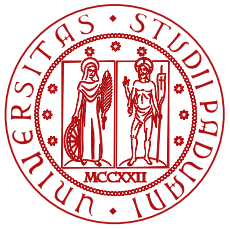


Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Master of Science in ICT for Internet and Multimedia

Development of machine learning-based regression models to predict inter-individual differences from MEG-based resting-state functional connectivity

Master Candidate

Enrico Fongaro

Student ID 2017403

Supervisor

Prof. Giulia Cisotto

Co-supervisors

Prof. Viviana Betti

Dr. Ottavia Maddaluno

Academic Year 2022/2023

April, 3rd 2023

Abstract

Recent research has demonstrated that functional connectivity profiles act as a unique fingerprint that can accurately identify subjects from a large group. As a result, brain connectomics has emerged as a rapidly growing research field that focuses on identifying individuals based on inter-individual variability in brain connectivity during resting-state and task-evoked responses. The main objective of this study, which is part of the ERC *HANDmade* project (SH4, ERC-2017-STG) led by Prof. Betti Viviana, is to investigate and predict inter-individual differences in motor tasks based on resting-state functional connectivity measured with magnetoencephalography (MEG). Several regression-based models are discussed and compared to improve the identification rate on 51 subjects from the Human Connectome Project. A residualisation approach is also presented to enhance model performance by increasing variability across subjects. Results indicate that inter-individual differences in brain connectomes during spontaneous and task-evoked activity can be accurately predicted. However, it is observed that no improvement can be obtained in the identification rate considering the muscle activity, compared to Vettoruzzo's MSc thesis. Regardless of the model complexity and the number of the EMG features considered, the results show the inadequacy of predicting muscular activity from individual brain connectivity. Therefore, the investigation is addressed to understand the relationship between the brain activity and muscular activity, and non-linear approaches employing neural networks are considered the best choices for this type of motor tasks. The study's findings can have implications in clinical research and rehabilitation fields for comprehending the neural mechanism underlying various neurological and muscular disorders and for developing personalized treatment approaches.

Sommario

Recenti ricerche hanno dimostrato che i profili di connettività funzionale agiscono come una impronta digitale univoca in grado di identificare con precisione i soggetti di un gruppo. Di conseguenza, la connettomica, lo studio delle connessioni tra aree diverse del cervello umano, è emersa come un campo di ricerca in rapida crescita che si concentra sull'identificazione della variabilità interindividuale nella connettività cerebrale, durante lo stato di riposo e le risposte evocate dal task. L'obiettivo principale di questo studio, che fa parte del progetto ERC *HANDmade* (SH4, ERC-2017-STG) coordinato dalla Prof.ssa Betti, è indagare e predire le differenze interindividuali nei task motori a partire dalla connettività funzionale spontanea, calcolata a partire da dati acquisiti tramite magnetoencefalografia (MEG). Diversi modelli basati sulla regressione vengono discussi e confrontati per migliorare il tasso di identificazione su 51 soggetti del Human Connectome Project. Viene inoltre presentata una tecnica di residualizzazione per migliorare le prestazioni dei modelli aumentando la variabilità interindividuale nelle attività motorie. I risultati indicano che le differenze interindividuali nei connettomi cerebrali durante l'attività spontanea ed evocata dal task motorio possono essere predette con precisione. Tuttavia, si osserva che non si ottiene alcun miglioramento, rispetto ai risultati della tesi di laurea magistrale di Vettoruzzo, nel tasso di identificazione considerando l'attività muscolare. Indipendentemente dalla complessità del modello e dal numero delle features EMG considerate, i risultati mostrano l'inadeguatezza della previsione dell'attività muscolare a partire dalla connettività cerebrale del soggetto. Pertanto, lo studio si è indirizzato a comprendere la relazione tra l'attività cerebrale e l'attività muscolare, e gli approcci non lineari che impiegano reti neurali sono considerati le scelte migliori per questo tipo di task motori. I risultati dello studio possono avere implicazioni nei campi della ricerca clinica e della riabilitazione per comprendere il meccanismo neurale alla base di vari disturbi neurologici e muscolari e per sviluppare approcci terapeutici personalizzati.

Contents

List of Figures

List of Tables

List of Acronyms

AN Auditory Network

ANN Artificial Neural Network

BLP Band Limited Power

BOLD Blood Oxygen Level-Dependant

CNS Central Nervous System

CONT Control Network

DAN Dorsal Attentional Network

DMN Default Mode Network

ECG Electrocardiography

EEG Electroencephalography

EMG Electromyography

EOG Electro-oculography

FC Functional Connectivity

FPN Fronto-Parietal Network

fMRI Functional magnetic resonance imaging

HCP Human Connectome Project

ICA Independent Component Analysis

LF Left Foot

LH Left Hand

LOOCV Leave-One-Out Cross Validation

LN Language Network

MEG Magnetoencephalography

MA Moving Average

ML Machine Learning

MLR Multiple Linear Regression

MN Motor Network

MNE Minimum-Norm Estimation

MSE Mean Squared Error

MSR Multilayered Shielding Room

MUAP Motor Unit Action Potential

NN Neural Network

PSP Postsynaptic Potential

ReLU Rectified Linear Unit

ROI Region of Interest

RF Right Foot

RH Right Hand

SQUID Superconducting QUantum Interference Device

TKEO Teager-Kaiser Energy Operator

VAN Ventral Attention Network

VFN Visual Foveal Network

VPN Visual Peripherical Network

1

Introduction

Neuroscience researchers have long been trying to understand what causes individuals to be different from one another. Thanks to the recent advancements in neuroimaging technology and methodologies, high-quality measures of the brain can now characterize an individual's personality traits and behaviour [1]. One of the most informative ways to understand individual differences in the brain is by investigating its functional connectivity (FC), which is obtained by analyzing the temporal statistical dependencies between the activities of different brain regions [2]. By modeling functional connectivity using network science, an area of research usually referred to as brain connectomics, researchers can gain a better understanding into the underlying neural mechanisms that give rise to individual differences in these processes.

The concept of *brain fingerprints* has become prominent following the influential research conducted by Finn and colleagues (2015) [3]. This research demonstrated that an individual's functional connectome profile is both distinctive and reliable, similar to a fingerprint. This discovery has encouraged neuroimaging investigations to concentrate on results derived on individual subjects, by exploiting unique properties of functional network organization across cognitive tasks and resting state [4]. The focus on individual subjects has led to a rapidly evolving research field that aims to predict inter-individual differences in behaviour using personalized approaches to describe brain organization at the subject level. One such approach, proposed by Tavor et al. (2016) [5], relies on fMRI data to identify individuals from large study populations. The method uses a machine learning-based technique to predict individual differences in task-evoked brain responses from spontaneous functional connectivity measurements taken during rest. The model highlights a close relationship between brain connectivity and tasks (includ-

ing movements) that can be captured at the level of individual subjects. However, most research in this field focuses on fMRI data, at slower time intervals, and the same questions can be addressed to MEG data, at higher temporal resolution [6]. Furthermore, exploring the correlation between brain connectivity and muscle activation during motor tasks may provide valuable insights into the neural mechanisms underlying individual differences in motor function.

The study presented in this MSc thesis is part of the ERC *HANDmade* project (StG, SH4, ERC-2017-STG; start date: 2018-02-01, end date: 2023-07-31) [7], led by Prof. Viviana Betti at Sapienza University of Rome. The project's main objective is to investigate the hypothesis that the intrinsic brain activity creates and maintains an internal model of the environment by integrating information from visual and sensory inputs, with a specific focus on the hand's role in this process [8]. The thesis work aims to address the questions regarding the correlation between brain connectivity during rest and muscle activation, employing MEG data in a cohort of 51 subjects from the Human Connectome Project (HCP) [9]. Two different motor tasks, i.e. finger tapping and toe squeezing, are used to explore the extent of inter-individual variability in brain activation, and to focus on whether differences in FC acquired during rest state can predict individual variations in motor task-evoked brain activity. To achieve this, the study employs several machine learning regression-based models, including neural networks, which utilize resting state FC as features to predict individual task-based functional activity, and identify subjects. Model training is performed by mapping FC maps from resting state to task condition, and subsequently, the trained models are applied to predict unseen subjects using a Leave-One-Out Cross Validation (LOOCV) approach. The study also examines the connection between brain connectivity and electromyography (EMG) features that are extracted from muscle activation, conducting a similar study on predicting muscle activity from brain connectivity at rest. These investigations aim to offer insights into how the neural mechanisms underlie individual variations in motor function. To summarize, this thesis seeks to systematically integrate the findings of A. Vettoruzzo and J. Fagotto's previous MSc thesis works [10, 11], supervised by Prof. Giulia Cisotto, with particular focus on improving the performance of the proposed models and exploring the inter-individual variability of FC extracted from MEG data. To address this, the thesis proposes a residualisation approach to enhance variability across individuals, by decomposing individual FC data in two independent components, namely the group-averaged and residual components.

The thesis is organized as follows:

Chapter 2 - Background This chapter provides the basic concepts of brain and muscle anatomy and physiology. It covers the generation of the brain activity, as well as how it can be acquired using MEG instrumentation. Additionally, it presents how muscle activity can be measured via EMG and concludes with an introduction to the mathematical tools used in the thesis, including linear regression models and Leave-One-Out approach.

Chapter 3 - Materials and methods In this chapter, data and techniques used in this study are described. The HCP is introduced as the source of the data. The preprocessing steps, including degree-normalization [12] and the method for active EMG detection [11], are explained to clean the MEG and EMG signals. The regression models used in this study are then discussed, with a focus on predicting inter-individual differences in motor tasks using rMEG (rest) and tMEG (task) functional connectomes, as well as EMG features and rMEG FC matrices. The chapter also highlights the importance of the residualisation step in the prediction process. Additionally, the study explores the relationship between brain connectivity at rest and muscular activity, and discusses the regression models used for this purpose.

Chapter 4 - Results This chapter presents the findings of the study. It begins with the functional connectivity analysis and explores the linear transformation of connectomes from resting state to task condition, including also the link-by-link transformation. It follows with the performance comparison of the models employed in the study, focusing on original and residualisation-based models. The study also assesses the inadequate model performance in predicting muscle activity from rMEG functional connectivity using LOOCV-based models. Finally, the accuracy of the regression models, examining the connection between FC at rest and muscular activity, is evaluated.

Chapter 5 - Discussion In this chapter, the inter-individual variability of brain connectivity and network analysis are discussed. Model performance, specifically linear regression models and neural network models, are evaluated. The limitations of the study are also mentioned.

Chapter 6 - Conclusions This final chapter provides a brief summary of the research conducted in the thesis and the key findings, by also discussing the possible future works for further investigations in this area and potential applications.

Chapter A - Appendix The Appendix contains additional and supporting material to the main content of the thesis providing further details and clarification.

2

Background

2.1 Brain anatomy and physiology

The human brain is the most complex and fascinating organ of the human body. It exerts centralized control over body's organs, from vital functions up to the most sophisticated movements, and interprets information from the outside world, allowing rapid and coordinated responses to changes in the environment. The brain is located inside the bony covering called the *cranium*, which protects it from injury, and, together with the *spinal cord*, composes the *central nervous system* (CNS). Three gross divisions characterize the brain anatomy: the *cerebrum*, the *brain stem* and the *cerebellum*. The cerebrum is the rounded gray section of the brain responsible for processing sensory stimuli, including sight, touch, and hearing. In addition to these basic functions, it also plays a critical role in higher-order cognitive processes such as memory and attention. The cerebellum is a small part of the brain, located at the bottom of the cerebrum near the back of the head, whose function is to coordinate muscle movements, maintain posture, and balance. It also plays an important role in other cognitive functions such as language processing and memory. The brainstem acts as a relay center connecting the cerebrum and cerebellum to the spinal cord. It performs many automatic functions such as breathing, heart rate and body temperature regulation, wake and sleep cycles. The *cerebral cortex* is the outermost layer of the cerebrum. Its surface has many folds, giving it a wrinkled appearance. The folds consist of many deep grooves called *sulci* and raised areas called *gyri* which enable large amounts of information to be processed by more nerve cells. The cerebrum is also divided in two halves, the right and left hemispheres, joined by a bundle of nerve fibers called the *corpus callosum*

that transmits messages from one side to the other. Each hemisphere controls the opposite side of the body and not all functions of the hemispheres are shared between them. In general, the left hemisphere of the brain is responsible for language and speech, and is called the *dominant* hemisphere. The right hemisphere plays an important role in interpreting visual information and spatial processing. The two halves have distinct fissures, which divide each hemisphere into four lobes: frontal, temporal, parietal and occipital. The frontal lobe is located in the forward part of the brain and it is involved in voluntary movements (motor strip), expressive language and for managing high level cognitive functions, e.g. concentration and problem solving. The parietal lobe, positioned behind the frontal one, is vital for sensory perception and integration, including the management of taste, hearing, sight, touch and smell. The temporal lobe, inferior to the parietal lobe and in the lateral part of the head, is associated to encoding of memory as well as receiving and processing auditory information. Last, the occipital lobe sits at the back of the head and is responsible for visual perception, including colour, form and motion [13, 14].

Neuroglia and *neurons* represent the two main types of cells in the nervous system. The function of neuroglia cells is to ensure structural support, nourishment and neuron protection. Neurons are highly specialized units for the processing and transmission of cellular signals. They have important roles in reacting to chemical and sensory stimuli, conducting the impulses and emitting specific chemical regulators. On average, the human brain contains about 100 billion neurons and many more neuroglia [15]. The *soma*, defined as the cell body of the neuron, contains the *nucleus*, from where most protein synthesis occurs. The *dentrites* of a neuron are cellular extensions with many branches, coming off of the cell body, that receive the signals from sensory receptors and other neurons. The *axon* is a finer, cable-like projection which can extend even ten of thousands of times the diameter of the soma in length. Finally, *synapses* are junctions that allow information to be transferred from one neuron to another. Each neuron may be connected to up to 10,000 other neighbours, passing signals to each other via as many as 1,000 trillion synapses [15].

2.1.1 The nerve impulse

The presynaptic neuron stores neurotransmitters in vesicles at the axon terminal. When the impulse reaches the axon terminal, it triggers a release of neu-

rotransmitters which diffuse through the synaptic cleft and bind to the receptors in the dendrites of the postsynaptic neuron. As a result, the receptors open up ion channels through the cell membrane which allow specific ions to pass. The opening of the ion channels will alter the transmembrane potential of the postsynaptic cell from its equilibrium (-70 mV). When the ion channels open, the ions are able to move along their concentration gradients so that the transmembrane voltage changes locally, i.e., a postsynaptic potential (PSP) is generated. If the resulting potential is less/more negative than the equilibrium potential, PSP is called excitatory/inhibitory. The local summation of excitatory/inhibitory PSPs at the root of the axon, or axon hillock, may reach around -55 mV. This condition allows the neuron to fire an *action potential* (AP), a threshold phenomena which consists of depolarizing front followed by repolarization. The AP burst propagates fast along the axon until it reaches the axon terminal and triggers the release of neurotransmitters [16, 17]. This moving change in membrane potential has three phases, which occur over just a few milliseconds: depolarization, followed by repolarization and a short period of hyperpolarization.

- **Depolarization**, also called the rising phase, is caused when positively charged sodium ions (Na+) suddenly rush through open voltage-gated sodium channels into a neuron. As additional sodium rushes in, the membrane potential actually reverses its polarity developing a positive value for a moment ($+40$ mV).

- **Repolarization**, or falling phase, is caused by the slow closing sodium channels and the opening of voltage-gated potassium channels. As a result, the membrane permeability to sodium declines to resting levels and potassium ions rush out of the cell. This expulsion acts to restore the localized negative membrane potential of the cell to its equilibrium.

- **Hyperpolarization** is the final phase where some potassium channels remain open and sodium channels reset. A period of increased potassium permeability, resulting in excessive potassium efflux before the potassium channels close, is seen as a slight potential dip following the spike [18].

2.2 Magnetoencephalography

Several non-invasive techniques provide researchers with valuable information about brain functions and can be used to diagnose and monitor certain neurological disorders. *Functional magnetic resonance imaging* (fMRI), based on blood oxygen level-dependant (BOLD) image contrast, monitors the haemodynamics in the brain. It can achieve almost sub-millimetre accuracy in the spatial localisation of neuronal activity. However, this relatively high spatial resolution is not matched in temporal accuracy because the slow speed of haemodynamic changes limits the timescale to seconds. In *electroencephalography* (EEG), the electric potential generated by the synchronized activity of neurons, mainly from the pyramidal cells, is recorded by a certain number of electrodes positioned on subject's scalp. EEG is extensively used in psychophysiological research and in clinical application. However, one limitation of EEG concerns the distorting effects of the intermediate structures, which seriously hinder a precise localization of the signal source. Overall, electroencephalography lacks spatial detail as the potential is smoothed by the different tissues of the head but provides high temporal resolution in the millisecond range [19, 17].

Magnetoencephalography (MEG) is a non-invasive neuroimaging technique that provides an adequate understanding of the function of the brain because of its ability to localise sources of electric and magnetic activity with an excellent temporal and ever-improving spatial resolution [16]. The typical MEG timescale is better than 1 ms and the spatial discrimination is, under favorable circumstances, $2-3$ mm for sources in the cerebral cortex [20]. Unlike EEG, the accuracy of MEG source mapping is immune to the signal distortions caused by complex layering of head tissues.

2.2.1 MEG signal generation

MEG allows to estimate the sources of intracranial currents by measuring the neuromagnetic fields outside the head. The electrical signals from a group of firing neurons constitute a small, but detectable, electrical current across the scalp which produces magnetic induction, whose strength can be measured remote from the current source, e.g. with a pick-up coil. The magnetic flux across the coil surface induces an electrical current in the coil wiring material, whose amplitude is instantaneously proportional to that of the magnetic induction and is readily

measurable. In MEG, the sources of magnetic induction that can picked up outside the head are the postsynaptic potentials (PSPs), slow ($\sim$ 10 ms) monophasic potential deflections, and action potential (APs), faster and biphasic axon discharges ($\sim$ 2 ms) [10]. The slower components of the PSPs are substantially smaller in amplitude than the APs. However, at the scale of cell assemblies, the mass effect of slower PSPs is stronger than that of APs owing to their greater overlap in time without requiring rigorous synchronization. The AP spikes, instead, are not perfectly synchronized and their summation cancels each other out. Therefore, the neuronal activity captured by MEG is not produced by the axonal action potential of pyramidal cells, but rather by the net contributions of excitatory and inhibitory dendritic postsynaptic potentials [19, 10]. However, the position and the geometric configuration of the pyramidal cells, including the shape and orientations of their dendrites, can affect the strength and synchrony of their PSPs. Particularly, MEG is most sensitive to the activity of neurons located in the fissures or sulci of the cerebral cortex, whose tangential orientation with respect to the skull produces strong magnetic fields which project radially. Radial sources are thus externally silent. MEG is also less sensitive to deeper (including subcortical) sources as magnetic field change decreases rapidly as $1/r^3$ with the distance of r. As a consequence, when the spatial and temporal summation of a sufficient number of neurons, about 10^6, along the cortex is activated in relative synchrony, the overall change in the magnetic field is strong enough for detection by highly sensitive MEG sensors.

2.2.2 MEG instrumentation

In practical terms, the extracranial magnetic induction produced by neural currents (10^{-9} A) is extremely weak, typically measured on a scale of femtoteslas (10^{-15} T) [21]. This reality imposes the need for sensitive sensor technology and requires a significant reduction in interference from environmental and other sources of disturbance. The present industry standards rely on pick-up coils coupled with *Superconducting QUantum Interference Devices* (SQUIDs) [20, 16]. SQUIDs operate by exploiting the quantum mechanical phenomena of superconductivity (zero resistance at very low temperatures) for the detection of small electrical currents. Cryogenic temperatures (4 K or -269 °C) serve to minimize thermal noise and therefore optimize data quality. Additionally, sensor arrays have evolved to provide whole-head coverage via a helmet containing more than 300 independent

sensor sites, sampled at up to 30 kHz simultaneously and enveloped in a thermally insulated dewar which is filled with $70-100$ L of cooling liquid helium (LHe). The subject's head is positioned underneath the helmet as close as possible to the sensors. Essentially, a SQUID sensor consists of a superconducting loop interrupted by one or two weak links, or Josephson junctions, made of two superconducting layers separated by a thin insulating layer which limits the flow of supercurrents. Alterations of the magnetic flux result in voltage changes across the junctions of the SQUID circuitry, which can be detected and used to determine the strength of the magnetic field. Since SQUIDs are very sensitive to variations in magnetic field, the magnetic flux is fed into it by means of a transformer, which shields the SQUIDs from unwanted noise. This consists of a pick-up coil, which coupled with the SQUIDs, forms an entirely superconducting loop called the *flux transformer*. In general, the extreme sensitivity of MEG sensing technology (field sensitivities of a few $\mathrm{fT}/\sqrt{\mathrm{Hz}}$ have been reported [20]) is challenged by many electromagnetic noisy sources. Any moving metal object or electrically powered instrument generates magnetic induction that is orders of magnitude stronger than the brain's. Their influence can be reduced by combining magnetic shielding and *gradiometer* devices to emphasize brain signals with respect to background noise. Such gradiometers, particularly sensitive to gradients in magnetism from neuronal sources and arranged either in a radial or planar fashion with respect to the head surface, greatly improve spatial resolution of MEG signal but still require magnetic shielding from the environment. Therefore, the installation of a heavy (20-ton) multilayered shielding room (MSR) to host the MEG instrument represents the best solution to remove the overwhelming disturbances and magnetic noise caused by environmental factors. Inside it, one or more layers of a high permeability alloy, such as μ-metal, and high conductivity metal, e.g. aluminium, provide effective ferromagnetic shielding against low and high frequencies of urban magnetic noise. Considerable care must also be taken in ensuring that all stimulus presentation and response devices within the magnetically shielded room are themselves electromagnetically silent. Experimental task are typically designed to avoid eye movements during relevant measurement periods in order to minimize artefacts, but remaining eye movements and blinks can still be identified with electro-oculography (EOG) [19]. An ECG is typically recorded to enable identification and subsequent exclusion of cardiac electrical activity, which can otherwise create large contaminating artefacts in the MEG signal. Other physiological signal, such as EMG, can be recorded during the MEG measurement to detect the electrical activity of muscles

as they contract or relax.

2.2.3 Brain rhythms

Oscillations in the brain, also called *brain waves* or *brain rhythms*, are produced by coordinated electrophysiological activity in large groups of neurons in both space and time. One important indicator to assess brain waves is frequency. Frequency allows to assess abnormalities in brain activity and to understand functional behaviours in cognitive research. The useful frequency band of MEG signals is about $0.5 - 1000$ Hz, with $0.5 - 80$ Hz being the most typical range [21]. It is a common practice to divide the latter interval into five brain rhythms, each reflecting different mental states and functions:

- **Delta** δ $0.5 - 4$ Hz: delta waves are slow, deeply penetrating, loud brainwaves characterized by low frequency and high amplitudes. They are generated in deepest meditation and dreamless sleep, as healing and regeneration are stimulated in this state. Suppression of delta waves leads to inability of brain revitalization and poor sleep [22], whereas an excess may result in learning disabilities, e.g. ADHD (*Attention-Deficit Hyperactivity Disorder*), or in severe neurological disorders, such as dementia and schizophrenia.

- **Theta** θ $4 - 8$ Hz: theta waves underlie various aspects of cognition and behaviour, including learning, memory consolidation and sleep. They refer to states of deep relaxation, meditation and drowsiness but they may be involved in the regulation and the processing of emotional information [23].

- **Alpha** α $8 - 13$ Hz: alpha waves are neuronal oscillations dominating the moments of relaxation and calm alertness. They are predominantly recorded from the occipital lobes during wakeful relaxation with closed eyes and play an important role in the suppression of irrelevant information and the processing of task-relevant information [24].

- **Beta** β $13 - 35$ Hz: beta brainwaves are active when attention is directed towards cognitive tasks and the outside world. They deal with problem-solving, judgement, decision making, or focused mental activity. Beta rhythms are further divided into two bands. Low beta (β_{low}) can be thought as a 'fast idle', or musing, while high beta (β_{high}) is associated to highly complex thought, integrating new experiences, high anxiety, or excitement [24].

- **Gamma** γ $35-80$ Hz: gamma rhythms are correlated with large brain network activity and cognitive phenomena including working memory, attention and awareness, as well as processing complex information and heightened perception. Altered gamma activity has been observed in many cognitive disorders such as Alzheimer's disease, epilepsy and schizophrenia [25].

2.2.4 Brain networks

Information processing in the cerebral cortex involves interactions among distributed areas. A defining aspect of the organization of brain areas is its spatial heterogeneity, namely the variation in the functional and structural properties of different brain regions, which give rise to multiple topographies at different scales. Brain parcellation is thus fundamental for understanding brain organization and function, as well as being a data reduction strategy [26].
Within the formal framework of graph theory, a *brain network*, or brain graph, refers to the interconnected neural structures within the brain (also referred to the nodes of the network) that work together to perform specific functions and provide a comprehensive description of the brain's structural connectivity, also called *human connectome* [27, 28]. The connectome essentially comprises a complete map of the brain's anatomical connections, according to different spatial levels. These structural connections shape large-scale neuronal dynamics which can be captured as patterns of *functional connectivity* (FC) during spontaneous brain activity (i.e., brain activity that occurs in absence of any external input or stimulation [29]) as well as in the course of stimulus or task-evoked perturbations. Functional connectivity describes statistical patterns of dynamic interactions among regions, also called functional networks. Its profile is both unique and reliable and acts as *fingerprint* that can accurately identify the subject from a large group of individuals [30]. Functional brain connectivity is generally derived from time series observations, extracted with a variety of techniques, including EEG, MEG and fMRI, and can be computed in multiple ways, e.g. through cross-correlation. While the presence of a statistical relationship between two neural elements is often considered as a sign of functional coupling, it must be noted that this does not imply a causal relationship [31]. Once all functional couplings are estimated, the full set of all pairwise couplings can then be aggregated into a connection matrix, the so called *functional connectivity matrix*, or FC matrix. Creating maps of structural and functional connections can be perfromed at different resolutions and with different

criteria. As a result, there is no a unique standard parcellation scheme, or *atlas*, which selects the Regions of Interest (ROIs) for further analysis. Depending on each employed atlas, the number of nodes as well as links of each network can vary.
In this study, the organization of networks in the human cerebrum was explored using resting-state MEG-based functional connectivity. A parcellation method is employed, consisting in a whole brain network based on 164 nodes divided into 10 different networks (Figure A.1 and A.2 in the Appendix) each reflecting specific properties:

- **Auditory Network (AN)** plays an important role in different aspects of processing auditory information, including pitch, and location, and in interaction with the environment. Multiple cortical areas, such as the inferior frontal gyrus, superior temporal gyrus have been implicated in this processing [32].

- **Control Network (CONT)** is primarily involved in cognitive control, decision making, and attention. This network includes the dorsal lateral and medial prefrontal cortex, the anterior cingulate cortex, and the superior parietal lobule [33].

- **Dorsal Attentional Network (DAN)** consists of regions such as the dorsal parietal cortex and the superior temporal sulcus, and is concerned in attention and spatial cognition [34].

- **Default Mode Network (DMN)** regions exhibit deactivation during a wide variety of resource demanding tasks. DMN is thought to be involved in a wide range of processes such as self-referential thinking, episodic memory and mind-wandering and comprises the posterior cingulate cortex, the medial prefrontal cortex, and the inferior temporal gyrus [35].

- **Fronto-Parietal Network (FPN)** is primarily involved in attention, working memory and cognitive control, such as problem-solving, planning and decision-making. It includes the prefrontal cortex and the parietal cortex [36].

- **Language Network (LN)** includes Broca's area, concerned in language production, and Wernicke's area, involved in language comprehension. The language network includes also other regions such as the superior temporal

gyrus (processsing auditory information) and the angular gyrus, involved in semantic processing [37].

- **Motor Network (MN)** comprises the primary motor cortex, the premotor cortex, and the supplementary motor area. It is responsible for the direct control, the execution and the coordination of body movements [38].

