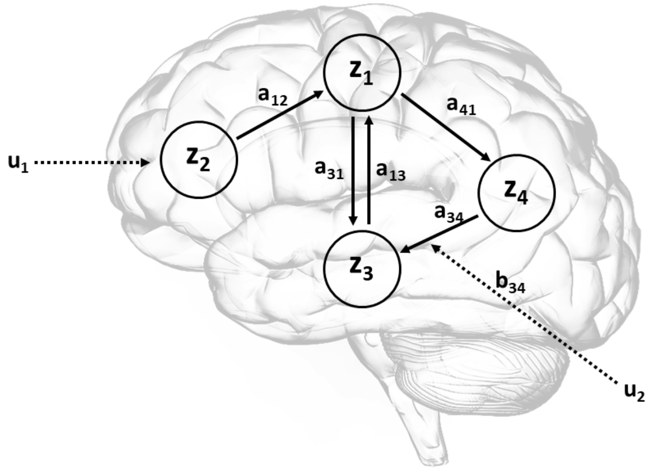


Figure 2.2: A schematic showing a system of connected brain regions, having one input that influence the system directly (u_1) and one that modulates connections between regions (u_2).

state equation

$$f(z,u) \approx f(z_0,u_0) + \frac{\partial f}{\partial z}z + \frac{\partial f}{\partial u}u + \frac{\partial^2 f}{\partial z \partial u}zu \quad \Longrightarrow$$

$$\dot{z} = (A + \sum u_j B^j)z + Cu, \tag{2.2}$$

where

$$A = \frac{\partial f}{\partial z} = \frac{\partial \dot{z}}{\partial z}$$

$$B^j = \frac{\partial^2 f}{\partial z \partial u_j} = \frac{\partial}{\partial u_j}\frac{\partial \dot{z}}{\partial z}$$

$$C = \frac{\partial f}{\partial u} = \frac{\partial \dot{z}}{\partial u}$$

This gives a system of differential equations that describes how activity in any region can be driven by activity in any other region (matrix A), directly by external inputs u (matrix C), and by activity in other regions that is context dependent (matrix B) on the j^{th} input. A response is defined as the change in activity over time, and so the units of connections are per unit time (Hz), and thus a strong connection is one that exerts its effect over a small time. **Figure 2.2** provides a representation of some arbitrary model of interacting regions z, with inputs u. To understand the model in terms of the bilinear state equations, it can be written as a system of differential equations.

Relatively to the example system represented in **Figure 2.2**, the equations are:

$$\begin{aligned}
\dot{z_1} &= a_{11}z_1 + a_{12}z_2 + a_{13}z_3 \\
\dot{z_2} &= a_{22}z_2 + c_{12}u_1 \\
\dot{z_3} &= a_{33}z_3 + a_{31}z_1 + (a_{34} + b_{34}u_2)z_4 \\
\dot{z_4} &= a_{44}z_4 + a_{12}z_2 + a_{13}z_3
\end{aligned}$$

These equations can be represented in the matrix form as

$$\begin{pmatrix} \dot{z_1} \\ \dot{z_2} \\ \dot{z_3} \\ \dot{z_4} \end{pmatrix} = \left[\begin{pmatrix} a_{11} & a_{12} & a_{13} & 0 \\ 0 & a_{22} & 0 & 0 \\ a_{31} & 0 & a_{33} & a_{34} \\ a_{41} & 0 & 0 & a_{44} \end{pmatrix} + u_2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & b_{34} & 0 \end{pmatrix} \right] \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ c_{12} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} u_1$$

This form has the advantage of allowing to see how the parameters are arranged into matrices.

Matrix A represents the intrinsic connectivity which contains forwards, backwards and self-connections between regions. Matrix B embodies changes in connectivity that are context dependent with regards to the experimental design. Matrix C represents the connectivity induced by the experimental stimulus.

2.3 Hemodynamic model

The models of neuronal states form the basis of DCM. The aim is then to estimate its parameters, in order to make inferences about the effective connectivity. However, in order to make inferences about hidden neuronal states, a phenomenologically accurate forward model is needed, that can translate synaptic activity into the BOLD signal that is measured in fMRI, as schematised in **Figure 2.1**. This is done in DCM using a hemodynamic model that is an extended version [72] of the ”Balloon model” [73].

Neural activity in each region is the cause of the BOLD response measured in that region with fMRI. The BOLD response is specific to fMRI and so, in other modalities (e.g. EEG, MEG) the forward model is a different one,

reflecting how the data are generated. Because this work is based on fMRI data, the focus will be on the forward model based on hemodynamic response, shown in **Figure 2.3**.
Regional changes in synaptic activity are known to cause changes in local blood volume and dHb (deoxyhemoglobin) concentration. This means that, for each region, besides the primary state variable z, that corresponds to the regional neural activity, there are four secondary state variables that correspond to the biophysical state variable of the hemodynamic forward model, which was first presented by Friston and colleagues in 2000 [72].

In brief, for a generic region i, neuronal activity z_i causes an increase in a vasodilatoatory signal s_i that is subject to autoregolatory feedback. Inflow f_i responds in proportion to this signal with concomitant changes in blood volume ν_i and deoxyhemoglobin content q_i.

$$\begin{aligned} \dot{s} &= z - \kappa s - \gamma(f-1) \\ \dot{f} &= s \\ \tau\dot{\nu} &= f - \nu^{1/\alpha} \\ \tau\dot{q} &= fE(f, E_0)/E_0 - \nu^{1/\alpha} q/\nu \end{aligned} \tag{2.3}$$

Associated with this model are a set of parameters, of which there is a subset of biophysical related parameters, κ, γ, τ, α, and E_0, which correspond to the rate of signal decay, the rate of flow dependent elimination, the hemodynamic transit time, Grubb's exponent, and the resting oxygen extraction fraction respectively.
Outflow, $f_{out}(\nu) = \nu^{1/\alpha}$, is related to volume through Grubb's exponent, α [74]. The oxygen extraction is a function of flow $E(f, E_0) = 1 - (1 - E_0)^{1/f}$, where E_0 is resting oxygen extraction fraction. The BOLD signal is taken to be a static nonlinear function of volume and deoxyhemoglobin concentration, that comprises a volume-weighted sum of extra- and intra- vascular signals

$$\begin{aligned} y = g(q, \nu) &= V_0(k_1(1-q) + k_2(1-q/\nu) + k_3(1-\nu)) \\ k_1 &= 4.3\ \theta_0 E_0 TE \\ k_2 &= \varepsilon\ r_0 E_0 TE \\ k_3 &= 1 - \varepsilon \end{aligned} \tag{2.4}$$

Here, $TE = 50ms$ is the echo time, $\theta_0 = 40.3s^{-1}$ is the frequency offset at the outer surface of the magnetized vessel for fully deoxygenated blood at $1.5T$, and ε is the ratio of intra- and extra-vascular signal, also treated as free parameter. This equation has been obtained throughout subsequent

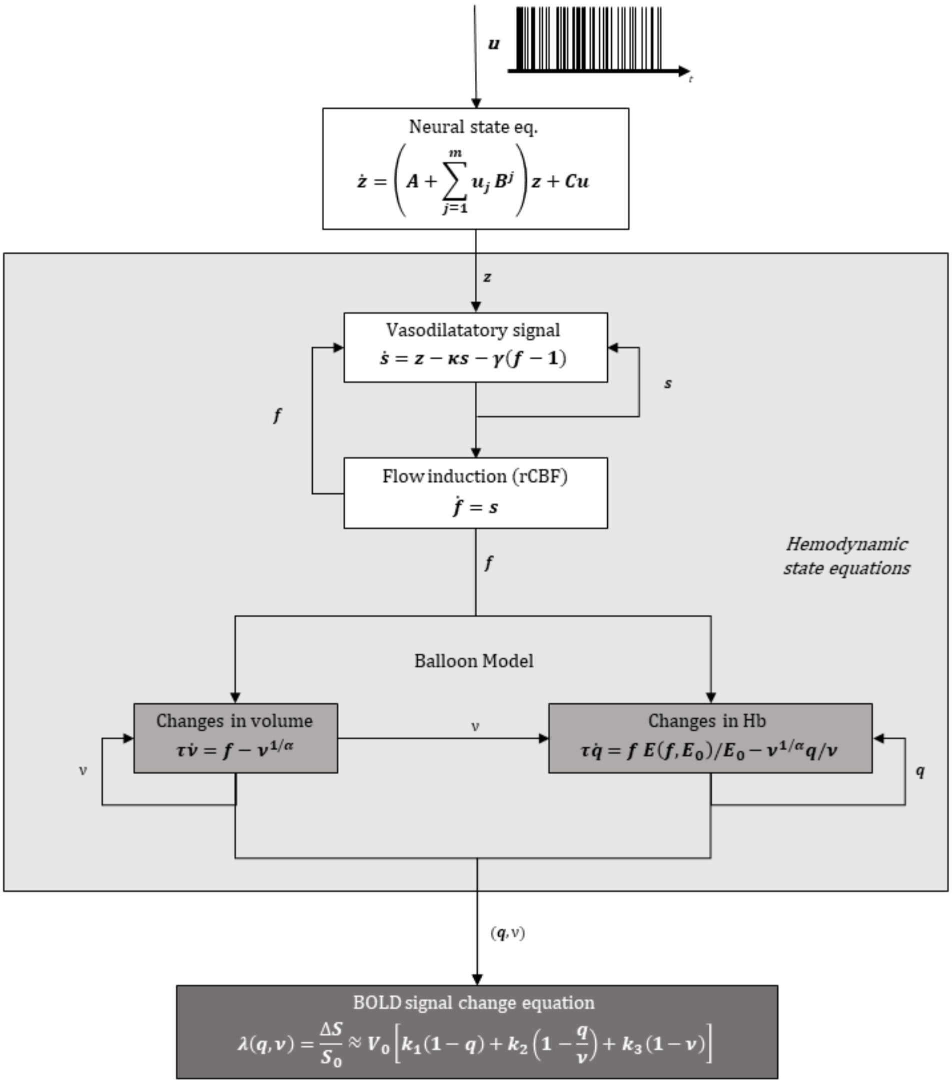


Figure 2.3: Schematic of the complete forward model.

Parameter	Description	Prior mean η_θ	Prior Variance C_θ
κ	Rate of signal decay	0.65 per s	0.015
γ	Rate of flow-dependent elimination	0.41 per s	0.002
τ	Hemodynamic transit time	0.98 s	0.0568
α	Grubb's exponent	0.32	0.0013
E_0	Resting oxygen extraction fraction	0.34	0.0024
ε	Ratio of the intra- to extra- vascular components of the gradient echo signal	-1	0.0078

Table 2.1: Priors of biophysical parameters

modifications of the original Balloon Model equations, to account for fMRI acquisition parameters and newly updated biophysical constants [75] and slice timing [76].
In Table 2.1, a list of the biophysical parameters $\theta^h = \{\kappa, \gamma, \tau, \alpha, E_0, \varepsilon\}$, relative to the SPM12 version, is provided.

2.4 Parameters Estimation

The neuronal state equations and the hemodynamic model combined provide an explanation of how the data were generated, and they are therefore referred to as a "generative model".

$$\begin{aligned} x &= \{z, s, f, \nu, q\} \\ \dot{x} &= f(x, u, \theta) \\ \theta &= \{\theta^c, \theta^h\} \end{aligned} \tag{2.5}$$

For given inputs u, neuronal state parameters θ^c and hemodynamic parameters θ^h, a predicted response $h(u, \theta)$ can be obtained by integrating Eq. (2.5) and passing it through the output non-linearity (Eq. (2.4)).

Integrating those equations and applying the output non-linearity is equivalent to a generalized convolution of the inputs with the systems' *Volterra kernels.* These kernels are easily derived from the Volterra expansions of Eq. (2.5) (Bendat, 1990),

$$h_i(u,\theta) = \sum_k \int_0^t \cdots \int_0^t \kappa_i^k(\sigma_1,\ldots,\sigma_k)u(t-\sigma_1),\ldots,u(t-\sigma_k)d\sigma_1,\ldots,d\sigma_k \tag{2.6}$$

where $h_i(u,\theta)$ is the predicted response of i^{th} region (considering only one input in this case), and

$$\kappa_i^k = \frac{\partial^k y_i(t)}{\partial u(t-\sigma_1 \ldots \partial u(t-\sigma_k))}$$

is the k^{th} order kernel of the i^{th} region.

These kernels are simply a representation of the model and can be seen as a generalization of the *impulse response* of the system; hence, the systems output can be obtained by convolution of the inputs with the systems kernels. The Volterra kernels are a time-invariant characterization of the input-output behavior of the system and can be thought of as generalized high-order convolution kernels that are applied to a stimulus function to emulate the observed BOLD response.

Volterra expansion is used as a model for non-linear behaviour similar to Taylor series, from which it differs for its capacity of capturing the memory effect. In fact, Taylor series approximates the response of a non-linear system for a given input in a fixed time point. The Volterra series, instead, approximates the response of a non-linear system that depends on the whole time course of the input.

The observed data y can then be modelled as the sum of the predicted response $h(u,\theta)$, confounding effects $X(t)$ (with coefficients β), and an error ε.

$$y = h(u,\theta) + X\beta + \varepsilon$$

Low frequency drifts can be filtered out from time series during processing, thus the model becomes

$$y = h(u,\theta) + \varepsilon \tag{2.7}$$

under Gaussian assumptions about the parameters θ and errors $\varepsilon \sim \mathcal{N}(0, C_\varepsilon)$. These high dimensional equations cannot be solved analytically and it would be computationally very costly to use a brute force numerical method [60]. Therefore when DCM was introduced [11], a Variational Bayesian (VB) technique was also introduced [11]. Using this Bayesian inversion scheme, parameters for the complete model are estimated (inverted), given the data and the prior distributions on the parameters. Using Bayes theorem, the posterior probability of the parameters is expressed mathematically as

$$
\begin{aligned}
p(\theta \mid y, m) &= \frac{p(y \mid \theta, m)\, p(\theta, m)}{p(y \mid m)} \\
log(p(\theta \mid y, m)) &= log(p(y \mid \theta, m)) + log(p(\theta, m)) - log(p(y \mid m))
\end{aligned}
\tag{2.8}
$$

This way, the dynamical system is introduced in Bayes' world, defining a likelihood model. As described above, in fact, the forward model can be made into an observation model by adding a normally distributed error with zero mean; hence, data are assumed to be a random variable, normally distributed around the prediction from the dynamic model:

$$
\begin{aligned}
p(y(t)|\theta, m) &= \mathcal{N}(y(t), \theta_\sigma) \\
p(y|\theta, m) &= \prod_t p(y(t)|\theta, m)
\end{aligned}
$$

The maximum posterior distribution of the parameters is then approximated using the iterative optimization Expectation Maximization (EM) algorithm, details of which are given in **Appendix C**.

2.5 Model priors

Due to the complexity of DCMs, model inversion needs to be more dependent on constraints, which is why DCMs are inverted within a Bayesian scheme. Each parameter is constrained by a prior distribution which is based on empirical knowledge, and the estimation procedure produces a posterior distribution. Placing DCM within a Bayesian framework is a necessity due to its complexity, but it also has many advantages compared to inference based on classical statistics. Using classical statistics such as p-values we are estimating the probability of observing the data given no effect, which is a problem as we can never say for certain that an observation has not occurred. Bayesian inference however, produces posterior distributions that are the probability of the effect given the data observed [77].

There are properties of neuronal dynamics that can be used as priors on the parameters of the neuronal state model. Neural activity, for example, cannot increase to infinitely high values and in the absence of an external input the dynamics are likely to return to a stable mode. These concepts are used to constrain DCMs through shrinkage priors on the coupling parameters that place a small probability on self-excitation and high values of regional activity. The priors used for the five biophysical parameters of the haemodynamic model are based on empirical values that have been obtained [78], and are reported in Table 2.1 in Section 2.3. Priors for the remaining haemodynamic model parameters, which cannot be biophysically informed, are identified as those which minimise the sum of squared differences between the Volterra kernels they imply and the Volterra kernels derived directly from data [72]. Additional constraints can be added to optimise a particular DCM if information about the anatomical structure are available. Previous studies have used structural connectivity information obtained via invasive tract tracing in macaque monkeys to inform the structure of models for effective connectivity studies [66]. Although these data are of high resolution, they are not necessarily relevant for human studies due to inter-species differences in connectivity. This problem can be overcome by using structural information obtained via DTI. Despite the fact the data are less detailed and do not contain directional information, they have still been successfully integrated into DCM as priors by Stephan and colleagues [79]. In this study, probabilistic tractography based on data collected via diffusion weighted imaging (DWI) was used to calculate the probability of anatomical connections existing between visual areas of the brain.

2.6 Inference

There can be two types of inference in DCM: inference about parameter space and inference about model space [66]. If, for example, we are interested in the specific effect of a connection, such as whether it exhibits an excitatory or inhibitory effect, it requires inference about the parameters of a model. Alternatively, we may wish to make inferences about model structure, for example to determine the presence of feedback connections. Early DCM studies tended to be more focused on inference about parameter space [67], however following a proliferation of methodology papers devoted to model selection [80, 81, 82, 83, 84], inference about model structure has become more common. One area in which inference about parameter space is still dominant, is the one involving group studies between patients and controls [85].

2.6.1 Bayesian Model Selection (BMS)

The problem of model selection, encountered in any modelling approach, tries to identify which, of a set of competing models, is most likely given the data.

The problem is confounded by the fact that the model fit, alone, is not enough to infer which model is best. Model complexity also needs to be considered to ensure that the model is not over-fitting the data (i.e. it has good generalisation capability) [86]. When it was first introduced, it was proposed that DCMs should be compared using a combination of Akaike's Information Criterion (AIC), and Bayesian Information Criterion (BIC) [83]. In another approach, log differences in model evidence (also known as log Bayes factor) were used to compare competing models and a value greater than 3 was suggested as the threshold for accepting one model over another [81]. Since then, this method, denominated Bayesian Model Selection (BMS), has become the preferred method of comparing models using an approximation of the Free Energy as model evidence [87]. The model evidence is given in Eq.(2.9).

$$p(y \mid m_i) = \int p(y \mid \theta, m_i) p(\theta \mid m_i) d\theta \tag{2.9}$$

This integral cannot be solved analytically, but it can be approximated, as detailed in Penny et al. (2004) [83]. The approximation is given in Eq. (2.10) as the log model evidence and consists of an accuracy and complexity term.

$$log\, p(y \mid m) = Accuracy(m) - Complexity(m) \tag{2.10}$$

Two models, m_1 and m_2, can therefore be compared using the Bayes factor [88] given in Eq. (2.11).

$$B_{12} = \frac{p(y \mid m_1)}{p(y \mid m_2)} = log\, p(y \mid m_1) - log\, p(y \mid m_2) \tag{2.11}$$

The Bayes factor is simply the difference between the log model evidences for $model1$ and $model2$. This means that the most likely model is the one with the greatest log evidence. The AIC and BIC provide simple approximations to the log evidence, and they were used in early DCM studies; however a free energy approximation is now preferred. As shown in Eq.(2.10), the complexity term penalises a model based on its complexity. In AIC and BIC, the complexity term is simply a function of the number of model parameters. In the free energy approach, the model evidence is approximated by Eq. (2.12), where F is known as the free energy, and the last term is the Kullback-Leibler ($\mathcal{D}_{\mathcal{KL}}$) divergence between the true posterior $p(\theta \mid y, m)$ density and the approximate posterior density $q(\theta)$ [60].

$$log\, p(y \mid m) = F(m) + \mathcal{D}_{\mathcal{KL}} \left(q(\theta) \mid\mid p(\theta \mid y, m) \right) \tag{2.12}$$

Due to the Gibb's inequality, the $\mathcal{KL}$ divergence is always positive, meaning the free energy provides a lower bound on the log model evidence. When the $\mathcal{KL}$ divergence is equal to zero then the true and approximate posterior

densities are the same and the free energy is equal to the log model evidence. Thus the EM optimisation scheme serves to maximize the free energy, implicitly decreasing the $\mathcal{KL}$ divergence, and making the approximate posterior distribution as close to the true one as possible, simultaneously providing an approximation of model evidence. Unlike AIC and BIC, in which each parameter in the model is penalised equally by the complexity term [81], the free energy approach has a complexity term that is the $\mathcal{KL}$ divergence between prior and approximate posterior. Thus, parameters are not penalised equally, and so the more a parameter deviates from its prior, the greater the penalty. This extra sensitivity has been empirically shown to make the free energy a better approximation to model evidence [82].
For multiple subjects analyses, two options exist depending on how parameters are considered to be distributed across subjects [66]. In the Fixed-Effects (FFX) approach it is assumed that model structure is the same for each subject in the population, and in the Random-Effects (RFX) approach it is allowed for the possibility that different subjects have different models, assuming model structure somehow as a random variable.

Fixed Effects Analysis

In the FFX approach, since every subject is assumed to have the same model, the model evidence given a dataset Y composed of independent data for individual subjects y_n is simply the sum of the log model evidences for each subject, given by Eq. (2.13).

$$p(Y \mid m) = \prod_{n=1}^{N} p(y_n \mid m) log\, p(Y \mid m) \quad = \sum_{n=1}^{N} log\, p(y_n \mid m) \tag{2.13}$$

Random Effects Analysis

The RFX approach assumes that for each subject different models generate the observed data. Assuming that the data are generated by models drawn from a probability distribution, this is achieved using a Bayesian hierarchical approach that can be inverted to obtain an estimate of the distribution [84]. A prior distribution of model probabilities is given by a Dirichlet distribution given by Eq. (2.14), where r_m is the probability of model m from a set of M total models, α_m are the number of times model m is selected in the population, and so can be viewed as the number of subjects for whom that model generated the data [84], and $Z(\alpha)$ is a normalisation term.

$$p(r \mid \alpha) = Dir(\alpha) = \frac{1}{Z(\alpha)} \prod_{m=1}^{M} r_m^{\alpha_m - 1} \tag{2.14}$$

The inversion of the model produces an approximation to the posterior distribution $P(r \mid Y)$. This was previously achieved using a VB approach [84] but since then a Gibb's sampling method has been suggested [83] and is the preferred method when comparing large numbers of models, i.e. more models than subjects.

2.6.2 Model Space

Clearly defining a plausible model space should be a fundamental component of any DCM study [66]. The problem is a general one in that for any experimental data, there are an infinite number of models that could explain it, which vary in both structure and parameter values. For this reason one always has to place limitations on model space to constrict it to a set of plausible alternatives. This is already an inherent part of the DCM framework which is based on Bayesian statistics. As already noted, prior distributions on parameters aim to constrain the parameters, which describe neural activity and the hemodynamic response, to values which are biophysically realistic. However given any number of regions, even with constraints on parameters, there are still a vast number of model structures that could explain the data. The problem of defining a plausible model space is not a trivial one and the main issue being highlighted [89, 90] is the problem of so called "combinatorial explosions".
Given a number of brain regions n, the number of possible models in bilinear DCM is determined by the Eq. (2.16), where j is the number of experimental manipulations and k is the number of connections between nodes n which is equal to $n(n-1)$.

$$m = (2^{nj} - 1)\left(\sum_{i=0}^{k} 2^{ij}\binom{k}{i}\right) \tag{2.15}$$

$$\binom{k}{i} = \frac{k!}{k!(i-k)!} \tag{2.16}$$

As the number of regions and experimental conditions increases the number of possible models rises very rapidly; one approach is to impose some limitations, usually based on intrinsic connectivity, and then estimate all possible models within a greatly reduced model space [91, 92]. Others have chosen to adopt a hierarchical approach [93, 94] by first defining a model space of varying intrinsic connectivity and then using the winning model to define a new model space of varying modulatory effects. Pyka et al. developed a genetic algorithm to search model space, and found that it was computationally more efficient than a brute force search [95]. Many studies comparing healthy controls to patient groups have omitted a model space search altogether and instead chosen to use classical statistics to compare parameters on a hypothesised model between groups [96, 97, 98]. This is a particularly popular approach for group studies [85] but has recently been discouraged except for cases when one has very strong a priori knowledge concerning model structure [66].

Model Families

Family level inference for DCM is an innovation introduced by Penny and colleagues [83] as a way of removing uncertainty in model structure. They showed that comparing large numbers of models in the traditional manner can be problematic, because, when the number of models to e compared increases, there could be the so called dilution of evidence (when many models share close values of free energy), and therefore no clear winning model is identified; they showed, instead, that grouping models into families according to some characteristic, e.g. input location, is a more robust approach. A family partition is defined and models are classified as belonging to one of the subsets which must be non-overlapping. The partitioning of the model set into families reflects the question being asked by the researcher.

2.6.3 Post-hoc BMS (model reduction)

The original purpose of DCM was as a hypothesis driven approach, in which a limited number of carefully selected models were compared in order to test a specific hypothesis about how the data were generated. It is still primarily used in this fashion, though there has been a recent trend for comparing an ever increasing numbers of models [82, 99] and thus using DCM as a more exploratory method.

The problem with comparing large numbers of models is that it is computationally intensive due to the need to fully invert each model. To face this problem, Friston and Penny have recently proposed a solution [80] in which only a single model, for each subject, is fitted to the data. Known as post-hoc BMS, only the largest of a set of models need be inverted and then model evidence for all the reduced models within this set is approximated. In addition to model evidence, the connectivity parameters can also be estimated from the posterior distribution of parameters in the full model [100]. In this context, ”reduced” models are referred to as models that are obtained from the full model by switching off one or more parameters in a full model by setting their prior mean and variance to zero, resulting in a reduced number of parameters with respect to the full model.

The important aspect of Bayesian model reduction is that models differ only in their priors, which means that the posterior of a reduced model can be derived from the posterior of the full model.

Based on the ability of post-hoc BMS to score large numbers of models, Friston and colleagues [99] outlined a method for "network discovery" using DCM, along with the additional constraint that connections are bidirectional. This work has been expanded by Seghier and colleagues [101] for a large DCM network containing twenty nodes, using the principle components of the functional connectivity network as constraints on the intrinsic connectivity.

This scheme is also referred to as *Bayesian model reduction (BMR)*, and refers to the Bayesian inversion of reduced models using only the posterior densities of a full model. Thus, large numbers of (reduced) models can be scored extremely efficiently. The post-hoc model evidence shows good agreement with the model evidence computed by classical (and computationally expensive) model estimation [100]. This means post-hoc BMS or BMR could be used to infer network structure in the entire model space, and a validation test could be performed much faster. Following, a rough description of this procedure is provided.For a more in-depth analysis see [80, 100]. As mentioned before, the key feature of post-hoc inversion is the assumption that the full and reduce models differ only in their priors over parameters; in fact, reduced models are obtained by imposing some parameters prior expectation and covariance to zero. This means that the full model shares the same likelihood whit the set of reduced models:

$$\begin{aligned} p(y \mid \theta, m_i) &= p(y \mid \theta, m_F) \\ \frac{p(y \mid m_i)}{p(y \mid m_F)} &= \frac{p(\theta \mid y, m_F)}{p(\theta \mid y, m_i)} \frac{p(\theta \mid m_i)}{p(\theta \mid m_F)} \end{aligned} \tag{2.17}$$

where m_i indicates the reduced model and m_F indicates the full model. Under Laplace approximation (Gaussian priors and posteriors of full and reduced models), an approximation of the evidence of the reduced model can be obtained. This is expressed by

$$\begin{aligned} log\, p(y \mid m_i) =& \frac{1}{2}\, log \frac{|\Pi_i|\, |P_F|}{|P_i|\, |\Pi_F|} \\ & - \frac{1}{2} \left(\mu_F^T P_F \mu_F + \eta_i^T \Pi_i \eta_i - \eta_F^T \Pi_F \eta_F - \mu_i^T P_i \mu_i \right) + F^F \end{aligned} \tag{2.18}$$

where $\eta_{i,F}$ and $\Pi_{i,F}$ are the prior means and precisions for the reduced and full models, while $\mu_{i,F}$ and $P_{i,F}$ are the posterior means and precisions. This is useful because the required means and precisions of the reduced model can be derived in a straightforward way from the means and precisions of the full model. To obtain the post-hoc parameters estimates, it is again assumed that the models differ only in the specification of the priors, i.e. they share the same likelihood. Using this assumption the linearised approximation can be subtracted to the conditional precision of the full model from the precision of the reduced model and eliminate the terms that do not depend on the priors obtaining an expression of the estimates of the parameters (means and precision) for any reduced model:

$$\begin{aligned} P_i &= P_F + \Pi_i - \Pi_F \\ \mu_i &= C_i(P_F\mu_F + \Pi_i\eta_i - \Pi_F\eta_F), \end{aligned} \tag{2.19}$$

where $C_i = P_i^{-1}$. Once the log-evidence of every reduced model has been estimated using its free energy approximation, the best model in the model space can be selected using BMS. One alternative approach, instead of explicitly defining the reduced models to be scored and then comparing them through their Bayes factors (to identify the best), the log-evidence in Eq.(2.18) could be optimised with respect to the prior distribution. Furthermore, we can consider any hyperparameterisation of the prior $p(\theta \mid m(\lambda)) = N(\eta(\lambda), \Sigma(\lambda))$ that induces a model. Here the hyperparameters λ control the mean and precision to produce a log-evidence, $F(\lambda)$. This perspective takes us away from the notion of discrete models and into a continuous model space supported by hyperparameters, $m(\lambda)$. In this the optimum model and posterior are:

$$\begin{aligned} \lambda^* &= argmax_\lambda\, \mathcal{F}(\lambda) \\ q(\theta \mid m^*) &= \mathcal{N}(\mu^*, C^*) \\ P^* &= P_F + \Pi(\lambda^*) - \Pi_F \\ \mu^* &= C^*(P_F\mu_F + \Pi(\lambda^*)\eta(\lambda^*) - \Pi_F\eta_F). \end{aligned} \tag{2.20}$$

The use of one or the other of the two alternative methods depends on whether we want (or are able to) to specify all the competing models that need to be tested, or if we want to perform an exhaustive search over all possible combinations of connections.

