Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

INGEGNERIA BIOMEDICA, ELETTRICA E DEI SISTEMI

Ciclo 35

Settore Concorsuale: 09/G2 - BIOINGEGNERIA

Settore Scientifico Disciplinare: ING-INF/06 - BIOINGEGNERIA ELETTRONICA E INFORMATICA

ASSESSING BRAIN CONNECTIVITY THROUGH ELECTROENCEPHALOGRAPHIC SIGNAL PROCESSING AND MODELING ANALYSIS

Presentata da: Giulia Ricci

Coordinatore Dottorato	**Supervisore**
Michele Monaci	Mauro Ursino
	Co-supervisore
	Elisa Magosso

Esame finale anno 2023

ABSTRACT

Brain functioning depends on the interaction among several neural populations that are linked through complex connectivity networks, enabling the transmission and integration of information. In the past decades, questions regarding the activation of different regions during cognitive processes and their reciprocal roles have attracted the interest of researchers in the field of systems neuroscience. Thanks to the development of advanced techniques for functional brain imaging and the tuning of sophisticated signal-processing methods, the interest in brain connectivity estimation has experienced a massive expansion.

Among different neuroimaging techniques, Electroencephalography (EEG) is receiving increasing attention in the current literature due to its capacity to detect rapid temporal changes in neural patterns, together with the possibility to reconstruct the cortical sources activity, leading to the study of functional brain circuits. However, although interest in brain networks is growing, the significance and functioning of several brain connectivity estimators remains controversial and insufficiently understood. Furthermore, a number of mathematical methods are employed in the literature to estimate connectivity from data, ranging from the simple statistical dependencies to sophisticated model-based methods that derive causal interactions, thus complicating the connectivity outcomes interpretations. Indeed, depending on the adopted methodologies and the context of application, important differences in the significance of these measures should be considered.

The first part of this PhD work addresses the methodological limitation described above exploring the reliability of the main data-driven connectivity estimation techniques and their accuracy in different conditions, as well as the neurophysiological significance of the connectivity estimates and the neural phenomena they can grasp.

To this aim, Neural Mass Models (NMMs), which simulate the EEG activity of a cortical region, represent a unique tool to reproduce networks of interconnected regions of interest (ROIs) and to test the performances of different functional connectivity (FC) metrics in controlled conditions. In this thesis, NMMs have first been employed to test the behaviour of Transfer Entropy and Granger Causality FC estimators in linear and non-linear conditions. Importantly, results underline that connectivity estimates reflect the amount of information transmitted from one region to another, a quantity that is significantly different from the connectivity strength. This phenomenon is driven by non-linearities which can influence the effect that one region exerts on a second region, leading, for instance, the activity of the second region to saturation and thus to a loss of information transmission. This is a crucial aspect that should be carefully taken into account when analysing brain connectivity estimates. The importance of non-linearities have also been emphasised in a NMMs study on a stroke patient, where task-dependent motor network changes have been characterized by varying the working point of the simulated cortical regions, keeping the connection values between ROIs fixed. Moreover, NMMs were used to compare the performances of different functional connectivity estimators (*Pearson correlation coefficient*, *Delayed correlation* *coefficient, Coherence, Lagged Coherence, Phase Synchronization and Transfer Entropy*). The analysis showed that Granger causality, in both temporal and spectral domains, outperforms the other estimators.

In addition to the methodological aspects described above, a second objective of this thesis was to assess brain connectivity changes on EEG experimental data. In this case, based on the results obtained with NMMs, Granger Causality was chosen for FC estimation. Starting from EEG scalp recordings, cortical sources have been reconstructed in order to extract the time series of brain ROIs, according to standardized atlases. Here, the basic assumption is that connectivity changes may incorporate important neuromarkers of behaviour and cognition, as well as of brain disorders, even at subclinical levels. Hence, EEG-based connectivity analysis has been carried out with two purposes: a) to detect task-dependent functional connectivity network changes, b) to identify resting-state network alterations between classes of individuals. The task-dependent connectivity changes have been investigated in an internal-external attentional task and in a Pavlovian fear conditioning-reversal experiment. Resting-state network alterations have been analysed in a non-clinical population with different autistic traits. In the latter case centrality measures of Graph theory have been employed to identify the main cortical regions and mechanisms involved in the characterization of the network.

In conclusion, connectivity-based neuromarkers, compared to the canonical EEG analysis, can provide deeper insight into the functioning of brain mechanisms in healthy and pathological conditions, and may drive future innovative diagnostic methods as well and therapeutic interventions. However, methodological studies to achieve a complete understanding of connectivity estimates, their accuracy and the kind of information they are able to grasp still require further investigation. Indeed, although Granger Causality was found to be the most reliable estimator, a number of questions remain open, especially concerning nonlinear phenomena.

Table of Contents

1 Introduction

1.1 Brain Connectivity

One of the fundamental and long-standing debates in neuroscience concerns function localization in the brain. Two opposite paradigms relate to the problem: a localizationist view, which emphasizes the specificity and modularity of brain organization, against a holist view, which stresses the global principles of brain functioning[1]. This controversy between localizationism and holism mirrors two contrasting properties that coexist in the brains: the functional segregation of different brain areas and their integration in perception and behaviour. The understanding of these two aspects of brain organization is central to any theoretical description of brain function.

Since the beginning of modern neuroscience, the brain has generally been viewed as an anatomically differentiated organ whose many parts and regions are associated with the expression of specific mental faculties, behavioural traits, or cognitive operations. Functional specialization has become one of the enduring theoretical foundations of cognitive neuroscience. The idea that individual brain regions are functionally specialized and make specific contributions to mind and cognition is supported by a wealth of evidence from both anatomical and physiological studies as well as from non-invasive neuroimaging[2]. These functionally specialized regions are capable to manage diverse responses and represent focal points of convergence of more specialized neural information. Further evidence for functional segregation is provided by the analysis of the specific deficits produced by localized cortical lesions. Specialization alone, however, cannot fully account for most aspects of brain function.

In the beginning of the 20th century, a paradigm shift occurred that changed the trend of research towards a more holistic view. In contrast to local specialization, brain activity is understood as globally integrated at many levels, ranging from the neuron to intercortical interactions to overall behavioural output.

Indeed, neurons do not work in isolation; rather, they form dynamical assemblies that tend to work in synchrony, a concept popularized by Hebb (1949)[3]. A multitude of neuronal assemblies are usually simultaneously active in the brain, occupying different cortical areas[4]. Neuronal assemblies that are functionally interconnected constitute a functional brain workspace. The brain has to integrate distributed sets of neuronal assemblies spread over multiple cortical domains to achieve coherent representation of events and to perform coordinated actions. Indeed, mounting evidence suggests that integrative processes and dynamic interactions across multiple distributed regions and systems underpin different cognitive processes such as visual recognition, language, cognitive control, emotion, and social cognition[5].

Hence, the cognitive functions delineated above require the brain to integrate a wide range of incoming stimuli from the environment and seamlessly 'bind' this complex stream of information into meaningful internal representations that are then used to plan for the next action.

Integration depends on neural communication among specialized brain regions, unfolding within a network of interregional projections, which gives rise to large scale patterns of synchronization and information flow between neural elements. Contemporary models suggest that brain functioning requires a trade-off between functional specialization and global communication[6]. The cortical infrastructure supporting a brain function may then involve many specialised areas whose union is mediated by the functional integration among them. Therefore, functional specialisation and integration are not exclusive, but rather complementary. Indeed, functional specialisation is only meaningful in the context of functional integration and vice versa. For instance, consciousness is thought to require brain-wide information broadcasting by a "global workspace", whereby segregated component processes are integrated and made available for undertaking higher cognitive functions, producing a unitary experience[7]. This balance between segregation and integration is manifest through dynamically changing patterns of correlated activity, constrained by the brains' structural backbone.

The most direct way to discover the brain mechanisms that underlie segregation and integration would be to use neuroimaging methods to map whole-brain structure and function. The critical issue is not simply to localize cognitive functions to some site in the brain but also to find out the patterns of dynamic interaction between different brain systems underlying cognitive processes; indeed, to unravel these processes it is essential to understand the dynamics of the workspaces that constitute the material core of any cognitive process.

A considerable amount of experimental evidence gathered in the last two decades supports the notion that electroencephalographic (EEG) signals can provide relevant insights into dynamic brain processes responsible for specific cognitive functions. Indeed, through EEG it is possible to grasp neuronal oscillations which have two main roles: (1) coding specific information, (2) assuring the communication between neuronal populations such that specific dynamic network may be created.

Electroencephalography and other neuroimaging techniques have firmly established functional specialization as a principle of brain organisation. Although functional integration has proved more challenging to assess, significant progress has recently been made in this field concerning the development of sophisticated technologies and algorithms.

In particular, a new approach has been developed in recent years to evaluate the anatomical and functional organization of the human brain. This approach aims to identify and classify brain connectivity networks with a number of neurobiologically meaningful and easily computable measures.

Indeed, to date it is usually postulated that localism and holism have been replaced by "connectionism," with many studies nowadays trying to assess interactions between specialized brain regions and not the function of these regions by themselves. The neuroscientific community started to approach the brain as an interconnected system where firing patterns of neurons integrate through networks of neural populations in a coordinated manner enabling communications between brain regions necessary for physiologic functions[8]. By addressing how connectivity mediates both segregation and integration, brain network approaches not only reconcile these seemingly opposing perspectives, but also suggest that their coexistence is fundamental for brain function. Nowadays, the study of brain regions interaction in cognitive processes is playing a crucial role, not only to understand mechanisms at the basis of normal brain functions, but also to identify alterations in pathological states.

Communication between brain regions can be assessed through connectivity measures depending on the level of analysis. Brain connectivity (or connectomes) is defined at three interrelated but distinct levels: (i) structural (or anatomical) connectivity, representing the presence and density of axonal connections, (ii) functional connectivity, defined as statistical dependencies among neural assemblies, and (iii) effective connectivity, enhances the information from functional connectivity and refers to an explicit model of causal inference, usually expressed in terms of differential equations[9].

These three dimensions of connectivity are strongly interdependent. Structural connectivity is the main determinant of functional connectivity, which in turn primes effective connectivity; moreover, through plasticity effects, functional and effective connectivity exert influences on the development of structural connections. Integration within a distributed system is usually better understood in terms of functional or effective connectivity since these two measures allow to infer the influence that one neuronal system exerts over another[10]. Indeed, for a complete understanding of how information is integrated in the brain, it is crucial to consider the causality of brain interactions.

Recently, numerous studies have embraced network science as a theoretical framework for brain connectivity. In this case, graph theory is employed to describe neural systems in terms of nodes (brain regions) and edges (connections) to explain how network topology shapes and modulates brain function. Complex network analysis enables the reliable quantification of brain networks with a small number of neurobiologically meaningful and easily computable measures[11,12]. These measures can describe both local and global features of brain network topology, allowing the extraction of the main brain regulation mechanisms, such as bottom-up or top-down.

In addition to brain imaging techniques, one way to study the synchronization between different brain areas, and thus information transmission, is through dynamical mathematical models inspired by brain functioning. Specifically, much attention has been devoted to Neural Mass Models (NMMs), which describe the average activity of macro-columns, or even cortical areas, using just a few state variables[13]. By connecting several NMMs together, it is possible to simulate a network of interconnected brain regions and to investigate how information is transmitted. Moreover, neurocomputational models represent a unique tool to test different connectivity estimation techniques against a series of ground-truth conditions.

1.2 Critical discussion

Since their introduction in the early nineties, neuroimaging techniques have primarily focused on the spatial localization of brain function. Although this model has provided tremendous insights into the workings of the brain, it was soon shown to be insufficient to explain higher-level functions that require coordinated action of many brain areas. To this aim, an alternate model of brain function based on distributed information processing has been formulated. As described above, this model does not exclude the spatial localization mode, rather it hypothesizes that spatially localized activated regions do encode simple properties, and that the interactions between these regions contribute to the encoding of more complex properties, giving rise to a repertoire of experiences. Importantly, this model highlights that interactions between brain regions are critical ingredient to understand brain function.

Since the early 1960s[14], part of neuroscientific research has focused on brain interactions, with a considerable increase in recent years. Throughout this time, the development of methods to efficiently and accurately quantify brain connectivity has been, and still remains, a challenging issue. Indeed, the problem of assessing connectivity from data is a difficult one, since this concept has ambiguous definitions and results depend crucially on how connectivity is defined and on the mathematical instruments employed[15].

The first who tried to shed light on the concept of brain connectivity was Friston in 1994, who introduced the useful analytical categories of anatomical, functional, and effective connectivity into brain functional imaging literature[10].

Anatomical connections may be determined by a variety of tract-tracing methods that can provide a description of structural geometry. Indeed, structural connectivity (the connectome) reflects the physical relationships between neural elements; that is, their synaptic connections or inter-regional projections (the 'wiring diagram'). These methods typically do not include EEG, so we will not discuss anatomical connectivity further, except for a few brief observations. Anatomical information may be an obviously useful starting point for subsequent physiological investigation. It may be represented by graphs that are either directed (if derived from a reliable invasive anatomical methods) or undirected (if derived from e.g., diffusion spectrum imaging). However, anatomy cannot tell us how regions are coupled dynamically, except perhaps on very slow (e.g., neurodevelopmental) time scales. However, while everyone agrees on the definition of structural connectivity, a still open debate concerns the definitions of functional and effective connectivity.

According to the aforementioned definition of Friston, functional connectivity (FC) is based on the estimation of "temporal correlations between remote neurophysiological events" [10]. It describes the set of pairwise statistical dependencies between observational data (time courses) recorded from individual different neural elements. Consequently, the resulting connectivity matrices are undirected (and therefore symmetric), unless time-lagged correlations are considered. Still, according to the traditional point of view, effective (or causal) connectivity is based on an estimate of the "influence that one neural system exerts on another"[10], and involves an explicit neurophysiological model describing the underlying process. Mathematically, the resulting connectivity matrices are directed and asymmetric and support causal information flow inference.

However, the previous definitions of functional and effective connectivity suffer from a variety of problems and are often misunderstood. The main issue arises from the failure to distinguish between theoretical properties of interest and the methods used to infer those properties. The modern point of view proposed recently by Reid et al. underlines that, although methods for FC estimation are based on the computation of some forms of statistical or information association between signals, the target is always to understand the causal interactions among the different neural populations[16]. This clearly runs counter to the typical definition of FC as the non-causal "statistical association" between measured brain signals.

The point raised by Reid et al. is supported by the recent definition of new FC methods that estimate valid causal inferences from observational data. Indeed, among all different methods to estimate FC, some of them are based on the concept of causality, as originally introduced by Wiener (1956)[17] and subsequently by Granger (1969)[18]. The ultimate objective of such approaches is the construction of a network in order to understand how brain regions causally interact to produce cognition, and how network's alterations may affect behavioral outcomes (for instance, in pathological states), although this network is derived from statistical dependencies between observational data.

In the following, we will especially refer to functional connectivity estimators within the wider point of view proposed by Reid et al. Now that this has been clarified, the main aspect that distinguish functional from effective connectivity is summarised below.

Functional connectivity measures are statistical dependences among measured time series that have recently been supplemented with causal information. These are data-driven estimates of connectivity that do not rely on physiological models, but rather on empirical models. In contrast, effective connectivity approaches are based on well-defined biophysical models of neuronal dynamics and should be understood as the simplest possible circuit diagram that would replicate the observed responses[19,20]. The main method to assess effective connectivity in neuroscience is dynamic causal modeling (DCM)[21] which assumes that the signals are produced by a state space model. These models were introduced in 2003 for fMRI, and only later they were extended to EEG and MEG. The key to Dynamic Causal Modeling (DCM) technique is that the response of a dynamic system can be modeled by a network of discrete but interacting neuronal sources described in terms of neural-mass or conductance-based models.

This distinction between functional and effective connectivity clarifies why methods like Granger Causality and DCM are not competitive but rather complementary: they make different assumptions, and they permit different interpretations. However, DCM concerns the comparison of the performances of distinct mechanistic models in accounting for observed data. In this case one should choose the best model and predefine or experiment with a large number of parameters. The uncertainty in predefining these parameters and the large number of their possible combinations is the main drawback of these techniques. Moreover, it is possible that no single model exists but rather multiple models may be equally appropriate for a given data set. In contrast, directional (or causal) functional connectivity methods have the advantage of being exploratory approaches that do not assume any specific underlying spatial or temporal relationship. Such methods can be used in assessing connectivity when no a-priori anatomical or physiological knowledge is available.

For the reasons outlined, this thesis focuses on understanding of the main functional connectivity estimators as intended by Reid et al. However, as highlighted in a recent study, a large number of metrics for estimating functional connectivity exist, which are based on different underlying mathematical measures of dependency and lead to different outcomes[22]. For this reason, simulated data using neurocomputational models, such as Neural Mass Models (NMMs), play a crucial role in evaluating competing functional connectivity methods against a "ground truth." Indeed, the analysis of generated data from a known network architecture enables parameters optimization, dependencies assessment as well as sensitivity analysis. The output of functional connectivity estimators are connectivity matrices characterized by the strength of brain region interactions.

More synthetic characterizations of the whole network as well as of specific nodes can be provided through graph-theory. Characterising the brain from an integrative point of view and studying its characteristics as a complex system can be of great impact in the field of cognitive and systems neuroscience. Recently, connectivity studies have been conducted to investigate human behavioural and cognitive performance, as well as the role of different large-scale brain networks in various conditions[23–26]. Comparing the brain topological alterations during a cognitive task and resting-state helps identify areas that affect human behavioural performance. For instance, DeSalvo et al. used a graph-based approach to explore variations in functional brain organization during semantic decision making compared with rest in healthy participants[27]. Moreover, Davison et al. demonstrated that changes in brain network properties of individuals correspond to task performance[28]. Furthermore, Cole et al. found that fronto-parietal network's functional connectivity pattern shifts across a variety of tasks, and that these patterns could be used to identify the specific task[29].

In addition to capturing various cognitive aspects in healthy subjects, task-dependent and resting-state brain connectivity analysis may be a unique tool for discriminating brain diseases. Indeed, brain disconnection results in functional impairment, associated with atypical integration of distributed brain areas. Studies in the field of complex brain networks have demonstrated that analysing the network properties and metrics derived from brain topology can help neurologists distinguish patient groups from control subjects in a variety of mental disorders, such as epilepsy, Alzheimer's disease (AD), multiple sclerosis (MS), autism spectrum disorder (ASD), and attention-deficit/hyperactivity disorder (ADHD) [30–35]. However, other mental disorders were also found in recent graph-based literature, including schizophrenia, Parkinson's disease, insomnia[36–39].

1.3 Scientific proposal

Despite the relevance of connectivity estimators in the field of computational neuroscience, their functioning and performances are still insufficiently understood.

The first objective of this thesis was to fill this gap by testing the performances, as well as parameters dependencies, of the main functional connectivity metrics in different working conditions. To this aim, Neural Mass Models were employed to simulate electroencephalographic data and to recreate a 'ground truth' network scenario of interconnected brain regions.

Particular emphasis was paid on: a) the performance comparison of different FC estimator, b) the neurophysiological meaning of the connectivity values provided by FC estimators, and c) the difference between true connectivity strength and information transmission.

Given the significance of causality in information integration among brain regions, directional FC algorithms, such as Granger causality[18] and Transfer Entropy[40], has been the focus of this analysis. Light has been shed on the limitations and advantages of these estimators by investigating their reliability under linear and non-linear working conditions. In addition, NMMs were employed to study how brain rhythms and information are transmitted in a physiologically plausible network of cortical regions, as well as to simulate the task-dependent changes of the cortical motor network in a stroke patient.

Once untangled all the aspects, functional connectivity metrics were computed on different EEG signal recordings. Hence, accepted the computational limitations identified via NMMs, and based on the above analysis outcomes, Granger Causality was chosen to estimate directed functional connectivity on experimental data. Indeed, the second objective of this study was to investigate brain connectivity changes during attentional and fear conditioning/reversal tasks, and to understand how autistic traits may be linked to resting-state connectivity alterations. For this purpose, EEG signals were reconstructed at the source level using various software and open-source toolboxes, and the functional connectivity has been computed and compared between conditions. Finally, in some cases, graph theory and complex network analysis have also been employed to extract and quantify the main topological features of brain networks.

1.4 Outline of the Thesis

The core of the thesis is organised as follows. An overview of the main methods and mechanisms on which this work is firmly grounded is provided in Chapter 2. Within this section, a general description of the main neuroimaging techniques is given, dwelling on electroencephalography and cortical source reconstruction. Subsequently, neural oscillations detectable with EEG are described highlighting their role in brain interactions. Furthermore, the main classes of functional connectivity estimators are presented, as well as the main graph theory metrics for the quantification of network topology. Finally, an overview on Neural Mass Models, together with their benefits in understanding connectivity metrics, is outlined.

Chapter 3 collects three studies carried out employing NMMs to simulate brain regions interactions in controlled conditions. In the first study, the performance of Transfer Entropy in detecting true connectivity strength is tested for different working conditions (i.e. linear or non-linear) and parameters (i.e. network size, pure delay, signal length). The second study compares the reliability of eight different functional connectivity estimators using ROC curves. Furthermore, it aims to investigate how brain rhythms are transmitted in a physiologically plausible network. For this purpose, Granger Causality is employed to estimate directional connectivity of the network under both linear and nonlinear conditions. In the third study, a network of NMMs has been fitted to experimental EEG data in order to assess the task-dependent changes in functional connectivity between the motor/premotor areas of a stroke patient.

Chapter 4 presents three studies where Granger Causality is applied on experimental EEG data to assess directional functional connectivity. In this section, advanced EEG processing techniques are employed for cortical sources reconstruction, to extract the time series of functionally significant cortical Regions of Interest (ROIs) and compute brain connectivity. The first two studies investigate task-dependent spectral connectivity changes respectively during: 1) internal-external attention competition tasks; 2) Pavlovian fear conditioning and reversal experiment. Whereas, the third study investigates the resting-state connectivity alterations in individuals with high autistic traits. In the latter study, some directional measures of graph theory are also used to extract the main features of the networks.

Finally, the conclusions of this thesis are summarized and critically commented in chapter 5.

2 Fundamentals

This chapter contains the fundamentals on which functional brain connectivity is grounded. First, the main neuroimaging techniques are presented, with particular emphasis on electroencephalography and cortical source reconstruction techniques. Furthermore, neural oscillatory mechanisms are described, underlying their fundamental role in ensuring communication between brain regions and integration of information. Moreover, methods for measuring brain connectivity are discussed, and in particular those for estimating directional functional connectivity. Then, principles of graph theory for the analysis of the brain as a complex network are presented, including the extraction of its main features. Finally, neural mass models are presented as a promising tool for the study of brain rhythms transmission in a network, as well as to assess the performances of FC estimators.

2.1 **Neuroimaging methods**

The answers to many questions we raise on brain organisation depend on the quality of data we are able to extract on the location, dynamics, oscillations, amplitude and types of brain activity, and thus depend on the sophistication of the neuroimaging technology employed[41]. These techniques have become an essential tool for the neuroscientist seeking to understand the brain on a spatial and temporal scale. Recent years have seen rapid growth in the field of neuroimaging technology and methodology and, subsequently, in understanding the structural and functional brain organization, as well as thriving clinical applications. Correspondingly, the general level of sophistication with which neuroimaging tools are used has transformed neuroscience research, becoming the predominant technique employed in behavioural and cognitive neuroscience.

Neuroimaging has established functional segregation as the foundation of brain organisation, as well as the integration between different brain areas in terms of functional and effective connectivity. Functional neuroimaging techniques, including positron emission tomography (PET), functional magnetic resonance imaging (fMRI), electroencephalography (EEG), magnetoencephalography (MEG), and other modalities, provide powerful tools to study human brain and serve as methodological foundations for system neuroscience.

In PET, a radioisotope tracer can be injected into the subject while performing a task. The brain areas participating in functional activation demand a higher level of oxygen and glucose energy. The regional cerebral blood flow increase and the metabolism are proportional to the neuronal activation. Over the years, fMRI has grown largely since it does not require the administration of tracers and provides a higher spatial resolution than PET. Furthermore, the advantages of such technology concern non-invasiveness, relative ease of implementation, high spatial resolution, and importantly, signal fidelity. Functional MRI measures brain activity by recording concomitant changes in cerebral perfusion (neurovascular coupling). This technique uses blood oxygenation level-dependent (BOLD) signals to highlight areas of active neuronal activity. The fMRI signal is robust and for the most part, highly reproducible and consistent. However, if fMRI can very effectively identify what brain areas - at a specific spatial scale – are active in association with specific tasks or in resting state, it is not so reliable in addressing mechanisms related to cognitive processes. Indeed, fMRI provides a high spatial resolution (1–10 mm), but has only limited temporal resolution (1 s), primarily due to the limitations of the hemodynamic response. However, cognitive processes are dynamic. Therefore, techniques such as PET and fMRI are not the most adequate to reliably grasp this fundamental feature of brain functioning.

Besides PET and fMRI, there are other neuroimaging methods such as electroencephalography (EEG) and magnetoencephalography (MEG) that allow temporal aspects to be reliably investigated. Moreover, since the computation of functional connectivity depends on the correspondence of neural signals over time, techniques such as EEG and MEG, which have excellent temporal resolution, are optimal for investigating brain interactions. However, the main limitation of these techniques concerns the spatial resolution which is in the order of several millimetres or even centimetres, but it can increase with the help of multielectrode (high-density) recordings and source imaging techniques. Furthermore, EEG and MEG poorly measure activity arising below superficial neural structures, whereas fMRI records activity within the entire brain volume.

Nevertheless, EEG and MEG remain the most appropriate tools for the study of brain network interactions. These techniques offer direct, real time, monitoring of spontaneous and evoked brain activity, but also allow for spatiotemporal localization of underlying neuronal generators. EEG and MEG show the following common characteristics: 1) they have a millisecond temporal resolution; 2) potential differences and magnetic fields are linear functions of source strengths and nonlinear functions of source locations; 3) they reflect the same elementary neuronal phenomena, since they are both caused by currents from synchronously activated neural populations, and thus both can be used for the localization of neuronal generators; 4) can be analysed in both temporal and spectral domains.

The EEG field is a scalar and a relative measurement; it is sensitive to both tangential and radial components of dipolar sources. Theoretically, a radially oriented dipolar source does not give rise to a magnetic field outside a spherical volume conductor; consequently, the MEG is not sensitive to radial components of dipolar sources but to the tangential components [4,42].

The main advantage of MEG is its good spatial resolution in separating cortical sources due to less spatial smearing than in the EEG. Indeed, MEG provides better spatial resolution of source localization (2-3 mm) than EEG (7-10 mm). However, literature confirmed that electric and magnetic measurements provide comparable information. For instance, it was demonstrated that, using the novel concept of the half sensitivity volume, EEG and MEG record the electric activity in a very similar way[43]. More recently, applied pattern recognition techniques were used to decode hand movement directions from simultaneous EEG/MEG measurements, concluding that the inference of movement direction works equally well for both techniques[44]. Besides the technical aspects, it may be beneficial to consider also the cost effect of the recording modality. The MEG instrumentation costs about 20 times more than the EEG instrumentation with the same number of channels. Considering both technical and economic aspects, and keeping in mind the purpose of the work, which focuses on the study of brain connectivity, EEG was the neuroimaging technique employed for this thesis and is detailed below.

2.1.1 Electroencephalography

The first known neuroelectrical recording on animals dates back to 1875 and was performed by Richard Caton. The advent of recording neural activity on humans took another half century to occur. Hans Berger, a German psychiatrist, pioneered the EEG in humans in 1924. He recorded neurophysiological signals that fluctuated rhythmically when eyes were closed, but which became far less rhythmic and of generally smaller amplitude when eyes were open. After more than 85 years of development and use in clinical practice, the electroencephalogram (EEG) remains the most-common non-invasive tool for the investigation of the electrophysiological activity of the brain.

The electric potentials registered by EEG is generated by the activity of neurons, which are the electrically excitable cellular units of the brain and nervous system. However, conventional scalp or cortical surface EEG recordings are unable to detect the activation of a single neuron. Instead, EEG is primarily generated by extracellular current flow of cortical pyramidal neurons in the cerebral cortex that are oriented perpendicularly to the brain's surface. Hence, the neural activity detectable by EEG is the summation of the excitatory and inhibitory postsynaptic potentials of relatively large groups of neurons thanks to their parallel organization. The strength of the current flow is directly proportional to the number of activated neurons and produces a signal which is detectable at the scalp level.

EEG instrumentation consists of a set of scalp electrodes coupled to high-impedance amplifiers and a digital data acquisition system. Typically, the resistive contact between the electrode and skin is improved using electrolytic gels or abrasive pastes. Another, more recent approach has been so-called 'dry' electrodes that capitalize on innovations in material sciences as well as electronics to minimize the setup time.

EEG has been traditionally employed considering the standard 10–20 system to define electrodes' position, which includes only 21 measurement electrodes. However, it has been widely acknowledged that the spatial resolution of the 10–20 system is not sufficient for modern brain research[41,45,46]. To improve the spatial resolution of EEG a higher number of electrodes must be used. Today, the market responds to this need providing commercially available systems with up to 256 electrodes.

An intrinsic issue of EEG is that to reach the scalp, neuronal signals generated in the cortex cross several layers of tissues with different electrical properties and a complex geometry. This implies that what is recorded at the scalp is an attenuated and distorted image of the cortical sources. In other words, this means that EEG is not sensitive to deep cortical activation. Inverse methods and approaches such as LORETA[47] claim to detect deep sources.

However, there is still the possibility that a lot of information from deep structures, above all belonging to the higher frequency domains (lower amplitudes) could be lost. Even if these methods allow a reliable source reconstruction, their theoretical limitations must be kept in mind.

Another limit of EEG recordings concerns the aspect that registered brain activity is overwhelmed by other signals generated by the body (i.e., eye movements, cardiac activity, and scalp muscle contraction) or in the environment. Moreover, temporary detachments of the recording electrodes can further erode the signal and interfere with the relevant physiological EEG activity. Fortunately, artifacts possess many distinguishing characteristics which are visually identifiable by well-trained experts. Through independent component analysis (ICA), based on blind sources separation, it is possible to visualize and localize independent components of EEG signals and thus to reject the artefacts and reconstruct a cleaned version of EEG signals. Data can be analyzed with a number of different Matlab (MathWorks, Natick, MA) toolboxes, such as EEGLAB (Swartz Center for Computational Neurosciences, http://www.sccn.ucsd.edu/eeglab). ICA in EEGLAB can be performed using the Infomax ICA algorithm[48].

The advent of digital technology has led to greater sophistication and multiple software applications to extend the applicability of EEG beyond the confines of the laboratory. Indeed, systematic improvements have been made in the portability of EEG systems, allowing recordings in real-world environments. In this case, EEG headset is connected wirelessly to a signal receiver allowing both movement and change in posture and orientation.

There are, of course, limitations to what this approach can tell us about cognition. These include the relatively poor spatial resolution of EEG, the fact that the dendritic field potentials of the cortical pyramidal neurons recorded on the scalp constitutes only part of the brain's relevant dynamics, and various more specialized technical problems such as volume conduction. Despite such limitations, EEG currently represents a widely employed neuroimaging technique to investigate complex brain mechanisms in non-clinical population, but also to monitor and diagnose brain disorders, such as epilepsy, autism and schizophrenia.

2.1.2 Cortical Sources Reconstruction

Since the discovery of electroencephalography (EEG), when it was hoped that EEG would offer "a window into the brain," researchers and clinicians have attempted to localize the neuronal activity that generates the scalp potentials measured noninvasively through EEG. As described above, the primary source of the EEG potentials is the current flow induced by pyramidal cells in the cerebral cortex which exhibit synchronous activity. The cortical sources are generally modelled through the mathematical abstraction of an equivalent current dipole (ECD), representing the postsynaptic currents flowing through the apical dendritic trees of cortical pyramidal cells at a given moment[49].

Since the head is a conducting medium, volume conduction allows the propagation of these current flows to the scalp surface, where they give rise to electric potential differences between electrodes placed at different positions on the scalp[50]. Thus, the electrical potential generated by each current dipole propagates through the conducting volume of the brain (grey matter, cerebrospinal fluid, meninges, skull and skin) to reach the scalp where the EEG signal can be recorded. Therefore, EEG signal is a mixture of many different contributions from a very large number of neural sources located in different regions of the cerebral cortex.

In order to interpret EEG data in neuroanatomical terms, the reconstruction of the sources from the recorded data is required (Fig. 2.1). Source modelling involves the estimation of the location in the brain of neural electrical sources, starting from the EEG signals recorded on the scalp. This requires the solution of an 'ill-posed' inverse problem, for which infinitely many solutions exist, which means that different combinations of cortical sources can result in the same potential distribution on the scalp. The common approach to tackle this problem is to make assumptions about the intracerebral sources. The solution to the inverse problem may also vary depending on head geometry, tissue conductivities, and electrode placement. Indeed, a priori constraints, preferentially incorporating anatomical, physiological, and biophysical knowledge, needs to be defined.

Early explorations in the 1950s using electric field theory to infer the location and orientation of the current dipole in the brain starting from scalp potential distribution, triggered considerable efforts to quantitatively deduce these sources. Fostered by the increasing availability of magnetic resonance imaging, which allows detailed realistic anatomy to be incorporated, the precision of sources localization approaches drastically increased. For source reconstruction researchers typically rely on one of the publicly available toolboxes such as LORETA-KEY©®, Brainstorm[51], Fieldtrip [52], EEGLAB[53] and MNE[54], which provide ready-made anatomical templates, methods for electrical forward calculations, and implementations of inverse solutions.

Today, source localization has reached a level of consistency and precision that allows these methods to be included in the family of brain imaging techniques. This technique allows not only precise source localization at given time instances, but the high time resolution of EEG also permits looking at information flow within large-scale brain networks and gaining new insights in the way the areas of such networks communicate with each other. This chapter gives an overview of EEG source imaging methods.

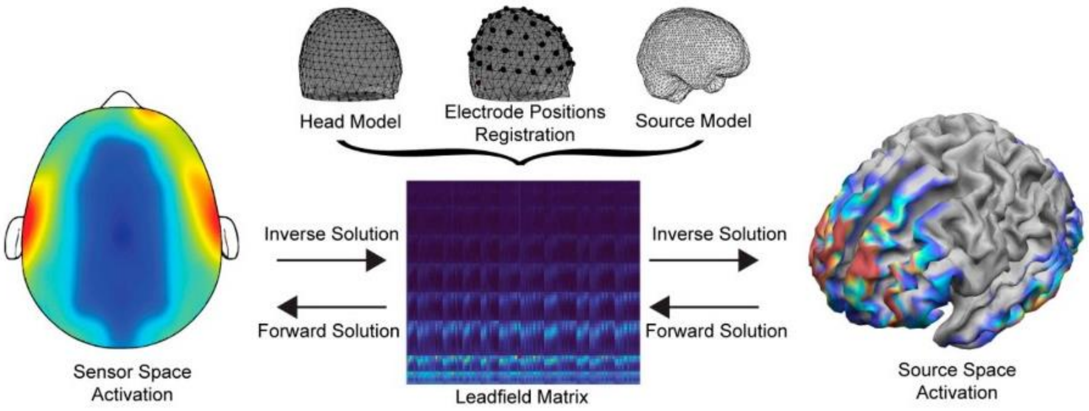


Figure 2.1 - Process flow to forward problem solution: three main configurations are required—(a) the head and source models (i.e., the location of the solution points in the brain), (b) the electrode alignment on the head model, and (c) the leadfield matrix using the channel locations in relation to the anatomical information of the head model. As such, the solution's accuracy highly depends on the efficient generation and composition of the points above[55].