- **Ventral Attention Network (VAN)** is a functional brain network primarily related to attention and spatial processing. It is also thought to be involved in stimulus evaluation, such as emotion regulation, and comprises the inferior parietal lobule, the temporo-parietal junction, and the ventral prefrontal cortex [39].

- **Visual Foveal Network (VFN)** is engaged in tasks that involves the processing of high-resolution visual information in the center of the visual field, also known as the fovea. It is composed of several brain regions, including the primary visual cortex, the middle temporal area and the lateral occipital complex [40].

- **Visual Peripherical Network (VPN)** is responsible in the processing of low-resolution visual information in the periphery of the visual field, such as the recognition of scenes and places. The lateral occipital complex, the parahippocampal place area and retrospinal cortex are engaged in these tasks [41].

2.3 Human neuromuscular system

The human body is a well-designed system whose form, support, stability, and movements are provided by the *musculoskeletal system*, also called locomotor system. The term musculoskeletal refers to two broad systems [42]:

- *Muscular system*, which contains all types of muscles in the body, including the tendons which attach the muscles to the bones. Skeletal muscles, in particular, are the ones that act on the body points to produce movements.

- *Skeletal system*, whose main component is the bone. Bones articulate with each other and form the joints, providing our bodies with a hard-core, yet mobile, skeleton, supported by articular cartilage and ligaments.

Additionally, neurons and muscles compose the neuromuscular system, which refers to the complex network of nerves and tissues that controls the tension and length of the muscle, to generate the required movements [43]. The neuromuscular system includes the *motor neurons*, the smallest functional units responsible for transmitting signals from the brain to the muscles fibers of the desired effector. The interface between the motor neurons and the muscle fibers is called the *neuromuscular junction*, where the electrical signal from the motor neuron triggers the release of neurotransmitters which leads the depolarization of the muscle fibers, causing them to contract. The strength and the duration of the muscle contraction is determined by the number of motor units that are activated and the frequency of nerve impulses sent from the brain. The resulting electrical activity of the muscle fibers during a contraction generates the electromyography (EMG) signal.

2.3.1 Electromyography signal

The EMG signal is the electrical manifestation of the neuromuscular activity associated with a contracting muscle. Under normal conditions, an electrical impulse propagating down a motor neuron activates all the muscle fibers attached to its branches, depolarizing the postsynaptic membranes. This depolarization, accompanied by a movement of ions, generates an electrical field which produces brief distinguishable bursts of electrical potential, known as *muscle fiber action potentials* [44]. Thus, the overall signal detected at the sensor site constitutes a spatial-temporal superposition of the individual action potentials, namely the *motor unit action potential* (MUAP). The motor unit action potential can be detected by a skin surface electrode (surface electromyography, sEMG), or by inserting a fine wire (or needle) electrode into the muscle tissue to measure the electrical activity directly [45]. The former, being non-invasive and relatively inexpensive, makes it the preferred method for many clinical and research applications but is susceptible to interference from electrical noise. The second technique is more invasive than sEMG, but provides a much more accurate and reliable measurement of muscle electrical activity. The shape and the amplitude of the MUAP are affected by the geometric arrangement of the active muscle fibers with respect to the electrode site as well as several external factors, including tissue characteristics, skin impedance, and external noise, can alter its characteristics. In general, the recruitment of MUAP and the firing frequency are the main control strategies to adjust the contraction process and modulate the force output of the involved

muscle [46]. In normal muscles, the peak-to-peak amplitude of a MUAP detected with indwelling electrodes (needle or wire) may range from a few microvolts to 5 mV, depending on muscle fibers' force of contraction, with a typical value of 500 μV [44]. During muscle relaxation, the averaged baseline noise is not higher than $3 - 5$ μV, even if several sources of noise affect the signal, including power line interference, ambient noise, motion and other biological activity artifacts [45]. The frequency content is typically between 6 Hz and 500 Hz, showing most frequency power between 20 and 150 Hz [46].

2.4 Mathematical tools

2.4.1 Linear regression

In statistics, linear regression, also referred to General Linear Model (GLM) [47], is a linear approach for modelling the relationship between a scalar response (target) and one or more explanatory variables (predictors). This statistical method is widely used in machine learning (ML), specifically, it can be considered as a baseline model of supervised learning algorithm performing a regression task. This technique is used for forecasting future outcomes, time series modelling and finding cause-effect relationship between variables. Given a data set $\{y_k, x_{k1}, \ldots, x_{kp}, \}_{k=1}^{K}$ of K statistical units, a linear regression model assumes that the relationship between the dependant variable y and the $p-$dimensional vector of regressors $\mathbf{x}$ is linear and modeled through an error variable ϵ, also referred to the residual. Thus the model takes the form

$$y_k = \beta_0 + \beta_1 x_{k1} + \cdots + \beta_p x_{kp} + \epsilon_k = \mathbf{x}_k^\top \boldsymbol{\beta} + \epsilon_k \qquad k = 1, \ldots, K \tag{2.1}$$

where $^\top$ denotes the transpose, so that $\mathbf{x}_k^\top \boldsymbol{\beta}$ is the inner product between vectors $\mathbf{x}_k$ and $\boldsymbol{\beta}$. In matrix notation, the set of K equations can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.2}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_K \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_K^\top \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{K1} & \dots & x_{Kp} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_K \end{bmatrix} \tag{2.3}$$

The parameter $\boldsymbol{\beta}$ is a $(p+1)-$dimensional vector whose elements are known as regression coefficients and, specifically, β_0 is the intercept term. In simple linear regression, $p = 1$, and the regression coefficient is known as slope. On the other hand, in multiple linear regression, the response variable can be predicted through a linear combination of explanatory variables and parameters ($p > 1$). Fitting a linear model to a given dataset usually requires estimating the regression coefficients $\boldsymbol{\beta}$ such that the sum of squared errors $\|\boldsymbol{\epsilon}\|_2^2$, with $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, is minimized. This approach is called the *least squares* method and, specifically, its solution for the multiple linear regression model can be written as:

$$\boldsymbol{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \tag{2.4}$$

This formula finds the values of the $\boldsymbol{\beta}$ parameters that minimizes the sum of the squared errors between the predicted values, lying on the estimated line, and the actual values [48].

2.4.2 Pearson's correlation coefficient

In statistics, the Pearson correlation coefficient is a measure of the strength and direction of the linear relationship between two sets of data. Formally, it is defined as the ratio between the covariance of two variables, X and Y, and the product of their standard deviations, as follows:

$$r_{XY} = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{2.5}$$

where $\operatorname{cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$ with μ_X and μ_Y the mean of X and Y respectively [49]. Pearson correlation r_{XY} ranges from -1 to 1, $r_{XY} \in [-1, 1]$. A value of -1, or 1, indicates a perfect negative, or positive, correlation, i.e. as one variable increases, the other variable decreases, or increases, linearly. A value of 0 refers to no correlation. Typically, a hypotheses test is performed to verify the

evidence against the null hypotheses, which states that there is no relationship between two compared variables. The p-value reflects the probability that the correlation between the two variables occurred only by chance. If the p-value is less than a pre-determined level of significance (such as $\alpha = 0.05$), the null hypotheses is rejected and there is evidence of a significant correlation between the two variables. Conversely, if the p-value is greater than the level of significance, there is not enough evidence to reject the null hypotheses and to support the claim that there is a significant correlation.

2.4.3 Leave One Out Cross Validation

In neuroscience research, great interest is placed on building predictive models to understand how the brain process information and generates movements. These models can be used to make predictions about future behaviours, to simulate the effects of perturbations on brain function, and to test hypotheses about the underlying neural mechanism. In this project thesis, the aim is to develop predictive models of inter-individual differences during motor tasks based on resting-state functional connectivity. This allows to characterize each individual's brain connectome as a unique pattern of connectivity or, in other words, as a brain fingerprint for the subject. One way researchers can evaluate performance of the model is by splitting the available data into a training set and a test set, and use them to train the model and to evaluate its performance, respectively. However, a single split of the data can be biased and unstable, especially if the dataset is limited, as is often the case in neuroscience. Cross-validation helps to overcome these issues by performing multiple iterations of this process on different subsets of the data, and then aggregating the results. The most common type of cross-validation is k-fold cross-validation, where the data is divided into k equally sized folds. The model is trained on $k-1$ folds and tested on the remaining fold, and this process is repeated k times, with each fold serving as the test set once. The performance of the model is then averaged over the k iterations. A similar technique, used in this project thesis, is Leave-One-Out Cross-Validation (LOOCV) [50]. It is k-fold cross validation with k equal to N, the number of data points in the set, meaning that N separate times, the model is trained on all the data except for one point, whose prediction is made for that left-out sample. By repeatedly training the model on a subset of the available data and testing it on a left-out sample, LOOCV can provide an estimate of how well the model is likely to perform on

new, unseen data. Specifically, in this project work, linear regression models are trained to map between the MEG-based resting-state FC pattern and their corresponding task-evoked FC data in a cohort of $N-1$ subjects, and subsequently the trained model is evaluated to the unseen subject to predict its task activation. This LOOCV approach is based on the model proposed by Tavor (2016) [5] in the field of fMRI study but readapted and modified to suit the MEG data characteristics.

2.4.4 Neural Networks

In its most general form, a *neural network* (NN), also known as artificial neural network (ANN), is a computational model which is inspired by the highly complex structure and function of biological brain networks. It can be used for a variety of regression tasks, including modeling complex and non-linear relationships between input and output, where the goal is to predict a continuous variable y, or a vector $\mathbf{y}$, based on a set of input features $\mathbf{x}$, or a matrix $\mathbf{X}$ [51]. As brain networks, it consists of interconnected nodes, the neurons, that process information and transmit signals through a network of connections, the synapses. Each neuron in a NN receives input signals from other neurons, properly modulated by weights and biases which determine the strength of the connections and control the overall activation level of the neuron. Hence, the neuron processes that input by summing up all the weighted inputs it receives and applying a non-linear activation function, and finally transmits an output signal to other neurons in the network. Training a neural network can be achieved by implementing a *backpropagation* algorithm. It is a well-known method of computing the gradient of the loss function with respect to the weights of the neural network, which is necessary for updating the weights in order to minimize the loss between predictions and observations. After training, the network can be used to predict output values for a new input data. In the context of this study, the neural network will be used to investigate the relation between the functional connectomes during rest and task conditions, as well as the non-linear relationship between muscle activity and FC patterns.

3

Materials and methods

3.1 Human Connectome Project

The Human Connectome Project (HCP) was launched in 2010 by the National Institutes of Health (NIH) as an ambitious effort to accelerate advances in human neuroimaging and studies in brain connectivity [52]. This interdisciplinary research project combines various imaging techniques and data analysis methods to construct comprehensive maps of the connections between different brain regions, and how they relate to cognitive and behavioural processes. The *WU-Minn-Ox* HCP consortium, one of the two consortia awarded by NIH, centered at Washington University, the University of Minnesota, and University of Oxford, succeeded in its core objectives by acquiring and analyzing high-quality MRI and MEG data together from more than 1100 HCP participants, and freely sharing the data and analysis via the ConnectomeDB database.

The *HCP Young Adult* dataset [9], the reference dataset for this thesis work, is a large-scale imaging and behavioural dataset of the human brain that analyses a population of 1200 healthy young adults, aged between $22 - 35$. The focus is directed on MEG data, acquired at 7 Tesla during resting-state (rMEG) and motor tasks (tMEG). MEG data from all HCP subjects are scanned on a whole head MAGNES 3600 (4D Neuroimaging, San Diego, CA) system housed in a magnetically shielded room, located at the Saint Louis University (SLU) medical campus. The MEG system includes 248 magnetometer channels together with 23 reference channels and data are recorded at 2034.51 Hz sampling rate. The setup requires HCP participants to be positioned in the MEG scanner supine, with the crown of the head touching but not pressing into the back of the MEG dewar. This

position leaves the eyes and the rest of the face free of the scanner. During resting-states scans, rMEG data are acquired in three runs of approximately 6 minutes each. Subjects are instructed to remain still during scanning and relax with eyes open and maintain fixation on a projected red crosshair on a dark background. ECG and EOG electrodes are attached to the subject for cardiac and oculomotor monitoring for artifact rejection. Following completion of rMEG, subjects are asked to perform finger tapping, with left or right hand, and toe squeezing, with left or right foot, that are designed to activate a variety of cortical and subcortical networks. During task acquisitions, tMEG data are extracted. Movements are paced with a visual cue, which is presented in a blocked design. Each block started with an instruction screen indicating the side (left, right) and the limb (hand, foot) to be used by the subject in the current block, resulting in a 3 second cue. Then, 10 pacing stimuli are presented in sequence, each one instructing the participants to make a brisk movement. The interval between consecutive stimuli is fixed to 1200 msec, for a total of 12 sec for each block. The visual stimulus stays on the screen for 150 msec and for the remaining 1050 msec the screen was black. In addition to the blocks of limb movements there are 10 interleaved resting blocks, each one lasting 15 sec. The experiment is performed in 2 runs, namely *Motort-10* and *Motort-11*, each consisting of 42 blocks: 10 of these blocks are resting blocks, the other 32 are divided in 16 hand movements (8 right and 8 left), and 16 foot movements (8 right and 8 left). This yields in a total of 80 movements per motor effector (LH, RH and LF, RF) for each run. In addition to the recorded MEG channels, EMG activity was recorded from each limb. EMG electrode stickers are applied to the skin to the lateral surface on the extensor digitorum brevis muscle and near the medial malleolus for the foot, also the first dorsal interosseus muscle between thumb and index finger, and the styloid process of the ulna at the wrist for the hand [9].

3.2 MEG pre-processing

MEG recordings for 51 subjects[1] have been properly downloaded from *ConnectomeDB* data management platform by the *Cognitive and Systems Neuroscience lab (Cosync lab)* [53]. The data collection includes raw MEG signals acquired for each subject, during resting states and task scans, and the related pre-preprocessing functions. Raw data has undergone a series of transformations and manipulations in order to remove noise, artifact, or other sources of interference, and to provide a cleaned and refined representation of the magnetic signals generated by brain activity. Two different pre-processing pipelines have been used based on rMEG or tMEG datasets. Although the pipelines differ in terms of specific processing steps, many operations associated to noise removal and artifact rejection as well as spatial and temporal filtering, are shared between them. Firstly, data filtering is performed by means of a band-pass Butterworth filter (cutoff frequency 1.3 Hz and 150 Hz). Notch filters are then applied to the band-pass filtered data in order to remove the power line noise (cutoff frequencies 59-61 Hz and 119-121 Hz). Then, artifacts, bad channels, and bad segments are identified and removed from rMEG and tMEG data by an Independent Component Analysis (ICA)-based approach, which allows to decompose MEG raw data into environmental non-brain components and subject's brain components.

Regarding the rMEG pre-processing, the resulting non-artifact sensor maps of the brain ICs, produced in the channel-level, are projected into the source space by means of Tikhonov-regularized Minimum-Norm Estimator (MNE) and Linearly Constrained Minimum Variance (LCMV) Beamformer filters. The source space onto which the IC topographies are projected consists of the individual surface-registered cortical sheet comprising 8004 vertices. Subsequently, for each vertex of the cortical sheet and for each of the three frequency bands (α, β_{low} and β_{high}), the Band Limited Power (BLP) envelope is estimated by a moving average integrating the squared amplitude of the vertex signal, with a window duration of 400 msec and a sliding time step of 20 msec. The final product is a dense connectome at the source-level, corresponding to the functional connectivity matrix, obtained

[1]Subjects ID: 105923, 106521, 108323, 109123, 113922, 116726, 133019, 153732, 156334, 162026, 162935, 164636, 169040, 175237, 177746, 185442, 189349, 191841, 198653, 204521, 205119, 212318, 212823, 221319, 250427, 255639, 257845, 287248, 293748, 353740, 358144, 406836, 559053, 568963, 581450, 599671, 601127, 660951, 662551, 667056, 679770, 680957, 706040, 707749, 725751, 735148, 783462, 814649, 891667, 898176, 912447

Network	Network Acronym	#Nodes	#Links
All nodes	AllNtw	164	13530
Auditory Network	AN	15	120
Control Network	CON	6	21
Dorsal Attention Network	DAN	24	300
Default Mode Network	DMN	28	406
Fronto-Parietal Network	FPN	12	78
Language Network	LN	13	91
Motor Network	MN	23	276
Ventral Attention Network	VAN	14	105
Visual Foveal Network	VFN	12	78
Visual Peripherical Network	VPN	17	153

Table 3.1: Brain networks and the corresponding number of nodes and links.

through the Pearson correlation coefficient and estimated on pairs of slowly varying BLP time series for each frequency band. Specifically, for every vertex pair on the surface-registered cortical sheet comprising 8004 vertices, Pearson's correlation is estimated across epochs of 25 seconds of non-overlapping data. The correlation coefficients estimated over these epochs in each run are subsequently averaged across epochs and runs. The 164-ROIs parcellation scheme is then applied to the surface-registered cortical sheet: the resulting FC matrix is described as a 164×164 symmetric matrix where each entry represents the connectivity strength between the BLP time series of two nodes. However, to avoid the spurious co-dependence of closer nodes, a leakage-correction method, based on the geometric correction scheme (GSC) approach [54], is employed. This step is necessary, as closer nodes may be more likely to be influenced by spurious noise or artifact than more distantly located nodes. Therefore, values in the FC matrix corresponding to pairs of nodes that are less than 35 mm apart are replaced by NaN, which stands for *Not a Number*. Finally, in order to reduce redundancy, the symmetry of the FC matrix allows to consider only its upper triangular part where each element corresponds to a link for a couple of nodes. The number of nodes, as well as the number of links, is reported in Table 3.1.

In the case of tMEG source-level preprocessing, the pipeline used is the same employed for rMEG preprocessing but integrated with the event-related fields (ERFs), which are MEG neural signals time-locked to specific events, such as motor responses [9]. Overall, the most relevant steps of tMEG preprocessing in source space and for each motor task are: compute the power series of the average

ERFs, compute time courses of frequency BLP and, finally, estimate time-resolved dense connectomes in the three frequency bands.

3.2.1 Degree-Normalization

The study of brain functional connectome, described as a weighted and undirected network or an adjacency matrix in graph theory context [3], aims to understand how distributed neural regions (ROIs) interact with each other during resting-state and task conditions [55]. To characterize the topology of functional networks, numerous statistics may be introduced. One of the most fundamental measure is the weighted degree (or the strength) of a node, defined as the sum of the weights of its neighbouring edges. In addition, the weighted degree sequence denotes the vector gathering the weighted degree of all nodes in the network. *Degree-normalization* uses the information encoded in the weighted degree sequence in order to reduce the strength between strongly connected nodes (*hubs*) comparatively to others, thereby balancing their excessive influence in the network [12]. Denoting M as the number of nodes for the network considered (see Table 3.1), the $M \times M$ functional connectivity matrix $\mathbf{FC}$ contains the Pearson's correlation coefficients between BLP time series associated to pairs of nodes, i and j:

$$\mathbf{FC} = [\mathbf{FC}_{ij}] \tag{3.1}$$

where $\mathbf{FC}_{ij} \in [-1, 1]$ and $\mathbf{FC}_{ij} = \mathbf{FC}_{ji}$. In order to not incur in complex numbers as a result of the degree-normalization, the absolute value of each correlation coefficient in $\mathbf{FC}$ is considered, namely $|\mathbf{FC}|_{ij} \in [0, 1]$. The degree d_i of node i is therefore defined as the sum of the weights of its neighboring edges:

$$d_i = \sum_{j=1}^{N} |\mathbf{FC}|_{ij} \tag{3.2}$$

As a result, the degree matrix $\mathbf{D}$ is the $M \times M$ matrix comprising the degree sequence on its diagonal and zeros elsewhere:

$$\mathbf{D}_{ii} = d_i \qquad \mathbf{D}_{ij} = 0 \quad \forall i \neq j \tag{3.3}$$

Therefore, the mathematical equation representing the degree-normalization is:

$$\mathcal{FC} = \mathbf{D}^{-1/2} |\mathbf{FC}| \mathbf{D}^{-1/2} \tag{3.4}$$

where $\mathcal{FC}$ is the symmetric adjacency matrix resulting from normalization.

3.3 EMG pre-processing

Electromyography preprocessing pipeline typically consists of several steps to prepare the raw EMG signal for subsequent analysis. Starting from signal acquisition, raw EMG signals have been amplified and recorded using surface electrodes placed on the skin overlying the muscle of interest during finger tapping and toe squeezing. The four channels corresponding to each limb are highly affected by various types of physiological and non-physiological noise and artifacts that originate mainly at the skin-electrode interface and in the amplifier electronics. Therefore, filtering is necessary to maximize signal-to-noise ratio and to isolate the frequency band of interest. Specifically, a band-pass 4° order Butterworth filter, with 15 Hz and 450 Hz cut-off frequencies, is applied to remove low-frequency and high-frequency disturbances and retain the most of unaffected EMG content. Low-frequency band is generally addressed to motion artifacts, high impedance of the analog circuit and electrocardiogram interference [56], while high-frequency band is mainly associated to different types of high-frequency contraction noise [57]. The resulting EMG signal, however, is still affected by the strong impact of environmental power line interference (PLI), generated by the linear alternative 60 Hz sinusoidal current, and by large artifacts around 180 Hz. In order to filter out these spectral components, two stop-band notch filters with 60 Hz and 180 Hz cut-off frequencies are applied to EMG signals [10]. After these preprocessing steps, a further cleaning operation has been performed based on *bad epochs*. An EMG epoch is defined as the temporal segment between two consecutive stimuli, or flashing crosses, in which the subject is required to execute the movement [11]. Bad segments consisted in short time epochs where the tMEG signals recorded is abnormal and cannot be used for further processing. During these epochs, sources lying outside the brain create a magnetic field which is several times larger than the brain magnetic field, probably generated by eye squints from the subject, or related to external sources, such as power supply bursts [9].

3.3.1 Method for active EMG detection

A threshold-based method is used to identify and extract the onsets and offsets of muscle activation in EMG signals. This approach considers a double threshold,

based on the amplitude of the signal, to distinguish bursts of muscle activity from background noise. A moving average (MA) filter is then applied to smooth the EMG signal by calculating its average amplitude over a sliding window of fixed length. A moderate window duration, such as 200 ms, is commonly used to balance noise reduction and signal capture, while shorter or longer window duration may lead to false positives or negatives and inaccurate burst properties [58]. As a result, an onset is detected when the value of the moving average exceeds a predefined threshold for the first time, and similarly, an offset is captured when the moving average value falls below the threshold after the period of muscle activity. More specifically, in this thesis work, the EMG detection procedure involves the following steps, slightly modified with respect to Fagotto's MSc thesis [11]:

1. *Rectification*: the process of rectification involves taking the absolute value of the EMG filtered time series $x(n)$ at each time point n, namely $y(n) = |x(n)|$. This has the effect of removing the negative component of the EMG signal, and producing a unipolar signal that only contains positive values.

2. *Thresholds*: the definition of the thresholds is based on the observation that muscle contractions typically produce a large, transient increase in EMG activity, which can be detected by looking for peaks in the EMG signal that exceed a certain threshold value. Specifically, given the maximum value $max = \max|y(n)|$ of all the peaks considered in the EMG rectified time series, the first threshold value thr_1 is defined as the mean of all the values with amplitude greater than 10% of max. Formally:

$$thr_1 = \frac{\sum_{n=1}^{N} z(n) \cdot \delta[z(n) = y(n)]}{\sum_{n=1}^{N} \delta[z(n) = y(n)]} \tag{3.5}$$

$$\text{where} \quad z(n) = \begin{cases} y(n) & \text{if } y(n) > 0.1 \cdot max \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

with N representing the total number of samples, $y(n)$ the rectified EMG value after filtering operation, and with indicator function δ, which takes the value 1 if the condition inside the bracket is true and 0 otherwise:

$$\delta[z(n) = y(n)] = \begin{cases} 1 & \text{if } z(n) = y(n) \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

This definition for the first threshold thr_1 allows to better discriminate muscle activity with respect to small amplitudes of the baseline. Following this step, a second threshold thr_2 is introduced and is defined experimentally as

$$thr_2 = 0.2 \cdot thr_1 \tag{3.8}$$

These threshold values are set based on the observation of amplitude and variability of the EMG signal. However, for all runs, all subjects and all muscle types, it turns out that the choice of these threshold parameters is suitable even if it requires some degrees of experimentation.

3. *Sliding window*: a sliding window embedding a moving average filter, with 200 ms duration, helps to smooth out the signal and reduce the effect of noise and artifacts. Whenever the MA of the sliding window passes above the threshold value thr_2, then the first sample of the window is considered an onset. Similarly, if it passes below the threshold value thr_2, then the last sample of the window is considered an offset.

Figure A.3 in the Appendix presents a visual representation of the method for detecting the onset and offset of EMG signal.

Considerations about the EMG onset/offset detector

This method for active EMG detection allows to extract segments corresponding to muscle activation. However, the detector may fail to detect all bursts in the signal, particularly those with low amplitudes, considered less significant for the analysis. Additionally, in some cases, subjects may perform more movements than required or anticipate them before the flashing cross. In such cases, strictly adhering to the criterion of 80 movements within each of the 8 block windows would mean to discard any bursts occurring beyond it, resulting in a loss of valuable data and potentially biasing the results. To address these issues, it may be appropriate to consider all bursts that occur during the recording session. This approach can help to ensure that as much data as possible is captured and used in the analysis, providing a more comprehensive understanding of the subject's EMG activity. Moreover, it is worth noting that some bursts, especially in foot movements, can be noisy and composed of unusual long contraction movements. This may make them difficult to automatically detect using the threshold-based method, as it may not meet the standard criteria for onset and offset detection. As a result, manual

segmentation turns out necessary to ensure accurate identification and extraction of the EMG bursts related to the movements of interest and excessively noisy.

3.3.2 EMG feature extraction

Feature extraction involves computing a set of quantitative measures or descriptors that capture relevant information about the EMG signal and its underlying physiological processes [59]. Three main categories of EMG features can be adopted, those being the time domain, frequency domain and the time-frequency domain. In the current project, only time domain features have been extracted for a total of 40 quantities [60, 61]. However, the focus is primarily on 10 time-domain features, widely used in electromyography, that are thought to be relevant for characterizing the muscle activity and for distinguishing the different movement classes. The inclusion of the remaining 30 features ensures that the results are robust, avoiding the possibility of bias in the analysis, and not driven solely by the choice of features. The list of 10 time-domain features with their formulas is presented in Table 3.2, whereas the 30 additional features are described in Appendix A.2 to the study. For each burst with duration N, the EMG signal, after preprocessing operations described in Section 3.3, is defined as $z(n)$.

3.4 Regression models

This thesis work includes two different investigations, both exploring different approaches to predict individual differences in motor tasks by analyzing brain activity. The first analysis focuses on identifying unique brain fingerprints based on subjects' functional connectomes during resting state and motor tasks. Three machine-learning linear regression models are presented, providing insights on their implementations and interpretations, whose structures are modified adaptations of Tavor's well-known LOOCV approach [5]. The second analysis aims to investigate whether individual differences in motor tasks can be directly observed in muscle activity. Inter-individual differences between brain activity and electromyography would suggest the existence of a motor fingerprint, consistent with findings of personal recognition based on EMG signal [62, 63]. Hence, a modified model, based on the three regression techniques of the first analysis, is created based on the resting-state functional connectomes and the extracted EMG features. Secondly, the analysis aims to explore the relationship between EMG features and brain

No.	EMG Feature	Formula
1	Average Amplitude Change	$AAC = \frac{1}{N}\sum_{n=1}^{N-1}\lvert z(n+1) - z(n)\rvert$
2	Average Energy	$AE = \frac{1}{N}\sum_{n=1}^{N}[z(n)]^2$
3	Coefficient Of Variation	$COV = \frac{\sigma}{\mu}$
4	Integrated EMG	$IEMG = \sum_{n=1}^{N}\lvert z(n)\rvert$
5	Kurtosis	$KURT = \frac{1}{N}\sum_{n=1}^{N}\frac{[z(n)-\mu]^4}{\sigma^4}$
6	Mean Absolute Value	$MAV = \mu = \frac{1}{N}\sum_{n=1}^{N} z(n)$
7	Root mean square	$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N}[z(n)]^2}$
8	Skewness	$SKEW = \frac{1}{N}\sum_{n=1}^{N}\frac{[z(n)-\mu]^3}{\sigma^3}$
9	Standard Deviation	$SD = \sigma = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}[z(n)-\mu]^2}$
10	Variance	$VAR = \sigma^2$

Table 3.2: 10 most relevant EMG time-domain features and their formulas

connectivity by developing a neural network. Specifically, the neural network is designed to predict EMG features in a new session of rest, providing insights into the underlying connections between the two.