2.6.4 Inference on parameter space

When making inferences on model parameters, the same decision to make as with group-level BMS emerges, i.e. FFX or RFX. A number of FFX methods exist, such as Bayesian Parameter Averaging (BPA), in which the posterior parameter distributions for each subject are combined according to Bayes

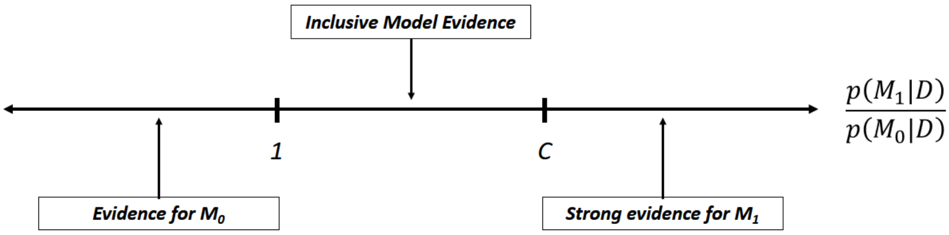


Figure 2.4: Schematic occam's window

theorem [102, 103]. A comparison of different FFX methods for Bayesian parameter inference in group studies can be found in Kasess and collegues [102]. However, a RFX approach in which subject specific parameter estimates are compared using a second level analysis using classical frequentist tests such as t-test or ANOVA are more common [66]. Another approach is Bayesian Model Averaging (BMA) [82] in which parameter estimates are not dependent on a single model but are averaged across multiple models within a set and are weighted according to the probability of each model, as shown in Eq. (2.21).

$$p(\theta \mid y) = \sum_{m_i} p(m_i \mid y) p(\theta \mid y, m_i) \tag{2.21}$$

Parameter averaging is performed among a reduced set of models, obtained through an algorithm called "Occam's windows". In brief, this algorithm uses the Bayes factor, i.e. the ratio between two models evidences, to evaluate which of them should be included/discarded from the subset of model space, absed on a trade-off between model evidence and complexity. This principle concerns the interpretation of the ratio of posterior model probabilities $p(M_1 \mid D) \,/\, p(M_0 \mid D)$. Here M_0 is a model with one less parameter than M_1. A representation of the essential idea is shown in **Figure 2.4**. If there is evidence for M_0, then M_1 is rejected, but to reject M_0 a storng evidence for the larger model is required (having a higher complexity). If the evidence is inclusive (falling within Occam's window), neither model is rejected. BMA is useful for scenarios in which there is no clear winning model or for comparison between groups in which model structure may not be equal such as patients and controls [66].

2.6.5 Parametric Empirical Bayes (PEB)

A common experimental aim is to test whether effective connectivity is different between groups of subjects, or is different according to a behavioural

measure (e.g. test scores) within a group. One approach is to take DCM connectivity parameters and apply a classical statistical test (e.g. t-tests, CVA).
However, the disadvantage of this approach is that it throws away the estimated uncertainty (variance) about the connection strengths. Alternatively, a hierarchical model over the parameters can be constructed, describing how group level effects constrain parameter estimates on a subject-by-subject basis. This approach produce a so called Parametric Empirical Bayes (PEB) model, which makes it possible to evaluate group effects and between-subjects variability on parameters, by modelling (first level) parameters as random variables oscillating around a group mean.
In fact, these sorts of models are used (either implicitly or explicitly) in the analysis of multisubject studies that contain both within and between subject effects [104, 105, 106].

Empirical Bayes refers to the Bayesian inversion or fitting of hierarchical models. In hierarchical models, constraints on the posterior density over model parameters at any given level are provided by the level above, after they have "seen" the data. These constraints are called empirical priors because they are informed by empirical data. These procedures permit the analysis of data from group studies using nonlinear models, such as dynamic causal models of neurophysiological timeseries. The inversion or fitting of hierarchical models in a nonlinear setting is achieved by using Bayesian Model Reduction (BMR), previously described in section 2.6.3.
This allows to compute posterior densities over model parameters, under new prior densities, without explicitly inverting the model again. For example, a nonlinear (causal) model can be inverted for each subject in a group and then evaluate the posterior density over group effects, using the posterior densities over parameters from the single-subject inversions.
This can be regarded as a generalisation of the standard summary statistic approach; however, instead of just using point estimators as summaries of first (within-subject) level effects, we take the full posterior density to the second (between-subject) level. In case of DCM, the posterior density over parameters at the first level is assumed to be approximately Gaussian, and thus mean and covariance are propagated at the second level.

In this discussion, will be considered an empirical Bayesian approach to any hierarchical model that can be expressed in terms of an arbitrary (nonlinear) model at the first level and a standard (parametric) empirical Bayesian (PEB) model at higher levels. Crucially, these procedures are very efficient because each hierarchical level of the model requires only the posterior density over the parameters of the level below. This means, deep hierarchical models can be inverted without having to revisit lower levels. This aspect of the scheme rests on Bayesian Model Reduction.

Bayesian model reduction unravels a number of issues in the inversion and interpretation of group DCM studies. These include the problem of local maxima when evaluating different models for Bayesian model comparison, and the fundamental distinction between random (between-subject) effects at the level of models and their parameters.

In contrast to previous treatment of random model effects at the between-subject level [84], this work considers random parameter effects in the setting of parametric empirical Bayes.

As reported by Friston and colleagues [19], Bayesian model reduction provides more robust estimates of posterior probabilities than fitting models to the data separately, because it is less susceptible to violations of (e.g., Laplace) assumptions. This application of Bayesian model reduction provides Bayesian model averages that could be used for classical inference with the standard summary statistic approach. However, we can go further in terms of model comparison and classification, using empirical Bayesian model reduction, as it has been done in this work. For DCM studies with N subjects and M parameters per DCM, we have a hierarchical model, where the responses of the i^{th} subject and the distribution of the parameters over subjects can be modelled as

$$\begin{aligned} y_i &= \Gamma_i^{(1)}(\theta^{(1)}) + \varepsilon_i^{(1)} \\ \theta^{(1)} &= \Gamma^{(2)}(\theta^{(2)}) + \varepsilon^{(2)} \\ \theta^{(2)} &= \eta + \varepsilon^{(3)}. \end{aligned} \tag{2.22}$$

In this hierarchical form, empirical priors encoding second (between-subject) level effects place constraints on subject-specific parameters. The implicit generative model is defined in terms of multivariate Gaussian distributions (assuming the data for each subject are conditionally independent):

$$\begin{aligned} ln\,(y, \theta^{(1)}, \theta^{(2)} \mid m) &= \sum_i ln\, p(y_i \theta^{(1)}) + ln\, p(\theta^{(1)} \mid \theta^{(2)}) + ln\, p(\theta^{(2)} \mid m) \\ p(y_i \mid \theta^{(1)}, m) &= \mathcal{N}\left(\Gamma_i^{(1)}(\theta^{(1)}), \Sigma_i^{(1)}(\theta^{(1)})\right) \\ p(\theta^{(1)} \mid \theta^{(2)}, m) &= \mathcal{N}\left(\Gamma^{(2)}(\theta^{(2)}), \Sigma^{(2)}(\theta^{(2)})\right) \\ p(\theta^{(2)} \mid m) &= \mathcal{N}(\eta, \Sigma) \end{aligned} \tag{2.23}$$

Generally, the second level would be a linear model where the random effects are parameterised in terms of their precision.

$$\begin{aligned}\Gamma^{(2)}(\theta^{(2)}) &= (X \otimes W)\beta \\ \Pi^{(2)}(\theta^{(2)}) &= I_N \otimes (Q_0 + \sum_j e^{-\gamma_j} Q_j)\end{aligned} \tag{2.24}$$

Here, $\beta \subset \theta$ are group means or effects encoded by a design matrix with between, X, and within-subject, W, components. The between-subject part encodes differences among subjects or covariates such as age, while the within-subject part specifies mixtures of parameters that show random effects. When every parameter can express different group effects, as in this work, $W = I_M$. Intuitively, the Kronecker product $X \otimes W$ models the fact that one or more parameters can show one or more group effects. We will assume that the first column of the design matrix is a constant term, modelling group means, and subsequent columns encode group differences or covariates such as age or clinical scores.

The second (between-subject) level precision is parameterised by log precisions $\gamma \subset \theta$ of (positive definite) precision components Q_j that are added to a lower bound on precision Q_0. These components specify whether the parameters are random or fixed effects. Formally, the difference between a random and fixed effect rests upon the prior variance at the second level. Random effects have an informative prior that shrinks subject-specific estimates towards their (second level) mean. Conversely, fixed effects have a relatively flat or uninformative prior Q_0 such that they are less influenced by parameter estimates from other subjects.

We now wish to find the approximate posterior, over second level parameters, that maximises free energy, where the free energy at the first level has already been optimised for each subject under full priors.

More precisely, we need the sufficient statistics (mean and covariance) of the approximate posterior, given the priors and approximate posteriors for each subject at the first level.

$$\tilde{q}^{(2)} = (\mu^{(2)}, C^{(2)}) = argmax_{q^{(2)}} F^{(2)} \tag{2.25}$$

As shown in Eq.(2.25), the sufficient statistics of the approximate posterior over second level parameters are obtained by maximising the second level free energy. The second level free energy comprises the expected (reduced) first level free energy summed over subjects and the complexity attributable to the posterior over second level parameters. At second level, the conditional precision that maximises negative free energy is the solution to:

$$\partial_{C^{(2)}} F^{(2)} = \sum_i \frac{1}{2} \partial^2_{\mu^{(2)}} F_i^{(1)} + \frac{1}{2} P^{(2)} = 0 \implies P^{(2)*} = -\sum_i \partial^2_{\mu^{(2)}} F_i^{(1)} \tag{2.26}$$

The second level expectations can now be optimised using gradient ascent as described in Friston et al. [19].
Note that the overall scheme can be applied recursively: once the second level parameters have been optimized they can be treated as first level parameters for efficient (recursive) inversion of deep hierarchical models.
Moreover, the summary statistics passed from one level to the next include not only the point estimators, or expectations, but the full (Gaussian) posterior density over parameters at the lower level.

The reduced free energy is a function of the approximate posterior over second level parameters and the (known) approximate posterior over the first level parameters, under the full model. Effectively, this means the expected (reduced) free energy is the free energy that we would get if we replaced the full priors with the empirical model. In short, we never needs to actually optimise the first level posterior, when optimising the posterior at the second level.

Summming up, optimization at the second level proceeds recursively (until convergence), updating expectations by an amount that is a function of gradient and curvature of the second level free energy function; in turn these terms depend on the sufficient statistics of reduced priors and posteriors, that are calculated exploiting the useful scheme of BMR (at the very first iteration the empirical priors are initialized using the full posteriors of the first level).

In summary, with a straight forward application of Bayes rule, the posterior density of any (reduced) model can be expressed in terms of the posterior of its parent or full model. This affords an efficient way to evaluate posterior densities under empirical priors, leading to the notion of hierarchical or empirical Bayesian model reduction. This form of hierarchical model inversion and comparison is interesting because it only requires the forward passing of the posterior density, from a lower level to a higher level, to generalise the standard summary statistical approach. In this generalisation, all the sufficient statistics of the posterior are passed to higher levels (as opposed to just passing the maximum likelihood or a posteriori parameter estimates). This scheme enable inference under hierarchical models of (non-linear) within-subject effects and (linear) between-subject effects. These can be applied in a number of different contexts. For example, we could simply assume that between-subject differences are attributable to random variations in their parameters and use a simple design matrix with $X = 1$ to provide empirical shrinkage priors, which shrink first level estimates towards the group mean. However, we can also consider more elaborate second level models

that contain information about subjects and the groups from which they were sampled.
This permits inference about group effects directly at the second or between-subject level.

Chapter 3

Materials and Methods

3.1 Subjects

The subjects from which fMRI data were taken and used in this work, comprised a group of patients and a group of control subjects. Patients were recruited from the outpatient and inpatient services at the Villa Beretta Rehabilitation Centre (Costa Masnaga, LC, Italy from 2010 to 2013). All patients had suffered from first-ever stroke > 6 months previously, resulting in weakness of at least the tibialis anterior muscle (to $< 4+$ on the Medical Research Council (MRC) scale [107]). Exclusion criteria consisted of (i) responsiveness of less than 10° in FES-induced ankle dorsiflexion; (ii) language or cognitive deficits sufficient to impair cooperation in the study; (iii) inability to walk even if assisted; (iv) high spasticity at ankle joint plantar flexor as measured by the modified Ashworth scale index, MAS > 2. The patients dataset available and used in this study, was relative to 8 patients and composed of 2 fMRI sessions per patient, taken before (t_1) and after (t_2) the rehabilitation treatment, for a total of 16 brain volumes, 2 for each patient (PRE, POST).

The rehabilitation treatment was based on the functional electrical stimulation (FES) of the Tibials Anterior muscle for the recovery of the drop foot. Patient characteristics, along with the degree of functional recovery at the time of scanning, as measured by the selected outcome measures, are listed in **Table 3.1**. Mean age was 40.4 ± 17.2 years, and mean time post-ictus was 31 ± 16.2 months.

The site of cerebral infarction was determined from the T1-weighted structural MRI, reported in **Figure 3.1** The age-matched control group was composed of 16 healthy volunteers with no neurological or orthopaedic impairment. Mean age was 36.4 ± 13.8. Their results have been fully reported

Subject	Age [$years$]	Sex	Site of lesion	Type of stroke	Time [$months$]	Training device	Gait velocity (pre-post) [m/s]	Endurance velocity (pre-post) [m/s]	Paretic step length (pre-post) [m/s]	TAAI (pre-post) [$0-1$]	MRC (pre-post) [$1-5$]	Capacity Score (pre-post)	Carry-over effect
PZ01	23	M	R MCA	H	23	Walkaid	0.64-0.71	0.82-0.92	486-538	0.60-0.38	3-4	20.30-21.32	yes
PZ02	38	F	R GP	I	23	Bioness	0.25-0.27	0.28-0.24	194-196	0.48-0.40	2-3	8.15-8.17	no
PZ03	64	F	L MCA	H+I	13	Walkaid	0.32-0.60	0.50-0.54	345-445	0.56-0.36	3-3	14.33-18.15	yes
PZ04	19	M	L MCA	H	44	Walkaid	0.82-0.86	1.01-1.10	561-567	0.44-0.53	3-4	23.42-23.82	no
PZ05	47	F	L GP	H	44	Walkaid	0.52-0.55	0.82-0.91	513-554	0.46-0.44	3-4	21.26-23	yes
PZ06	25	F	R MCA	I	30	Walkaid	0.70-0.65	0.98-0.96	591-544	0.62-0.71	3-3	24.50-22.7	no
PZ07	46	M	R GP	I	13	Walkaid	0.49-0.48	0.58-0.63	420-410	0.43-0.22	3-3	17.31-16.99	no
PZ08	61	F	R MCA	H	58	Walkaid	0.40-0.55	1.36-1.33	430-460	0.69-0.75	2-3	18.50-19.76	yes

Table 3.1: Patients individual characteristics with pre-post clinical scores and carry-over evaluation

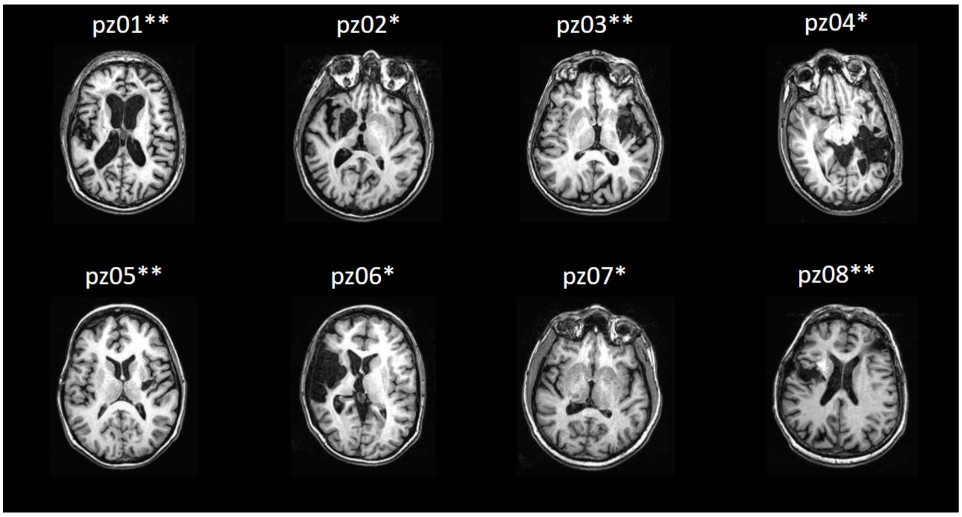


*Figure 3.1: The site of cerebral infarction as determined from the T1-weighted structural MRI. ** indicates patients with FES carryover; * indicates patients with no FES carryover.*

previously [6], and basic personal informations are reported in **Table 3.2**. Villa Beretta Rehabilitation Centre Ethics Committee and all subjects gave informed written consent in accordance with the Declaration of Helsinki.

3.2 Capacity Score (CS) and Carryover Effect (CE)

One of the purposes of this study is to evaluate FES-induced mechanisms of action, and possible predictive biomarkers for the carry-over effect in terms of brain connectivity. Ideally, the outcome of a rehabilitation treatment is successful if it induces a stable improvement on the targeted task and, thus, understanding the key features and causes of carry-over effect would be fundamental for the evaluation of FES-based treatment effectiveness on different patients. In fact, analysing the differences in effective connectivity between patients that exhibited the carry-over effect and those who did not, could help us to find predictive biomarkers for the rehabilitation outcome and understand a possible predisposition of patients to it, prior to treatment. This wold open the way to a process of customization of the treatment based on subjective connectivity features, which is already thought to be potentially an influential factor on the outcome of motor recovery.

In order to investigate the differences between patients exhibiting or not the carry-over effect, an evaluation of the improvements after the attendance of FES-based motor rehabilitation is needed. Dealing in particular with neuro-

#	Age	Sex	Handedness
1	24	M	R
2	24	M	R
3	26	F	R
4	25	M	R
5	25	M	R
6	47	F	R
7	61	M	R
8	45	M	R
9	50	F	R
10	59	F	R
11	51	F	R
12	39	M	R
13	30	F	R
14	22	F	R
15	22	M	R
16	33	F	R

Table 3.2: Healthy subjects individual characteristics

motor rehabilitation trainings, different aspects of motor improvement can be nowadays accurately evaluated with a wide battery of measures, that are quantitative, reliable, and safe.

Gandolla and colleagues [10] defined a quantitative and comprehensive method of combining multiple measures in order to assess a stable functional improvement. This approach: (i) encompasses different aspects of motor function (e.g., kinematic, muscle activity, clinical indices, etc.); (ii) provides a method that can evaluate treatment-induced improvement in a single subject; (iii) is simple to interpret in that it provides a binary outcome in relation to improvement; (iv) is validated with respect to standard clinical evaluation. The patients were classified in a binary fashion, based on whether or not the patient achieved the carry-over effect: this was done by combining the outcomes of 5 different clinical and instrumentation measures, commonly used by clinicians to asses motor ability and improvements of impaired patients.

The set of outcome measures was designed to assess different aspects of patients' functional condition:

1. GV — self-selected gait velocity as measured during the gait analysis test;

2. EV — endurance velocity, as calculated during the 6-min walking test;
3. PSL — paretic step length as measured during the gait analysis test;
4. TAAI — tibialis anterior activation index of the tibialis anterior muscle between toe off and toe strike and during the whole gait cycle as detected by electromyography measures;
5. MRC index at ankle joint.

The carryover effect was determined using a novel algorithm based on variables minimum detectable change that combines the outcome measures to obtain a unique parameter, Capacity Score (CS), where a higher Capacity Score indicates higher residual ability. In particular, for each assessment session, patients were assigned to a point in the 5-dimensional space, identified by the 5 outcome measures listed above. The 5-dimensional space was centred on the outcome measures derived from healthy subjects, and therefore the further away the patient is from the origin, the more impaired he/she is. Moreover, the outcome variables have been normalised with respect to the corresponding minimum detectable change. The difference in the 5-dimensional space between (i) the Euclidean distance of "subject zero" (a patient that scores zero in all outcome measures, i.e., the most impaired patient in our space) with respect to the origin (i.e., distance of the given patient from the healthy control group) and (ii) the Euclidean distance of each patient with respect to the origin (i.e., distance of the "subject zero" from the healthy control group) is defined as Capability Score. The difference between Capacity Scores at different timing (i.e., post-pre) is thresholded (difference $>$ 5% of initial value) to obtain carryover effect assessment. The algorithm has been validated against clinical evaluation [10]. This way the patients group has been subdivided into two subgroups of 4 patients each: one including subjects that successfully achieved a carry-over effect (CE) and the other composed by those subjects that did not gain a durable improvement (nCE).

A more detailed description of the algorithm can be found in Gandolla et al. (2015) [10].

3.3 Experimental set-up

The experimental setup was composed of 1.5 T MRI scanner (GE Cv/I), a motion capture system (Smart μg; BTS), and an electrical stimulator (RehaStim proTM; HASOMED GmbH), as previously described and validated [108, 109], and reported in **Figure 3.2**.

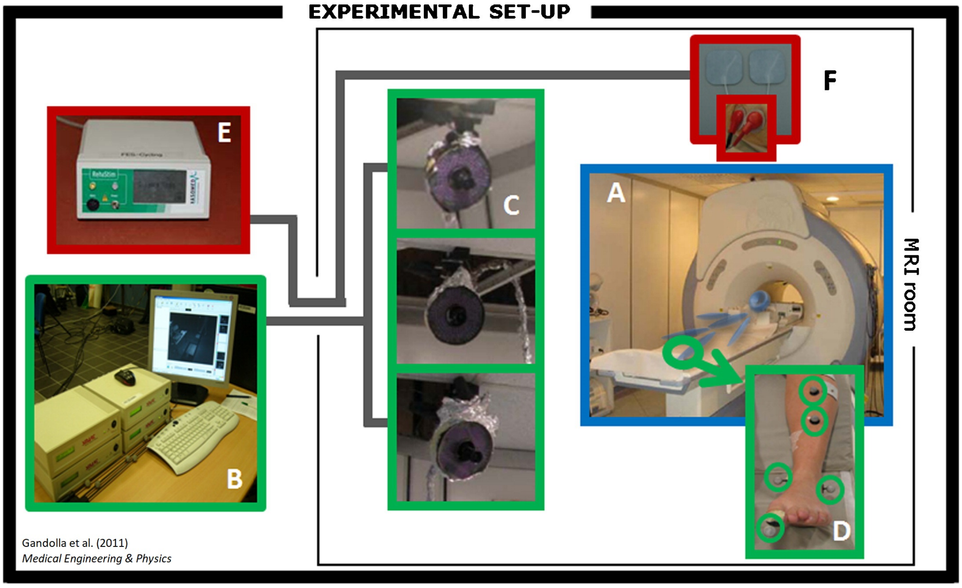


Figure 3.2: Experimental set-up. (A) MRI scanner CV/I 1.5 T; (B) motion capture system Smart μg with (C) three cameras and (D) reflective markers; (E) FES current-controlled stimulator RehaStim pro and (F) stimulation electrodes.

3.4 fMRI task design

The fMRI data used in this work were available thanks to the following experimental design conceived and employed in previous studies [6][8].
A 2 × 2 fMRI design was performed using right ankle dorsiflexion (ADF). Experimental factors included volitional intention to perfrom the movement [V: with the levels "volitional" and "passive"] and FES [F: with the levels "present" and "absent"]. During a continuous 10 minutes scanning session, subjects performed 20 alternate 9-seconds OFF and 21-seconds ON blocks. The 4 conditions that constituted the factorial design were performed during the ON blocks in a semi-randomised order: (i) FV: FES-induced ADF concurrently with voluntary movement by the subject; (ii) FP: FES-induced ADF, with no voluntary intention to perform the movement; (iii) V: voluntary ADF; (iv) P = passive dorsiflexion (by the experimenter) of the subject's ankle without FES. Subjects were specifically instructed to remain completely relaxed during FP and P conditions and to equally voluntarily contribute during V and FV conditions. Dorsiflexion was paced every $3.5s$ (for 6 repetitions within a block) with an auditory cue. The auditory cues were presented through an earphone. Prior to scanning, subjects practiced the protocol until they were comfortable with the task; the experimenter was assisting the training to check the correct execution of the protocol. All sub-

jects were free to choose the amplitude of their active movement to preclude fatigue. Subjects were instructed to keep eyes closed and head movements were minimized with rubber pads and straps. To ensure minimum transmission of movements to the head, knees were bent with the subject's legs lying on a pillow.

3.4.1 FES stimulation paradigm

Functional electrical stimulation was applied to the peroneal nerve through superficial self-adhesive electrodes, with biphasic balanced current pulses at $20Hz$ fixed frequency. The pulse width had a trapezoidal profile (maximum pulse width 400 μs) and the current amplitude was set subject by subject so as to reproduce the same movement amplitudes as during voluntary movements, within the tolerance threshold. Current amplitude and pulse width were kept the same for both FP and FV conditions.

3.5 fMRI pre-processing

Imaging data were analyzed using Statistical Parametric Mapping (SPM8, Wellcome Department of Imaging Neuroscience, `http://www.fil.ion.ucl.ac.uk/spm/`) implemented in Matlab (Matlab R2010b). A skull stripping procedure, on the structural image for each subject, was performed to improve the co-registration of functional and structural images. Participants with right-sided infarcts (left-leg weakness) were flipped about the mid-sagittal line, such that all subjects were considered to have left-sided infarcts. All fMRI volumes were then realigned and unwarped to suppress task-related motion artifacts [110]. Realignment parameters were assessed for excessive motion after unwarping procedure. A threshold of 4 mm in translation and 5 deg in rotation was applied [111]. The skull stripped structural image was then coregistered to the mean image of the functional realigned volumes, and segmented. The spatial normalization transformation (to the Montreal Neurological Institute (MNI) reference brain in Talairach space [112]) was then estimated using the segmented structural image. The structural image and functional volumes were normalized and resampled to $2mm \times 2mm \times 2mm$ voxels. Functional normalized images were then smoothed with an isotropic $8mm$ full-width half-maximum kernel [113]. The time series in each voxel were high pass filtered at 1/128 Hz during subsequent modeling to remove low frequency confounds.

3.6 Statistical analysis and DCM analysis

A detailed description of the construction and results of the statistical analysis, relative to the experimental factors (FV, FP, V, P), and their interaction, can be found in Gandolla et al. [6, 8].

In this work, we are interested in extracting BOLD response time-series, representative of specific regions of interest (ROIs), resulting from experimental inputs conveniently suited for our model to build a convenient model space to test our experimental hypotheses.

Namely, the general linear model was reformulated to specify the driving and modulatory experimental inputs. These comprised (i) a stimulus function representing the effect of descending voluntary signals, V (combining onsets from V and FV conditions, i.e. u_1); (ii) a second input encoding the contribution of ascending functional electrical stimulation to proprioceptive input, E (combining onsets from FV and FP conditions, i.e. u_2), and (iii) a third input representing underlying proprioceptive input from all movements, P (combining onsets from all conditions, i.e. u_3). All ADF onsets belonging to the same condition were defined as a single event type and modeled as delta (stick) functions in the corresponding stimulus function; all onset stimulus functions were then convolved with a canonical hemodynamic response function and used as regressors in the general linear model of the observed fMRI time series. Thus, for each subject, voxel-wise parameter estimates for each regressor were obtained.

To summarize the regional activity of each subject, an F-contrast was performed across all covariates of the new design matrix for each subject, obtaining a contrast image of brain activation for all three experimental inputs.

3.6.1 ROIs selection

The choice of the ROIs to be included in the (dynamic causal) model, was made on the basis of a priori knowledge about the role of cortical areas in motor control loop. In fact, given the purpose of the project, and depending on information found in literature, seleceted ROIs included contralateral leg primary sensorimotor cortices (M1 and S1), supplementary motor area (SMA) and the angular gyrus (AG). A representation of the ROIs approximate locations is reported in **Figure 3.3** As previously done by Gandolla and colleagues (2014) [6] in their DCM analysis, performed on healthy subjects with the same experimental control, M1 and S1 have been included in the model, as they have a fundamental role in the sensory-motor coupling, and they have been observed to be active in both healthy subjects and patients

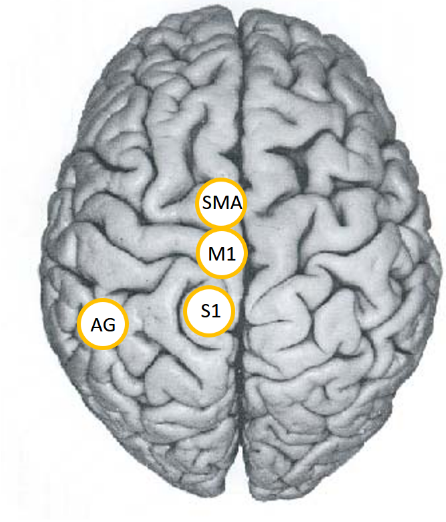


Figure 3.3: Representation of approximate ROI locations.

during the experiment in Gandolla et al. (2016) [8]. Sensory-motor integration allows an individual to take sensory information and use it to make motor actions that, in turn, can be used to modify the sensory system's response to future stimuli [114]. Flexible sensorimotor integration would allow an individual the ability to correct for errors [115]. To produce the desired flexibility the nervous systems employ the use of internal models and efference copies [114][115][116].