2.1.2.1 Forward modelling:

In order to estimate the cortical sources of scalp potentials, we must first be able to solve the associated forward problem, that is, we need a forward model that maps a source of known location, strength, and orientation to an array of EEG sensors. Since EEG's frequencies of interest are relatively low (typically <100 Hz), this model is governed by the quasi-static versions of Maxwell's equations[56]. Moreover, the permeability of biological tissue is approximately that of free space, thus, the main factors determining the forward model are the electrical conductivity and geometry of the head.

In the discussion below, the following notation has been used: lowercase letters for scalar quantities, lowercase letters with arrow for vector and uppercase letters for matrices. The total current density within the brain volume $\vec{j}$ (whose unit is in A/m^2) is given by the sum of two distinct components, represented by the primary current density or impressed current $\vec{j}_p$ and the volume current density $\vec{j}_v$:

$$\vec{j} = \vec{j}_p + \vec{j}_v \qquad (2.1)$$

The primary current $\vec{j}_p$ flows within the cortical macro-column and is assumed to be the physical correlate of neuronal activity. It can be modelled by means of an equivalent current dipole (ECD) represented by a point source:

$$\vec{j}_p(\vec{r}) = \vec{q}\delta(\vec{r} - \vec{r}_q) \qquad (2.2)$$

In Eq. (2.2), $\vec{q} = \int \vec{j}_p(\vec{r})\, dr$ is the current dipole moment and is measured in Am, $\delta(.)$ is the Dirac function and is measured in $1/m^3$, $\vec{r}$ is a generic spatial position with respect to a generic reference and $\vec{r}_q$ identifies the position of the current dipole. The current dipole moment $\vec{q}$ represents the model of the postsynaptic potential in a cortical macro-column.

The volume current (or return current) results from the effect of the electric field in the volume on extracellular charge carriers, and is represented by Ohm's law:

$$\vec{J}_v = \sigma(\vec{r})\vec{e}(\vec{r}) = -\sigma(\vec{r})\nabla v(\vec{r}) \tag{2.3}$$

In Eq. (2.3), σ indicates the electrical conductivity of the medium surrounding the macro-column, and $\vec{e}$ represents the electric field generated by the current dipole; the latter can be expressed as the negative gradient of the electric potential v. Thus, by substituting equations (2.2) and (2.3) into (2.1) we obtain:

$$\vec{J} = \vec{q}\delta(\vec{r} - \vec{r}_q) - \sigma(\vec{r})\nabla v(\vec{r}) \tag{2.4}$$

Note that even though the currents of interest are the primary currents (i.e. the current dipoles $\vec{q}$), as they represent cortical activations, volume currents must also be considered when solving the direct problem since they contribute to EEG scalp potentials. The application of divergence to all terms in equation (2.4) leads to the following expression:

$$\nabla \cdot (\vec{J}) = \nabla \cdot (\vec{q}\delta(\vec{r} - \vec{r}_q)) - \nabla \cdot (\sigma(\vec{r})\nabla v(\vec{r})) \tag{2.5}$$

Considering the quasi-stationarity condition $\nabla \cdot (\vec{J}) = 0$, this leads to the Poisson equation:

$$\nabla \cdot \left(\vec{q}\delta(\vec{r} - \vec{r}_q)\right) = \nabla \cdot (\sigma(\vec{r})\nabla v(\vec{r})) \tag{2.6}$$

Equation (2.6) is a partial derivative equation that provides a complete description of the direct problem[49]. Electrical potentials can be calculated from the previous equation from a given primary current density ($\vec{J}_p(\vec{r})$ or dipole current $\vec{q}$) and a conducting volume which, in the applications of our interest, is represented by the head. The problem is completed by imposing the appropriate boundary conditions and fixing the potential at the reference point.

Crucial aspects for the forward model solution concern the assumed shape and conductivity of the head. Signals generated by synchronized postsynaptic potentials do not propagate homogeneously in the brain. The electrical characteristics of many biological tissues are inhomogeneous, anisotropic, dispersive, and nonlinear[57]. Different tissues such as the scalp, skull, cerebrospinal fluid, and grey and white matter have different conductivity characteristics and therefore attenuate the current to a different extent.

The electrical conductivity of the head, $\sigma(\vec{r})$, is typically estimated by segmenting an anatomical MR image into its different biological tissues and head structures (e.g., scalp, skull, cerebrospinal fluid, grey and white matter). Among all tissues, a good estimate of the conductivity and shape of the skull[58] is of great relevance due to the large difference in conductivity between skull and soft tissue, which impacts EEG potentials. Ideally, an individual geometric model should be created from a structural MRI of the participant's head and digitized electrode positions. However, the acquisition of individual MRI is not always possible and generally comes at a high cost. Therefore, it is common practice in EEG source analysis to use template anatomies such as the Colin27 head[59] (a detailed MR image made of 27 scans of a single individual head) or the ICBM152 head[60] (a non-linear average of the MR images of 152 individual heads). Then, a conductivity value is associated to each tissue, typically considering standard values measured in vitro using excised tissue[61]. Alternatively, impedance tomography (EIT) can be employed to estimate conductivity values in vivo, by injecting a small current (in the order of 1-10 μA) through one electrode and measure the potential differences in all the others. Most of the forward models consider isotropic tissue conductivity. However, the anisotropic conductivity information can also be incorporated into the forward model employing the Finite Element Methods (FEM).

A full mathematical solution of the EEG forward problem is achieved assuming as a boundary condition that the electric potential v is fixed at some reference point and that no currents leave the head volume. Spherical head models can be employed when a simple and approximated solution is sufficient. In this case, the head model consists of three concentric spheres corresponding to the brain, the skull and the scalp. Each sphere is characterised by a homogeneous and isotropic conductivity value and the brain-skull, skull-scalp and scalp-air are the interface surfaces. Thus, Eq. (2.6) can be solved analytically. However, spherical head models lead to good representations of reality only in the superior regions of the brain, where the head has a more spherical shape. To obtain more accurate solutions, it is necessary to adopt more complex and realistic head models and to solve the forward problem by using numerical methods such as Boundary Element Methods (BEM), Finite Element Methods (FEM), or Finite Difference Methods (FDM).

BEM is the most popular method introduced to solve the forward problem and can be used to compute a numerical solution for Eq. (2.6) under the assumption that $\sigma(\vec{r})$ is piecewise homogeneous. In this case, boundaries can be discretized considering the interface surfaces of the different conductivity tissues, so that the solution of the problem is achieved by solving a set of linear equations. The surfaces are discretized employing a mesh of triangles so that the complexity of the system depends on the total number of nodes in all meshes.

The FEM, on the other hand, models tissue conductivity inhomogeneity and even conductivity anisotropic distributions within the white matter. It discretizes Eq. (2.6) over the entire head volume. By defining a different conductivity for each element, the model allows the incorporation of an anisotropic conductivity tensor instead of scalar values for $\sigma(\vec{r})$. The FEM employs tetrahedrons to connect the nodes of an irregular grid, or cubes in the case of a regular grid, while the potential $v(\vec{r})$ is interpolated over the nodes. The discretization for FEM method also leads to a system of linear equations. In principle, a FEM model incorporating spatially variable anisotropic conductivity values provides the most accurate forward model.

However, it is important to note that the accuracy of these methods depends on both the knowledge of real tissue conductivities and on the numerical details required, such as the resolution of the mesh in which the solution is calculated. Currently no method exists that is capable of producing high-resolution images from which to extract conductivity values. Furthermore, due to the computational costs, the grids of the BEM and FEM have relatively low resolution. For these reasons, the spherical model could represent a valid alternative solution to reduce the potential numerical instability associated with BEM and FEM methods. The forward model is typically computed using automated software such as OpenMEEG[62].

Having introduced the head model and the source model (ECD), a useful notation for the treatment of the inverse problem can be expressed. Denoting $m(\vec{r},t)$ the scalp potential at time t at the electrode at position $\vec{r}$, due to the moment dipole $\vec{q}$ located at $\vec{r}_q$; we can write:

$$m(\vec{r},t) = \vec{g}(\vec{r},\vec{r}_q)\ \vec{q}(t) \tag{2.7}$$

In Eq. (2.7) $\vec{q}(t)$ is a vector of dimension 3×1 containing the three components of the dipole along the three space dimensions, i.e. $\vec{q}(t) = (q_x, q_y, q_z)$ or even $\vec{q} = \ ||\vec{q}|| \cdot \vec{e}_q$ where $||\vec{q}||$ represents the intensity or amplitude of the dipole and $\vec{e}_q = \frac{\vec{q}}{||\vec{q}||}$ is the dipole's orientation vector. Instead, $\vec{g}(\vec{r},\vec{r}_q)$ is a vector of size 1×3 and represents the solution of the forward problem for a dipole with unit amplitude and orientation $\vec{e}_q$, and expresses how each component of dipole has an effect on the potential m measured by the electrode at position $\vec{r}$. In Eq.(2.7), it was taken into account that the scalp electric potential is linear with respect to the dipole moment $\vec{q}$, while it is non-linear with respect to the position $\vec{r}_q$ of the dipole[63].

2.1.2.2 Inverse Model:

The fundamental problem of determining intracranial sources that generate EEG scalp potentials represent the challenge of solving the inverse problem. From the forward model and the potential measurements on the scalp at time t, the inverse problem consists in finding the current distribution $\vec{J}_p(\vec{r},t)$ that generates the data measured on the scalp at time t. Assuming that the current density distribution $\vec{J}_p(\vec{r},t)$ arises from an ns number of current dipoles $\vec{q}_1, \vec{q}_2, \dots \vec{q}_{ns}$ at positions $\vec{r}_{q1}, \vec{r}_{q2}, \dots \vec{r}_{qns}$, we can write:

$$\vec{J}_p(\vec{r},t) = \sum_{j=1}^{ns} \vec{q}_j(t)\delta(\vec{r} - \vec{r}_{qj}) \tag{2.8}$$

Due to the superposition of effects, the potential $m(\vec{r},t)$ of the electrode at the generic position $\vec{r}$ and at time t is given by:

$$m(\vec{r},t) = \sum_{j=1}^{ns} \vec{g}(\vec{r},\vec{r}_{qj})\ \vec{q}_j(t) \tag{2.9}$$

Eq. (2.9) relates to a single electrode recording the potential at the scalp in a single position $\vec{r}$. Considering a number of electrodes equal to ne at positions $\vec{r}_1, \vec{r}_2, \ldots \vec{r}_{ns}$ and the electrodes' potential measured in position $\vec{r}_i$ and at time t, the formulation of Eq. (2.9) becomes:

$$\vec{m}(t) = G\left(\{\vec{r}_i, \vec{r}_{qj}\}\right)\vec{q}(t) \qquad (2.10)$$

In Eq. (2.10) the matrix $G\left(\{\vec{r}_i, \vec{r}_{qj}\}\right)$ has dimension $ne \times\ 3ns$, indeed, $\vec{g}(\vec{r}_i, \vec{r}_{qj})$ is a vector of dimension 1×3, while $\vec{q}_j(t)$ is a vector of dimension 3×1 that contains the dipole components along the three directions x, y and z. Such a matrix $G\left(\{\vec{r}_i, \vec{r}_{qj}\}\right)$ is called leadfield matrix, or gain matrix, whose generic column represents the electrical potential at the scalp (at the sensors) generated by a unit current dipole at a given position and oriented in one of three orthogonal directions. The leadfield matrix is generally assumed to be time invariant (i.e. time-independent). The vector $\vec{q}(t)$ has dimension $3ns \times 1$ and $\vec{m}$ has dimension $ne \times 1$.

It is important to notice that the potentials at the electrodes are sampled at t time points $t_1, t_2, \ldots, t_t$ and that they are assumed to be linked to the sources at the same time points. Thus, a time invariant model is considered and the system described in Eq. (2.10) becomes:

$$M = GQ \qquad (2.11)$$

The matrix to the left of the equal contains the data measured at different time points and has dimension $ne \times t$, while the matrix of dipole moments at different time points has dimension $3ns \times t$.

In the previous formulation, it was assumed that both the orientation and amplitude of the dipoles were unknown. However, since apical dendrites producing the measured potentials are oriented perpendicular to the cortical surface, the orientation of the dipoles is often fixed orthogonal to the cortex surface[64]. In this condition only amplitude (or intensity) of the dipoles varies over time; as a consequence, a generic moment of dipole $\vec{q}_j(t)$ can be written as:

$$\vec{q}_j(t) = s_j(t)\,\vec{n}_j \qquad (2.12)$$

In Eq. (2.12) $s_j(t) = \left\|\vec{q}_j(t)\right\|$ represents the dipole intensity and $\vec{n}_j = \frac{\vec{q}_j(t)}{\|\vec{q}_j(t)\|}$ is the orientation vector of the dipole of dimension 3×1. It is possible to incorporate the dipoles' orientation versors within the G-matrix, obtaining $A\left(\{\vec{r}_i, \vec{r}_{qj}, \vec{n}_j\}\right)$ with dimensions $ne \times ns$, where the generical element $\vec{g}(\vec{r}_i, \vec{r}_{qj})\vec{n}_j$ of the A matrix is a scalar. Below, the following abbreviated notation is used:

$$M = AS \qquad (2.13)$$

where M is the matrix of EEG data measurements at different time points and has dimension $ne \times t$, A is the matrix of ns dipoles and ne electrodes and has dimension $ne \times ns$, and S is the matrix of source amplitudes at different time points and has dimension $ns \times t$.

Typically, a noise or perturbation matrix, denoted by N, is added to the system of dimension $ne \times t$, so that the data matrix can be written in the following form:

$$M = AS + N \tag{2.14}$$

At this point, starting from Eq. (2.14) the inverse problem consists, in the most general case, to estimate the number ns, the position $\vec{r}_{qj}$, and the time evolution of each current dipole, given the electrode positions, the matrix A calculated in the forward problem and the recordings M measured on the scalp.

The EEG inverse problem is ill-posed. This means that for all admissible output potentials, the solution is non-unique (since the number of sources >> number of electrodes) and unstable (it is highly sensitive to small changes in the noisy data)[64]. A solution to this problem can only be found if a priori assumptions about the sources are considered. Neurophysiologic, biophysical and anatomic knowledge, as well as the assumptions about distributions of neuronal activity are all contributors to such a priori constraints. Many different constraints have been introduced over the years and new constraints and assumptions are continuously formulated in literature based on new available knowledge of signal generation. Various mathematical inverse models have been formulated depending above all on the number of dipoles considered and whether dipole position, magnitude and orientations are kept fixed or assumed to be known.

There are two main approaches to the inverse solution: *parametric* approach (or Dipolar Solution) and *non-parametric* approach (or Distributed Inverse Solution).

Parametric methods: are known as Dipolar solution approaches and assume that electric potentials on the scalp arise from a limited number of ECDs of unknown location and orientation. When dipoles orientation is unknown, Eq. (2.11) needs to be considered and, for each dipole and each time point, it is necessary to estimate the three dipole components. In this classical approach, the hypothesis is that only one or few areas in the brain are active and generates the scalp potential. This results in an overdetermined problem with more data than unknowns.

These methods can differ in minimization algorithms and efficiency to escape local minima, measures of goodness of fit as well as the use of physiological and/or mathematical constraints often required in the solution estimation process[41]. They need a priori assumptions on the number and location of the brain source, giving a unique solution provided by the identification of the global minimum exists. Moreover, such approaches require a model order search in addition to a source parameter optimization[65]. In this case, the inverse problem is non-linear (as the matrix A is a non-linear function of the positions $\vec{r}_{qj}$). Dipolar models range in complexity depending on the number of dipoles that can be reliably considered, which is limited by the number of scalp electrodes and by the nonlinear complexity of the search algorithms with multiple sources. The main parametric methods are: least-squares source estimation, beamforming approaches and the multiple-signal classification algorithm (MUSIC) [66].

Non-parametric methods: are known as Distributed Source Imaging techniques and assume a very large number of dipole sources with fixed location and possibly also with fixed orientation, distributed over the entire brain volume or on the cortical surface. Indeed, no constraint is imposed on the number of sources. Instead, a large number (usually more than 5000) of equivalent dipoles are distributed in over the whole source space and the strength of each of these dipoles is estimated. Using anatomical information from the individual or a template MRI, the source space is usually constrained to the grey matter. As it is assumed that sources are the current flow induced by pyramidal cells in the cerebral cortex, which are normally oriented to the cortical surface, fixed orientation dipoles are generally set to be normally aligned[67]. Hence, the reverse problem consists in considering Eq. (2.13) and estimating the dipole amplitude alone. Moreover, since the dipole source location is known, the problem is a linear one. However, the problem is strongly underdetermined due to the fact that the number of ne electrodes is much smaller than the number of dipoles ns, so regularisation methods are required to compensate for depth bias.

This class of methods produce solutions that show activity over large portions of the brain surface. This is due to the low resolution that results from mapping $c(\approx 10^2)$ electrodes onto $p(\approx 10^4)$ dipoles on the cortical surface. Several imaging methods are developed for the solution of EEG inverse problem keeping in mind low localization error, low computational complexity and validation of the achieved results. For instance, methods that impose $l_2 - norm$ constraints on the source distribution are particularly popular, as they lead to solutions that are linear in the sensor data and therefore efficient to compute. Among these methods, the most popular are minimum norm method (MN)[68], low resolution brain electromagnetic tomography (LORETA)[69], standardized LORETA (sLORETA)[70]and exact LORETA (eLORETA)[71].

The choice of the inverse solution strongly determines how the user interprets the data. Parametric methods can lead to precise and accurate results in the case of focal activation, e.g. in somatosensory stimulations or in analyses of epileptic brain activity. However, converging evidence suggest that brain process can be considered as an integrated network of distributed neural activity. Indeed, an extended activations of neuronal tissue in some conditions cannot be disregarded. In application to cognitive experiments, where the number of active regions in the brain cannot be predicted and large areas of the brain may be involved in the response, dipole models can perform poorly, while distributed models might be more suitable.

Nowadays, parametric approaches have been largely replaced by non-parametric methods. Indeed, also in this work, distributed source imaging techniques has been employed to solve the inverse solution.

In the following, the non-parametric estimation approach to the inverse problem is illustrated with the Bayesian method. This method is based on the Bayes' theorem and consists of finding an estimate of the matrix S that maximises the a posteriori probability of S given the measures M. Denoting by $p(S|M)$ the a posteriori probability of S given the measurements M, by $p(M|S)$ the likelihood function, i.e. the probability of observing the data M given the sources S, and by $p(S)$ the a priori probability of the sources S (which reflects the a priori knowledge of the statistical properties of S), Bayes' theorem states that:

$$p(S|M) = \frac{p(M|S)\, p(S)}{p(M)} \tag{2.15}$$

Knowing posterior distribution, many statistical properties of S could be computed using Markov Chain Monte Carlo methods. In practice, the majority of inverse methods simply look for the S that maximises the posterior density. Bayesian interference estimates S sources as:

$$\hat{S} = arg\ \max_{S}\{p(S|M)\} = arg\ \max_{S}\{p(M|S)\, p(S)\} \tag{2.16}$$

In Eq. (2.16), the term $p(M)$ was neglected as it does not depend on S, so in the maximisation with respect to S it is considered as a constant.

The measurement noise model is described by $p(M|S)$, while all prior information about S is encoded in $p(S)$. Assuming that there are no uncertainties in the direct model and that the only errors are due to noise superimposed on the data, $p(M|S)$ can be expressed by $p(M - AS) = p(N)$.

When no forward model uncertainty is considered, the noise measurement is a white Gaussian process, and the current density is assumed Gaussian with covariance C_s, the maximum a posteriori estimate is the minimizer of the following function:

$$||M - AS^T||_2^2 + tr(SC_s^{-1}S^T) \tag{2.17}$$

The Tikhonov regularized version of the inverse problem[72] considers $C_s = \lambda^{-1}I$, with solution:

$$\hat{S} = (A^TA + \lambda I)^{-1}A^TM \tag{2.18}$$

Where $\lambda^{-1} = \frac{\gamma^2}{\sigma^2}$ is the signal-to-noise ratio, since $\gamma^2$is the variance of the sources while $\sigma^2$is the variance of the noise. Different choices yield to other commonly used linear methods, such as column weighted minimum norm, which is designed to reduce the preference for superficial cortical sources. In this case, $C_s = \lambda^{-1}W$ with $W_{ii} = ||a_i||^2$, where a_i is the *i*-th column of A. In low-resolution electromagnetic tomography (**LORETA**), which employes

Laplacian weighting to regularize the solution, C_s is defined by $C_s^{-1} = \lambda K K^T$ where K is the inversion matrix and:

$$K_{ij} = \begin{cases} 1 & i = j \\ -1/n & j \, \epsilon \, \mathcal{N}(i) \\ 0 & else \end{cases} \qquad (2.19)$$

with $\mathcal{N}(i)$ the set of nearest neighbours of the source location i on the discrete grid and n the cardinal number of $\mathcal{N}(i)$.

However, minimization of Laplacian of the sources leads to a smooth (low resolution) distribution of the 3D activity. This constraint has been justified by the physiologically plausible assumption that activity in neighbouring voxels is correlated.

LORETA provides smooth and better localization for deep sources with less localization errors. However, it leads to low spatial resolution and blurred localized images of a point source with dispersion in the image, which is undesirable in some cases. Thus, improvements of this algorithm have been proposed by this and other authors, leading to more sophisticated algorithms, such as sLORETA[70] or eLORETA[71].

sLORETA and eLORETA: in 2002 a new tomographic method for electric neuronal activity was introduced, where localization inference is based on images of standardized current density. The standardized LORETA also known as sLORETA is based upon the assumption of standardization of the current density. This implies that not only the variance of the noise in the EEG measurements is taken into account, but also the biological variance in the actual signal. This biological variance is considered as independent, as uniformly distributed across the brain, resulting in a linear imaging localization technique having exact, zero-localization error[73].

There have been many attempts to minimize the localization error by choosing the weight matrix in a more adequate way, in order to give more relevance to the deeper sources with reduced localization error. For instance, exact-LORETA, or eLORETA, achieves depth weighting with reduced localization error from 12 to 7 mm. eLORETA is an inverse solution which provides exact localization with zero error in the presence of measurement and structured biological noise. This method suffers from the disadvantage of low resolution like other members of LORETA family. Due to low resolution, undesired blurring is caused in resultant localization images when the space is subjected to regularization for EEG inverse problem.

2.1.2.3 Cortex parcellation:

Once the cortical sources have been reconstructed using the aforementioned algorithms, cortex can be parcellated based on its functio-anatomical organization. Such parcellation reflects the subdivision of the cortex (or other grey matter structures) into functionally and structurally distinct areas.

The idea that the functional similarity structure in the brain is properly represented by structural criteria is rooted in the following two notions: 1) there is a certain parallelism between structural properties and functional specialization in the brain; 2) mapping structure and function onto the cortical surface follows the smoothness principle, since topologically closer parts of the cortex tend to be more similar to each other, leading to areas of relatively homogeneous structural and functional properties[74].

Brodmann's map (1909) represents one of the first attempts of cortex parcellation, and reflects the specific variation in size and packing density of cell bodies in the layers of the cortical surface. A more recent and widely used atlas is, for instance, the Desikan-Killiany[73], which starting from 40 MRI scans identifies 34 cortical ROIs in each hemisphere (see Fig. 2.2).

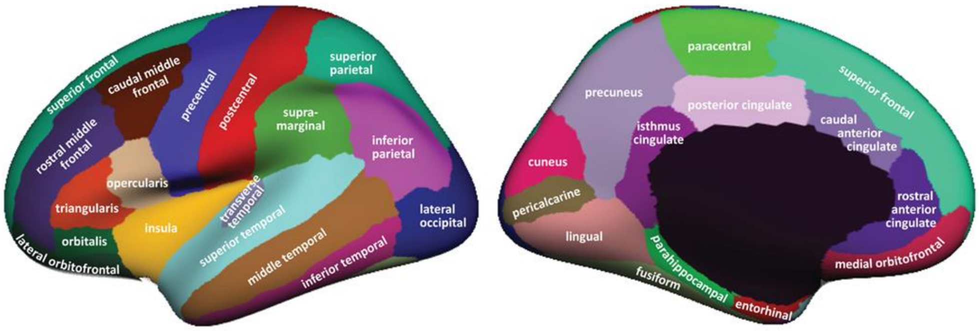


Figure 2.2 – Desikan-Killiany atlas: the left panel illustrates the lateral view of the left hemisphere while the right panel shows the medial view of the left hemisphere[75]

2.2 Neural oscillations and cognitive processes

The human brain is a complex adaptive system in which a vast array of behaviours arises from coordinated neural activity across diverse spatial and temporal scales. Cognition is the most complex function of the brain as requires high selectivity, integration and flexibility[76]. Indeed, an overarching goal of systems neuroscience is to understand the relationship between cognitive functions and underlying neural activity.

Analysis of brain activity at a mesoscopic scale reveals the presence of synchronous oscillations, which cover a large spectrum of frequencies and can be detected using different neuroimaging techniques such as EEG or MEG. These oscillations are not merely an epiphenomenon, but play a crucial functional role in many cortical processes[77]. Indeed, neural oscillations are generated by the rhythmic neural activity that appears throughout different structures of the nervous system such as cerebral cortex, hippocampus, subcortical nuclei and sense organs[78,79]. The rhythmic and synchronized activity of large numbers of neurons (neural populations) can give rise to macroscopic oscillatory electric fields, which can be observed in the EEG. The brain spontaneously generates neural oscillations with a large variability in frequency, amplitude, duration, and recurrence.

A substantial number of studies in literature links the large-scale oscillatory activity of the brain with the dynamics of the fundamental cognitive processes such as memory, attention, and consciousness[80].

The first human EEG oscillation ever described was a Berger alpha rhythm, followed by significant clinical and basic research. From scalp electric potential recordings, researches identified various other oscillatory patterns, particularly obvious during sleep and rest. Over the course of time, almost every cognitive process has been associated with an event-related EEG oscillation[81].

It is generally assumed that EEG signals can be decomposed into distinct brain rhythms (delta: 1–4 Hz, theta: 4–8 Hz, alpha: 8–12 Hz, beta: 12–30 Hz, gamma: 30–70 Hz). However, the frequency spectrum of the EEG signal can also be much broader, ranging from the infraslow, 0.1 Hz, to the very fast, reaching values of several hundreds of Hz. In general, EEG oscillations at low frequencies, such as δ, tend to engage large spatial domains and may represent the cooperative activity of spread neuronal networks in the brain, whereas high-frequency oscillations may predominantly reflect the activity of local neuronal populations — a phenomenon that gives rise to the observed 1/f amplitude characteristics in EEG frequency spectra[82]. Oscillations at intermediary frequencies, such as in the θ and α ranges, are optimal to gate the transfer of information across neural populations, such as those of the hippocampal formation and associated cortical areas in the case of θ[83] and of thalamocortical systems in the case of α[84]. Oscillations at the higher frequencies, in the β and γ range, are especially adequate to engage relatively discrete populations in achieving transfer of packets of specific information among neuronal assemblies[85]. Thus, specific oscillations have different kinds of functional connotations. However, there are many more different cognitive processes than the five well-established frequency bands (δ, θ, α, β and γ). The contribution of neural oscillations to cognitive processes strictly depends not only on the brain region in which they occur but also on their amplitude, frequency and phase. Some examples of the main brain rhythms and the involved cognitive processes are described below.

Delta oscillations (1-4 Hz): The most prominent cognitive correlate of event-related delta activity is detecting a target stimulus in a series of distractors. When a stimulus requiring attentional resources is processed, the Event Related Potential (ERP) reveals a so-called P300 component which is mainly characterised by delta and theta oscillations[86]. Sources of delta oscillations have been observed in frontal and cingulate cortex, and in line with their low frequency these oscillations span a rather wide region of neural networks — possibly in an inhibitory manner[87]. This assumption is in line with a role in cognitive processes such as attention, since attending to one stimulus or location can be achieved by inhibiting other stimuli or locations. These slow brain oscillations also represent the characteristic EEG signature during non-REM sleep.

Theta oscillations (4-8 Hz): Human EEG theta oscillations are most commonly associated with memory processes[88]. It has been assumed that the cortical theta oscillations reflect the communication with hippocampus — a region that is known to serve memory functions and to exhibit oscillations in the theta range[89]. A prominent increase in EEG theta has been consistently reported especially in the fronto-medial region in various cognitive tasks[90], and the increase is more pronounced in the most demanding tasks. For instance, the increase of the frontal midline of the theta is most observable in tasks requiring internally sustained attention, such as working memory tasks[91] and mental arithmetic tasks[92]. These tasks share the need to update, organise and keep online multiple information, for their manipulation and retrieval. Moreover, a source localization EEG study revealed that increased theta activity in the anterior midcingulate cortex is evoked by fear-conditioned stimuli compared to non-fear-conditioned stimuli [93]

Alpha oscillations (8-12 Hz): Many recent studies have overturned the traditional interpretation of alpha activity as reflecting a state of cortical inactivity, focusing more on its central role in attentional processes. Indeed, it exerts a functional inhibition of task-irrelevant processes and brain areas that may interfere with a successful task performance[90,94]. The decrease/increase in alpha power has been associated to cortical excitation/inhibition respectively, based on the adaptive alpha response to task demands. Since oscillations are ideally suited to serve as pacemakers, a recent hypothesis suggests how the two functions of inhibition and timing of the alpha rhythm are crucial for cognitive processes that require both suppression and selection[95]. EEG alpha oscillations are mainly modulated during sensory stimulation[96]. For instance, alpha power decreases especially in bilateral occipital areas to enhance visual processing[97]. In our recent papers[98,99], which investigated alpha power modulations in virtual reality (VR) environment, we found that performing an internally oriented attentional task while immersed in a (highly stimulating) VR setting increased alpha power to the same level as in rest condition, whereas stimulation alone induced a strong decrease. These results suggest that alpha rhythm is crucial to isolate a subject from the environment, and move attention from external to internal cues. Furthermore, we also found that both alpha power and connectivity are modulated by the subjects' feeling of higher/lower comfort in the virtual environment.

Beta oscillations (12-30 Hz): Modulation of human EEG beta oscillations has mainly been observed when subjects perform motor tasks[100]. These oscillations are associated with top-down controlled processing, carry information about task rules, reflect attention to upcoming motor tasks, and may also reflect premotor mechanisms guiding motor actions[101].

Furthermore, beta oscillations are also modulated during cognitive tasks requiring sensorimotor interaction[102]. A recent hypothesis integrating aspects of motor and cognitive processes suggested that beta activity reflects whether the current sensorimotor state is expected to remain stable or to change in due course[103].

Gamma oscillations (30- up to 140 Hz): While many of the low-frequency oscillations have been associated with functional inhibition, faster gamma-band oscillations are believed to reflect cortical activation[104]. Depending on the exact cortical region, gamma oscillations are closely related to attentive processing of information[105], active maintenance of memory content. A prominent gamma rhythm provides a signature of engaged networks. In sensory cortex, gamma power increases with sensory drive[106], and with a broad range of cognitive phenomena, including attention[107]. At a given recording site, gamma is stronger for some stimuli than others, generally displaying selectivity and a preference similar to that of nearby neuronal spiking activity[108]. Interestingly, irregular gamma activity has been observed in neurological disorders such as Alzheimer's disease, Parkinson's disease, schizophrenia, and epilepsy[31] .

However, most often oscillations at different frequencies work in a cooperative, integrated way, so that more than one brain rhythm can temporally coexist in the same or different structures of the brain and influence each other. Indeed, through long range connections the oscillation in one region may be transmitted to other regions, facilitating information integration in the brain.

Detailed biophysical studies revealed that neurons are endowed with complex dynamics, including their intrinsic abilities to resonate and oscillate at different frequencies, suggesting that the precise timing of their activation within a brain network may be a crucial factor in information transmission. Other studies demonstrated that perception, memory and consciousness, which are grounded on information integration, result from synchronized neural patterns distributed among the brain. Indeed, coherent EEG oscillations in two distant brain regions may reflect the functional cooperation of these two regions.

The synchronous activity of oscillation networks is now viewed as the 'middle ground' of information integration, linking neural activity to behaviour and cognition.

2.3 Brain Connectivity Estimation

Nowadays, a key challenge in neuroscience and neuroimaging is to move beyond identification of regional activations toward the characterization of functional circuits underpinning perception, cognition, behaviour and consciousness.

If we understand the brain as a functional distributed network of interacting neural populations, the connectivity of a particular cortical area with the others is crucial to its functionality[109]. This consideration leads to the idea that the functional organisation of the brain is reflected in its connections. Recently, there has been a growing interest in studying how brain areas connect in both normal and pathological conditions[110].

Two characteristics of brain connectivity play a major role: synchronization and causal influence. First, communication of neurons within an assembly is achieved through the synchronous activity of the participating neurons. General network synchronization may lead to simultaneous and coherent activity in many brain regions. Indeed, EEG synchronization in the alpha-, beta-, theta-, and-gamma bands has been associated with memory, sensory integration, attention, and consciousness, respectively[111]. Second, information flow leads to a causal relationship between activities in different regions. For example, it is well documented that visual evoked potentials measured through electroencephalography (EEG) source localization seem to have a larger latency in the downstream areas as compared to the upstream ones[112]. This suggests a causal relationship between activities in these brain areas.

There are two approaches to functional coupling: functional and effective connectivity. Functional connectivity is defined as the statistical dependence among measured time series, and until recently was usually assessed in terms of correlations or mutual information. These measures have recently been supplemented with causal metrics such as Granger causality and Transfer Entropy which provide information about the causality of the interaction. In contrast, effective connectivity quantifies the causal influence of one neuronal system over another and relies upon a model of neuronal coupling[10,113], such as dynamic causal modeling (DCM) and structural equation modelling. This framework requires strong a priori knowledge about the input to the system and the connectivity network. To overcome this limitation, the more suitable network is often chosen among various possible alternatives using Bayesian selection methods.

However, as previously discussed in Chapter 1, the understanding of directional functional connectivity estimators and their limitations has been the focus of this PhD thesis. Below, a general overview of the main FC measures and their mathematical formulation is presented.

2.3.1 Functional Connectivity

As described in Section 2.2, neuronal oscillations reflect synchronized rhythmic activity patterns of local neuronal ensembles and establish the dynamic coordination in the brain. Indeed, the synchronisation of neural oscillations between brain regions facilitates the information flow in the cortical network. The brain can dynamically coordinate the flow of information by changing the strength, pattern, or the frequency with which different brain areas are engaged in oscillatory synchrony.

In the field of neuroimaging, functional connectivity (FC) describes the statistical relationship between the signals of anatomically separated brain regions, reflecting the level of functional communication between them. Examining the brain as a network of functionally interacting brain regions provides a cue to examine how functional connectivity relates to human behaviour, and how the resulting network organization may be altered in neurodegenerative diseases[11].

Literature provides a huge number of metrics to quantify FC[22], and each one has its advantages and drawbacks. In order to classify these variety of metrics, a first subdivision can be made by considering the causality feature of some estimators, which are able to discriminate the direction (or causality) of the interaction (Fig. 2.3). It should be noted that this differentiation exists only due to the definition of functional connectivity proposed by Reid et al. in Chapter 1.

Non-directed FC metrics capture some form of interdependence between signals, without reference to the direction of the interaction. In contrast, *directed* measures assess a statistical causation from the data, based on the principle that causes precede and help predict their effects. Indeed, if a signal can be predicted by the past information from a second signal better than the past information from its own signal, then the second signal can be considered causal to the first signal. In neuroscience, a rich and growing literature has evolved in the use of directed functional connectivity estimators to quantify neuronal interactions [110,114]. The hypothesis of the study is that perceptual, motor and cognitive functions reflect the synchronization of neural assemblies in local or distant regions of the brain and their consequent causal (directional) interaction. Therefore, the causality of these interactions is a central aspect that allows us to grasp a broader spectrum of features of brain networks. Indeed, this thesis mainly focuses on directional FC estimators.