3.4.1 Predicting inter-individual differences in motor tasks using rMEG and tMEG functional connectomes

The goal of the first analysis is to predict inter-individual differences in motor tasks using MEG-based functional connectomes during both muscular activity (tMEG) and resting-state (rMEG). To achieve this goal, three linear regression models are developed based on Tavor's LOOCV approach. The first two models are based on the hypothesis that the total connectome transforms almost linearly from rest to task for each subject, regardless of the node considered. The third model, instead, is based on the hypothesis that the overall connectome does not transform

linearly, but that each single link or connection undergoes a linear transformation from rest to task. This hypotheses will be discussed in Section 3.4.1, where the near-linear transformations will be explored and their validity will be examined. For each model presented, the *degree-normalization* operation (see Section 3.2.1) has been applied to individual FC symmetric matrices, and their triangular parts are flattened in order to extract FC arrays, also referred to as FC vectors or maps, whose elements represent specific connections between pairs of nodes in the network. For each subject, a mean FC vector is computed for each rest and task conditions by averaging the normalized vectors between corresponding runs. These feature vectors may be used as input to the regression models. However, one of the innovations in this study with respect to Vettoruzzo's and Fagotto's MSc thesis [10, 11], is to try to explicitly capture individual variations in FC arrays through *residualisation*. By residualing the input data of the models with respect to the group-averaged values, the accuracy and the reliability of the predictions can be improved. This enables to gain a deeper understanding of the factors that underlie the inter-individual differences in motor performance. The use of residualised data is consistent with recent studies in the field [64].

Residualisation of the resting-state and task functional connectomes

Residualisation is the process of separating each FC array into two components: the group map, denoted by subscript g, and the subject residual, denoted by superscript r. Thus, each individual resting-state and task maps can be written as the sum of the group-averaged activation map (scaled by some factor) and a further residual map specific to the given individual. The group-averaged rest $\mathbf{R}_g$ and task FC maps $\mathbf{T}_g$ are computed as follow:

$$\mathbf{R}_g = \frac{1}{N}\sum_{i=1}^{N}\mathbf{R}_i \qquad \mathbf{T}_g = \frac{1}{N}\sum_{i=1}^{N}\mathbf{T}_i \tag{3.9}$$

where $\mathbf{R}_i$ and $\mathbf{T}_i$ are the resting-state and task FC arrays for the i-th subject, and N is the total number of subjects in the dataset. The group-averaged FC arrays represent the typical FC patterns across the entire population of subjects. Hence, the decomposition of the i-th FC maps can be represented as follow [64]:

$$\mathbf{R}_i = \mathbf{R}_i^r + \alpha_i \mathbf{R}_g \qquad \mathbf{R}_i^r \perp \mathbf{R}_g \tag{3.10}$$

$$\mathbf{T}_i = \mathbf{T}_i^r + \beta_i \mathbf{T}_g \qquad \mathbf{T}_i^r \perp \mathbf{T}_g \tag{3.11}$$

where

$$\alpha_i = \frac{\mathbf{R}_i \cdot \mathbf{R}_g}{\|\mathbf{R}_g\|^2} \qquad \beta_i = \frac{\mathbf{T}_i \cdot \mathbf{T}_g}{\|\mathbf{T}_g\|^2} \tag{3.12}$$

are the scale factors that, multiplied with the group-averaged FC maps $\mathbf{R}_g$ and $\mathbf{T}_g$, represent the group-averaged components for i-th subject during resting-state and task conditions. The residual component of each FC array is obtained by subtracting the corresponding group-averaged component from the original FC array, as follow:

$$\mathbf{R}_i^r = \mathbf{R}_i - \alpha_i \mathbf{R}_g \qquad \mathbf{T}_i^r = \mathbf{T}_i - \beta_i \mathbf{T}_g \tag{3.13}$$

It is important to note that $\mathbf{R}_i^r$ and $\mathbf{R}_g$, as well as $\mathbf{T}_i^r$ and $\mathbf{T}_g$, are orthogonal, meaning that their dot product is zero. This ensures that the residual component captures inter-subject variability that is not explained by the group-averaged component. This technique aims to demonstrate that training and evaluating the model using residualised data (i.e, data where the group-averaged maps have been regressed out) would be of more value than training a model on the original resting-state and task maps. In the remainder of the study, the residualised resting-state and task maps are also referred to as resting-state and task *variation* maps.

Model 1: residual and group rMEG components to predict tMEG

The first model is a multiple linear regression model, inspired by Tavor's LOOCV approach (2016) [5]. The independent variables in the model are the resting-state FC maps, which are separated into residual and group components, while the dependant variables are the FC maps during motor tasks. The multiple linear equation is described by $\mathbf{y}_n = \mathbf{X}_n \boldsymbol{\beta}_n + \boldsymbol{\epsilon}_n$, where $\mathbf{y}_n$ is the task-based FC map for subject n, $\mathbf{X}_n$ is the design matrix that includes the constant term 1 and the residual and group FC vectors for subject n, and $\boldsymbol{\epsilon}_n$ is the error term for subject n. The design matrix $\mathbf{X}_n$ is defined as follows:

$$\mathbf{X}_n = [\mathbf{1}, \mathbf{x}_n^r, \mathbf{x}_n^g] = \begin{bmatrix} 1 & x_{n1}^r & x_{n1}^g \\ 1 & x_{n2}^r & x_{n2}^g \\ \vdots & \vdots & \vdots \\ 1 & x_{nK}^r & x_{nK}^g \end{bmatrix} \tag{3.14}$$

where x_{nk}^r and x_{nk}^g denote the residual and group FC vectors for subject n and connection k (with $k = 1, \ldots, K$), respectively. Before training the model, the regression coefficient vectors $\boldsymbol{\beta}_n = [\beta_{n0}, \beta_{n1}, \beta_{n2}]^\top$ for each subject $n = 1, \ldots, N$

can be obtained by using the equation 2.4. Then, the LOOCV approach is applied to train the model. Specifically, for each subject n, the model is trained using the regression coefficients of the remaining $N-1$ subjects and the average of these regression coefficients is then calculated as follows:

$$\overline{\boldsymbol{\beta}_n} = \sum_{i=1}^{N, i \neq n} \frac{\boldsymbol{\beta}_i}{N-1} \tag{3.15}$$

This average regression coefficient vector $\overline{\boldsymbol{\beta}_n}$ is then used to predict the FC map during the motor task for the n-th subject using the multiple linear regression equation:

$$\widehat{\mathbf{y}_n} = \mathbf{X}_n \overline{\boldsymbol{\beta}_n} \tag{3.16}$$

where $\widehat{\mathbf{y}_n}$ represents the predicted FC map during the motor task. The first model assumes that the FC map during motor task can be modeled as a linear combination of the group and residual components of resting-state FC maps, namely the average FC pattern across all subjects and individual variations from the group pattern. By including both components in the model, the first model aims to capture both the commonalities and the individual differences in the FC patterns. It is also possible to include a z-normalization that involves scaling the features of FC maps, corresponding to the links $k = 1, \ldots, K$ with K dependant on the network considered (see Table 3.1), so that they have zero mean and unit variance. In this case, z-normalization can be applied to the resting-state and motor task FC arrays before training the regression model. This ensures that the features have the same scale, and helps to prevent any individual differences in the FC values from affecting the model's prediction as well as to improve performance. The z-normalization can be presented mathematically as:

$$x_{nk}^{z-norm} = \frac{x_{nk} - \mu_k}{\sigma_k} \quad \text{with} \quad \mu_k = \sum_{i=1}^{N} \frac{x_{ik}}{N} \quad \text{and} \quad \sigma_k = \sum_{i=1}^{N} \frac{(x_{ik} - \mu)^2}{N-1} \tag{3.17}$$

where x_{nk}^{z-norm} is the standardized FC value of the n-th subject and k-th link, x_{nk} is the original value of that element, μ_k is the mean of the N values corresponding to the k-th FC link, and σ_k is the standard deviation related to it. The same approach can be performed on y_{nk}.

Model 2: residual rMEG component to predict tMEG variation map

This model is based on the *baseline model* proposed by Zheng et al. (2022) [64]. It assumes that the individual task variations (i.e., residualised activation maps) can be represented though a linear relation with respect to the variations in resting-state FC arrays. Specifically, the linear equation can be described as $\mathbf{y}_n = \mathbf{X}_n\boldsymbol{\beta}_n + \boldsymbol{\epsilon}_n$ where:

$$\mathbf{y}_n = \begin{bmatrix} y^r_{n1} \\ y^r_{n2} \\ \vdots \\ y^r_{nK} \end{bmatrix} \quad \text{and} \quad \mathbf{X}_n = [\mathbf{1}, \mathbf{x}^r_n] = \begin{bmatrix} 1 & x^r_{n1} \\ 1 & x^r_{n2} \\ \vdots & \vdots \\ 1 & x^r_{nK} \end{bmatrix} \tag{3.18}$$

The variables $\mathbf{y}_n$ and $\mathbf{X}_n$ are, respectively, the task and the resting-state variation maps for subject n (the latter integrated with a column of ones which represent the intercept term). After the computation of the regression coefficient vector $\boldsymbol{\beta}_n = [\beta_{n0}, \beta_{n1}]^\top$ for each subject n through the equation 2.4, the training and the prediction of the model is performed on residualised data and consists on the same steps seen in model 1. The regression coefficient vector $\overline{\boldsymbol{\beta}_n}$ is estimated using the same LOOCV approach as the first model. The regression coefficients are averaged over the $N-1$ remaining subjects, excluding subject n, and used to predict the task FC map for subject n:

$$\overline{\boldsymbol{\beta}_n} = \sum_{i=1}^{N, i\neq n} \frac{\boldsymbol{\beta}_i}{N-1} \qquad \widehat{\mathbf{y}_n} = \mathbf{X}_n\overline{\boldsymbol{\beta}_n} \tag{3.19}$$

Model 3: residual link-by-link model

The third model differs from the previous two models in that it assumes a linear relationship between the N residual FC values for each link k rather than assuming a linear relationship between the total connectome during rest and task. Specifically, for each subject n and link k, the corresponding tMEG and rMEG residual FC values, denoted as y_{nk} and x_{nk} can be described by a simple linear regression model. Mathematically, this can be expressed as:

$$y_{nk} = \beta_{0k} + \beta_{1k}x_{nk} + \epsilon_{nk} \quad \text{with} \quad n = 1, \ldots, N \tag{3.20}$$

where β_{0k} and β_{1k} are the intercept and the slope coefficient for link k respectively, and ϵ_{nk} represents the deviation from the linear relationship of the n-th subject for link k. To estimate the regression coefficients for each link k and for each subject n, the n-th subject is excluded from the system of equations and the coefficients are solved using the remaining $N-1$ subjects, as follow:

$$\begin{bmatrix} y_{1k} \\ \vdots \\ y_{(n-1)k} \\ y_{(n+1)k} \\ \vdots \\ y_{Nk} \end{bmatrix} = \begin{bmatrix} 1 & x_{1k} \\ \vdots & \vdots \\ 1 & x_{(n-1)k} \\ 1 & x_{(n+1)k} \\ \vdots & \vdots \\ 1 & x_{Nk} \end{bmatrix} \begin{bmatrix} \widehat{\beta_{0k}} \\ \widehat{\beta_{1k}} \end{bmatrix} + \begin{bmatrix} \epsilon_{1k} \\ \vdots \\ \epsilon_{(n-1)k} \\ \epsilon_{(n+1)k} \\ \vdots \\ \epsilon_{Nk} \end{bmatrix} \quad \text{with} \quad n = 1, \ldots, N \tag{3.21}$$

where $\widehat{\beta_{0k}}$ and $\widehat{\beta_{1k}}$ are the regression coefficients estimated on $N-1$ subjects through equation 2.4. This process is repeated for all N subjects, resulting in N sets of regression coefficients for each link k. These coefficients can then be used to predict the FC residual value during task for the n-th subject based on its residual FC value during rest, for each link k:

$$\widehat{y_{nk}} = \widehat{\beta_{0k}} + \widehat{\beta_{1k}} x_{nk} \tag{3.22}$$

Performance evaluation

Once the predictions are estimated, according to the model considered, a method to evaluate the model performance turns out necessary. The goodness of a model may be assessed by the mean squared error (MSE) which measures the average squared difference between the observed and predicted task-related FC maps, $\mathbf{y}_n$ and $\widehat{\mathbf{y}_n}$, defined as:

$$MSE = \frac{1}{K} \sum_{k=1}^{K} (y_{nk} - \widehat{y_{nk}})^2 \tag{3.23}$$

However, the purpose of the study is to understand whether the model is able to capture the inter-individual variability among subjects during their motor tasks. Thus, it is important to highlight that this purpose is not strictly related to the ability of the model in predicting FC values as close as possible to the actual ones. The inter-subject identifiability, in fact, is the ability to capture differences among individual FC patterns with the aim of comprehending which subject they belong to. A more discriminant way to identify individuals, widely used in many research

publications, is to use Pearson's correlation coefficient, rather than MSE [64, 5]. The Pearson correlation coefficient, as seen in Section 2.4.2, is a representative way to measure similarity between predicted and observed FC maps: it can actually measure whether the two patterns have a similar change trend so as to analyze the linear or non-linear relationship between them. In the context of brain fingerprints, a tool used to assess the performance of a classification model is the identifiability matrix, *IDmatrix* in short [5, 6]. It displays the number of correct and incorrect predictions made by the model, measuring the similarity between predicted and actual FC patterns during motor tasks. The IDmatrix is often represented as a square $N \times N$ matrix, where rows and columns represent the actual and predicted subject functional connectomes, respectively. Each cell contains the Pearson correlation coefficient $r_{\widehat{\mathbf{y}}_i \mathbf{y}_j}$ between the predicted values for i-th subject and the actual values for j-th subject, with diagonal elements representing the correlation for the same i-th subject, $r_{\widehat{\mathbf{y}}_i \mathbf{y}_i}$. Correlation coefficients in the diagonal of the IDmatrix higher than out-diagonal values indicate a better identification of the subject. Once the IDmatrix is visualized, two different metrics can be extracted in order to quantitatively assess the model performance:

1. *Prediction score*, defined as the average of the diagonal elements of the IDmatrix,

$$Pscore = \frac{1}{N} \sum_{i=1}^{N} r_{\widehat{\mathbf{y}}_i \mathbf{y}_i} \tag{3.24}$$

   It reflects the average level of similarity, intended as the average of linear correlation, between the actual and predicted values.

2. *Identification rate* (*IDrate*) [6], defined as the number of subjects *correctly identified* ($N_{correct}$), in percentage, or, in other words, whose diagonal elements are higher than the off-diagonal values,

$$IDrate = \left(\frac{N_{correct}}{N} \right) \cdot 100\% \tag{3.25}$$

   This metric allows to investigate on brain fingerprints among the subjects.

3.4.2 Predicting inter-individual differences in motor tasks using EMG features and rMEG FC matrices

The idea of this study is to investigate whether, replacing EMG features to the tMEG functional connectivity maps, inter-individual differences in motor tasks can

be predicted as well. To do so, a new model, whose implementation is very similar to the previous three models, has been developed and applied to EMG features and rMEG FC maps. However, a mathematical adaption is needed to accommodate the difference in dimensions between the EMG quantities and rMEG maps, which is why the use of the three original methods is not possible. Specifically, while the rMEG FC arrays have a different number of features K (i.e., edges), according to the network considered, the number of EMG features extracted from electromyography signal is 10, or at most 40 (see Tables 3.2 and A.2)[2]. For each EMG feature type f and for each subject n, a multiple linear equation is used as follow:

$$y_{nf} = \mathbf{X}_n \boldsymbol{\beta}_{nf} + \epsilon_{nf} \tag{3.26}$$

where y_{nf} and $\mathbf{X}_n = [1, \mathbf{x}_n^\top] = [1, x_{n1}, x_{n2}, ..., x_{nK}]$ are the EMG value and rMEG functional connectome (preceded by 1) for the feature type and subject considered, respectively. The regression coefficient $\boldsymbol{\beta}_{nf} = [\beta_{n0}, \beta_{n1}, ..., \beta_{nK}]_f^\top$ can be obtained by solving the multiple linear equation though standard least-squares regression technique (Equation 2.4). Therefore, for each EMG feature type and for each subject, the regression model involves $K + 1$ predictors and the first predictor is always a constant term. As a result, the regression coefficients $\boldsymbol{\beta}_{nf}$ for each subject now have $K + 1$ elements, instead of 2 or 3 elements as in the previous models. The LOOCV training technique and the prediction of the EMG feature value for the n-th subject is performed in the same way as described before:

$$\overline{\boldsymbol{\beta}_{nf}} = \sum_{i=1}^{N, i \neq n} \frac{\boldsymbol{\beta}_{if}}{N-1} \qquad \widehat{\mathbf{y}_{nf}} = \mathbf{X}_n \overline{\boldsymbol{\beta}_{nf}} \tag{3.27}$$

Note that this model may be applied not only on original resting-state FC maps but also on variations of rMEG functional connectomes, where the group-averaged components have been regressed out.

Additionally, this study shows the performance of a neural network model, whose characteristics are described in Section 4.3.2, applied to EMG features and rMEG functional connectivity. In this context, the LOOCV approach is still adopted: the NN is trained on $N - 1$ pairs of EMG feature values and the corresponding resting-state FC maps in order to predict the EMG feature value of the

[2]The EMG features are averaged across repetitions of the task and between the two sessions of muscular activity (Motort-10 and Motort-11)

n-th subject (left out from training) from its rMEG functional connectome.

The performance evaluation of the linear regression and NN models in this study involves compunting the Pearson's correlation and visualizing the IDmatrix, as shown in the previous section. The only difference lies on the size of the arrays being compared: the arrays contain now only 10 or at most 40 elements representing the number of EMG features being predicted rather than the number of links K in the FC maps.

3.4.3 Predicting muscular activity from brain connectivity in a new session of rest

This study aims to investigate the relationship between brain connectivity during resting-state and muscular activity, and to develop a NN model that can accurately predict EMG features (target outputs) based on rMEG FC patterns (input features). The purpose is to explore whether the non-linear NN model is more performant in predicting muscular activity than a Multiple Linear Regression (MLR) model, originally developed by Fagotto for the same task [11]. The EMG features are represented as an $N \times 1$ array, where N is the number of subjects, and are averaged between the number of repetitions and the two sessions of muscular activity (Motort-10 and Motort-11). The rMEG FC maps are represented as an $N \times K$ matrix, where K is the number of links, and are averaged between 2 out of 3 runs of resting-state. Thus, the NN model is trained on N subjects without excluding any of them, and used to predict the target outputs based on all rMEG FC patterns for a new session of rest. The network output is the estimated set of EMG features values, for the feature type and motor task considered. It is important to note that this study does not involve any LOOCV approach. Its aim is not to predict the muscle activity for a specific subject based on a model trained on the remaining $N-1$ individuals, but rather to investigate on the relation between brain connectivity at rest and muscle activity, and to test it on a new session of resting-state in order to confirm its validity.

Performance evaluation

To evaluate the performance of the NN model, an experimental method proposed in Fagotto's MSc thesis has been adopted, although slightly modified in order to make it more reliable. For each feature type f, effector and subject n, the predicted values are compared with the actual ones by assessing whether they fall

in the range defined as follow:

$$\mu_{nf} - \sigma_{nf} \leq \widehat{y_{nf}} \leq \mu_{nf} + \sigma_{nf} \tag{3.28}$$

where μ_{nf} (also referred to y_{nf}) and σ_{nf} are the intra-subject mean value and the standard deviation, calculated on the set of EMG feature values extracted across repetitions and between runs. A further analysis evaluates the significance of the prediction. Specifically, z-score hypotheses test determines whether the predicted value $\widehat{y_{nf}}$ belongs to a normal distribution with mean μ_{nf} and standard deviation σ_{nf}. The null hypotheses, stating that the prediction is a sample of the normal distribution, is rejected if the p-value is lesser than $\alpha = 0.05$. Hence, the predicted feature value $\widehat{y_{nf}}$ for the subject n and feature type f is considered correct whether it is within the typical range of values $[\mu_{nf} - \sigma_{nf}, \mu_{nf} + \sigma_{nf}]$ and passes the statistical test. The accuracy of the model is determined by calculating the rate of correct predictions, which is obtained by dividing the number of correct estimations ($N_{correct}$) by the total number of predictions ($N = 51$):

$$Accuracy = \left(\frac{N_{correct}}{N}\right) \tag{3.29}$$

In order to improve the performance of the NN model, an additional step is taken to normalize the target feature values. This is done using the following formula:

$$y_{nf}^{norm} = \frac{y_{nf} - \mu_f}{\sigma_f} \tag{3.30}$$

where μ_f and σ_f are the mean and the standard deviation of the target feature values across all individuals for the corresponding feature f. This z-normalization helps to scale the target feature values to a common distribution, represented by zero mean and unit variance.

4

Results

In this chapter, the results of the various studies conducted on the topic are presented. Due to the various networks and frequency settings that must be considered in the analysis for each movement type, the analysis is mainly oriented in finger tapping, particularly in right hand (RH). In many cases, β_{low} frequency is the typical band used for visualization purposes, as it is particularly related to muscle activity and brain reorganization during task [65, 66]. Furthermore, four networks − AllNtw, MN, VPN, and CON − are mostly selected as they represent a range of performance extremes and provide a clear illustration on differences across networks. Additional findings and details will be provided in Appendix A.

4.1 Functional connectivity analysis

This section provides a comprehensive analysis of the model's underlying hypotheses based on the functional connectivity maps. An example of the functional connectivity patterns during resting-state and right finger tapping condition (RH) for subject '106521' is shown in Figure 4.1. The focus is on assessing the existence of a linear relation between the functional connectomes measured during rest and task, and on exploring this relation on a link-by-link basis with particular investigation on the inter-individual variability of FC profiles. Additionally, the impact of degree-normalization on FC values is examined, and the effect of residualisation is discussed.

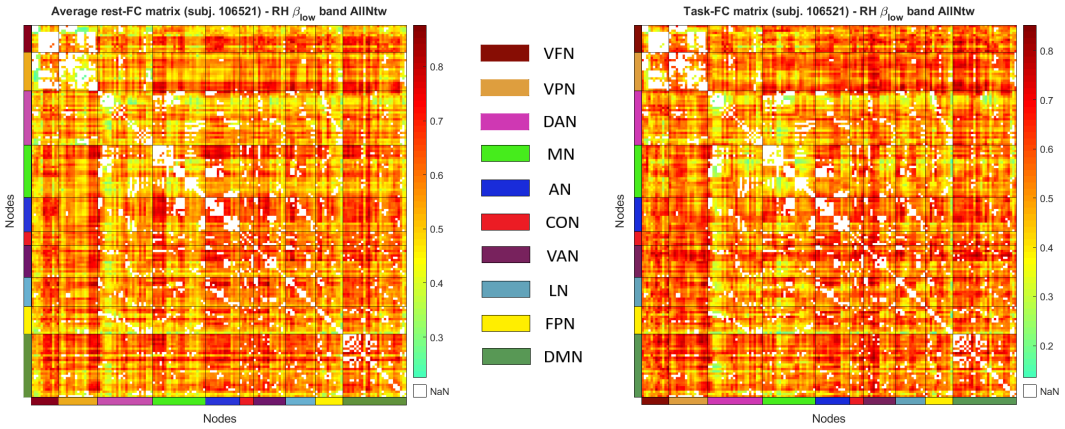


Figure 4.1: Functional connectivity maps during resting-state (left) and task condition (right) for subject '106521' considering all 164 nodes composing AllNtw in β_{low}.

4.1.1 Exploring the linear transformation of connectomes from resting-state to task

Figure 4.2 explores the linearity between functional connectomes, underlying rest and task conditions. To examine this relation, a straightforward approach is to display the actual FC values during the two conditions, and fit a linear regression line that minimizes the least square error. In this regard, just as an example of consideration, subject '559053' is shown in α, β_{low} and β_{high} frequency bands, RH and AllNtw. The first row of Figure 4.2 reveals that, for the subject considered, the FC values between rest and task tend to follow a near-linear relation in AllNtw. The correlation values, shown on each plot, indicate the degree of linear relationship of tMEG with respect to rMEG, with higher correlation values expressing a stronger linear relation. An alternative approach to examine this correlation for each subject is through a matrix, which exploits the correlations both within and between individuals in a range $[-0.2, 1]$. This approach is shown in the second row of Figure 4.2. Each cell of the matrices correspond to the correlation coefficient obtained by comparing the functional connectomes of the i-th and j-th subjects during rest (columns) and task (rows) conditions. The diagonal elements represent the intra-subject correlation, indicating the level of linear similarity between the two connectomes of the same subject. It is important to note that higher diagonal values do not imply identical transformations from rest to task across subjects, but rather suggest that the relation between rMEG and

tMEG is approximately linear for each individual. In contrast, the out-diagonal elements represent the correlation between subjects' functional connectomes, which captures the inter-individual variability in FC values during resting-state and task conditions. Smaller out-diagonal values indicate the presence of unique brain fingerprints on both rMEG and tMEG-based functional connectivities for different subjects. Conversely, higher out-diagonal values suggest greater similarity (intended as linear relationships) between connectomes of different individuals in both rest and task conditions.

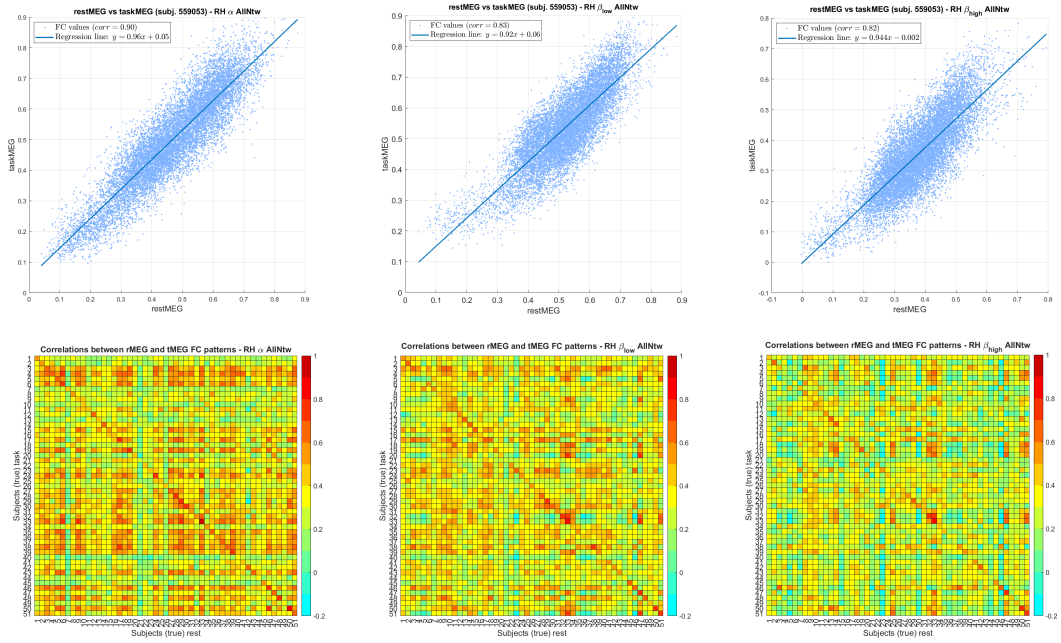


Figure 4.2: Linear relationship between functional connectomes during rest and task conditions, as well as the correlation matrix of the connectomes across the subjects. The first row displays the scatter plots of the FC values for subject '559053' in the α, β_{low}, and β_{high} frequency bands, right hand (RH) and AllNtw. The blue lines represent the linear regression of the data, while the correlation coefficient is reported in the legend. The second row shows the correlation matrices for the same bands and network, with the diagonal and out-diagonal elements representing the intra-subject and inter-subject correlations, respectively.