SMA is also known to contribute to the control of movement [12][13][14]; in fact, neurons in the SMA project directly to the spinal cord and may play a role in the direct control of movement. Possible functions attributed to the SMA include the postural stabilization of the body, the coordination of both sides of the body such as during bimanual action, the control of movements that are internally generated rather than triggered by sensory events, and the control of sequences of movements. Moreover, changes in activation in both contralesional SMA and sensorimotor cortices have been observed in post-stroke patients during motor rehabilitation [8][15].

AG is a fairly versatile and polyvalent region located in the posterior part of the inferior parietal lobule, that has been shown in numerous meta-analysis reviews to be consistently activated in a variety of motor and non-motor tasks . It is reported to be involved in different functions, including cognitive tasks, multi-sensorial integration, memory retrieval, attention and motor control [16][17]. In fact, AG is associated with high complexity tasks involving all the functions cited above, such as language processing

and speech production [18]. In the context of the motor control loop, it is believed that AG is responsible of both awareness of discrepancy between intended action and movement consequences, and awareness of action authorship. This region is assumed to be involved with higher order aspects of motor control, that allows one to consciously access different aspects of one's own actions. Hence, this region processes discrepancies between intended action and movement consequences in such a way that these can be consciously detected by the subject [117][8]. Although these are not the only cortical areas potentially contributing to the control of movement, SMA and AG were inserted into the model as they have been proven to be of particular interest in the context of observing differences in groups of subjects (i.e. CE, nCE and healthy subjects);
Indeed Gandolla and colleagues (2016) [8] observed different activation patterns in SMA and AG regions between healthy subjects, CE and nCE patients, during the experiment. Also for this reason they have been included as nodes in the dynamic model, with the hope to differentiate the behaviour of the two groups of patients and recognise markers for the carry-over effect.

Functionally, the choice of subject-specific ROI coordinates was informed by the SPM12 atlas of maximum probability tissue labels provided by Neuromorphometrics, Inc. (`http://Neuromorphometrics.com/`) under academic subscription. Subject-specific maxima were selected, for each region, inside the activation map afterwards the corresponding mask, taken from the atlas, was applied. Crucially, the SPM12 atlas does contain masks labeled for SMA and AG, while specific masks for M1 and S1 ”leg area” are not included; in these cases (M1 and S1) statistical parametric maps were masked using ”pre-central gyrus” and ”post-central gyrus” labeled masks (left sided), and then activation maxima were selected within these masked maps, with the researcher paying particular attention that those selected coordinates felt within the leg area of the corresponding ROI being selected. A further precaution consisted in ensuring the selected maxima, from the 4 different ROIs, had an euclidean distance of at least $8mm$ from one another, in order to avoid subsequent spatial superimposition of the $4mm$ radius spheres with which the volumes of interest have been approximated.
For each subject, regional responses were then summarized with the first eigenvariate of a sphere (4 mm radius) centered in the subject-specific maxima.

Crucially, we allowed E (altered proprioception) to modulate different connections or combination of connections, where modulation of self-connections corresponds to a modulation of intrinsic excitability. The rationale for modeling the modulatory effects of E is that we wanted to examine both the

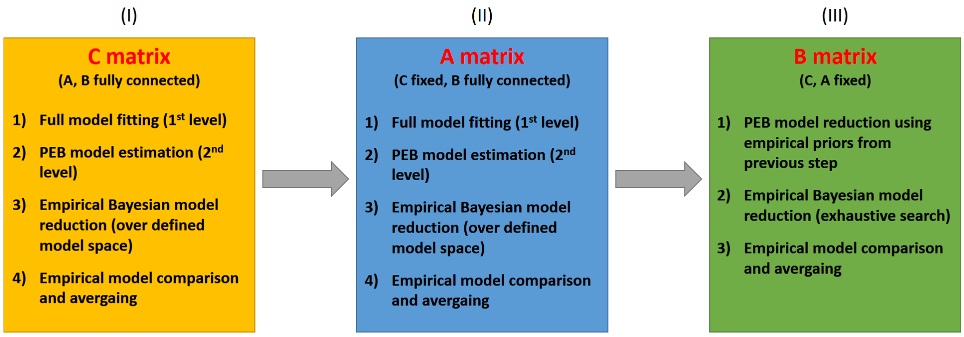


Figure 3.4: Schematic representation of steps involved in the procedure for model structure identification.

driving effects of stimulation and activity-dependent effects on cortical excitability or gain.

3.6.2 Model Selection procedure - inference about model structure

The basis of the DCM is represented by the bilinear state equation

$$\dot{z} = (A + \sum u_j B^j)z + Cu$$

as previously described in chapter 4.
Here, the parameters of interest used to assess effective connectivity are identified by the A, B and C matrices.
In this work, we were interested in investigating all three matrices structure, and in comparing their estimated values between the groups of recruited subjects. As mentioned in the previous chapter, the high number of parameters (connections) to be tested give rise to such a vast model space that would be difficult even to define all possible competing models it contains. To identify the three matrices structures, we proceeded with a customized hierarchical approach, analysing each matrix structure in succession, by exploring the corresponding model subspace. At each new step we fixed the structure of the matrices already identified in previous steps. This procedure is illustrated in a schematic in **Figure 3.4**.

Before proceeding with a more detailed description of this procedure, it is necessary to anticipate that it was decided to start by identifying model structure for healthy subjects; in subsequent analysis of patients, the structure of matrix C (inputs) was not investigated, but instead it was assumed to be the same as for healthy subjects. In the same way, during testing of matrix A (intrinsic connections), the structure identified for the healthy

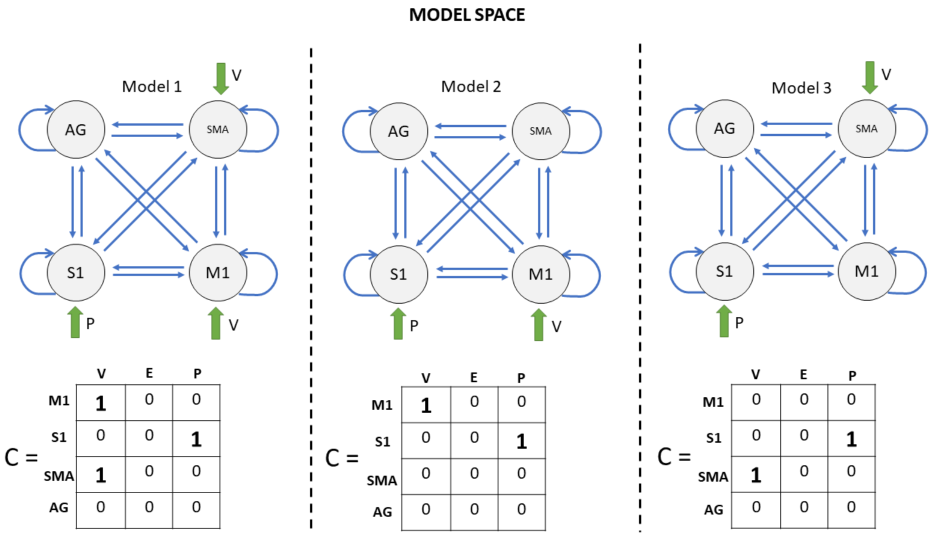


Figure 3.5: Model space relative to C matrix, containing 3 competing models

group was used as a starting point, exploring then the model space resulted from combinations of remaining connections. Matrices B (modulatory effects) were fully investigated in both groups.

In this procedure we decided to employ the Parametric Empirical Bayes (PEB) framework creating a hierarchical model, consisting of first level within-subject effects (subject specific DCM) and a second level between-subject group effects (GLM encoding mean and group effects), that permitted us to easily observe differences in groups of patients. Inference on model space was carried out, one matrix at a time, following Empirical Bayesian model reduction, comparison and averaging, similarly to the approach described by Friston and colleagues [19] in their technical note about Bayesian procedures for the analysis of group studies. This way we needed to fit only the full model at the first level, sparing to perform a (computationally expensive and time consuming) model inversion at the first level for each of the models in the model space. The results of inference about model structure, as well as results on inference about parameters, will be presented in the results chapter.

C matrix identification - Healthy subjects

The effects of experimental inputs were based upon prior knowledge about functional anatomy: V was assumed to drive either M1 or SMA or both, modeling top-down intentional signals during voluntary movements; P was assumed to drive S1, modelling the proprioceptive and somatosensory consequences of movement (e.g. ascending afferents from muscle spindles and Golgi tendon) that are known to convey information to sensory areas; E, instead, was not included between the driving inputs, as its onset functions (combination of FV and FP onsets) were already included in the P input (all onsets). On the other hand, E was considered only as modulatory input, as to isolate the contribution of electrical stimulation and better observe the modulatory effect of the latter, over intrinsic connections.

Therefore, 3 competing models have been defined to address the identification of the C matrix structure. These models differed only in terms of matrix C, as A and B matrices were, in this first step, assumed as fully connected. The 3 competing models are represented in **Figure 3.5**: model 1 had input P in S1 and input V both in SMA and M1, model 2 had input P in S1 and input V in M1, model 3 had input P in S1 and input V in SMA.

A hierarchical PEB model requires to define a design matrix at the second level, modeling between-subject effects. In this case, since we were analysing model structure of healthy subjects, there were no subgroups of subjects and thus the GLM at the second model has been modeled using only a single regressor, i.e. group mean (a 16 × 1 vector of ones):

$$\begin{bmatrix} Healthys \\ H01 \\ H02 \\ H03 \\ H04 \\ H05 \\ H06 \\ H07 \\ H08 \\ H09 \\ H01 \\ H11 \\ H12 \\ H13 \\ H14 \\ H15 \\ H16 \end{bmatrix} \longrightarrow \theta_i = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \cdot \beta_{mean} + \varepsilon, \tag{3.1}$$

with $\varepsilon \sim \mathcal{N}(0, \Sigma)$

Matrix A identification - Healthy subjects

In this step, C matrix has been fixed as resulting from the previous identification step. In the process of identification of the A matrix structure, the main issue was again the numerosity of model space. In fact, a model composed of 4 nodes, would result in thousands of competing models to test, supposing we want to test the presence/absence of all connections. To solve this problem, we introduced a number of constraints on model connectivity, based on prior knowledge about anatomical and functional connections, in order to reduce the size of the model space to explore. M1 and S1 bidirectional conncections ($M1 \longleftrightarrow S1$) have been fixed and, thus, excluded from the analysis, by virtue of their anatomical connections [118][119]. This constraint was also based on previous DCM studies involving M1 and S1 areas [6].

A second constraint was placed on the bidirectional connections between SMA and M1 ($M1 \longleftrightarrow SMA$); indeed, anatomical projections between these two areas are reported in literature [120][14]. Regarding the connection from M1 to SMA, this was assumed to be a functional feedback connection, most likely via reentrant feedback from M1 to basal-ganglia-thalamo-cortical circuits, which link the SMA and are known to be crucial for the initiation and control of voluntary action [14]. All the other bidirectional connections between remaining regions ($M1 \longleftrightarrow SMA, SMA \longleftrightarrow S1, SMA \longleftrightarrow AG, S1 \longleftrightarrow AG$) were tested, as no information was found in literature about the presence of anatomical projections and little knowledge is reported about indirect functional connectivity between those areas; thus, we considered reasonable to properly test the presence or absence of such connections in the model.

A schematic illustration of connections to be tested is reported in **Figure 3.6**.

The remaining connections to be tested, would still produce a fairly wide model space; for this reason, we decided to follow a 2-step identification of matrix A:

1. The first step consisted in defining the model space by forcing connections to be bidirectional (i.e. considering 2 nodes, for example, either both connections are discarded, or are present in both directions); this led to a quite small model space of 16 competing models.

2. Once the models at the first step were compared, and possibly one or more bidirectional connections were discarded, we proceeded with the definition of a model space, composed of models obtained by combina-

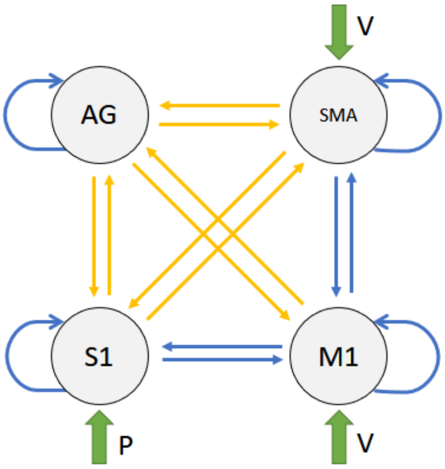


Figure 3.6: Schematic illustration of the A matrix connections subjected to test; in blue are connections that have been fixed based on prior knowledge, in order to reduce the model space, in yellow are the connections which combination determine the model space. Note that C matrix has been fixed has from the winning model from the previous step

tions of the remaining monodirectional connections; having discarded one bidirectional connection at the first step, this led to a new model space composed of 64 competing models.

Matrix B identification - Healthy subjects

As described before, input E (electrical stimulation) was the only input allowed to modulate intrinsic connections, while not being considered as driving input. Moreover, no constraints were placed on modulatory effects, i.e. all connections (excluded the missing ones in A) were allowed to be modulated by E and thus subjected to test.

Matrix identification in Patients

Model structure of patients' group, as discussed previously, has been investigated starting from few assumptions, derived from the structure identification of healthy subjects. In particular, C matrix structure was assumed fixed (same structure for patients and healthy subjects), while the A matrix structure was still reviewed, allowing to possibly discard one or more intrinsic connections, but using as starting point the structure obtained for healthy subjects (i.e. we hypothesised that patients can not exhibit effective connections that are missing in controls' group). On the other hand, no constraints were placed on the A matrix structure, i.e. no connections were forced to be existent, as patients, due to neurological damage, could have different connectivity topology with respect to healthy ones. Consequently, instead of explicitly define all models, an exhaustive search at second level over the entire model space was applied.

The B matrix structure was inspected starting from a structure corresponding to that of matrix A (i.e. intrinsic connections that have been discarded can not be modulated).

Basically, this procedure was aimed at making inference about model structure and model parameters, with the purpose of comparing the structure and values of effective connectivity among the different groups of subjects. Unlike the healthy subjects group, patients were divided in CE and nCE subgroups. Each of these two groups was composed of 4 subjects, for each of which 2 fMRI scans, related to different sessions, were available, i.e. one before, PRE, and one after POST the rehabilitation.

For this reason, as to better evaluate connectivity differences, we decided to construct the design matrix of group effects at the second level as follows; besides a group mean regressor, we included a carry-over regressor, encoding the subject classification (CE = 1, nCE = -1) and a time regressor, encoding the session time point (PRE = 1, POST = -1). Moreover, having a comprehensive clinical score to assess a functional improvement for each patient, a further capacity score regressor has been included, as to take into account for the inter-subject variability, in the connectivity parameters values, due to the residual motor capability (potentially being source of confounding effects). Subject-specific capacity scores were mean-corrected and normalized between [-1,1], before being included in the design matrix.

The resulting second level linear model was:

$$\begin{bmatrix} Patients \\ P01_{t1} \\ P02_{t1} \\ P03_{t1} \\ P04_{t1} \\ P05_{t1} \\ P06_{t1} \\ P07_{t1} \\ P08_{t1} \\ P01_{t2} \\ P02_{t2} \\ P03_{t2} \\ P04_{t2} \\ P05_{t2} \\ P06_{t2} \\ P07_{t2} \\ P08_{t2} \end{bmatrix} \longrightarrow \theta_i = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \cdot \beta_{mean} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix} \cdot \beta_{CE} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix} \cdot \beta_{time} + \begin{bmatrix} 0.49 \\ -0.24 \\ 0.60 \\ 0.27 \\ 0.61 \\ 0.22 \\ 0.82 \\ 0.42 \\ -1 \\ 0.87 \\ 1 \\ 0.12 \\ -1 \\ 0.92 \\ 0.78 \\ 0.08 \end{bmatrix} \cdot \beta_{CS} + \varepsilon \quad (3.2)$$

with $\varepsilon \sim \mathcal{N}(0, \Sigma)$.

3.7 Objectives of the study and starting hypotheses

The main objective of this work is to investigate and possibly identify the effective connectivity features that ma discriminate patients able to achieve the carry-over effect from other that do not show stable improvements.

In general, we want to gather further insights in the motor control circuits, with the aim of corroborating the Active Inference alternative account for motor control. In this context, borrowing the idea from the predictive coding framework, we assumed our model as a hierarchy, with different regions (or nodes) operating at different hierarchical levels, each acting as interface between converging bottom-up errors and top-down predictions. This recurrent exchange proceeds until prediction error is minimised at all levels of the system.

Also, in order to better understand motor loop updating mechanism related to brain plasticity processes during motor recovery, we are interested in observing the longitudinal changes in effective connectivity, in patients attending FES-based motor rehabilitation treatment.

Based on the work done and the results reported in previous studies, and based on researches in the literature, we have some starting hypotheses:

- The first hypothesis, as anticipated, is to assume the Active Inference account for motor control. In this view, we expect to observe an "update" of the information about the proprioceptive prediction sent from SMA and/or M1 toward S1, as the efference copy of the signal sent to spinal cord centres of motor control. "Update" here means we expect FES to have a modulatory effect over this connection (B matrix). Conversely, an abnormal behaviour of these connections could be observed in nCE group as possible cause of the missed of achievement of carryover effect.

- Secondly, we expect to observe a possibly negative value of the estimated mean value of the intrinsic (A matrix) and extrinsic (B matrix) connection from SMA to M1, as the result of the suppressive activity that SMA exerts over M1 [14, 121, 12], in abnormal environmental conditions, in order to suppress standard motor programs in favour of alternative ones.

- A third hypothesis is based on the fact that the AG is thought to compute action awareness [8] and to be responsible for detecting discrepancies between intended action and movement consequences: we expect to observe, at least in a healthy control circuit, non-zero connections from SMA/M1 and S1 toward AG, as this region should integrate both sensory proprioceptive information and motor planning information. From the same infromation, and based on previous findings about abnormal activation of AG in nCE patients, we expect to detect differences in parameters between CE/nCE patients, indicating different activation in this region (in nCE there could be a less precise mechanism of motor planning resulting in increased computed prediction error, i.e., discrepancy between intended and performed action).

- Finally, we expect to observe similarities between healthy subjects and CE patients, indicating similar activation in SMA between the two groups, as reported in [8].

Chapter 4

Results and Discussion

4.1 Results

4.1.1 Participants

Activation maps analysis for the healthy controls group and the patients have been previously descrived in [6][8], and their analysis goes beyond the goal of this work.
MNI coordinates of mean ROI baricenters, along with their variability (±) in terms of standard deviation are reported for healty group and patients, in **Table 4.1** and **Tabel 4.2**, respectively.

4.1.2 DCM results

In the previous chapter, the procedure employed to define and explore the model space has been discussed along with the methods for assessing parameters estimates at the group level. Below are the results about inference on model structure, first, and inference about parameters of the resulting model, then.

Inference about model structure

As described in Chapter 5, the identification of model structure was carried out in a step-wise process, defining at first healthy subjects model structure (one matrix at a time) and then reviewing patients model structure.

Concerning C matrix structure identification, the posterior probabilities of the three specified competing models are shown in **Figure 4.1a**. As can be seen, the posterior probability of $Model$ 1 over competing models was 0.81, resulting in a strong evidence supporting this model. Therefore, the

	x	y	z
M1	-4.3 ± 3.9	-30.8 ± 9.9	64.8 ± 10.9
S1	-7.6 ± 3.8	-43.9 ± 10.6	66.8 ± 9.2
SMA	-4 ± 2.5	-10.1 ± 13.4	65.8 ± 6.4
AG	-37.8 ± 10.1	-67.5 ± 6.4	43 ± 9.4

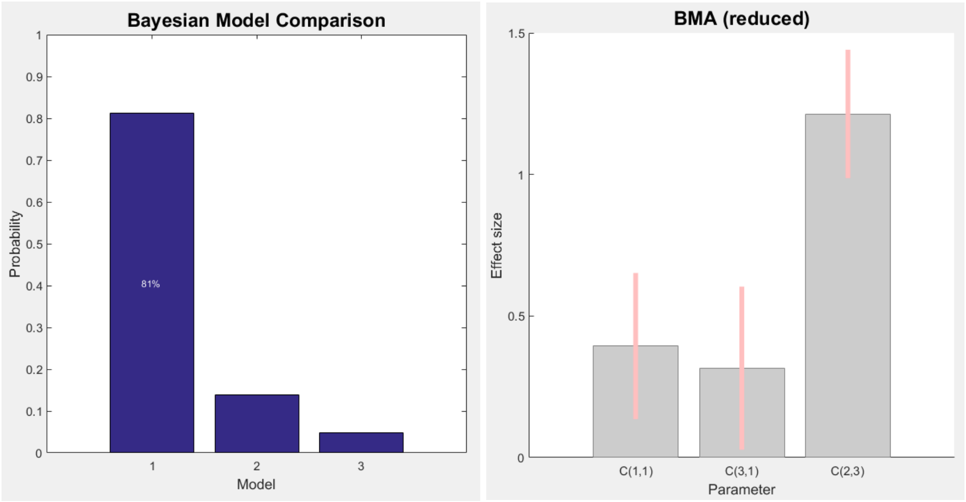
Table 4.1: Healthy ROI coordinates (mean ± standard deviation).

	x	y	z
M1	-4.3 ± 3.2	-23 ± 10	67 ± 8.1
S1	-4.4 ± 4.3	-35.4 ± 9.3	67.4 ± 6.5
SMA	-3.3 ± 2.7	-11.5 ± 9.4	67.4 ± 6.1
AG	-46.9 ± 7.2	-59.1 ± 9	33.6 ± 12

Table 4.2: Patients ROI coordinates (mean ± standard deviation).

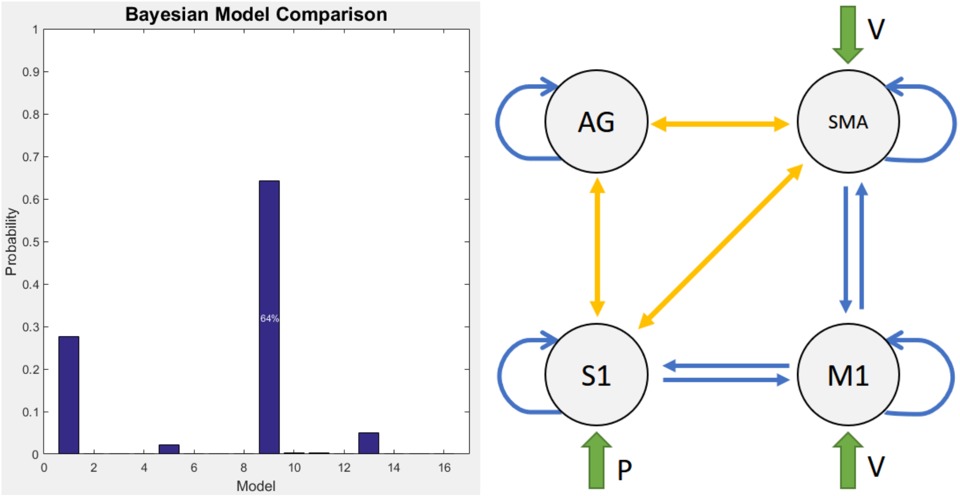
resulting parameters estimated values were strongly affected by estimates of *model* 1, being the model with far greater weight in the subsequent Bayesian model average, reported in **Figure 4.1b**.
$Model$ 1 indicates the presence of input V on both M1 and SMA, suggesting that the voluntary intention to execute a movement is processed by the two cooperating areas.

During the study of A matrix structure, a 2-phase method has been used. The first phase consisted in the definition (and scoring) of a model subspace which elements were formed by combinations of bidirectional connections. Posterior model probabilites for each of the 16 specified competing models are reported in **Figure 4.2a**. As shown, the best model had a posterior probability of 0.64 over competing models (positive evidence), and was the one where the bidirectional connection between M1 and AG has been discarded, as represented in **Figure 4.2b**. Although not equipped with a strong evidence, the "winning" model revealed a poor contribution of the bidirectional connection $M1 \longleftrightarrow AG$ to the model evidence. This result is in agreement with the lack of evidence in the literature supporting the presence of functional or anatomical connections between the two areas.
In the second phase, the former result was applied, i.e. the bidirectional connection $M1 \longleftrightarrow AG$ was discarded from the model; combinations of the remaining 6 connections produced a model space of 64 competing models. Posterior probabilities for the 64 model are shown in **Figure 4.3a**. The model with the grater posterior probability corresponds to the "full" model, that, however, had a posterior probability lower than 0.5 (weak evidence).



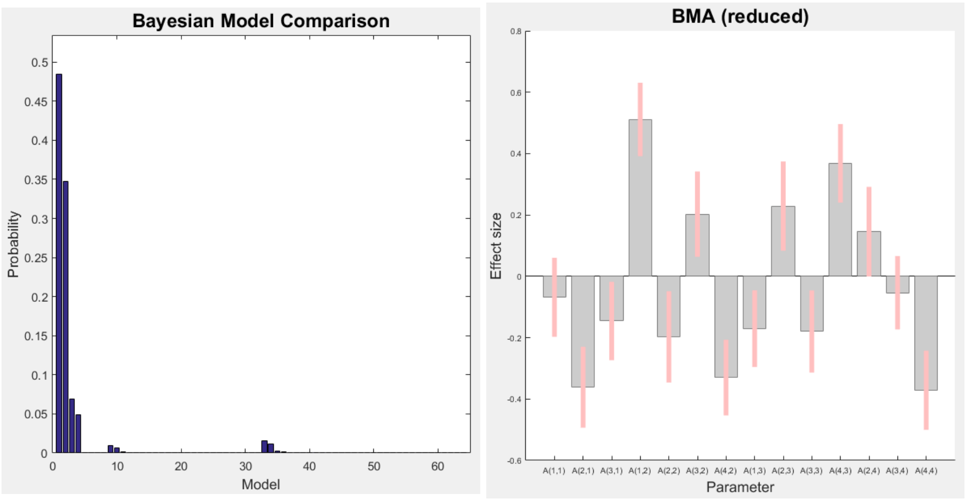
(a) *(b)*

Figure 4.1: (a) Posterior probabilities among 3 competing models; (b) Representation of posterior expectations (grey bars) and their 95% confidence intervals (pink bars), after BMA.



(a) *(b)*

Figure 4.2: (a) Posterior probabilities among 16 competing models; (b) schematic of resulting best model: yellow connections are those who were subjected to test and survived.

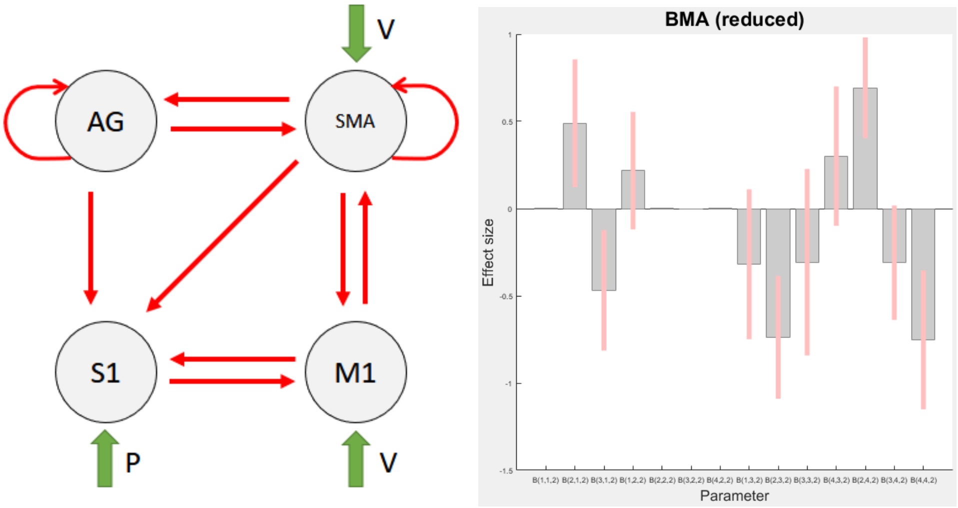


(a) *(b)*

Figure 4.3: (a) Posterior probabilities over 64 alternatives for the A matrix; (b) representation of posterior expectations (grey bars) and their 95% *confidence intervals (pink bars), after BMA.*

Therefore, in the subsequent Bayesian model averaging (**Figure 4.3b**) over the models within the Occam's window, no further connections were discarded from the model, even though some of them had distibutions whose means were not significantly different from zero. This will be further discussed in the next paragraph about inference on model parameters.

Concerning the analysis of B matrix structure, model space was not explicitly defined, as no a priori hypotheses were made about modulatory effects of input E. Thus, an exhaustive search over all possible combinations of B matrix connections has been performed. A schematic representation of the resulting model is showed in **Figure 4.4a** along with the representation of parameters estimates after BMA (**Figure 4.4b**). In red are represented the connections modulated by input E. As described in Chapter 5, model structure of patients group was obtained starting from the results coming from Healthy subjects group, i.e. with respect to matrices A and C, they were initially defined as obtained from healthy subjects analysis and then further inspected, by performing a greedy search over all possible combinations of remaining parameters (B matrix was initialised as full and then tested). Furthermore, in patients group, PEB model has been defined using a design matrix containing several regressors, modeling different group effects, such as carry-over effect and time, and clinical measures (i.e. capacity score).