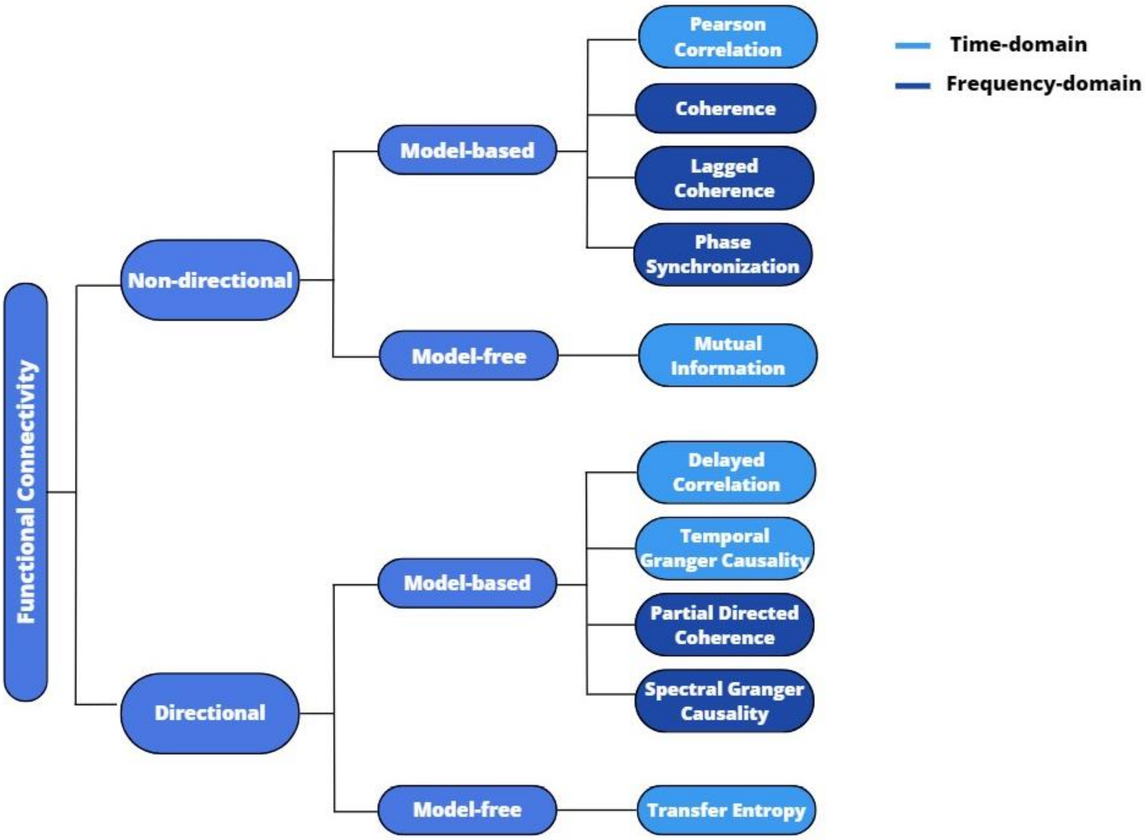


Figure 2.3 – Taxonomy of the most popular Functional Connectivity estimators. A first subdivision is based on whether the metric quantifies the direction of the interaction. Then, within both directional and non-directional types of estimates, a distinction can be made between model-free (based on information theory) and model-based (i.e. assumption of linear interaction) approaches. Further, another important differentiation can be made between metrics that are computed from the time (light blue) or frequency (dark blue) domain of the signals.

Furthermore, within both *directed* and *non-directed* types of estimators, a distinction can be made between *model-based* and *model-free* approaches. Model-based approaches make the strong assumption of linear relationships between time series of two neural populations, while model-free make no assumptions about the type of interaction and are generally based on Information Theory[115]. However, the main advantage of model-based methods is that they are easily convertible to the frequency domain, which entails the advantage of allowing the study of functional connectivity for a given brain rhythm.

In the following, the main *model-based* and *model-free* functional connectivity measures are presented.

Model-based:

The simplest measures of **non-directed model-based** interactions are correlation and coherence. The Pearson correlation coefficient measures the linear relationship between two random variables. In the general linear modeling framework, the squared correlation coefficient R^2 represents the fraction of the variance of one of the signals that can be explained by the other, and vice versa.

Coherence is a frequency-domain measure that allows the spatial correlations between signals to be measured in different frequency bands. It provides an information about the stability of relationship between two signals, with respect to power asymmetry and phase.

Correlation and coherence are only few examples of the metrics that have been used in the electrophysiological literature to estimate non-directed model-based functional connectivity. For instance, other non-directed model-based measures used in this thesis are Lagged Coherence and Phase Synchronization, both formulated in the frequency-domain.

However, as discussed above, our interest primarily focuses on causal measures of connectivity. A first example of **directed model-based** methods, is the delayed version of correlation. Yet, when we shift two time series with respect to one another, and evaluate the correlation as a function of time lag, we can infer directed interactions. Hence, both the time lag that maximize the correlation and the magnitude of correlation can be informative about information flow between brain areas. However, in the case of bidirectional interactions, which is the dominant scenario in cortico-cortical connections, some issues may arise. Indeed, outcomes typically lack a clear peak, and have significant values at both positive and negative lags, indicating complex, bi-directional interactions that occur at multiple delays. To address this limitation, more sophisticated estimators, such as Partial Directed Coherence and Granger Causality, have been formulated.

The Partial Directed Coherence [116] is a linear spectral quantifier which reveals the directed functional relationship between any given pair of signals in a multivariate data set. Whereas, the idea beyond Granger Causality can be traced to Norbert Wiener[17], but the method was finalised by the econometrician Clive Granger[18] in terms of linear autoregressive (AR) modeling of stochastic time series. AR models are simple mathematical models in which the value of a variable at a particular time is described as a linear weighted sum of its own past, considering a number of time-steps, plus a random noise. Each variable represents a stochastic process, or the time series of a neural population (e.g. signal of a specific brain region). Hence, fitting an AR model means to find the optimal weights in order to minimize the estimation errors. The AR coefficients (weights) are derived such that the corresponding linear combination of the past values of the signal provides for the best possible (in the least squares sense) linear prediction of the current value. In practice, the AR approach reduces to a method for estimating these coefficients and using those to compute the interaction measures.

According to Wiener's maxim, a variable *X* causes a second variable *Y,* if the past of *X* contains information that helps predict the future of *Y*, over and above the information already in the past of *Y* itself. In the case of Granger Causality, a bivariate (*X*-*Y*) and univariate (*Y*) AR model is estimated, as well as the related prediction errors. If the prediction error of the bivariate representation is lower than the one in the univariate condition, we can say that *X* causes *Y*. An important generalization of Granger Causality in the frequency domain was provided by Geweke in 1982[117], that enabled the coupling assessment between different EEG frequency bands that have a well-known biomedical significance.

When using Granger Causality, one needs to bear in mind that it is based on the comparison between model errors and thus its use only makes sense when considering random (or stochastic variables). Another intrinsic assumption is that the data are (weakly) stationary, which means that the means and variances of the variables are stable over time.

Given these assumptions, Granger causality has different advantages. First, it is easy to compute since there are several standard algorithms for optimal estimation of AR. Second, because these models are very general, no a priori assumptions on the underlying physical mechanisms are needed. Third, as all model-based (or linear) measures, AR models can be easily transformed into the frequency domain, allowing spectral estimations of Granger causality. Indeed, this metric offers a simple yet powerful means for characterizing information flow both in time and frequency domains, making minimal assumptions about the underlying generative mechanisms.

However, Granger Causality also suffers from some limitations. Indeed, AR methods requires estimation of model order - i.e., the number of past observations (time-steps) - to be included in the models. This aspect is crucial since the use of different orders may result in different conclusions. Recently, regularization-based algorithms have been implemented that estimate the optimal model order from experimental data[118].

Furthermore, Granger's classical formulation involves a bivariate approach, where connectivity is estimated separately for each pair of brain regions in the network. However, brain activity measured on different sites is highly correlated. Indeed, there exists a multitude of relations between different brain regions. In such a situation it is difficult to judge if two given regions interact with each other, or if they are driven by a third region. As Granger Causality developed, the formulation was generalized from bivariate to multivariate approach[119,120]. Indeed, through the multivariate approach one can either fit a full model where all variables (brain regions) are taken into account. The classic bivariate approach typically yields more stable results since it involves the fitting of fewer parameters. However, the advantage of the multivariate approach is that information from all regions is considered when estimating the interaction terms between any pair of sources, thus enabling the distinction between direct and indirect interactions.

Although Granger approach has been widely used for causality estimation from EEG signals, it is limited to modeling only the linear (i.e., Gaussian) component of the interactions[121]. This clearly imposes a limitation when applied to nonlinear systems like the brain. For instance, significant physiological processes such as epilepsy[122] violate the Gaussianity assumption. In these cases, AR models may either misallocate the non-linearities, or ignore them entirely. Nevertheless, many nonlinear systems have linear or quasi-linear domains of applicability, and within this domain, AR models are able to capture significant properties of the system behaviour.

Importantly, since Granger Causality is not grounded on an underlying model of how the signals are generated, it is a phenomenological (data-driven) method that does not consider the representation of a neurobiological process. While this may seem at first glance to be a limitation, it brings with it the great advantage of assessing directional influences of a region on another without any *a priori* hypothesis.

Model-free:

Model-free methods are more generalized approaches that do not assume linear relationship between time series. These methods are Information-based techniques and provide FC measures that are sensitive to both linear and nonlinear statistical dependencies between two time series. However, the main disadvantage of this technique applied to EEG data is that no spectral formulation exists. Before describing some of the information theoretic measures, we first provide a brief background on the key concepts that underlie the specific information theoretic measures of interest.

Information theory (IT) sets a powerful framework for the quantification of information and communication[115] and provides an ideal basis to precisely formulate causal hypotheses. In the context of information theory, the key measure of information of a discrete random variable is its Shannon entropy. This entropy quantifies the reduction of uncertainty obtained when one actually measures the value of the variable. First attempts to obtain **non-directed model-free** measures of the interdependence between two random variables were made through mutual information (MI)[123]. This measure quantifies the amount of information that can be achieved on a random variable, by observing a second variable. MI is based on probability distributions and is sensitive to second and all higher order correlations. However, MI does not bring information about causality: it is symmetric and captures the amount of information that is shared by two signals.

To achieve a **directed model-free** measure, the quantification of information has been linked to Wiener's definition of causal interactions, based on an increase of prediction power. Bearing in mind his definition, if we can associate a prediction enhancement to an uncertainty reduction, a causality measure may naturally be expressible in terms of information theoretic concepts. A rigorous derivation of a Wiener causal measure within the information theoretic framework was published by Schreiber under the name of Transfer Entropy (TE)[40]. This estimator is a more generic implementation of the maxim that causes must precede and predict their effects, and can detect non-linear forms of interaction, which may remain invisible to linear approaches like Granger causality. It has been shown that when the data are Gaussian (i.e., normally distributed) Granger causality is an approximation to Transfer Entropy[124]. This means that under gaussian conditions Granger causality values can be interpreted in terms of information flow.

Transfer Entropy, as Granger Causality, is of great relevance to understand how the brain exchange information in the brain, but in some conditions this may be intrinsically different from causal strength[125]. Indeed, information is not a direct measure of coupling strength, and should be used with extreme caution to measure a coupling parameter (such as the weight of synapses among two neural populations). This aspect is discussed in detail within Chapter 3.

2.3.1.1 Mathematical formulation

The mathematical formulation of the main functional connectivity estimators is given below. Let us assume that the presynaptic and postsynaptic signals are described by two discrete stochastic processes (say ***x***[*n*] and ***y***[*n*], respectively, where we use the boldface to denote a random variable). In the following, we will use *x*[*n*] and *y*[*n*] (without bold) to represent two particular realizations of the stochastic processes, where *n* is the discrete time (*n* = 0, 1, …, *N*-1)[126].

Pearson correlation coefficient: the expression of this is:

$$r_{yx} = \frac{\sum_{i=0}^{N-1}(x[i]-\bar{x})(y[i]-\bar{y})}{\sqrt{\sum_{i=0}^{N-1}(x[i]-\bar{x})^2\sum_{i=0}^{N-1}(y[i]-\bar{y})^2}} \tag{2.20}$$

where $\bar{x}$ and $\bar{y}$ represent the average values of the corresponding quantity. Of course, this estimator is nondirected (i.e., $r_{yx} = r_{xy}$). Moreover, it can be positive or negative to discriminate between excitatory or inhibitory connections.

Delayed correlation coefficient: this coefficient differs from the previous one since the postsynaptic signal is delayed, assuming that a finite time is necessary to propagate information from *x* to *y*. Hence, we can write:

$$dr_{yx} = \max_{d}\left|\frac{\sum_{i=0}^{N-d-1}(x[i]-\bar{x})(y[i+d]-\bar{y})}{\sqrt{\sum_{i=0}^{N-d-1}(x[i]-\bar{x})^2\sum_{i=0}^{N-d-1}(y[i+d]-\bar{y})^2}}\right| \tag{2.21}$$

where *d* is the delay (expressed as the number of samples); hence, if Δ*t* is the sampling period, the overall temporal delay is *d*·Δ*t*. It is worth noting that, *d* is usually chosen as the value that maximizes the absolute value in Expression (2.21). Thus, the delayed correlation coefficient can assume a positive or negative value and is a directional measure of connectivity.

Coherence: this estimator is computed as the magnitude squared coherence function:

$$C_{yx}(f) = \frac{\left|P_{yx}(f)\right|^2}{P_{xx}(f)P_{yy}(f)} \tag{2.22}$$

where f is frequency, $P_{yx}(f)$ is the cross-spectral density of the two signals, $P_{xx}(f)$ is the power spectral density of *x* and $P_{yy}(f)$ is the power spectral density of *y*. Coherence, of course, is a nondirected estimator ($C_{yx}(f) = C_{xy}(f)$) and provides an estimate at each frequency of the discrete power spectra.

Lagged coherence: a possible limitation in the use of coherence is that it is affected by zero-lag (instantaneous) correlations, which can artificially inflate the estimated values. Among the measures proposed to mitigate this issue, the lagged coherence was developed by Pascual-Marqui at al. [127] to detect physiological lagged connections between brain regions and is not affected by volume conduction and by low spatial resolution. It is defined as follows

$$LC_{yx}(f) = \frac{\left(Im\{P_{yx}(f)\}\right)^2}{P_{xx}(f)P_{yy}(f) - \left(Re\{P_{yx}(f)\}\right)^2} \tag{2.23}$$

where $Im\{\cdot\}$ and $Re\{\cdot\}$ denote the imagery and real part of the corresponding complex-valued argument and the remaining symbols on the right have the same meaning as in Equation (2.22). Lagged coherence is a nondirected connectivity measure ($LC_{yx}(f) = LC_{xy}(f)$).

Phase synchronization: to estimate phase synchronization (see [128,129]), the analytical signal is computed as:

$$Z_x[n] = x[n] - \bar{x} + jH[x[n] - \bar{x}] = A_x[n]e^{j\varphi_x[n]} \tag{2.24}$$

where H[.] denotes the Hilbert transform and $j = \sqrt{-1}$. Usually, the average value of the signal is subtracted before computing the analytical form. Of course, a similar expression holds for the *y* signal too. Then, the phase difference between the two signals at any discrete time *n* is:

$$\Delta\varphi_{yx}[n] = \varphi_y[n] - \varphi_x[n] \tag{2.25}$$

Phase synchronization is finally obtained by estimating the quantity: $\left|E\left\{e^{j\Delta\varphi_{yx}}\right\}\right|$, where E{.} represents the statistical mean value. The latter is estimated as follows:

$$PS_{yx} = \left|\frac{1}{N}\sum_{i=0}^{N-1} e^{j\Delta\varphi_{yx}[i]}\right| \tag{2.26}$$

which provides a scalar nondirected quantity $PS_{yx} = PS_{xy}$.

Time-Domain Granger Causality: this estimate is based on the autoregressive (AR) modeling framework and compares the prediction ability of two AR models of the same process ***y***[*n*]—i.e., a univariate AR model and a bivariate AR model; in the latter, the current value of the process ***y***[*n*] was predicted not only based on its past values (as in the univariate case), but also on the past values of the other process ***x***[*n*]. Specifically, we can write

$$\boldsymbol{y}[n] = \sum_{k=1}^{p} a\,[k]\boldsymbol{y}[n-k] + \boldsymbol{\eta}_{\boldsymbol{y}}[n] \tag{2.27}$$

$$\boldsymbol{y}[n] = \sum_{k=1}^{p} b\,[k]\boldsymbol{y}[n-k] + \sum_{k=1}^{p} c\,[k]\boldsymbol{x}[n-k] + \boldsymbol{\varepsilon}_{\boldsymbol{y}}[n] \tag{2.28}$$

for the univariate and bivariate AR model, respectively, where *p* is the order of the model. $\boldsymbol{\eta}_{\boldsymbol{y}}[n]$ and $\boldsymbol{\varepsilon}_{\boldsymbol{y}}[n]$ are white noise processes and represent the model's residual (or prediction error) in each case. The variance of the residual (let us say γ and σ_{yy}, respectively) quantifies the quality of the model fit. The Granger causality from *x* to *y* in the time domain is defined as [130,131]:

$$GC_{yx} = \ln\frac{var(\boldsymbol{\eta}_{\boldsymbol{y}}[n])}{var(\boldsymbol{\varepsilon}_{\boldsymbol{y}}[n])} = \ln\frac{\gamma}{\sigma_{yy}} \tag{2.29}$$

where, of course, in practice the variances will be estimated on the particular realizations of the residuals. A substantial reduction in the variance of the residual in case of the bivariate compared to univariate model means that including the past values of *x* provides a better prediction model for *y*, and GC_{yx} is substantially larger than 0—i.e., *x* casually influences *y* in the Granger sense. Similarly, Granger causality from *y* to *x*, GC_{xy}, was computed via the same procedure, building the AR models for the process ***x***[*n*]. Granger causality is a directed connectivity estimator ($GC_{yx} \neq GC_{xy}$).

Frequency-domain (spectral) Granger causality: Granger causality can be formalized in the spectral domain [131,132] starting from the joint bivariate autoregressive representations of the two processes:

$$\sum_{k=0}^{p} A[k]\begin{bmatrix}\boldsymbol{x}[n-k]\\ \boldsymbol{y}[n-k]\end{bmatrix} = \begin{bmatrix}\boldsymbol{\varepsilon}_{x}[n]\\ \boldsymbol{\varepsilon}_{\boldsymbol{y}}[n]\end{bmatrix} \tag{2.30}$$

Equation (2.30) is derived from Equation (2.28) and the analog one expressing the bivariate model of $\boldsymbol{x}[n]$; $A[k]$ are 2 × 2 coefficient matrices (identity matrix at time lag 0). After Fourier transforming Eq. (2.30), we manipulated it to obtain

$$\begin{bmatrix} X(f) \\ Y(f) \end{bmatrix} = \begin{bmatrix} H_{xx}(f) & H_{xy}(f) \\ H_{yx}(f) & H_{yy}(f) \end{bmatrix} \begin{bmatrix} \mathrm{E}_x(f) \\ \mathrm{E}_y(f) \end{bmatrix} = H(f) \begin{bmatrix} \mathrm{E}_x(f) \\ \mathrm{E}_y(f) \end{bmatrix} \qquad (2.31)$$
$$H(f) = A^{-1}(f)$$

This is the transfer function matrix. By right multiplying each side of Equation (2.31) by its conjugate transpose (*), the cross-spectral density matrix $S(f)$ for signals *x* and *y* can be expressed as

$$S(f) = \ H(f)\Sigma H(f)^{*} \qquad (2.32)$$

where $\Sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix}$ is the covariance matrix of the residuals (white noise processes) in Equation (2.30). The spectral Granger causality from *x* to *y* is computed as (for further mathematical details see [131])

$$sGC_{yx}(f) = \ln \frac{P_{yy}(f)}{P_{yy}(f) - \left(\sigma_{xx} - \frac{\sigma_{xy}}{\sigma_{yy}}\right) \left|H_{yx}(f)\right|^{2}} \qquad (2.33)$$
$$= \ln \frac{P_{yy}(f)}{P_{yy}(f) - \widetilde{\sigma_{xx}} \left|H_{yx}(f)\right|^{2}}$$

The numerator expresses the total power spectrum of *y* at frequency f, while the denominator is the difference between the total power spectrum and the “causal” power exerted by signal *x* on signal *y* at the same frequency. Accordingly, the quantity sGC_{yx} at a given frequency *f* is zero when the causal power of *x* onto *y* at *f* is zero and increases (>0) as the causal power increases. The spectral Granger causality from *y* to *x*, $sGC_{xy}(f)$, was obtained from Equation (2.33) by exchanging the subscripts *y* and *x*. Of course, this connectivity measure is directional ($sGC_{yx} \neq sGC_{xy}$).

Transfer Entropy: to calculate the transfer entropy from *x* to *y*, one first needs to construct the embedded vectors.

$$X^{m}[n] = [x(n)\, x(n-\Delta n)\, x(n-2\Delta n) \ldots\, x(n-(m-1)\Delta n)]$$
$$Y^{h}[n] = [y(n)\, y(n-\Delta n)\, y(n-2\Delta n) \ldots\, y(n-(h-1)\Delta n)]$$

In previous equations, *m* and *h* are the embedding dimensions, defining the number of past samples used, and Δn represents the embedding delay. These parameters serve to approximately reconstruct the state spaces of the pair of time series. Each vector, $X^m[n]$ and $Y^h[n]$, comprises the present and *m-1* (or *h-1*) past samples of the particular realization of the random process.

The concept behind TE is that, in case of a causal influence from *x* to *y*, the probability of **y**[n], conditioned by its past $\boldsymbol{Y}^h[n-\Delta n]$ only, should be lower than the probability of **y**[n] conditioned by both its past $\boldsymbol{Y}^h[n-\Delta n]$ and the past of the other signal $\boldsymbol{X}^m[n-\Delta n]$. This concept can be formalized by computing the corresponding reduction in entropy as the Kullback–Leibler divergence between the two probability distributions [133]. However, as discussed by Wibral et al. [134], the influence of a neural signal on another takes some time (e.g., tens of milliseconds) to be effective due to the traveling time of the action potential along the axons from the presynaptic to the postsynaptic region—that is, a pure delay (say *d*) in the neural interactions must be taken into account. If we assume that *d* can be approximated by *l* embedding delays (*d* = *l*·Δ*n*), $\boldsymbol{X}^m[n-\Delta n]$ can be replaced by the delayed signal $\boldsymbol{X}^m[n-l\Delta n]$ in the definition of TE. In practice, *l* is generally unknown, and it needs to be estimated from the available data (see below). Based on this description, we acquired:

$$TE_{yx} = \sum_{\substack{y[n] \\ Y^h[n-\Delta n] \\ X^m[n-l\Delta n]}} p(y[n], Y^h[n-\Delta n], X^m[n-l\Delta n]) * \\ * \, log_2 \frac{p(y[n]/Y^h[n-\Delta n], X^m[n-l\Delta n])}{p(y[n]/Y^h[n-\Delta n])} \tag{2.34}$$

Of course, transfer entropy is directional—i.e., $TE_{yx} \neq TE_{xy}$. A fundamental problem in the evaluation of Equation (2.34) is that various joint and marginal probability distributions (with very large dimensionality, up to *m* + *h* + 1) must be evaluated starting from the finite data samples. Moreover, several parameters (such as the embedding dimensions *m* and *h*, the embedding delay Δ*n* and the overall transmission delay, *l*Δ*n*) are unknown and require estimation from the data. In this study, we used the software package Trentool [135,136] to estimate TE from the outputs of the neural mass model. More details on the implementation can be found in our work [137].

Some of the metrics adopted to estimate functional connectivity are frequency-dependent (coherence, lagged coherence, spectral Granger causality). In these cases, in order to derive a single value for each bivariate connection, the mean values of the estimated connectivity profile over the entire range of frequencies are usually extracted.

Taken together, methods such as Granger Causality (AR models) and Transfer Entropy form the basis for causal (directed) functional connectivity estimation from EEG data. Therefore, Chapters 3 and 4 mainly focuses on these estimation metrics.

2.4 **Graph Theory**

Graph theory, developed by Euler in 1736, is rooted in the physical world and represents a mathematical method to formally describe and analyse graphs. A graph is defined as a set of nodes (vertices) linked by connections (edges). When describing a real-world system, a graph provides a representation of the network's elements and their interactions[11].

Since the mid-1990s, developments in the physics of complex systems have led to the rise of network science as a transdisciplinary effort to characterize network structure and function.

Following the emergence of promising results in electrical circuits and chemical structures in its early applications, graph theory has now become influential in addressing a large number of practical problems in other disciplines, such as transportation systems, social networks, big data environments, the internet of things, electrical power infrastructures, and biological neural networks. In all this cases, a complex real-world system is shaped by a collection of pairwise interconnected elements, where complexity arises in the macroscopic behaviour of a system of interacting elements.

Neural systems have long been described as sets of discrete elements linked by connections. Indeed, the human brain is a complex system, in which approximately 86 billion neurons interact through approximately 150 trillion synapses[138]. We have known since the nineteenth century that the neuronal elements of the brain constitute a formidably complex structural network. Since the twentieth century it has also been widely acknowledged that this anatomical substrate supports the dynamic emergence of coherent physiological activity, that can span the multiple spatially distinct brain regions that make up a functional network[139]. Nonetheless, the turning point of the complex brain network studies using graph theory goes back to the introduction of the "Human Connectome" [140] [141].

Like many complex systems, the brain exhibits a very wide range of dynamic activity and connectivity patterns that are thought to be instrumental for enabling the integration and processing of information in the course of behaviour and cognition[142]. Indeed, its information processing system needs to be highly flexible and adaptive in order to control body functions, interpret information from the outside world and embody the essence of mind.

Over the past two decades, the combination of non-invasive neuroimaging techniques with graph theory has allowed to map brain networks (i.e., the connectome) at the macroscopic level. There is a growing number of studies in which the brain is modelled as a complex network based on neural units (usually brain regions) connected by structural connectivity (structural pathways) or functional connectivity (time-dependent activities) [6,11]. According to literature there is an interest in understanding the brain's functional connectivity patterns when the subject is at rest, in the absence of specific cognitive demands, as well as its reorganization during task performance, by building graph models of brain networks[24,143]. Recent studies demonstrated that brain network organization undergoes changes during development and ageing[144], as well as in the case of brain disorders[30] , providing novel insights into the neurophysiological mechanisms in health and disease. Analytical approaches capable of capturing the properties of brain networks can enhance our ability to make inferences from functional MRI, EEG and MEG data.

Graph theory offers a wide range of theoretical tools to quantify specific features of brain network architecture (topology) that can provide information complementing the anatomical localization of areas responding to given stimuli or tasks (topography)[9]. Explicit modelling of the interactions among areas can furthermore reveal peculiar topological properties that are conserved across diverse biological networks, and highly sensitive to disease states. The field is evolving rapidly, partly fuelled by computational developments that improved brain imaging techniques and enabled the study of connectivity among multiple brain regions.

2.4.1 Complex Brain network construction:

Complex networks are represented as an interconnected sets of nodes and their pairwise edges, often summarized in the mathematical form of adjacency matrix (also known as a connectivity matrix). The anatomical location of the nodes constituting a given network, referred to as topography, is integrated with a representation of the architecture of their connections, referred to as topology. The latter aspect can be studied through structural and functional connectivity, which defines the relationships between the nodes.

Graph theory approach allows the characterization of intrinsic topological organisation of the network, capturing aspects such as highly connected or centralised nodes, small-worldness and modular organisation[11]. The brain topology is important for the overall function, performance and behaviour. Its biological relevance derives from the fact that certain structural properties of connectivity emerge as optimal trade-offs maximizing information-processing capability and speed with respect to the physiological ‘cost’ of synaptic and axonal metabolism.

Through EEG data we can generate a graph characteristic of the brain network, which can then be topologically described extracting graph theory measures. The first step to generate the graph is to define the network nodes. Depending on the spatial scales of interest, a brain network can involve tens to hundreds of nodes, and can be defined in various ways: neurons, neuronal populations, electrodes or brain regions. Usually, the nodes are anatomically defined as brain regions, which depend on the choice of the atlas and cortex parcellation.

The second step involves the edges definition. In the case of structural connectivity it refers to the anatomical pathways between brain regions, while in the case of functional connectivity is typically defined as the statistical dependency between time series, recorded by EEG, MEG or functional MRI (fMRI)[10], and represents respectively the structural or functional connectivity linking the nodes. Once the connectivity matrix has been estimated (e.g., Granger Causality, Transfer Entropy) it has to be converted in the adjacency matrix, indicating the edges between each pair of nodes in a graph. Thus, a network with N nodes can be represented by an $N \times N$ adjacency matrix, in which the non-zero elements reflect the presence or strength of an edge between two nodes.

A graph may be categorized as directed or undirected, depending on whether the edges between nodes contain directional information (e.g., causal interaction). For undirected graphs the adjacency matrix is symmetric, while for directed graphs it is asymmetric. A graph can also be classified as binary or weighted[145]. Binary links denote the presence or absence of connections, while weighted links also contain information about connection strengths. In functional networks, weights may represent respective magnitudes of correlational or causal interactions. Furthermore, weak and non-significant connections may represent spurious connections induced by noisy signals and tend to obscure the topology of strong and significant connections. For these reasons they are often discarded, by applying an absolute, or a proportional weight threshold value. A common threshold method is for instance to retain only those connections that statistically differ between two conditions ($p - value < 0.05$). The threshold is applied to each element of the matrix in order to obtain a sparse adjacency matrix.

Once we have established an adjacency matrix of a brain network, we can assess the topological properties of the network using the metrics developed in graph theory (Fig 2.4). Then, the extracted brain network metrics can be compared to the equivalent parameters of another population/condition or of random networks. Statistical testing of network parameters is typically conducted by permutation - or resampling-based methods of non-parametric inference, given the lack of statistical theory concerning the distribution of most network metrics.

Each step can affect the final results, including the parcellation scheme (number of nodes and edges), the functional connectivity estimator (directional/ non directional), the chosen threshold (sparsity, fixed/variable) and the extracted network's parameters.

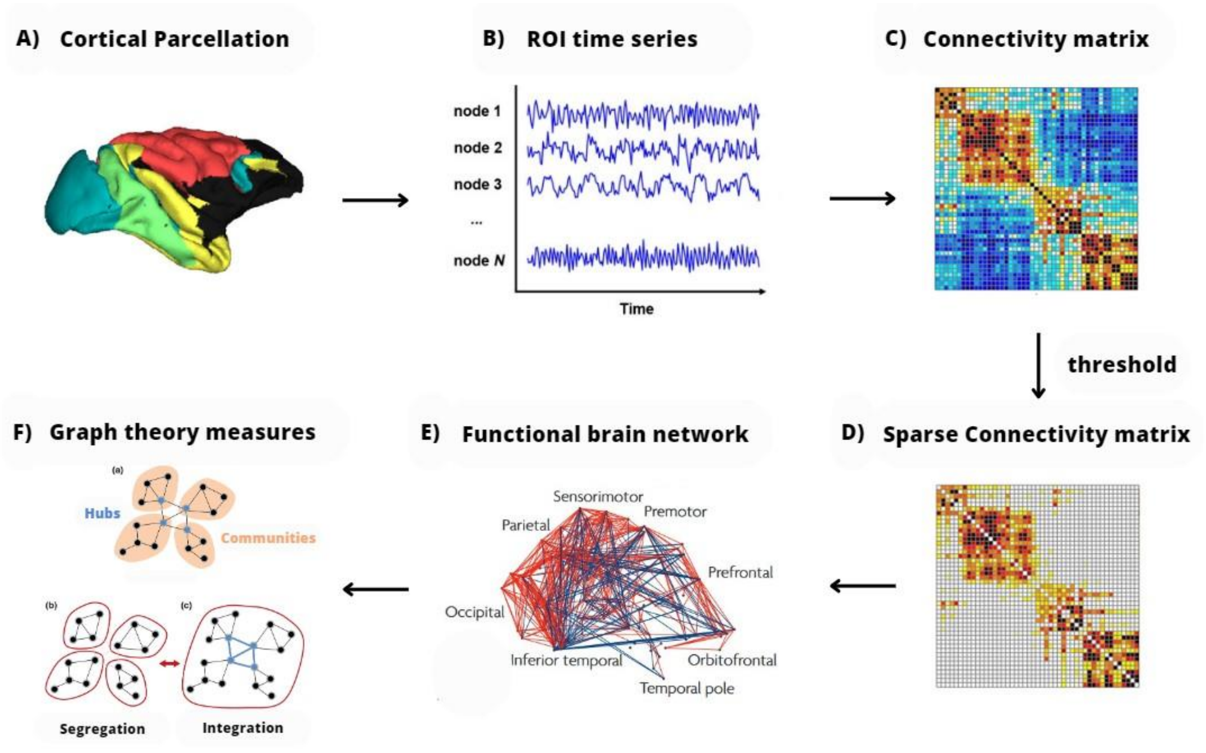


Figure 2.4 - Schematic representation of brain network construction and graph theoretical analysis using EEG data. After source reconstruction and subdivision of the brain into different cortical areas (ROIs) (A), a time course is extracted from each ROI (B) so that they can create the connectivity matrix (C). To reduce spurious connections and noise effects, the sparse connectivity matrix is extracted (D), and the corresponding functional brain network (E) is constructed. Eventually, by quantifying a set of topological measures, graph analysis is performed on the brain's connectivity network (F).

2.4.2 Brain Network Measures

Graph topology can be quantitatively described by a wide variety of measures. Some of them are discussed below.

Measures of Centrality*: c*rucial brain regions (nodes of the graph) often interact with many other regions, thus supporting functional integration. In general, a node with high centrality is crucial to efficient network communication. One of the most common measures of centrality is the *degree* of a node, which is equal to the sum of links connected to that node. The degree has a straightforward neurobiological interpretation: nodes with a high degree are interacting, structurally or functionally, with many other nodes in the network.

Since edges direction may be informative for some networks, degree centrality can be easily adjusted to recognize the existence of a directed network, where two different indices are distinguished: *in degree* defined as the sum of connection strengths entering to a node and *out degree* defined as the sum of connection strengths leaving a given node. Their mathematical description is given below.

In the following, A indicates a generic adjacency matrix (i.e., a matrix containing all edges' weights). In particular, the element $A_{i,j}$ of the matrix will represent the weight of the edge connecting node *i* to node *j*.

$$In\ degree_i = \sum_j A_{j,i} \tag{2.35}$$

$$Out\ degree_i = \sum_j A_{i,j} \tag{2.36}$$

Two analogous but more specialized measures of centrality, *hubness* and *authority*, can provide additional information to characterize directionality. Due to their definition these two measures are highly circularly interdependent. The hub's index of a node is defined as the weighted sum of the authority's indices of all its successors; hence, this measure summarizes the capacity of a node to send information to other critical, authoritative nodes. The authority's index of a node is defined as the weighted sum of the hub's indices of all its predecessors and summarizes the capacity of a node to receive essential information from hubs. Their mathematical formulation is the following one.