However, when examining individual networks, e.g. CON, MN and VPN (other networks can be found in Figure A.6 and A.8 in the Appendix), a linear relation between rMEG and tMEG functional connectivities is not always evident. This can be attributed to the limited number of nodes in each network, in comparison to the entire brain network, which can hinder the characterization of the system through linear regression relationships. As depicted in Figure 4.3, the diagonal elements

do not exhibit notably higher correlation values, and the out-diagonal elements indicate that the inter-subject variability is evident. Thus, the assumption of a

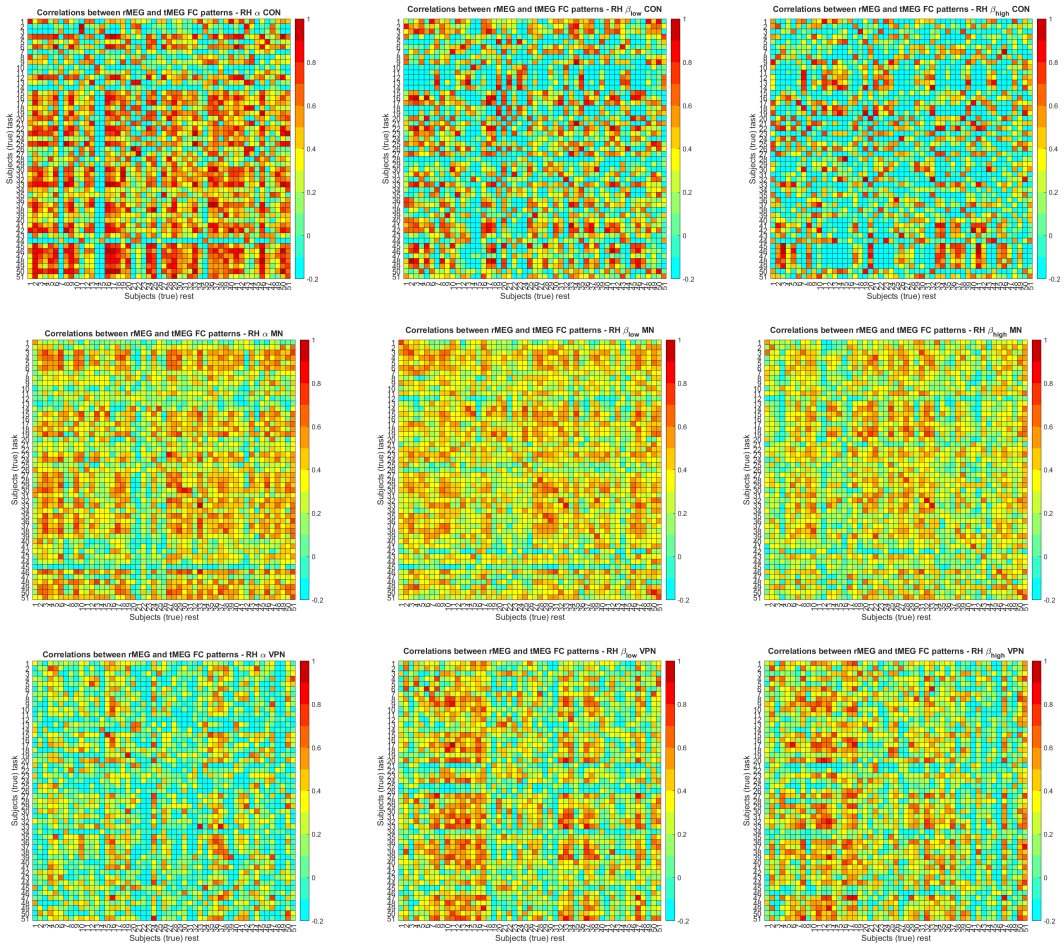


Figure 4.3: Correlation matrices of the functional connectomes during rest and task conditions for the networks CON, MN and VPN. Unlike the whole brain network (AllNtw), the limited number of nodes in these networks makes it difficult to characterize the system through linear relationships between rMEG and tMEG functional connectivities. The diagonal elements do not show higher correlation values, and conversely the out-diagonal elements suggest that the inter-subject variability is very high for these networks.

linear relationship between rMEG and tMEG functional connectivities is valid only when considering the whole brain network, and not its subdivisions. In order to ensure a linear relationship between rMEG and tMEG functional connectivities for all networks considered and the three frequency bands, it is necessary to apply a transformation to the input data. By utilizing degree-normalization (as discussed in Section 3.2.1), the resulting diagonal correlations are enhanced but they come

with the drawback of increasing the off-diagonal correlations.

Assessing the impact of degree-normalization

The application of degree normalization allows for the balancing of weights of highly connected nodes (hubs) and low-connected nodes, effectively scaling the data. Figure 4.4 displays the resulting matrices obtained from this transformation for AllNtw, CON, MN, and VPN (other networks in Figure A.7 and A.9) across the three frequency bands. As a result, it helps to establish a linear relationship between rMEG and tMEG functional connectivities for all networks considered. Degree-normalization has an additional important consequence on the results. As shown in Figure 4.5, which displays the color-coded FC values for each subject in the β_{low} frequency band for AllNtw, MN, and VPN networks, the application of degree-normalization causes the FC values to cluster more tightly around a common linear pattern, which is almost indistinguishable from a unique linear regression line. Specifically, the normalized data tend to be closer to the average regression line (the solid black line) than the original ones. In contrast, the original data appear to be more dispersed than those obtained from the normalized data. This indicates that, in absence of normalization, the linear transformation from rest to task varies across subjects, and regression lines may differ across individuals. Conversely, degree-normalization yields almost to the same linear transformation from rest to task for most individuals, resulting in regression lines that are close to each other, and implying a similar map from rest to task between individuals.

Visualizing group-average and residuals components

Residualisation refers to the process of decomposing each functional connectivity (FC) map into two components, namely, the residual and the group average components (Section 3.4.1). The present study aims to investigate in detail the group-average and residual decomposition and to establish the rationale for considering the FC variation maps as the optimal input data for improving model performance. The two components are subjected to the same step of the analysis performed previously, considering AllNtw, MN, and VPN in β_{low} frequency band, as depicted in Figure 4.6 and 4.7 (other networks in β_{low} shown in Figure A.10 and A.11). The residual component captures the variability in FC maps among subjects and is characterized by higher FC values in the diagonal and lower values in the out-diagonal elements compared to the results obtained solely by apply-

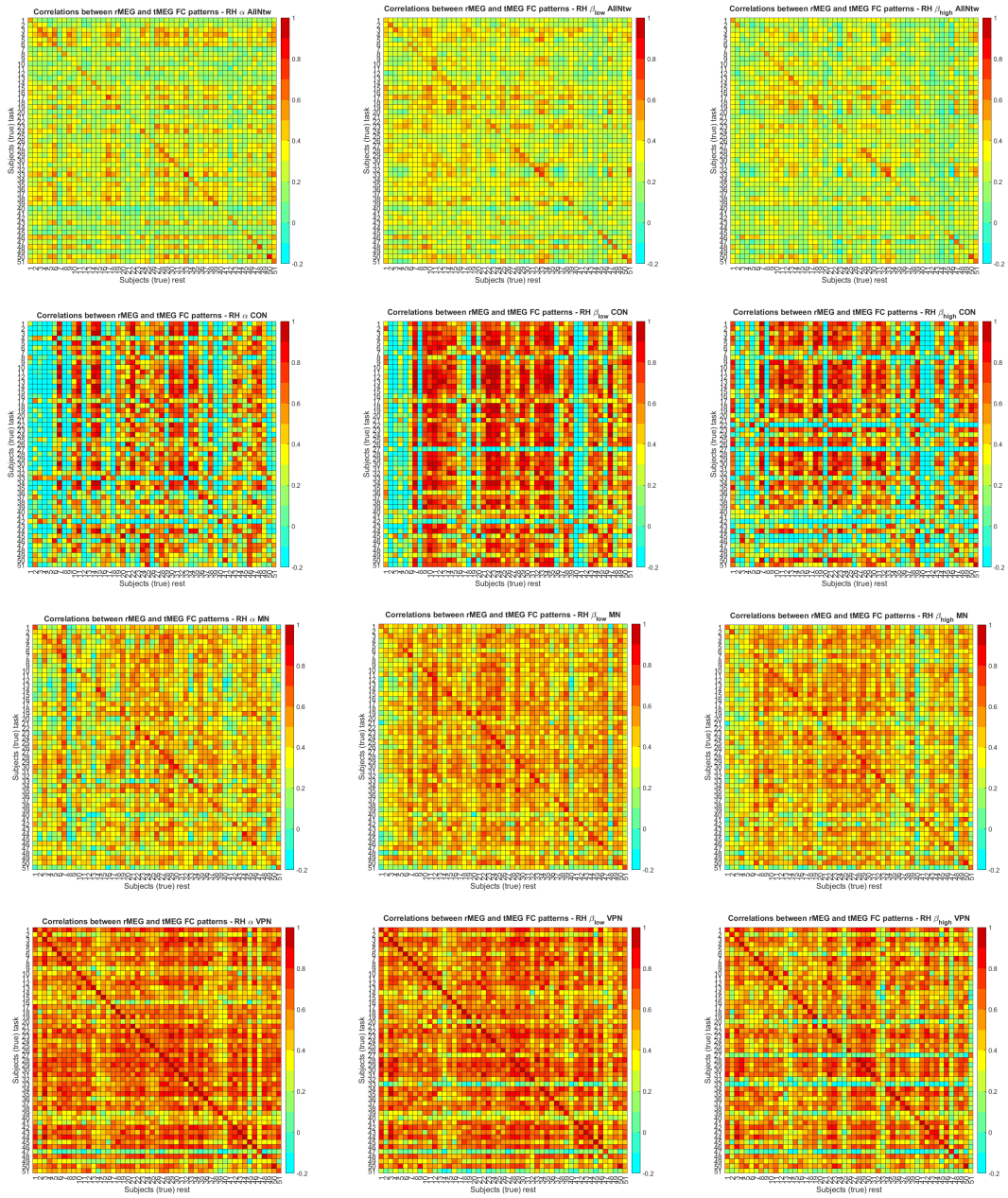


Figure 4.4: Matrices of correlation coefficients between rMEG and tMEG functional connectivities after applying degree-normalization to the input data, for four brain networks (AllNtw, CON, MN, and VPN) and three frequency bands. Before degree-normalization, the matrices show lower diagonal and out-diagonal correlations, indicating a weaker linear relationship between rMEG and tMEG functional connectivities. After degree-normalization, the matrices exhibit higher diagonal and out-diagonal correlations, suggesting a stronger linear relationship.

ing degree-normalization to the data. Conversely, the group-averaged component is marked by very high correlation values on both diagonal and out-diagonal el-

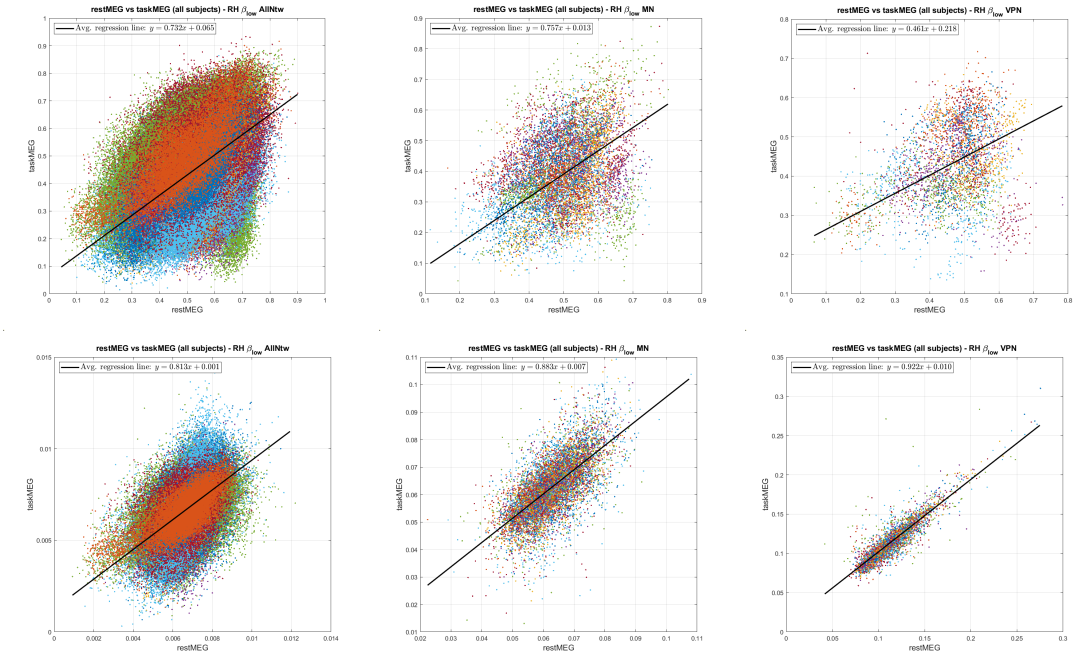


Figure 4.5: Effect of degree-normalization on the linear relationship between rMEG and tMEG functional connectivities for AllNtw, MN, and VPN networks in the β_{low} frequency band. Each dot represents the FC value for a network's link and for a single subject during rest and task conditions, color-coded by subject, without (first row) and with (second row) degree-normalization. The solid black line represents the averaged linear regression line for all the subjects, equivalent to the mean of individual linear regression lines. The effect of degree-normalization is visible as it compacts the data points around the averaged regression line, indicating a more consistent linear relationship across the subjects.

ements, indicating minimal inter-subject variability. This observation confirms that the residual component and the group-averaged component are independent of each other, and that the residual component is the optimal choice for optimizing the model as it maximizes inter-subject variability, while preserving linearity between rMEG and tMEG FC maps for each individual, with almost the same linear relation for each subject.

4.1.2 Exploring the linear link-by-link transformation of FC values from resting-state to task

In the current study, the hypotheses underpinning the third model, presented in Section 3.4.1, are investigated by focusing on each link connecting two nodes. The goal is to test whether the functional connectivity link transforms linearly

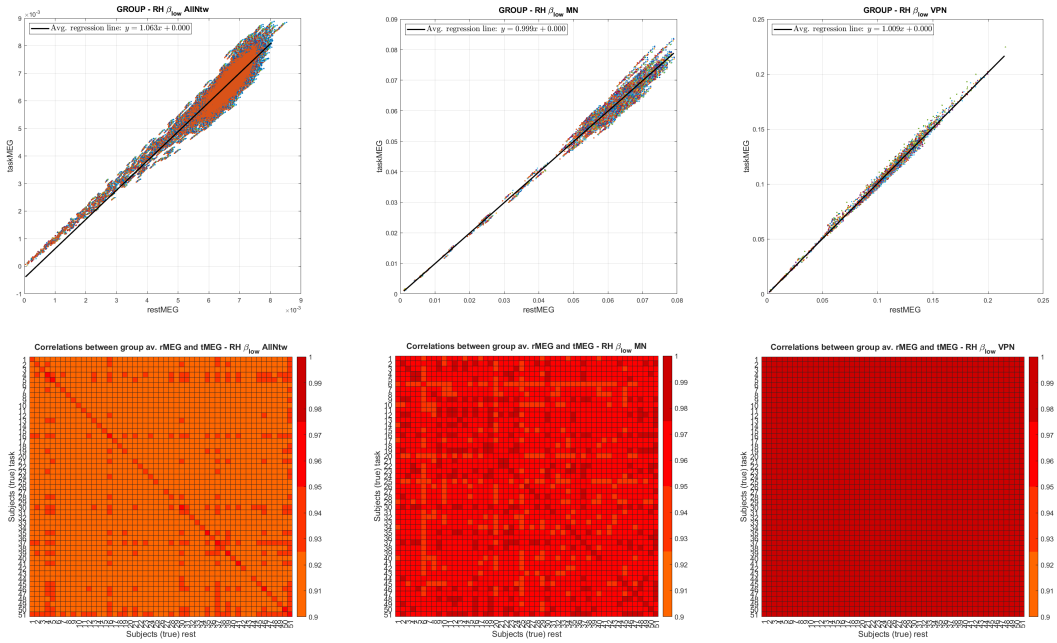


Figure 4.6: FC values and matrices of correlations associated to the group-averaged components after residualisation approach. Each panel represents a different network (AllNtw, VPN, and MN) in β_{low} frequency band. The group-averaged components reflect the common patterns of FC maps across subjects, characterized by high values in both diagonal and off-diagonal elements.

across subjects. To achieve this, the 51 sample values for a randomly chosen link − original rMEG and tMEG FC values (without degree-normalization) − corresponding to the 51 subjects are plotted. The Figure 4.8 displays three panels corresponding to one random link (e.g. link 501) in the three frequency bands of AllNtw, where the 51 values are fitted by a solid regression line. The regression coefficients, p-value and correlation for each panel are provided in the legend. As presented, the p-value for the examined link falls below the significance threshold, indicating that a linear transformation is applicable to describe this relation. Despite the significant p-value, the level of correlation is very weak, meaning that there is almost no linear relationship between the values considered. To visualize the behaviour of the total number of links in AllNtw, comprising all networks considered, a histogram of correlations can be employed. The histogram displays the count of correlation values in a particular bin, providing a clear visualization of the correlation distribution. The median, mean, and standard deviation of the distribution are presented in a textbox within the plot. As observed, the mean correlation values are low, measuring 0.25 in α, 0.23 in β_{low}, and 0.31 in β_{high}. In

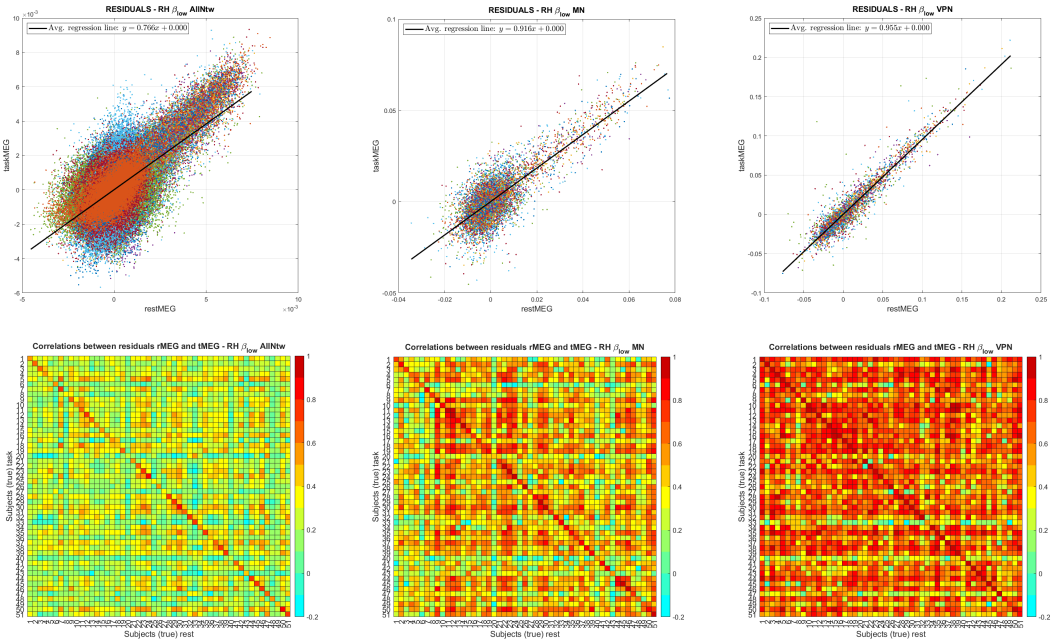


Figure 4.7: FC values and matrices of correlations associated to the residual components after residualisation technique. Each panel represents a different network (AllNtw, VPN, and MN) in β_{low} frequency band. The residual components capture the inter-subject variability in FC maps, characterized by higher values in the diagonal and lower values in the off-diagonal elements.

accordance with the previous section, the degree-normalization and residualisation techniques can be also applied to the FC values in the present study. Figure 4.9, in the first row, illustrates the degree-normalized FC values, which exhibit a slight improvement in the mean of the distribution compared to the original FC values. However, in the second row, the residual FC values are displayed, revealing that no additional improvement can be obtained compared to the normalized values. Based on these observations, it can be concluded that constructing a model based on a linear transformation for each link considered may not fully adhere to the behaviour of the data. As previously demonstrated, utilizing group-averaged components can substantially enhance the correlation between the data. However, it fails to accurately capture individual patterns as each subject's values are projected onto a common pattern, making it difficult to differentiate across subjects. Figure A.12 in the Apppendix illustrates that the use of group-averaged components results in a narrow distribution pattern in the correlation histogram, indicating that the overall connectome for each subject behaves similarly to the other individuals, making it challenging to distinguish across them.

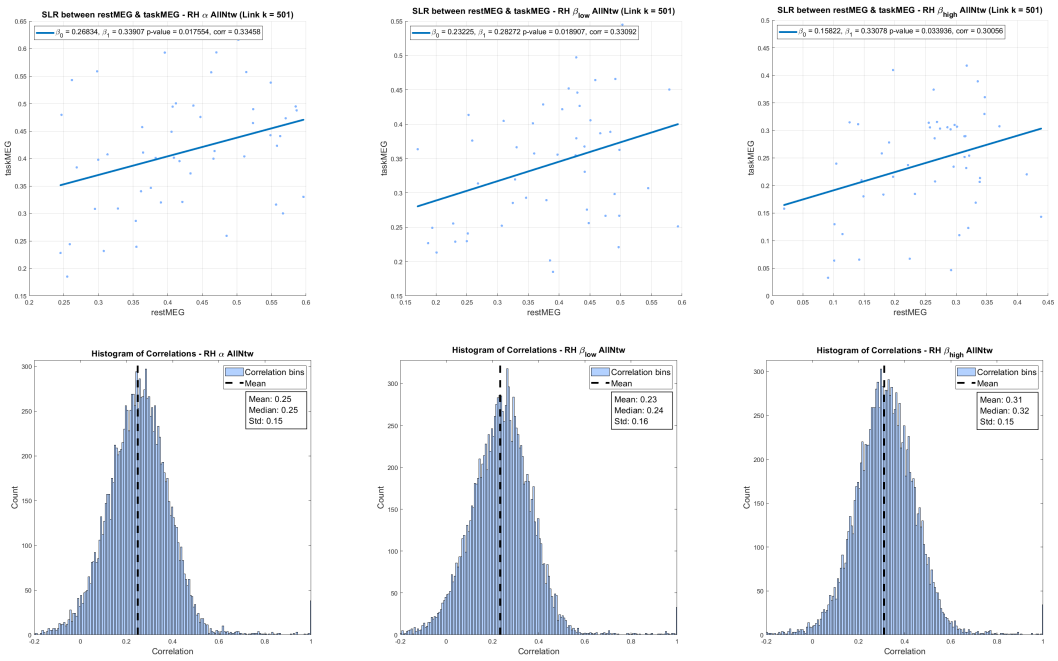


Figure 4.8: The first row of the figure shows the original rMEG and tMEG FC values, corresponding to a random link (e.g. 501) of AllNtw, plotted for each of the 51 subjects in α, β_{low}, and β_{high} frequency bands. Each plot is fitted with a solid regression line. The second row displays a histogram of the correlations between all links in AllNtw, showing the number of counts for each bin of correlation. The mean, median, and standard deviation of the correlation are also reported in a textbox on the plot.

4.2 Modelling tMEG from rMEG-based functional connectivity

The objective of this study is to predict individual variations in tMEG functional connectivities using rMEG FC maps. This goal is strictly related to detect a unique brain fingerprint for each subject that enables accurate identification. In this section, an evaluation of the three models presented in Section 3.4.1 will be conducted. The performance of each model will be assessed through the IDmatrix, which provides an indication of the number of individuals correctly identified based on the IDrate and p-score (respectively Equation 3.24 and 3.25). The comparison of the models will also involve a comparison with the original, normalised and residual data, allowing for an assessment of their relative efficacy. Additionally, the performance of each model will be investigated with respect to individual networks and frequency bands, with the aim of determining which networks as well

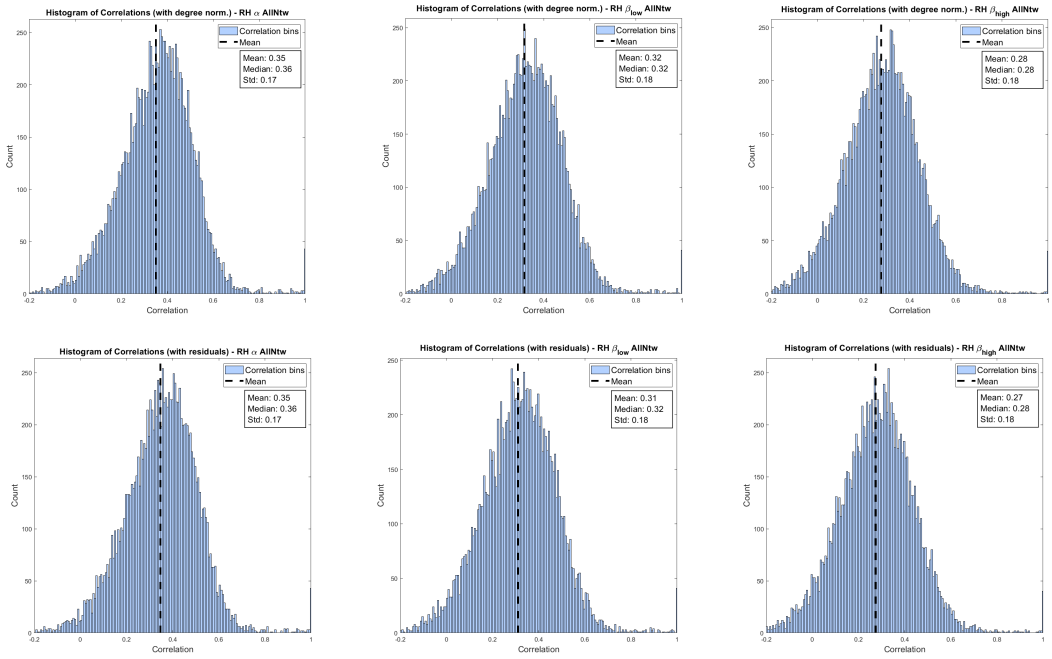


Figure 4.9: Application of the degree-normalization and residualisation to the functional connectivity data. The first row displays the distribution of correlation between degree-normalized FC values, while the second row shows the equivalent betweeen residual FC values (after degree-normalization) for AllNtw in α, β_{low}, and β_{high}. In both cases, the mean of correlation is slightly improved compared to the original FC values.

as frequency bands exhibit the best performance.

4.2.1 Analyzing model performance across networks

Figure 4.10 depicts the IDmatrices obtained by the three different models in β_{low} and AllNtw, CON, MN and VPN (additional results can be found in Figure A.14 and A.13 in the Appendix). Each column corresponds to a model, and each row represents a network. A correct identification of the individuals is indicated by higher values on the diagonal compared to off-diagonal elements within the row. Table 4.1 presents the list of p-score and IDrate values for each network and frequency band. An IDrate greater than 70% is considered a good performance, highlighted in yellow cells. It can be observed that Model 2 (built on mapping resting to task residual FC maps) performs better than the others in many networks, such as AllNtw, DMN, MN, and VPN. Regarding frequency bands, α and β_{low} show comparable and better performance than β_{high}, indicating their greater involvement in brain reorganization during this type of task [65,

66]. Conversely, β_{high} exhibits the poorest performance in many of the networks considered, suggesting its limited involvement in the processing of sensorimotor information during motor tasks. Notably, the matrices obtained from all the models are mostly diagonal-dominant in each network, with the VPN network showing the highest performance. The Control Network (CON), presents a less evident diagonal, making it the only exception in this task. The network analysis will be further discussed in the next chapter.