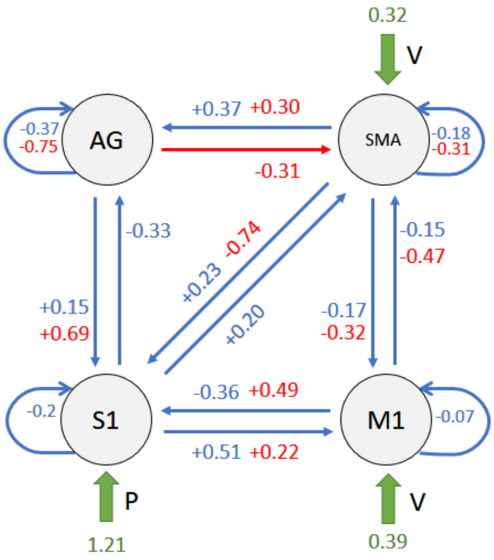


(a) (b)

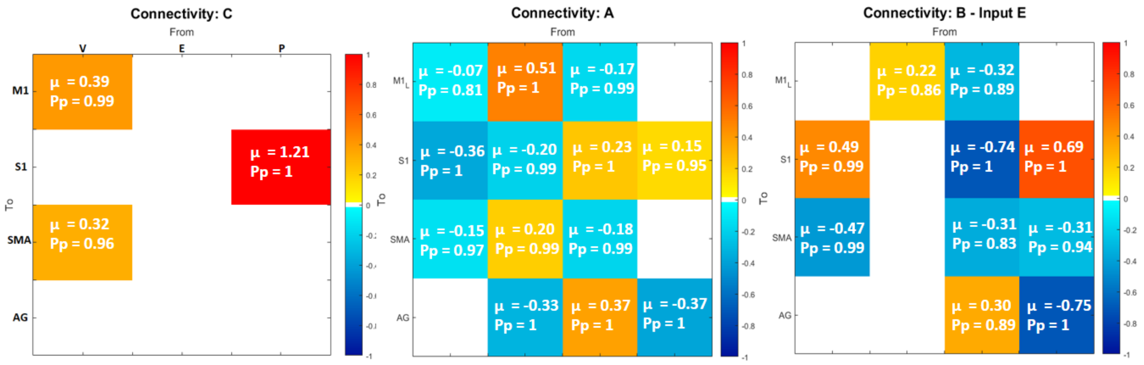
Figure 4.4: (a) Representation of surviving B matrix connections; (b) representation of posterior expectations (grey bars) and their 95% *confidence intervals (pink bars), after BMA.*

Inference about parameters

For what concerns healthy subjects a representation of the resulting model is reported in **Figure 4.5a**. Here a schematic of the model is shown, with blue values indicating intrinsic connectivity estimates and red values indicating modulatory effects of input E. Green values indicate contributions of voluntary and proprioceptive inputs (V and P). In **Figure 4.5b** a graphic representation of individual matrices is reported, with associated parameters estimates and their significance (i.e. probability of being different from zero). Reported values have been thresholded for a significance $P_p \geq 0.8$, i.e. estimates with lower significance has not been reported.
Results about patients model structure and parameters group estimates are reported in **Figure 4.64.74.8**. Here parameter estimates are reported (with color levels) for each first level matrix, highlighting each regressor contribution (second level estimate). As anticipated, in patients analysis second level GLM was modeled using several regressors included in the design matrix; assuming there are n parameters at the first level (A, B, C matrices), and l regressors at the second level, the inversion of a PEB model would produce a set of $n \times l$ parameters. Basically, that corresponds to split each of the three matrices at the first level into l matrices at the second level, representing estimates of the group effects included. In our case we had 4 regressors



(a) Model connectivity structure. Overall representation of final resulting model for healthy subjects; blue values are relative to intrinsic connections in matrix A, red values are relative to modulatory effects of input E (FES) in matrix B, green values indicate driving inputs (V, E, P) contribution in matrix C.



(b) Matrix representation. Estimated posterior means μ and their significance level P_p, expressed for each first level matrix (from left to right: C - driving inputs; A - intrinsic connectivity; B - modulatory effects of input E).

Figure 4.5

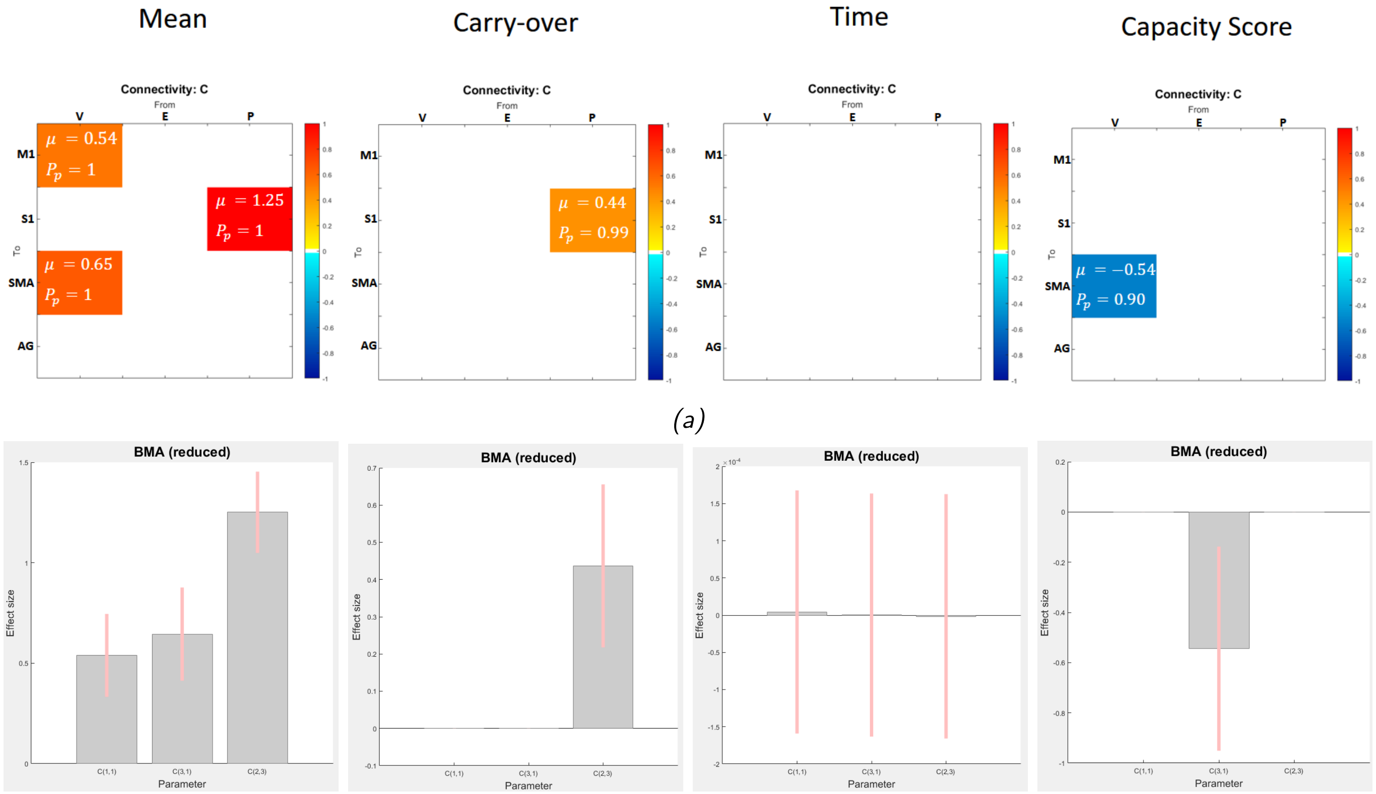
modelling mean, carry-over, time and capacity score; among these, we are mostly interested in the carry-over and time regressors, in order to differentiate the behaviour within 4 patients subgroups, i.e. CE-pre, CE-post, nCE-pre and nCE-post. Schematics of the models for these 4 groups are reported in **Table 4.9**, along with the overall parameters values, resulting from the contribution different regressors; here the contribution of the capacity score regressor has been intentionally ignored, as to isolate and neglect the inter-subject variability component associated to residual capability, that could behave as a confounding effect. In order to address punctual values of first level parameters, considering the contribution of all group effects (except for capacity score), estimated second level parameters have been thresholded, using a significance threshold of $P_p \geq 0.80$. Then, the overall distribution for each parameter at the first level have been calculated as sum of Gaussian distributions, i.e. having a mean equal to the sum of means and a variance equal to the sum of variances; after this step a further significance threshold has been applied on the resulting overall distribution. It must be taken into account that distribution means of second level estimates are multiplied by the regressor, that indeed express different contributions across subjects. We remember that the design matrix regressors, reported in Chapter 5, were defined as follows: the carry-over regressor contained $1s$ for CE patients and $-1s$ for nCE subjects; the time regressor contained $1s$ for PRE sessions and $-1s$ for POST sessions. Note also that, in recent versions of SPM software, as the one used during this work (SPM12), self connections are hyper-parameterised as $S = -0.5 \times e^{(A_{ii}+B_{ii})}$, in order to ensure them to have negative values and therefore maintain the stability of the dynamic system.

Inside **Figures 4.6, 4.7, 4.8**, self connectivity log-values are reported, as estimated by the SPM software. In **Table 4.9**, instead, self connectivity values have been transformed (through the exponential function) and reported as if they were normal rates of change of activity in the regions of the dynamic system.

4.2 Discussion

The aim of this thesis was to examine the longitudinal changes in motor control loop mechanisms, influenced by FES treatment during motor tasks, and to investigate the underlying features of carry-over effect, by evaluating differences among healthy subjects and groups of patients.

One fundamental hypothesis we started from, and we hoped to confirm, was the assumption of the Active Inference account for motor control as

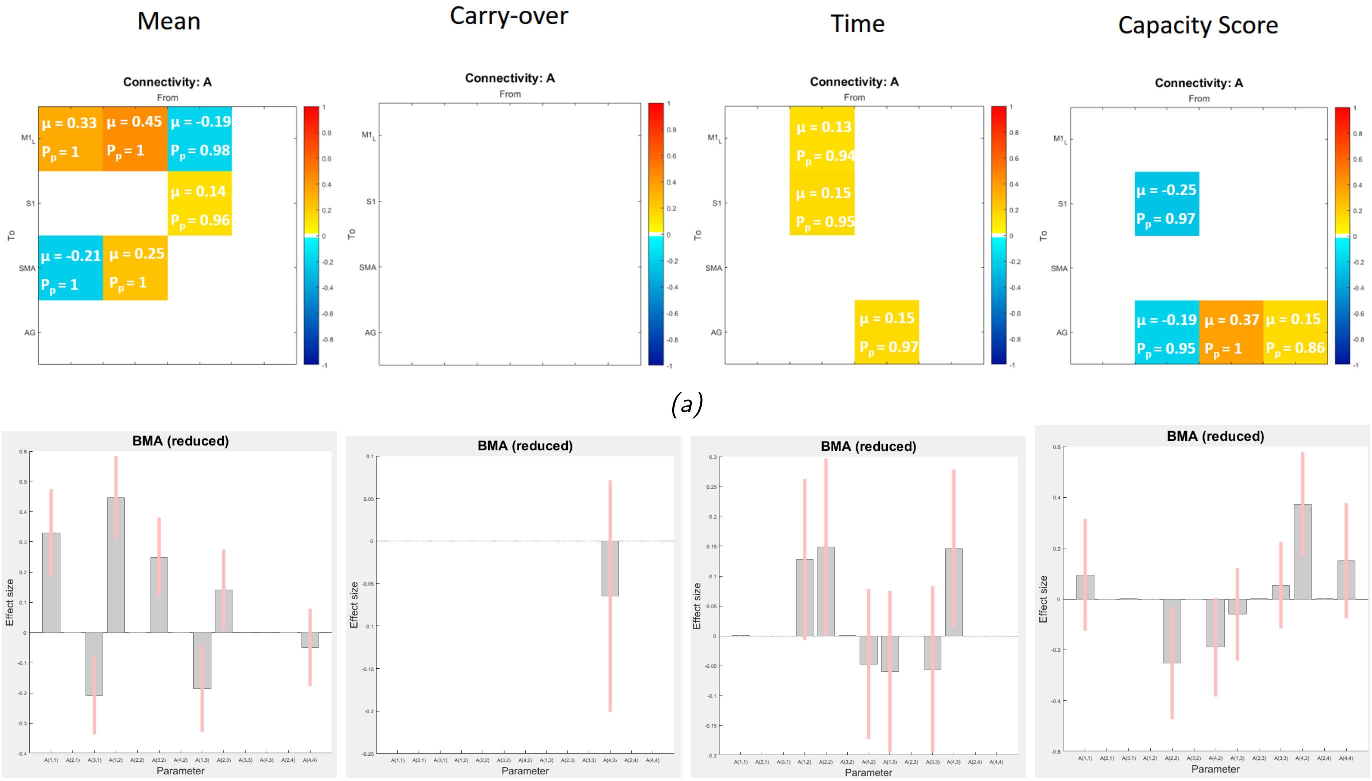


(a)

(b)

Figure 4.6: (a) Second level estimated parameters (thresholded for $P_p \geq 0.8$) relative to C matrix, explicited for each regressor (μ is the parameter posterior mean, P_p is the probability of the parameter to be different from zero); (b) estimated means (grey bars) along with their 95% confidence intervals (pink bars).
From left to right panels represent: group mean (β_{mean}); carryover effect ($\beta_{carryover}$); time (β_{time}); capacity-score (β_{CS}). Colour levels express the magnitude of the value, i.e., red tones are for positive values, blue tones are for negative values.

underlying working mechanism. In this context, we borrowed the idea form the popular predictive coding framework, as reviewed by Friston and collegues in 2010 [9], that models the brain as a hierarchical inference engine, trying to optimize probabilistic representations of what caused its sensory inputs. From this perspective, different brain regions operate at different hierarchical levels, acting as interfaces between the converging bottom-up errors and the top-down predictions. Top-down predictions are conveyed by backward connections and are compared with the representations being generated within the region, with the difference between the two reflecting the prediction error. This prediction error is then forwarded to higher levels to



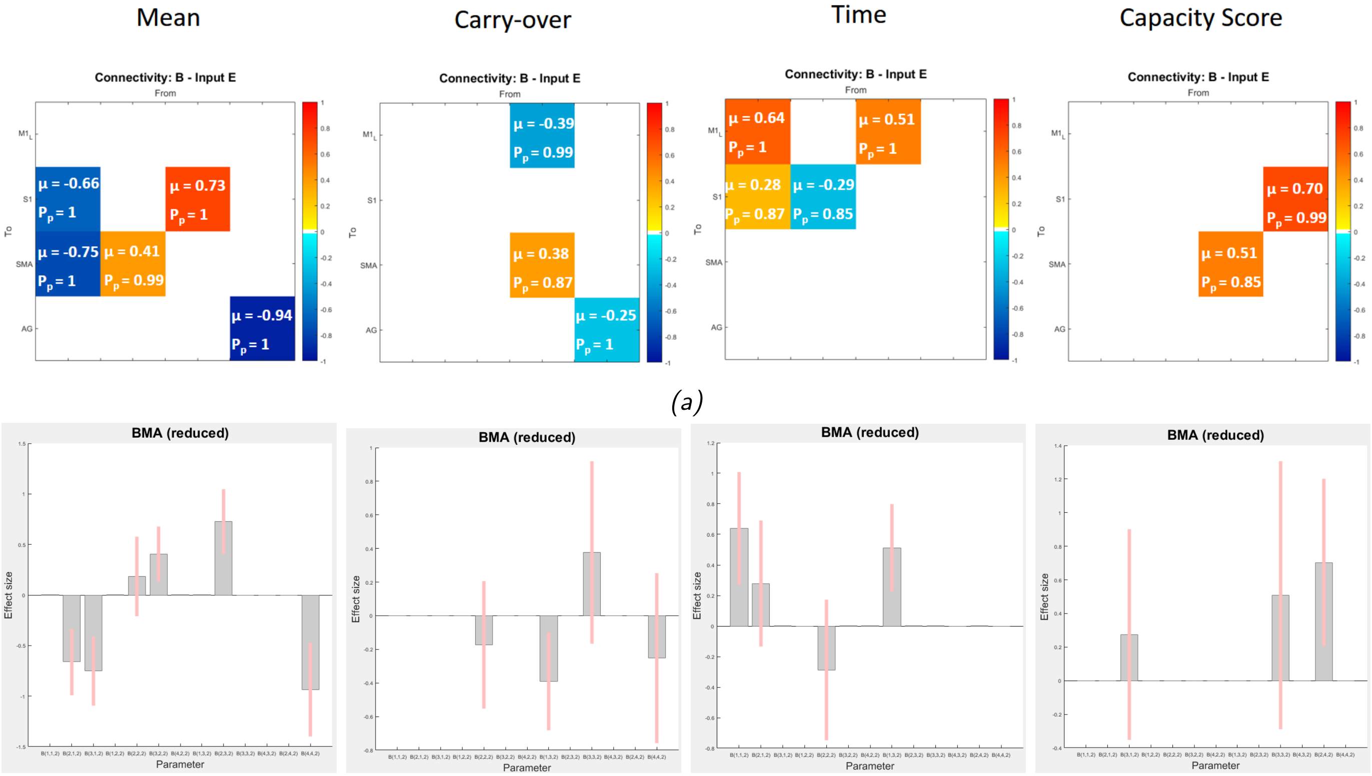
(a)

(b)

Figure 4.7: (a) Second level estimated parameters (thresholded for $P_p \geq 0.8$) relative to A matrix, explicited for each regressor (μ is the parameter posterior mean, P_p is the probability of the parameter to be different from zero); (b) estimated means (grey bars) along with their 95% confidence intervals (pink bars).
From left to right panels represent: group mean (β_{mean}); carryover effect ($\beta_{carryover}$); time (β_{time}); capacity-score (β_{CS}). Colour levels express the magnitude of the value, i.e., red tones are for positive values, blue tones are for negative values.

adjust and optimize the predictions. The recurrent exchange of bottom-up prediction errors and top-down predictions proceeds until prediction error is minimized at all levels of the system. Each hierarchical level contains both error units, computing prediction error, and state units, processing the desired prediction. Therefore, error-units receives messages from the state-units of the same level and the level above, and state-units are driven by error-units of the same level and the level below.

In our model, M1 and S1 act as low level areas, in that they represent input-output nodes of the system: input is ascending sensory feedback from the peripheral proprioceptors (in muscles, tendons, joints) reaching S1, while



(a)

(b)

Figure 4.8: (a) Second level estimated parameters (thresholded for $P_p \geq 0.8$) relative to B matrix, explicited for each regressor (μ is the parameter posterior mean, P_p is the probability of the parameter to be different from zero); (b) estimated means (grey bars) along with their 95% confidence intervals (pink bars).
From left to right panels represent: group mean (β_{mean}); carryover effect ($\beta_{carryover}$); time (β_{time}); capacity-score (β_{CS}). Colour levels express the magnitude of the value, i.e., red tones are for positive values, blue tones are for negative values

output is descending prediction signal (desired proprioceptive consequences) sent from M1 toward skeletal muscles, through the pyramidal tract. SMA an AG are assumed to be higher order areas, operating at the upper level in the hierarchy, responsible for higher processing of preditions and prediction errors. In fact, the supplementary motor area (SMA) has long been thought to have a special role in the internal generation of complex movements [122][123][124][125] and so to be responsible for higher order aspects of motor planning. SMA is also known to be involved in the organization of movements through a suppressive influence of standard motor planning in favor of an alternative one, when environmental conditions changes are

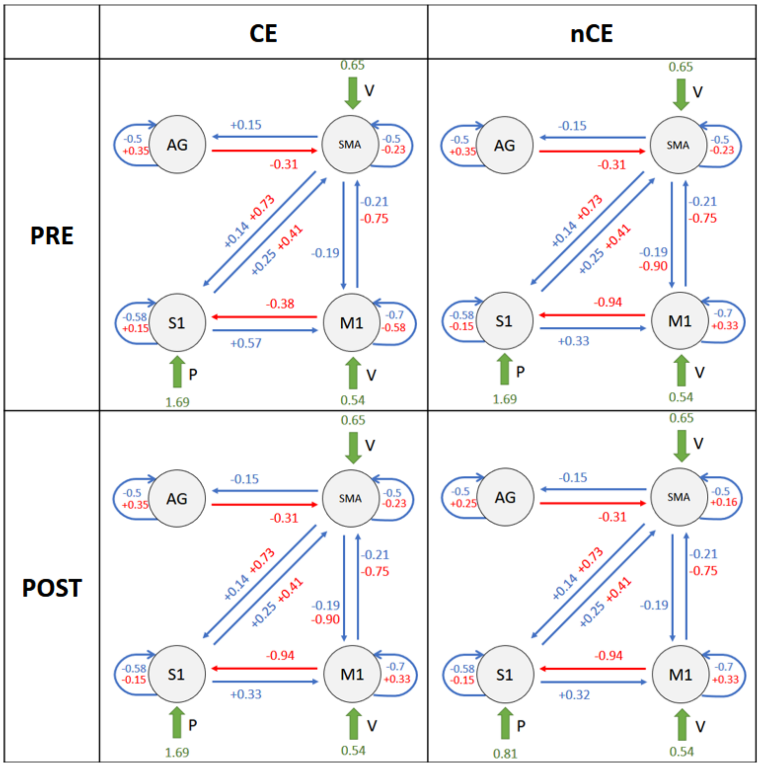


Figure 4.9: 2×2 factorial representation of patients model structure, along with estimated parameter posterior means. Values are obtined as $\theta_i = \beta_{mean} \pm \beta_{carryover} \pm \beta_{time}$. Blue values are relative to intrinsic connections in matrix A, red values are relative to modulatory effects of input E (FES) in matrix B, green values indicate driving inputs (V, E, P) contribution in matrix C. Values are reported for each subgroup, localised as follows: in the upper left panel is the CE-PRE subgroup; in the upper right panel is the nCE-PRE subgroup; in the lower left panel is the CE-POST subgroup; in the lower right panel is the nCE-POST subgroup.

in place [14][126][127][12][121]. AG is assumed to be a multi-sensorial integration hub [101], also able to process discrepancies between intended action and movement consequences, in such a way that they con be consciously detected by the subject [117].

In such a working architecture, the top-down predictions are based on previous knowledge of the external world, similar learned experiences that can be retrieved, and the awareness of own action (sense of agency). They may also come from other subsystems that maintain the intention (i.e., the planned action/decision to be made) and the saliency and the priority given to particular events of interest.

$M1/SMA \longrightarrow S1$ connections

Regarding the assumption of Active Inference account for motor control, important parameters to analyse are those identifying connections directed from primary and supplementary motor areas toward primary somatosensory area ($M1/SMA \longleftrightarrow S1$), in that these regions (SMA, M1) are the main responsible for the processing of motor plans and the aforementioned connection should carry the proprioceptive prediction of the voluntary movement to S1, within which this prediction should be compared with the true propiroceptive information coming from the periphery, to generate an error. According to Active Inference theory, then, we would expect that the presence of FES, perturbing the proprioception, would results in an update of the prediction sent from motor areas toward S1.

Considering healthy patients group, we observed a mean value of -0.36 in intrinsic connectivity (i.e. A matrix) from M1 to S1 and a value of 0.49 of the modulatory contribution of input E over this connection (i.e. B matrix). On the other hand, a reverse behaviour has been observed on the connection from SMA to S1, which estimated values are 0.23 for intrinsic connectivity and -0.74 for modulatory effects. Our interpretation recalls the concept of sensory suppression, widely reported in the literature [128][129][130][131]. This phenomenon involves a reduction in sensory perception that occurs during movement, also called movement-related gating. This function of sensory suppression is most likely to suppress redundant movement-related feedback that can be predicted from the motor commands so that the perception of unexpected or novel inputs is enhanced. The central motor command plays a key role in generating sensory suppression. Peripheral feedback from the moving limb also contributes because sensory suppression is seen during passive movement as well, i.e. in the absence of a motor command. We suggest that, in healthy subjects, the observed suppressive influence that M1 exerts over S1 in normal conditions (no FES), it is due to the sensory suppression mechanism. During FES, instead, SMA exhibit a suppressive effect over S1 while M1 to S1 connection become positive. This behaviour can be explained thinking about the role of SMA in higher order processing of complex movements and its function of suppression of habitual motor programs in presence of environmental changes. We propose that, because ankle dorsiflexion is a fairly simple movement for subjects with no neurological impairment, in absence of FES, M1 is mostly in charge of movement programming (SMA has a minor contribution due to simplicity of movement), and, accordingly, seonsory suppression is exerted mostly by M1. Conversely, during FES, the contribute of SMA becomes important, resulting in the sup-

pression of M1 activity (see $M1 \longleftrightarrow SMA$ section for details) in favour of an alternative motor planning; hence, it is SMA responsible of the sensory suppression, of predicted sensory causes of the movement, over S1.
In patients, the intrinsic connection from M1 to S1 seems to be salient, while the modulation parameter has an estimated value of $-0.66 \pm 0.28_t$, where 0.28_t represents the estimated parameter of the time regressor ($PRE = 1$, $POST = -1$). The connection from SMA to S1 has been estimated to be positive in both intrinsic and modulatory connections, with values of 0.14 and 0.73 (no group effects), respectively. As can be seen, the behaviour of modulatory effects is opposite with respect to healthy subjects, indicating possibly a misbehaviour of the updating the proprioceptive prediction sent to S1 and, therefore, of the tactile suppression mechanism. In patients, these functions are compromised, probably due to the lack of self-confidence about motor program.

$S1 \longrightarrow M1/SMA$ connections

In the reverse direction, representing the prediction error sent back from S1, for the connection $S1 \longrightarrow M1$, estimated values are 0.51 (intrinsic) and 0.22 (extrinsic), for healthy subjects and $0.57 \pm 0.13_t$ (intrinsic) for patients, where 0.13_t is the estimated value for time regressor; extrinsic connection in patients was discarded. For connection $S1 \longrightarrow SMA$ estimates are 0.20 for intrinsic connectivity of healthy subjects (extrinsic parameter was discarded), while for patients intrinsic connectivity value is 0.25 and extrinsic connectivity value 0.41. FES has modulatory effects on $S1 \longrightarrow M1$ but not over $S1 \longrightarrow SMA$, in healthy subjects. This reconfirm the hypothesis of the different contributions that these two regions have in the motor planning between the two groups: for healthy subjects, ankle dorsi-flexion represents a simple movement, requiring little assistance from SMA in the programming of the movement, while for impaired patients, M1 contribute is not sufficient and the processing support of SMA is required. In fact, modulation of FES in patients can be observed over $S1 \longrightarrow SMA$ connection but is absent on S1 to M1. Moreover, in patients, a time effect is observed over the connection from S1 to M1, showing a slight decrease of the magnitude of this connection over time. This result might be explained by the fact that patients attending FES based rehabilitation, gets more used to electrical stimulation as time passes by, and the ”surprise” effect decreases as this procedure becomes habitual, i.e. the prediction error signal sent from S1 to M1 decreases as rehabilitation proceeds forward.

$M1 \longleftrightarrow SMA$ **connections**

A second aspect on which we focused, was the bidirectional connectivity between M1 and SMA. In fact, we know from literature that supplementary motor areas exert a suppressive action over primary motor areas, whenever a perturbation of standard external conditions is present [14][21][22]. This suppression of habitual motor planning happens in favor of alternative predictions planned by SMA. Concerning SMA to M1 connection, in healthy subjects the intrinsic value was estimated to be -0.17 , and the modulatory effect -0.32. The fact that the experimental context represents an unusual situation (experimental set-up with MRI, auditory cues, first time FES) could be the cause for the suppressive activity of SMA even in normal conditions, i.e. without electrical stimulation. Accordingly, the presence of FES, introducing a further disturbance of normal considitons, enhances this effect. Indeed, FES artificially induces part of the movement, in sinergy with muscle activation. It can thus help the completion of the desired action (for patients), or it can only introduce a disturbance.
In patients, the estimated mean value for intrinsic connectivity is -0.19 (no group effects), consistent with healthy subjects behavior. This is in agreement with the findings reported by Gandolla and collegues in 2016 [8] about the similarity in activation patterns of SMA between CE patients and healthy subjects. In modulatory effects, differences in both CE/nCE patients and PRE/POST sessions have been observed; in particular, estimated parameters values are $\mp 0.39_{ce}$ for the carry-over regressor and $\pm 0.51_{t}$ for the time regressor. This means that the component of modulatory effects relative to the time course of the treatment goes toward more negative values with the passing of time (i.e. in POST treatment sessions SMA contribute is higher, when concomitant FES is applied), meaning that the more the patients practice with FES treatment, the more this suppressive mechanism is enhanced, as they are more capable of to consider the disturbance of external conditions in their control loop, or from another point of view, they can take advantage of the "help" of FES. Carry-over estimated component values suggests that nCE patients have a decreased suppressive action of SMA over M1, while in CE patients this effect reflects the behavior of halthy subjects. Interestingly, if we observe the value of modulatory parameter, combining carry-over and time effects, we observe that CE-PRE and nCE-POST patients have similar values, close to zero and indeed with a low significance probability (i.e. $P_p < 0.80$), CE-POST patients have a negative value (suppression) in agreement with healthy subjects, and nCE-PRE patients have a highly positive value, indicating a clear faulty action of SMA

suppressive job.
Looking at connection from M1 toward SMA, we observed consistent negative values in both healthy a patients groups, either for intrinsic and extrinsic (modulatory) connections. Estimated values for the healthy group are -0.15 and -0.47 for intrinsic and extrinsic connections, respectively. For patients intrinsic connectivity parameter has a value of -0.21, while extrinsic connectivity parameter estimated values is -0.75. The interpretation on the role of this connection is unclear, due to the indirect nature of it: this is most likely via re-entrant feedback from M1 to basal-ganglia-thalamo-cortical circuits, which link the SMA and are known to be crucial for the initiation and control of voluntary action [14][132]. However, the estimated values ensure the presence of a negative feedback from M1 to SMA, in ageement with what Kasses and collegues reported in 2007 in a study on the suppressive influence of SMA over M1 during motor imagery [14].