Hubness (y_i) is proportional to the sum of the weights of edges exiting from a node, multiplied by the *authority* of the node the edge points to.

$$y_i = \beta \sum_j A_{i,j} x_j \tag{2.37}$$

Authority (x_i) is proportional to the sum of the weights of edges entering a node, multiplied by the *hubness* of the node the edge originates from.

$$x_i = \alpha \sum_j A_{j,i} y_j \tag{2.38}$$

Many other measures of centrality are based on the idea that central nodes participate in many short paths within a network, and consequently act as important controls of information flow. For instance, *closeness* centrality is defined as the inverse of the average shortest path length from one node to all other nodes in the network. A related and often more sensitive measure is *betweenness*, defined as the fraction of all shortest paths in the network that pass-through a given node. Bridging nodes that connect disparate parts of the network often have a high betweenness centrality. The notion of betweenness centrality is naturally extended to links and could therefore also be used to detect important anatomical or functional connections. However, in the case of betweenness, there is no formulation that takes directionality into account.

***Measures of Segregation**: f*unctional segregation in the brain is the ability for specialized processing to occur within densely interconnected groups of brain regions. Measures of segregation primarily quantify the presence of such groups, known as clusters or modules, within the network. These measures have straightforward interpretations in anatomical and functional networks. The presence of clusters in anatomical networks suggests the potential for functional segregation in these networks, while the presence of clusters in functional networks suggests an organization of statistical dependencies indicative of segregated neural processing. If the nearest neighbours of a node are also directly connected to each other, they form a cluster. The clustering coefficient quantifies the number of connections that exist between the nearest neighbours of a node as a proportion of the maximum number of possible connections. The mean clustering coefficient for the network reflects, on average, the prevalence of clustered connectivity around individual nodes. Random networks have low average clustering whereas complex networks are characterized by high clustering.

The mean clustering coefficient is normalized individually for each node and may therefore be disproportionately influenced by nodes with a low degree. A classical variant of the clustering coefficient, known as the transitivity, is normalized collectively and consequently does not suffer from this problem. Both the clustering coefficient and the transitivity have been generalized for weighted and directed networks[146].

***Measures of Integration**: f*unctional integration in the brain is the ability to rapidly process specialized information arising from distributed brain regions. Measures of integration formulate this concept by estimating the ease of communication between brain regions and are commonly based on the concept of a path, defined as a sequence of distinct nodes and edges. This measure has been generalized for weighted and directed networks. In anatomical networks a path represents potential routes of information flow between pairs of brain regions. Lengths of paths consequently reflect the potential for functional integration between brain regions, with shorter paths denoting stronger potential for integration. In the case of functional connectivity data, by its definition, already contain such information for all connections. In functional networks, paths represent sequences of statistical dependencies and may not correspond to information flow through anatomical connections. For this reason, network measures based on functional paths are less straightforward to interpret. In general, path lengths are inversely related to edges weights, as large weights typically represent strong associations, close proximity and shortest path.

The average shortest path length between all pairs of nodes in the network is known as the characteristic path length of the network[147] and is the most commonly used measure of functional integration. Random and complex networks have short average path lengths.

Small-worldness: originally described in social networks, the 'small-world' property combines high levels of local clustering among nodes of a network and short paths that globally link all nodes of the network[11]. This means that all nodes of a large system are linked through relatively few intermediate steps, despite the fact that most nodes maintain only a few direct connections — mostly within a clique of neighbours. Hence, small-world networks are formally defined as networks that are significantly more clustered (have many short-range links) than random networks, yet have approximately the same characteristic path length (few long-range links) as random networks[147]. More generally, small-world networks should be simultaneously highly segregated and integrated. These two characteristics are the result of a natural process to satisfy the balance between minimizing the resource cost and maximizing the flow of information among the network components[113]. Evidence for small-world attributes has been reported in a wide range of studies of genetic, signalling, communications, computational and neural networks. These studies indicate that virtually all networks found in natural and technological systems have small-world architectures and that the ways in which these networks deviate from randomness reflect their specific functionality.

Liao et al. explained in detail why the human brain network is expected to have a small-world architecture[141]. The metabolic and wiring costs in connections among anatomically adjacent brain areas are lower than those among distant brain regions[113].

Theoretical examinations have pointed out that the brain regions are more likely to interact with their neighbouring areas to reduce the whole metabolic costs, while at the same time they need to have a small number of long-distance connections among themselves to accelerate data transmission.

In agreement with theoretical studies, empirical investigations have also proved the dispersion of a few long connections among a plethora of short connections in the human brain network[148].

Moreover, recent studies demonstrated that the small-world property of brain networks experiences topological alterations under different cognitive loads and during development[143,149], as well as in neurological and mental disorders[150]. These alterations may provide novel insights into the biological mechanisms underlying human cognition, as well as health and disease.

Although of particular significance, measures of segregation, integration and small-worldness do not consider the fundamental problem of directionality in the processing pathway and the importance of top-down or bottom-up connectivity in several brain processing. Therefore, since the directionality of information flow is the focus of this thesis, the main measures employed have been directional centrality measures, such as *outdegree*, *indegree*, *hubness* and *authority*

2.5 **Neural Mass Models**

In addition to neuroimaging techniques, another promising method to investigate brain rhythms transmission in the cortex, and its arising functional connectivity, is through mathematical models simulating brain dynamics. These models can be subdivided in two main typologies. In the first case, models involve the description of individual neurons activation (generally spiking neurons) and the explicit inclusion of ionic channels, axons and dendrites [151]. This class of models adopts the approach of Hodgkin and Huxley—namely, careful empirical observations to understand and model the response of the system to its inputs. Indeed, Hodgkin and Huxley carefully measured the conductance of a single axon in response to changes in membrane potential. This approach respects the notion that complex systems can exhibit specific rules at different levels of organization and that large-scale activity may hence be more than the sum of its parts[152]. Although these models are able to grasp the mechanisms that cause oscillations in a network of neurons at a microscopic scale, they are computationally demanding and unsuitable for capturing the behaviour of entire cortical regions at a mesoscopic level.

Since the 1970s much attention has been devoted to Neural Mass Models (NMMs), which describe the average activity of macro-columns, or even cortical areas, using just a few state variables (differential equations). Indeed, the complexity of neural networks generating EEG signals is such that this second approach is usually more viable and suitable to catch the key neural mechanisms. These models summarize the behaviour of millions of interacting neurons, under the assumption that neurons in the same population share similar inputs and have synchronized activity. We decided to focus on NMMs since they represent a good compromise between accuracy and simplicity.

A single NMM describes a local population of interacting neurons, such as pyramidal and inhibitory, that are capable of producing various morphologic EEG-like waveforms and rhythmic activity for a given set of model parameters.

Basically, these models use two conversion operations: a wave-to-pulse operator, which is generally a sigmoid function, and a linear pulse-to-wave conversion implemented at a synaptic level. The first operator converts the average postsynaptic membrane potential from other neural populations into an average firing rate. The breadth of the sigmoid function (lower threshold and upper saturation) implicitly incorporates the non-linear behaviour typical of neural interactions

The second operator depends on synaptic kinetics and models the average postsynaptic response as a linear convolution of incoming spike rate.

One of the first NMMs is the Wilson-Cowan oscillator (1972), still largely employed to investigate synchronization among neural oscillations[153]. Then, Lopes da Silva et al. (1974) implemented a simple model of one excitatory and one inhibitory population in feedback to simulate the α rhythm in the thalamus[154], and Freeman proposed a similar model to generate the neural dynamics in the olfactory cortex (1978)[155]. Subsequently, in the early nineties, an improvement of such models was proposed by Jansen and Rit (1995)[156]. Their model involves a more realistic representation of the single cortical column considering three different neural populations with different synaptic kinetics: pyramidal neurons, excitatory interneurons, inhibitory interneurons. These equations have been frequently employed to build networks of interconnected cortical areas to study EEG dynamics in large regions of the brain and functional connectivity. Friston et al. developed a theoretical framework known as Dynamic Causal Modeling (DCM)[21], combing a variant of the Jansen and Rit model to characterize the dynamics in cortical regions, together with a Bayesian approach for parameter estimation from data. Another variation of the Jansen and Rit model was implemented by Sotero et al. (2007), that investigated how effective connectivity among brain regions may affect the distribution of brain rhythms on the overall brain[157].

Recent studies have emphasised that the kinetics of inhibitory populations have a key influence on signals generation. In this regard, an important advancement concerning the use of NMMs was provided by a study of Wendling et al. (2002) on the hippocampal dynamics during epilepsy[158]. Specifically, the proposed model includes the addition of a fourth population to the model, representing $GABA_A$ interneurons with fast synaptic kinetics. The strength of the Wendling model is that it is able to simulate the dynamics of high-frequency (γ) EEG signals recorded through intracerebral electrodes in the hippocampus during the transition from interictal to fast ictal activity.

However, as highlighted by Ursino et al., 2007 this model has some limitations[159]. Considering a single mass model, and white noise as input, it is able to produces a single narrow band rhythm or, more rarely, a wide band spectrum. Conversely, real EEG spectra measured during motor or cognitive tasks show the concomitant presence of multiple rhythms in the same cortical region and this phenomenon plays a key role in brain dynamics and processes.

Some authors have attempted to overcome this limitation by trying to simulate multimodal spectra assuming that there are several subpopulations in the same brain region (single mass model), each with a different synaptic kinetics. However, this requires the integration of a specific population of neurons into the model for each brain rhythm, becoming computationally demanding. A more parsimonious approach was proposed by Cona et al., 2009 which hypothesized that due to the internal dynamics, each cortical region can produce just one intrinsic rhythm, but it can receive additional rhythms contributions from other regions via long-range excitatory connections[160]. However, even this model showed some limitations: 1) it was still necessary to modify synaptic kinetics between regions, 2) it remained difficult to simulate more than two rhythms in the same cortical region.

The real turning point for these NMMs arrived with the work of Ursino et al. in 2010 where they proposed the addition of a new feedback loop, through which fast inhibitory interneurons can produce a γ rhythm without the need to modify synaptic kinetics[13].

This short summary underlines the rising importance that neural mass models are acquiring for the study of brain dynamics and the incredible advances that have been made over the years.

The NMM described above concerns a local population of interacting neurons, such as pyramidal and inhibitory cells. Despite the dimension achieved, there still exist several orders of magnitude to reach the large-scale systems that support brain function. This can be obtained by coupling an ensemble of NMMs into brain circuits. According to present neurophysiological knowledge, a network of interconnected NMMs can be realized assuming that long-range synapses emerge from pyramidal neurons of the source cortical region. Dynamics within each neuronal population node (that is, each NMM) consequently reflects the local population activity plus the influences from other regions (other nodes), received through long-range synapses from other populations. Furthermore, each population may also receive inputs from the external environment and superimposed Gaussian white noise.

This thesis focuses on NMMs since they allow a mechanistic description of brain rhythms whilst maintaining a limited number of state variables.

3 Simulated Brain Networks through NMMs

The variety of functional connectivity estimators found in the literature is grounded on different underlying mathematical formulations and address diverse aspects of connectivity features. Since there is no ground truth for EEG signals, the problem of choosing the most suitable connectivity method for a particular type of data, as well as the evaluation of its reliability, is a difficult one.

Neural Mass Models represent a unique tool to evaluate the reliability of different connectivity results in ground truth conditions. Three studies are presented in this section, whose aim is to shed light on the strengths and limitations of the main functional connectivity estimators using data simulated by NMMs. Indeed, these models have been employed to: a) test the performance of FC estimators in linear and nonlinear conditions; b) compare the reliability of different FC estimators; c) study how brain rhythms are transmitted in the brain; d) simulate the task-dependent connectivity network of a stroke patient.

Since the same NMM was used for the three studies, its description and mathematical formulation is detailed below. First, equations of a single region of interest (ROI) are described. Then, a model of several interconnected ROIs is built from these equations.

Model of a single Region of Interest

The model of a single Region of Interest (ROI) consists of the feedback arrangement among four neural populations: pyramidal neurons (subscript *p*), excitatory interneurons (subscript *e*), inhibitory interneurons with slow and fast synaptic kinetics ($\text{GABA}_{\text{A,slow}}$ and $\text{GABA}_{\text{A,fast}}$, subscripts *s* and *f*, respectively). Each population receives an average postsynaptic membrane potential (say *v*) from other neural populations, and converts this membrane potential into an average density of spikes fired by the neurons (say *z*). This conversion is simulated with a static sigmoidal relationship, which reproduces the non-linearity in neuron behavior (the presence of a zone where neurons are silent (below threshold) and an upper saturation, where neurons fire at their maximal activity).

To model dynamics in a whole ROI, the four populations are connected via excitatory and inhibitory synapses, according to the schema in Fig. 3.1. Each synaptic kinetics is described with a second order system, but with different parameter values. We assumed three types of synapses: glutamatergic *excitatory* synapses with impulse response $h_e(t)$, assuming that synapses from pyramidal neurons and from excitatory interneurons have similar dynamics; GABAergic inhibitory synapses with *slow* dynamics (impulse response $h_s(t)$); GABAergic inhibitory synapses with *faster* dynamics (impulse response $h_f(t)$). They are characterized by a gain (G_e, G_s, and $G_{f,}$ respectively) and a time constant (the reciprocal of these time constants denoted as ω_e, ω_s, and $\omega_{f,}$, respectively). The average numbers of synaptic contacts among neural populations are represented by eight parameters, C_{ij}, where the first subscript represents the target (post-synaptic) population and the second refers to the pre-synaptic population.

In a previous work [161] a sensitivity analysis was performed on the role of connections linking different ROIs, and was found that the most influential connections are "from pyramidal to pyramidal" and "from pyramidal to fast inhibitory". Accordingly, in this thesis we assume that inputs to each ROI (say *u*) target only pyramidal and fast-inhibitory populations (see Fig. 3.1). The equations of a single ROI are written below:

Pyramidal neurons

$$\frac{dy_p(t)}{dt} = x_p(t) \tag{3.1}$$

$$\frac{dx_p(t)}{dt} = G_e \omega_e z_p(t) \quad - 2\omega_e x_p(t) - \omega_e^2 y_p(t) \tag{3.2}$$

$$z_p(t) = \frac{2e_0}{1 + e^{-rv_p}} - e_0 \tag{3.3}$$

$$v_p(t) = C_{pe} y_e(t) - C_{ps} y_s(t) \quad - C_{pf} y_f(t) \tag{3.4}$$

Excitatory interneurons

$$\frac{dy_e(t)}{dt} = x_e(t) \tag{3.5}$$

$$\frac{dx_e(t)}{dt} = G_e \omega_e \left(z_e(t) + \frac{u_p(t)}{C_{pe}} \right) \quad - 2\omega_e x_e(t) - \omega_e^2 y_e(t) \tag{3.6}$$

$$z_e(t) = \frac{2e_0}{1+e^{-rv_e}} - e_0 \tag{3.7}$$

$$v_e(t) = C_{ep} y_p(t) \tag{3.8}$$

Slow inhibitory interneurons

$$\frac{dy_s(t)}{dt} = x_s(t) \tag{3.9}$$

$$\frac{dx_s(t)}{dt} = G_s \omega_s z_s(t) \quad - 2\omega_s x_s(t) - \omega_s^2 y_s(t) \tag{3.10}$$

$$z_s(t) = \frac{2e_0}{1+e^{-rv_s}} - e_0 \quad (3.11)$$

$$v_s(t) = C_{sp}y_p(t) \quad (3.12)$$

Fast inhibitory interneurons

$$\frac{dy_f(t)}{dt} = x_f(t) \quad (3.13)$$

$$\frac{dx_f(t)}{dt} = G_f\omega_f z_f(t) \quad - 2\omega_f x_f(t) - \omega_f^2 y_f(t) \quad (3.14)$$

$$\frac{dy_l(t)}{dt} = x_l(t) \quad (3.15)$$

$$\frac{dx_l(t)}{dt} = G_e\omega_e u_f(t) \quad - 2\omega_e x_l(t) - \omega_e^2 y_l(t) \quad (3.16)$$

$$z_f(t) = \frac{2e_0}{1+e^{-rv_f}} - e_0 \quad (3.17)$$

$$v_f(t) = C_{fp}y_p(t) - C_{fs}y_s(t) \quad - C_{ff}y_f(t) + y_l(t) \quad (3.18)$$

The inputs to the model, *$u_p(t)$* and *$u_f(t)$* (Eqs. 3.6 and 3.16) represent all exogenous contributions coming from external sources (either from the environment or from other brain regions) filtered through the low-pass dynamics of the excitatory synapses (Eqs. 3.5 - 3.6 and Eqs. 3.15 - 3.16, respectively). In fact, a common assumption in neurophysiology is that long-range connections in the brain are always mediated via excitatory glutamatergic synapses. In particular, *$u_p(t)$* is the input to pyramidal cells and *$u_f(t)$* the input to GABA$_{A,fast}$ interneurons. These terms will be described below.

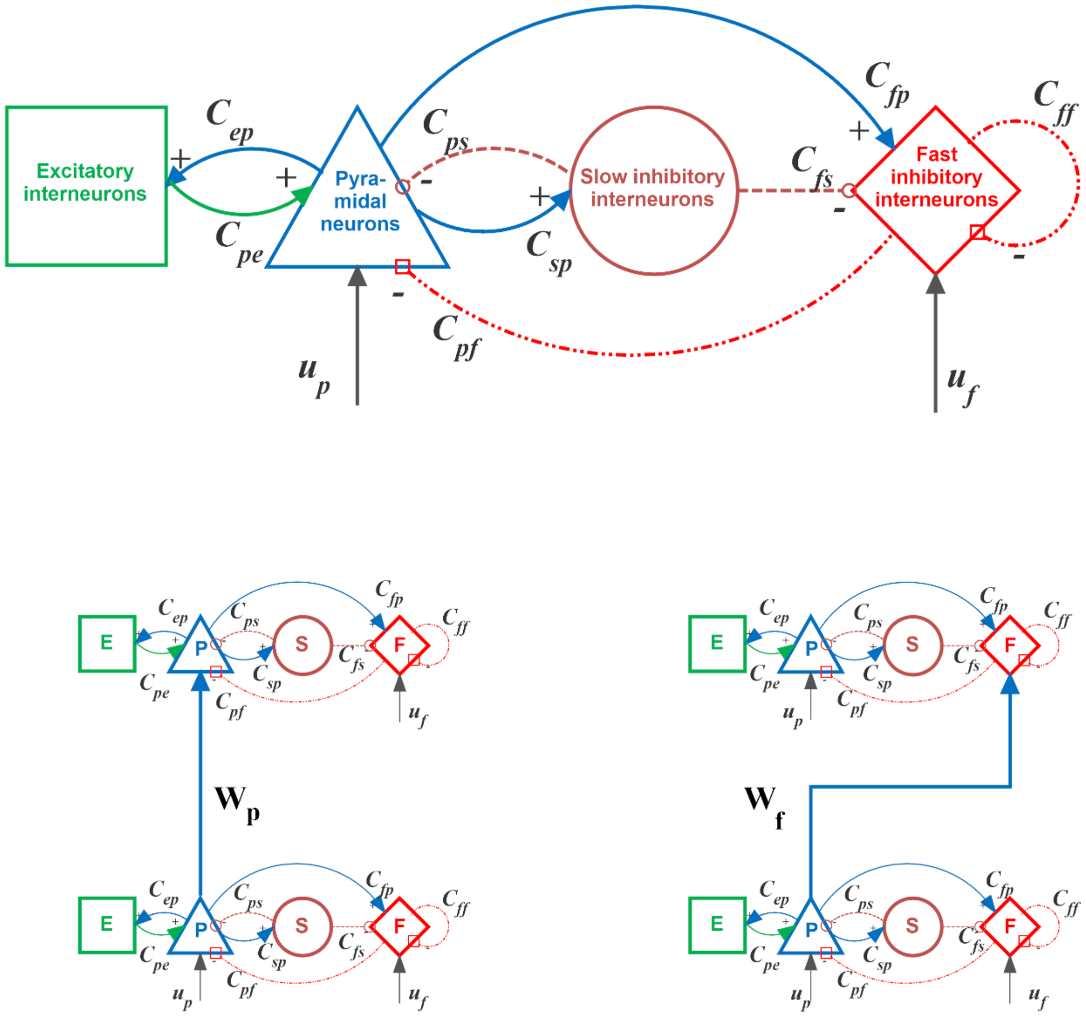


Figure 3.1 – Block diagram of the neural mass model (upper panel) used to simulate activity in a single region of interest (ROI). Continuous lines denote excitatory synapses (from pyramidal neurons, blue lines, or from excitatory interneurons, green lines), magenta dotted lines denote slow inhibitory synapses, and red dash-dotted lines denote fast inhibitory synapses. The bottom panels show two exempla of connections among ROIs: *excitatory* (pyramidal-pyramidal) in the left and bi-synaptic *inhibitory* (pyramidal-fast inhibitory-pyramidal) in the right.

Model of several interconnected ROIs and connectivity parameters

In order to study connectivity between regions, let us consider two ROIs (each described via Eqs. 3.1- 3.18), which are interconnected through long-range excitatory connections. The presynaptic and postsynaptic regions will be denoted with the superscript k and h, respectively. The generalization to more than two regions is trivial. Throughout the manuscript, we will use the first superscript to denote the target ROI (post-synaptic) and the second superscript to denote the donor ROI (pre-synaptic).

To simulate connectivity, we assumed that the average spike density of pyramidal neurons of the presynaptic area (z_p^k) affects the target region via a weight factor, W_j^{hk} (where *j* = *p* or *f*, depending on whether the synapse targets to pyramidal neurons or fast inhibitory interneurons) and a time delay, *T*. This is achieved by modifying the input quantities u_p^h and/or $u_f^h$of the target region.

Hence, we can write

$$u_j^h(t) = n_j^h(t) + W_j^{hk} z_p^k(t-T) \qquad j = p, f \tag{3.19}$$

where $n_j(t)$ represents a Gaussian white noise (in the present work, if not explicitly modified, we used: mean value $m_j = 0$ and variance $\sigma_j^2 = 9/dt$, where *dt* is the integration step) which accounts for all other external inputs not included in the model.

It is worth noting that the synapses W_p^{hk} have an excitatory role on the target region *h*, since they directly excite pyramidal neurons. Conversely, synapses W_f^{hk} , although glutamatergic in type, have an inhibitory role, via a bi-synaptic connection. In particular, both connections go from the source ROI *k* to the target ROI *h*, but in the inhibitory case this is composed of two synapses (from pyramidal neurons in the source ROI *k* to inhibitory interneurons in the target ROI *h* and then from inhibitory interneurons in target ROI *h* to pyramidal neurons still in ROI *h*). Hence, the second synapse is internal to ROI *h* and has not been modified throughout this work. Hence, in the following the general terms “excitatory connection” and “inhibitory bi-synaptic connection” will be used to describe these two different connections, although both glutamatergic in type. In particular, we wish to stress that the Dale principle is always satisfied in our model, since individual neural populations within each ROI are either excitatory or inhibitory, and this distinction is established a priori in the model.

3.1 **Evaluation of Transfer Entropy Performance with NMMs**

The study reported in this chapter refers to the published journal paper entitled "Transfer Entropy as a measure of Brain Connectivity: A Critical Analysis with the Help of Neural Mass Models", Mauro Ursino*, Giulia Ricci, Elisa Magosso, *Frontiers in computational neuroscience*, (2020).

In this section, NMMs have been used to study the performance of Transfer Entropy in estimating the true connective strength imposed by the model. The reliability of the estimator was tested under the following conditions: 1) increasing network size, 2) linear and non-linear conditions, 3) increasing pure delay between ROIs, 4) increasing signal length.

The main result that emerged from this study concerns the significance of the changes in connectivity detected by the estimator, which do not always reflect a true change in the connection strength, but rather a change in the transmission of information due to nonlinear effects.

Objective *- Assessing brain connectivity from electrophysiological signals is of great relevance in neuroscience, but results are still debated and depend crucially on how connectivity is defined and on mathematical instruments utilized. Aim of this work is to assess the capacity of bivariate Transfer Entropy (TE) to evaluate connectivity, using data generated from simple neural mass models of connected Regions of Interest (ROIs).* ***Approach*** *- Signals simulating mean field potentials were generated assuming two, three or four ROIs, connected via excitatory or by-synaptic inhibitory links. We investigated whether the presence of a statistically significant connection can be detected and if connection strength can be quantified.* ***Main Results*** *- Results suggest that TE can reliably estimate the strength of connectivity if neural populations work in their linear regions, and if the epoch lengths are longer than 10 s. In case of multivariate networks, some spurious connections can emerge (i.e., a statistically significant TE even in the absence of a true connection); however, quite a good correlation between TE and synaptic strength is still preserved. Moreover, TE appears more robust for distal regions (longer delays) compared with proximal regions (smaller delays): an approximate a priori knowledge on this delay can improve the procedure. Finally, nonlinear phenomena affect the assessment of connectivity, since they may significantly reduce TE estimation: information transmission between two ROIs may be weak, due to non-linear phenomena, even if a strong causal connection is present.* ***Significance*** *- Changes in functional connectivity during different tasks or brain conditions, might not always reflect a true change in the connecting network, but rather a change in information transmission. A limitation of the work is the use of bivariate TE. In perspective, the use of multivariate TE can improve estimation and reduce some of the problems encountered in the present study.*

3.1.1 Introduction

Cognitive phenomena originate from the interaction among several mutually interconnected, specialized brain regions, which exchange information via long range synapses. Consequently, the problem of assessing brain connectivity during different cognitive tasks is playing a crucial role in neuroscience nowadays, not only to understand mechanisms at the basis of normal cognitive functions, but also to identify alterations in pathological states. Connectivity is often estimated from fMRI neuroimaging techniques[162–164]. However, thanks to their higher temporal dynamics, electrophysiological data, obtained from electro- or magneto-encephalography, joined with methods for cortical source localization [165–168] are receiving increasing attention.

The problem of assessing connectivity from data, however, is a difficult one, since the concept of connectivity has ambiguous definitions (see [163]) and results depend crucially on how connectivity is defined and on the mathematical instruments utilized.

Although there are different ways to define connectivity, in the following we will refer to *functional connectivity* (FC) defined as "the statistical dependence or mutual information between two neuronal systems" [169]. A distinct definition of connectivity, stronger than FC (see [170] and [169])*,* is *effective connectivity*: it refers to the influence that one neural system exerts on another and is based on an explicit model of causal inference, usually expressed in terms of differential equations. The most popular method to evaluate effective connectivity is dynamical causal modelling (DCM). DCM assumes that the signals are produced by a state space model (see Tab. 1 in [170] for a list of possible equations used in recent papers). However, this framework requires strong a priori knowledge about the input to the system and the connectivity network. To overcome this limitation, the more suitable network is often chosen among various possible alternatives using Bayesian selection methods [171].

However, despite these dichotomous definitions, the fundamental interest in all FC research is still "understanding the casual relationship among neural entities", as stressed by Reid [172] recently. Although the kind of causal inference that can be inferred with FC methods is limited and only indirect, several FC measures can provide some useful information in regard to causality (recent assessment papers are [172–174]). Indeed, among the different ways to calculate FC, some of them are based on the concept of causality (although without using state-space models), as originally introduced by Wiener [175] and subsequently by Granger [176]. According to their definition, we can say that a temporal series X has a causal influence on a second temporal series Y if the prediction on the future of Y is improved by knowledge on the past of X. Interestingly, the technical links between Granger causality and DCM have also been recently incorporated in the state space framework [177] with reference to functional magnetic data in resting states. Results of these authors indicate a qualitative consistency between Granger causality and DCM, and show that both can be used to estimate directed functional and effective connectivity from fMRI measurements in a reliable way.

One of the most promising methods to infer FC from data is Transfer entropy (TE). TE implements the causal principle expressed above within the framework of information theory, by using conditional probabilities [133,135] for more details): if a signal X has a causal influence on a signal Y, then the probability of Y conditioned on its past is different from the probability of Y conditioned on both its past and the past of X. The same idea can be expressed observing that entropy on the present measurement of Y is reduced if knowledge of the past of X is added to knowledge of the past of Y. A great advantage of TE compared with the other methods is that it does not require any prior assumption on data generation (i.e., it is model-free).

For this reason, TE is largely used in neuroscience today to assess connectivity from EEG/MEG data sets in conditions lacking any prior assumption. Also some variants of TE (such as Phase Transfer Entropy [178]) have been proposed recently.

Nevertheless, the use of TE to assess connectivity may also exhibit some drawbacks, besides definite advantages. First, as recognized in several papers [135,179,180], estimation of TE from data can be affected by various elements of the estimation procedure; among the others: the embedding dimension and the delay in the reconstruction of the state space, the quantity of data samples available, the method adopted to estimate high-dimensional conditional probabilities.

Second, it is unclear how much TE is affected by spurious information, such as that arising from shared inputs or from a cascade among several populations, or due to a redundancy in the population processing [181]. To reduce the previous aspects, multivariate TE methods have also been proposed recently [182].

Third, and maybe more important, TE is not a direct measure of coupling strength, and should be used with extreme caution to measure a coupling parameter (such as the weight of synapses among two neural populations). Actually, TE measures how much information is transferred from X to Y: this concept is of the greatest value to understand how the brain performs its computation by exchanging information between different regions (see also [183]) but in some conditions may be intrinsically different from causal strength.

Once established that TE is a valid tool to investigate the computational aspects of the brain, i.e. the transfer of information between different areas, in the present study we wish to critically analyse how good it may be at estimating a biophysical coupling property too, i.e. the connection strength between Regions of Interest (ROIs). To this end, a powerful way is to challenge TE with the use of simulated data. These should mimic real neuroelectric signals (especially for what concerns their frequency content), and should be generated via biologically inspired models with assigned coupling terms among neural units.

Indeed, many such studies have been published in the last decade, to compare FC estimated values with a “true” connectivity topology incorporated in simulation models, providing quite a large set of validation information. In the following, we will first encompass a synthetic analysis of the recent literature, to point out the present major gaps and elements which, in our opinion, deserve further analysis (especially, with reference to TE). Then, the aim of this work is better delineated, as it emerges from the absences in the present literature.

3.1.2 Literature Critical Review

3.1.2.1 Summary of previous studies

Several studies have been performed in recent years, to compare the results obtained with the FC estimation methods, with the “true” connectivity values incorporated in simulation models (used as a sort of “ground truth”). However, most of these studies were aimed at exploring whether FC can discover the presence of connectivity links (ON/OFF), using receiver operating characteristic (ROC) curves. Just in a few of them, the relationship between the FC index and the connectivity strength was explored, although generally in rather a qualitative way.

Two main classes of studies will be considered in the following, depending on the simulation model adopted: those which use spiking neurons, mainly aimed at analysing connectivity in neural cultures, and those using neurons with continuous outputs, more oriented to the analysis of connectivity among larger regions of interest. As to the studies with spiking neurons, in the following we will limit our analysis just to those which use TE. A wider approach is used for the selection of studies simulating larger cortical ROIs.

Studies with spiking neurons - Ito et al. [184] applied the TE with multiple time delays to a network model containing 1000 Izhikevich’s neurons. They observed that their measures generally increase with synaptic weights, but there is substantial variability in the obtained results. Moreover, in their work the synaptic weights were bimodally distributed around just two values (0 and a positive one).

Garofalo et al. [185] compared the estimation obtained with various methods (Transfer Entropy, Joint Entropy, Cross Correlation and Mutual Information) first in a neuronal network model made up of 60 synaptically connected Izhikevic neurons, and then in cultures of neurons. In the model they also included inhibitory connections. The comparison was performed with ROC curves. Their results suggest that TE is the best method, both with the excitatory and excitatory+inhibitory models, but it recognizes also some strong indirect connections not classified as true positive. Moreover, it exhibits problems in identifying inhibitory connections.

Orlandi et al. [186] also used realistic computational models that mimicked the characteristic bursting dynamics of neural cultures, and extended previous works by attempting the inference of both excitatory and inhibitory connectivity via TE. The quality of the reconstruction was quantified through a ROC analysis. They showed that the most difficult aspect is not the identification of a link, but rather its correct labelling (excitatory or inhibitory). Hence, they suggested a two-step analysis (for instance before and after the use of pharmacological blocking inhibitory connections).

Timme and Lapish [187] analysed the strength of information theory methods using both small networks of neurons and larger 1000 neuron models of Izhikevich type (800 excitatory, 200 inhibitory). They concluded that TE can be used to measure information flow between neurons. More important, they suggested the use of partial information decomposition to move beyond pair of variables to group of variables, and found that this method can be used to break down encoding by two variables into redundant, unique and synergistic parts. The last aspect will be commented in the Discussion session of the present paper.

All previous studies suggest that TE can be a powerful instrument to infer the existence of connections among neurons. However, the application of these studies is limited to cultures of about a thousand of units. Of course, the neuroelectric dynamics of an entire ROI, resulting from millions of neurons, is largely different. To study this aspect, higher levels models, with just a few states variables per ROI, are generally used, particularly neural mass models (NMMs).

Studies using Neural Mass Models - A pioneering study which evaluated functional connectivity using neural mass models was performed by David et al. [188]. The authors used cross-correlation, mutual information, and synchronization indices (hence, they did not evaluate TE). For simplicity, they used a symmetric configuration and did not consider the problem of inhibitory connections among ROIs. The results suggest that each measure is sensitive to changes in neuronal coupling, with a monotonic dependence between the functional connectivity measures and the coupling parameter, and that the statistical power of each measure is an increasing monotonic function of the signal length.

Studies quite similar to the present one, although with a simpler aim, and without TE as a target, have been performed by others [128,189] (indeed, they used either regression methods or synchronization indices). Various models were employed to generate signals: among the others, two NMMs (but with only two populations each) connected with an excitatory coupling parameter. The authors explored the relationship between the coupling parameter and the estimated FC and observed that the regression methods exhibit good sensitivity to the coupling parameter. However, in that study the characterization of the *direction* of coupling was not dealt with, inhibitory connections were not incorporated, and the authors did not test TE accuracy.

A systematic study on the performance of various methods for FC estimation was performed by Wang et al. [16]. They compared the performance of 42 methods (including, among the others, the pairwise directed TE and the partial TE), using five different models to generate signals (including a three population NMM). Moreover, they used a connectivity structure with 5 nodes. Although this is the most complete study presently available, it limits the analysis to the performance of the connectivity estimate on an ON/OFF basis, using ROC curves (i.e., they did not evaluate whether the estimated FC values are sensitive to the strength of the coupling parameters). Their results suggest that, for the NMM simulations, Granger causality and TE are able to recover the underlying model structure, with TE much less time consuming. However, TE failed when simulations were performed with highly nonlinear (Rossler or Hénon) equations.

The previous summary highlights several important points. First, various methods do exist to infer FC from signals, each with its own virtues and limitations (but see also [172]). However, in many studies TE emerges as one of the most effective methods, which joins benefits of good sensitivity and efficient computation time. However, despite the excellent works performed until now, several problems are still insufficiently clarified, which justify further studies. These questions are stressed below.

First, no study analysed carefully the relationship between the TE metrics and the connectivity strength using NMMs to simulate neuroelectrical activity of entire ROIs. Actually, neither David et al. [188] nor Wendling et al. [128] used TE in their analysis, whereas Wang et al. [174] (who tested TE) did not evaluate the sensitivity vs. the connection strength.

Second, despite some authors analysed the presence of excitation + inhibition in models of spiking neurons [185,186] , we are not aware of any study in which inhibition between ROIs is properly taken into account in the analysis of FC. Indeed, although long-range connections between ROIs are mediated by synapses from pyramidal neurons (hence, they are all excitatory in type) one region can inhibit another region by targeting into the population of inhibitory interneurons. In particular, it is known that lateral connections in the cortex target all population types, in different layers of a cortical column [190,191]. Although the role of various connections types in the propagation of brain rhythms has been carefully studied with NMMs [161,191–195] we are not aware of any NMM study which investigates the role of long-range inhibition on FC estimation.