4.2.2 Comparing original and residualisation-based models

In order to determine the effectiveness of the residualisation step, a comparison is made between the IDrate and the p-score values obtained with and without the decomposition technique. Table 4.2 illustrates the performance of a LOOCV model, equivalent to the model 2 (which is found to be the most performant network among the models analyzed) without residualisation step, and with and without degree-normalization. The IDmatrices for AllNtw, MN, CON, and VPN are provided in Figure A.15 in the Appendix. As expected, the results demonstrate that the performance of the model without residualisation is lower than that obtained by residualising the data, with normalized data performing better than original ones. This observation is independent of the model used and is only related to the data characteristics. The residualised data increase the inter-individual differences compared to the normalized and original data, even though the linear transformation from rMEG to tMEG functional connectivity can be considered almost the same across individuals. This can be explained by the high variability of FC residual maps during resting-state or task state, which enhances the identifiability of subjects. Further details about the characteristics of FC variation maps (residualised data) will be discussed in the next chapter.

4.3 Predicting EMG muscular activity from brain connectivity

In this section, the central topic is to explore the possibility of predicting inter-individual differences in muscular activity directly from EMG features based on rMEG functional connectivity. To accomplish this task, a new machine learning-based LOOCV model, presented in Section 3.4.2, has been trained on EMG features and rMEG FC maps to predict muscular activity in a new unseen subject.

		Model 1		**Model 2**		**Model 3**	
Network	Band	$Pscore$	$IDrate$	$Pscore$	$IDrate$	$Pscore$	$IDrate$
AllNtw	α	0.36 ± 0.14	78.43%	0.62 ± 0.13	90.20%	0.61± 0.12	74.51%
	β_{low}	0.32 ± 0.16	72.55%	0.61 ± 0.14	88.24%	0.58± 0.13	66.67%
	β_{high}	0.30 ± 0.15	58.82%	0.50 ± 0.15	82.35%	0.48± 0.15	70.59%
AN	α	0.44 ± 0.22	39.22%	0.73 ± 0.19	49.02%	0.73± 0.19	39.22%
	β_{low}	0.44 ± 0.24	41.18%	0.76 ± 0.20	43.14%	0.76± 0.19	35.29%
	β_{high}	0.38 ± 0.22	27.45%	0.68 ± 0.23	45.10%	0.66± 0.23	19.61%
CON	α	0.27 ± 0.32	9.80%	0.37 ± 0.39	17.65%	0.33± 0.40	11.76%
	β_{low}	0.33 ± 0.33	7.84%	0.37 ± 0.37	11.76%	0.30± 0.38	5.88%
	β_{high}	0.27 ± 0.39	7.84%	0.31 ± 0.42	11.76%	0.23± 0.41	3.92%
DAN	α	0.37 ± 0.18	45.10%	0.65 ± 0.17	52.94%	0.64± 0.17	39.22%
	β_{low}	0.37 ± 0.21	47.06%	0.65 ± 0.17	56.86%	0.63± 0.16	39.22%
	β_{high}	0.33 ± 0.20	47.06%	0.57 ± 0.18	58.82%	0.54± 0.19	35.29%
DMN	α	0.39 ± 0.20	58.82%	0.70 ± 0.16	70.59%	0.69± 0.15	58.82%
	β_{low}	0.34 ± 0.17	60.78%	0.70 ± 0.16	74.51%	0.68± 0.15	52.94%
	β_{high}	0.32 ± 0.17	45.10%	0.57 ± 0.17	76.47%	0.56± 0.16	50.98%
FPN	α	0.41 ± 0.20	35.29%	0.70 ± 0.15	47.06%	0.68± 0.16	27.45%
	β_{low}	0.41 ± 0.24	43.14%	0.68 ± 0.19	49.02%	0.65± 0.20	31.37%
	β_{high}	0.28 ± 0.20	19.61%	0.57 ± 0.20	37.25%	0.55± 0.19	21.57%
LN	α	0.40 ± 0.20	29.41%	0.53 ± 0.21	29.41%	0.51± 0.21	25.49%
	β_{low}	0.28 ± 0.25	25.49%	0.45 ± 0.29	39.22%	0.42± 0.27	19.61%
	β_{high}	0.23 ± 0.22	19.61%	0.37 ± 0.26	25.49%	0.32± 0.27	15.69%
MN	α	0.42 ± 0.21	52.94%	0.77 ± 0.13	74.51%	0.75± 0.12	52.94%
	β_{low}	0.45 ± 0.18	64.71%	0.81 ± 0.11	72.55%	0.78± 0.12	52.94%
	β_{high}	0.42 ± 0.15	60.78%	0.74 ± 0.13	66.67%	0.72± 0.15	43.14%
VAN	α	0.30 ± 0.16	37.25%	0.70 ± 0.16	56.86%	0.68± 0.18	29.41%
	β_{low}	0.36 ± 0.20	37.25%	0.69 ± 0.16	37.25%	0.65± 0.20	33.33%
	β_{high}	0.31 ± 0.19	19.61%	0.56 ± 0.17	41.18%	0.53± 0.19	33.33%
VFN	α	0.34 ± 0.24	25.49%	0.73 ± 0.20	33.33%	0.71± 0.22	11.76%
	β_{low}	0.43 ± 0.22	23.53%	0.73 ± 0.22	31.37%	0.69± 0.23	15.69%
	β_{high}	0.26 ± 0.23	21.57%	0.62 ± 0.24	35.29%	0.54± 0.26	15.69%
VPN	α	0.74 ± 0.18	88.24%	0.95 ± 0.07	94.12%	0.93± 0.10	76.47%
	β_{low}	0.73 ± 0.18	84.31%	0.94 ± 0.09	90.20%	0.93± 0.07	74.51%
	β_{high}	0.63 ± 0.23	70.59%	0.89 ± 0.17	86.27%	0.88± 0.15	62.75%

Table 4.1: P-score and IDrate for each network and frequency band, accordingly to the model considered. The p-score is presented as $mean \pm std$. The cells highlighted in yellow correspond to the IDrate values greater than 70%. Model 2 shows better performance in many networks, such as AllNtw, DMN, MN, and VPN. The frequency bands, α and β_{low} exhibit similar and better performance than β_{high}, indicating their stronger involvement in brain reorganization during motor tasks.

		Model 2 (norm. data)		**Model 2 (original data)**	
Network	Band	*Pscore*	*IDrate*	*Pscore*	*IDrate*
AllNtw	α	0.52 ± 0.14	70.59%	0.58 ± 0.15	62.75%
	β_{low}	0.52 ± 0.14	64.71%	0.54 ± 0.15	54.90%
	β_{high}	0.47 ± 0.14	66.67%	0.50 ± 0.16	47.06%
AN	α	0.80 ± 0.11	21.57%	0.56 ± 0.31	21.57%
	β_{low}	0.77 ± 0.16	33.33%	0.41 ± 0.32	25.49%
	β_{high}	0.65 ± 0.20	23.53%	0.36 ± 0.29	17.65%
CON	α	0.53 ± 0.38	11.76%	0.55 ± 0.29	17.65%
	β_{low}	0.56 ± 0.40	13.73%	0.45 ± 0.38	11.76%
	β_{high}	0.50 ± 0.36	7.84%	0.34 ± 0.46	11.76%
DAN	α	0.57 ± 0.18	37.25%	0.62 ± 0.22	33.33%
	β_{low}	0.58 ± 0.16	33.33%	0.60 ± 0.18	31.37%
	β_{high}	0.53 ± 0.18	43.14%	0.52 ± 0.21	37.25%
DMN	α	0.69 ± 0.11	49.02%	0.57 ± 0.16	39.22%
	β_{low}	0.68 ± 0.14	49.02%	0.57 ± 0.16	39.22%
	β_{high}	0.58 ± 0.15	31.37%	0.53 ± 0.18	33.33%
FPN	α	0.56 ± 0.17	33.33%	0.53 ± 0.20	21.57%
	β_{low}	0.54 ± 0.19	39.22%	0.53 ± 0.21	21.57%
	β_{high}	0.42 ± 0.22	25.49%	0.45 ± 0.25	21.57%
LN	α	0.56 ± 0.19	21.57%	0.49 ± 0.24	21.57%
	β_{low}	0.52 ± 0.24	19.61%	0.35 ± 0.30	27.45%
	β_{high}	0.43 ± 0.22	25.49%	0.30 ± 0.25	19.61%
MN	α	0.62 ± 0.19	41.18%	0.51 ± 0.23	17.65%
	β_{low}	0.69 ± 0.13	52.94%	0.53 ± 0.19	33.33%
	β_{high}	0.66 ± 0.13	45.10%	0.51 ± 0.18	31.37%
VAN	α	0.70 ± 0.14	41.18%	0.60 ± 0.19	23.53%
	β_{low}	0.65 ± 0.14	21.57%	0.52 ± 0.19	11.76%
	β_{high}	0.53 ± 0.17	27.45%	0.45 ± 0.20	15.69%
VFN	α	0.71 ± 0.15	29.41%	0.60 ± 0.20	5.88%
	β_{low}	0.72 ± 0.16	27.45%	0.67 ± 0.19	19.61%
	β_{high}	0.59 ± 0.23	23.53%	0.55 ± 0.28	15.69%
VPN	α	0.90 ± 0.12	90.20%	0.35 ± 0.29	15.69%
	β_{low}	0.88 ± 0.15	82.35%	0.41 ± 0.26	11.76%
	β_{high}	0.81 ± 0.22	76.47%	0.40 ± 0.25	7.84%

Table 4.2: Pscore and IDrate values for each network and frequency band obtained using a LOOCV model, equivalent to Model 2 without the residualisation step. The p-score is presented as $mean \pm std$. The cells highlighted in yellow correspond to the IDrate values greater than 70%. The results indicate that the model without residualisation performs lower than the model with residualisation, and that normalized data performs better than original data.

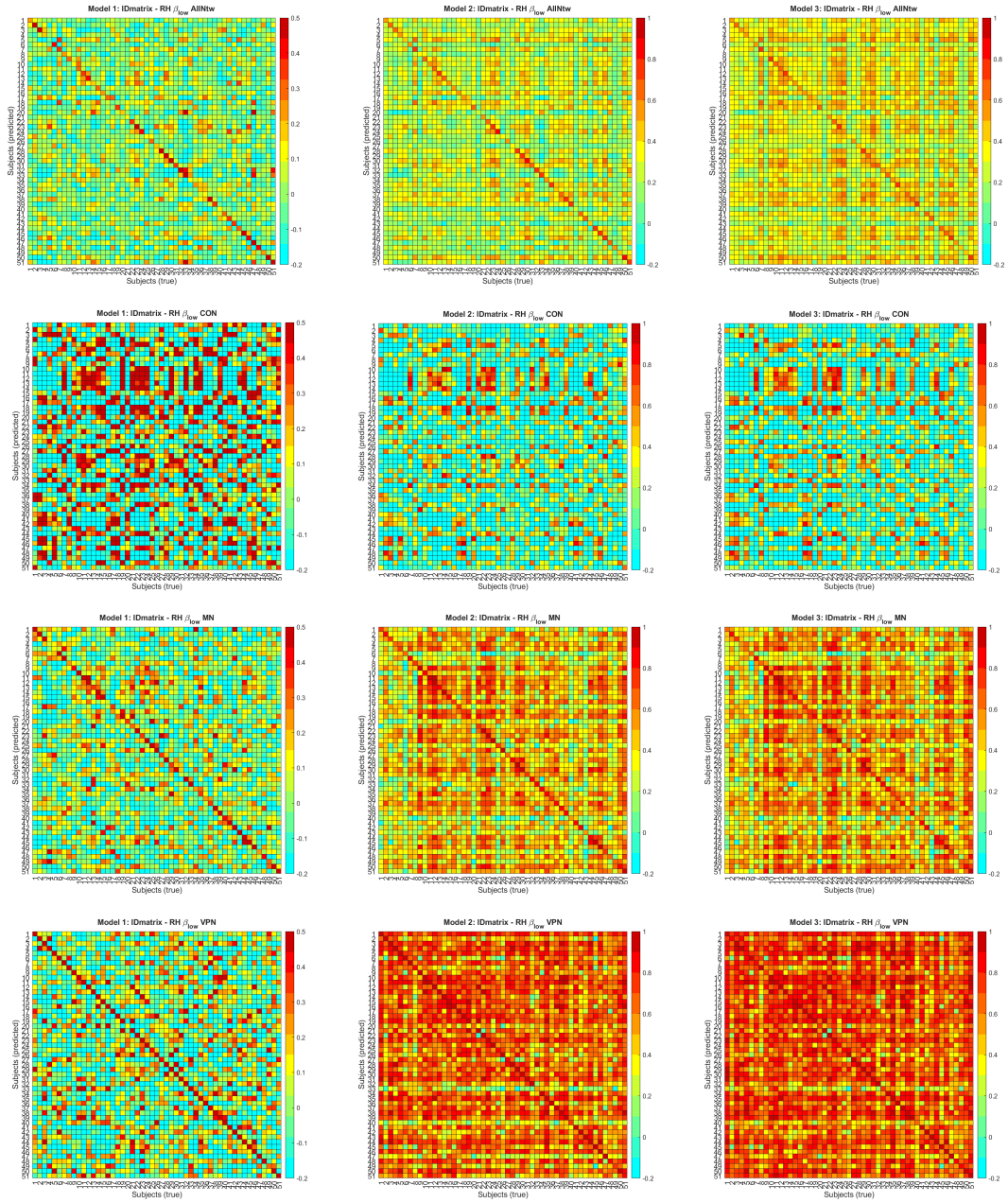


Figure 4.10: IDmatrices obtained with the three models for β_{low} frequency band, and four different networks (AllNtw, CON, MN and VPN). Each cell in the matrix represents the correlation coefficient between the predicted tMEG FC (rows) and the actual connectivity values (columns) for the same subject (diagonal elements) and for different subjects (off-diagonal elements). Diagonal elements higher than the corresponding off-diagonal elements in the row indicate correct identification of the individuals.

The 40 EMG features for the right hand are extracted using the method described in Section 3.3.1 and are reported in Table 3.2 and A.2. It is worth noting that

Right Hand (RH)

EMG feature	$mean \pm std$	EMG feature	$mean \pm std$
AAC [V]	$(2.58 \pm 1.27) \times 10^{-5}$	MAV [V]	$(6.88 \pm 3.45) \times 10^{-5}$
AE [V^2]	$(1.74 \pm 1.56) \times 10^{-8}$	RMS [V]	$(1.16 \pm 5.66) \times 10^{-4}$
COV	1.35 ± 0.20	SD [V]	$(9.23 \pm 4.53) \times 10^{-5}$
$IEMG$ [V]	0.08 ± 0.05	$SKEW$	2.61 ± 0.48
$KURT$	12.70 ± 3.79	VAR [V^2]	$(1.12 \pm 0.99) \times 10^{-8}$

Table 4.3: Mean and standard deviation of the 10 most relevant EMG features extracted from all subjects (V = Volt).

these features are averaged between the number of significant repetitions (detected by the method) and the two motor sessions (Motort-10 and Motort-11).

4.3.1 Inter-subject variability for EMG features

Table 4.3 shows the mean and standard deviation of the 10 most significant EMG features extracted across subjects, providing insight into the variability of these features among individuals.

Another approach to better capture the variability of EMG features across subjects is boxplot representations. Boxplots provide a visualization of the summary statistics of the data, where the central mark represents the median, and the bottom and the top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, which are plotted individually using the '+' marker symbol. The data presented in Figure 4.11 shows that there is not a lot of variation in most of the EMG features among different individuals. This lack of distinctiveness can make it challenging to accurately identifying individuals based solely on the muscular activity generated by right finger tapping. Figure A.5 provides an additional boxplots displaying the variability of the further 30 features.

4.3.2 Assessing the inadequate model performance in predicting muscle activity

After extracting the EMG features and training the LOOCV model, the model's performance is evaluated using IDmatrix. Each cell represents the correlation coefficient computed between actual and predicted EMG features for the same or dif-

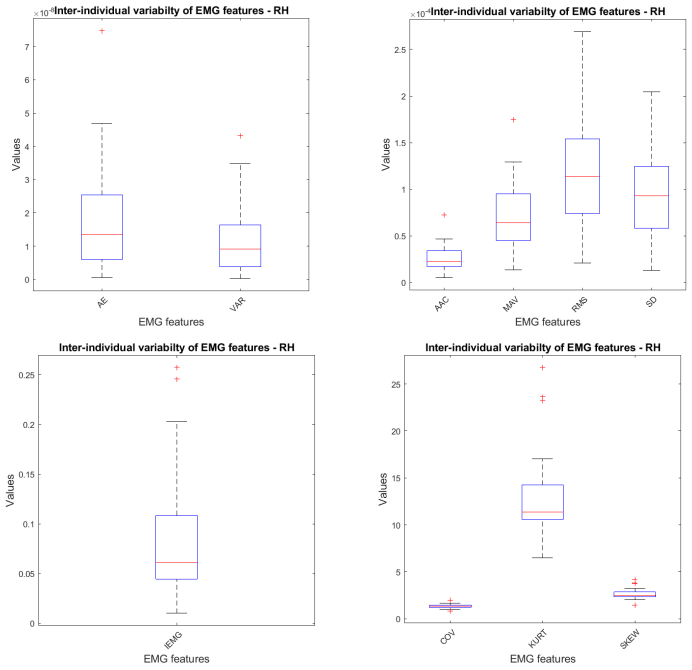


Figure 4.11: Boxplots of 10 most relevant EMG feature values across individuals. The central mark in each boxplot refers to the median, while the bottom and the top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points that are not considered outliers, which are plotted individually using the '+' marker symbol.

ferent individuals. However, the results obtained from the linear regression model are not satisfactory considering the 10 most relevant features, as depicted in the first IDmatrix of Figure 4.12 in β_{low} and MN. This poor performance,regarding the non-diagonal dominance of IDmatrices, is observed across all types of networks and frequency bands.

Thus, based on this outcome, it can be assumed that the number of EMG features used in the model is insufficient to capture the inter-individual variability in the muscular activity of subjects. In response, the model is retrained with the additional 30 features. However, despite this modification, the IDmatrices obtained still do not show any clear diagonal, indicating poor performance of the model, as reported in Figure A.16. Furthermore, the same unsatisfactory results (not shown in this thesis project) are obtained when considering only those features with higher inter-individual differences such as KURT, SKEW, and other related features, and also when replacing tMEG with rMEG during the model training. Thus, the idea is to apply data transformations, such as the normalization of features or the residualisation technique for both rMEG functional connectivities

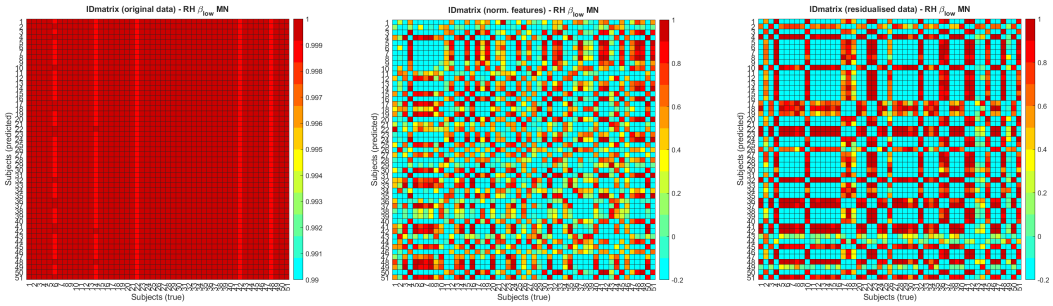


Figure 4.12: IDmatrices obtained using a LOOCV model based on linear regression, trained on rMEG and 10 EMG features in MN and β_{low}. The correlation coefficient between actual and predicted features is calculated for each cell. Despite applying data transformations such as normalization of EMG features or residual decompositions of both rMEG and EMG features, there is no evidence of a diagonal structure.

and EMG features, in order to increase their inter-individual variability. However, no good results can be achieved, as depicted in the second and third IDmatrix in Figure 4.12. These findings suggest that the features used in the model may not effectively capture and differentiate muscle activity across individuals, and that the structure of a linear regression model may not adequately adhere to the relationship between the functional connectomes, both during resting-state and task state, and the muscle activity. Hence, a more complex non-linear model has been developed to address the poor performance of the linear regression model. A feed-forward neural network is used with the built-in MATLAB function, *fitrnet* [67], to fit a non-linear regression model to the data. The network has two fully connected layers, with 10 neurons each. The first layer uses a rectified linear unit (ReLU) activation function, and the second uses a hyperbolic tangent activation function. The objective function minimized during the training of the neural network is the mean squared error (MSE) loss. The aim of the optimization process is to find the weights and biases of the network that minimize the MSE loss computed on EMG features of $N - 1$ individuals, thus making predictions on new, unseen subject. However, even with this more complex model, and with tuning of parameters such as the number of nodes and layers, the results are still poor, as depicted in Figure 4.13 and A.17. Therefore, it can be concluded that EMG features extracted during this simple motor task (RH) cannot sufficiently characterize a motor fingerprint for the individual.

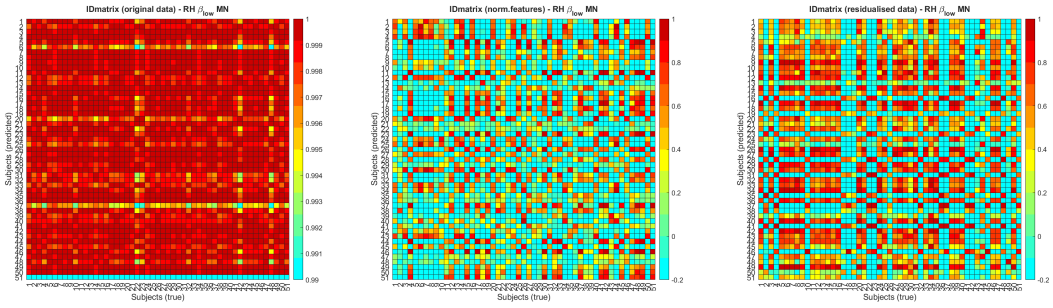


Figure 4.13: IDmatrices obtained using a NN model based on *fitrnet* MATLAB function, trained on rMEG and 10 EMG features in MN and β_{low}. Even applying data transformations such as normalization of EMG features or residual decompositions of both rMEG and EMG features, there is no evidence of a diagonal structure.

4.4 Predicting muscular activity from brain connectivity in a new session of rest

Due to the poor performance in predicting inter-individual differences in muscular activity based on functional connectomes, this section explores the potential of predicting muscle activity based on a new, unseen session of resting-state. To achieve this objective, two regression models have been employed. The first model used for this purpose is the Multiple Linear Regression (MLR) model, which was originally developed by Fagotto for the same task [11]. The second model, on the other hand, is an implementation of a non-linear neural network. Both models have been trained on the average of two resting-state sessions, namely the mean of the corresponding rMEG FC maps, and the extracted EMG features, which are averaged between repetitions and motor sessions (Motort-10 and Motort-11)[1]. Additionally, z-normalization is applied to the EMG features, as shown in Equation 3.30.[2] Testing is performed on the third session of resting-state and performance

[1]It is worth noting that the standard deviations used in this study are calculated using propagation error theory. Denoted $\mu_{10} \pm \sigma_{10}$ and $\mu_{11} \pm \sigma_{11}$ as the means and the standard deviations across repetitions in Motort-10 and Motort-11 respectively, the average across sessions $\mu_{10,11} \pm \sigma_{10,11}$ is calculated as $\mu_{10,11} = (\mu_{10} + \mu_{11})/2$ and the corresponding standard deviation through the error propagation formula $\sigma_{10,11} = 1/2(\sigma_{10}^2 + \sigma_{11}^2)^{1/2}$

[2]Error propagation is used to calculate the normalized data. For each feature $f = 1, \ldots, 10$ and for each individual $n = 1, \ldots, N$, the average feature value across repetitions and sessions is denoted as $y_{nf} \pm \sigma_{nf}$. The normalized data, indicated as $y_{nf}^{norm} \pm \sigma_{nf}^{norm}$, are calculated as $y_{nf}^{norm} = (y_{nf} - \mu_f)/\sigma_f$ with μ_f and σ_f describing the mean and the standard deviation across individuals for the corresponding feature f, and error propagation is used to calculate the

Right Hand (RH), β_{low}

	AllNtw	AN	CON	DAN	DMN	FPN	LN	MN	VAN	VFN	VPN
AAC	0.49	0.10	0.18	0.08	0.2	0.04	0	0.31	0.12	0.16	0.41
AE	0.61	0.12	0.22	0.14	0.35	0.02	0.08	0.33	0.06	0.16	0.47
COV	0.90	0.16	0.51	0.39	0.31	0.08	0.08	0.63	0.08	0.29	0.73
IEMG	0.65	0.12	0.39	0.24	0.41	0.06	0.02	0.33	0.06	0.16	0.65
KURT	0.98	0.25	0.80	0.80	0.39	0.25	0.10	0.78	0.20	0.33	0.98
MAV	0.49	0.10	0.16	0.08	0.12	0.04	0.04	0.31	0.10	0.10	0.43
RMS	0.63	0.06	0.16	0.14	0.25	0.02	0.02	0.35	0.10	0.12	0.55
SD	0.69	0.06	0.20	0.25	0.25	0	0.02	0.41	0.10	0.14	0.59
SKEW	0.98	0.22	0.78	0.67	0.35	0.25	0.06	0.73	0.16	0.35	0.94
VAR	0.69	0.12	0.25	0.20	0.37	0.02	0.10	0.43	0.06	0.18	0.59

Table 4.4: Performances of the Multiple Linear Regression (MLR) model applied to the 10 EMG feature types extracted during right hand activity (RH) and rMEG functional connectomes in β_{low} frequency band. Each cell displays the accuracy calculated as the number of correctly identified subjects among the total ones, for each feature type and brain network considered. Accuracy values higher than 0.7 are highlighted in yellow.

of the regression models is evaluated by determining whether the predicted EMG feature values, denoted as $\widehat{y_{nf}}$ for each individual n and feature f, fall within the range described by the actual features and their standard deviations, specifically $[y_{nf} \pm \sigma_{nf}]$, as outlined in Section 3.4.3 and Equation 3.28. A statistical z-test is then conducted on the predicted values to confirm that they belong to the normal distribution with mean y_{nf} and standard deviation σ_{nf}.

4.4.1 Analysis of low predictive accuracy of linear regression models

Table 4.4 displays the performances of the MLR model applied to the 10 EMG feature types extracted during right hand activity (RH) and rMEG functional connectomes in β_{low} frequency band. Each cell of the table shows the accuracy calculated with the Equation 3.29, describing the number of correctly identified subjects among the total, for each feature type and brain network considered.

The table 4.4 shows that the MLR model performs poorly in predicting subjects correctly, with only *KURT*, and *SKEW* features exhibiting good accuracy rates

standard deviation of the normalized data, represented as $\sigma_{nf}^{norm} = \sigma_{nf}/\sigma_f$.

in certain networks, namely AllNtw, CON, DAN, MN, and VPN. These findings are consistent with those reported by Fagotto J., which identified Kurtosis and Skewness as the most effective features. Nevertheless, the low accuracy rates suggest that a well-defined linear relationship between the resting-state functional connectivity and the EMG feature values does not exist, indicating that the MLR model is not suitable for this task. Therefore, an alternative approach, such as a non-linear neural network, might be needed to improve performance.