C matrix discussion

C matrix overall parameters have all positive values for both patients and healthy groups. However, a significant difference between CE and nCE patients is observed, i.e., the estimated parameter relative to the carry-over regressor has a negative value: the proprioceptive input results to be lower in nCE patients, indicating that these subjects are less able to exploit and integrate the proprioceptive information in the motor control loop. This reduced sensitivity of nCE patients to proprioceptive information could be partly responsible for their inability to achieve stable improvements, due to the fundamental role of sensory-motor integration in the motor control mechanism. Another interesting result is about the Capacity-Score related component, which has been estimated to have a negative contribution over the voluntary input acting on SMA ($V \longrightarrow SMA$), i.e., the more a patient has residual capacity, the less is the effect of voluntary input on SMA. This result can be justified by the involvement of SMA in higher order processing of complex movements. This translates to the fact that the contribution of SMA in the motor planning gets lower, as the residual motor ability gets higher, as indicated in previous longitudinal and cross-sectional studies on activation changes during motor recovery [133][134].

AG connections

The interpretation of effective connectivity related to AG is not trivial, due to the versatility and complexity of the functions performed by AG and due to the absence of direct structural connections with other regions of the

model.
A major observation we can make, is about effective connections coming from SMA and S1. These connections, at least in healthy subjects, have not been discarded from the model, indicating that indeed AG has at some extent a role in calculating the discrepancy between intended action (proprioceptive prediciton from SMA) and movement consequences (real prorprioception from S1). Similarly, in patients, the estimated parameters for this connection, associated to the capacity-score regressor, have values that are consistent with those of healthy subjects, i.e. the more the patient has residual capability, the more these connections behave as those of a health subject.
Moreover, concerning the self-connection of AG, we observed a difference between CE and nCE patients in extrinsic connectivity; in fact, the estimated value for the second level parameter of the carry-over regressor is $-0.94 \mp 0.25_{ce}$, where $\mp 0.25_{ce}$ is the carry-over related component (multipliers are: $CE = 1$, $nCE = -1$). It is not clear how this result should be interpreted in terms of activation patterns, due to ambiguity of self-connections role. However, this group difference can be linked with what Gandolla and colleagues (2015) [10] reported about an increased activation of AG in nCE with respect to CE an healthy subjects, during the same FES treatment.

Self Connections

Relatively to other self-connections, as anticipated, their interpretation remains controversial, due to the requisite of constraining these parameters to be negative, in order to mathematically ensure the stability of the dynamical system. However, differences in self-connectivity have been observed between CE and nCE group of patients; in particular, these differences involved primary and supplementary motor areas, reflecting a similar activations of these regions of CE patients with healthy subjects, and abnormal ones in nCE patients, in accordance with observed activation patterns in Gandolla et al. (2016) [8]

Chapter 5

Conclusions and Future research directions

In conclusion, we presented a customised hierarchical approach for the identification of brain network structure (including M1, S1, SMA and AG) when investigating multiple matrices of the bilinear equation.
Following is reported a summary list of the hypotheses we started with and we wanted to verify:

1. Active Inference account: we expected modulation of $SMA/M1 \longrightarrow S1$ by input E (i.e. during FES) in healthys and CE;
2. Suppressive activity of SMA over M1: both in normal conditions and during FES we expected estimated negative values in intrinsic and extrinsic connectivity parameters;
3. AG computes action awareness: we expected non-zero intrinsic connections $SMA/M1 \longrightarrow AG$;
4. Similar activation in SMA, between healthy subjects and CE patients: we expected differences between CE and nCE in parameters relative to this region.

This study has highlighted a profound interaction between artificially altered sensory feedback and volitional movement in healthy subjects. This interaction, involving both motor areas and the somato-sensory area, is reflected by changes in coupling between these regions, supporting an active inference account of motor control, in which sensorimotor integration rests upon the context-sensitive assimilation of descending motor predictions. However, in patients this mechanism seems to be disturbed, indicating some sort of malfunction in proprioceptive prediction propagation toward S1, even in those

who presented stable improvements after the rehabilitation.
Secondly, the hierarchical nature of the motor processing in the brain has been exposed. Results have highlighted the role of SMA as higher order motor processing unit responsible for complex movements, and its superintendance role in suppressing standard motor plans as external conditions changes occur.
This latter aspect resulted to be an important feature, underlining both a role in the achievement of carry-over effect, and the changes this connection undergoes throughout the rehabilitation program. A further interesting and important characterization of nCE patients points toward their inhability to integrate sensorial proprioceptive information within the motor control loop, as highlighted by the results regarding the contribution of proprioceptive (P) experimental input over S1.
Even though the angular gyrus seemed to be a potentially interesting region to inspect, it did not reveal much about its role within the motor control loop; further studies should be made to have a better insight about this area. Still, a difference in patients with and without carry-over has been observed, in terms of self-connectivity, pointing to abnormal responses in AG as indicator that the FES carry-over effect is unlikely, as previously reported in literature.

In conclusion, we were able to confirm some of the initial hypotheses we had, while other have not been verified; on the other hand, we observed some interesting behaviours, not completely expected, that, however, are in agreement with the literature.

Some of the most important limitations of this study are related to the AG, in that, due to the wide range of functions this area is associated with, and the lack of direct connections with the other considered cortical regions (M1, S1, SMA), the model we built seems to be too simple, from the point of view of the number of regions considered, to obtain useful information about effective connectivity of the angular gyrus, and its role in motor control. Other limitations concern available fMRI data: brain volumes used in this work and acquired during previous studies was missing the cerebellum area, which is considered to give a major contribution in the generation of motor commands. Thus, in future developments, it would be appropriate to use novel images that includes this region, or, alternatively, to incorporate in the model deeper cortical regions, such as the cingulate gyrus.
In this direction, since increasing the number of nodes can considerably increase the complexity of the model and consequently it can make estimation and interpretation less precise, it would be certainly interesting to create

separated DCMs for subcircuits of motor control loop and combine inferred effective connectivity information thus found.

One consideration has to be made about repeated measures: the fact that for each patient two sessions have been acquired, at different times, it was not properly taken into consideration, if not inserting the time regressor in the design matrix. Another approach, could be that of using a third level in the hierarchy, i.e. creating a PEB model for each subject modelling the time course variations, around a subjec-specific mean, and then creating a further PEB of PEBs modelling effective group effects between subjects.

Another important limitation regards the number of subjects involved. All the procedures described in this work try to make inferences about the "population" of post-stroke patients (or healthy subjects), based on certain statistics calculated from a sample of subjects. Thus, it would be very important to have a data-set as wide as possible, so that the resulting conclusions were as robust as possible. In future studies, then, new subjects (both healthy and impaired) should be recruited for the collection of new data.

Finally, the same approaches and procedures employed in this work, for the investigation of motor control loop in healthy subjects and post-stroke patients, could be used to gain new insights in other types of pathologies.

Bibliography

[1] Morton Glanz, Sidney Klawansky, William Stason, Catherine Berkey, and Thomas C. Chalmers. Functional electrostimulation in poststroke rehabilitation: A meta-analysis of the randomized controlled trials. *Archives of Physical Medicine and Rehabilitation*, 77(6):549 – 553, 1996.

[2] WT Liberson, HJ Holmquest, D Scot, and M Dow. Functional electrotherapy: stimulation of the peroneal nerve synchronized with the swing phase of the gait of hemiplegic patients. *Archives of physical medicine and rehabilitation*, 42, 1961.

[3] D.N Rushton. Functional electrical stimulation and rehabilitation - an hypothesis. *Medical Engineering and Physics*, 25:1350–4533, 2003.

[4] Simona Denisia Iftime-Nielsen, Mark Schram Christensen, Rune Jersin Vingborg, Thomas Sinkjaer, Andreas Roepstorff, and Michael James Grey. Interaction of electrical stimulation and voluntary hand movement in sii and the cerebellum during simulated therapeutic functional electrical stimulation in healthy adults. *Human Brain Mapping*, 33(1):40–49, 2012.

[5] Mark Schram Christensen and Michael James Grey. Modulation of proprioceptive feedback during functional electrical stimulation: an fmri study. *European Journal of Neuroscience*, 37(11):1766–1778, 2013.

[6] Marta Gandolla, Simona Ferrante, Franco Molteni, Eleonora Guanziroli, Tiziano Frattini, Alberto Martegani, Giancarlo Ferrigno, Karl Friston, Alessandra Pedrocchi, and Nick S. Ward. Re-thinking the role of motor cortex: Context-sensitive motor outputs? *NeuroImage*, 91:366 – 374, 2014.

[7] Wulfram Gerstner. Hebbian learning and plasticity. 2011.

[8] Franco Molteni Eleonora Guanziroli Giancarlo Ferrigno Marta Gandolla, Nick S. Ward and Alessandra Pedrocchi. The neural correlates of long-term carryover following functional electrical stimulation for stroke. *Neural Plasticity*, 2016.

[9] Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11:127–138, 2010.

[10] Marta Gandolla, Franco Molteni, Nick S. Ward, Eleonora Guanziroli, Giancarlo Ferrigno, and Alessandra Pedrocchi. Validation of a quantitative single-subject based evaluation for rehabilitation-induced improvement assessment. *Annals of Biomedical Engineering*, 43(11):2686–2698, 2015.

[11] K.J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *NeuroImage*, 19(4):1273 – 1302, 2003.

[12] Parashkev Nachev, Henrietta Wydell, Kevin O'Neill, Masud Husain, and Christopher Kennard. The role of the pre-supplementary motor area in the control of action. *NeuroImage*, 36:T155 – T163, 2007. Cortical Control of Higher Motor Cognition.

[13] Ross Cunnington, John L. Bradshaw, and Robert Iansek. The role of the supplementary motor area in the control of voluntary movement. *Human Movement Science*, 15(5):627 – 647, 1996. Movement Control: A Southern Hemisphere Perspective.

[14] Christian H. Kasess, Christian Windischberger, Ross Cunnington, Rupert Lanzenberger, Lukas Pezawas, and Ewald Moser. The suppressive influence of sma on m1 in motor imagery revealed by fmri and dynamic causal modeling. *NeuroImage*, 40(2):828 – 837, 2008.

[15] Judith D Schaechter. Motor rehabilitation and brain plasticity after hemiparetic stroke. *Progress in Neurobiology*, 73(1):61 – 72, 2004.

[16] Mohamed L Seghier. The angular gyrus: multiple functions and multiple subdivisions. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, 19(1):43–61, 2013.

[17] Michael F. Bonner, Jonathan E. Peelle, Philip A. Cook, and Murray Grossman. Heteromodal conceptual processing in the angular gyrus. *NeuroImage*, 71:175–86, 2013.

[18] Mohamed L. Seghier, Elizabeth Fagan, and Cathy J. Price. Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. *Journal of Neuroscience*, 30(50):16809–16817, 2010.

[19] Karl J. Friston, Vladimir Litvak, Ashwini Oswal, Adeel Razi, Klaas E. Stephan, Bernadette C.M. van Wijk, Gabriel Ziegler, and Peter Zeidman. Bayesian model reduction and empirical bayes for group (dcm) studies. *NeuroImage*, 128:413 – 431, 2016.

[20] Karl Friston, Peter Zeidman, and Vladimir Litvak. Empirical bayes for dcm: A group inversion scheme. *Frontiers in Systems Neuroscience*, 2015.

[21] Adriaan R. E. Potgieser, Bauke M. de Jong, Michiel Wagemakers, Eelco W. Hoving, and Rob J. M. Groen. Insights from the supplementary motor area syndrome in balancing movement initiation and inhibition. In *Front. Hum. Neurosci.*, 2014.

[22] Christian Grefkes, Simon B. Eickhoff, Dennis A. Nowak, Manuel Dafotakis, and Gereon R. Fink. Dynamic intra- and interhemispheric interactions during unilateral and bilateral hand movements assessed with fmri and dcm. *NeuroImage*, 41(4):1382 – 1394, 2008.

[23] D. Mozaffarian et al. Heart disease and stroke statistics - 2015 update: a report from the american heart association. *Heart disease and stroke statistics*, (131):322–329, 2015.

[24] D. Wade. Measurement in neurological rehabilitation. *Oxford University Press*, 1992.

[25] C. Warlow, J. van Gijn, and M. Dennis et al. Stroke: practical management. *Oxford: Blackwell Publishing, 3rd edn.*, 2008.

[26] Lisa D. Alexander, Sandra E. Black, Fuqiang Gao, Gregory Szilagyi, Cynthia J. Danells, and William E. McIlroy. Correlating lesion size and location to deficits after ischemic stroke: the influence of accounting for altered peri-necrotic tissue and incidental silent infarcts. *Behavioral and Brain Functions*, 6(1):6, 2010.

[27] T. Naoyuki and I Shin-Ichi. Rehabilitation with post stroke motor recovery: A review with a focus on neural plasticity. *Stroke Res Treat.*, 2013.

[28] Dora Y.L. Chan, Chetwyn C.H. Chan, and Derrick K.S. Au. Motor relearning programme for stroke patients: a randomized controlled trial. *Clinical Rehabilitation*, 20(3):191–200, 2006.

[29] P.M. Rossini, C. Calautti, F. Pauri, and JC Baron. Post-stroke plastic reorganisation in the adult brain. *The Lancet Neurology*, 2(8):493–502, 2003.

[30] Neeraj Jain. Adult brain plasticity - what is revealed is exciting, what is hidden is critical. 27:439–42, 10 2002.

[31] I.J. Hubbard, M.W. Parsons, C. Neilson, and L.M Carey. Task-specific training: evidence for and translation to clinical practice. *Occupational Therapy International*, 16(3-4):175–189, 2009.

[32] TK Chan, CO Ho, WK Lee, YK Fung, YF Law, and CY Tsang. Functional outcome of the hand following flexor tendon repair at the 'no man's land'. *Journal of Orthopaedic Surgery*, 14(2):178–183, 2006.

[33] Mindy F. Levin, Jeffrey A. Kleim, and Steven L. Wolf. What do motor "recovery" and "compensation" mean in patients following stroke? *Neurorehabilitation and Neural Repair*, 23(4):313–319, 2009.

[34] A.M. Davis, Z. Agnidis, E. Badley, A. Kiss, J.P. Waddell, and A.E. Gross. Predictors of functional outcome two years following revision hip arthroplasty. *The Journal of Bone and Joint Surgery-American*, 88(4):685–691, 2006.

[35] B. B. Johansson. Current trends in stroke rehabilitation. a review with focus on brain plasticity. *Acta Neurologica Scandinavica*, 123(3):147–159, 2011.

[36] L. Brewer, F. Horgan, A. Hickey, and D. Williams. Stroke rehabilitation: recent advances and future therapies. *QJM: An International Journal of Medicine*, 106(1):11–25, 2013.

[37] Edward S.C., Chandramouli K., and Sandeep P.K. Emerging treatments for motor rehabilitation after stroke. *Neurohospitalist.*, 5(2):77–88, 2015.

[38] Won Hyuk Chang and Yun-Hee Kim. Robot-assisted therapy in stroke rehabilitation. *Journal of Stroke*, 15(3):174–181, 2013.

[39] V.M. Pomeroy, L. King, A. Pollock, A. Baily-Hallam, and P. Langhorne. Electrostimulation for promoting recovery of movement or functional ability after stroke. *The Cochrane database of systematic reviews*, (2):CD003241, 2006.

[40] Sukanta K. Sabut, Chanda Sikdar, Ramkrishna Mondal, Ratnesh Kumar, and Manjunatha Mahadevappa. Restoration of gait and motor recovery by functional electrical stimulation therapy in persons with stroke. *Disability and Rehabilitation*, 32(19):1594–1603, 2010.

[41] JH Burridge, PN Taylor, SA Hagan, DE Wood, and ID Swain. The effects of common peroneal stimulation on the effort and speed of walking: a randomized controlled trial with chronic hemiplegic patients. *Clinical Rehabilitation*, 11(3):201–210, 1997.

[42] Malcolm H. Granat, Douglas J. Maxwell, Amanda C.B. Ferguson, Kennedy R. Lees, and Joseph C. Barbenet. Peroneal stimulator: Evaluation for the correction of spastic drop foot in hemiplegia. *Archives of Physical Medicine and Rehabilitation*, 77(1):19 – 24, 1996.

[43] David F. Collins. Central contributions to contractions evoked by tetanic neuromuscular electrical stimulation. *Exercise and Sport Sciences Reviews*, 35(3):102–109, 2007.

[44] David Burke, Simon C. Gandevia, and Brian McKeon. The afferent volleys responsible for spinal proprioceptive reflexes in man. *The Journal of Physiology*, 339(1):535–552, 1983.

[45] A.Arturo Leis, Gert J. Grubwieser, John H. Schild, M.Stewart West, and Dobrivoje S. Stokic. Control of ia afferent input to triceps surae (soleus) locomotor nucleus precedes agonist muscle activation during gait. *Journal of Electromyography and Kinesiology*, 5(2):95 – 100, 1995.

[46] A. J. Bergquist, J. M. Clair, O. Lagerquist, C. S. Mang, Y. Okuma, and D. F. Collins. Neuromuscular electrical stimulation: implications of the electrically evoked sensory volley. *European Journal of Applied Physiology*, 111(10):2409, 2011.

[47] Roberto Merletti, A Andina, M Galante, and I Furlan. Clinical experience of electronic peroneal stimulators in 50 hemiparetic patients. 11:111–21, 1979.

[48] Emilia Ambrosini, Simona Ferrante, Alessandra Pedrocchi, Giancarlo Ferrigno, and Franco Molteni. Cycling induced by electrical stimulation improves motor recovery in postacute hemiparetic patients: a randomized controlled trial. *Stroke*, 42 4:1068–73, 2011.

[49] D. M. Wolpert and J. R. Flanagan. Motor prediction. *NeuroImage*, 11(18):R729–R732, 2001.

[50] T. J. Sejnowski and G. J. Tesauro. The hebb rule for synaptic plasticity: implementations and applications. *Neural Models of Plasticity*, pages 94–103, 1989.

[51] Noelle Leonard, Amishi Jha, Bethany Casarjian, Merissa Goolsarran, Cristina Garcia, Charles Cleland, Marya Gwadz, and Zohar Massey. Mindfulness training improves attentional task performance in incarcerated youth: a group randomized controlled intervention trial. *Frontiers in Psychology*, 4:792, 2013.

[52] Karin Rosenkranz and John C. Rothwell. Modulation of proprioceptive integration in the motor cortex shapes human motor learning. 32(26):9000–9006, 2012.

[53] C. E. Chapman and S. A. Ageranioti-Bélanger. Discharge properties of neurones in the hand area of primary somatosensory cortex in monkeys in relation to the performance of an active tactile discrimination task. *Experimental Brain Research*, 87(2):319–339, 1991.

[54] Tim Genewein and Daniel Braun. A sensorimotor paradigm for bayesian model selection. *Frontiers in Human Neuroscience*, 6:291, 2012.

[55] Daniel M. Wolpert and Zoubin Ghahramani. Computational principles of movement neuroscience. 3 Suppl:1212–7, 12 2000.

[56] D.M. Wolpert and M. Kawato. Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7):1317 – 1329, 1998.

[57] Konrad Kording and Daniel Wolpert. Bayesian integration in sensorimotor learning. 427:244–7, 02 2004.

[58] Karl J. Friston, Jean Daunizeau, and Stefan J. Kiebel. Reinforcement learning or active inference? *PLOS ONE*, 4(7):1–13, 2009.

[59] Rick A. Adams, Stewart Shipp, and Karl J. Friston. Predictions not commands: active inference in the motor system. 2012.

[60] J. Daunizeau, O. David, and K.E. Stephan. Dynamic causal modelling: A critical review of the biophysical and statistical foundations. *NeuroImage*, 58(2):312 – 322, 2011.

[61] A. Razi and K. J. Friston. The connected brain: Causality, models, and intrinsic dynamics. *IEEE Signal Processing Magazine*, 33(3):14–35, 2016.

[62] Olivier David, Stefan J. Kiebel, Lee M. Harrison, Jérémie Mattout, James M. Kilner, and Karl J. Friston. Dynamic causal modeling of evoked responses in EEG and MEG. *NeuroImage*, 30(4):1255–72, 2006.

[63] Marie-Helene Boudrias, Carla Sa Goncalves, William Penny, Chang-Hyun Park, Holly Rossiter, Penelope Talelli, and Nick Ward. Age-related changes in causal interactions between cortical motor regions during hand grip. 59:3398–405, 11 2011.

[64] Hae-Jeong Park, Chongwon Pae, Karl Friston, Changwon Jang, Adeel Razi, Peter Zeidman, Won Seok Chang, and Jin Woo Chang. Hierarchical dynamic causal modeling of resting-state fmri reveals longitudinal changes in effective connectivity in the motor system after thalamotomy for essential tremor. *Frontiers in Neurology*, 8:346, 2017.

[65] Christian Grefkes, Dennis A. Nowak, Ling E. Wang, Manuel Dafotakis, Simon B. Eickhoff, and Gereon R. Fink. Modulating cortical connectivity in stroke patients by rtms assessed with fmri and dynamic causal modeling. *NeuroImage*, 50 1:233–42, 2010.

[66] K.E. Stephan, W.D. Penny, R.J. Moran, H.E.M. den Ouden, J. Daunizeau, and K.J. Friston. Ten simple rules for dynamic causal modeling. *NeuroImage*, 49(4):3099 – 3109, 2010.

[67] Klaas Enno Stephan and Alard Roebroeck. A short history of causal modeling of fmri data. *NeuroImage*, 62(2):856 – 863, 2012.

[68] Klaas Enno Stephan, Lars Kasper, Lee M. Harrison, Jean Daunizeau, Hanneke E.M. den Ouden, Michael Breakspear, and Karl J. Friston. Nonlinear dynamic causal models for fmri. *NeuroImage*, 42(2):649 – 662, 2008.

[69] A.C. Marreiros, S.J. Kiebel, and K.J. Friston. Dynamic causal modelling for fmri: A two-state model. *NeuroImage*, 39(1):269 – 278, 2008.

[70] J. Daunizeau, K.J. Friston, and S.J. Kiebel. Variational bayesian identification and prediction of stochastic nonlinear dynamic causal models. *Physica D: Nonlinear Phenomena*, 238(21):2089 – 2118, 2009.

[71] R.J. Moran, K.E. Stephan, T. Seidenbecher, H.-C. Pape, R.J. Dolan, and K.J. Friston. Dynamic causal models of steady-state responses. *NeuroImage*, 44(3):796 – 811, 2009.

[72] K.J. Friston, A. Mechelli, R. Turner, and C.J. Price. Nonlinear responses in fmri: The balloon model, volterra kernels, and other hemodynamics. *NeuroImage*, 12(4):466 – 477, 2000.

[73] Richard B. Buxton, Eric C. Wong, and Lawrence R. Frank. Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magnetic Resonance in Medicine*, 39(6):855–864, 1998.

[74] RL Grubb, ME Raichle, JO Eichling, and MM Ter-Pogossian. The effects of changes in paco2 on cerebral blood volume, blood flow, and vascular mean transit time. *Stroke*, 5(5):630–639, 1974.

[75] Klaas Enno Stephan, Nikolaus Weiskopf, Peter M. Drysdale, Peter A. Robinson, and Karl J. Friston. Comparing hemodynamic models with dcm. *NeuroImage*, 38(3):387 – 401, 2007.

[76] Stefan J. Kiebel, Stefan Kloppel, Nikolaus Weiskopf, and Karl J. Friston. Dynamic causal modeling: A generative model of slice timing in fmri. *NeuroImage*, 34(4):1487 – 1496, 2007.

[77] K.J. Friston and W. Penny. Empirical bayes and hierarchical model. *Statistical parametric Mapping: the analysis of functional brain images*, 2007.

[78] K.J. Friston and W. Penny. Bayesian inversion of dynamic models. *Statistical parametric Mapping: the analysis of functional brain images*, 2007.

[79] Klaas Enno Stephan, Marc Tittgemeyer, Thomas R. Knosche, Rosalyn J. Moran, and Karl J. Friston. Tractography-based priors for dynamic causal models. *NeuroImage*, 47(4):1628 – 1638, 2009.

[80] Karl Friston and Will Penny. Post hoc bayesian model selection. *NeuroImage*, 56(4):2089 – 2099, 2011.

[81] W.D. Penny. Comparing dynamic causal models using aic, bic and free energy. *NeuroImage*, 59(1):319 – 330, 2012.

[82] Will D. Penny, Klaas E. Stephan, Jean Daunizeau, Maria J. Rosa, Karl J. Friston, Thomas M. Schofield, and Alex P. Leff. Comparing families of dynamic causal models. *PLOS Computational Biology*, 6(3):1–14, 03 2010.

[83] W.D. Penny, K.E. Stephan, A. Mechelli, and K.J. Friston. Comparing dynamic causal models. *NeuroImage*, 22(3):1157 – 1172, 2004.

[84] Klaas Enno Stephan, Will D. Penny, Jean Daunizeau, Rosalyn J. Moran, and Karl J. Friston. Bayesian model selection for group studies. *NeuroImage*, 46(4):1004 – 1017, 2009.

[85] Mohamed L. Seghier, Peter Zeidman, Nicholas H. Neufeld, Alex P. Leff, and Cathy J. Price. Identifying abnormal connectivity in patients using dynamic causal modeling of fmri responses. In *Front. Syst. Neurosci.*, 2010.

[86] Mark A. Pitt and In Jae Myung. When a good fit can be bad. *Trends in Cognitive Sciences*, 6(10):421 – 425, 2002.

[87] Karl Friston, Jérémie Mattout, Nelson Trujillo-Barreto, John Ashburner, and Will Penny. Variational free energy and the laplace approximation. *NeuroImage*, 34(1):220 – 234, 2007.

[88] Robert E. Kass and Adrian E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

[89] J.D. Ramsey, S.J. Hanson, C. Hanson, Y.O. Halchenko, R.A. Poldrack, and C. Glymour. Six problems for causal inference from fmri. *NeuroImage*, 49(2):1545 – 1558, 2010.

[90] Gabriele Lohmann, Kerstin Erfurth, Karsten Muller, and Robert Turner. Critical comments on dynamic causal modelling. *NeuroImage*, 59(3):2322 – 2329, 2012.

[91] Alexander P. Leff, Thomas M. Schofield, Klass E. Stephan, Jennifer T. Crinion, Karl J. Friston, and Cathy J. Price. The cortical dynamics of intelligible speech. *Journal of Neuroscience*, 28(49):13209–13215, 2008.

[92] Sukhbinder Kumar, Klaas E Stephan, Jason D Warren, Karl J Friston, and Timothy D Griffiths. Hierarchical processing of auditory objects in humans. *PLOS Computational Biology*, 3(6):1–9, 2007.

[93] Nia Goulden, Shane McKie, Emma J. Thomas, Darragh Downey, Gabriella Juhasz, Stephen R. Williams, James B. Rowe, J.F. William Deakin, Ian M. Anderson, and Rebecca Elliott. Reversed frontotemporal connectivity during emotional face processing in remitted depression. *Biological Psychiatry*, 72(7):604 – 611, 2012.

[94] R.G.M. Schlosser, G. Wagner, K. Koch, R. Dahnke, J.R. Reichenbach, and H. Sauer. Fronto-cingulate effective connectivity in major depression: A study with fmri and dynamic causal modeling. *NeuroImage*, 43(3):645 – 655, 2008.

[95] M. Pyka, D. Heider, S. Hauke, T. Kircher, and A. Jansen. Dynamic causal modeling with genetic algorithms. *Journal of Neuroscience Methods*, 194(2):402 – 406, 2011.

[96] Jorge Renner Cardoso de Almeida, Amelia Versace, Andrea Mechelli, Stefanie Hassel, Karina Quevedo, David Jerome Kupfer, and Mary Louise Phillips. Abnormal amygdala-prefrontal effective connectivity to happy faces differentiates bipolar from major depression. *Biological Psychiatry*, 66(5):451 – 459, 2009. Medical Consequences and Contributions to Depression.

[97] Nicolas A. Crossley, Andrea Mechelli, Paolo Fusar-Poli, Matthew R. Broome, Pall Matthiasson, Louise C. Johns, Elvira Bramon, Lucia Valmaggia, Steven C.R. Williams, and Philip K. McGuire. Superior temporal lobe dysfunction and frontotemporal dysconnectivity in subjects at risk of psychosis and in first-episode psychosis. *Human Brain Mapping*, 30(12):4129–4137.

[98] Andrea Mechelli, Paul Allen, Edson Amaro, Cynthia H.Y. Fu, Steven C.R. Williams, Michael J. Brammer, Louise C. Johns, and Philip K. McGuire. Misattribution of speech and impaired connectivity in patients with auditory verbal hallucinations. *Human Brain Mapping*, 28(11):1213–1222.

[99] Karl J. Friston, Baojuan Li, Jean Daunizeau, and Klaas E. Stephan. Network discovery with dcm. *NeuroImage*, 56(3):1202 – 1221, 2011.

[100] M.J. Rosa, K. Friston, and W. Penny. Post-hoc selection of dynamic causal models. *Journal of Neuroscience Methods*, 208(1):66 – 78, 2012.

[101] Mohamed L. Seghier and Karl J. Friston. Network discovery with large dcms. *NeuroImage*, 68:181 – 191, 2013.