Finally, and more important, several studies underline the difficulty of FC methods (and in particular, TE) to deal with strongly non-linear problems. For instance, Wang et al.[174] observed that TE fails to find a proper connectivity topology when signals are generated with Rössler equations or Hénon systems, i.e., with strongly non-linear models. By comparing TE computed at various time lags to values computed with surrogate linearized data, Nichols et al. [196] observed that TE is quite sensitive to the presence of non-linearity in a system. Indeed, although NMMs have been frequently used in the domain of FC assessment, to our knowledge all previous papers used these models in “quite linear conditions”, i.e. without inducing strong alterations in the working point and/or moving dynamically from linear vs. saturation activity regions. We speculate that the same model (with assigned connectivity strength) can produce largely different values of FC estimation depending on the working conditions, on noise variance and on the amplitude of the input changes.

3.1.2.2 Objectives and work organization

Taking in mind the previous limitations of former works, the present study was conceived with the following major aims: i) to analyse the relationship between the TE metrics and the strength of the connectivity parameters using NMMs, in order to assess whether changes in TE from one trial to another can be used to infer an underlying *change* in connectivity between ROIs; ii) to study the role of synapses targeting to excitatory vs. inhibitory populations in affecting FC; iii) to reveal how non-linearities can dramatically affect the inference of connection strength, leading to different conclusions on connectivity among regions depending on the particular working condition. This point is of value to highlight that TE is actually a powerful metric to assess information transfer and computation in the brain, but in some cases may be different from coupling strength. We think that neither of these points has been thoroughly assessed in previous papers.

To reach these objectives, we evaluated FC with bivariate TE using data generated from simple NMMs of connected populations. In particular, the values of TE between two ROIs estimated from simulated data were compared with the strength of the coupling terms used in the model, at different values of this strength. We investigated different network topologies (with two, three or four ROIs) and the role of time delay, signal length, and changes in external input (mean value and noise variance). The latter aspect is of pivotal value to assess the role of non-linearities.

TE was estimated using Trentool, a software package implemented as a Matlab toolbox under an open source licence [136]. Simulated data were generated using the model of neural masses described in Ursino et al. and Cona et al.[161,194] which represents a good compromise between biological reliability and simplicity, and is able to simulate realistic spectra of neuroelectric activity in the cortex (including alpha, beta and gamma bands). In particular, in this work the internal parameters of this model were assigned to simulate spectra with a strong component in the beta band and some component in the gamma band, as often measured in motor, premotor and supplementary motor cortices [161,194].

The paper is structured as follows. First, the main theoretical aspects of transfer entropy are described. Subsequently, equations of the neural model are given, with parameter numerical values.

In section results, TE estimates obtained with Trentool on simulated data were used not only to test the performance of this metric in detecting the presence or absence of a connection (ON/OFF evaluation by means of statistical tests against surrogate signals), but also to compare the TE values of the detected connections with the strength of the coupling terms in the model. Results are then critically discussed to emphasize in which conditions TE can provide reliable indications on connectivity, and in which conditions information transfer is different from connection strength. Limitations of this work (such as the use of a bivariate estimator) are also debated and lines for further work delineated.

3.1.3 Transfer Entropy: Theoretical and Practical Aspects

In the following, we first summarize the main theoretical aspects of transfer entropy, as a model-free method to estimate connectivity. Then, some practical issues of the estimation procedure adopted by Trentool are discussed.

3.1.3.1 General theory

Throughout this section, we will use a lower case letter to denote a single (scalar) variable, and an upper case letter to denote a vector. Moreover, we will use the boldface to represent a random variable (or a random vector) and no-bold to represent the realization of these variables during the experiment.

Let us consider a discrete random variable **x**, with realization $x \in S_x$ and probability distribution *p*(*x*) over its outcomes. The amount of *information* gained by observation of the event *x* is

$$h(x) = log_2 \frac{1}{p(x)} = - \, log_2 \, p\,(x) \tag{3.20}$$

For instance, if a discrete event has probability *p*(*x*) = 1/8 = 2^{-3}, its realization provides three bits of information.

Shannon entropy of the random variable **x** is computed as the average value of the information over all possible realizations of *x*, i.e.

$$S(\boldsymbol{x}) = \sum_{x \in S_x} p(x) \, log_2 \frac{1}{p(x)} = - \sum_{x \in S_x} p(x) \, log_2 \, p\,(x) \tag{3.21}$$

The same definition of Shannon entropy, of course, can be applied in case of conditional probability. Let us assume that we observe the outcome of a discrete random variable **y** (with probability distribution *p*(*y*), and $y \in S_y$) after we have already observed a realization *x* of the other random variable **x**. The amount of information gained by the observation *y* becomes

$$h(y/x) = - \, log_2 \, p\,(y/x) \tag{3.22}$$

and, by computing the average value over all possible realization of *x* and *y*, we have

$$\begin{aligned} S(\boldsymbol{y}/\boldsymbol{x}) &= - \sum_{x \in S_x} p(x) \sum_{y \in S_y} p(y/x) \, log_2 \, p\,(y/x) \\ &= - \sum_{\substack{x \in S_x \\ y \in S_y}} p(x,y) \, log_2 \, p\,(y/x) \end{aligned} \tag{3.23}$$

Mutual information of **x** and **y** is evaluated by computing the difference between the entropy of **y**, and the conditional entropy of **y/x.** Of course, the entropy of **y** must be greater (or at least equal) than the entropy of **y/x,** since observation of a realization *x* can reduce the amount of information provided by the observation *y*. The difference between the two entropies is considered as a sort of information that **x** and **y** share. Accordingly, we can define mutual information as follows

$$I(\boldsymbol{y}, \boldsymbol{x}) = S(\boldsymbol{y}) - S(\boldsymbol{y}/\boldsymbol{x}) \tag{3.24}$$

Using the Bayes theorem, one can demonstrate that

$$I(\boldsymbol{y},\boldsymbol{x}) = I(\boldsymbol{x},\boldsymbol{y}) = S(\boldsymbol{x}) - S(\boldsymbol{x}/\boldsymbol{y}) \tag{3.25}$$

i.e., mutual information does not contain any directional evidence.

The same concept of mutual information can be restated assuming that both **x** and **y** are conditioned by the value of a third random variable, **z**. We obtain the conditioned mutual information

$$I(\boldsymbol{y},\boldsymbol{x}/\boldsymbol{z}) = S(\boldsymbol{y}/\boldsymbol{z}) - S(\boldsymbol{y}/\boldsymbol{x},\boldsymbol{z}) \tag{3.26}$$

Let us now apply the same concepts to two time series generated by two stochastic processes. From each process we can define a time-dependent random state vector ($\boldsymbol{X}^m(t)$ and $\boldsymbol{Y}^n(t)$, respectively), whose particular observation can be written as follows (see Takens 1980))

$$\begin{gathered} X^m(t) = [x(t) \;\; x(t-\Delta t) \;\; x(t-2\Delta t) \dots x(t-(m-1)\Delta t)\,] \\ Y^n(t) = [y(t) \;\; y(t-\Delta t) \;\; y(t-2\Delta t) \dots y(t-(n-1)\Delta t)\,] \end{gathered} \tag{3.27}$$

where *m* and *n* are the embedding dimensions, describing how many past samples are used (these are the dimensions of the so-called delay embedding space) and *Dt* is the embedding delay. According to the previous equations, $X^m(t)$ and $Y^n(t)$ contain the present and *m-1* (or *n-1*) past samples of the random process.

Let us now consider the random variable **y**(*t*) representing a present sample of the stochastic process, conditioned by its *n* past samples; the conditional probability is p $(y(t)/Y^n(t-\varDelta t))$ and Shannon entropy is S $(y(t)/Y^n(t-\varDelta t))$. The idea is that, in case of causality from X to Y, the probability of **y**(*t*) conditioned by both $\boldsymbol{X}^m(t-\varDelta t)$ and $\boldsymbol{Y}^n(t-\varDelta t)$ should be different from the probability of **y**(*t*) conditioned by its past only. This effect can be quantified as a difference in Shannon entropy, i.e., by evaluating the additional information that the past of X provides on the present of Y. This leads to the following definition of Transfer Entropy

$$\begin{aligned} TE(X \to Y) &= I(\boldsymbol{y}(t), \boldsymbol{X}^m(t-\varDelta t)/\boldsymbol{Y}^n(t-\varDelta t)) \\ &= S(\boldsymbol{y}(t)/\boldsymbol{Y}^n(t-\varDelta t)) \\ &\quad - S(\boldsymbol{y}(t)/\boldsymbol{X}^m(t-\varDelta t), \boldsymbol{Y}^n(t-\varDelta t)) \end{aligned} \tag{3.28}$$

TE is asymmetric and naturally incorporates direction of information transfer from X to Y.

The previous equation considers the influence that the past of Y and X can have on the present sample of Y. However, as rigorously demonstrated by Wibral et al. (2013) this equation cannot be used to express any causal relationship. In particular, in neural problems, the influence of a signal on another is often characterized by a pure delay (say *d*) which represents the time necessary for action potentials to travel along axons from the pre-synaptic region to the post-synaptic one. Assuming that the time delay can be approximated by *l* sampling periods (i.e. *d* = *l*·D*t*), we can use the delayed signal $\boldsymbol{X}^m(t-d) = \boldsymbol{X}^m(t-l\cdot\varDelta t)$ in the definition of TE instead of $\boldsymbol{X}^m(t)$. In most cases, *l* is not known, and represents a parameter that should be estimated from data (see below).

Thus we can write

$$TE(X \to Y, l) = I\left(\boldsymbol{y}(t), \frac{\boldsymbol{X}^m(t - l \cdot \Delta t)}{\boldsymbol{Y}^n(t - \Delta t)}\right) \\ = S(\boldsymbol{y}(t)/\boldsymbol{Y}^n(t - \Delta t)) - S(\boldsymbol{y}(t)/\boldsymbol{X}^m(t - l \cdot \Delta t), \boldsymbol{Y}^n(t - \Delta t)) \tag{3.29}$$

Wibral et al. (2013) [134] rigorously demonstrated that the predictive information transfer from X to Y over a time delay *d* is properly captured by this equation (aligning with Wiener's principle).

The previous equation can be rewritten as the Kullback-Leibler divergence between the two probability distributions [133]

$$TE(X \to Y, l) = \sum_{\substack{y(t+\Delta t) \\ Y^n(t) \\ X^m(t-l\cdot\Delta t)}} p(y(t), Y^n(t - \Delta t), X^m(t - l \cdot \Delta t)) * \\ * \log_2 \frac{p(y(t)/Y^n(t - \Delta t), X^m(t - l \cdot \Delta t))}{p(y(t)/Y^n(t - \Delta t))} \tag{3.30}$$

or also as a representation of four Shannon entropies:

$$TE(X \to Y, l) = S(\boldsymbol{X}^m(t - l \cdot \Delta t), \boldsymbol{Y}^n(t - \Delta t)) \\ - S(\boldsymbol{y}(t), \boldsymbol{X}^m(t - l \cdot \Delta t), \boldsymbol{Y}^n(t - \Delta t)) \\ + S(\boldsymbol{y}(t), \boldsymbol{Y}^n(t - \Delta t)) - S(\boldsymbol{Y}^n(t - \Delta t)) \tag{3.31}$$

where S(**X**,**Y**) is used to denote the Shannon entropy of the joined probability of **X** and **Y**.

3.1.3.2 Practical aspects on TE estimation

As it is clear from the last equation, the estimation of TE from finite data samples requires the evaluation of various joint and marginal probability distributions. This may be a difficult task, since the probability densities implicated in this equation can have a very large dimensionality (up to *n* + *m* + 1). Moreover, several parameters are not known a priori and must be estimated; in particular, the estimate of TE can be seriously affected by the choice of the embedding dimensions (*n* and m), of the sampling period (D*t*) and of the delay (*d* = *l*·D*t*). Furthermore, TE estimation can have a residual bias. To eliminate this bias, it is important to compare the TE estimated from empirical data, with that obtained from surrogate data sets. Surrogate data sets should incorporate no information transfer from X to Y, but maintain the same statistical properties as the original data. Comparison between the TE obtained from the original data and those obtained from surrogate data also allows computation of a p value to test the statistical significance of the obtained TE value.

The estimates of TE from the outputs of our neural mass model (see section 3), and their statistical significance were performed using the software package Trentool [135,136]. The same package also provides an estimation of the time delay, *l*, and of the embedding dimensions, *m* and *n*, as the values which maximize TE. In this tool, joint and marginal probability distributions are computed using a k-th nearest neighbour estimator. Furthermore, the method contains two additional parameters: the mass for the nearest-neighbour search and a correction to exclude autocorrelation effects from the density estimation. Specifically, the estimate of the bivariate TE for each model configuration was performed as follows. The same model configuration (corresponding to a specific pattern of connectivity among a few ROIs) was run 10 times creating 10 trials of signals; each trial contained the temporal patterns of the local field potentials (over a given time interval, see below) in the involved ROIs, and affected by a random noise. The ten trials were given as input to Trentool that computed the TE values of the fed signals and of the surrogate data and provided the *p* value (permutation test) to assess whether the TE of the simulated signals was significantly different from that of surrogate data. Furthermore, in all simulations with more than two ROIs, the results were subjected to a partial correction of spurious information flow that may be introduced by the bivariate analysis of a highly multivariate system. Namely, this correction works on cascade effects and simple common drive effects. To this end, we used the Trentool Graph Correction function described in the manual (see http://www.trentool.de/ for more details).

In section Results, we will always report the *difference* between TE estimated on simulated signals, and that obtained from surrogate data. Whenever no statistical significance was achieved ($p > 0.05$) the difference was set at zero (i.e. no connection detected). Otherwise (connection detected), the true difference is used as an estimate of connectivity strength.

All details on the version of Trentool used and a table with all parameters adopted in the Trentool functions can be found in the Supplementary Material Part 1.

3.1.4 Model Description

Equations of a single region of interest (ROI) as well as the model of several interconnected ROIs is described at the beginning of Chapter 3 and represented in Fig. 3.1.

In line with the notation used for the inter-region synapses, in the following we will denote with TE^{hk} the transfer entropy from ROI *k* to ROI *h*, that is $TE^{hk} = TE(ROI\ k\ \rightarrow ROI\ h)$.

3.1.4.1 Assignment of model parameters

Parameters within each ROI were given to simulate a power spectral density with a significant activity in the beta range (about 20 Hz) and some activity in the gamma range (above 30 Hz), as shown in Fig. 3.2. This power density is typical of supplementary and premotor cortical areas (see also [161,192,194]. Power spectral density was computed by applying the Welch method on the post-synaptic membrane potential of pyramidal neurons (i.e, on quantity v_p in Eq. 3.4, which is representative of local mean field potentials). Fig. 3.2 was obtained assuming that the two ROIs are linked with excitatory connections. Of course, power density can change if other kinds of synaptic connections among ROIs are implemented, still keeping these two main rhythms as they depend on the internal parameters of each ROI.

A list of parameters for the average numbers of intra-region synaptic contacts C_{ij}, and for the reciprocal of synaptic time constants, w_i, is reported in Table 3.1.

These internal parameters have been maintained constant and equal for all ROIs throughout the following simulations.

Table 3.1: Parameters setting used in the Neural Mass Model to simulate dynamics in a single ROI.

Internal Parameters	
Connectivity constants:	
C_{ep}	40
C_{pe}	40
C_{sp}	40
C_{ps}	50
C_{fs}	20
C_{fp}	40
C_{pf}	60
C_{ff}	20
Reciprocal of synaptic time constants:	
ω_e	75 s^{-1}
ω_s	30 s^{-1}
ω_f	300 s^{-1}
Synaptic gains:	
G_e	5.17 mV
G_s	4.45 mV
G_f	57.1 mV
Saturation value of the sigmoid:	
e_0	2.5 Hz
Slope of the sigmoid:	
r	0.56 mV^{-1}

As stated previously, for each model configuration ten simulations were repeated and the model output signals (post-synaptic membrane potentials v_p of each ROI) of these ten trials fed as input to software Trentool, for TE estimation and comparison with surrogate data. The length of simulated signals was 60 seconds in general, but the effect of signal length on TE estimation was also assessed (see section Results). Finally, it is important to remark that for each model configuration, the simulations were performed using always the *same ten seeds* to realize white noise; hence TE differences among model configurations can be ascribed only to differences in synapses (or differences in the input mean value or variance), not to individual random noise realizations.

In order to gain a deeper understanding of the virtues and limitations of TE, in all cases the results of TE estimates were compared with those obtained with a linear delayed correlation coefficient (DCC). For the sake of brevity, all results of the DCC are reported in the Supplementary Material part 2.

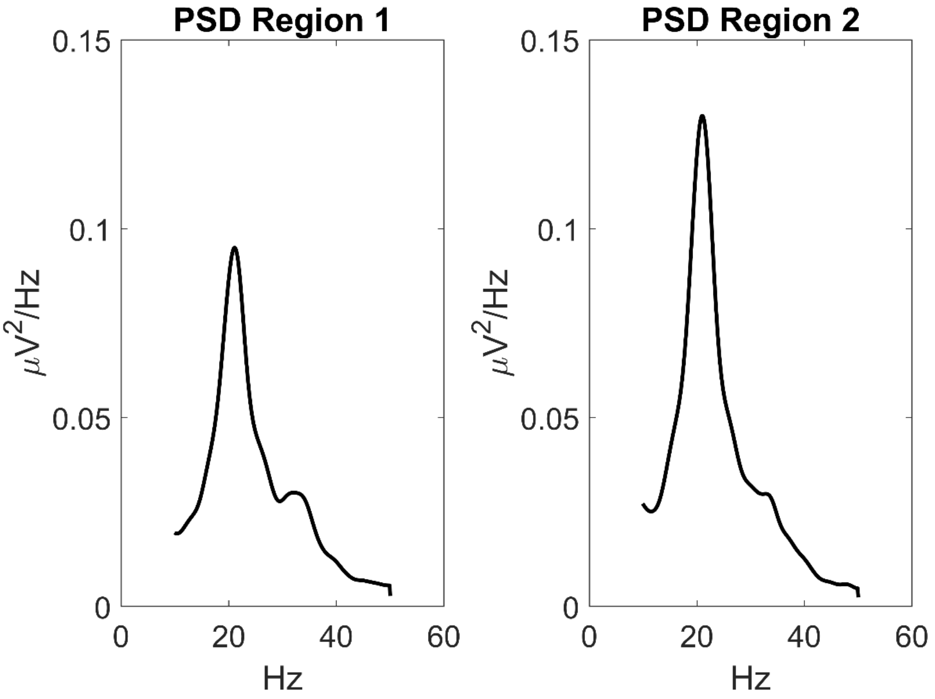


Figure 3.2 – Power spectral density simulated with the model assuming two regions interconnected via excitatory synapses ($W_P^{12} = 40$ and $W_P^{21} = 60$) and no inhibitory connections ($W_F^{12} = 0$ and $W_F^{21} = 0$). Parameters within the neural mass models are reported in Table 1, and maintained for all simulations in this work. The input to the two regions was a random white noise with zero mean values and variance 9/dt (where dt is the integration step, hence the power density of random noise is 9). In these conditions, the two regions exhibit a clear oscillation in the beta range (about 20 Hz) with a contribution in the gamma range too (about 35 Hz). This pattern is similar to the one observed in premotor and supplementary motor areas (see [161,192,193]). It is worth noting that the second region exhibits greater power, since it receives higher excitation.

3.1.5 Results

3.1.5.1 Two interconnected ROIs

A first set of simulations was performed by using two ROIs, linked by means of reciprocal inhibitory and/or excitatory connections.

Fig. 3.3 depicts the TE estimated when the two ROIs are linked via two *excitatory* connections, realized by means of pyramidal-pyramidal synapses W_P^{12} and W_P^{21}, in the absence of any reciprocal inhibitory link. Some aspects of the results are noticeable: i) when the synapse is zero, the relative TE is negligible (i.e., not significantly different from that of surrogate data); ii) TE increases quite linearly with the strength of the synapse; iii) TE from region 2 to 1 increases moderately when the reciprocal synapse (i.e W_P^{21} from ROI1 to ROI2) increases. This effect is made evident by the greater slope in the linear relationships of Fig. 3.3c. By comparing the results of the TE with those obtained with the DCC (see Fig. 3.3S in Supplementary Material part 2) one can observe that synapse strength estimation with TE is more reliable and less affected by the changes in the other synapse; DCC can discriminate between the two synapse strengths (i.e., it is a bidirectional estimator) but estimation of one synapse tends to increase more markedly with the increase in the other.

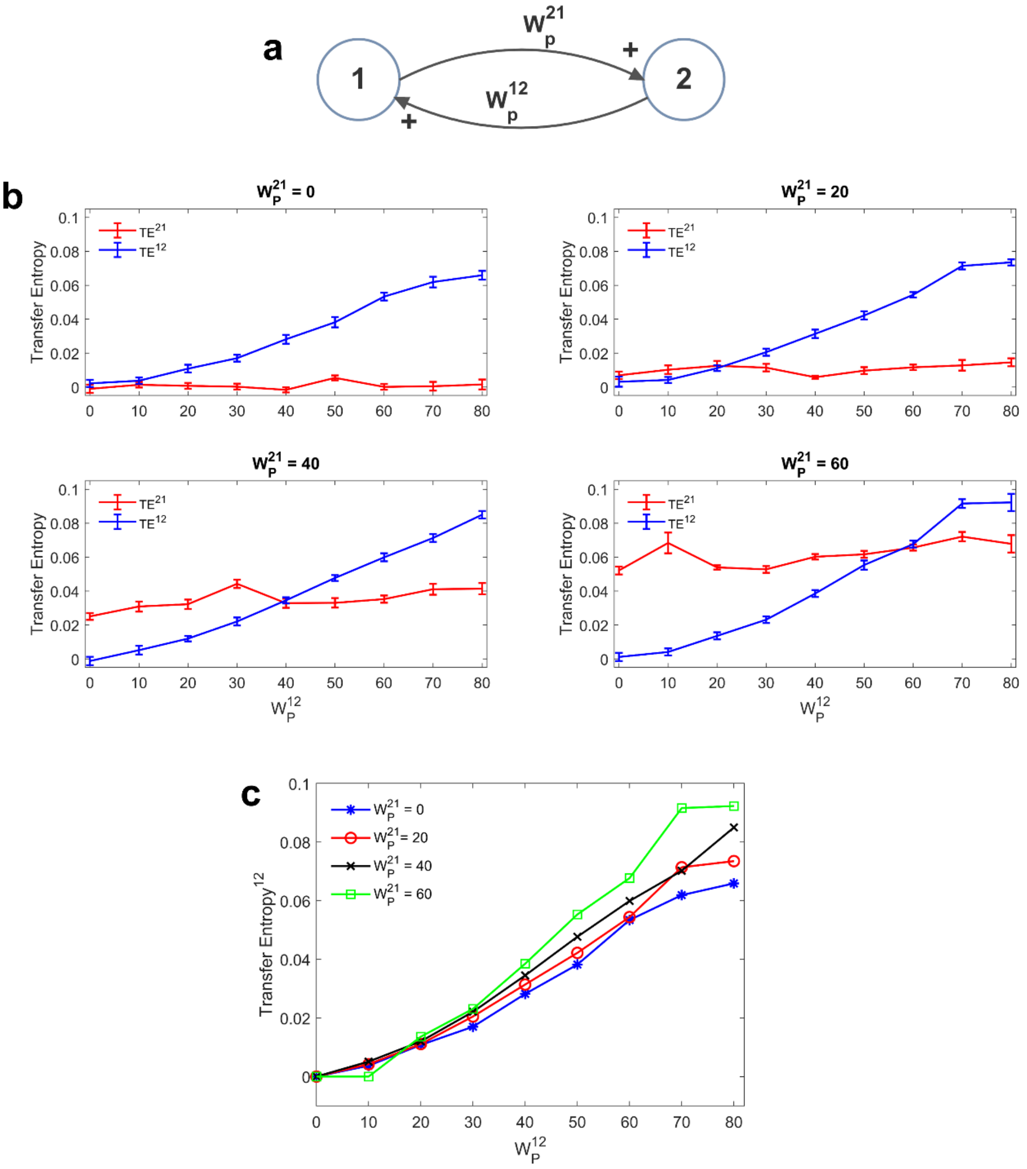


Figure 3.3 – Dependence of Transfer Entropy on feedback, realized assuming two regions interconnected with reciprocal excitatory synapses (Fig. 3.3 a). In particular the synapse from region 2 to region 1 (W_P^{12}) was progressively varied between 0 and 80, at different values of the synapse from region 1 to region 2 (W_P^{21}). Inhibitory synapses were set at zero. Fig. 3.3 b reports the individual values (TE^{12} and TE^{21} ± Standard Error of the Mean (SEM)), obtained with all combinations of synapses. Results concerning the transfer entropy TE^{12} from region 2 to region 1 are further summarized in Fig. 3.3 c. As it is clear, TE increases quite linearly with the value of the excitatory synapse in the direction under study, and it is also moderately affected by the value of the excitatory synapse in the other direction.

Fig. 3.4 depicts the TE estimated when the two ROIs are linked via reciprocal bi-sinaptic *inhibitory* connections. Results substantially confirm that TE increases quite linearly with the synapse strength. However, some differences are evident compared with the excitatory case.

First, the effect of an inhibitory connection on TE is much more efficacious than the effect of an excitatory link. In fact, an increase in the synapse W_F^{12} from 0 to 25 - 30 causes an increase in TE from 0 to approximately 0.06 in our simulated data. To produce the same effect, an excitatory synapse (say W_P^{12}) should be increased from zero to approximately 60. Hence, in our particular model realization, inhibitory connections are about twofold more efficacious in information transmission compared with the excitatory connections. Second, we observed a peak in the estimation of TE when one synapse (either W_F^{12} or W_F^{21}) is set at zero and the other has a value as high as 30. Assuming that this peak represents a failure in the algorithm accuracy, we repeated the estimations of TE using a greater number of trials (30 instead of 10). We observed that, with 30 trials the peak in Fig. 3.4c disappears (i.e., we have a TE value as low as 0.0723 for W_F^{12}= 30 and W_F^{21}=0, while the other values remain very similar to those computed with 10 trials).

In some other cases (when W_F^{21}= 30 or 40 and W_F^{12}greater than 25) Trentool fails to find a correct solution; the problem here is related with the reconstruction of states from scalar time series using time-delay embedding. In particular, TRENTOOL tries to optimize both the embedding dimension and the embedding delay according to Ragwitz' criterion (see Trentool manual); this procedure provides an error in these particular cases.

Comparison with DCC (Fig. 3.4S in Supplementary Material part 2) shows that correlation can be used to detect the sign of the synapse (i.e., DCC provides negative value in case of inhibitory connections, whereas TE is always positive) and is more regular (i.e., it does not exhibit sudden peaks). However, in this case too, as in Fig. 3.3S, the synapse strength estimation by DCC increases markedly with an increase in the reciprocal synapse.

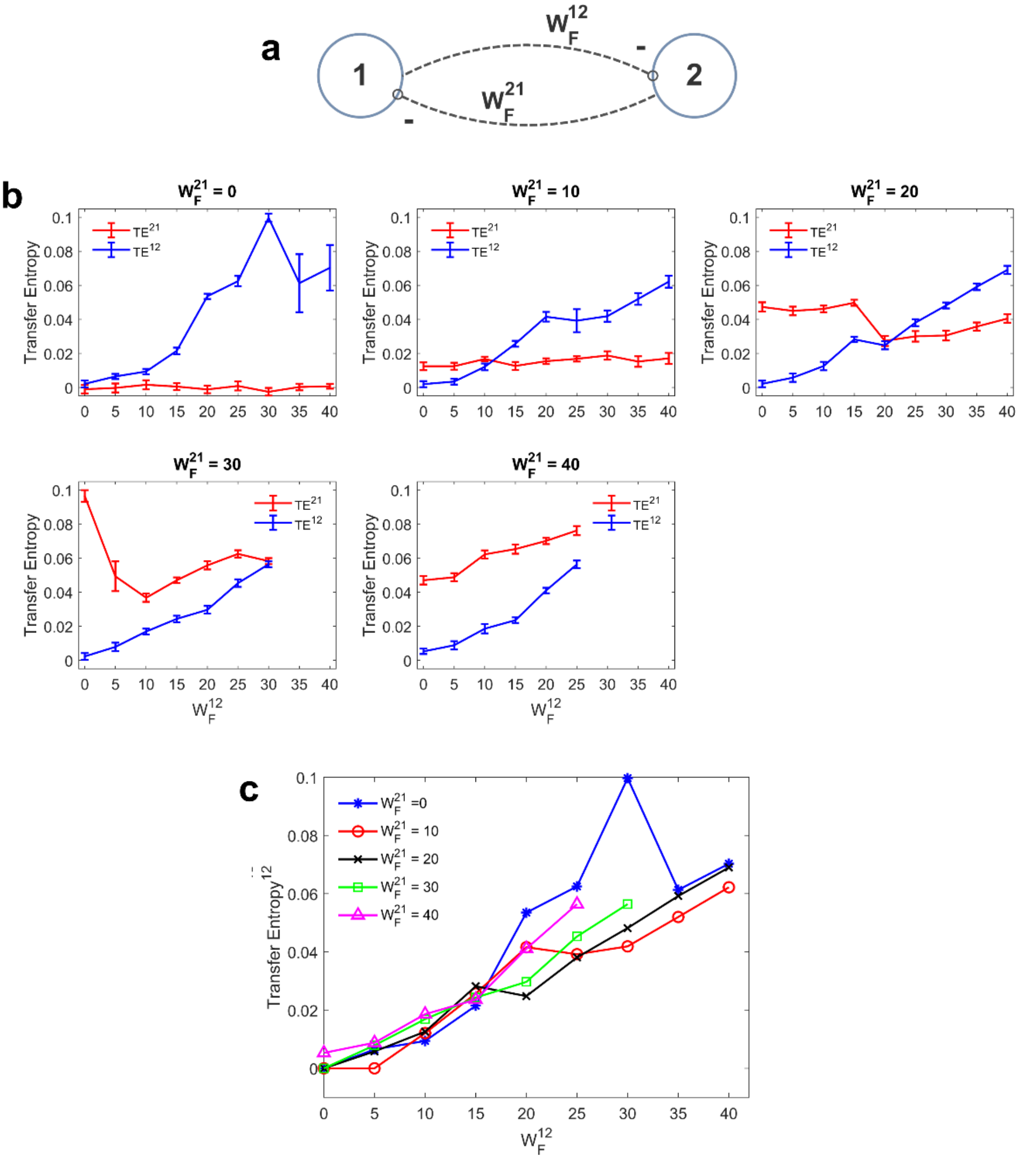


Figure 3.4 – Dependence of Transfer Entropy on feedback realized assuming two regions interconnected with reciprocal inhibitory synapses (Fig. 3.4a). In particular the synapse from region 2 to region 1 (W_F^{12}) was progressively varied between 0 and 40, at different values of the synapse from region 1 to region 2 (W_F^{21}). If both inhibitory synapses are too high, the algorithm fails to compute acceptable values of TE. Excitatory synapses were set at zero. Fig. 3.4b reports the individual values (TE^{12} and TE^{21} ± SEM), obtained with all combinations of synapses. The results concerning the transfer entropy TE^{12} from region 2 to region 1 are further summarized in Fig. 3.4c. As it is clear, TE increases with the value of the inhibitory synapse in the direction under study; but the value of the inhibitory synapse in the other direction affects the estimation significantly. It is worth noting that the effect of inhibitory synapses on TE is stronger than the effect of excitatory synapses (let us compare results of Figs. 3.4b, 3.4c with those in Figs. 3.3b, 3.3c).

3.1.5.2 Three connected ROIs

Various simulations were performed by assuming three interconnected ROIs (named 1, 2 and 3 in the following) with reciprocal connections (either inhibitory or excitatory). This is a multivariate condition; for instance, the estimated TE^{12} from ROI2 to ROI1 may be affected also by the connections between ROI1 and ROI3 and between ROI2 and ROI3. Hence, we expect that results will be much less linear than in the previous case.

A first simulation was performed assuming that ROI1 and ROI2 receive a common input from a third region (ROI3). The schema is depicted in Fig. 3.5a. The strength of this common input was then progressively raised (Figs. 3.5b and 3.5c). Results suggest that estimation of TE is only moderately affected by the presence of a shared input. An increase of this input causes just a moderate reduction in TE, which endangers linearity especially at low values of the synapse W_P^{12}. A similar independence on the shared input can be observed looking at the DCC, too (Fig. 3.5S in Supplementary Material part 2). However, it is worth noting that, in these simulations, we used the same delays (16.5 ms) for all connections: this could induce a resonance. More complex conditions, using different delays, can be tested in future studies.

Further three-ROIs simulations were performed using the more complex schema depicted in Fig. 3.6b, where ROI2 and ROI3 are in competition via reciprocal inhibitory synapses, and exchange excitation with ROI1. Thirty-eight different combinations of excitations and inhibitions were tried. Since results are quite numerous, we do not describe all cases in detail, but just a global summary is reported in the plots of Fig. 3.6a. In these plots we show the value of TE estimated in a single pathway, as a function of the synapse strength used in that path, while the others synapses are varied (for instance, in the upper left panel in Fig. 3.6a, W_P^{21} is varied from 0 to 60, while the other synapses are varied, for a total of 38 different simulations). As it is clear from this figure, despite the multivariate condition, quite a linear relationship is maintained between the estimated value of TE and the synapse value in that pathway; however, the correlation between the two quantities decreases significantly compared with the univariate case.

In seven cases out of thirty-eight in Fig. 3.6, at least one synapse was set at zero. However, due to the presence of a multivariate condition, a residual TE is computed by the algorithm despite the absence of a direct causal link. The situation is summarized in the seven snapshots of Fig. 3.6c, where the “spurious” TE value (i.e., the value associated with the null synaptic connection) is reported, together with the network generating such a false estimate. It is worth noting that the spurious TE is always the consequence of a bi-synaptic link (highlighted in red), and its value quite regularly reflects the strength of this link. However, the spurious TE is always quite small (< 0.025). Only in the last snapshot, where two synapses are simultaneously set at zero, spurious TE values increase to approximately 0.04 or more. However, they are still much smaller than TE values associated with “true” synapses.

By comparing the results in Fig. 3.6 with those obtained with the DCC (Fig. 3.6S in Supplementary material part 2) one can observe a similar behaviour in the estimation of most synapses. The main difference is that TE provides a much better estimation of the inhibitory synapse W_F^{32} than DCC. DCC, however, is able to discriminate between excitatory connections and inhibitory bi-synaptic connections, providing negative values in the last case.