4.4.2 Evaluate the effective performance for NN model

A non-linear neural network is implemented using the MATLAB function *fitrnet* as previously employed, with two hidden layers comprising 10 neurons each, and ReLU together with *tanh* activation functions. Table 4.5 presents the network's performance, which indicates that higher accuracy rates are achieved by using a non-linear relation between rMEG functional connectivity and EMG features. The most performant networks are AllNtw and VPN, followed by DAN, DMN, and MN. The latter shows an average accuracy equal to 0.68. In term of best features, beyond KURT and SKEW, other features such as AE, and COV can be used for a correct prediction of their values. Focusing on AE, it is possible to visualize the performance of the two models by presenting the predicted and the actual values, along with their corresponding standard deviations, in a plot. Figure 4.14 illustrates how many AE normalised feature values are correctly predicted across subjects and between the two regression models, NN and MLR, in β_{low}, AllNtw and RH [11]. Plots indicate whether normalised feature predictions fall within the range described by the true values and the standard deviations. While this study does not report on the impact of using different rest sessions for training and testing the NN model, it is worth noting that these results are still considered reliable and robust, regardless of which rest sessions are used. This suggests that the findings are not dependant on a particular selection of rest sessions, and that the non-linear neural network outperforms the MLR model being a more suitable approach for this task.

Right Hand (RH), β_{low}

	AllNtw	AN	CON	DAN	DMN	FPN	LN	MN	VAN	VFN	VPN
AAC	0.80	0.22	0.20	0.47	0.37	0.27	0.22	0.43	0.25	0.31	0.78
AE	0.78	0.18	0.24	0.49	0.43	0.27	0.20	0.78	0.37	0.20	0.84
COV	0.92	0.49	0.39	0.88	0.84	0.45	0.29	0.63	0.47	0.31	1.00
IEMG	0.84	0.31	0.25	0.55	0.61	0.31	0.31	0.63	0.45	0.33	0.90
KURT	1.00	0.88	0.76	1.00	0.96	0.75	0.65	1.00	0.73	0.59	1.00
MAV	0.69	0.16	0.14	0.49	0.24	0.14	0.14	0.57	0.08	0.29	0.86
RMS	0.80	0.25	0.18	0.49	0.37	0.04	0.10	0.51	0.31	0.12	0.88
SD	0.86	0.18	0.18	0.51	0.47	0.25	0.20	0.63	0.18	0.18	0.86
SKEW	1.00	0.82	0.55	1.00	0.90	0.71	0.55	0.96	0.55	0.53	1.00
VAR	0.80	0.31	0.18	0.53	0.43	0.35	0.16	0.65	0.25	0.25	0.86

Table 4.5: Performances of the Neural Network (NN) model applied to the 10 EMG feature types extracted during right hand activity (RH) and rMEG functional connectomes in β_{low} frequency band. Each cell displays the accuracy calculated as the number of correctly identified subjects among the total ones, for each feature type and brain network considered. Accuracy values higher than 0.7 are highlighted in yellow.

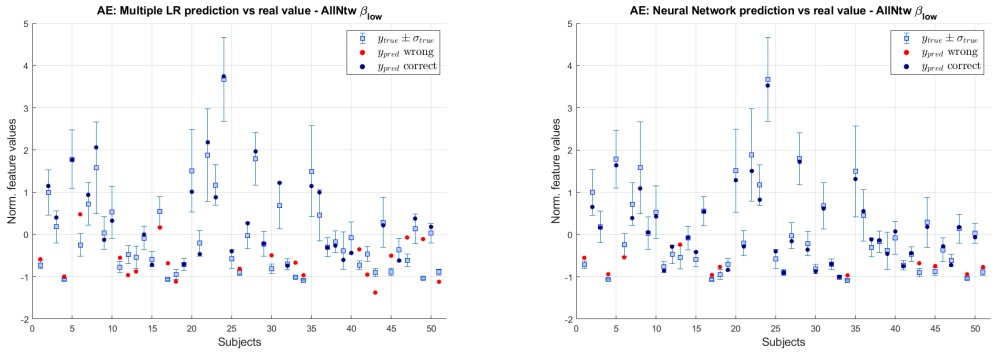


Figure 4.14: Comparison between the actual and predicted normalized values for the *AE* feature type. Each dot represents the prediction specific for the subject, and the square box and corresponding error bar indicate the actual value and its standard deviation for the same feature type and subject. The dark blue dots represent subjects whose predictions fall within the range described by the actual feature value and standard deviation. The plots illustrate the ability of the non-linear neural network to predict feature values across subjects.

5

Discussion

In this chapter, the hypotheses and results of the models presented in the previous section will be discussed, along with exploring the findings emerged from the analysis of the FC data and the limitations of the study. The aim is to provide a deeper understanding of the brain fingerprint through MEG-based functional connectivity.

5.1 Brain connectivity analysis

5.1.1 Inter-individual variability in brain connectivity

For the purpose of this study, it is important to define the notions of intra and inter-subject variability in brain connectivity that have been employed within the context of this research. Intra-subject variability refers to the stability of FC values at rest across different scans, regardless of the network and frequency band considered. Vettoruzzo has shown that almost all subjects exhibit similarity values above 0.90 [10]. On the other hand, inter-individual variability arises from the differences in the FC matrices during resting-state and motor task, as well as the existing relationships between them. Exploring inter-individual variability, and possibly applying data transformations to enhance it, can lead to the identification of a unique pattern in brain connectivity for each individual, i.e. the brain fingerprint. This concept is closely tied to identifying individuals based on their differences in motor tasks and resting state. However, the data must exhibit a common pattern that establishes a mathematical framework for analyzing these differences and which serves as the foundation for the model performing the subjects' identification. In this study, two data transformation techniques are pre-

sented, both of which preserve the inter-individual differences in brain connectivity while creating a linear pattern among all functional connectomes of the subjects. It is important to highlight that both techniques are mutually exclusive and have been implemented sequentially. The first technique, degree-normalization, extracts a common linear pattern to the FC values, thereby enabling the application of linear regression models. The second technique, residual decomposition, maintains the linearity of the transformation from resting-state to task state, while increasing variations among individuals' functional connectivity matrices. This means that the brain reorganization dynamics may not be similar among subjects, and the common linear pattern is just a result of applying the degree-normalization for the purpose of using a linear regression model to improve identification rates. Furthermore, it is important to note that even though data transformations can create a common linear pattern among functional connectomes of different individuals, it does not necessarily imply that the connectivity strengths, for the same link and for the network considered, are similar across subjects. Instead, it means that there is a common pattern of connectivity changes that occur from rest to task across individuals, but the strength of those changes can vary, confirming the brain fingerprint. The study by Tik et. al [68] is consistent with this idea, as it shows that even though a schizophrenic patient may have different connectivity strengths compared to control subjects, it still exhibits similar patterns of connectivity changes from rest to task as the controls. Additionally, if the objective of the study is to localize the actual activation regions during the task condition, instead of identifying individuals based on their FC values, it is necessary to transform the predictions by applying the inverse of degree-normalization and evaluate the performance of the model using MSE. This helps to ensure that the predictive values can be compared to the actual ones. However, since the aim of this study is to identify individuals based on their functional connectivity values, it is better to use correlation coefficient to improve the model's performance by capturing the linear relation existing between predicted and actual values. Nevertheless, the strength of correlation obtained by comparing FC data depends on several factors, including the number and distribution of NaN values resulting from leakage correction. In resting-state FC matrices, the numbers of NaN values is consistent within the same subject, but can vary between subjects due to differences in the structure of brains. Although the positions of NaNs are relatively similar across subjects, with similarity values ranging from 90% to 99% in most networks [10], the number of NaN values can reduce the number of valid links on which Pear-

Network	#NaNs	#Valid links	#Valid/#total links (%)
AllNtw	1423 ± 118	12107 ± 118	89.5 ± 0.9
AN	54 ± 2	66 ± 2	55 ± 1.7
CON	8 ± 0	13 ± 0	62 ± 0
DAN	59 ± 4	241 ± 4	80.3 ± 1.3
DMN	87 ± 4	319 ± 4	78.6 ± 1.0
FPN	19 ± 1	59 ± 1	75.6 ± 1.3
LN	25 ± 1	66 ± 1	72.5 ± 1.1
MN	95 ± 6	181 ± 6	65.6 ± 2.2
VAN	25 ± 2	80 ± 2	76.2 ± 1.9
VFN	50 ± 1	28 ± 1	35.9 ± 1.3
VPN	83 ± 6	70 ± 6	45.7 ± 3.9

Table 5.1: Comparison of the mean number of NaNs generated by leakage correction across subjects and the mean number of valid links used to compute the correlation coefficients across networks. In addition, the ratio (as a percentage) between the mean number of valid links to the total number of links is calculated. Values, together with their standard deviations, are subject to approximation.

son's correlation can be computed. To provide a more detailed understanding of network performance, Table 5.1 shows the average number of NaN values with their standard deviations and the number of links (and the corresponding ratio) on which correlation coefficients can be computed.

5.1.2 Evaluation of network and frequency band performance

In terms of the performance of frequency bands, α and β_{low} are found to be almost comparable across networks, which is consistent with their involvement in the processing of sensorimotor information [65, 66]. On the other hand, β_{high} returns the worst performance across networks, which may be due to its less prominent role in the execution of a specific motor task, as supported by Vettoruzzo and Fagotto's MSc thesis [10, 11] and [65]. In terms of network performance, it is important to consider several factors, such as the number of valid links to compute correlations and the strength of correlation between FC matrices from rest to task, after residualisation and normalisation steps. The best results are obtained with the VPN network, which suggests that the more parallel the regression lines extracted for each connectome from rest to task after degree-normalization, the

better the model performs. This can be visualized in Figure 4.5, where FC values cluster more tightly to the average of the regression lines extracted from each individual. Conversely, the worst network performance is observed in CON, although not reported in this study, due to the few nodes available to build regression lines and the small correlations observed between rest and task. Some individuals in CON even have negative slopes compared to most of the other regression lines, implying negative correlation with the prediction. Regarding AllNtw, the high number of links allows functional connectomes to be more distinguishable from the others. The other networks considered in this study show varying degrees of performance, with particular attention to MN network, which exhibit a sufficient number of valid links and a near-linear pattern between functional connectomes during rest and task, resulting in good accuracy rates.

5.2 Examination of performance based on model complexity

5.2.1 Linear regression models

Regarding model constructions, two hypotheses have been proposed to identify inter-individual variability in brain connectivity: one assumes a linear relationship between the functional connectomes during resting state and task conditions (4.1.1), while the other investigates the linear relationship between the FC values for each link of the network during the same states (4.1.2). The results of the study suggest that the link-by-link hypothesis, i.e. Model 3, does not properly capture the behaviour of brain connectivity. Instead, the first assumption appears to be the optimal approach for addressing the identification objective (Model 1 and Model 2). Compared to the other two models that assume the first hypotheses, the use of the link-by-link hypothesis increases the complexity of the model and shows lower performance, as demonstrated in Table 4.1. It is worth noting that these models show larger IDrate values than the ones proposed by Vettoruzzo, which assumes a link-by-link behaviour of the data. It is observed that Model 2 outperforms Model 1 as it is constructed solely by utilizing the independent residual components, without integrating the linearly dependant group-averaged components as built in Model 1.

5.2.2 Neural Network model

The question remains whether the relatively simplistic GLM (General Linear Model) approach [47], intended as the linear regression model, is the best choice for predicting brain connectivity. To explore more complex approaches, non-linear neural network models have been employed in this study. Several results presented in literature made use of NNs to map brain connectivity from resting-state to task conditions [69, 70]. To expand the study, a feed-forward neural network is implemented using the *feedforwardnet* function [71] with 1 hidden layer and 10 neurons, as suggested in a previous work by Cohen et al [72]. The neural network is used to map resting-state variation FC maps (residuals) to the corresponding maps during task. Although the neural network performs well, as reported in Table A.3 in the Appendix, Model 2 slightly outperforms it (Table 4.1). This lack of improvement in IDrate performance can be attributed to the fact that degree-normalization enforces a nearly linear pattern in the FC data, indicating that the Model 2 is capable to properly capture the relation between rest and task conditions. As a result, the use of the neural network, which is more complex and computational demanding than the regression model, does not significantly improve the performance compared to Model 2.

5.3 Assessment of EMG muscle activity

5.3.1 Evaluation of EMG data preprocessing

Regarding EMG signals, it is important to note that the preprocessing pipeline is designed to minimize the number of filtering operations and transformations to the data, in order to preserve as much of the original signal as possible and to maximize inter-individual variability among EMG features which are extracted. This is particularly important for finger tapping, as it is a simple task that does not require any additional preprocessing steps, beyond those described in Section 3.3. Applying further data transformations could result in a loss of information about the muscular activity and affect the feature extraction. For example, the Teager Kaiser Energy Operator[1] (TKEO), which was proposed in Fagotto's MSc

[1]Teager-Kaiser Energy Operator method applied to EMG signals is typically used to enhance the detection and analysis of muscle activity by estimating its instantaneous energy. The formula for TKEO applied to EMG signals is $\psi[x(n)] = x^2(n) - x(n-1) \cdot x(n+1)$ where $x(n)$ represents the

thesis [11], is not employed in this study because the most of the variability in EMG bursts is canceled out by applying it, and the onsets and offsets may not be detected or correspond to the real ones, as shown in Figure A.4 in the Appendix. It is worth noting that the number of repetitions considered during each motor session does not strictly adhere to 80 bursts, as subjects sometimes do not correctly follow the rules during the recording, such as by anticipating or postponing the movements, making more movements than necessary or missing some of them.

5.3.2 Inadequacy of EMG features in identifying subjects

The EMG features are extracted using an automatic detection method proposed by Fagotto but adjusted in terms of sliding window duration and definition of the offsets. Compared to Fagotto's work, different types and number of features have been extracted. To predict them from brain connectivity during resting-state or task state, it is necessary to extract robust quantities with low intra-subject variability compared to the inter-subject variability, which should be as high as possible. However, it can be challenging to find features that are both robust and have high inter-subject variability, particularly for a task like finger tapping as demonstrated in the EMG boxplot representations (4.11 and A.5). This was also confirmed by an investigation on the EMG features extracted in Fagotto's work[2]. The reason for the low performance of the LOOCV model applied to rMEG funtional connectivity and EMG features is mainly due to the strong assumption that there exists a clear and well-defined relationship between different rMEG FC matrices and their corresponding EMG features. Thus, it was not possible to consider EMG features as a reliable identifier or a motor fingerprint for individuals performing the finger tapping task, unlike FC matrices.

However, a significant finding by Vettoruzzo is the intra-subject consistency of spontaneous brain connectivity, where FC values were found to be similar among the three scans of resting state, as described before. This result enables to explore the relationship underlying the brain connectivity during rest and EMG features, by training models (Multiple Linear Regression and Neural Network) on two ses-

EMG amplitude at time n, and $\psi[x(n)]$ refers to the TKEO value at the same time. Compared with other methods, TKEO is relatively robust to noise and less sensitive to artifacts [73, 74].

[2]Zorzetto A. and Mohammad H. A. from the University of Padua conducted an investigation on the EMG features extracted in Fagotto's study, under the supervision of Prof. Cisotto. The analysis showed that most of the features had very small inter-subject variability, which supports the difficulty of finding robust features in a task like finger tapping.

sion of rest and verify the relation on the third session of resting state. The reason why the MLR model shows lower performance than the NN model is that the MLR assumes a linear relationship between the predictors and the targets for all subjects, employing a single vector of regression coefficients for all functional connectomes. In contrast, the NN model assumes a non-linear relation between the input and output, leading to a better performance.

5.4 Limitation of the study

This study presents a limitation in the treatment of NaN (Not a Number) values in neural networks. The practice of replacing NaNs with 0 was implemented across the different NNs, following the approach used in the MLR developed by Fagotto. This assumption poses a potential challenge that needs to be considered when interpreting the results of the NN models. However, this assumption was not used in other models such as Model 1, Model 2, or Model 3. Two reasons make this hypotheses necessary. Firstly, due to the inter-subject variability in the position of NaNs, their removal makes it impossible to train a NN model built on the connectivities of the other $N - 1$ subjects who do not share the same position of NaNs. Secondly, the performance of the NN model trained and tested on a smaller FC pattern, created by removing all FC values associated to a link containing NaNs for a particular subject, is very low[3]. Nevertheless, it is important to note that the application of leakage correction on the data was intended to prevent the possibility that closer nodes may be more likely influenced by spurious noise or artifact than more distantly located nodes [54]. By removing the link connecting these nodes, the influence of the co-dependence of closer nodes on the performance can be discarded. The assumption made in this study is that the leakage-correction is not fundamentally different from considering these nodes as independent of each other, with their connectivity strength (i.e. correlation) being zero. In graph theory context, brain functional connectivity can be represented by an adjacency matrix, where the elements indicate the connection strengths between the nodes. If an element is set to 0, it means that there is no connection between the corresponding nodes. This would imply that the connectivity strength of closer nodes does not affect the model performance, which aligns with the intended effect

[3]This was also proved by Zorzetto A. and Mohammad H. A. using Fagotto's MLR model

of applying leakage correction to the outset. While the assumption made in the implementation of the NN model may help to overcome certain limitations of the data, it is not an ideal solution in this particular domain of application. Valid oversampling approaches among links of subjects could provide a more robust and accurate approach to dealing with these issues.

Although not considered a limitation, there is an important aspect of Model 2 that requires clarification. Specifically, the correlation between the actual and predicted task FC variation maps in Model 2 is equivalent to the correlation between the actual rest and task FC variation maps, up to a positive or negative sign. This relationship can be observed by comparing the IDmatrices of Model 2 in Figure 4.10 with the rest-task correlation matrices in Figure 4.7. This is a consequence of the mathematical properties of Pearson's correlation coefficient. Specifically, when considering two general sets of data, Y and $\widehat{Y}$, with $\widehat{Y}$ lying on the line described by $\widehat{Y} = aX + b$, where a and b are the slope and the intercept term respectively, it can be demonstrated that $r_{\widehat{Y}Y} = \pm r_{XY}$ according to whether the slope a is positive or negative, as shown in [75]. Similarly, the predicted task FC residual values lie on the line described in equation 3.19, which is obtained by averaging the regression coefficients over $N-1$ subjects, and subsequently will be compared with the actual ones through Pearson's correlation. As a result, in Model 2, computing the Pearson correlation coefficient between the actual rest and task FC variation maps is sufficient to identify subjects. However, it is important to highlight that this property is a specific characteristic of Pearson's correlation coefficient and does not reflect the accuracy of predicting the FC variation maps during the task state. Different predicted regression lines will result in different MSE values, which indicate the degree of accuracy in predicting the residual FC maps. Additionally, it is important to emphasize that these findings are specific to Model 2 and do not apply to Model 1 or Model 3, as these models have different implementations.

6

Conclusions

The present work represents a methodological exploration into the inter-individual variability of functional connectomes and muscular activity for a cohort of 51 individuals from the Human Connectome Project. Through a comparison of several regression models, including neural networks, the goal is to predict task-evoked brain response and muscular activity from resting-state functional connectivity. The results suggest that linear regression models are capable of fulfilling this task, particularly when the residualisation is employed to decompose functional connectivity values. The α and β_{low} frequency bands show the highest performance in line with the domain previous knowledge [66, 65], and certain networks, such as All-Ntw, DMN, MN and VPN, consistently perform well across models (identification rates greater than 70%), as shown in Table 4.1. On the other hand, the prediction of muscular activity from brain connectivity does not produce satisfactory results due to the low inter-subject variability of EMG features. Nonetheless, the analysis highlights the importance of identifying the relationship between functional connectomes and muscle activity, particularly in a new unseen session of resting-state. Neural networks appear to be more effective in capturing the non-linearity of this relationship.

Future research could focus on improving these approaches by better understanding the underlying neural mechanisms and processes. For instance, linearly independent components, like the features used as predictors by Tavor et al. (2016), could be used to decompose FC patterns and produce a prediction of the FC task-evoked response. Similarly in the muscular activity, using Discrete Wavelet Transform (DWT) to extract the approximation and detailed coefficients from the entire time series of the EMG signal could lead to better performance when predicting muscle activity based on brain connectivity, in line with [62].

Additionally, dealing with the NaN values present in the dataset to be analyzed is important for accurate and reliable analyses. Oversampling or undersampling methods can be effective strategies to address this issue and prevent biases that may affect the generalizability of the results. These techniques can enhance the robustness of the findings and increase the validity of the conclusions drawn from the data.

Overall, this work, conducted as part of the ERC *HANDmade* project, represents a small step towards better understanding the individual variability in brain organization and its relationship with motor functions. This can have implications in clinical research for comprehending the neural mechanisms underlying various neurological and muscular disorders and for developing personalized treatment approaches [68]. Secondly, the study highlights the importance of understanding the relationship between functional connectomes and muscular activity, which can be used to develop more effective rehabilitation and intervention strategies for patients with motor impairments. Finally, the study emphasizes the need for continued efforts to improve the understanding of inter-individual variability in brain functions and muscular activity, as well as the methods for predicting and characterizing this variability. This can lead to the development of more precise and effective personalized medicine approaches for various neurological and muscular disorders.

A

Appendix

A.1 Brain Networks

The brain division into 10 different functional networks is displayed in Figures A.1 and A.2. Each network serves a different cognitive or behavioral function and interacts with other networks to support overall brain function. The visualization of these networks using BrainNet Viewer [76] allows for a better understanding of the organization of the brain and the interactions between different regions.

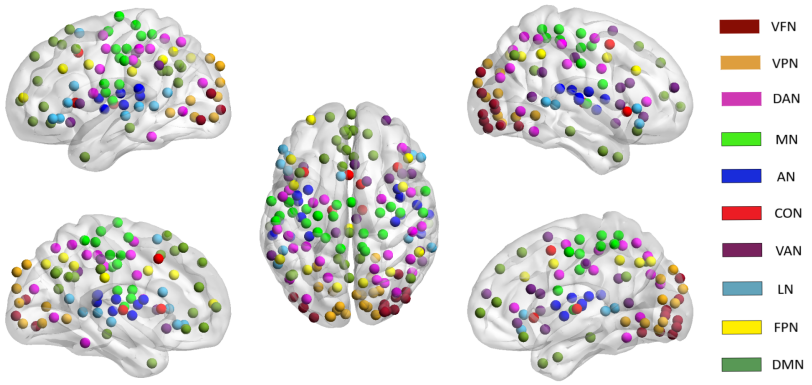
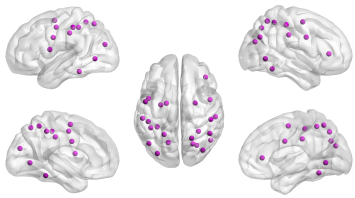


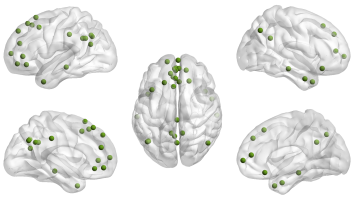
Figure A.1: 164-ROIs parcellation network (AllNtw) using BrainNet Viewer

(a) AN

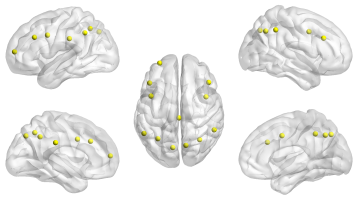
(b) CON



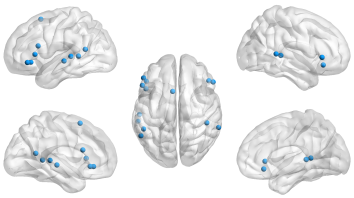
(c) DAN



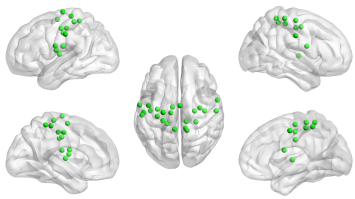
(d) DMN



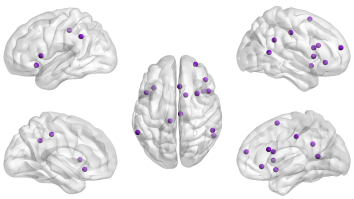
(e) FPN



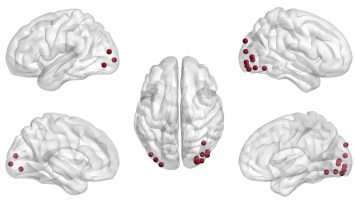
(f) LN



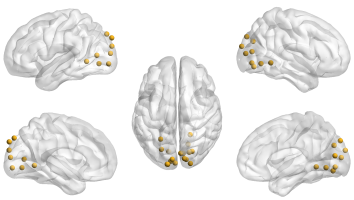
(g) MN



(h) VAN



(i) VFN



(j) VPN

Figure A.2: Functional brain networks visualized using BrainNet Viewer

A.2 Supplementary material on EMG activity

A.2.1 Further details about EMG onset/offset detection

This section provides a visual representation of the method used for detecting the onsets and offsets of EMG signal. Specifically, the method is demonstrated considering subject '105923' and right finger tapping (RH) in a block of 10 repetitions for the same movements. Figure A.3 displays the filtered signal, which is preprocessed, using a Butterworth 4° filter with cutoff frequencies [15,450] Hz, as well as two Notch filters with cutoff frequencies 60 Hz and 180 Hz. Figure A.3 shows also the rectified signal, which is obtained using the method described in Section 3.3.1, and highlights the extraction of EMG signal onsets/offsets using the threshold values thr_1 and thr_2 (see Section 3.3.1).

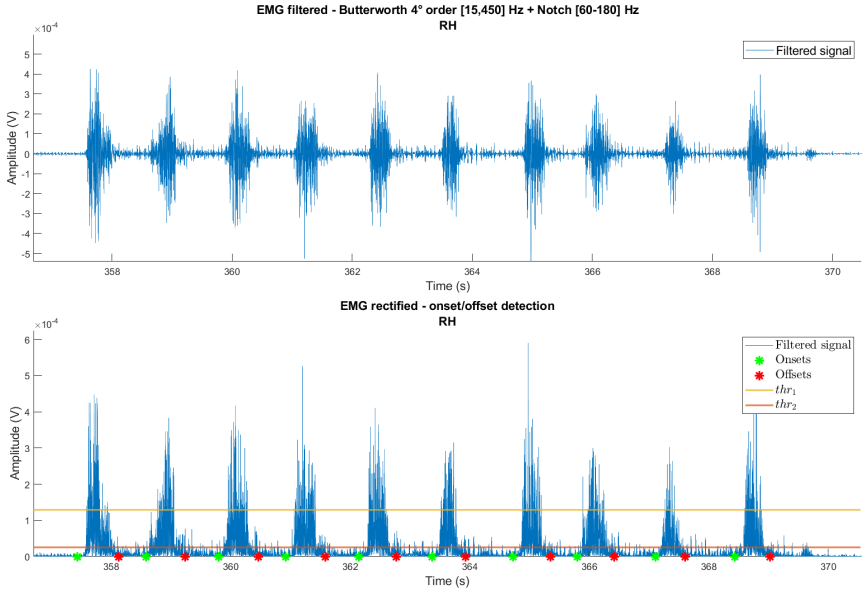


Figure A.3: In the upper panel, the filtered EMG signal for subject '105923' during right finger tapping (RH). In the lower panel, the rectified EMG signal for the same subject and task showing the method for extracting the onsets/offsets of EMG bursts using the thresholds thr_1 and thr_2.

Additionally, Figure A.4 illustrates the application of Teager-Kaiser Energy operator to the EMG data. However, this operation is not performed in the current study due to its potential to cause a loss of valuable information in signal, which could decrease the inter-individual variability of features and standardize burst patterns. Moreover, the automatic method for detecting onsets and offsets of

EMG bursts may fail in some cases due to the lower signal energy within the bursts.