[102] Christian Herbert Kasess, Klaas Enno Stephan, Andreas Weissenbacher, Lukas Pezawas, Ewald Moser, and Christian Windischberger. Multi-subject analyses with dynamic causal modeling. *NeuroImage*, 49(4):3065 – 3074, 2010.

[103] Jane Neumann and Gabriele Lohmann. Bayesian second-level analysis of functional magnetic resonance images. *NeuroImage*, 20(2):1346 – 1355, 2003.

[104] K.J. Friston, A.P. Holmes, C.J. Price, C. Buchel, and K.J. Worsley. Multisubject fmri studies and conjunction analyses. *NeuroImage*, 10(4):385 – 396, 1999.

[105] Mark W. Woolrich, Timothy E.J. Behrens, Christian F. Beckmann, Mark Jenkinson, and Stephen M. Smith. Multilevel linear modelling for fmri group analysis using bayesian inference. *NeuroImage*, 21(4):1732 – 1747, 2004.

[106] Mark W. Woolrich, Saad Jbabdi, Brian Patenaude, Michael Chappell, Salima Makni, Timothy Behrens, Christian Beckmann, Mark Jenkinson, and Stephen M. Smith. Bayesian analysis of neuroimaging data in fsl. *NeuroImage*, 45(1, Supplement 1):S173 – S186, 2009. Mathematics in Brain Imaging.

[107] Medical Research Council/Guarantors of Brain. Aids to the examination of the peripheral nervous system. London, UK, 1986.

[108] Claudia Casellato, Simona Ferrante, Marta Gandolla, Nicola Volonterio, Giancarlo Ferrigno, Giuseppe Baselli, Tiziano Frattini, Alberto Martegani, Franco Molteni, and Alessandra Pedrocchi. Simultaneous measurements of kinematics and fmri: compatibility assessment and case report on recovery evaluation of one stroke patient. *Journal of NeuroEngineering and Rehabilitation*, 7(1):49, 2010.

[109] Marta Gandolla, Simona Ferrante, Claudia Casellato, Giancarlo Ferrigno, Franco Molteni, Alberto Martegani, Tiziano Frattini, and Alessandra Pedrocchi. fmri brain mapping during motion capture and fes induced motor tasks: Signal to noise ratio assessment. *Medical Engineering & Physics*, 33(8):1027–1032, 2011.

[110] Jesper L.R. Andersson, Chloe Hutton, John Ashburner, Robert Turner, and Karl Friston. Modeling geometric deformations in epi time series. *NeuroImage*, 13(5):903 – 919, 2001.

[111] Tom Johnstone, Kathleen S Ores Walsh, Larry L. Greischar, Andrew L. Alexander, Andrew S. Fox, Richard J. Davidson, and Terrence R. Oakes. Motion correction and the use of motion covariates in multiple-subject fmri analysis. *Human brain mapping*, 27 10:779–88, 2006.

[112] J. Talairach and P. Tournoux. Co-planar stereotaxic atlas of the human brain. 1988.

[113] Karl. J. Friston, J. Ashburner, C. D. Frith, J.-B. Poline, J. D. Heather, and R. S. J. Frackowiak. Spatial registration and normalization of images. *Human Brain Mapping*, 3(3):165–189.

[114] Martha Flanders. What is the biological basis of sensorimotor integration? *Biological Cybernetics*, 104(1):1–8, 2011.

[115] Reza Shadmehr, Maurice A. Smith, and John Krakauer. Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience*, 33:89–108, 2010.

[116] Daniel M. Wolpert, Zoubin J. C. Ghahramani, and Michael I. Jordan. An internal model for sensorimotor integration. *Science*, 269 5232:1880–2, 1995.

[117] Chlöé Farrer, Scott H. Frey, John Darrell Van Horn, Eugene Tunik, David J. Turk, Sara K. Inati, and Scott T. Grafton. The angular gyrus computes action awareness representations. *Cerebral cortex*, 18 2:254–61, 2008.

[118] Amanda K Kinnischtzke. Cell type specific connections from primary motor to primary somatosensory cortex. July 2013.

[119] Maxime Guye, Geoffrey J.M Parker, Mark Symms, Philip Boulby, Claudia A.M Wheeler-Kingshott, Afraim Salek-Haddadi, Gareth J Barker, and John S Duncan. Combined functional mri and tractography to demonstrate the connectivity of the human primary motor cortex in vivo. *NeuroImage*, 19(4):1349 – 1360, 2003.

[120] Francesco Vergani, Luis Lacerda, Juan Martino, Johannes Attems, Christopher Morris, Patrick Mitchell, Michel Thiebaut de Schotten, and Flavio Dell'Acqua. White matter connections of the supplementary motor area in humans. *Journal of Neurology, Neurosurgery & Psychiatry*, 85(12):1377–1385, 2014.

[121] Franz-Xaver Neubert, Rogier B. Mars, Ethan R. Buch, Etienne Olivier, and Matthew F. S. Rushworth. Cortical and subcortical interactions during action reprogramming and their related white matter pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 107 30:13240–5, 2010.

[122] Nathalie Picard and Peter L. Strick. Activation of the supplementary motor area (sma) during performance of visually guided movements. *Cerebral Cortex*, 13(9):977–986, 2003.

[123] Ray M. Chu and Keith L. Black. Chapter 8 - current surgical management of high-grade gliomas. In *Schmidek and Sweet Operative Neurosurgical Techniques (Sixth Edition)*, pages 105 – 110. W.B. Saunders, sixth edition edition, 2012.

[124] JM Orgogozo and B Larsen. Activation of the supplementary motor area during voluntary movement in man suggests it works as a supramotor area. *Science*, 206(4420):847–850, 1979.

[125] Ingo Hertrich, Susanne Dietrich, and Hermann Ackermann. The role of the supplementary motor area for speech and language processing. *Neuroscience & Biobehavioral Reviews*, 68:602–610, 2016.

[126] Claire Wardak. The role of the supplementary motor area in inhibitory control in monkeys and humans. *Journal of Neuroscience*, 31(14):5181–5183, 2011.

[127] J. Yang and H. Shu. The causal interactions between bilateral m1 and sma during verb comprehension, motor imagery and hand motion. *Archives of Neuroscience*, 1(3), 2014.

[128] Kazuhiko Seki and Eberhard E. Fetz. Gating of sensory input at spinal and cortical levels during preparation and execution of voluntary movement. *Journal of Neuroscience*, 32(3):890–902, 2012.

[129] Patrick Haggard and Ben Whitford. Supplementary motor area provides an efferent signal for sensory suppression. *Cognitive Brain Research*, 19(1):52 – 58, 2004.

[130] Manuel A. Castro-Alamancos and Elizabeth Oldford. Cortical sensory suppression during arousal is due to the activity-dependent depression of thalamocortical synapses. *The Journal of Physiology*, 541(1):319–331.

[131] Georgiana Juravle, Gordon Binsted, and Charles Spence. Tactile suppression in goal-directed movement. *Psychonomic Bulletin & Review*, 24(4):1060–1076, 2017.

[132] G E Alexander, M R DeLong, and P L Strick. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, 9(1):357–381, 1986.

[133] N. S. Ward, M. M. Brown, A. J. Thompson, and R. S. J. Frackowiak. Neural correlates of motor recovery after stroke: a longitudinal fmri study. *Brain*, 126(11):2476–2496, 2003.

[134] N. S. Ward, M. M. Brown, A. J. Thompson, and R. S. J. Frackowiak. Neural correlates of motor recovery after stroke: a cross-sectional study. *Brain*, 126(6):1430–1498, 2003.

[135] F. Bloch, W. W. Hansen, and M. Packard. The nuclear induction experiment. *Phys. Rev.*, 70:474–485, 1946.

[136] F. Bloch, W. W. Hansen, and Martin Packard. Nuclear induction. *Phys. Rev.*, 69:127–127, 1946.

[137] E. M. Purcell, H. C. Torrey, and R. V. Pound. Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.*, 69:37–38, 1946.

[138] P. C. Lauterbur. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242:190, 1973.

[139] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4):189–210, 1994.

[140] K.J. Worsley, Alan Evans, Sean Marrett, and P Neelin. A three-dimensional statistical analysis for cbf activation studies in human brain. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, 12:900–18, 12 1992.

[141] K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4(1):58–73, 1996.

[142] David Parker, Xueqing Liu, and Qolamreza R. Razlighi. Optimal slice timing correction and its interaction with fmri parameters and artifacts. *Medical Image Analysis*, 35:434 – 445, 2017.

[143] Karl J. Friston, Steven Williams, Robert Howard, Richard S. J. Frackowiak, and Robert Turner. Movement-related effects in fmri time-series. *Magnetic Resonance in Medicine*, 35(3):346–355, 1996.

[144] In Richard S.J. Frackowiak, Karl J. Friston, Christopher D. Frith, Raymond J. Dolan, Cathy J. Price, Semir Zeki, John T. Ashburner, and William D. Penny, editors, *Human Brain Function (Second Edition)*. Academic Press, Burlington, second edition edition, 2004.

[145] John Ashburner Stefan Kiebe Thomas Nichols William Penny, Karl Friston. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, 2006.

[146] Karl J. Friston, Christopher D. Frith, Peter F. Liddle, and Richard S. Frackowiak. Comparing functional (pet) images: the assessment of significant change. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, 11 4:690–9, 1991.

[147] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4):189–210, 1995.

[148] Ludwig Fahrmeir and Gerhard Tutz. Multivariate statistical modelling based on generalized linear models / ludwig fahrmeir, gerhard tutz. *SERBIULA (sistema Librum 2.0)*, 01 1994.

[149] Hinton G.E. Neal R.M. A view of the em algorithm that justifies incremental, sparse, and other variants. 1998.

[150] K.J. Friston, D.E. Glaser, R.N.A. Henson, S. Kiebel, C. Phillips, and J. Ashburner. Classical and bayesian inference in neuroimaging: Applications. *NeuroImage*, 16(2):484 – 512, 2002.

Appendix A

MRI and fMRI principles

A.1 MRI principles

Nuclear magnetic resonance (NMR) is a non-invasive means of obtaining clinical images and of studying tissue metabolism in vivo. Bloch and Purcell independently discovered NMR in 1946 [135] [136] [137]. Six years later they were awarded the Nobel Prize for their achievements. Since then, the development of NMR spectrometers and NMR scanners has led to the opening up of whole new branches of physics, chemistry, biology and medicine, allowing to observe in-vivo human tissues up to that time unknown. The process of acquiring two and 3D images by NMR, known as magnetic resonance imaging (MRI), was first illustrated by Lauterbur in 1973 [138] who produced a 2D MR image of a phantom. Over the last 30 years, Fourier transform imaging techniques have tremendously accelerated the development of MRI.

The basic underlying physics of MRI is the same as that of NMR. At the atomic level, nuclei exhibit a property known as nuclear spin, an intrinsic angular momentum that can be visualized as a rotating motion of the nucleus about its own axis. In quantum mechanics spin is represented by a magnetic spin quantum number. Certain isotopes of elements have an odd number of protons and/or neutrons such as ^{1}H, ^{13}C, and ^{15}N, and have an overall positive charge. As a result of having both electrical charge and nuclear spin, the nuclei will possess an intrinsic magnetic moment μ in the direction of the spin axis, and the relation can be represented as follows

$$|\vec{\mu}| = \gamma|\vec{I}| \tag{A.1}$$

where I is the angular momentum and γ is the gyromagnetic ratio. The strength of the magnetic moment is a property of the type of nucleus. Hydrogen nuclei ^{1}H, as well as possessing the strongest magnetic moment, are present in water molecules and consequently are in high abundance in biological material. For this reason hydrogen imaging is the most widely used MRI procedure.

Consider a collection of ^{1}H nuclei (spinning protons). In the absence of an externally applied magnetic field, the magnetic moments have random orientations and, therefore, the net magnetization is zero. However, if an externally supplied magnetic field B_0 is imposed, the magnetic moments have a tendency to align with the external field.

The magnetic moments or spins are constrained to adopt one of two orientations with respect to B_0, denoted *parallel* and *anti-parallel*. Nuclei with higher magnetic spin quantum number than ^{1}H will adopt more than two orientations. The spin axes are not exactly aligned with B_0, they precess around B_0 with a characteristic frequency. The rate of precession ω_0 is called the Larmor frequency and it is directly proportional to the strength of the magnetic field as expressed below

$$\omega_0 = \frac{\gamma B_0}{2\pi} \tag{A.2}$$

where γ is the gyromagnetic ratio.
For a hydrogen nucleus, $\gamma_H = 4257\ Hz/Gauss$. Thus at 1.5 Tesla (15,000 Gauss), $F = \frac{\omega_0}{2\pi} = 63.855\ MHz$.

A.1.1 The net Magnetisation Vector

Considering a collection of ^{1}H nuclei, when an external magnetic field is applied, a certain amount of protons will align in the parallel state (high energy) and others in the anti-parallel state (low energy) with respect to the direction of the magnetic field. The reason why some spins adopt the higher energy anti-parallel state is that spins may move from one state to the other if an exact amount of energy, $\Delta E = E_1 - E_2$ is supplied to the system. This amount of energy could be provided in the form of thermal energy. In fact, if the temperature of the system were absolute zero, all spins would adopt the parallel orientation. Thermal energy thus will cause few spins to assume the anti-parallel configuration. Let $P1$ and $P2$ be the number of spins adopting the parallel and anti-parallel states respectively; at room temperature in a 1.5 Tesla magnetic field there will be a population ratio $P2 : P1$ equal to $100,000 : 100,006$.

The magnetic moments of a collection of protons can be represented as vectors. Each one of these vectors can be described by its components per-

pendicular (M_{xy}) and parallel (M_z) to B_0(supposing the direction of B_0 is aligned with the z-axis of $\mathbb{R}^3$).
The net magnetization vector can be expressed as $\vec{M} = \hat{\imath}\, M_z + \hat{\jmath}\, M_x + \hat{k}\, M_y$. For a large enough number of spins, each of which can assume one out of two possible energy states which have anti-parallel orientation, individual components perpendicular to B_0 cancel, leaving only components in the direction parallel to the magnetic field. As most spins adopt the parallel rather than the anti-parallel state, the net magnetization $\vec{M}$ is in the direction of the B_0 field.

A.1.2 Effects of Radio frequency Pulses

In order to obtain the magnetic resonance signal, radio frequency (RF) energy must be applied. Magnetic resonance will occur only when an electromagnetic field (in the radio frequency region of the electromagnetic spectrum) is applied with a frequency equal to the Larmor precession rate, in order to match the energy difference ΔE between the nuclear spin-levels under a constant field B_0. The energy difference between the two states can be expressed as

$$\begin{aligned} \Delta E &= E_{anti-parallel} - E_{parallel} \\ \Longrightarrow \Delta E &= \frac{\gamma \hbar B_0}{4\pi} - \frac{-\gamma \hbar B_0}{4\pi} = \frac{\gamma \hbar B_0}{2\pi} = \hbar \omega_0 \end{aligned} \tag{A.3}$$

where $\hbar$ is the Planck's constant. RF pulses at the Larmor frequency will produce resonance and an efficient transfer of energy will occur, causing nuclear spins to swap between parallel and anti-parallel states. The energy transmitted by the RF pulse can be represented as follows:

$$\begin{aligned} \hbar f &= \Delta E = \frac{\gamma \hbar B_0}{2\pi} \\ \Longrightarrow f &= \frac{\gamma B_0}{2\pi} \quad \text{or} \quad \omega = \gamma B_0 \end{aligned} \tag{A.4}$$

Like all other electromagnetic radiations, RF energy has electric and magnetic field components. If the magnetic field component of RF, represented by B_1, lies in the xy plane, it will produce a torque when acting in cohesion with the net magnetization M. The $x - y$ components of M will be made coherent by the B_1 field giving a net $x - y$ component to M and hence effectively causing M to tilt from the z direction into the xy plane.
After the RF pulse is applied, the magnetisation vector $\vec{M}$ is flipped by an angle α, which is proportional to both the intensity of the RF pulse and the duration τ_{B_1}. If the pulse is long and strong enough to rotate the net

magnetisation $\vec{M}$ onto the transverse xy plane, then is termed as a 90° RF pulse. Similarly, a 180° RF pulse would rotate net magnetisation towards the -z direction. The flip angle α can be expressed as $\alpha = \gamma B_1 \tau_{B_1}$. Once the magnetisation vector is flipped by 90° and begins to precess about the B_0 axis, the NMR signal can be detected thanks to a receiver coil, within which an electromotive force is induced by the precessing magnetisation vector, according to Faraday's law of magnetic induction. From this received signal an MR image can be reconstructed. The waveform the voltage or signal induced in the receiver coil is termed free induction delay (FID). The magnitude of the generated signal depends on the number of nuclei contributing to produce the transverse magnetisation and on the relaxation times.

A.1.3 MR relaxation processes

Following termination of an RF pulse, the return of M to its equilibrium state (the direction of the $z - axis$) is known as relaxation. In this phase nuclei will dissipate their excess energy as heat to the surrounding environment (or lattice) and revert to their equilibrium position. There are three factors that influence the decay of M: magnetic field inhomogeneity, longitudinal T1 relaxation and transverse T2 relaxation.

T1 relaxation (also known as *spin-lattice relaxation*) is the realignment of spins (and so of M) with the external magnetic field B_0 (z-axis), that leads to a gradual increase in the longitudinal magnetization, and is defined as the time the longitudinal magnetization takes to reach 63% of its final value, following a 90° RF pulse. The process of equilibrium restoration follows a growing exponential trend, described by the equation

$$M_z = M_0(1 - e^{-\frac{t}{T1}}) \tag{A.5}$$

T2 relaxation (also known as transverse relaxation or *spin-spin relaxation*) is the decrease of the x-y component of magnetisation. Under the influence of the RF pulse, protons begin to precess together or "in phase", leading to the formation of the transverse magnetization; immediately after the 90° RF pulse, these protons will rotate about their z-axis and start to "dephase". This occurs as a result of protons processing at slightly different frequencies due to spin-spin interactions. Due to dephasing, the transverse magnetization keeps decreasing until the protons are completely dephased, at which the transverse signal becomes zero. The T2 relaxation constant is defined as the time taken by the transverse magnetization to decay to 37% of its original value. Different tissues have different values of T2 which gives rise to T2 contrast. The transverse magnetization M_{xy} is given by the following equation

$$M_{xy} = M_0(e^{-\frac{t}{T2}}) \tag{A.6}$$

Regarding magnetic field inhomogeneities, it is practically impossible to construct an NMR magnet with perfectly uniform magnetic field strength B_0 and, despite the presence of additional hardware to assist the normalization of the field, it is inevitable that an NMR sample will experience different B_0's across its body so that nuclei comprising the sample will have different processional frequencies (according to the Larmor equation). Due to this difference in precessing frequency related to magnetic field inhomogeneities, following an RF pulse, phase differences between nuclei will increase with time and the vector addition of these phases will reduce M_{xy} as time goes on. Basically this inohomegeity of the B_0 field causes the dephasing effect to be accelerated. Hence, typically we do not measure a pure T2 relaxation time but rather a faster ralaxation called T2*, that characterises dephasing due to both B_0 inhomogeneity and transverse relaxation.

In order to obtain a signal with a T2 dependence rather than a T2* dependence, a pulse sequence known as the spin-echo sequence has been devised, which reduces the effect of B_0 inhomogeneity on (M_{xy}). A pulse sequence is an appropriate combination of one or more RF pulses and gradients with intervening periods of recovery. A pulse sequence consists of several components, of which the main ones are the repetition time (TR), the echo time (TE), flip angle, the number of excitations (NEX), bandwidth and acquisition matrix. Following is a brief description of how the spin echo pulse sequence works.

After a 90° pulse, a MR signal is formed which decays with T2* characteristics. At time TE/2, a 180° pulse is applied to the sample which causes the spins to invert. After inversion, the order of the spins is reversed with the fastest lagging behind the others. At time TE, the spins become coherent again so that a signal (known as the spin echo) is produced. If a further 180° pulse is applied at time TE/2 after the peak signal of the first spin echo, then a second spin echo is reduced from its previous peak amplitude due to T2 dephasing which cannot be rephased by the 180° pulses. A line drawn through the peak amplitude of a large number of spin echoes describes the T2 decay, while individual spin echoes exhibit T2* decay.

Signal strength decays with time to varying degrees depending on the different materials in the sample. Different organs have different T1s and T2s and hence different rates of decay of signal. When imaging anatomy, some degree of control of the contrast of different organs or parts of organs is possible by varying TR and TE. The intensity of a spin echo signal, I, can be approximated as

$$I = N(H) \cdot f(V) \cdot (1 - e^{-TR/T1}) \cdot e^{-TE/T2} \tag{A.7}$$

where $N(H)$ is the proton density and $f(V)$ is a function of flow.

A.1.4 MR Imaging

The NMR signal does not contain any spatial information. In order to determine the actual location within the sample from which the RF signal was emitted, magnetic field *gradients* are superimposed on the magnet generating the otherwise (almost) homogeneous external magnetic field B_0. Hence, gradient coils are used in MRI machines to vary the magnetic field in x, y and z directions. Depending on their function, they are called the *slice-select* gradient, the readout or *frequency encoding* gradient, and the *phase encoding* gradients. Based on their directions, they can be termed as G_x, G_y, and G_z respectively. According to equation (A.2), the magnetic field gradient causes identical nuclei to precess at different Larmor frequencies.

Slice selection

Let consider the process of slice selection. When the gradient magnetic field (G_z) is present, applying a single fraquency RF pulse to the whole sample, only a narrow plane perpendicular to the longitudinal axis will match the resonant frequency and consequently will absorb the RF energy. Everywhere else in the sample is receiving the wrong frequency of excitation for resonance to occur. This technique allows a slice, with thickness determined by the magnetic field gradient strength, to be selected from a sample. Let ω_s be the frequency of the applied RF pulse applied over an axial slice (which perpendicular to the long axis of the body) with a bandwidth of $\Delta\omega_s$. Then the nuclei which get excited will have precession frequencies between $\omega_s + \Delta\omega_s$ and $\omega_s - \Delta\omega_s$. The thickness of the axial slice selected, is given by the following equation where T is the slice thickness and G_{slice} is one of the magnetic field gradients.

$$T = \frac{2\Delta\omega_s}{\gamma G_{slice}} \tag{A.8}$$

Frequency encoding

The slice selection is followed by processes known as frequency encoding and phase encoding. Three magnetic field gradients, placed orthogonally to one another inside the bore of the magnet, are required to encode information in three dimensions. With a slice selected and excited as described above, current is switched to one of the two remaining gradient coils (referred to as the frequency encoding gradient) This has the effect of spatially encoding the excited slice along one axis, so that columns of spins perpendicular to the axis precess at slightly different Larmor frequencies. For a homogeneous sample, the intensity of the signal at each frequency is proportional to the number of protons in the corresponding column. The frequency encoding gradient (G_x) is turned on just before the receiver is gated on and is left on while the signal is sampled or read out. For this reason the frequency encoding gradient is also known as the readout gradient. The resulting FID is a graph of signal (formed from the interference pattern of the different frequencies) induced in the receiver verses time. If the FID is subjected to Fourier transform, a conventional spectrum in which signal amplitude is plotted as a function of frequency can be obtained. Thus, a graph of signal versus frequency is obtained which corresponds to a series of lines or views representing columns of spins in the slice.

Phase encoding

Similarly, by the application of phase encoding gradient (G_y), we are able to allocate different phase angle of spins to different spatial locations. A phase encoding gradient is applied orthogonally to the other two gradients after slice selection and excitation, but before frequency encoding. It serves as a phase memory, remembering relative phase throughout the slice selection. With both the gradients G_x and G_y acting together, the matrix of voxels of the axial slice could be represented by unique groups of frequencies and phase angles, one for each voxel in the slice. A Fourier transformation allows phase information to be extracted so that a voxel (x,y) in the slice can be assigned the intensity of signal which has the correct phase and frequency corresponding to the appropriate volume element. The signal intensity is then converted to a grey scale to form an image.

K-space

Applying the gradients mean that each voxel is spatially differentiated. As the slice data is collected, it is slotted into a 2D representation called ”K-space”, which is in the spatial frequency domain, by using two-dimensional Fourier transform. Each point in the k-space represents a particular frequency and phase and can be represented as (k_{FE}, k_{PE}): each column of k-space contains the data obtained during one frequency encoding step. Each row is filled in by repeating the phase-encoding steps. The two-dimensional inverse Fourier transform of k-space data $S(k_x, k_y)$, which is nothing but the spatial frequency with units cm^{-1}, gives a complex image of $\rho(x, y)$, which has both real and imaginary components.

$$\rho(x, y) = \iint S(k_x, k_y) e^{i2\pi(k_x x + k_y y)} dk_x dk_y \tag{A.9}$$

MR images are represented as the magnitude of these real and imaginary components. In practice, since the data measurements are made discreetly over a finite region, the discrete Fourier transform is used. The spatial resolution of the image, in fact, depends on the number of k-space measurements made and therefore there is a sort of trade-off between the number of measurements (time of acquisition) and spatial resolution.
The most two popular ways of acquiring data in k-space are Echo-Planar Imaging (EPI), which basically sample k-space in a sort of a cartesian ”zig-zag” grid, and Spirals, which starts from the center and measures outward following a spiral trajectory. It’s important to note that is not a one-to-one relationship between the image space and k-space; this means that there is not a single measurements in the k-space that gives all the information about a single voxel of the brain, but rather all the points in k-space contain a little information about every voxel so each individual point in the image space depends on all the points contained in k-space.

A.1.5 MRI sequences

An MRI sequence is a combination of RF and gradient pulses which are designed and sequentially arranged to acquire the data to form the desired image. MRI signal intensity depends on many parameters, including proton density, T1 and T2 relaxation times. Different pathologies can be selected by the proper choice of pulse sequence parameters. Repetition time (TR) is the time between two consecutive RF pulses measured in milliseconds. For a given type of nucleus in a given environment, TR determines the amount of T1 relaxation. The longer the TR, the more the longitudinal magnetisation is recovered. Tissues with short T1 have greater signal intensity than tissues with a longer T1 at a given TR. A long TR allows more magnetisation to recover and thus reduces differences in the T1 contribution in the image contrast. Echo time (TE) is the time from the application of an RF pulse to the measurement of the MR signal. TE determines how much decay of the transverse magnetisation is allowed to occur before the signal is read. It therefore controls the amount of T2 relaxation. The application of RF pulses at different TRs and the receiving of signals at different TEs produces variation in contrast in MR images.

Spin Echo (SE) pulse sequence

The spin echo (SE) sequence comprises two radiofrequency pulses - the 90° pulse that creates the detectable magnetisation and the 180° pulse that refocuses it at TE. The selection of TE and TR determines resulting image contrast. In T1-weighted images, tissues that have short T1 relaxation times (such as fat) present as bright signal. Tissues with long T1 relaxation times (such as cysts, cerebrospinal fluid and edema) show as dark signal. In T2-weighted images, tissues that have long T2 relaxation times (such as fluids) appear bright. In cerebral tissue, differences in T1 relaxation times between white and grey matter permit the differentiation of these tissues on heavily T1-weighted images. Proton density-weighted images also allow distinction of white and grey matter, with tissue signal intensities mirroring those obtained on T2-weighted images. In general, T1-weighted images provide excellent anatomical detail, while T2-weighted images are often superior for detecting pathology.

Gradient Recalled Echo (GRE) pulse sequences

Gradient recalled echo (GRE) sequences, which are significantly faster than SE sequences, differ from SE sequences in that there is no 180° refocusing RF pulse. In addition, the single RF pulse in a GRE sequence is usually switched on for less time than the 90° pulse used in SE sequences. The scan time can be reduced by using a shorter TR, but this is at the expense of the signal to noise ratio (SNR) which drops due to magnetic susceptibility between tissues. At the interface of bone and tissue or air and tissue, there is an apparent loss of signal that is heightened as TE is increased. Therefore it is usually inappropriate to acquire T2-weighted images with the use of GRE sequences. Nevertheless, GRE sequences are widely used for obtaining T1-weighted images for a large number of slices or a volume of tissue in order to keep scanning times to a minimum. GRE sequences are often used to acquire T1-weighted 3D volume data that can be reformatted to display image sections in any plane.

Summing up, spins or protons immersed in an external magnetic field precess around their axes as well as around the axis of the static magnetic field and they have a position either parallel (low energy) or anti-parallel (high energy) along the axis of the static magnetic field. A radio frequency pulse is then used to knock that spin out of the alignment of the static magnetic field after which precession occurs back to its original or resting state and by using a radio frequency receiver one can measure the energy that is sent out by this precession process and measure either the longitudinal or transverse relaxation time, that is the time that is necessary for that spins system to go back to its relaxed or low energy state. By measuring that we can create a T2 signature and different type of tissue have different relaxation times and we can use that information to create an image of the structure that we're trying to create an NMRI image of. Basically MRI studies brain anatomy; typically, structural images, or T1 images, are taken and these images have high spatial resolution and can thus be able to distinguish different types of tissues. Functional MRI instead studies brain function; here, functional images, or T2* images, are taken and these images have lower spatial resolution but higher temporal resolution. Here one can relate changes in signal to experimental manipulations. An fMRI experiment consists of a sequence of individual MR images, where one can study oxygenation changes in the brain across time and then correlate it with whatever experimental manipulation has been done.