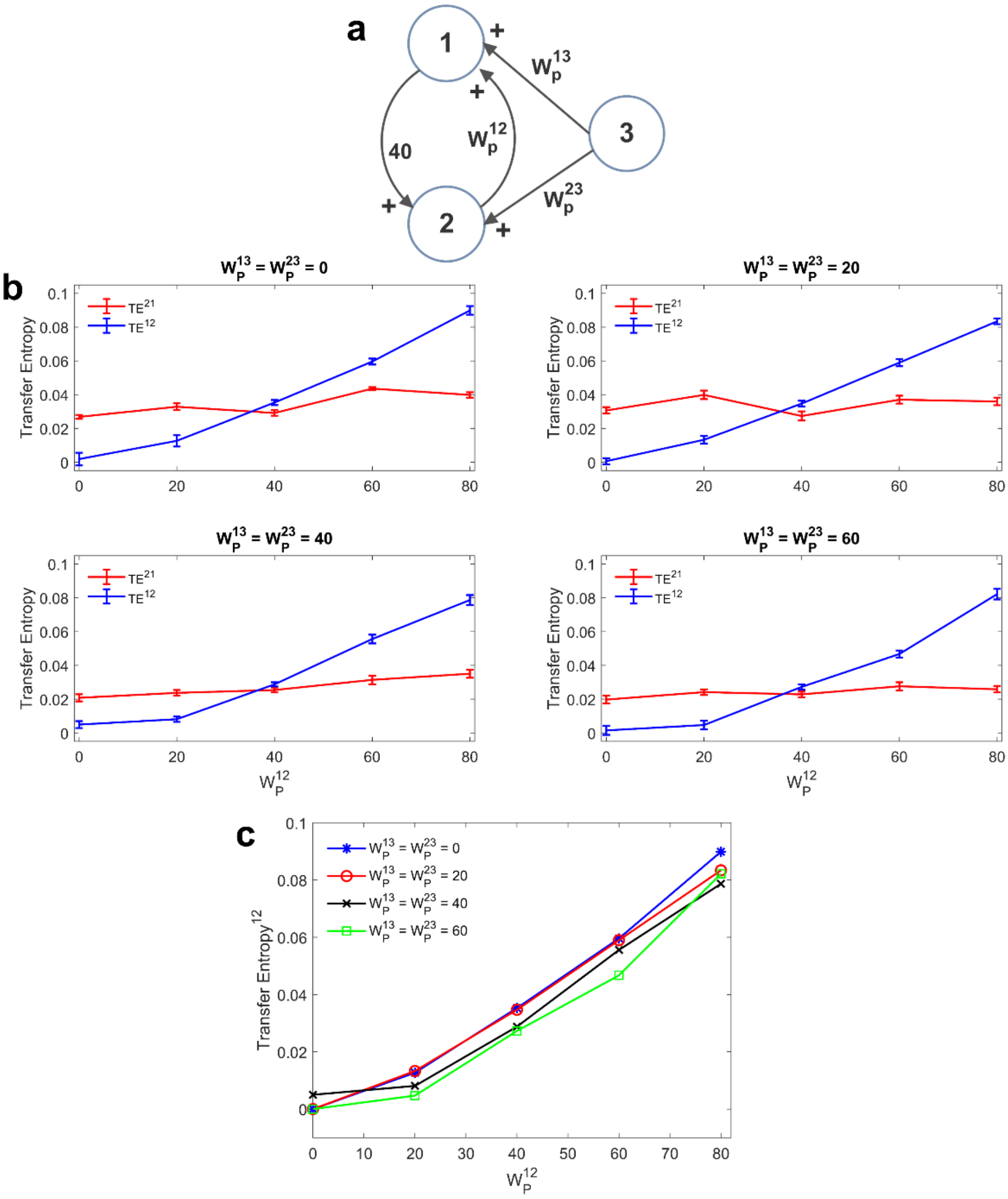


Figure 3.5 – Influence of a common external source on TE estimation. Simulations were performed assuming three regions (see Fig. 3.5a) interconnected via an excitatory synapse from region 2 to region 1 (W_P^{12}), which was progressively varied between 0 and 80, and a constant excitatory synapse in the other direction set at the value $W_P^{21} = 40$. The two regions 1 and 2 also receive a shared input coming from the third region, via equal excitatory synapses $W_P^{13} = W_P^{23}$. Fig. 3.5b reports the individual values of TE (TE^{12} and TE^{21} ± SEM), obtained at different strength of the input from region 3. The results concerning the transfer entropy TE^{12} from region 2 to region 1 are further summarized in Fig. 3.5c. TE increases linearly with the value of the excitatory synapse and is quite independent of the presence of an external shared input.

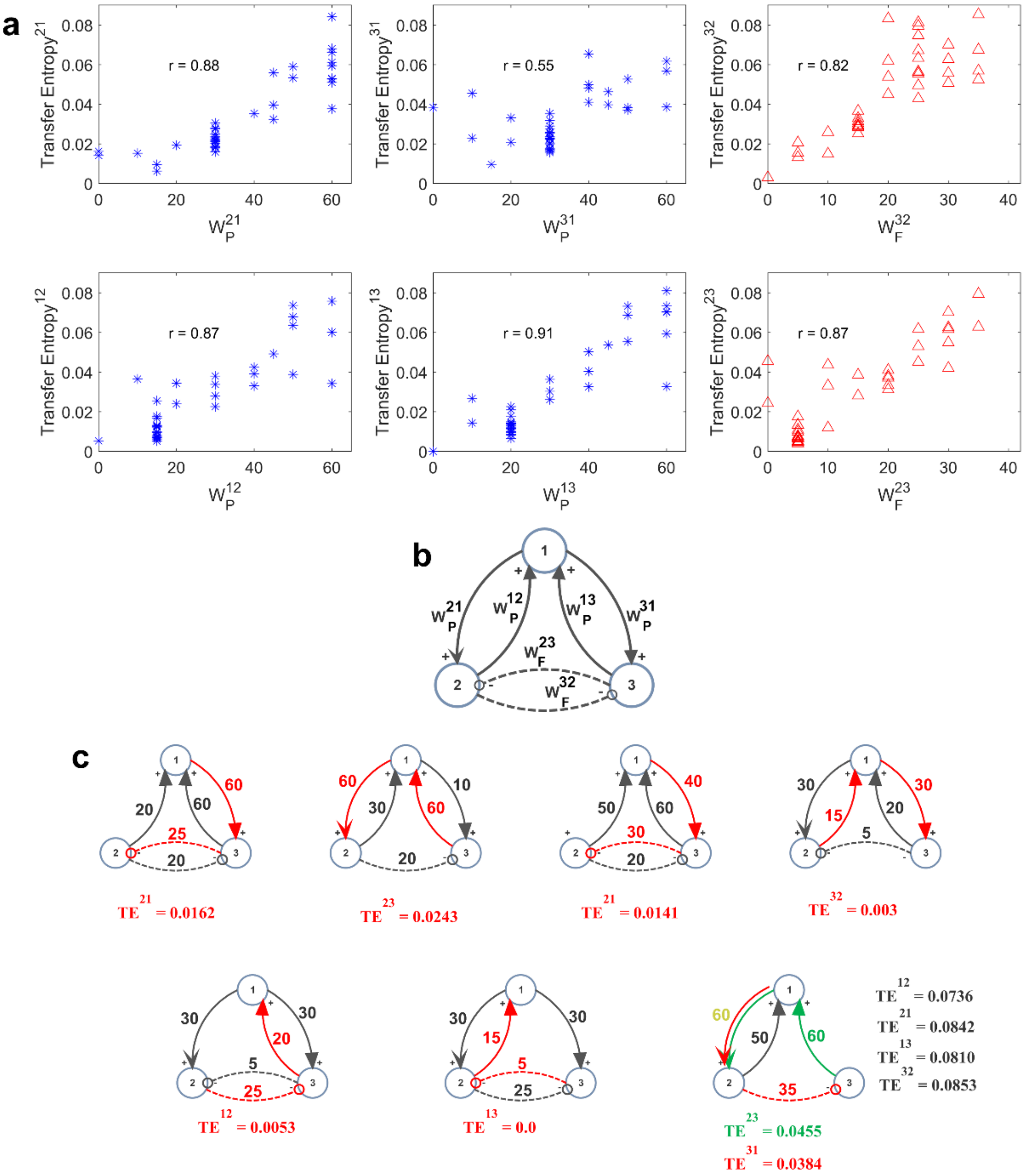


Figure 3.6 – Effect of different combinations of synapses on TE in a model of three interconnected regions. Simulations were performed by using the schema depicted in Fig. 3.6b, where regions 2 and 3 are in competition via inhibitory synapses (W_F^{32} and W_F^{23}), and are linked via excitatory synapses to region 1 (synapses W_P^{12} , W_P^{21} , W_P^{13} , and W_P^{31}). All other synapses are set at zero. Thirty-eight simulations were performed with various combinations of the six synapses described above. The panels in Fig. 6a show the TE in a given direction, as a function of the synapse in the same direction, while the other synapses were varied (SEM are not reported here for simplicity, but are of the same order as in the other figures). Quite a high positive correlation is evident; however, the effect of the other synapses has a strong role in modulating the value of TE. The snapshots in Fig. 3.6c summarize all simulations performed with at least one synapse set at zero. A spurious TE can be ascribed to the presence of a bi-synaptic link (red or green lines). Only in the last snapshots (right bottom), all values of TE are reported, to compare the spurious values of TE with those of real synaptic links.

3.1.5.3 Four connected ROIs

A further set of simulations was performed using four interconnected ROIs. In order to mimic a physiological schema, we assumed that two ROIs represent regions located in the left hemisphere (ROIs 1 and 3) and the other two represent regions in the right hemisphere (ROIs 2 and 4). Moreover, we assumed that excitation in one cortex can lead to inhibition of the symmetrical area in the other cortex and vice versa, according to the Theory of Inhibition (see [198]) whereas feedback excitations can be present between the previous layer and the subsequent layer. A similar schema may occur, for instance, considering the connections between the two Supplementary Motor and the two Primary Motor areas [199,200]. Six different networks, which differ as to the number and strength of connections were simulated (Fig. 3.7a). A comparison between the estimated TE values and the model synaptic strengths is reported in Fig. 3.7b (in all cases, to allow a direct comparison, the values are normalized to the maximum for each configuration). As it is clear from this figure, just in a few cases TE can produce some spurious connections (*one* W_P^{32} in the configuration n. 2, *two* W_P^{23} and W_P^{32} in the configuration n. 5 and *one* W_P^{23} in the configuration n. 6, if we consider a threshold as low as 0.1 to discriminate between the presence or the absence of a synapse). In most cases, TE overestimates the synapse W_P^{32}. Synapses W_P^{14} and W_P^{41} are underestimated in the configurations 3 and (especially W_P^{41}) in the configuration 5. In general, however, the overall behaviour is satisfactory, with a high correlation between the normalized synaptic strengths of the models and the normalized TE values (Fig. 3.7c). It is worth-noting that estimation of the inhibitory by-synaptic connections is more reliable than the estimation of the excitatory connections.

A comparison with Fig. 3.7S in the Supplementary Material part 2 shows that TE is much more reliable compared with the DCC in the evaluation of 4 interconnected ROIs. Briefly, the number of spurious connections is higher, the difference between the normalized DCC values vs. the normalized model synaptic values is higher, and the correlation between the DCC values and the true synaptic weights much poorer when using DCC than TE.

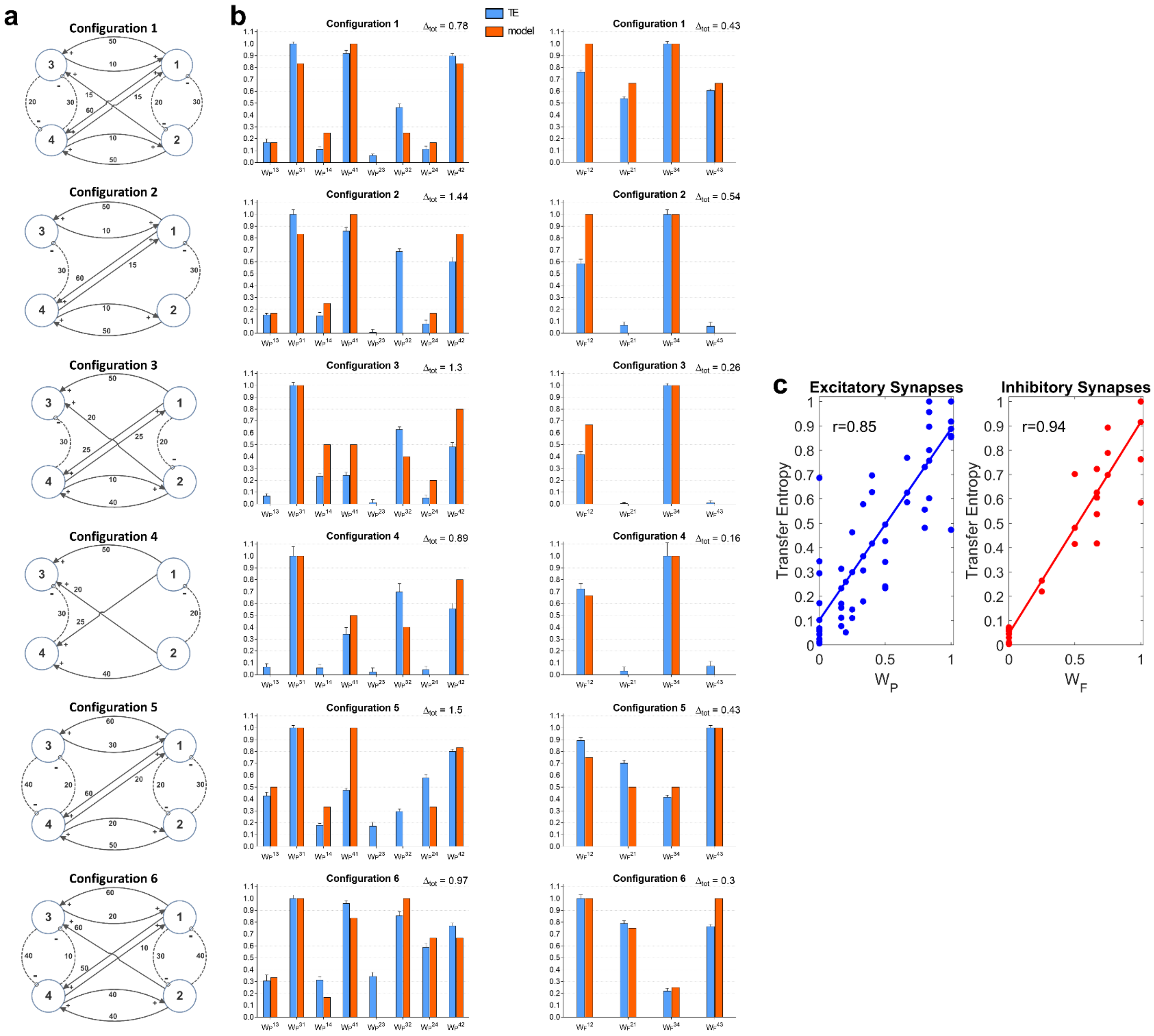


Figure 3.7 – Estimation of the connectivity strength obtained during six different simulations, each performed with four interconnected ROIs. Each row refers to a different network configuration (see reported in Fig. 3.7a) in which the ROIs 1 and 3 belong to one hemisphere, and the ROIs 2 and 4 to another hemisphere. Each ROI exchanges *inhibitory* connections with the adjacent ROI at the same layer in the other hemisphere (1 vs. 2 and 3 vs. 4), and feedback excitatory connections with ROIs of the other layer (1 and 2 vs. 3 and 4). Bars in the two columns of Fig. 3.7b compare the estimated TE values (± SEM) with the true connectivity values in each circuit, normalized to the maximum (the graph bar in the first column considers the eight excitatory synapses, the graph bar in the second column the four inhibitory synapses). Finally, Fig. 3.7c reports the correlation between all the estimated TE values and the true connectivity values, normalized to the maximum, for the excitatory and the inhibitory synapses.

3.1.5.4 Effect of the input mean value and SD

The neural mass model used in this work is intrinsically non-linear (due to the presence of sigmoidal relationships which mimic the dependence of spike density on post-synaptic membrane potential in individual neural populations). Conversely, most methods used to assess connectivity from data (like PDC or most implementations of DCM) assume a linear model in the estimation process. TE does not assume any model behind signals, but simply computes information transfer from the source to the target.

However, it is to be stressed that, due to non-linear effects, information transfer may not reflect true anatomical connectivity. Hence, it is of the greatest value to assess how TE estimation may change in conditions when all synapses are fixed (i.e., a constant anatomical connectivity is used) but the working point or the global activity in the neural populations is modified. Results of this analysis are reported in Fig. 3.8 and Fig. 3.9 as detailed below. For the sake of simplicity, we show results obtained using two-ROIs model, with the two ROIs connected via reciprocal excitatory synapses (schemes in Fig. 3.8a and Fig. 3.9a).

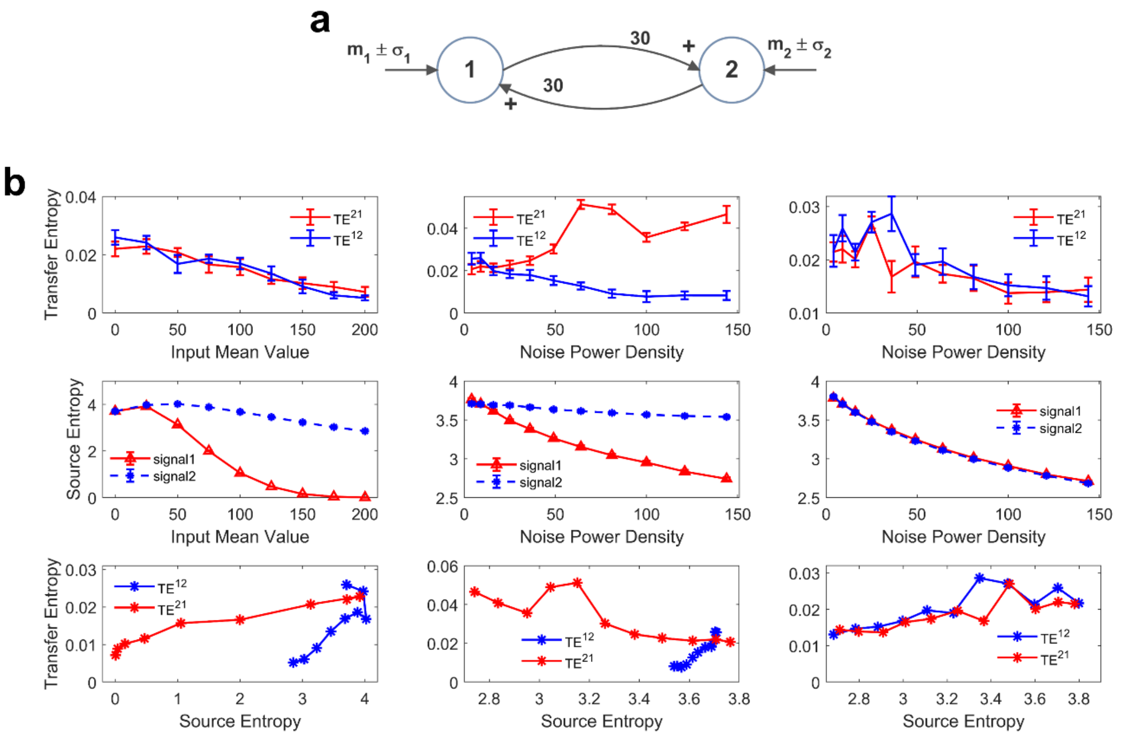


Figure 3.8 – Effect of the mean value and standard deviation of the input noise on the estimation of Transfer Entropy. The simulations were performed using two interconnected regions (Fig. 3.8a), with synapses $W_P^{12} = W_P^{21} = 30$, $W_F^{12} = W_F^{21} = 0$ and by varying the mean value m₁ (left panels in Fig. 3.8b) and standard deviation σ₁ of noise (middle panels in Fig. 3.8b) of the input to ROI1. Standard deviation (σ) of the noise was computed as $\sigma = \sqrt{\rho/dt}$ where *dt* is the integration step and ρ is the noise power density. Finally, the right panels in Fig. 3.8b show the case when noise standard deviation was increased in both populations altogether (both parameters σ₁.and σ₂). In Fig. 3.8b, the first row shows TE ± SEM, while the second row shows entropy (± SEM) of the two signals, vs. the input values. Finally, the third row in Fig. 3.8b plots the TE vs. the entropy of the source

signal. It is worth noting the presence of quite a linear dependence of TE on the source entropy, with the only significant exception of TE^{21} in the middle panel.

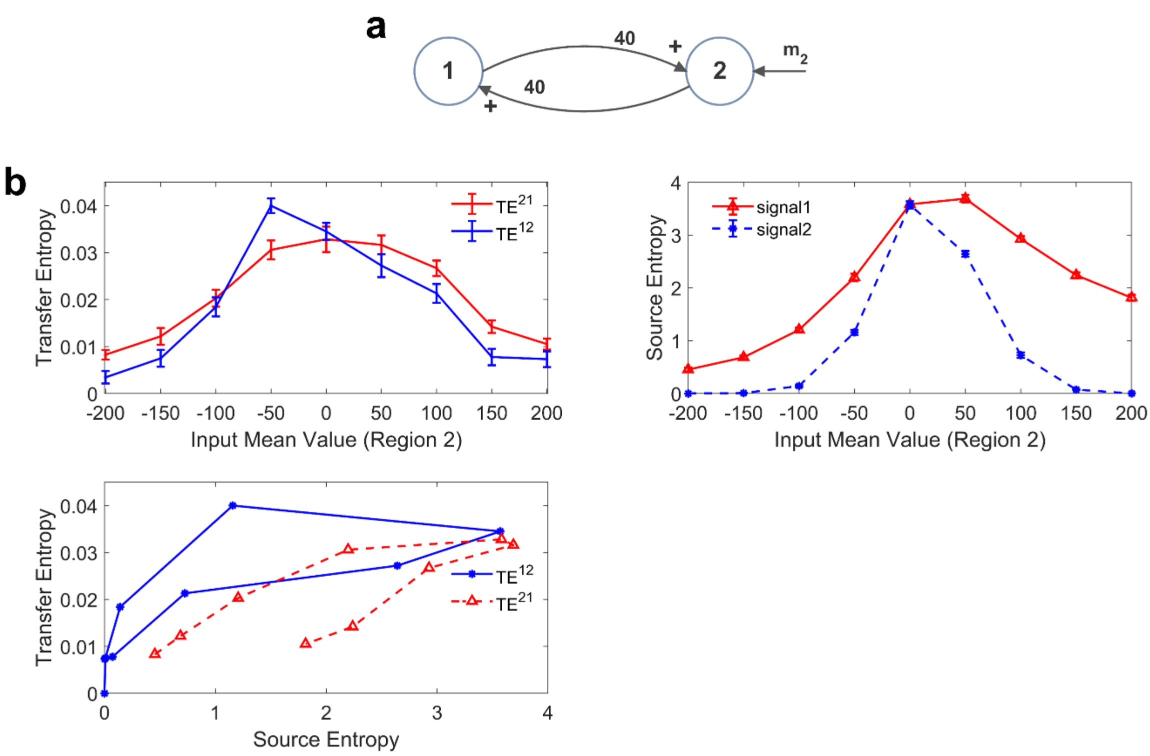


Figure 3.9 – Effect of the region working point on the estimation of Transfer Entropy. The simulations were performed using two interconnected regions (Fig. 3.9a), with synapses $W_P^{12} = 40$, $W_P^{21} = 40$, $W_F^{12} = 0$, and $W_F^{21} = 0$ (i.e., just a reciprocal excitation). In this case we assumed that input to pyramidal neurons in region 1 have a zero-mean value, whereas mean value m$_2$ of the input to pyramidal neurons of region 2 is progressively increased from −200 (strong inhibition) to +200 (strong excitation). In Fig. 3.9b, the left plot shows TE ± SEM, while the right top plot shows entropy (± SEM) of the two signals, vs. the input values. Finally, the bottom left plot shows the TE vs. the entropy of the source signal. As it is clear, transfer entropy reaches a maximum value when the first region exits from the inhibition zone to the central zone. Furthermore, TE declines when regions enter into the upper saturation, due to excessive excitation. It is worth noting the presence of quite a linear dependence of TE on the source entropy, although with a hysteresis.

In order to test non-linear phenomena, first we modified either the mean value or the standard deviation of the noise entering to pyramidal neurons in one ROI. Indeed, a change in the mean value shifts the working point along the sigmoidal relationship. An increase in SD causes large oscillations in neuronal activity, which may be partly cut-off by the saturation levels of the sigmoid. We also tested the effect of changing the standard deviation of the noise to both ROIs. We remark that all previous simulations were performed with zero mean values and variance $\sigma^2 = 9/dt$, where dt is the integration step.

Since these changes may induce a change in the entropy of the source, we also computed the entropy of the two signals and we evaluated the relationship between TE and the entropy of the source signal.

The left panels in Fig. 3.8b show the effect of an increase in the mean value of the input to ROI1 (parameter m_1 in Fig. 3.8a). Increasing this value causes a significant decline in the estimated value of the TE from 1 to 2. The reason is that activity of pyramidal neurons in ROI1 approaches the upper saturation, hence its entropy is dramatically reduced, and the quantity of information transmitted from 1 to 2 is reduced too. It is worth noting that also TE from 2 to 1 decreases. The reason is that also ROI2 exits from the central linear region, as a consequence of the strong excitation coming from ROI1, thus causing a moderate reduction in its entropy.

The middle panels in Fig. 3.8b show that an increase in the standard deviation of noise entering into ROI1 (parameter σ_1 in Fig. 3.8a, whereas noise to ROI2 is maintained at the basal level) causes a dramatic increase of TE from 1 to 2, despite a reduction in the entropy of signal 1. At the same time, TE from 2 to 1 is reduced. This result indicates that the values of TE do not only reflect the connectivity strength, but also the reciprocal level of noise in the two ROIs.

Finally, in the right panel of Fig. 3.8b we tested the case when both standard deviations (to ROI1 and to ROI2, i.e. parameters σ_1 and σ_2 in Fig. 3.8a) are progressively increased altogether. In this condition, both TEs show a similar decrease, but the changes are quite moderate compared with the previous cases, and reflect the moderate decrease in both source entropies.

Looking at the bottom row in Fig. 3.8b, we can observe the presence of quite a linear relationship between TE and the entropy of the source. There is only one remarkable exception; the increase in noise of signal 1 in the middle panel reduces its entropy (due to a saturation) but causes an increase in information transmission from 1 to 2, which is reflected in a negative relationship between TE and the source entropy.

Similar results can be obtained using the DCC, as shown in Fig. 3.8S of the Supplementary Material part 2.

The effect of the input mean value is further illustrated in Fig. 3.9b where we modified the input mean value entering to ROI2(parameter m_2 in Fig. 3.9a). Here we assume that ROI2 starts from a condition of strong external inhibition (obtained with a negative input mean value). In this initial state, TE is almost zero despite the presence of a strong reciprocal connectivity. Then, the input to ROI2 is progressively increased (i.e., the pyramidal population is progressively excited). As a consequence of this excitation, ROI1 is excited too. In this situation, TE initially increases as a consequence of progressive reciprocal excitation in the network, which corresponds to an increase in the entropy of both signals. When excitation becomes excessive, however, TE starts to decrease (as in the example of Fig. 3.8) since both regions enter into the upper saturation zone, and the source entropies decrease again. The relationship between TE and the entropy of the source (bottom left panel in Fig. 3.9b) is quite linear, although it exhibits a kind of hysteresis.

In this case too, the patterns obtained with the DCC are similar (Fig. 3.9S in the Supplementary Material part 2).

3.1.5.5 Estimation of the pure delay

In our model we included a pure delay in the connectivity among the different ROIs. This simulates the time necessary for spikes to travel along axons and reach a target region starting from a source region. During the previous simulations we used a pure delay as high as 16.5 ms in all synapses. The software package Trentool provides an estimation of this delay, assuming the value which maximizes TE.

In order to assess the role of pure delay, we repeated some simulations with two interconnected ROIs by varying the delay. We examined whether: i) the value is correctly estimated by the algorithm (at least approximately); ii) the estimated value of TE is affected by this delay.

Results, summarized in Fig. 3.10 (with W_P^{12} varying and W_P^{21} fixed), show that the estimated value of TE is reduced, and the linearity in the relationship "TE vs. synapse strength" worsens if a small value is used for the delay (10 ms). Conversely, larger values (16.5 ms or 23 ms) provide robust results, with a moderate increase in TE with larger delays. The values of delay estimated by the algorithm are 8 ms, when we used a delay as low as 10 ms, 15 ms when we use a delay as large as 16.5 ms, and 25 ms when we use a delay as large as 23 ms.

In the left panel of Fig. 3.10b we can observe an anomalous peak in TE when W_P^{12} = 40 In this case too, as in the case of Fig. 3.4c, this peak could be eliminated using 30 trials in the computation of TE (TE = 0.0199).

It is worth noting that 10 ms are about 1/5 of the resonant period of the present model (see the spectra in Fig. 3.1), whereas 16.5 ms is about 1/3 of this period and 23 ms close to ½. This may have an impact in the synchronization of the two circuits.

Results obtained with the DCC (Fig. 3.10S in Supplementary Material part 2) are similar. DCC appears more robust than TE when using a small delay, but even in this case the sensitivity of the estimator (i.e., the slope of the relationship between the metrics and the synapse strength W_P^{12}) increases with the delay. Moreover, as usual, DCC fatigues to assess a constant synapse (W_P^{21} fixed) when the other synapse is varying, at all values of the delay.

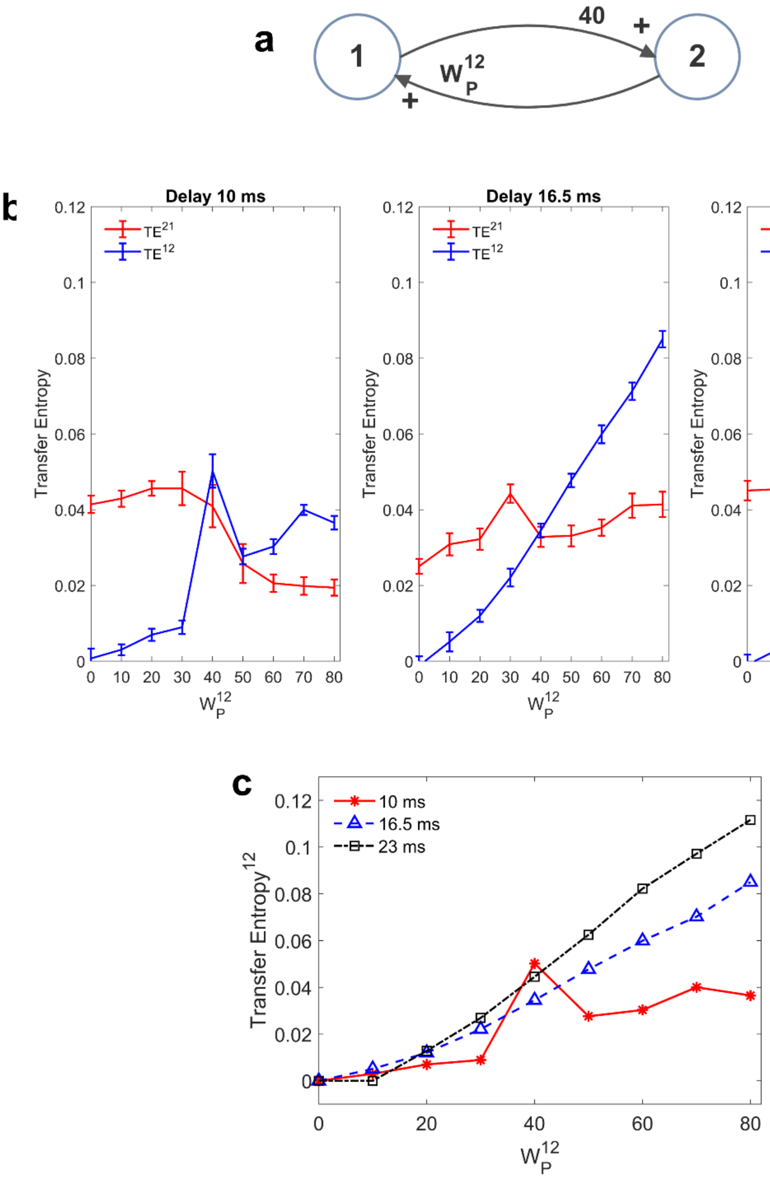


Figure 3.10 – Effect of the delay between the two regions on the estimation of Transfer Entropy. The simulations were performed using two interconnected regions (Fig. 3.10a), with synapses, $W_P^{21} = 40$, $W_F^{12} = 0$, and $W_F^{21} = 0$, and by changing the value of synapse W_P^{12} (hence we have a reciprocal excitation). The simulations were repeated with different delays. It is worth noting that the delay value was estimated by the algorithm together with TE. Fig. 3.10b reports the individual values (TE^{12} and TE^{21} ± SEM), obtained with all combinations of synapses. The results concerning the transfer entropy TE^{12} from region 2 to region 1 are further summarized in Fig. 3.10c. The estimation of TE increases (both as to its strength and linearity) at high values of time delay, and worsens when time delay is reduced.

3.1.5.6 Effect of the signal length

An important aspect in the estimation of TE is the signal length. In neural problems non-stationarity often precludes the use of long signals; the use of short signals, in turn, may jeopardize the estimation accuracy. In all previous simulations we used long stationary signals (60 s of simulations with fixed parameters and a stationary random noise, evaluated after settling the initial transient phenomena). We repeated simulations using a useful signal length (after elimination of the initial transient period) as low as 30s, 10s, 4s, 3s and 2s. It is worth noting that, even when using the shorter temporal window, at least 40 cycles of network oscillations are contained within the examined portion of the signal. Results are summarized in Fig. 3.11. We can observe that, in the range 3s – 60 s, a reduction in signal length causes a reduction in the estimated value of TE, but the linearity in the relationship "TE vs. synapse strength" is approximately preserved even with the use of shorter signals (although linearity becomes more evident for signal length ≥ 10 s). Hence, caution must be taken when comparing TE values obtained from signals with different length. Conversely, when the signal is as short as 2s, TE becomes quite insensitive to the synapse strength, although a value of TE significantly different from that of surrogate data is still detectable.

Conversely, DCC (Figure 3.11S in Supplementary Material part 2) is almost no affected by the signal length. Once again, however, DCC fails to assess that a synapse is constant when the other is varying, whereas TE can recognize a constant synapse even when using short signals.