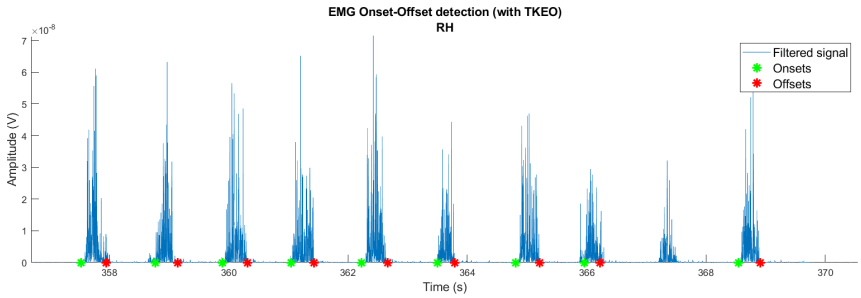


Figure A.4: Rectified EMG signal for subject '105923' during right finger tapping, after the application of Teager-Kaiser Energy Operator. TKEO may lead to a standardization of EMG bursts and, in addition, the automatic method may fail to detect onsets/offsets due to the lower signal energy within the bursts.

A.2.2 Additional EMG Features

In the present study, a total of 40 EMG features have been extracted from the segmented EMG signals. These features are commonly used in the literature to describe various aspects of EMG signals. Among these 40 EMG features, 10 time-domain features are selected as the main focus of the study. Additionally, the remaining 30 ones, are used to enforce the robustness of the results and to reject the possibility of bias occurrence. These additional features, together with the most relevant ones, are included in the EMG Feature Extraction Toolbox project available in the GitHub platform of Jingwei Too [61]. Researchers interested in implementing these quantities can refer to the GitHub project for guidance on how to compute them. In Table A.2, a list of these 30 additional EMG quantities, numbered from 11 to 40, is presented. The figure A.5 shows the inter-individual variability of the additional 30 EMG features studied, as represented by boxplots. Each boxplot contains 51 values, one for each of the 51 subjects studied. The boxplot for a given EMG feature represents the distribution of values for that feature across all subjects. The figure notes that for most of the 30 EMG features studied, the inter-individual variability is low, meaning that the right hand muscle activity cannot be discriminated among subjects. This indicates that the extracted EMG features during the motor task studied may not be useful in distinguishing individuals properly, as they are too similar among subjects.

No.	Abbreviations	EMG features
11	ASS	Absolute Value of Summation of Square Root
12	ASM	Absolute Value of Summation ofăexp root
13	AR	Auto-Regressive Model
14	CARD	Cardinality
15	DAMV	Difference Absolute Mean Value
16	DASDV	Difference Absolute Standard Deviation Value
17	DVARV	Difference Variance Value
18	EMAV	Enhanced Mean absolute value
19	EWL	Enhanced Wavelength
20	IQR	Interquartile Range
21	LCOV	Log Coefficient of Variation
22	LD	Log Detector
23	LDAMV	Log Difference Absolute Mean Value
24	LDASDV	Log Difference Absolute Standard Deviation
25	LTKEO	Log Teager Kaiser Energy Operator
26	MFL	Maximum Fractal Length
27	MAD	Mean Absolute Deviation
28	MSR	Mean Value of Square Root
29	MMAV	Modified Mean Absolute Value
30	MMAV2	Modified Mean Absolute Value 2
31	MYOP	Myopulse Percentage Rate
32	FZC	New Zero Crossing
33	SSI	Simple Square Integral
34	SSC	Slope Sign Change
35	TM	Temporal Moment
36	VARE	Variance of EMG
37	VO	V-Order
38	WL	Waveform Length
39	WA	Willison Amplitude
40	ZC	Zero Crossing

Table A.2: Additional 30 EMG features

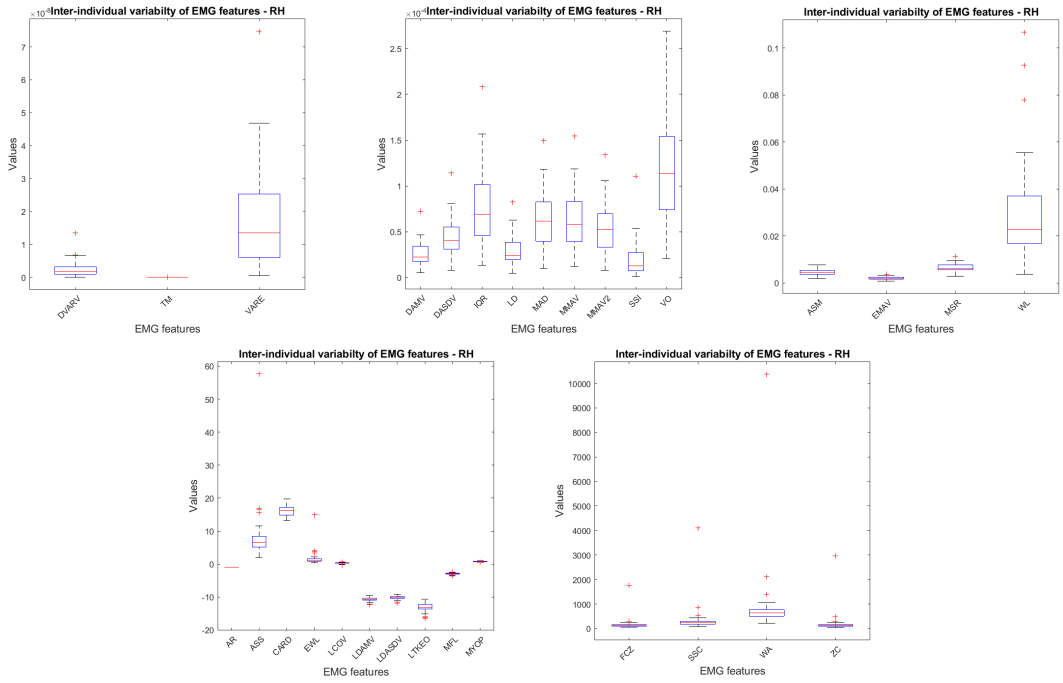


Figure A.5: Boxplots of additional 30 EMG feature values across individuals. The central mark in each boxplot refers to the median, while the bottom and the top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points that are not considered outliers, which are plotted individually using the '+' marker symbol.

A.3 Supplementary results on connectivity

A.3.1 Relation between rest and task functional connectivity

This section presents additional findings on the linear transformation between resting-state and task-based MEG functional connectomes, as well as the link-by-link transformation. The impact of applying degree-normalization and residualisation techniques on the transformed data is also examined. The results provide insights into the consistency and variability of the connectomes across different brain networks, as well as the potential utility of using such transformations to increase inter-individual variability and improve subject's identification. Figures A.6 and A.8 display the matrices of correlation obtained between functional connectomes during resting-state and task condition in seven brain networks (AN, DAN, DMN, FPN, LN, VAN, and VFN) for the frequency bands α, β_{low} and β_{high}, without degree-normalization of the FC values. The corresponding matrices after applying degree-normalization to the FC values are depicted in Figure A.7 and A.9. The same correlation matrices for group-averaged and residual components are shown in Figures A.10 and A.11. Finally, Figure A.12 displays the distribution of correlation between group-averaged FC values (after degree-normalization) for AllNtw in α, β_{low}, and β_{high}. As shown, the mean correlation is very high compared to the one extracted from residual FC values.

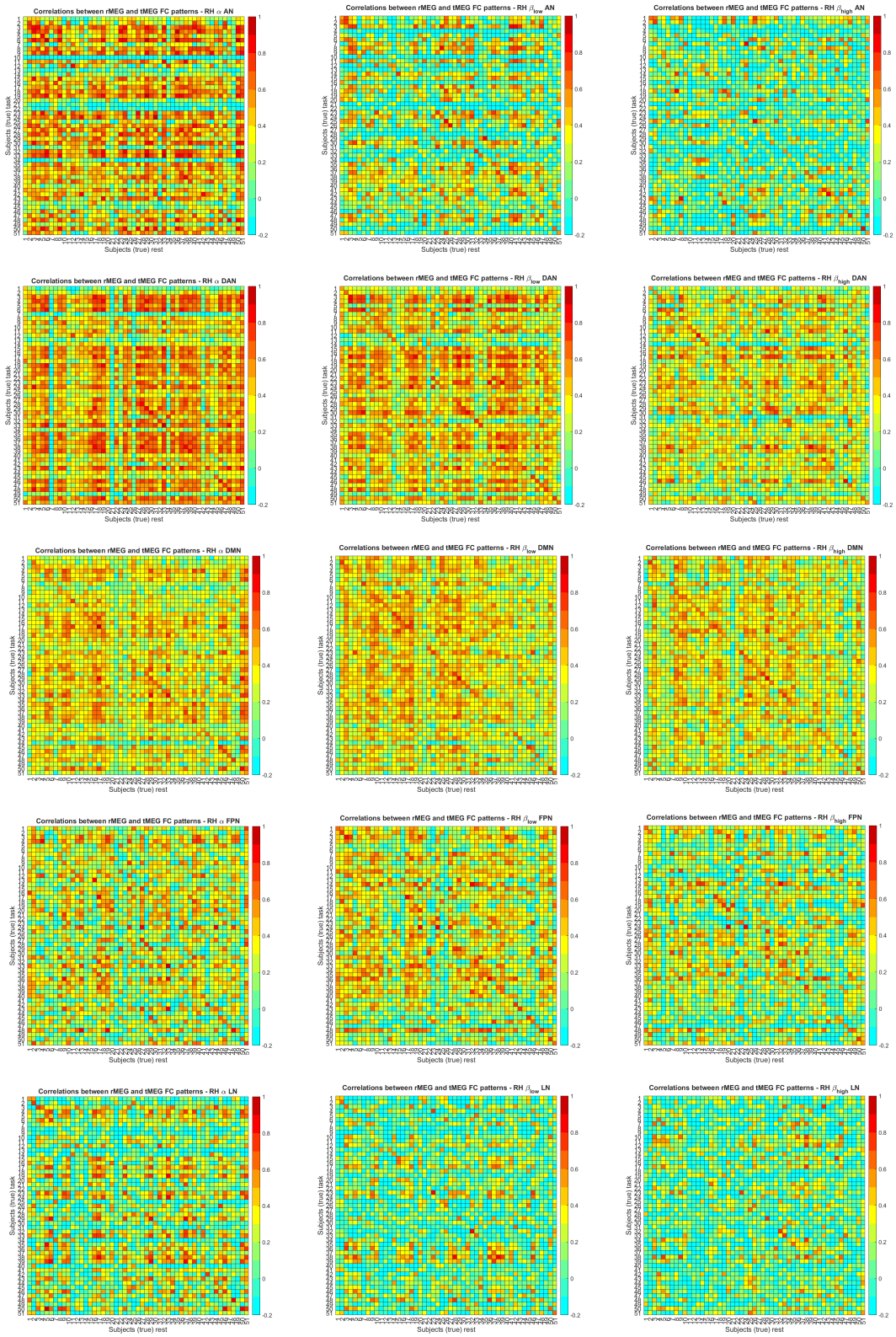


Figure A.6: Matrices of correlation coefficients between original rMEG and tMEG functional connectivities to the input data, for the other brain networks (AN, DAN, DMN, FPN, and LN) and three frequency bands.

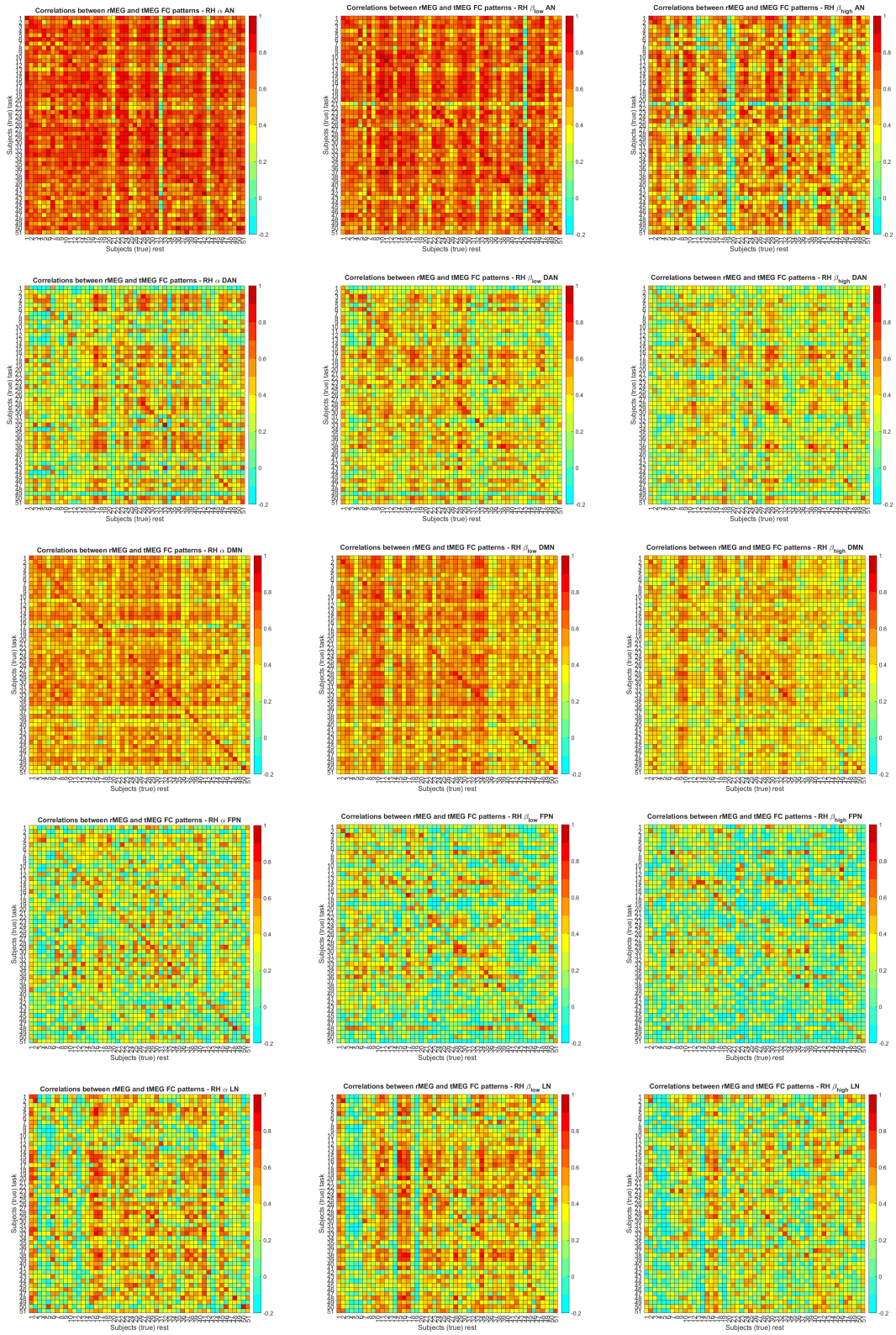


Figure A.7: Matrices of correlation coefficients between rMEG and tMEG functional connectivities after applying degree-normalization to the input data, for the other brain networks (AN, DAN, DMN, FPN, and LN) and three frequency bands.

Figure A.8: Matrices of correlation coefficients between original rMEG and tMEG functional connectivities to the input data, for the other brain networks (VAN and VFN) and three frequency bands.

Figure A.9: Matrices of correlation coefficients between rMEG and tMEG functional connectivities after applying degree-normalization to the input data, for the other brain networks (VAN and VFN) and three frequency bands.

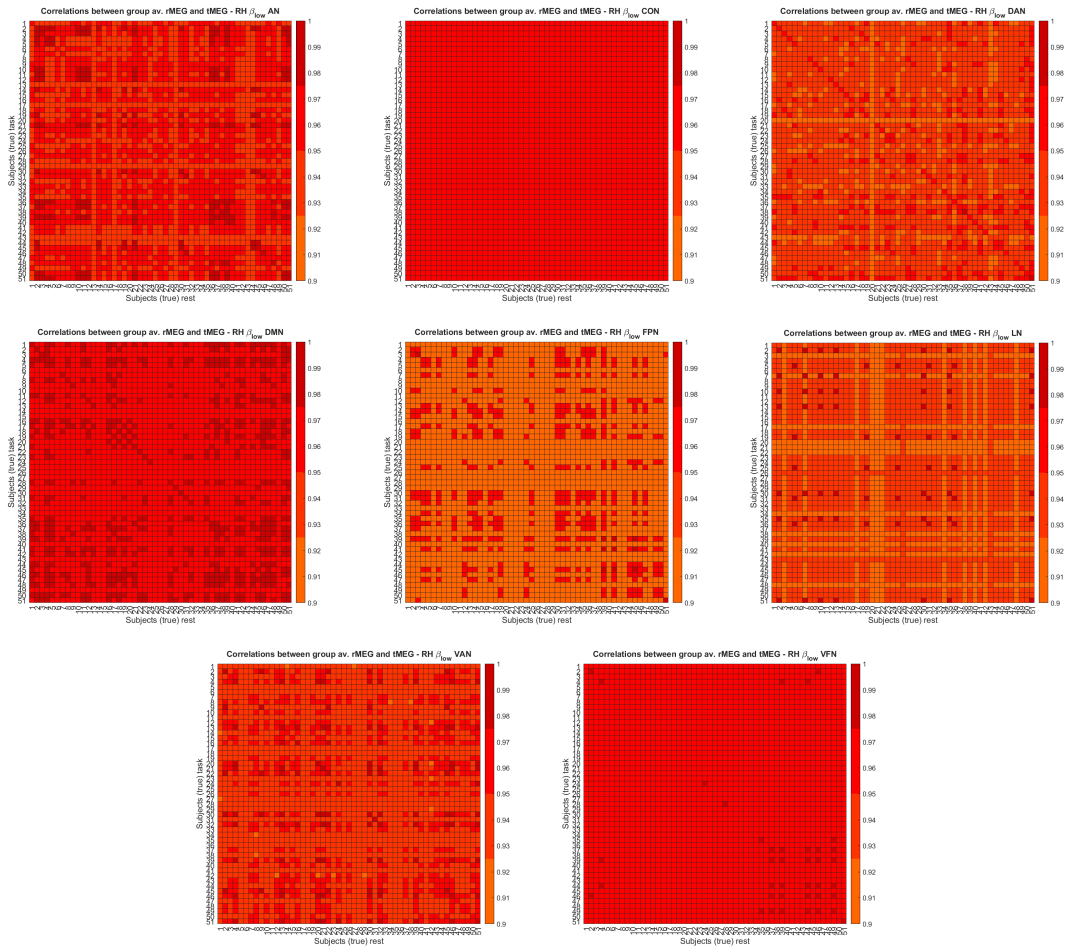


Figure A.10: Matrices of correlation coefficients between group-averaged rMEG and tMEG functional connectivities to the input data, for the other brain networks (AN, CON, DAN, DMN, FPN, LN, VAN, and VFN) and β_{low} frequency band.

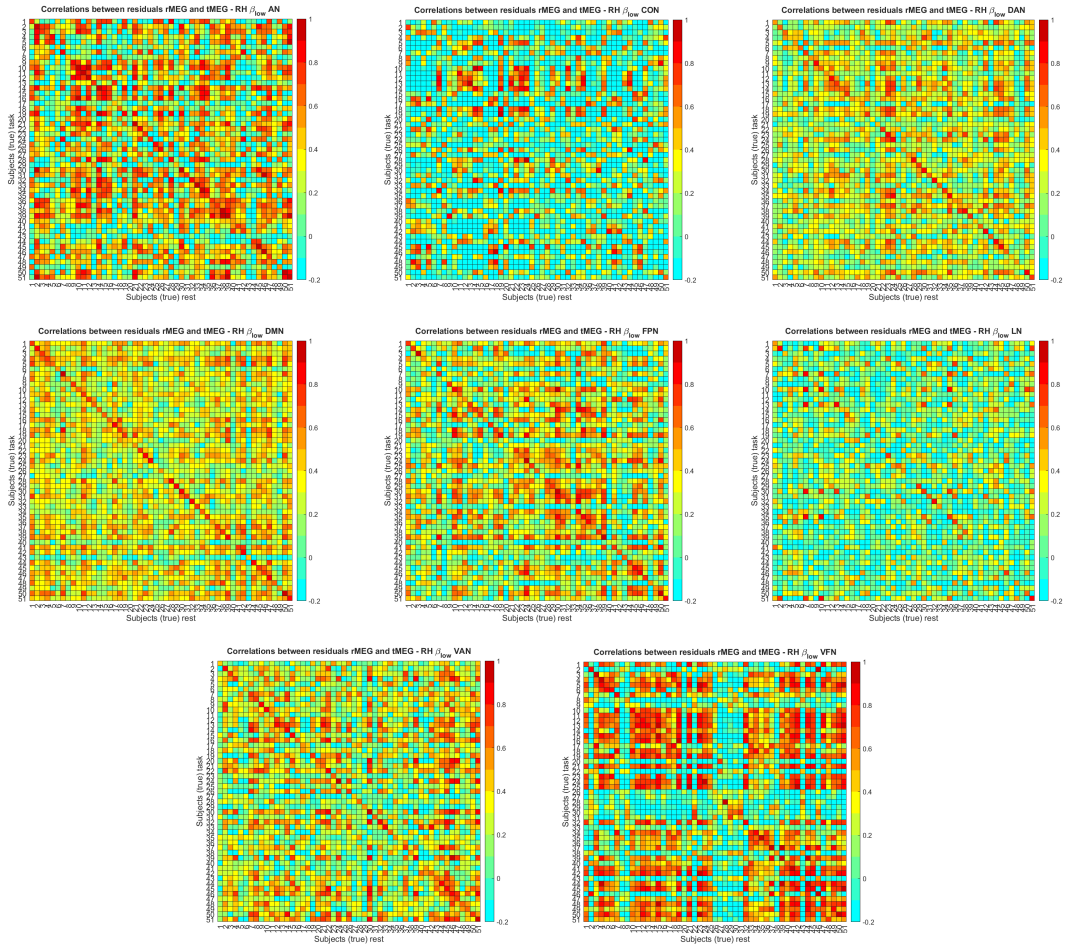


Figure A.11: Matrices of correlation coefficients between residual rMEG and tMEG functional connectivities to the input data, for the other brain networks (AN, CON, DAN, DMN, FPN, LN, VAN, and VFN) and β_{low} frequency band.

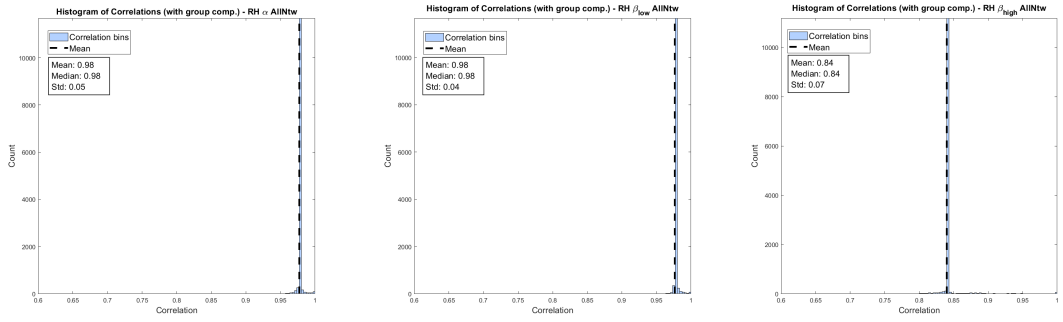


Figure A.12: Distribution of correlation between group-averaged FC values (after degree-normalization) for AllNtw in α, β_{low}, and β_{high}. As shown, the mean correlation is very high compared to the one extracted from residual FC values.

A.3.2 Modelling task from rest-based functional connectivity

This section displays additional results regarding the application of the three machine learning-based models to predict the individual task maps based on the resting-state functional connectivity. Supplementary results investigates the performance of the Model 2 with respect to residualised, normalised and original data. Figures A.14 and A.13 depict the IDmatrices obtained with the three models for β_{low} frequency band, and seven networks (AN, DAN, DMN, FPN, LN, VAN and VFN). The performance evaluation of Model 2 with residualised, normalised, and original data for the β_{low} frequency band, and four networks (AllNtw, CON, MN, and VPN) is visualized in the IDmatrices of Figure A.15.

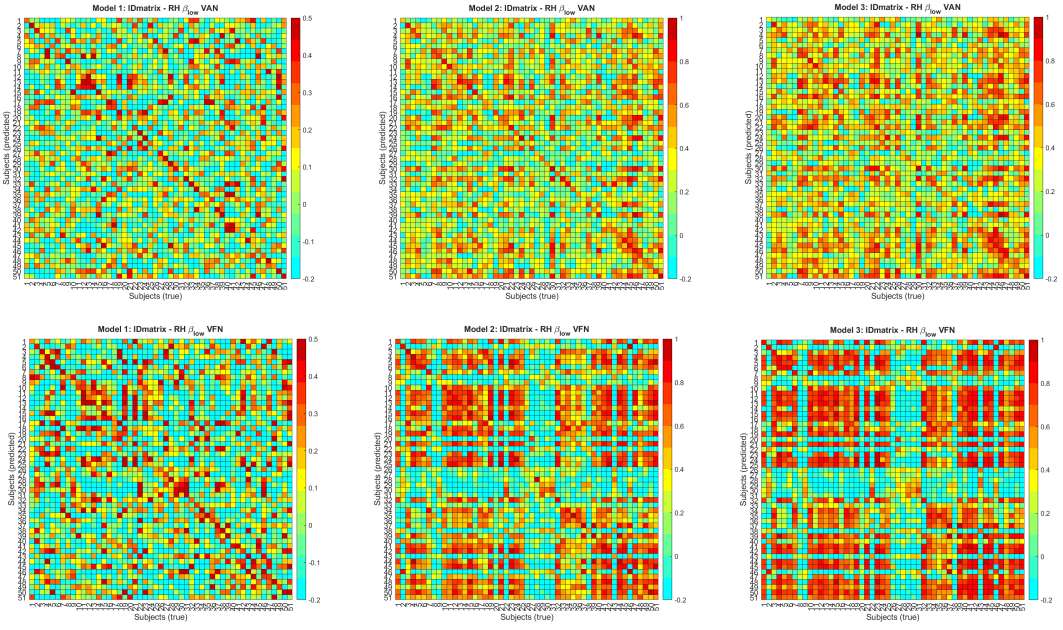


Figure A.13: Additional IDmatrices obtained with the three models for β_{low} frequency band, and networks (VAN and VFN). Each cell in the matrix represents the correlation coefficient between the predicted tMEG FC (rows) and the actual connectivity values (columns) for the same subject (diagonal elements) and for different subjects (off-diagonal elements).