A.2 BOLD signal

For what concerns functional MRI, it is based on very similar principles but its focused on a slightly different aspect. The most common approach towards fMRI uses the Blood Oxygenation Level Dependent (BOLD) contrast. It allows to measure the ratio of oxygenated to deoxygenated hemoglobin in the blood. This method exploits the difference in magnetic properties of hemiglobin in its two states, oxygenated and deoxygenated. In its basic resting state, the brain has capillaries, arteries and veins that manage the blood supply to the brain, and there is a certain ratio of oxygenated versus deoxygenated hemoglobin present. When neurons are active, or when a particular area of the brain is active, there is an energy consumption and consequently and increasing demand of oxygen and glucose to sustain this activity; these substances are supplied to the brain by the vascular system in order to replenish for this request. Therefore, at first oxygen is consumed by neurons and there is more deoxygenated hemoglobin present locally than there is oxygenated hemoglobin. Later, an influx of additional oxygenated blood is supplied to the area for that replenishment, increasing and changing again the ratio of oxygenated versus deoxygenated hemoglobin. Oxygenated and deoxygenated hemoglobin have different effects on dephasing effect, with deoxygenated hemoglobin causing more dephasing than oxygenated hemoglobin does. Basically, fMRI measures the change in the homogeneity in the magnetic field within a particular volume, due to varying oxygenated/deoxygenated blood ratio, which is referred to as T2*. When the T2 relaxation is measured, the oxygenated blood can be differentiated from the deoxygenated blood for a particular area in the brain. There's a slight change in the frequency that is measured at each precession individually and by focusing on that variation, which is a measure of the phase decay that's happening, one can take an estimate of the local distortion of the magnetic field which is thought to be the result of a change in the oxygenated versus deoxygenated blood ratio. The reason for this different behavior is that deoxygenated blood is paramagnetic and it introduces inhomogeneity. It distorts the local magnetic field that is measured by the T2* measurement. Oxygenated hemoglobin, instead, is weakly diamagnetic and has very little effect on the magnetic field. So, essentially it does not distort the signal in the considered volume. When oxygen is absorbed by the astrocytes to replenish oxygen and glucose metabolism in the cell that has been firing, it causes hemoglobin-induced dephasing which causes a change in the MRI signal that one is measuring. After a certain period of time, deoxygenated blood will cause more distortions locally in that area than oxygenated does, and by picking up that difference one can draw a conclusion that brain activity must occur in that area.

The basic idea is that a stimulus result in brain activation: in an experimental design for example, a person could be asked to tap their finger very specifically. Initially, oxygen necessary for that brain activation to occur is removed from the blood. This depletion of the oxygenation results in an initial dip in the MRI signal. In response to this brain activation, the blood supply system creates an influx of oxygenated blood that then gives rise to the BOLD signal, as it does not distort the MRI signal locally, until it reaches a top. At that point, the activation or the stimulus is removed, for example asking the person to stop tapping their finger; at that point the oxygenation and MRI signal drop as the cognitive task ends. It typically overshoots beyond the base line a little bit a few seconds until it comes back and the ratio of oxygenated and deoxygenated blood and MRI signal are back to baseline and essentially back into its resting state. This waveform just described is knwn as hemodynamic response funtion (HRF). A blood supply increase and drop in response to a cognitive function or cognitive task that the brain is executing, essentially producing the brain activation signal that we need to measure in a particular area of the brain. It is very important to note that BOLD fMRI does not measure neural activity directly, rather it measures metabolic demands, oxygen consumption of active neurons. The hemodynamic response function mentioned above represents the change in the fMRI signal triggered by this neural activity.

A.3 fMRI artifacts and Noise

Bold fMRI signal contains multiple sources of noise related to the hardware and the participants themselves, what they do, how much they move in the scanner. Sources of noise include thermal motion of free electrons in the system, gradient and magnetic field instability, causing spikes, head movement and its interactions with magnetic field, physiological effects including heartbeat and respiration and their effect on the movement of the chest wall and also on parameters that interact with the vasculature like CO_2 levels. When modeling fMRI time series data is important to understand certain non signal related components of the signal. One of the main components is *drift*. These are slow changes in voxel intensity over time, also called low frequency noise, which is always present in fMRI signal; one of the primary reasons for this is thought to be scanner instabilities as drift has been seen even in cadavers. It is important that experimental conditions that vary slowly not be confused with drift; for this reason experimental design should use high frequencies (so use more rapid alternations of stimulus on/off states). Another important source of confounds is motion: subject motion during the experiment can give rise to serious problems. Typically motion correction is performed in the pre-processing stages of the analysis. Respiration and heart beat, referred to as physiological noise, also give rise to noise to a particular frequency.

Appendix B

Statistical Parametric Mapping (SPM)

B.1 fMRI analysis

Statistical parametric mapping or SPM is a statistical technique created by Karl Friston for examining differences in brain activity recorded during functional neuroimaging experiments using neuroimaging technologies such as fMRI or PET. Brain mapping studies are usually analyzed with some form of statistical parametric mapping. This entails the construction of continuous statistical processes to test hypotheses about regionally specific effects. Statistical Parametric Maps (SPM) are images or fields with values that are, under the null hypothesis, distributed according to a known probability density function, usually the Student's t or F-distributions. These are known colloquially as t- or F-maps. The success of statistical parametric mapping is due largely to the simplicity of the idea. Namely, each and every voxel (i.e., image volume element) is analysed using any standard (univariate) statistical test, usually based on a General Linear Model (GLM) of the data. The resulting statistics are assembled into an image - the SPM. SPMs are interpreted as continuous statistical processes by referring to the probabilistic behaviour of random fields [139][140][141]. Random fields model both the univariate probabilistic characteristics of an SPM and any non-stationary spatial covariance structure. 'Unlikely' topological features of the SPM, like peaks or clusters, are interpreted as regionally specific effects, attributable to the experimental manipulation. In short, the GLM is used to explain continuous (image) data in exactly the same way as in conventional analyses of discrete data. Random Field Theory (RFT) is used to resolve the multiple-comparison problem when making inferences over

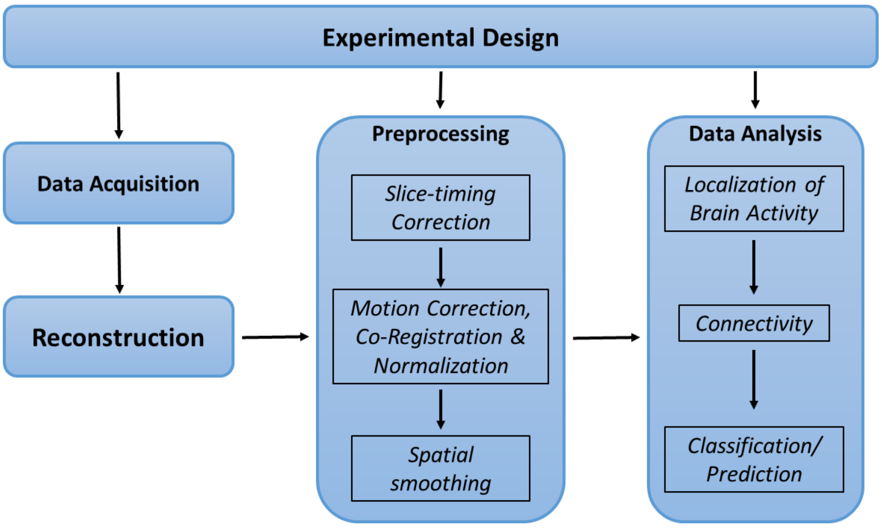


Figure B.1: Data processing pipeline.

the volume analysed. RFT provides a method for adjusting p-values for the search volume and plays the same role for SPMs as the Bonferroni correction for discrete statistical tests.

Researchers are often interested in examining brain activity linked to a specific psychological process or processes. An experimental approach to this problem might involve asking the question "which areas of the brain are significantly more active when a person is doing task A compared to task B?". Although each task might be designed to be identical, except for the aspect of behaviour under investigation, the brain is still likely to show changes in activity between tasks due to factors other than task differences (as the brain is involved with co-ordinating a whole range of parallel functions unrelated to the experimental task). Furthermore, the signal may contain noise from the imaging process itself.

To accommodate these random effects, and to highlight the areas of activity linked specifically to the process under investigation, statistics are used to look for the most significant difference above and beyond background brain activity. This involves a multi-stage process to prepare the data, and to subsequently analyse it using a statistical method known as the General Linear Model (GLM).

B.1.1 Image pre-processing

Before the analysis, fMRI data undergoes a series of pre-processing steps which are basically aimed at identifying and removing artifacts and val-

idating certain model assumptions. In general, there is a three-fold goal with pre-processing: one is to minimize the influence of data acquisition and physiological artifacts; a second one is to check statistical assumptions and transform the data to meet assumptions; the third is to standardize the locations of brain regions across subjects to achieve validity and sensitivity in group analysis. Pre-processing is performed both on the fMRI data and on the structural scans that are collected prior to the experiment.
Typically this pre-processing procedure involves the following steps:

- Visualisation and Artifact Removal
- Slice Time Correction
- Motion Corrections
- Physiological Corrections
- Co-registration
- Normalization
- Spatial Filtering
- Temporal Filtering

Visualisation and Artifact removal

This should always be the first part of any pre-processing pipeline. Exploratory data techniques are used to investigate the raw image data and detect possible problems and artifacts. For example, fMRI data often contain transient spike artifacts or slow drift over time. Any exploratory technique such as principal components can be used to look for these spike related artifacts.

Slice timing Correction

Usually, multiple slices of the brain are sampled during each individual repetition time to construct a brain volume. However, each slice is typically sampled at different time points, because they are acquired sequentially, and so even though a brain volume is treated as being scanned at a single time point, this is usually not the case, for example the top of the brain volume might be sampled a second or two later than the bottom of it, and this is something that need to be corrected before doing any statistical analysis. Basically, during slice time correction, each voxel's time series is shifted so they appear to have been sampled simultaneously. **Figure B.2** illustrate

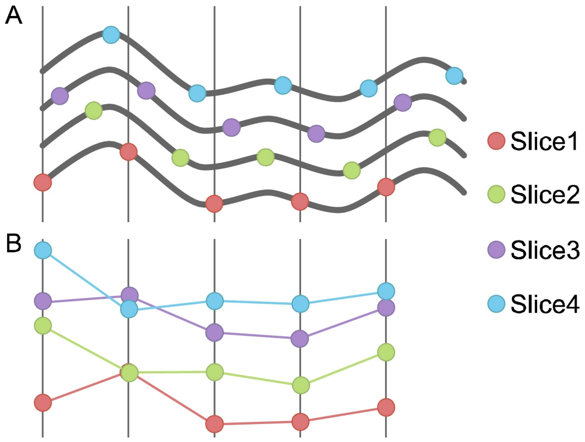


Figure B.2: The slice-timing problem: the same signal sampled at different offsets yields signals that do not look the same. (A) Five adjacent slices acquired with interleaved acquisition all sample the same underlying bold signal. (B) Without correction, reconstruction yields five different signals despite having the same underlying shape.

this situation in which the same single voxel is sampled, belonging to four different slices of the brain and they all have the same hemodynamic response signal. However, if a time course is extracted from Slice 1, Slice 2, Slice 3 and Slice 4 and without slice time correction, they are going to look quite different, even though they come from the same underlying signal. This can be corrected using temporal interpolation and other methods. In temporal interpolation, the information from nearby time points is used to estimate the amplitude of the MR signal at the onset of each TR. Conventional interpolation-based techniques attempt to operate on the discrete signals (**Figure B.2b**) and estimate the signal value between the sample points. Mathematically, this can be represented by the convolution of the sampled signal, $F[s,n]$ with an interpolation kernel h:

$$\hat{f}[s,n] = \sum_{k=0}^{N-1} F[s,k]\, h(nT_s - k - \phi(z)) \tag{B.1}$$

where $\hat{f}$ is the shifted/interpolated signal and $\phi(z)$ is the offset delay imposed for each slice by interleaved acquisition. Linear, cubic spline or sinc functions are commonly used to interpolate time points [142]. Alternatively one can use Fourier methods and the phase shift methods and slide the time course by applying a phase shift to the Fourier transform of the time course.

Motion Correction

Head motion represents a serious confound in neuroimaging, and particularly in fMRI studies. Folowing are several criticalities that can be introduced by head motion: movement artefacts add up to the residual variance and reduce sensitivity; data may be lost if sudden movements occur during a single volume; movements may be correlated with the task performed. Even very small movements of the head during an experiment can be a major source of error if not treated correctly. Despite restraints on head movement, co-operative subjects still show displacements of up to several millimeters.
When analyzing a time series associated with a single voxel, we always assume that it depicts the same region of the brain at every time point. However, if there is head movement, one voxel that was in one position might be in a completely different brain area because of head motion, and this is something that need correction. One first procedure for the correction of motion artifacts is *Realignment*: it involves the estimation parameters of an affine 'rigid-body' transformation that minimizes the differences between each successive scan and a reference scan (usually the first scan or the average of all scans in the time series); here the brain is assumed to be 'rigid' and thus brain shape is assumed not to change. A rigid-body transformation process is composed by two steps. The first step called 'Registration' involves the estimation of six parameters describing a spatial trasformation between the source images and reference image that "best" match the images together. The goodness of the match is based on an objective funtion, which is maximized or minimized using some optimization algorithm. We have three sets of translations (*mm*) in the x, y, and z direction (*degrees*), and three sets of rotations, one around each axis (pitch, yaw and roll).
The transformations can be represented as matrices, and are multiplied together:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & 0 & \sin\phi & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\phi & 0 & \cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi & 0 \\ 0 & \sin\psi & \cos\psi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Registering images that have been collected with the same modality allows a relatively simple obective function to be used, i.e. the mean squared difference between the images. Estimation of the affine transformation is usually effected with a first-order approximation of the Taylor expansion of the effect of movement on signal intensity, using the spatial derivatives

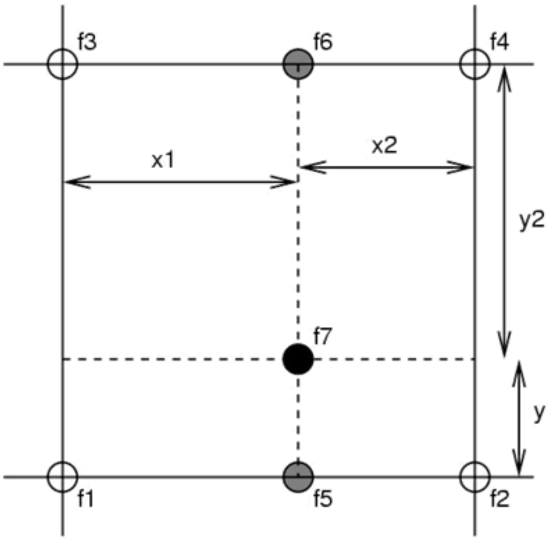


Figure B.3: Example of bi-linear interpolation. The values of the "new" voxel is computed as weights average of neighbouring voxels as:
$f_5 = f_1 x_2 + f_2 x_1$, $f_6 = f_3 x_2 + f_4 x_1$, $f_7 = f_5 y_2 + f_6 y_1$

of the images. This allows for a simple iterative least square solution that corresponds to a Gauss-Newton search ([113]).

– Mathematical description of Gauss-Newton algorithm and implementation on realignment ??? (see statistical parametric mapping libro)

After the registration step, the transformation parameters are estimated and applied to each slice to match the reference image (usually the first slice of the time series); at this point it is necessary to determine the values of the "new" transformed voxels. In the second step, 'Transformation' (or "reslicing") is then applied by resampling the data using an interpolation scheme. There are a number of interpolation methods commonly used, including Simple interpolation and B-spline interpolation. The simplest interpolation technique, known as nearest neighbour, consists in assigning a value of a specific voxel by taking the value of the nearest neighbour. A rather better option is Tri-linear interpolation, which consists of a weighted average of the neighbouring voxels. This is slower than nearest enighbour, but the resulting images are less "blocky". In **Figure B.3** is illustrated a bi-dimensional example of linear interpolation (bi-linear in this case). This method is not really optimal because by taking the weighted average of the neighbouring voxel we are introducing smoothness into the resampled image causing the loss of information. There are better ways of doing interpolation although such methods could be slower. This is the case of B-spline interpolation, a generalized interpolation method. In this technique the values of voxels in the original image are interpolated using a linear combination of gaussians basis functions. B-splines are a family of functions of varying degree. Interpolation using B-splines of degree 0 or 1 (first and second order) is almost identical to nearest neighbour or linear interpolation respectively.

The iterative minimization procedure used for fMRI motion correction schemes is typically a nonlinear least squares routine (e.g., Levenberg-Marquardt). Possible errors in this process may result from finding only a local (rather than global) minimum of the cost function, leading to suboptimal results. Once motion parameters for realignment have been determined, they are applied to create a new 3D motion-corrected data set. Creation of the corrected data set requires spatial interpolation, as the new data points typically fall in between the original uncorrected data points. This process may be computationally-intensive, so usually some combination of (fast) linear interpolation is used during the initial motion correction steps, followed by a more time-consuming interpolation method (such as windowed sinc) for the final spatial transformation. For most imaging modalities this realignment procedure is sufficient to realign scans to, in some instances, a hundred microns or so ([143]). Typically head motions are relatively small (¡2 mm) during normal fMRI experiments, so the assumptions underlying this rigid body approach are justified.

However, sudden abrupt head motion will violate these conditions and may not produce an appropriate correction for motion. In fMRI, in fact, even after perfect realignment, movement-related signals may still persist. This requires a further step in which the data are adjusted for residual movement-related effects. Rigid body transformations cannot compensate for nonlinear effects. These include field inhomogeneity effects, motion during slice acquisition, interpolation artifacts, and spin-excitation history effects. Field inhomogeneity effects may be the most important. Even though rigid head motion can be corrected in image space, head displacements affect magnetic field homogeneity and shimming, so even with perfect realignment some motion-related errors persist. These additional sources of residual motion error are often referred to as the residual variance and may be addressed in part by *unwarping.*

Field mapping and unwarping

In extreme cases, as much as 90% of the variance in fMRI time-series can be accounted for by the effects of movement after realignment ([143]). Causes of these movement-related components are due to movement effects that cannot be modelled using a linear model. These non-linear effects include: subject movement between slice acquisition, interpolation artefacts due to resampling (Grootoonk et al. 2000), non-linear distortion due to magnetic field inhomogeneities (Anderson et al. 2001) and spin-excitation history effects ([143]). One simple way to correct for this residual variance is to incorporate movement parameters as confounds in the statistical model of activations. However, this may remove activations of interest if they are correlated with the movement. Another approach for the correction of non-linear magnetic field inhomogeneities is "unwarping". Different tissues have different magnetic susceptibilities. The magnetic susceptibility of a substance is the measure of the extent to which the substance modifies the strength of the magnetic field passing through it or alternatively the degree of magnetization of a material in respoonse to an applied magnetic field. In the human body this difference of magnetic susceptibility among tissues causes distortions in the magnetic field, distortions that are most noticeable near air-tissue interfaces. These field inhomogeneities have the effect of "deflecting" locations of the image with respect to the real object. In fact, knowing the location at which ^{1}H spins will precess at a particular frequency and thus where the signal comes from is dependent upon correctly assigning a particular field strength to a particular location. If the field B_0 is homogeneous, then the image is sampled according to a regular grid and voxels can be localized to the same bit of brain tissue over subsequent scans by realigning. This is because the same transformation is applied to all voxels between each scan. If, instead, there are inhomogeneities in B_0, then different deformations will occur at different points in the field over different scan, giving rise to non-rigid deformation. Therefore the combination of field inhomogeneities with head movement will result in changes in the image that do not really follow the rigid body assumption. Filed inhomogeneities change with the position of the object in the field, so there can be non-rigid, as well as rigid distortion over subsequent scans. The movemente-by-inhomogeneity interaction can be observed by changes in the deformation field over subsequent scan. A deformation field indicates the directions and magnitudes of location deflections throughout the magnetic field (B_0) with respect to the real object. A fieldmap measures field inhomogeneity (potentially per every scan) and thus captures deformation field. Basic rapid field mapping is commonly performed as a normal part of automated prescan procedures. In the simplest case a low-resolution dual echo sequence with relatively short TE's is first performed, with computation of magnitude and phase-difference images. The phase-difference images require unwrapping as all phase measurements are "wrapped" (compressed) over the range of $0° - 360°$. After phase unwrapping field map values are used to calculate corrective pixel shifts in the phase-encode direction, a process known as unwarping. Using fieldmaps, unwarping can estimate changes in distortion from motion parameters obtained from alignment, and return the predicted change in deformation field with subject movement (estimated via iteration) at each time point. Basically, if a deformation field can be thought of as little vectors at each position in space showing how that particular location has been deflected, "derivative field" is then the rate of change of those vectors with respect to subject movement. Given these "derivative fields" we should be able to remove the variance caused by the susceptibility-by-movement interaction. This means that the time-series will be undistorted to some "average distortion" state rather than to the true geometry. If one wants additionally to address the issue of anatomical fidelity one should combine Unwarp with a measured field-map. The description above can be thought of in terms of a Taylor expansion of the field as a function of subject movement. Unwarp alone will estimate the first (and optionally second, see below) order terms of this expansion. It cannot estimate the zeroth order term (the distortions common to all scans in the time series) since that doesn't introduce (almost) any variance in the time series. The measured fieldmap takes the role of the zeroth order term. Once the deformation field has been modelled over time, the time-variant field is applied to the image. The image is therefore resampled assuming voxels, corresponding to the same bits of brain tissue, occur at different locations over time. Unwarp is of use when variance due to movement is large. It is particularly useful when the movements are task related as can remove unwanted variance without removing "true" activations. It is useful when high field strength or long readout time increases amount of distortion in images. However, it can be computationally intensive and increasing computational times.

Co-registration

After performing motion correction, the structural MRI image that is collected in the beginning of the session, is registered to the fMRI image in a process known as co-registration. Co-registration maximizes the mutual information between two images of different modalities. Co-registration refers to the alignment of functional and structural images from the same subject to map functional information into anatomical space. By coregistering the structural and the functional images to one another, this allows one to visualize single-subject task activations overlaid on the individual's anatomical information and can relate changes in BOLD signal due to experimental manipulation to anatomic al structures. So even though the statistical analysis is on low-resolution functional data, we might want to present the results of the analysis on a high-resolution structural scan where we can make out detail. This also simplifies later transformation of the fMRI images to a standard coordinate system, as will be described later in the "Spatial Normalization" section. The difference between the functional and structural images from the same brain is not trivial. By contrast to the high-resolution structural images with clear region boundary contours, functional images are normally blurry and suffered from geometric and intensity distortions - particularly in the phase encode direction (Jezzard and Clare, 1999; Jezzard, 2000). The basic idea regarding coregistration herein is similar to the realignment, i.e. defining a cost function with the goal to minimize the differences on image parameters among images. Essentially, coregistration is a process similar to realignment that aligns two images but from different modalities (i.e. T1 structural and T2 functional images) from the same individual (within subject). In the same way as the realignment process, coregistration follows the rigid-body transformation steps of registration, transformation, interpolation (nearest neighbour, linear or B-spline: the method used depends on the type pf adta and the research question; the default in SPM is 4^{th} order B-spline) and reslicing. However, because of distortions on functional images, the rigid-body transformation with six parameters (3 translations and 3 rotations) may be not enough to correct. Depending on the complexity of distortions, either a nine-parameter transformation with another three additional parameters accounting for scaling differences on x-, y- or z- axes or even more sophisticated algorithms could be used to quantify the cost function. The overall procedure changes the position without changing the value of the voxels and give correspondence between voxels. As the two images are of different modalities, a least squares approach cannot be performed because of the different shape and signal intensities (relative intensities of grey/white matter vary between funtional and structural, no voxel to voxel match, no simple subtraction signal intensities). Therefore, as a result of the different contrasts between functional and structural images, the mutual information is more suitable to act as cost function than the sum of squared differences (Collignon et al., 1995; Wells et al., 1996). The mutual information (MI) (also known as *Shannon information*) between two signals can be expressed as:

$$I(f,g) = H(f) + H(g) - H(f,g) \tag{B.2}$$

where $H(f,g)$ is the joint entropy of the images (also referred to as entropy of the joint probability), and $H(f)$ and $H(g)$ are the marginalized entropies:

$$H(f,g) = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P(f,g)\, logP(f,g)\, df dg$$

$$H(f) = -\int_{-\infty}^{+\infty} P(f)\, logP(f)\, df \tag{B.3}$$

$$H(g) = -\int_{-\infty}^{+\infty} P(g)\, logP(g)\, dg$$

MI is a measure of dependence of one image on the other, and can be considered as the distance (Kullbacl-Leibler divergence) between the joint distribution ($P(f,g)$) and the distribution assuming complete independence ($P(f)P(g)$). When the two distributions are identical, this distance (and the mutual information) is zero. Rearranging, the expression for MI becomes:

$$\begin{aligned} I(f,g) &= KL(P(f,g)||P(f)P(g)) \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P(f,g)\, log(\frac{P(f,g)}{P(f)P(g)})\, df dg \end{aligned} \tag{B.4}$$

The MI between two images is maximized when they are in register. The discrete representation of the probability distributions is from a joint histogram, which can be considered as an $IbyJ$ matrix P. The entropy is then computed from the histogram according to:

$$H(f,g) = \sum_{j=1}^{J}\sum_{i=1}^{I} p_{ij}\, log\, p_{ij} \tag{B.5}$$

Generating a joint histogram involves scanning through the voxels of the reference image and finding the corresponding points of the source. The appropriate bin in the histogram is incremented by one for each of these point pairs. Pairs are ignored if the corresponding voxel is unavailable because it lieas outside the image volume. The coordinate of the corresponding point rarely lies at an actual voxel centre, meaning that interpolation is required.

Spatial Normalisation

Spatial normalisation involves warping all the grey matter images to the same stereotaxic space, which is achieved by matching to a common template image. In general, all brains are different and the brain size of two subjects can differ up to 30%. There also may be substantial variation in the shapes of people's brains. What normalisation does, is to allow one to stretch, squeeze, and warp each brain image so that it matches some standard brain template. This is really important for group analysis, because one want to be able to compare different brains with one another, and if considering a single voxel one want to be able to look at that voxel across the entire population of the subjects. In order to do that, one need all the brains to be registered to the same standard brain space. Therefore, using the co-registration procedures, the structural MR image needs to be warped into the template space. The pros of this procedure are that the results can be reported and interpreted in a consistent manner: having images mapped to a standard space allows the activation sites to be reported according to their Euclidian coordinates within the standard space. Results can in this way be generalised with a larger population, because one can compare different subjects with one another. Also results can be compared across studies, if each study is normalised for the same brain, and they can also be averaged across subjects. The cons are that it reduces spatial resolution and introduces potential errors. In realignment, the generative model for within-subject movements is a rigid-body displacement of the first image. The generative model for spatial normalisation instead is a canonical image or template that is distorted to produce a subject-specific image. Spatial normalisation inverts this model by undoing the warp using a template matching procedure. After realigning the data, a mean image of the series, or some other co-registered (e.g. a T1-weighted) image, is used to estimate some warping parameters that map it onto a template that already conforms to some standard stereotaxic anatomical space like the Talairach space or the MNI space (Montreal Neurological Institute). Currently the most common type of spatial normalisation, implemented in the SPM12 software used in this work, is the *non-label based (intensity based)* normalisation. This adopts a two-stage procedure in which the first step is represented by a linear registration, which involves a 12-parameter affine transformation where the parameters constitute a spatial transformation matrix: the 3 translation and 3 rotations of a rigid-body transformation, and additionally 3 shears and 3 zooms. This estimation is performed automatically by minimizing squared distance between parameters and expected values. The second step involves a non-linear registration (warping), that accounts for smaller-scale anatomical diefferences (e.g head shape). In this step, warps are modelled by linear combinations of low-frequency cosine basis functions (sometimes polynomials are used).

Estimation of the parameters of all these models can be accommodated in a Bayesian framework, in order to obtain a more robust fit. This requires knowledge of the errors associated with the parameter estimates, and also knowledge of the a priori distribution from which the parameters are drawn. In this framework the idea then is to find the warping parameters θ that have the maximum posterior probability $p(\theta|y)$ given the data y, where $p(\theta|y)p(y) = p(y|\theta)p(\theta)$. Put simply, one wants to find the deformation that is most likely given the data. This deformation can be found by maximizing the probability of getting the data, given the current parameters, times the probability of those parameters. In practice, the deformation is updated iteratively using a Gauss-Newton scheme to maximize $p(\theta|y)$. This involves jointly minimizing the likelihood and prior potentials $H(y|\theta) = ln\, p(y|\theta)$ and $H(\theta) = ln\, p(\theta)$. The likelihood potential is generally taken to be the sum of squared differences between the template and deformed image and reflects the probability of actually getting that image if the transformation was correct. The prior potential can be used to incorporate prior information or constraints on the warp. Priors can be determined empirically or motivated by constraints on the mappings. Priors play a more essential role as the number of parameters specifying the mapping increases and are central to high-dimensional warping schemes (Ashburner et al. 1997). The deformations required to transform images to the same space are not clearly defined. Unlike rigid body transformations, where the constraints are explicit, those for nonlinear warping are more arbitrary. Without any constraints it is of course possible to transform any image such that it matches another exactly. The issue is therefore less about the nature of the transformation and more about defining constraints or priors under which a transformation is effected. The validity of a transformation can usually be reduced to the validity of these priors. In practice, most people use an affine or spatial basis function warps and iterative least squares to minimize the posterior potential. A nice extension of this approach is that the likelihood potential can be refined and taken as the difference between the index image and a mixture of templates (e.g. depicting grey, white and skull tissue pertitions). This models intensity differences that are unrelated to registration differences and allows different modalities to be co-registered ([113]).