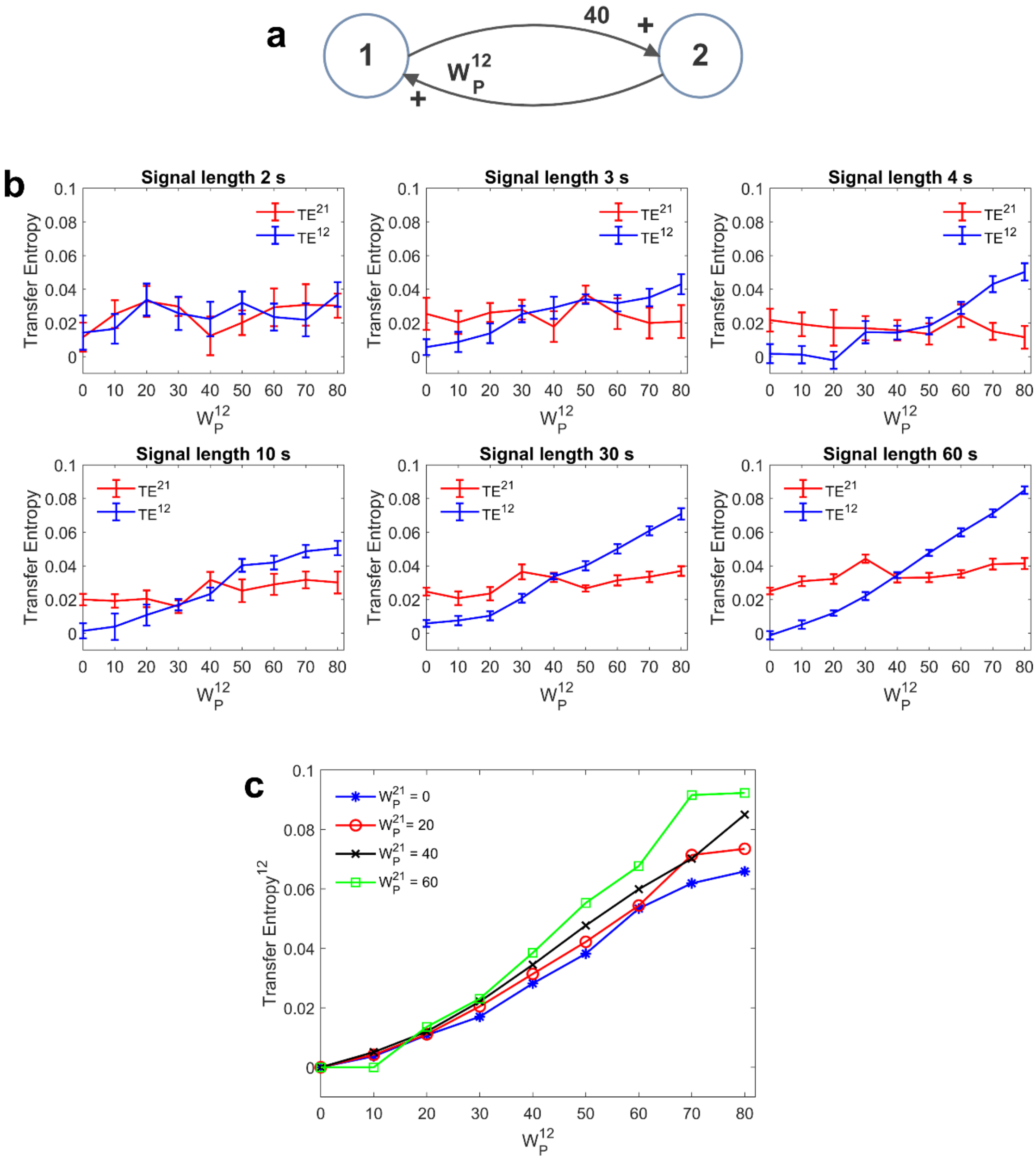


Figure 3.11 – Effect of the duration of the signal on the estimation of Transfer Entropy. The simulations were performed using two interconnected regions (Fig. 3.11a), with synapses, $W_P^{21} = 40$, $W_F^{12} = 0$, and $W_F^{21} = 0$, and by changing the value of synapse W_P^{12} (hence we have a reciprocal excitation). The simulations were repeated with different durations of the signals. Fig. 3.11b reports the individual values (TE^{12} and TE^{21} ± SEM), obtained with all combinations of synapses. The results concerning the transfer entropy TE^{12} from region 2 to region 1 are further summarized in Fig. 3.11c. The estimation of TE and its linear relationship with the synaptic strength are reduced when the signal length becomes as low as 3-4 s and, at shorter distance, the algorithm totally fails to detect a linear relationship between TE and connectivity (let us note that the relationship between TE^{12} and synapse W_P^{12} becomes flat when the signal length is as short as 2 s).

3.1.6 Discussion

In the last decade the use of methods to assess connectivity from neuroimaging data has received an enormous attention, as a fundamental aspect of cognitive neuroscience [165]. In fact, adequate understanding of brain functioning can be obtained only by considering the brain as a fully integrated system, the parts of which continuously exchange information in a dynamical reciprocal way [201,202]. However, much debate is still present in the literature on the reliability of methods used to assess connectivity, and on the true significance of the indices extrapolated from data [172].

Aim of this work is to assess the reliability of a non-linear measure (Transfer Entropy) used to investigate connectivity. Indeed, various recent papers underlined that TE represents an efficient method to estimate connectivity, which, compared with other methods, joins reliability and smaller computational time [174,184]. Nevertheless, the relationship between TE, as a measure of information transfer, and the true anatomical connectivity between regions is still debated. In order to clarify this problem, we evaluated the significance of the connectivity values derived from bivariate TE by using the data generated through realistic models of neural populations coupled with assigned connectivity parameters. In particular, we investigated whether: i) the method is able to discriminate between the presence or absence of connectivity between populations; ii) the method is sensitive to a progressive change in the strength of the connection; iii) the effects of non-linarites on TE estimation, in particular the effect of a change in populations working point; iv) the effect of signal length and time delay. Moreover, all results have been compared with those obtained with the linear delayed correlation coefficient (See Supplementary Material Part 2).

The importance of performing accurate validation studies for FC methods, based on simulation data, has been strongly emphasized in a recent perspective study by Reid et al. [172]. Both detailed neuron-level simulation models [203,204] or more abstract models, such as neural masses [161,194,205–210] can be of value to reach this objective, with alternative advantages and limitations. Our choice was to use NMMs which, as pointed out by Reidet al.[172], exhibit multiple advantages: among the others, computational efficiency, the possibility to generalize over multiple conditions, and the ease in the interpretation of results.

An important aspect to be recognized, however, is that NMMs are adequate to simulate (although with several approximations) the neuroelectrical activity in cortical columns, which may be significantly different from that measured on the scalp, due to propagation phenomena from the cortex to the skull through the interposed soft tissues. Hence, the present analysis images that the signals, obtained from scalp EEG/MEEG measurements, are first recreated on the cortex, via classic methods for source localization and reconstruction, before TE is calculated on them.

It is worth noting that the present study is focused on the bivariate algorithm for TE estimation (in particular, we used the open source toolbox Trentoool [136], which is largely used in Neuroscience problems today). The use of bivariate instead of multivariate TE surely represents the main limitation of the present study, and some of the errors encountered when simulating three or four populations (such as the presence of spurious connections) can be reduced using multivariate algorithms (such as those proposed in Montalto et al. [182]). This can be attempted in future works. For instance, Harmah et al. [211] in a recent paper, evaluated multivariate TE in people with schizophrenia, and found that multivariate TE outperformed bivariate TE and Granger causality analysis under various signal-to-noise conditions. However, it is to be stressed that this difference between our results and those obtained with multivariate TE are probably not so strong as in other works, since Trentool implements some tools for post-hoc corrections of multivariate effects, i.e. a partial correction of spurious information flow.

Another limitation of the present approach is that we did not use other indices (like the "Coincidence Index" (see [212])) to improve the performance of our estimator. Indeed, the only additional measure we used is the DCC (see Supplementary Material Part 2 and the last paragraph in the Discussion). Recently, Reid et al.[16] suggested the simultaneous use of alternative measures, and their integration into a comprehensive framework, to improve connectivity estimate. This may be the subject of future work.

In order to realize physiologically reliable neural signals, with a frequency content analogous to that measured in cortical regions, we used the model proposed by the authors in recent years [161]. This allows multiple rhythms (for instance in the beta and gamma range) to be simultaneously produced and transmitted between regions, as a consequence of the non-linear feedback between excitatory and inhibitory populations (with glutamatergic, slow-GABAergic and fast-GABAergic synaptic dynamics). In particular, in this study we chose synaptic connections within the ROI (i.e., parameters C_{ij} in Eqs. 3.1 -3.18) to have power spectral densities quite similar to those occurring in pre-motor and supplementary cortical areas during motor tasks [192,193].

It is worth noting that, in order to eliminate a possible bias in the estimation of TE, we always compared the TE value estimated from the model with that obtained on surrogate data (i.e., data with the same statistical properties of our signals but lacking of any connectivity). Hence, TE was set at zero whenever no statistical difference was observed between model signals and surrogate data; in all other cases, the (positive) difference between the model TE and that of surrogate data was assumed as an index of the synaptic strength.

i) *Detection of spurious connections* - A first important result of our study is that the TE algorithm is able to discriminate between the presence of a significant connectivity, and the absence of connectivity rather well, in conditions when two ROIs are interconnected. In particular, in all cases when connectivity in the model was set at zero or at an extremely low value ($W_P \leq 10$ or $W_F \leq 5$) , the algorithm provided no significant difference between model signals and surrogate data (Figs. 3.3-3.4 and 3.10-3.11). A similar result also holds when two populations receive a common signal from a third population, i.e., they have a common external source (Fig. 3.5). Only in one case (Fig. 3.5b left bottom panel) a very mild value of TE is obtained when the synapse W_P is zero. Conversely, in the more complex situations of Fig. 3.6 (three interconnected ROIs), some artefacts can be seen in the computation of TE: we can observe a significant value of TE even when the corresponding synapse is zero. This is generally quite small, with the exception of the last snapshot in Fig. 3.6c, when two synapses are at zero. It is worth noting that these “spurious” connections are always the consequence of a bi-synaptic link from the source region to the target one. Some spurious connections can be found, of course, also when simulating four interconnected ROIs (Fig. 3.7), but their number remains quite limited. It may be interesting in future studies to test whether these “spurious” connectivity values can be eliminated or reduced by using multivariate methods for TE estimation. In particular, Olejarczyk et al. [213] performed a comparison between the multivariate approach and the bivariate one for the analysis of effective connectivity in high density resting state EEG, and found that the multivariate approach is less sensitive to false indirect connections.

The previous results substantially agree with those by Wang et al. [174] who, using signals obtained from NMMs with different connection strengths, observed that the bivariate TE provides high values of the Area under the ROC curves (i.e., a high measure of separability), hence the method performs very well in detecting the underlying connectivity structure. By the way, no evident difference was reported by Wang et al. [174] when comparing the performance of the bivariate TE with that of the partial TE (see Fig. 3.9 in their work).

ii) *Dependence of TE estimation on synaptic strength* – An important result of our study is that quite a linear relationship can be observed between the estimated value of TE and the strength of the connection (either mono-synaptic excitatory or bi-synaptic inhibitory) in the same direction, provided the model is working in the linearity region. We are not aware of a similar analysis in the literature: indeed, most previous studies limit the investigation to the presence or absence of a connection (i.e., on its statistical significance, see also Vicente et al. [135]), or are based on ROC curves (see [174]). The linear relationship is quite straightforward in the case of excitatory synapses (Fig. 3.3), and less precise in case of inhibitory synapses (Fig. 3.4), but is still well evident when two populations are used. Furthermore, we also simulated conditions characterized by an excitation from ROI2 to ROI1 with a simultaneous inhibition from ROI1 to ROI2, and conditions in which ROI2 sends both an excitatory monosynaptic and a bi-synaptic inhibitory connection to ROI1. Sensitivity analysis on these cases, not reported from briefness, confirms what we observed in the other simulations: TE quite linearly depends on synapse strength, and inhibition in our model has a stronger effect than excitation.

In the more complex three-populations multivariate model (Fig. 3.6) a clear positive correlation between TE and synapse strength is still evident, although influenced by the other synapse values. A very good correlation is still evident when using four interconnected ROIs (Fig. 3.7) and, in this case, TE significantly outperforms the delayed correlation coefficient (see below). This result suggest that TE can be used (although with caution) not only to detect the presence or absence of a causal connection, but also to investigate whether this connection is stronger or weaker than another, or it is changing (reinforcing or weakening) with time. As commented below, however, a particular attention must be posed to any change in working conditions of a ROI (i.e., non linearity in the model), since it may affect TE dramatically.

Some authors recently emphasized that synapses among neurons exhibit a log-normal distribution [214] and so that a few strong synapses dominate network dynamics over a large amount of weaker synapses. Although this result has been obtained in networks of hundreds of neurons (whereas our study is concerned with connections among a few ROIs) it may still be of interest in the problem of connectivity estimate, revealing that the main point is the capacity to estimate large synapses correctly (with minor emphasis on the smaller ones). Results in Fig. 3.7 show that this aspect is well managed by TE.

A further important result (although well-expected) is that TE cannot discriminate between an excitatory or a bi-synaptic inhibitory connection among two populations (we remind here that an inhibitory connection denotes a by-synaptic connection, from pre-synaptic pyramidal neurons to post-synaptic fast-GABAergic interneurons, and then to pyramidal neurons in the target population). This is quite obvious, since TE is always positive, and so it detects a sort of "absolute value" for the synaptic strength. In our model, inhibitory synapses are twofold more powerful in affecting signal transmission (hence TE) than excitatory connections. This result, however, depends on the parameters we used to simulate the internal number of synapses within populations. A different choice of internal parameters may modify this result. A similar conclusion (i.e., the incapacity to discriminate between excitatory and inhibitory connections via TE) has been reported in previous studies by considering synapses linking spiking neurons [185,186]; we are not aware of a similar generalization considering the interactions among ROIs via NMMs simulations. However, as shown in Supplementary Material Part 2, the delayed correlation coefficient is able to discriminate between excitatory and inhibitory connections very well. Hence, the simultaneous use of both metrics may allow this limitation to be easily overcome.

Finally, we wish to stress that the present study is devoted to the analysis of interactions among brain regions, and the mathematical model used to generate data simulates neural population dynamics; hence the results are not immediately applicable to the interactions among individual neurons.

iii) *Effect of non-linearities* – A very important result of the present study is that TE strongly depends on the working point of the populations, and on the SD of the input noise. The first aspect is a consequence of the sigmoidal characteristics in the model, which describe the non-linear relationship linking post-synaptic membrane potential to spike density. In particular, whenever a population of pyramidal neurons enters into a saturation region, its capacity to transmit information towards other ROIs drastically decreases, despite the presence of a strong synapse. This is basically a consequence of a reduction in the entropy of the source signal (see Figs. 3.8 and 3.9, but see also Wollstadt et al.[215] as an example of a reduction in source entropy induced by isofluorane anaesthesia) although a significant exception can be found in one case in Fig. 3.8. This aspect is crucial in the interpretation of TE, and has been clearly recognized by Wibral et al [181]. These authors, when commenting on the relationship between TE and causality, underline the distinction between information transfer and causal interactions. In particular, they suggest that TE is not a measure of causal strength and that not all causal interactions serve the purpose to transmit information (see [181], pages 8-11). Hence, when using TE in the field of neuroscience, one must always take in mind that TE measures the amount of information that is transmitted from one region to another (or, in case of multivariate models, a certain amount of information transmitted through a bi-synaptic link). A high value of TE likely denotes the presence of a causal relationship, since information cannot be transmitted without coupling (care, however, should be taken in multivariate models to a shared information which may provide a spurious TE [187]). Conversely, a small value of TE does not necessarily indicate the absence of a causal link. Let us consider, for instance, the case in which activity in a pre-synaptic region shifts its working point reaching its upper saturation level (i.e., maximal activity of pyramidal neurons) and sends a strong synapse to a target population. Thanks to this coupling term, the second population can also enter into saturation (see for instance, Fig. 3.8b leftmost panel and Fig. 3.9b). At this point, TE is drastically reduced, and the appearance is that of poor information transmission. However, the first ROI may play an extremely important causal role on the second population even in this condition of poor information exchange. Let us consider, for instance, the case when a target region participates to a winner takes all dynamics against other regions: the causal link may be fundamental for it to win the competition, but is not detected (or just poorly detected) with the use of TE.

Moreover, TE is dramatically affected by the power level of input noise. In particular, it is not the level of noise per se which affects TE, but rather the relative contributions among the two populations. If noise to both populations rises together, TE does not increase, but rather exhibits a moderate reduction (Fig. 3.8b rightmost panel). We attribute this decrease, evident only at very elevated power, to the presence of saturation in the sigmoid, which cut-offs the entropy of the source signals. Conversely, if one population receives much stronger noise than the other (Fig. 3.8b middle panel), the amount of Entropy that it can transfer dramatically rises, while TE of the other is reduced, despite the presence of a similar reciprocal connectivity strength. Indeed, the first population can transmit much new information to the second, in the form of random fluctuations, while the information transmitted from the second to the first becomes quite negligible. In this particular case, we observed a surprising negative correlation between entropy in the first population (which decreases due to saturation) and TE it transfers to the second, that rather increases (see the bottom central panel in Fig. 3.8b). This is an important point to be recognized in the interpretation of physiological results.

In conclusion, we can say that TE is quite linearly related with coupling strength as long as the populations work in the linear region and input noise is stable; this is no longer true if saturation is reached, or if other strong non-linear effects become influential (let us think, for instance, to synchronization in non-linear oscillators). Moreover, the estimated connectivity is strongly affected by the amount of activation that a population receives from randomised external sources. We think that this aspect has not been sufficiently investigated in previous studies using NMMs, where populations are used in linear working conditions and with stationary noise levels, and the non-linear effects on connectivity estimates are negligible. To confirm the possible disruptive effect of non-linearities on connectivity estimates, we remind the result by Wang et al. [174]: these authors observed that Granger causality and TE fail to discover the correct network topology when data are produced with strongly non-linear equations.

However, we wish to stress that in many neurocognitive problems estimation of TE may be of the greatest value even when its value is uncorrelated with the true causal connectivity: in fact, transfer of information may be more useful than measures of synaptic strength if the goal is to understand how the brain performs its computation and how one region transmits data to the other (see also Lizier and Prokopenko [183] as a nice illustration why transfer entropy may be more interesting when trying to understand a computation, compared to measures of physical causality). Indeed, the two measures are complementary, and knowledge of both may provide the best approach to the problem.

iv) *The temporal duration and time delay* – Another important indication of the present study concerns the duration of the signals necessary to achieve quite a robust estimation of TE. This is an important point, since inconsistency in the length of the selected epochs can be found in the literature, which endangers a meaningful comparison between results. In particular, previous studies have shown that connectivity estimates are affected by the epoch length and that the severity of this bias varies for different connectivity metrics [188,213,216–220]. Our results suggest that TE increases with signal duration, especially above 10 s (hence, caution should be taken when comparing experiments with different length). Conversely, the estimations with DCC are just scarcely affected by the signal length (Supplementary Material Part 2). However, for signal lengths ≥ 10s a clear linear relationship between TE and synaptic strength is evident, and an approximately linear relationship can still be detected for signal lengths of 3-4s (Fig. 3.11 b). If lower durations are used (in particular, we used 2 seconds in Fig. 3.11) TE fails to detect a clear relationship between TE and synapse strength (while DCC still offers good results): TE appears pretty high even at very low values of synaptic strength, and does not increase if the coupling term is increased. This effect of signal duration on connectivity estimation agrees with a few other results in the literature. Olejarczyk et al. [213], who used multivariate TE for the analysis of connectivity in high density resting state EEG, used temporal windows as long as 20 s. Moreover, they observed that the smaller window which still ensures the quality of results is 10 s, and that the results do not change substantially if the epoch length is increased between 10 and 40 s; these results are quite comparable to our observations in Fig. 3.11b. Fraschini et al. [220] used two different measures of connectivity, i.e., the phase lag index (PLI) and the amplitude envelope correlation (AEC). Their results show that epoch length has an important impact on connectivity estimates: both mean PLI and AEC decrease with an increase in epoch length, with a tendency to stabilize at a length of 12 s for PLI and 6 s for AEC. The sensitivity of TE to the length of time series may represent a critical issue, especially in dynamic connectivity studies. Thus, when choosing TE as a functional connectivity estimator researchers should carefully consider the nature and the dynamics of the neural processes under analysis.

A final aspect concerns the time delay between signals. This is important in neuroscience, since spike transmission along axons in long-range connections can take several milliseconds to move from pre-synaptic to post-synaptic regions. Results in Fig. 3.10 are quite unexpected, pointing out that TE estimation increases (and becomes more linear) with the time delay. This increase is evident also when using DCC. This result seems at odd with the results by Wang et al. [174] who refer that TE was quite robust against variations of signal delay. However, the results of Wang et al. consider only the capacity to detect a given network topology, without investigating the relationship between TE and connectivity strength.

It is worth noting that, in our work, time delay is unknown to the algorithm, and is estimated within a given range assigned "a priori". In particular, differences in Fig. 3.10 cannot be significantly ascribed to an error in the evaluation of the time delay, since the values obtained by the algorithm (8 ms, 15 ms and 25 ms) are not too distant from the real ones (10 ms, 16.5 and 23 ms, respectively). If confirmed by other studies, this result may indicate that connectivity between proximal regions (assuming a smaller connection delay between them) can be somewhat underestimated compared with connectivity among more distal regions.

However, it is important to stress that, in many cases, an approximate value of the delay can be inferred from neurophysiologica/anatomical considerations, and this value should be used directly in the procedure. Indeed, we used a small range of values to drive Trentool algorithm. In the DCC estimates presented in Supplementary Material Part 2, we always delayed the target signal by the number of samples closer to the true delay.

v) *Comparison between TE and DCC* – All the estimates of functional connectivity obtained with TE have been replicated using the linear Delayed Correlation Coefficient, as shown in Supplementary Material Part 2. A few conclusions can be drawn from this comparison. a) TE provides a more reliable estimation of the connection strength. This is already evident when considering two ROIs connected in feedback, as in Figs. 3.3, 3.4, 3.10 and 3.11. In these cases, TE can recognize that one synapse remains constant while the other is varying, whereas the synapse strength estimated with DCC is significantly affected by a change in the other synapse. b) TE works better than DCC in a multivariate network. This is especially evident comparing the values estimated using 4 interconnected ROIs (Fig. 3.7), both for what concerns the correlation between the values estimated by the metrics and the model synapses strengths, and the number of spurious connections estimated by the metrics. c) DCC is able to discriminate between excitatory and inhibitory connections, whereas TE provides only the absolute value of the connection. d) DCC is less affected by noise, i.e., it exhibits a smaller standard deviation on repeated trials and less evident fluctuations. However, these differences are not so strong as to overcome differences noticed at point a. e) The computational time is smaller for DCC than TE. f) Both TE and DCC exhibit a similar behaviour in response to non-linear changes, as examined in Figs. 3.8 and 3.9. In other terms, both evaluate a computational property, rather than a true causal connection.

As suggested by Reid et al.[172] each metrics exhibits alternative virtues and limitations. The use of TE, integrated with a preliminary analysis with DCC, may represent a good approach to the study of network functional connectivity. DCC may provide a first rapid screening, able to discriminate between excitatory and inhibitory links, subsequently reinforced by a more accurate and reliable analysis with TE.

3.1.7 Conclusions

In conclusion, using the open-source toolbox Trentool, and neural mass models to generate biologically realistic signals, the present study provides indications on whether brain connectivity can be assessed from bivariate TE. In particular, we not only investigated whether the presence of a statistically significant connection can be detected (as in binary 0/1 network) but also if connection strength can be quantified. Results suggest that TE can be a promising method to estimate the strength of connectivity if neural populations work in the linear regions, and if the epoch lengths are longer than 10 s. In case of multivariate networks, some spurious connections can emerge (i.e., a statistically significant TE can be detected even in the absence of a true direct connection): however, quite a good correlation between TE and synaptic strength is still preserved in these cases, even when using four interconnected ROIs. A puzzling unexpected problem is the role of time delay: estimated TE appears higher for distal regions compared with proximal regions.

Finally, as well expected, nonlinear phenomena may play a dramatic role in the assessment of connectivity, since they may significantly reduce the estimation of TE. In fact, TE is an index of information transfer and not directly an index of connectivity strength. In particular, due to non-linear relationships between the connected regions, a strong causal strength may be present between two nodes in a network, even if the detected TE is very small. We claim that similar problems can be found not only with TE but also if other metrics of connectivity (in particular those based on autoregressive models) are used, as shown when using the Delayed Correlation Coefficient. This is perhaps the most important aspect of the present work, which deserves accurate ad hoc investigation. We suggest that changes in connectivity, often reported in the literature during different tasks, or in different brain conditions, might not always reflect a true change in the connecting network, but rather a change in information transmission due to a different working region of the involved populations. However, in conditions when linearity is a good approximation of the system, changes in TE can actually reflect true changes in connectivity. Hence, researchers need to carefully consider non-linearity to apply bivariate TE. Moreover, they should check bivariate vs. multivariate TE to improve their estimation.

3.1.8 Supplementary Material 1

Table 3.1S Parameters for the configuration structure *cfgTEP* of the functions *TEprepare* and *InteractionDelayReconstruction_calculate* (TRENTOOL Version 3.3)

Field Name	Data Type	Value	Description
TEcalctype	string	‘VM_ds’	Estimator guaranteeing optimal self-prediction
predictime_u	Integer (ms)	15	Assumed information transfer delay *u* between source and target time series
predicttimemax_u	Integer (ms)	18	Maximum u to be scanned
predicttimemin_u	Integer (ms)	12	Minimum u to be scanned
predicttimestepsize	integer	1	Time steps between u’s to be scanned
ensemblemethod	string	’no’	Use of the ensemble-method for (time-resolved) TE estimation
kth_neighbors	integer	4	Number of neighbours for fixed mass search (controls balance of bias/statistical errors)
TheilerT	string	’ACT’	Number of temporal neighbours excluded to avoid serial correlations (Theiler correction)
maxlag	Integer (samples)	1000	The range of lags for computing the ACT: from -MAXLAG to MAXLAG
trialselect	string	‘no’	Sets a minimum number of trials that have to survive trial selection
actthrvalue	integer	30	Max threshold for the ACT for trial selection
optimizemethod	string	‘ragwitz’	Define method for parameter optimization: ’ragwitz’
verbosity	string	’info_minor’	Defines the verbosity of console output of TRENTOOL
ragdim	integer	4:8	For Ragwitz: range of embedding dimensions to scan vector from 1 to n
ragtaurange	double	[0.8 1.8]	For Ragwitz: 1x2-vector of min and max embedding delays
ragtausteps	integer	10	For Ragwitz: number of equidistant steps in ragtaurange with a minimum of 5
flagNei	string	‘Mass’	For Ragwitz: ’Range’ or ’Mass’ type of neighbor search
sizeNei	integer	4	For Ragwitz: Radius or mass for the neighbor search according to flagNei
repPred	integer	100	For Ragwitz: repPred represents the number of sample points for which the prediction is performed

Field Name	Data Type	Value	Description
optdimusage	string	'indivdim'	’indivdim’ to use the individual optimal dimension for each channel
dim	integer	Output TEprepare	Value(s) for embedding dimension. This is automatically taken from the field TEprepare in the data
tau	integer	Output TEprepare	Embedding delay in units of act (x*act). This is automatically taken from the field TEprepare in the data
alpha	double	0.05	Significance level for statisatical permutation test
tail	integer	1	1 tail test of significance (for the permutation tests)
surrogatetype	string	'trialshuffling'	Strategy for surrogate data creation
extracond	string	'Faes_Method '	Perform conditioning in tansfer entropy formula on additional variables. Values: ’Faes_Method’
shifttest	string	'no'	’yes’ string Perform shift test to identify instantaneous mixing between the signal pairs.
MIcalc	integer	1	Determines whether mutual information is calculated additionally to TE (1) or not (0)
shifttesttype	string	’TE>TEshift’	The shift test can be calculated for the direction TE value of original data > TE values of shifted data (value = ’TE>TEshift’)
shifttype	string	’predicttime’	Shifting the length of the ’predicttime’
numpermutation	integer	190100/500a	Nr of permutations in permutation test
permstatstype	string	’indepsamples T	Type of the test statistic used: ‘indepsamplesT’ for distribution of the t-values
correctm	string	’FDR’	Correction method used for correction of the multiple comparison problem over all analyzed channel combinations - False discovery rate ’FDR’

Table 3.2S Parameters for the configuration structure *cfgTESS* of the functions *TEsurrogatestats* and *InteractionDelayReconstruction_calculate* (TRENTOOL Version 3.3)

3.1.9 Supplementary Material 2

In this Supplementary material, we provide the results obtained with the same analysis presented in the Manuscript, but using the delayed correlation coefficient (DCC) instead of Transfer Entropy. In particular, we computed the pairwise linear correlation coefficient between the source signal and the delayed target signal. The delay was chosen equal to two sampling periods (20 ms) for all tests. Only in Fig. 3.10, we used a delay as low as 1 sampling period (10 ms) for the left panel, and 3 sampling periods (30 ms) for the right panel.

All figures have the same meaning as the figures shown in the text. The only difference is that, for clarity, we show the results ± SD (instead of ± SEM) since SEM is very small when computed on DCC estimates, and so it would be scarcely visible in the figures. We just remind that, in our tests, we have SEM = SD / $\sqrt{10}$.

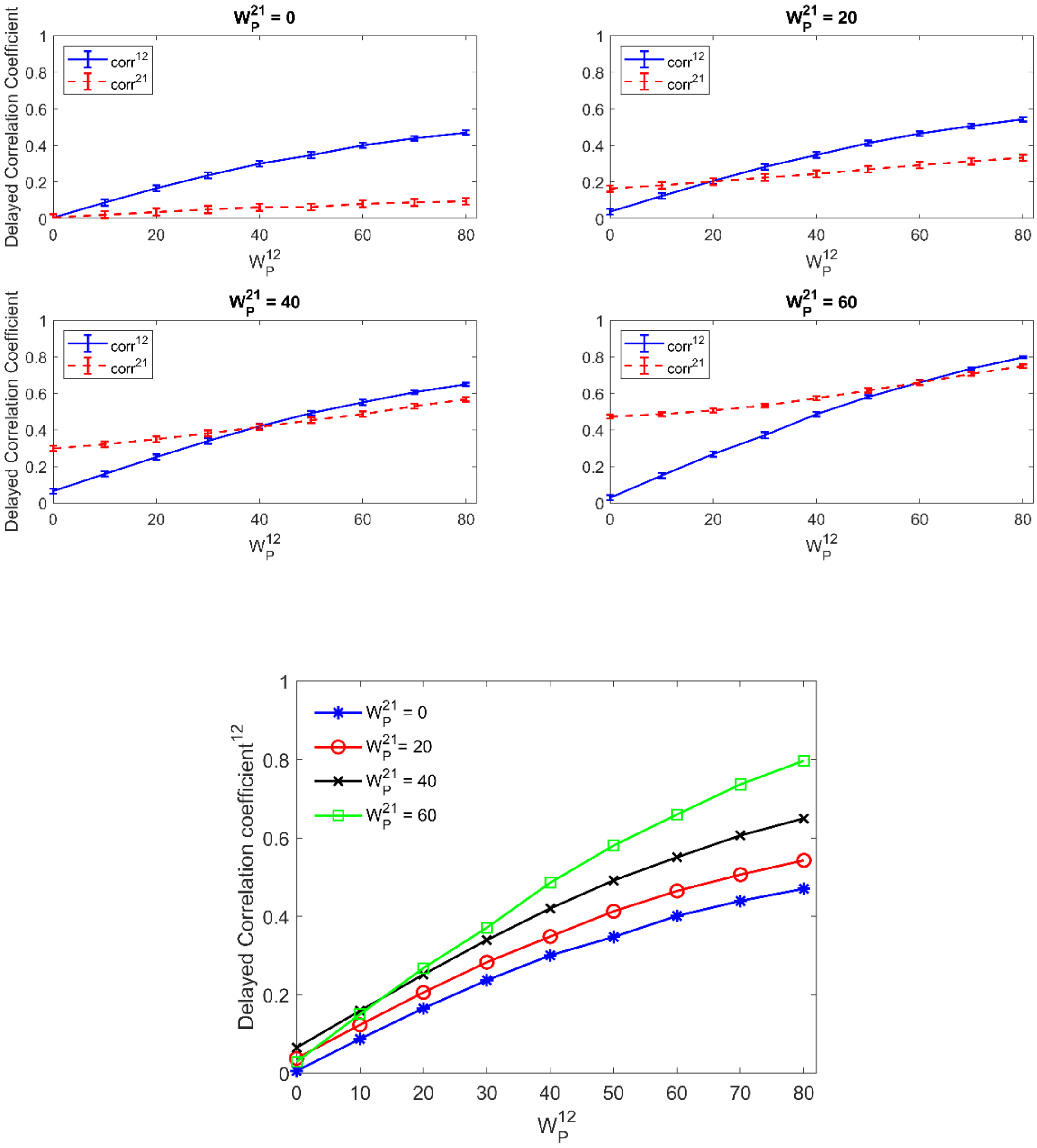


Figure 3.3S – Dependence of the delayed correlation coefficient (DCC) on feedback, realized assuming two regions interconnected with reciprocal excitatory synapses (upper panels). See text for more details.

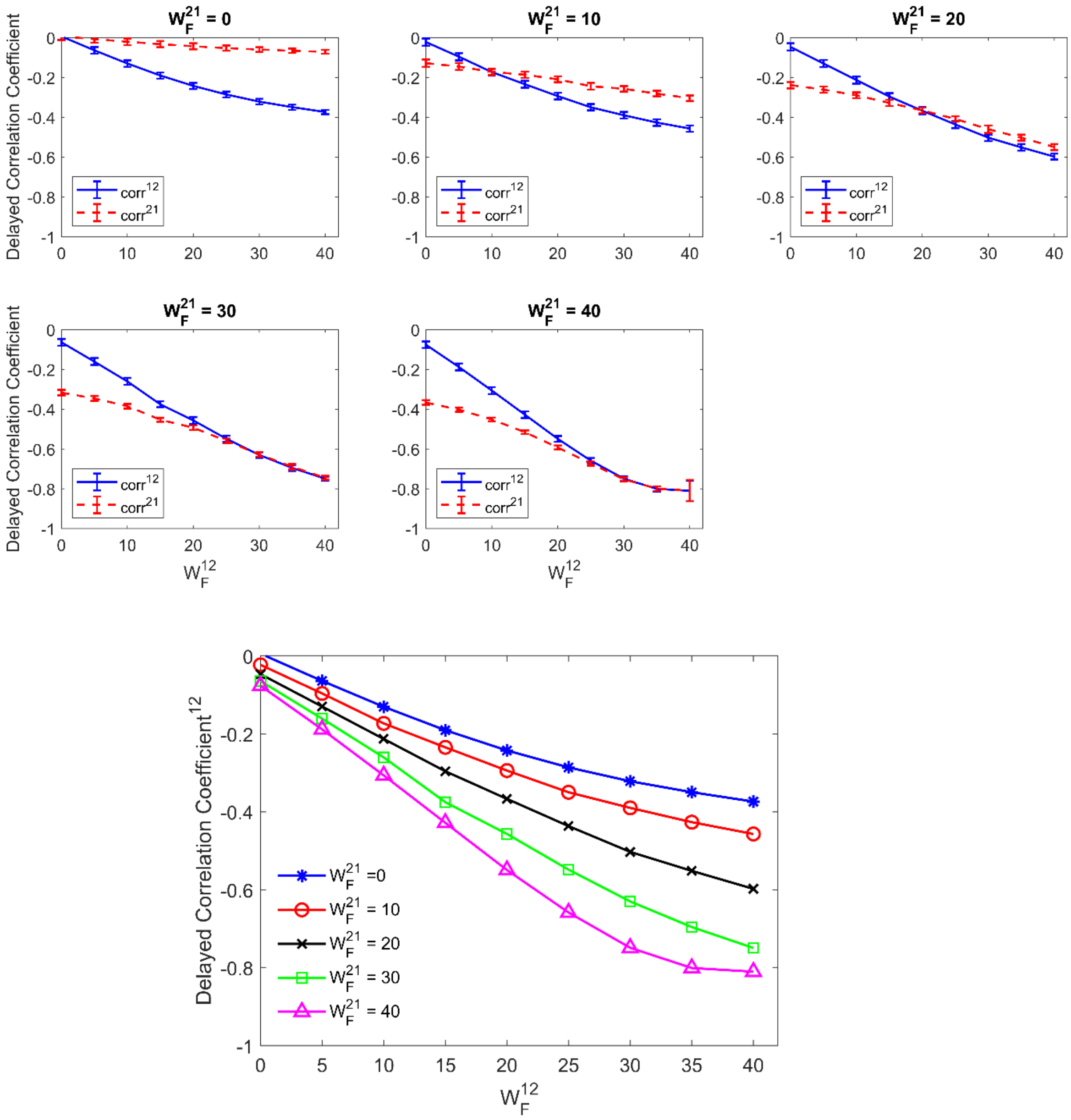


Figure 3.4S – Dependence of the delayed correlation coefficient (DCC) on feedback realized assuming two regions interconnected with reciprocal inhibitory synapses. See text for more details.