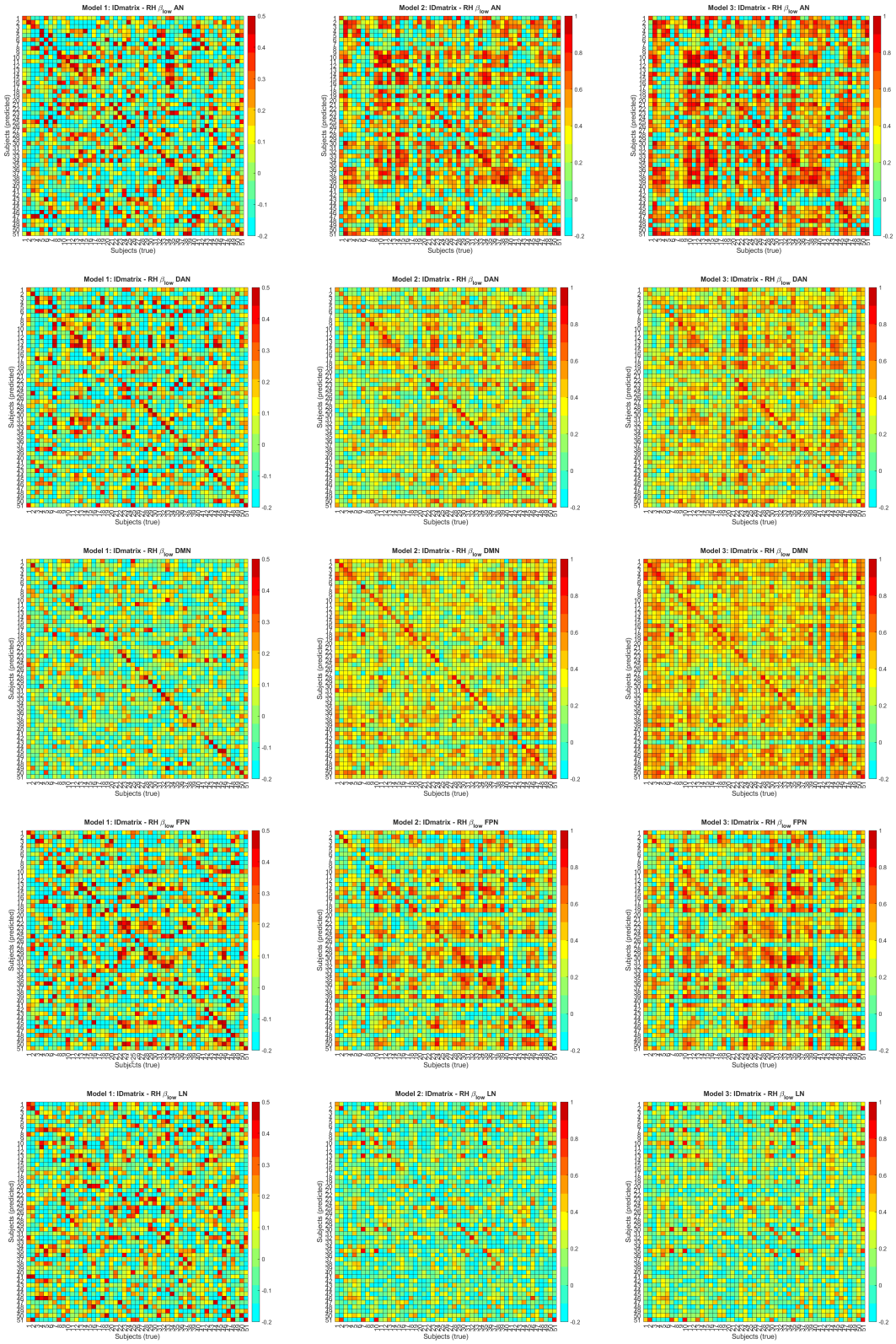


Figure A.14: Additional IDmatrices obtained with the three models for β_{low} frequency band, and networks (AN, DAN, DMN, FPN and LN). Each cell in the matrix represents the correlation coefficient between the predicted tMEG FC (rows) and the actual connectivity values (columns) for the same subject (diagonal elements) and for different subjects (off-diagonal elements).

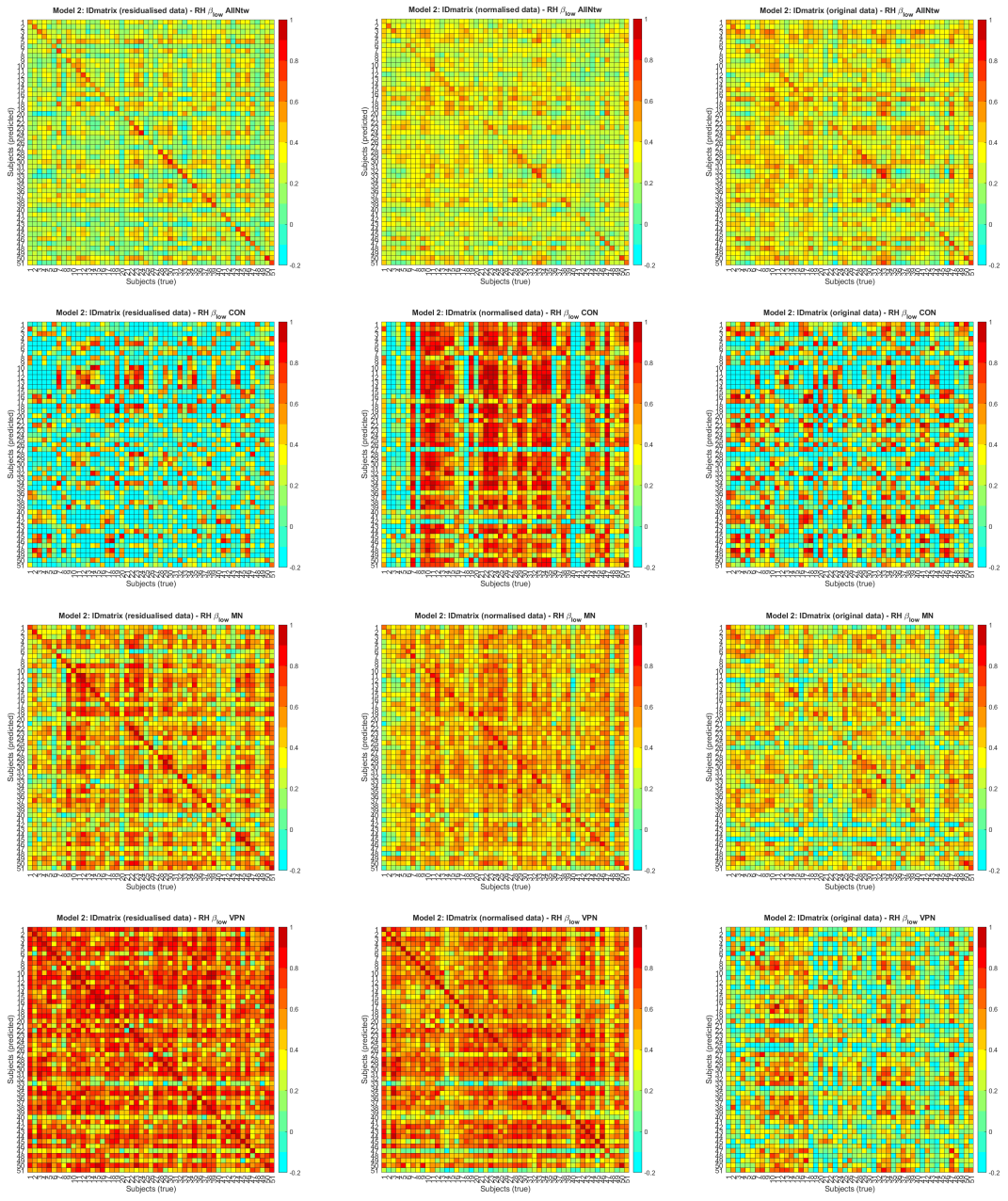


Figure A.15: IDmatrices obtained with the Model 2 trained respectively with residualised, normalised, and original data (respectively column 1,2 and 3) for the β_{low} frequency band, and four networks (AllNtw, CON, MN, and VPN).

A.3.3 Assessing the inadequate models performance in predicting muscle activity

In this section, the performance of the models developed to predict muscle activity using rMEG-based functional connectivity is presented. Specifically, these results focus on the inadequacy of the LOOCV linear regression and feed-forward neural network models, described in Section 3.4.2, to accurately predict individual muscle activity based on the further 30 extracted features, presented in Table A.2. Figure A.16 and Figure A.17 present the IDmatrices, resulting from the linear regression and feed-forward neural network models respectively, which displays the poor correlation between the predicted and actual EMG values.

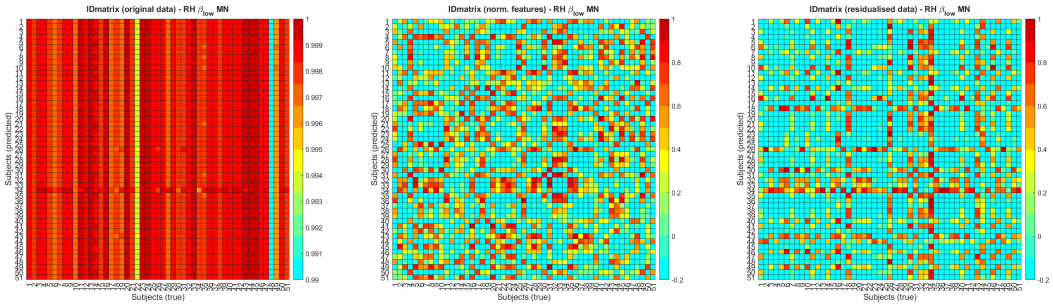


Figure A.16: IDmatrices obtained using a LOOCV model based on linear regression, trained on rMEG and 30 EMG features in MN and β_{low}. Even applying data transformations such as normalization of EMG features or residual decompositions of both rMEG and EMG features, there is no evidence of a diagonal structure.

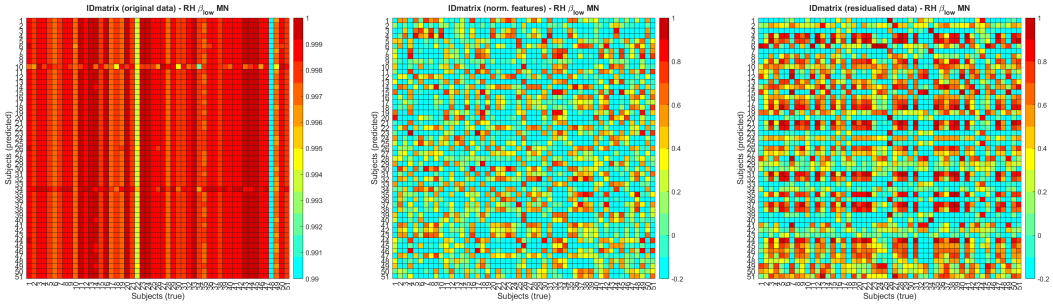


Figure A.17: IDmatrices obtained using a NN model based on *fitrnet* MATLAB function, trained on rMEG and 30 EMG features in MN and β_{low}. Despite applying data transformations such as normalization of EMG features or residual decompositions of both rMEG and EMG features, there is no evidence of a diagonal structure.

NN model							
Network	Band	$Pscore$	$IDrate$	Network	Band	$Pscore$	$IDrate$
AllNtw	α	0.62 ± 0.13	90.20%	LN	α	0.53 ± 0.21	29.41%
	β_{low}	0.58 ± 0.24	84.31%		β_{low}	0.45 ± 0.29	39.22%
	β_{high}	0.48 ± 0.22	80.39%		β_{high}	0.34 ± 0.30	25.49%
AN	α	-0.69 ± 0.30	1.96%	MN	α	0.77 ± 0.12	74.51%
	β_{low}	0.69 ± 0.17	9.8%		β_{low}	0.81 ± 0.11	72.55%
	β_{high}	0.23 ± 0.20	3.92%		β_{high}	0.74 ± 0.13	66.67%
CON	α	0.37 ± 0.39	17.65%	VAN	α	0.7 ± 0.16	56.86%
	β_{low}	0.37 ± 0.37	11.76%		β_{low}	0.69 ± 0.16	37.25%
	β_{high}	0.31 ± 0.42	11.76%		β_{high}	0.52 ± 0.25	37.25%
DAN	α	0.64 ± 0.17	52.94%	VFN	α	0.72 ± 0.20	33.33%
	β_{low}	0.65 ± 0.17	54.9%		β_{low}	0.73 ± 0.22	31.37%
	β_{high}	0.57 ± 0.18	56.86%		β_{high}	0.61 ± 0.25	35.29%
DMN	α	-0.67 ± 0.25	1.96%	VPN	α	0.95 ± 0.07	94.12%
	β_{low}	0.70 ± 0.16	74.51%		β_{low}	0.94 ± 0.09	90.20%
	β_{high}	0.58 ± 0.17	76.47%		β_{high}	0.89 ± 0.17	86.27%
FPN	α	-0.66 ± 0.26	1.96%				
	β_{low}	0.68 ± 0.19	49.02%				
	β_{high}	0.56 ± 0.2	35.29%				

Table A.3: Pscore and IDrate for each network and for each frequency band, obtained by the neural network model trained on FC variation maps. The neural network is implemented by the built-in *feedforwardnet* MATLAB function, containing one hidden layer with 10 neurons each. The model presents a slightly lower performance compared to the Model 2 reported in Table 4.1.

A.3.4 Neural network performance in predicting task-based brain connectivity

Table A.3 presents the IDrate and p-score values obtained by training a feed-forward neural network with one hidden layer and 10 neurons on the FC variation maps during rest and task conditions. The implementation of the network uses the built-in MATLAB function *feedforwardnet*. The training process is similar to that of Model 2, with the exception that the averages of biases and weights across subjects are computed instead of the average of regression coefficients of $N-1$ individuals. Nevertheless, despite the neural network's increased complexity and longer computational time, its performance is slightly lower than that of Model 2.

References

[1] Nitin Williams and Richard N. Henson. "Recent advances in functional neuroimaging analysis for cognitive neuroscience". In: *Brain and Neuroscience Advances* 2 (Jan. 2018), p. 239821281775272. DOI: 10.1177/2398212817752727. URL: https://doi.org/10.1177/2398212817752727.

[2] Dimitri Van De Ville et al. "When makes you unique: Temporality of the human brain fingerprint". In: *Science Advances* 7.42 (Oct. 2021). DOI: 10.1126/sciadv.abj0751. URL: https://doi.org/10.1126/sciadv.abj0751.

[3] Emily S Finn et al. "Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity". In: *Nature Neuroscience* 18.11 (Oct. 2015), pp. 1664–1671. DOI: 10.1038/nn.4135. URL: https://doi.org/10.1038/nn.4135.

[4] Michael W. Cole et al. "Intrinsic and Task-Evoked Network Architectures of the Human Brain". In: *Neuron* 83.1 (July 2014), pp. 238–251. DOI: 10.1016/j.neuron.2014.05.014. URL: https://doi.org/10.1016/j.neuron.2014.05.014.

[5] I. Tavor et al. "Task-free MRI predicts individual differences in brain activity during task performance". In: *Science* 352.6282 (Apr. 2016), pp. 216–220. DOI: 10.1126/science.aad8127. URL: https://doi.org/10.1126/science.aad8127.

[6] Ekansh Sareen et al. "Exploring MEG brain fingerprints: Evaluation, pitfalls, and interpretations". In: *NeuroImage* 240 (2021), p. 118331. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2021.118331. URL: https://www.sciencedirect.com/science/article/pii/S1053811921006078.

[7] Feb. 2023. URL: https://erc.europa.eu/homepage.

[8] URL: https://cordis.europa.eu/project/id/759651.

[9] HCP WU-Minn. *1200 subjects data release reference manual*. URL: https://humanconnectome.org/storage/app/media/documentation/s1200/HCP_S1200_Release_Appendix_III.pdf.

[10] Anna Vettoruzzo. "Modelling MEG-based resting-state functional connectivity to predict inter-individual differences in motor tasks". MA thesis. University of Padua, 2021.

[11] Jessica Fagotto. "Modelling MEG-based functional connectivity and EMG activity to predict inter-individual differences in motor tasks". MA thesis. University of Padua, 2022.

[12] Benjamin Chiêm et al. "Improving Functional Connectome Fingerprinting with Degree-Normalization". In: *Brain Connectivity* (Aug. 2021). DOI: 10.1089/brain.2020.0968. URL: https://doi.org/10.1089/brain.2020.0968.

[13] *Anatomy of the Brain*. MayfieldClinic. 2018. URL: https://mayfieldclinic.com/.

[14] M. Hunter Manasco. "Basic Brain Anatomy". In: *Introduction to Neurogenic Communication Disorders*. 2013, pp. 13–22.

[15] Jiawei Zhang. *Basic Neural Units of the Brain: Neurons, Synapses and Action Potential*. 2019.

[16] Peyman Adjiamian. "MEG: technical improvements in image co-registration and studies of visual cortical oscillations". PhD thesis. 2002, pp. 21–30.

[17] Joonas Iivanainen. "On-scalp magnetoencephalography: Theory, implementation and measurements". PhD thesis. 2020, pp. 15–17.

[18] LibreTexts. "The Action Potential and Propagation". In: *Anatomy and Physiology (Boundless)*. 2020.

[19] Malcolm Proudfoot et al. *Practical Neurology*. 2014, pp. 336–343.

[20] M. Hämäläinen et al. *Magnetoencephalographytheory, instrumentation, and applications to noninvasive studies of the working human brain*. 1993, p. 413.

[21] Sylvain Baillet. "Magnetoencephalography for brain electrophysiology and imaging". In: *Nature Neuroscience* (2017).

[22] Abhang Priyanka A. *Introduction to EEG- and Speech-Based Emotion Recognition*. 2016, pp. 51–79.

[23] Seager Matthew A. et al. *Oscillatory brain states and learning: Impact of hippocampal theta-contingent training*. 2002, pp. 1616–1620.

[24] Koudelková Zuzana et al. "Analysis of brain waves according to their frequency". In: *INTERNATIONAL JOURNAL OF BIOLOGY AND BIOMEDICAL ENGINEERING* (2018).

[25] Jia X. et al. "Gamma rhythms in the brain". In: *Biology* (2011).

[26] Simon B. Eickhoff. *Imaging-based parcellations of the human brain*. 2018.

[27] Olaf Sporns. *Structure and function of complex brain networks*. 2013, pp. 247–250.

[28] Olaf Sporns, Giulio Tononi, and Rolf Kötter. "The Human Connectome: A Structural Description of the Human Brain". In: *PLoS Computational Biology* 1.4 (2005), e42. DOI: 10.1371/journal.pcbi.0010042. URL: https://doi.org/10.1371/journal.pcbi.0010042.

[29] Michael D. Fox and Marcus E. Raichle. "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging". In: *Nature Reviews Neuroscience* 8.9 (Sept. 2007), pp. 700–711. DOI: 10.1038/nrn2201. URL: https://doi.org/10.1038/nrn2201.

[30] Emily S Finn et al. "Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity". In: *Nature Neuroscience* (2015).

[31] Smith SM et al. "Network modeling methods for FMRI". In: *Neuroimage* (2011).

[32] Joseph J. Kuiper et al. "A parcellation-based model of the auditory network". In: *Hearing Research* (2011).

[33] B. T. Thomas Yeo et al. "The organization of the human cerebral cortex estimated by intrinsic functional connectivity". In: *Journal of Neurophysiology* (2011).

[34] Christiane S. Rohr et al. "Functional Connectivity of the Dorsal Attention Network Predicts Selective Attention in 47 year-old Girls". In: *Cerebral Cortex* (2017).

[35] Hideya Koshino et al. "Coactivation of the Default Mode Network regions and Working Memory Network regions during task preparation". In: *Scientific Reports, Natureserach* (2014).

[36] Scott Marek et al. “The frontoparietal network: function, electrophysiology, and importance of individual precision mapping”. In: *Dialogues in Clinical Neuroscience* (2018).

[37] Ingo Hertrich et al. “The Margins of the Language Network in the Brain”. In: *Language Sciences* (2020).

[38] Hans-Peter Müller et al. “Motor network structure and function are associated with motor performance in Huntingtons disease”. In: *Motor network structure and function are associated with motor performance in Huntingtons disease* (2016).

[39] Simone Vossel, Joy J Geng, and Gereon R Fink. “Dorsal and ventral attention systems: distinct neural circuits but collaborative roles”. en. In: *Neuroscientist* 20.2 (Apr. 2014), pp. 150–159.

[40] Emma E. M. Stewart, Matteo Valsecchi, and Alexander C. Schütz. “A review of interactions between peripheral and foveal vision”. In: *Journal of Vision* 20.12 (Nov. 2020), pp. 2–2. ISSN: 1534-7362. DOI: 10.1167/jov.20.12.2.

[41] *Visual cortex*. Wikipedia. 2022. URL: https://en.wikipedia.org/wiki/Visual_cortex.

[42] Gordana Sendic MD. *Musculoskeletal system*. Oct. 2022. URL: https://www.kenhub.com/en/library/anatomy/the-musculoskeletal-system.

[43] Ruud Hosman et al. *The neuromuscular system*. June 2012. URL: https://arc.aiaa.org/doi/10.2514/6.2010-8354.

[44] C.J. De Luca. *Electromyography*. 2006, pp. 98–109.

[45] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin. “Techniques of EMG signal analysis: detection, processing, classification and applications”. In: *Biological Procedures Online* 8.1 (Dec. 2006), pp. 11–35. DOI: 10.1251/bpo115. URL: https://doi.org/10.1251/bpo115.

[46] Peter Konrad. *The ABC of EMG, A Practical Introduction to Kinesiological Electromyography*. Apr. 2005. URL: https://www.noraxon.com/wp-content/uploads/2014/12/ABC-EMG-ISBN.pdf.

[47] Wikipedia. *General linear model — Wikipedia, The Free Encyclopedia*. http://en.wikipedia.org/w/index.php?title=General%20linear%20model&oldid=1141339122. [Online; accessed 27-March-2023]. 2023.

[48] Wikipedia. *Linear regression — Wikipedia, The Free Encyclopedia*. http://en.wikipedia.org/w/index.php?title=Linear%20regression&oldid=1135734562. [Online; accessed 19-February-2023]. 2023.

[49] Wikipedia. *Pearson correlation coefficient — Wikipedia, The Free Encyclopedia*. http://en.wikipedia.org/w/index.php?title=Pearson%20correlation%20coefficient&oldid=1139788725. [Online; accessed 19-February-2023]. 2023.

[50] "Leave-One-Out Cross-Validation". In: *Encyclopedia of Machine Learning*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, 2010, pp. 600–601. ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_469. URL: https://doi.org/10.1007/978-0-387-30164-8_469.

[51] Simon Haykin. *Neural networks and learning machines*. Pearson India Education Services Pvt. Ltd, 2021.

[52] Jennifer Stine Elam et al. "The Human Connectome Project: A retrospective". In: *NeuroImage* 244 (Dec. 2021), p. 118543. DOI: 10.1016/j.neuroimage.2021.118543. URL: https://doi.org/10.1016/j.neuroimage.2021.118543.

[53] URL: https://www.cosynclab.it/.

[54] Vincent Wens et al. "A geometric correction scheme for spatial leakage effects in MEG/EEG seed-based functional connectivity mapping". In: *Human Brain Mapping* 36.11 (Sept. 2015), pp. 4604–4621. DOI: 10.1002/hbm.22943. URL: https://doi.org/10.1002/hbm.22943.

[55] Ed Bullmore and Olaf Sporns. "Complex brain networks: graph theoretical analysis of structural and functional systems". In: *Nature Reviews Neuroscience* 10.3 (Feb. 2009), pp. 186–198. DOI: 10.1038/nrn2575. URL: https://doi.org/10.1038/nrn2575.

[56] Agnese Sbrollini et al. "Surface electromyography low-frequency content: Assessment in isometric conditions after electrocardiogram cancellation by the Segmented-Beat Modulation Method". In: *Informatics in Medicine Unlocked* 13 (2018), pp. 71–80. ISSN: 2352-9148. DOI: https://doi.org/10.1016/j.imu.2018.10.006. URL: https://www.sciencedirect.com/science/article/pii/S2352914818301801.

[57] P.C. Bhaskar and M.D. Uplane. "High Frequency Electromyogram Noise Removal from Electrocardiogram Using FIR Low Pass Filter Based on FPGA". In: *Procedia Technology* 25 (2016), pp. 497–504. DOI: 10.1016/j.protcy.2016.08.137. URL: https://doi.org/10.1016/j.protcy.2016.08.137.

[58] Nurhazimah Nazmi et al. "Assessment on Stationarity of EMG Signals with Different Windows Size During Isotonic Contractions". In: *Applied Sciences* 7.10 (2017). ISSN: 2076-3417. DOI: 10.3390/app7101050. URL: https://www.mdpi.com/2076-3417/7/10/1050.

[59] Christopher Spiewak et al. "A Comprehensive Study on EMG Feature Extraction and Classifiers". In: *Open Access Journal of Biomedical Engineering and its Applications* 1 (Feb. 2018). DOI: 10.32474/OAJBEB.2018.01.000104.

[60] Jingwei Too et al. "EMG Feature Selection and Classification Using a Pbest-Guide Binary Particle Swarm Optimization". In: *Computation* 7.1 (Feb. 2019), p. 12. DOI: 10.3390/computation7010012. URL: https://doi.org/10.3390%5C%2Fcomputation7010012.

[61] Jingwei Too, Abdul Rahim, and Norhashimah Mohd. "Classification of Hand Movements based on Discrete Wavelet Transform and Enhanced Feature Extraction". In: *International Journal of Advanced Computer Science and Applications* 10.6 (2019). DOI: 10.14569/ijacsa.2019.0100612. URL: https://doi.org/10.14569%5C%2Fijacsa.2019.0100612.

[62] Lijing Lu et al. "A Study of Personal Recognition Method Based on EMG Signal". In: *IEEE Transactions on Biomedical Circuits and Systems* 14.4 (Aug. 2020), pp. 681–691. DOI: 10.1109/tbcas.2020.3005148. URL: https://doi.org/10.1109/tbcas.2020.3005148.

[63] Jiayuan He and Ning Jiang. "Biometric From Surface Electromyogram (sEMG): Feasibility of User Verification and Identification Based on Gesture Recognition". In: *Frontiers in Bioengineering and Biotechnology* 8 (2020). ISSN: 2296-4185. DOI: 10.3389/fbioe.2020.00058. URL: https://www.frontiersin.org/articles/10.3389/fbioe.2020.00058.

[64] Ying-Qiu Zheng et al. "Accurate predictions of individual differences in task-evoked brain activity from resting-state fMRI using a sparse ensemble learner". In: *NeuroImage* 259 (2022), p. 119418. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2022.119418. URL: https://www.sciencedirect.com/science/article/pii/S1053811922005353.

[65] Jacopo Barone and Holly E. Rossiter. "Understanding the Role of Sensorimotor Beta Oscillations". In: *Frontiers in Systems Neuroscience* 15 (May 2021). DOI: 10.3389/fnsys.2021.655886. URL: https://doi.org/10.3389/fnsys.2021.655886.

[66] Priyanka A. Abhang, Bharti W. Gawali, and S. C. Mehrotra. *Introduction to EEG- and speech-based emotion recognition*. Elsevier/AP, Academic Press is an imprint of Elsevier, 2016.

[67] *Fitrnet*. URL: https://it.mathworks.com/help/stats/fitrnet.html.

[68] Niv Tik et al. "Predicting individual variability in task-evoked brain activity in schizophrenia". In: *Human Brain Mapping* 42.12 (May 2021), pp. 3983–3992. DOI: 10.1002/hbm.25534. URL: https://doi.org/10.1002/hbm.25534.

[69] Shiyang Chen and Xiaoping Hu. "Individual Identification Using the Functional Brain Fingerprint Detected by the Recurrent Neural Network". In: *Brain Connectivity* 8.4 (May 2018), pp. 197–204. DOI: 10.1089/brain.2017.0561. URL: https://doi.org/10.1089/brain.2017.0561.

[70] Gokce Sarar, Bhaskar Rao, and Thomas Liu. "Functional connectome fingerprinting using shallow feedforward neural networks". In: *Proceedings of the National Academy of Sciences* 118.15 (Apr. 2021). DOI: 10.1073/pnas.2021852118. URL: https://doi.org/10.1073/pnas.2021852118.

[71] URL: https://it.mathworks.com/help/deeplearning/ref/feedforwardnet.html.

[72] Alexander D. Cohen et al. "Regression-based machine-learning approaches to predict task activation using resting-state fMRI". In: *Human Brain Mapping* 41.3 (Oct. 2019), pp. 815–826. DOI: 10.1002/hbm.24841. URL: https://doi.org/10.1002/hbm.24841.

[73] James F. Kaiser. "Some useful properties of Teager's energy operators". In: *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing* 3 (1993), 149–152 vol.3.

[74] Stanislaw Solnik et al. "TeagerKaiser energy operator signal conditioning improves EMG onset detection". In: *European Journal of Applied Physiology* 110.3 (June 2010), pp. 489–498. DOI: 10.1007/s00421-010-1521-8. URL: https://doi.org/10.1007/s00421-010-1521-8.

[75] A. G. Asuero, A. Sayago, and A. G. González. "The Correlation Coefficient: An Overview". In: *Critical Reviews in Analytical Chemistry* 36.1 (Jan. 2006), pp. 41–59. DOI: 10.1080/10408340500526766. URL: https://doi.org/10.1080/10408340500526766.

[76] Mingrui Xia, Jinhui Wang, and Yong He. "BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics". In: *PLoS ONE* 8.7 (July 2013). Ed. by Peter Csermely, e68910. DOI: 10.1371/journal.pone.0068910. URL: https://doi.org/10.1371/journal.pone.0068910.