When applying non-linear warping one risk is to incur in over-fitting, introducing unrealistic deformations in the image; usually is more preferable to have a slightly less-good match, that is still anatomically realistic. In order to avoid this, some regularization terms/constraints must be included in normalization process. Without regularization in the nonlinear registration it is possible to introduce unnecessary deformations that only reduce the residual sum of squares by a tiny amount. This could potentially make the algorithm very unstable. Regularization is achieved by minimizing the sum of squared difference between the template and the warped image, while simultaneously minimizing some function of the deformation field. This basically ensures voxels stay close to their neighbours and it involves setting limits to the parameters used in the flexible warp (affine transformation and weights for basis funtions) The normalization step has still some limitations. First of all it is difficult to attempt exact structural matches between subjects, due to individual anatomical differences. Moreover, even if anatomical areas were exactly matched, it does not mean fucntionally homologus areas are matched too. This is particularly problematic in patients studies with lesioned brains. One possible solution is to correct gross differences by applying spatial smoothing to normalized images.

Spatial filtering or Smoothing

In fMRI is common to spatially smooth the acquired data prior to statistical analysis. THe reasons for applying spatial filtering are multiple. It can increase the signal-to-noise ratio and therefore increase sensitivity: if there is a coherent region of activation, it's actually beneficial to average over that region because the same signal is maintained but the noise is decreased, and thus a higher signal to noise ratio is obtained. It can also validate certain distributional assumptions and remove artifacts. For example, it is often assumed that data is normally distributed, and by smoothing one can increase that likelihood by averaging over lots of the different voxels. This way it may increase the validity of statistical analysis, and finally it is also required for Gaussian random field theory, which is often used in multiple comparison. One other prop of spatial filtering is that it may overcome limitations in the normalization by blurring residual anatomical differences between subjects, allowing for better spatial overlapping. One major drawack is that the image resolution is reduced by spatial smoothing, and thus a portion of information in space is lost.

When doing spatial filtering, it is often used a Gaussian kernel. The size of the kernel is determined by something called the full width at half maximum, that measures the width of the kernel at 50% of its peak value. The relationship between the full-width half maximum and the standard deviation can be expressed as

$$\sigma = \frac{FWHM}{2\sqrt{2ln(2)}} \tag{B.6}$$

For what concerns the choice of the width of the filter, the matched filter theorem states that a filter that's matched to the signal extent will give optimal signal to noise ratio. So if one knows that the spatial extent is of certain size, one should smooth over those voxels, because in this way the signal is retained while reducing the noise.

However, typically the amount of smoothing is chosen at priori before we look at activations, and it's usually independent of the data. Hence, the likelihood what one get the smoothing kernel matched to the signal extent is very unlikely. Furthermore, the same amount of smoothing is applied throughout the whole image. And the spacial extent of activation is probably going to differ across the brain. In order to circumvent this, one could use adaptive smoothing methods. And there are such methods including non-stationary spatial Gaussian Markov random fields, and this allows the smoothing to vary across both space and time.

The warped grey-matter images are now smoothed by convolving with an isotropic Gaussian kernel. This makes the subsequent voxel-by-voxel analysis comparable to a region of interest approach, because each voxel in the smoothed image contains the average amount of grey matter from around the voxel (where the region around the voxel is defined by the from of the smoothing kernel). This is often referred to as grey-matter density, but should not be confused with cell packing density measured cytoarchitectonically. Critically, smoothing removes finescale structure from the data that is not conserved from subject to subject. This increases the sensitivity of VBM to differences that are expressed at a larger spatial scale. The smoothing conforms to the matched filter theorem, which says that the smoothing should match the scale of the differences in question. Normally, the smoothing kernel is Gaussian with a FWHM between 4 and 16 mm. By the central limit theorem, smoothing also has the effect of rendering the data more normally distributed, thus increasing the validity of parametric statistical tests.

By the matched filter teorem, the optimum smoothing kernel corresponds to the size of the effect that one anticipates. The spatial scale of haemodynamic responses is, according to high-resolution optical imaging experiments, about 2-5mm. Despite the potentially high resolution afforded by fMRI, an equivalent smoothing is suggested for most applications. By the central limit theorem, smoothing the data will render the errors more normal in their distribution and ensure the validity of inferences based on parametric tests. When making inferences about regional effects using random field theory the assumption is that the error terms are a reasonable lattice representation of an underlying continuous scalar field. This necessitates smoothness to be substantially greater than voxel size. If the voxels are large, then they can be reduced by sub-sampling the data and smoothing (with the original point spread function) with little loss of intrinsic resolution. In the context of inter-subject averaging it is often necessary to smooth data onto a spatial scale where homologies in functional anatomy are expressed among subjects.

Temporal filtering

The last step in the pre-processing phase of fMRI is, optionally, temporal filtering. The typical power spectrum of the signal coming from a subject at rest in the scanner, highlights the dominant presence of low frequency noise components. This residual noise arises mostly from physical sources, sometimes referred to as "scanner drift". However, by filtering the data with an appropriate high-pass filter, most of the noise can be removed, increasing the SNR. Ideally, the reamaining noise spectrum would be flat (i.e. "white noise", with equal power at all frequencies). The choice of the high pass cut-off would ideally maximize the signal-to-noise ratio. However, one cannot distinguish signal from noise on the basis of the power spectrum alone. Usually, a cut-off period of approximately $128s$ is used, based on observations that the noise becomes appreciable at frequencies below approximately $0.008Hz$ (though this may vary considerably across scanners and subjects). In other words, some loss of signal may be necessary to minimize noise. Experimental designs therefore try to avoid significant power at low frequencies. This step can actually be skipped in the pre-processing phase, including, instead, low frequencies variations as confounds in the definition of the GLM in order to take into account for low frequency components in the observed data.

B.1.2 The General Linear Model (GLM)

[144][145] There are multiple goals in the analysis of fMRI data and they include localising areas activated by a task or in relation to a process of interest, determining networks corresponding to brain function, functional connectivity, and effective connectivity, and making predictions about psychological or disease states or other outcomes from functional imaging data. All of these can be handled in the general linear modelling framework (GLM). The GLM analysis process is typically a two level hierarchical analysis, that involves a first level in which within subject effects are analysed, individual by individual, and a second level analysis across subjects or across groups, in a group analysis. This can be done in stages or, alternatively, hierarchical models can combine both those levels into one integrated model. Basically this procedure involves a stage of design specification, followed by

the building of the model. Secondly, that model is combined with real data and estimated: this is done at each single voxel. Finally, contrast images are calculated.

In fact, the statistical parametric mapping approach through which functional mapping studies are analysed, entails the construction of statistical processes to test hypotheses about regionally specific effects [146]. Statistical parametric maps are images which values are, under the null hypothesis, distributed according to a known probability density function, usually the Student's t or F-distributions. The idea behind this method consists in analysing each and every voxel using a standard (univariate) statistical test. The resulting statistical parameters are assembled into an image.

Statistical parametric mapping exploit the combined use of the General Linear Model (GLM) and Random Field Theory (RFT) to analyse and make classical inferences about topological features of the statistical parametric maps. The GLM is used to estimate some parameters that explain continuous data. RFT is used to resolve the multiple comparison problem that ensues when making inferences over the volume analysed, containing thousands of voxels. RFT provides a method for adjusting p-values for the search volume and plays the same role for continuous data (i.e. images) as the Bonferroni correction for a number of discontinuous or discrete statistical tests.

The GLM family of tests

The general linear model approach treats the data as a linear combination of model functions (predictors) plus a noise term (error). Essentially the data are broken up into the part that can be explained with the model, and the part that can't be explained. These model functions are assumed to have known shapes but their amplitudes, or slopes, are unknown and those are what needs to be estimated when fitting the model.

The GLM encompasses many of the commonly used techniques in fMRI data analysis. The entire GLM family embodies a wide collection of parametric statistics including: linear regression, multiple regression, t-tests, ANOVA, ANCOVA, correlation, F-tests and so on. All of these are different instances of the general linear model framework. In some cases, there is a simple close form algebraic solution. In this case one could solve the equations and estimate the model in one step, whereas in many other cases iterative solutions are required.

The simplest example is simple linear regression: one predictor, one outcome. In this modelling procedure, four stages can be identified. The first is to specify the model. In this case, the assumption is that there is a linear relationship between the predictor and the outcome. Secondly, the model is estimated. In this case, this means estimating the slope and the intercept of that model (linear). Third, the statistical inference is performed. This stage serves to cast the significance of that slope and get a p-value, which relates to how likely it is to observe a slope like this under the null hypothesis that there is no actual true relationship (that the line is actually flat). Finally, when significant effects are found, scientific interpretation hypothesises the meaning of this relationship.

All the GLM models are characterised by the use of one variable, which is a continuous variable, as the dependent variable of the outcome. In the context of fMRI studies, this observed response variable Y, is associated with time-series representing the BOLD response of each voxel of the brain. The general linear model equation for each single voxel can be expressed as

$$y_i = \beta_0 + \beta_1\, x_{i1} + \beta_2\, x_{i2} + \cdots + \beta_k\, x_{ik} + \varepsilon_i \tag{B.7}$$

In this notation i indicates time points, so this is relative to a single data point at time t_i. β_0 represent the intercept parameter, that capture the average across time. The beta's are regression slopes. When estimating the GLM then, the job is to solve for that beta vector.
Considering all time points, the previous equation can be expressed in matrix notation as

$$\begin{bmatrix} Y_1 \\ Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1k} \\ 1 & X_{21} & \cdots & X_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{nk} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

which becomes in compact form

$$Y = X\beta + \varepsilon \tag{B.8}$$

This equation expresses the observed response variable Y in terms of a linear combination of explanatory variables X plus an error term [147]. The matrix that contains explanatory variables (e.g. designed effects or confounds) is called the *design matrix*. Each column of the design matrix corresponds to an effect that has been built into the experiment or that may confound the results. These are referred to as explanatory variables, covariates or regressors. It is essential that the regressors included in the design matrix are linearly independent; in this way the design matrix will be of full rank. The effects on the response variable are modelled in terms of functions of the presence of these conditions (e.g. delta functions convolved with the haemodynamic response function) and constitute the first n columns of the design matrix. Then, usually follows a series of nuisance terms. The very first column is whole brain activity.

The residual errors $\varepsilon_i \sim N(0, \sigma^2)$ are assumed to be independent and identically (normally) distributed. This property is also known as *"sphericity"*. Under IID (Independent and Identically Distributed) assumptions the probability density function of the errors, from all observations, has spherical iso-contours, hence sphericity. Deviations from either of the IID criteria constitute non-sphericity. If the error terms are not identically distributed then different observations have different error variances. Correlations among errors reflect dependencies among the error terms (e.g. serial correlation in fMRI time series) and constitute the second component of non-sphericity. In neuroimaging both spatial and temporal non-sphericity can be quite profound.

The violation of sphericity is serious for the repeated measures ANOVA, with violation causing the test to become too liberal (i.e., an increase in the Type I error rate). Therefore, determining whether sphericity has been violated is very important. Luckily, if violations of sphericity do occur, corrections have been developed to produce a more valid critical F-value (i.e., reduce the increase in Type I error rate). This is achieved by estimating the degree to which sphericity has been violated and applying a correction factor to the degrees of freedom of the F-distribution.

Estimation

In the parameters estimation algorithm, the objective is to find the parameters β that minimize a cost function, usually identified by the sum of squared errors (in a least squares approach). Hence, assuming IID errors and using an Ordinary least squares (OLS) estimation approach, the parameters estimates can be obtained through a minimization procedure as described below. The sum of squared error is expressed as

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - x_{i1}\hat{\beta}_1 - \cdots - x_{ik}\hat{\beta}_k)^2$$

This is minimized when:

$$\frac{\partial S}{\partial \hat{\beta}_j} = 2\sum_{i=1}^{n} (-x_{ij})(y_i - x_{i1}\hat{\beta}_1 - \cdots - x_{ik}\hat{\beta}_k) = 0$$

This equation is the j^{th} row of matrix $X^T Y = (X^T X)\hat{\beta}$. For the general linear model, the least squares estimates are the maximum likelihood es-

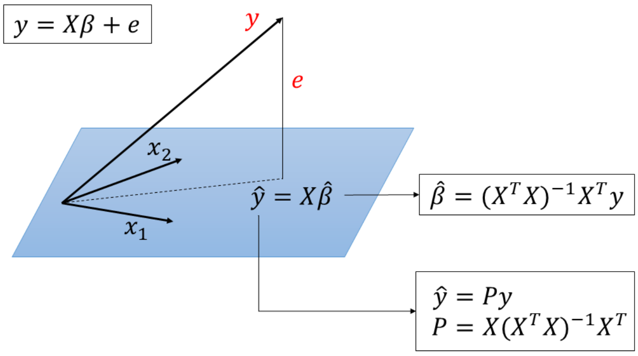


Figure B.4: Geometric perspective for the minimization of the sum of squared errors: the predicted values $\hat{y} = X\hat{\beta}$ are the projection (P is a projection matrix) of the data vector y onto the design plane X (in blue). The smallest errors are obtained when vector e is ortogonal to the X plane

timates, and the best linear unbiased estimates (Gauss-Markov theorem). That is, of all linear parameters estimates consisting of linear combinations of the data, whose expectation is the true value of the parameters, the least squares estimates have the minimum variance. If (X^TX) is invertible, which it is if, and only if, the design matrix X is of full rank (i.e. no regressor can be expressed as linear combination of other regressors), then the least squares estimates are:

$$\hat{\beta} = (X^TX)^{-1}X^TY \tag{B.9}$$

This equation can be derived also from the observation of a geometric perspective on the GLM, represented in **Figure B.4**. Looking at this geometric perspective, where the blue plane represents the design space defined by X, can be seen that the smallest errors (shortest error vector) are obtained when the vector e is ortogonal to the X plane. This implies a zero vector product between X and e, leading to

$$\begin{aligned} X^Te &= 0 \\ X^T(Y - X\hat{\beta}) &= 0 \\ X^TY &= X^TX\hat{\beta} \\ \hat{\beta} &= (X^TX)^{-1}X^TY \end{aligned}$$

Considering the design matrix, the relative contribution of each of its columns is assessed using standard maximum likelihood, as shown above, while inferences about these contributions are made using t or F-statistics, depending upon whether we are looking at a particular linear combination or all of them together. This will be discussed below. One aspect that needs to be mentioned before continuing with the discussion is related to the problems associated with the use of the GLM for fMRI data. First of all, BOLD responses have a delayed and dispersed form. To account for this ”shape” of BOLD responses, one could model haemodynamic responses convolving the stimulus function with a canonical hemodynamic response funtion (HRF), obataining an expected BOLD response. This step is essential for the minimisation of residuals. Beside that, the variability in the hemodynamic response, across subjects and brain regions, could optionally be compensated by modelling it with a canonical HRF plus its derivatives, with respect to time and dispersion. Another problem, is that the BOLD signal may include substantial amounts of low-frequency noise. As discussed before, this can be corrected in the pre-processing phase, by high-pass filtering in the frequency domain. In the time domain, a high-pass filter can be implemented by a discrete cosine transform (DCT) with harmonic periods up to the cut-off. These basis functions can be made explicit as confounds in the design matrix X or they can be viewed as part of a filter matrix, S, applied to both data and model. Finally, the data could be serially correlated (temporally autocorrelated) and this violates the assumptions of the noise model in the GLM. To face the problem of temporal autocorrelation, a couple of solution are commonly employed that will be briefly discussed below. One solution, known as *pre-colouring*, consists of imposing some known autocorrelation structure on the data (filtering with matrix W) and use Satterwaite correction for df’s. An alternative solution, known as *Pre-whitening*, consists in estimating the autocorrelation directly, to create a filter to ”pre-whiten” the data before fitting the GLM.

B.1.3 Statistical Inference (Contrasts)

Parametric statistical models are assumed at each voxel, using the general linear model to describe the variability of the data in terms of experimental and confounding effects and residual variability. Hypotheses expressed in terms of the model parameters are assessed at each voxel with univariate statistics. After fitting the GLM model, the estimated parameters are used to determine whether significant activation is present in a voxel or not. In fact, in functional imaging experiments, we are often interested in many sorts of effects such as the main effect of a factor or the possible interactions between factors. In order to extract the desired information, an appropriate contrast should be applied to the estimated parameters. This contrast can be either a T-test contrast or an F-test contrast depending on what the experimenter is interested in.

Inference is based on the fact that, if the design matrix X is full rank, then the estimates, $\hat{\beta}$, are normally distributed as

$$\hat{\beta} \sim \mathcal{N}(\beta,\ \sigma^2(X^T X)^{-1}) \tag{B.10}$$

Assuming independent and identical errors, the residual variance σ^2 is estimated by the residual sum of squares divided by the appropriate degrees of freedom

$$\hat{\sigma}^2 = \frac{\varepsilon^T \varepsilon}{n-p}$$

where n is the number of observations and $p = rank(X)$.
Using these results, t and F-statistics can be derived to perform tests on linear combinations of effects of interest.
From the normality of parameter estimates, $\hat{\beta}$, it follows that for a column vector c containing k weights

$$c^T \hat{\beta} \sim \mathcal{N}(c^T \beta,\ \sigma^2 c^T (X^T X)^{-1} c) \tag{B.11}$$

Furthermore, $\hat{\beta}$ and σ^2 are independent (Fisher's law). Thus, prespecified hypotheses concering linear compounds of the model parameters $c^T\beta$ can be assessed using

$$\frac{c^T \hat{\beta} - c^T \beta}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} \sim t_{n-p} \tag{B.12}$$

where t_{n-p} is a Student's t-distribution with $n-p$ degrees of freedom. For example, the hypothesis $\mathcal{H} : c^T\beta = d$ can be assessed by computing

$$T = \frac{c^T \beta - d}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} \tag{B.13}$$

and computing a p-value by comparing T with a t-distribution having $n-p$ degrees of freedom. In SPM package, all null hypotheses are of the form $c^T\beta = 0$ and tests based on this t-value are always one sided.

As described above, t-statistic follows a t distribution with a "degrees of freedom" parameter. This has been useful for hypothesis testing, allowing to test, say, the hypothesis that some explanatory variable has no effect on the dependent variable. All we need to do is to calculate a t-statistic for this null hypothesis and the data and see if that test statistic is unlikely under

the null distribution (the Student's t-distribution).
Unfortunately, when dealing with more complicated hypotheses, this test no longer works. Hypotheses involving multiple regression coefficients require a different test statistic and a different null distribution. These are: the test statistics F_0 and its null distribution, the F-distribution. Suppose we want to test the null hypothesis that a subset of the regression slopes of the GLM are zero (i.e. the corresponding regressor/effect does not explain the data). The null hypothesis then would be, for example,

$$\mathcal{H} : \beta_1 = \beta_2 = ... = \beta_i = 0$$

The key to implement a F-test is the notion of contrast matrices. A contrast matrix is a generalization of a contrast vector. Each column of a contrast matrix consists of one contrast vector. Importantly, the contrast matrix controls the partitioning of the design matrix. A user-specified contrast matrix c is used to determine a subspace of the design matrix, i.e. $X_c = Xc$. The orthogonal contrast to c is given by $c_0 = I_p - cc^-$ (this notation indicates the pseudo-inverse, a generalization of the inverse matrix for non-square matrices). Then, let $X_0 = Xc_0$ be the design matrix of the reduced model (model having some parameters imposed to zero, under the null hypothesis). We wish to compute what effects X_c explain, after first fitting the reduced model X_0. However, the subsequent fitting of two models is unnecessary, because one can construct a projection matrix from the data to the subspace of X_c, which is orthogonal to X_0. This subspace is denoted by X_a. The projection matrix M due to X_a can be derived from the residual forming matrix of the reduced model X_0. This matrix is given by $R_0 = I_n - X_0 X_0^-$. The projection matrix is then $M = R_0 - R$, where R is the residual forming matrix of the full model, i.e. $R = I_n - XX^-$. The F-statistic can then be written as

$$F = \frac{\hat{\beta}^T X^T M X \hat{\beta}}{Y^T R Y} \frac{n-p}{p_1} \sim F_{p_1, n-p} \tag{B.14}$$

where p_1 is the rank of X_a. This equation means that a F-statistic can be conveniently computed, for any user-specified contrast without any reparameterization.
In summary, the formulation of the F-statistic is a powerful tool, because by using a contrast matrix c, a subspace spanned by contrasts of the design matrix X, can be tested. Importantly, we do not need to reparameterise the model and estimate an additional parameter set, but we can use estimated parameters of the full model.

Multiple testing problem

SPM is a mass univariate method meaning that the same type of test is performed on each and every voxel in the desired mask of the brain; this leads to the so called multiple testing problem. Because many statistical tests are being conducted, adjustments have to be made to control for Type I errors (false positives) potentially caused by the comparison of levels of activity at a large number of voxels. In this case, a Type I error would result in falsely detecting background brain activity as activity related to the task. Assuming that there are N_v voxels in our brain mask, if we make inference with a significance level of α at each individual test, this would mean that the expected number of false positives is not α but $\alpha \times N_v$ which is known as the family-wise error (FWE) rate, α_{FWE}; thus, a correction for multiple comparison is in place. In representing statistical results, SPM uses a family-wise correction based on Gaussian Random Field theory [147].

With an anatomically open hypothesis (i.e. a null hypothesis that there is no effect anywhere in a specified volume of the brain), a correction for multiple dependent comparisons is necessary. The theory of random fields provides a way of adjusting the p-value that takes into account the fact that neighbouring voxels are not independent by virtue of continuity in the original data. Provided the data are smoothed, the RFT adjustment is less severe (i.e. more sensitive) than a Bonferroni correction for the number of voxels. As noted above, RFT deals with the multiple comparison problem in the context of continuous, statistical fields, in a way that is analogous to the Bonferroni procedure for families of discrete statistical tests. Let's consider the fundamental difference between an SPM and a collection of discrete t-values. When declaring a peak or cluster of the SPM to be significant, we refer collectively to all the voxels associated with that feature. The false positive rate is expressed in terms of peaks or clusters, under the null hypothesis of no activation. This is not the expected false positive rate of voxels. One false positive peak may be associated with hundreds of voxels, if the SPM is very smooth. Bonferroni correction controls the expected number of false positive voxels, whereas RTF control the expected number of false positive peaks. Because the number of peaks is always less than the number of voxels, RTF can use a lower threshold rendering it much more sensitive. In fact, the number of false positive voxels is somewhat irrelevant because it is a function of smoothness. The RTF correction discounts voxels size by expressing the search volume in terms of smoothness or resolution elements (*Resels*).

In brief, adjustments are made, based on the number of resels in the image and the theory of continuous random fields, in order to set a new criterion for statistical significance that adjusts for the problem of multiple comparisons.

Appendix C

Expectation-Maximisation (EM) algorithm

The following discussion is taken from the work of Friston and colleagues (2003) [11].

C.1 E-step

Bayesian inference is based on the conditional probability of the parameters given the data $p(\theta \mid y)$. Assuming this posterior density is approximately Gaussian (*Laplace approximation*), the problem reduces to finding its first two moments, the conditional mean $\eta_{\theta|y}$ and covariance $C_{\theta|y}$.
The i^{th} estimate of these moments will be denoted by $\eta_{\theta|y}^{(i)}$ and $C_{\theta|y}^{(i)}$.
Given the posterior density we can report its mode, i.e. the maximum a posterior (MAP) estimate of the parameters (equivalent to $\eta_{\theta|y}$), can be reported, or the probability that the parameters exceed some specific value (e.g. the probability that parameters are different from zero).
The posterior probability is proportional to the likelihood of obtaining the data, conditional to θ, times the prior probability of θ:

$$p(\theta \mid y) \propto p(y \mid \theta)\, p(\theta), \tag{C.1}$$

where the Gaussian priors are specified in terms of their expectation η_θ and covariances C_θ.

The likelihood can be approximated by expanding Eq.(2.7) about a working estimate of the conditional mean.

$$\begin{aligned} h(\theta, u) &\approx h(\eta_{\theta|y}^{(i)}) + J(\theta - \eta_{\theta|y}^{(i)}) \\ J &= \frac{\partial h(\eta_{\theta|y}^{(i)})}{\partial \theta}. \end{aligned} \tag{C.2}$$

Let $r = y - h(\eta_{\theta|y}^{(i)})$ such that $\varepsilon \approx r - J(\theta - \eta_{\theta|y}^{(i)})$.
Under Gaussian assumptions, the likelihood and prior probabilities are given by

$$\begin{aligned} p(y \mid \theta) &\propto exp\left\{-\frac{1}{2}\left(r - J(\theta - \eta_{\theta|y}^{(i)})\right)^T C_\varepsilon^{-1}\left(r - J(\theta - \eta_{\theta|y}^{(i)})\right)\right\} \\ p(\theta) &\propto exp\left\{-\frac{1}{2}(\theta - \eta_\theta)^T C_\theta^{-1}(\theta - \eta_\theta)\right\}, \end{aligned} \tag{C.3}$$

where C_ε is the error variance. Assuming the posterior density is approximately Gaussian, we can substitute Eq.(C.3) into Eq.(C.1) to give the expression of the posterior density

$$\begin{aligned} p(\theta \mid y) &\propto exp\left\{-\frac{1}{2}(\theta - \eta_{\theta|y}^{(i+1)})^T C_{\theta|y}^{-1}(\theta - \eta_{\theta|y}^{(i+1)})\right\} \\ C_{\theta|y} &= (J^T C_\varepsilon^{-1} J + C_\theta^{-1})^{-1} \\ \eta_{\theta|y}^{(i+1)} &= \eta_{\theta|y}^{(i)} + C_{\theta|y}\left(J^T C_\varepsilon^{-1} r + C_\theta^{-1}(\eta_\theta - \eta_{\theta|y}^{(i)})\right). \end{aligned} \tag{C.4}$$

Equation (C.4) can be expressed in a more compact form by augmenting the residual data vector, design matrix, and covariance components

$$\begin{aligned} C_{\theta|y} &= (\bar{J}^T \bar{C_\varepsilon}^{-1} \bar{J})^{-1} \\ \eta_{\theta|y}^{(i+1)} &= \eta_{\theta|y}^{(i)} + C_{\theta|y}(\bar{J}^T \bar{C_\varepsilon^{-1}} \bar{y}) \end{aligned} \tag{C.5}$$

where

$$\bar{y} = \begin{bmatrix} y - h(\eta_{\theta|y}^{(i)}) \\ \eta_\theta - \eta_{\theta|y}^{(i)} \end{bmatrix}, \quad \bar{J} = \begin{bmatrix} J \\ I \end{bmatrix}, \quad \bar{C_\varepsilon} = \begin{bmatrix} C_\varepsilon & 0 \\ 0 & C_\theta \end{bmatrix}.$$

The discussion done so far provides the basis for a recursive estimation of the conditional mean (and covariance) and corresponds to the *E-step* in the EM algorithm. The starting estimate of the conditional mean is generally taken to be the prior expectation. If Eq.(2.7) were linear, i.e. $h(\theta) = H\theta \implies J = H$, Eq.(C.4) would converge after a single iteration. However, when h is non-linear, J becomes a function of the conditional mean and several iterations are required. Note that in the absence of any priors, iterating Eq.(C.4) is formally identical to the Gauss-Newton method

of parameter estimation.
The conditional covariance of the parameters is assumed to be Gaussian. The validity of this assumption depends on the rate of convergence of the Taylor expansion of h in Eq.(C.2).
Because h is non-linear the likelihood density will be only approximately Gaussian.
However, the posterior or conditional density will be almost Gaussian, given a sufficiently long time series [148].

C.2 M-step

So far the error covariance C_ε has been assumed to be known. Clearly in many situations (e.g., serial correlations in fMRI) it is not. When the error covariance is unknown, it can be estimated through some hyperparameters λ_j, where $C_\varepsilon = \sum \lambda_j Q_j$.
$Q_j = \partial C_\varepsilon / \partial \lambda_j$ represents a basis set that embodies the form of the variance components and could model different variances for different blocks of data or indeed different forms of serial correlations within blocks.
Restricted Maximum likelihood (ReML) estimators of λ_j maximise the (marginal) log likelihood $log\ p(y \mid \lambda) = F(\lambda)$.
This log likelihood obtains by integrating over the conditional distribution of the parameters as described in Neal and Hinton (1998) [149]. Under a Fisher-scoring scheme [150] this gives

$$\begin{aligned} \lambda^{i+1} &= \lambda^i - \left\langle \frac{\partial^2 F}{\partial \lambda^2} \right\rangle^{-1} \frac{\partial F}{\partial \lambda} \\ \frac{\partial F}{\partial \lambda_j} &= -\frac{1}{2} tr\{PQ_i\} + \frac{1}{2} \bar{y}^T P^T Q_i P \bar{y} \\ \left\langle \frac{\partial^2 F}{\partial \lambda_{jk}^2} \right\rangle &= -\frac{1}{2} tr\{PQ_i P Q_j\} \\ P &= \bar{C_\varepsilon}^{-1} - \bar{C_\varepsilon}^{-1} \bar{J} C_{\theta|y} \bar{J}^T \bar{C_\varepsilon}^{-1} \end{aligned} \tag{C.6}$$

The above equations are quick to implement due to the sparsity structure of the covariance basis set Q_i. If the basis set is the identity matrix, embodying *i.i.d* assumption about the errors, then this is equivalent to the sum of squared residuals estimation used in classical analysis of variance.

To sum up, this recursive optimisation scheme, also known as Variational Laplace (VL), updates the moments of the posterior density, $q(\theta \mid y, m)$ by maximising the negative variational Free Energy, which provides a lower bound on the log model evidence, $log\ p(y \mid m)$, Beal and Ghahramani

(2003), being the Kullback-Leibler divergence between the approximate and true posterior, always positive or at least zero.