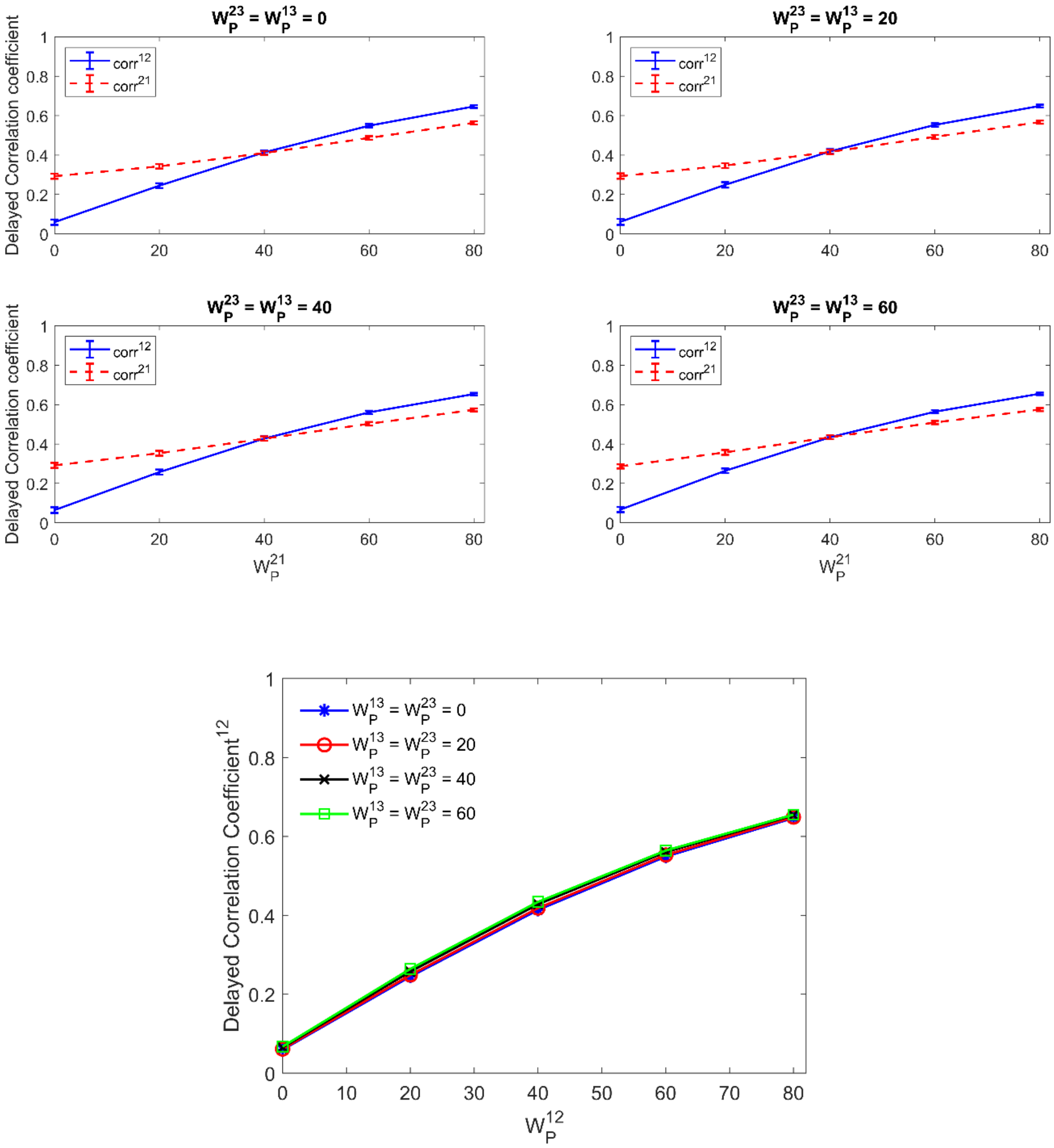


Figure 3.5S – Influence of a common external source on DCC estimation. Simulations were performed assuming three regions interconnected via an excitatory synapse from region 2 to region 1, which was progressively varied between 0 and 80, and a constant excitatory synapse in the other direction set at a constant value. See text for more details.

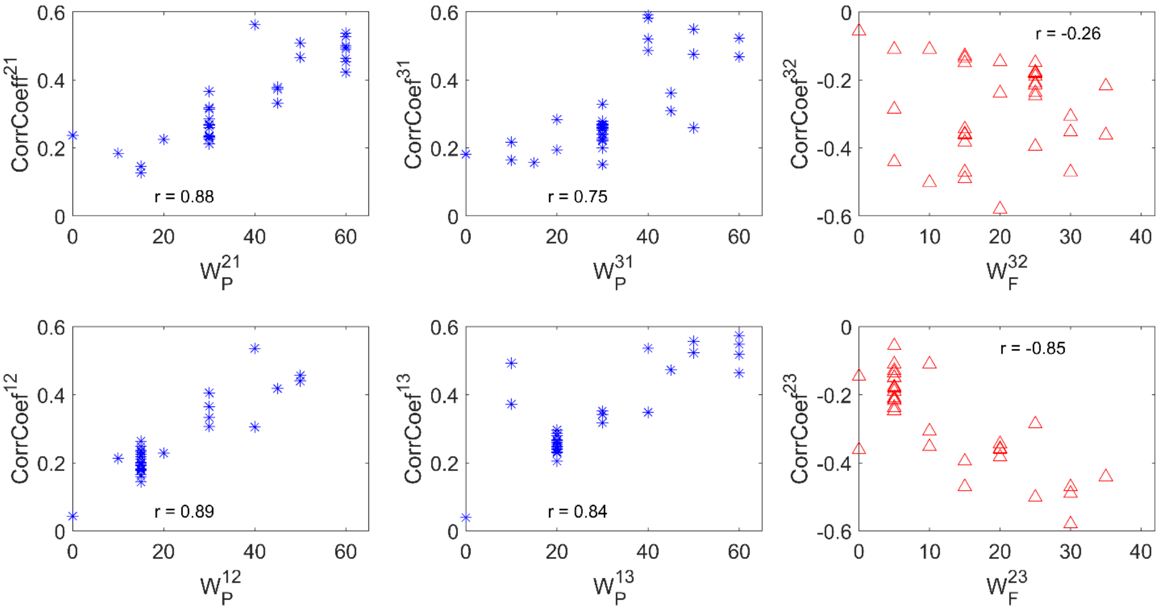


Figure 3.6S – Effect of different combinations of synapses on DCC in a model of three interconnected regions, where regions 2 and 3 are in competition via inhibitory synapses and are linked via excitatory synapses to region 1. All other synapses are set at zero. See text for more details.

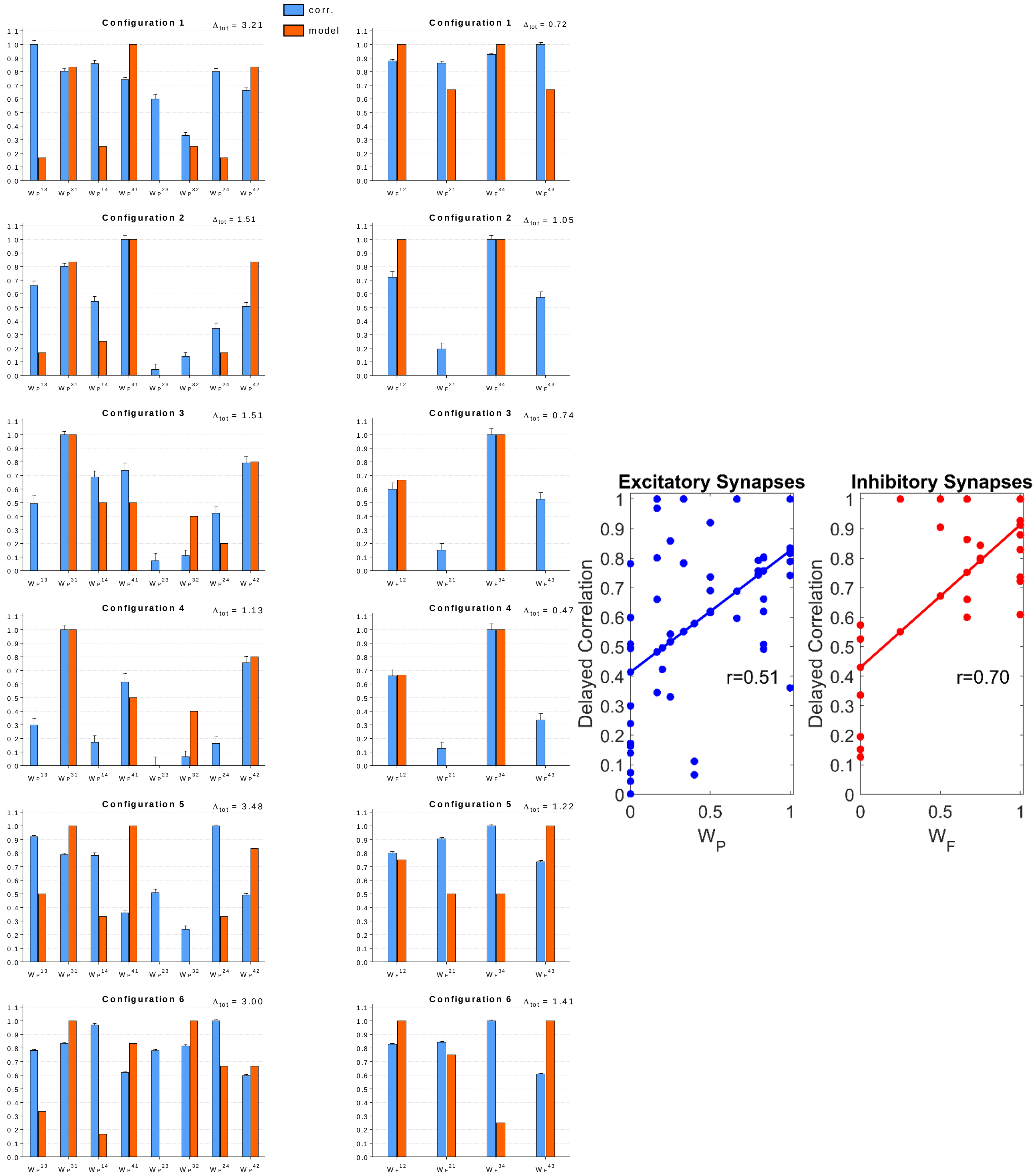


Figure 3.7S – Estimation of the connectivity strength with DCC obtained during six different simulations, each performed with four interconnected ROIs. Each row refers to a different network configuration. See text for details.

Figure 3.8S

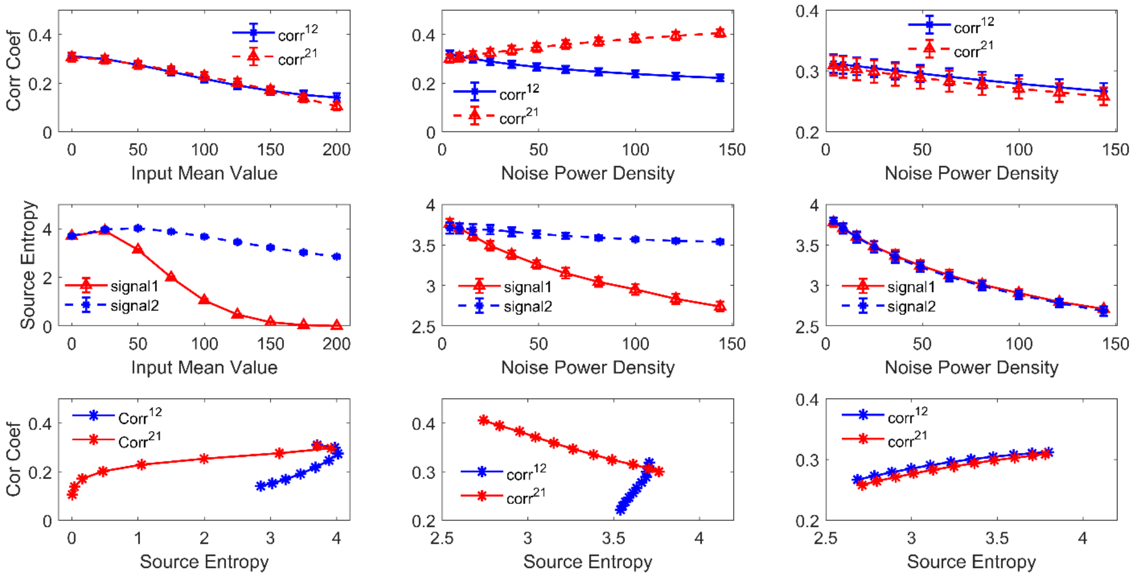


Figure 3.8S – Effect of the mean value and standard deviation of the input noise on the estimation of DCC. The simulations were performed using two regions connected with excitatory synapses and by varying the mean value m1 (left panels) and standard deviation σ1 of noise (middle panels) of the input to ROI1. Finally, the right panels show the case when noise standard deviation was increased in both populations altogether (both parameters σ1.and σ2). The first row shows DCC ± SD, while the second row shows entropy (± SD) of the two signals, vs. the input values. Finally, the third-row plots DCC vs. the entropy of the source signal. See text for more details.

Figure 3.9S

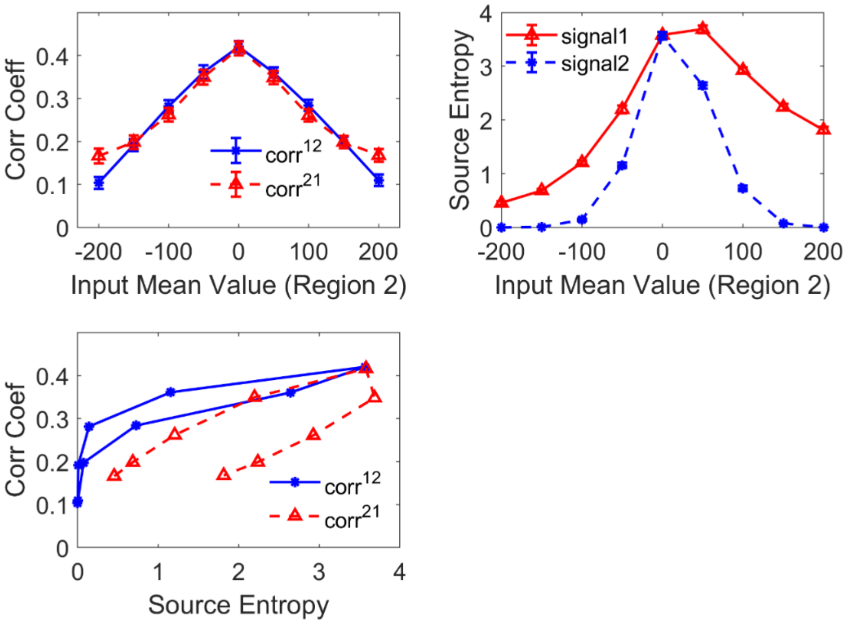


Figure 3.9S – Effect of the region working point on the estimation of DCC. The simulations were performed using two regions connected with excitatory synapses and by varying the input mean value to region 2. The meaning of the plots is the same as in Fig. 8S. See also the text for more details.

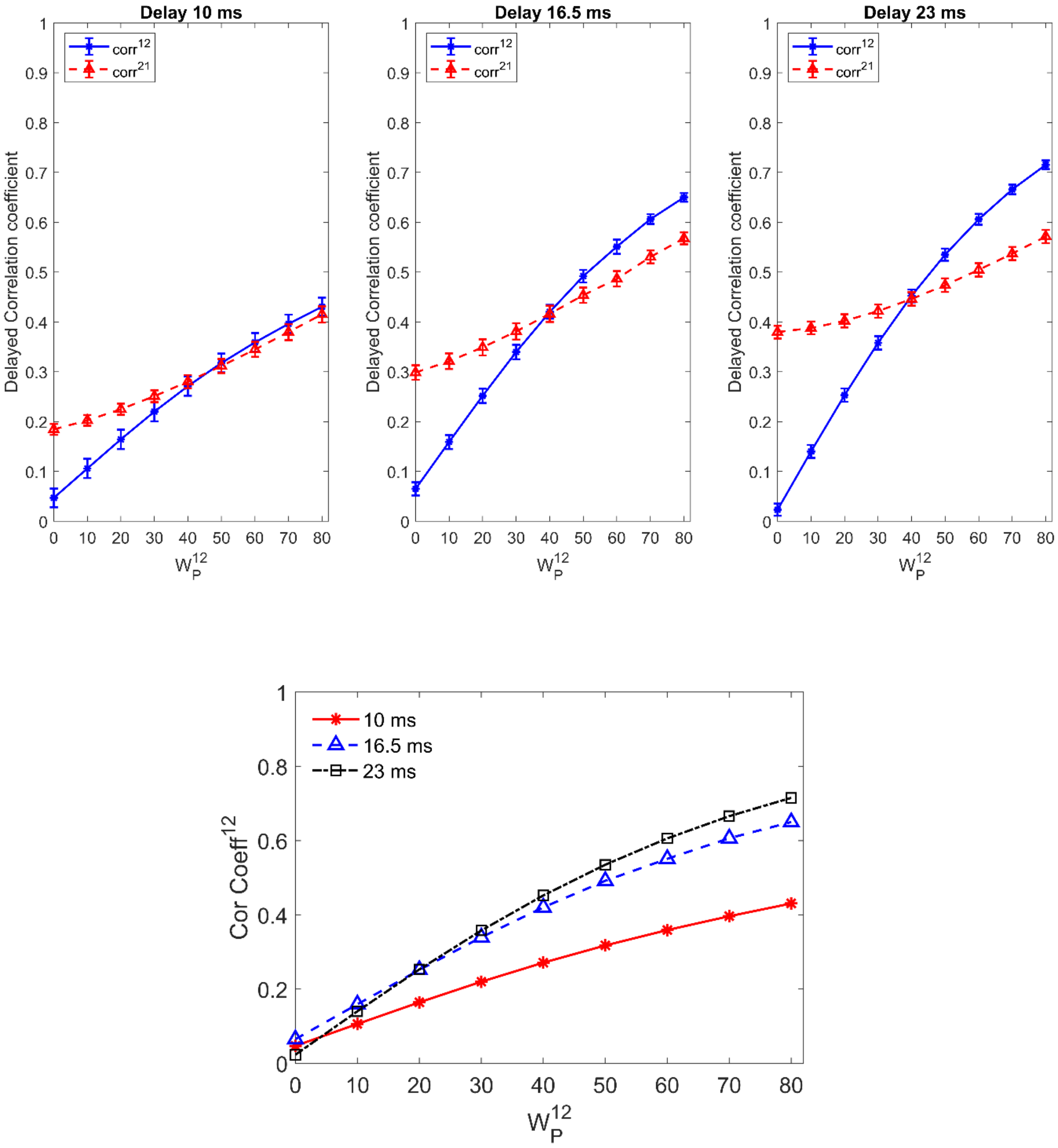


Figure 3.10S – Effect of the delay between the two regions on the estimation of DCC. The simulations were performed using two regions connected with excitatory synapses. The simulations were repeated with different delays. See the text for more details.

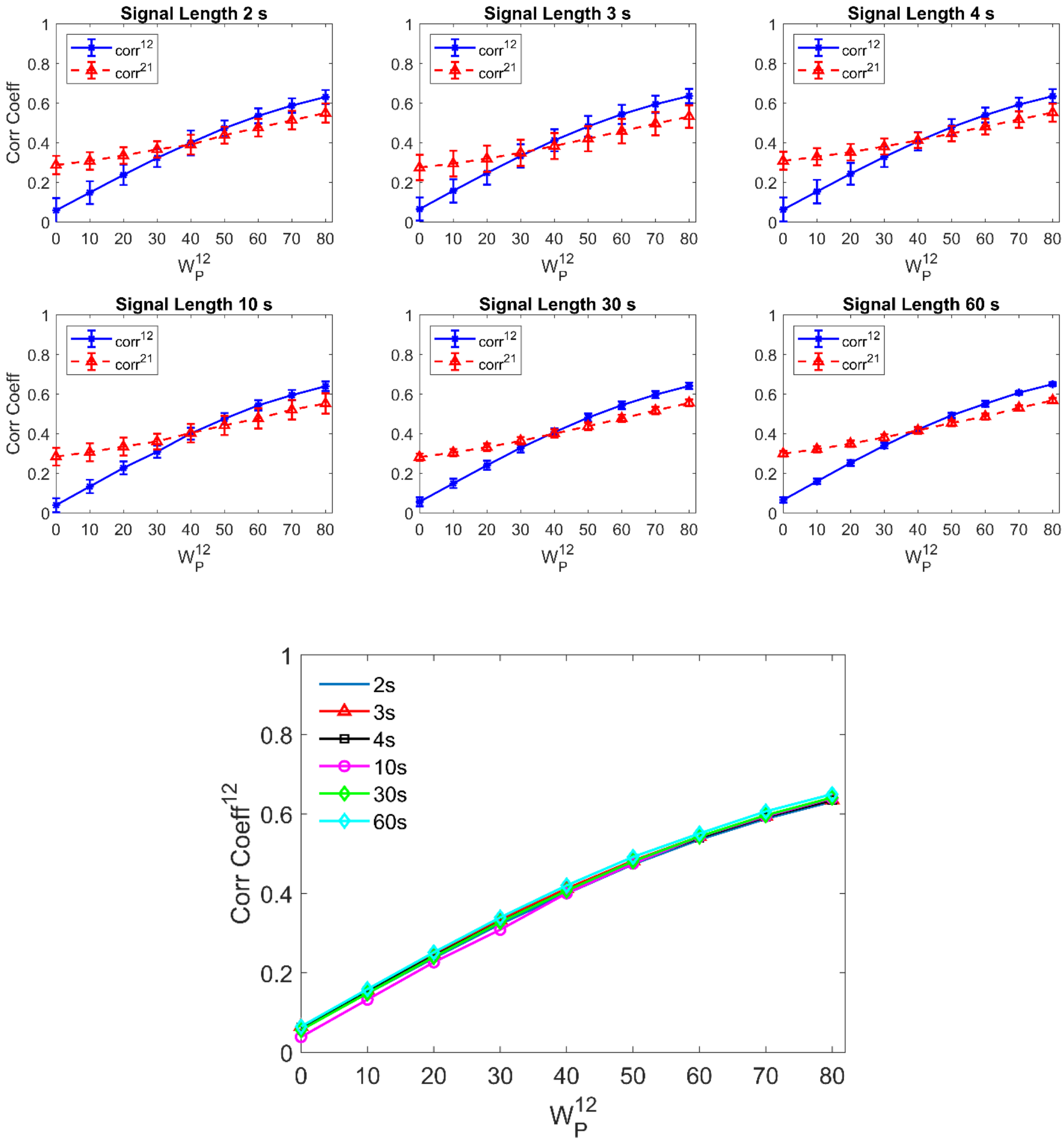


Figure 3.11S – Effect of the duration of the signal on the estimation of DCC. The simulations were performed using two regions connected with excitatory synapses. The simulations were repeated with different durations of the signals. See the text for more details.

3.2 Brain Rhythms Transmission and Functional connectivity estimation with NMMs

The study reported in this chapter refers to the published journal paper entitled "The Relationship between Oscillations in Brain Regions and Functional Connectivity: A Critical Analysis with the Aid of Neural Mass Models", Giulia Ricci, Elisa Magosso, Mauro Ursino*, *Brain Sciences* (2021).

In this section, NMMs were used to simulate different brain rhythms (theta, alpha, beta and gamma frequency bands) and to assess through FC estimators how these are transmitted in a physiologically plausible network. We also tested the reliability of eight different FC estimators: Pearson correlation coefficient, Delayed correlation coefficient, Coherence, Lagged Coherence, Phase Synchronization, Time-Domain Granger Causality, Frequency-domain (spectral) Granger Causality and Transfer Entropy. Since Granger Causality outperformed the other estimators, it was analysed in greater detail. Specifically, the relationship between Granger Causality and true connectivity strength was assessed in both linear and nonlinear conditions. The results highlight that changes in functional connectivity do not always reflect a physical change but rather a change in information transmission.

Background: *Propagation of brain rhythms among cortical regions is a relevant aspect of cognitive neuroscience, often investigated using functional connectivity (FC) estimation techniques. Aim of this work is to assess the relationship between rhythm propagation, FC and brain functioning using data generated from neural mass models of connected Regions of Interest (ROIs).* ***Method***: *We simulated networks of four interconnected ROIs, each with a different intrinsic rhythm (in theta, alpha, beta and gamma ranges). Connectivity was estimated using eight estimators and the relationship between structural connectivity and FC was assessed as a function of the connectivity strength and of the inputs to the ROIs.* ***Results:*** *show that Granger estimation provides the best accuracy, with a good capacity to evaluate the connectivity strength. However, the estimated values strongly depend on the input to the ROIs, hence on non-linear phenomena. When a population works in the linear region, its capacity to transmit a rhythm increases drastically. Conversely, when it saturates, oscillatory activity becomes strongly affected by rhythms incoming from other regions.* ***Discussion***: *Changes in functional connectivity do not always reflect a physical change in the synapses. A unique connectivity network can propagate rhythms in very different ways, depending on the specific working conditions.*

3.2.1 Introduction

Brain functioning depends on the interaction among different regions, which exchange information via a complex connectivity network and work together in a coordinated manner to realize cognitive tasks. Accordingly, the study of brain connectivity has been receiving increasing attention in cognitive neuroscience, as documented by the large amount of research in the field (see, among the others, [169,172,174,221]). Indeed, there is large consensus that connectivity is a primary mean for understanding brain function at different levels of organization. In fact, connectivity analysis has been assessed noninvasively in several recent studies, both starting from data obtained with magnetic resonance [199,222,223] or via neuroelectric imaging techniques (MEG or EEG)[224–226]; this analysis is of great value to encompass the relationships among the different areas involved, to unmask their specific role, and, ultimately, to understand how these interactions produce cognition in a coordinated fashion. In humans, invasive connectivity studies can also be performed with electrocorticography (ECoG) (i.e., placing electrodes at the brain surface [227]), for instance, in the presurgical evaluation of epilepsy or during deep-brain stimulation therapy. The significance of brain connectivity estimates, however, is still the subject of large debate in the literature. A traditional distinction (which, however, has been questioned recently; see [172]) discerns between functional connectivity (FC) and effective connectivity estimates. According to a traditional point of view, the first assesses "the statistical dependence or mutual information between two neuronal systems" [169], whereas the second represents the causal influence that one neural system exerts on another, based on an explicit model describing the underlying process. Both, however, are different from the structural connectivity, defined as the presence of a physical connection among the regions.

The previous definitions suffer from a variety of problems and are often misunderstood. In the following, we will especially refer to methods for the estimation of functional connectivity, but within the wider point of view proposed by Reid et al. [172] recently. These authors clearly underlined that, although methods for FC estimation are based on the computation of some forms of statistical or information association between signals, the target is always to understand the causal interactions among the different neural populations. Hence, the ultimate objective is to construct a causal network in order to comprehend how populations interact causally to produce cognition, and how alterations in these connections affect behavior (for instance, in pathological states). In other terms, the objective is to gain a mechanistic insight into brain functioning via a distributed process of structurally connected neural groups, although these networks are derived from associations among signals.

As underlined by Reid et al. [172], within this larger framework an essential role is now played by methods to validate FC estimation techniques against a series of ground-truth conditions. A typical way to implement validation is through the use of neurocomputational models inspired by brain functioning. Typically, simulated signals can be used to test whether methods for FC estimation are able to identify the structural connectivity imposed on the model, and, via a sensitivity analysis, to detect changes induced by parameter manipulation in the theoretical network.

Indeed, several such studies have been published in the last two decades, providing important confirmations but also widening the debate. Within these studies, it is possible to identify two main categories of employed models: those which simulate individual spiking neurons and ionic channels, mainly oriented to the study of connectivity in neuron cultures, and more abstract models, which employ neurons with continuous outputs and are more concerned with the study of connectivity among entire brain regions.

In the following, we will focus the attention on a particular kind of abstract model—Neural Mass Models (NMMs) (for a summary on some connectivity studies with spiking neurons, see Ursino et al., 2020 [137], and also [185]). Rather than modeling all individual neurons within a brain circuit, NMMs simulate averaged activity generated by a population of similar neurons.

We decided to focus on NMMs since they represent a good compromise between accuracy and simplicity. They allow oscillatory phenomena to be described in a clear mechanistic way (emphasizing, for instance, the role of the different subpopulations involved) whilst still maintaining a limited number of state variables. It is easier to understand the behavior of such models compared with detailed models with spiking neurons, and it is easier to perform a sensitivity analysis across a wide range of conditions (as carried out in the present work). Parameters have a more general meaning, and so it is easier to generalize the validation results over different conditions. Of course, these models also have some limits. They cannot be used to incorporate results taken from individual neurons (for instance, to simulate the effect of ionic channels or drugs as measured directly on individual neurons) and their dynamics may exhibit some differences compared with the exact dynamics resulting from large populations of spiking neurons.

David et al. [188] tested the capacity of some FC estimators (cross-correlation, mutual information, and synchronization indices) to detect changes in neural coupling in a symmetric configuration of two NMMs: each measure was found to be sensitive to variations in neuronal coupling, with a monotonic dependence between the functional connectivity measures and the coupling parameter. Ansari-Asl et al. [189] and Wendling et al. [128] employed various NMMs (but with only two populations each) connected with an excitatory coupling parameter and explored the relationship between the coupling parameter and various FC estimates. They suggested that there are no ideal methods and that it is strongly advised to compare the outcomes from different connectivity estimates. Wang et al. [174] performed a systematic analysis on the performance of 42 different FC estimators against five different generations models (including convolution NMMs) with a five-node connectivity structure. Their results suggest that, when using signals generated with NMMs, Granger causality and Transfer Entropy (TE) show the ability to retrieve the underlying model structure quite well. However, difficulty can be encountered when the generative models include stronger nonlinearities (such as Rossler and Henon equations). However, in the study by Wang et al., only the classification accuracy was tested, using Receiver Operating Characteristic (ROC) curves, without an analysis on the relationship between the connectivity strength and the FC metrics.

All previous studies provide important indications on the virtues and limitations of several FC estimators, but also exhibit several important limitations. First, they did not test networks whilst including inhibitory couplings. Nevertheless, inhibition plays an important role in brain functioning, both to avoid excessive uncontrolled excitation spreading in the brain, and to implement competition mechanisms among different brain regions. Second, the previous studies did not test the effect of nonlinear behavior carefully. We claim that the estimate of a connectivity network via FC methods is strongly affected by the specific functioning of the different neural units involved—above all by their working point in the nonlinear neuron characteristics. Third, brain rhythms (in the α, β, γ and θ frequency ranges) are known to play a fundamental and specific role in many cognitive tasks (such as working memory, episodic memory, internal and external attention, perceptual grouping [228–231]). In particular, the idea that different brain regions are characterized by different intrinsic rhythms, and that these rhythms are transmitted from one region to another to produce a sophisticate "system of rhythms", subserving various cognitive functions, has received important support in recent papers [232–237] and plays a fundamental role in neuroscience (see the final section of this work for a discussion on this aspect). Unfortunately, the problem of how these rhythms can propagate within a structural connectivity network and modulate their power has not been investigated in previous FC studies, despite its enormous relevance for the present cognitive neuroscience research.

In a recent work [137], we investigated the capacity of an important FC estimator (the bivariate transfer entropy [135,181]) to detect changes in connectivity using signals generated by NMMs of interconnected ROIs (with two to four coupled regions). The connectivity network included excitatory and inhibitory links, and simulations were performed both in linear conditions and altering the working point of the individual regions. We found that TE can consistently estimate the strength of connectivity if neural populations work in their linear regions, and if the signal lengths are longer than 10 s. However, nonlinear phenomena strongly alter the TE estimation results; indeed, TE describes the amount of information transferred from one region to another, which is different from a true causal relationship. In that work, however, all regions were characterized by populations with identical internal parameters, producing a similar rhythm typically in the β band (which is fundamental to study activity in supplementary motor-premotor-primary motor cortical areas [238]). Hence, the previous study did not address the problem of how different rhythms (in different frequency bands) can be transmitted within a causal network and whether this rhythm propagation can be detected via common FC estimation metrics.

The aim of the present paper is to significantly improve the previous research by analyzing the capacity of different and frequently used FC metrics to evaluate rhythm propagation and causal connectivity in a network of four interconnected ROIs, assuming that each ROI is characterized by a specific rhythm (in the bands θ, α, β or γ, respectively). Specifically, we tested eight different bivariate FC metrics, either nondirected (correlation, phase synchrony, coherence, lagged coherence) or directed (delayed correlation, temporal Granger causality, frequency Granger causality, transfer entropy). Moreover, while five of the previous metrics provide just a single value, three of them (coherence, lagged coherence and frequency Granger causality) provide a frequency-dependent estimation, allowing a discrimination among different frequency bands.

The paper is structured as follows. First, the NMM is qualitatively described, assigning parameters to each ROI to simulate the four different rhythms. Then, the eight methods used to assess FC are briefly described. The performances of these metrics are compared using 100 networks of connectivity among the four ROIs, generated randomly and including excitatory and inhibitory connections. Finally, a more complete analysis is performed using signals generated through a physiologically inspired connectivity network, simulating the interaction among rhythms in the occipital, parietal and frontal regions. In the main text, this analysis is performed using the Granger estimators, and concerns both changes in connectivity strength and alterations in the network working point. Results of additional simulations (performed with all estimators and on a different network) are presented in the Supplementary Material, together with NMM mathematical equations.

3.2.2 Method

Please note that the NMMs used is the one described at the beginning of section 3 represented in Figure 3.1.

As illustrated in Fig. 3.1, each ROI, in addition to receiving long-range synapses from other ROIs, may receive inputs from the external environment (I_p and I_f, entering into the pyramidal population and fast inhibitory population, respectively) and superimposed Gaussian white noise. Here, we assumed that only the external input I_p to the pyramidal population is non-null in the four simulated ROIs, while I_f is kept at 0 in each ROI.

However, it is worth noting that in our model the random noise is not applied directly to the LFP of pyramidal neurons (i.e., the quantity v_p in Equation 3.4), but it acts on pyramidal neurons through the typical low-pass dynamics of the glutamatergic synapses (this is described in Equations (3.5) and (3.6) of the *Equations 3.1 - 3.18 of section 3*). Hence, the effective noise on v_p exhibits a low-pass shape that resembles the true 1/f noise of LFPs (see Fig. 3.12S in the *Supplementary Material 2*). Of course, it may be of value, in future work, to directly use realistic noise extracted from real LFP recordings, which may contain other frequencies coming from different inputs. This may significantly contribute to making the model spectra more reliable.

All model equations can be found in the *Equations 3.1 - 3.18 of section 3*.

In the present work, to assess the performance of the eight different FC estimators, we first generated 100 random networks connecting the four ROIs, with the same external input (I_p =400) to each ROI. We assumed that each network can have a number of long-range synapses ranging between 3 and 9 (randomly chosen, the others are set a zero) with connection strengths as great as 10, 20, 30, and 40 (randomly chosen). Each connection can be either excitatory or inhibitory, with 50% probability of each.

Subsequently, we tested the FC estimators on the two connectivity networks depicted in Fig. 3.12 by varying both the connection strength and the inputs to the ROIs (the latter changes have been performed to test nonlinear effects). Only results on the first network (in Fig. 3.12A) are illustrated in the main text. All other results are summarized in the *Supplementary Material 2*.

Most of the results shown in the present work consider networks with a similar number of excitatory and inhibitory (bisynaptic) connections. Indeed, several authors stress that the excitatory and inhibitory ensembles are well balanced in the human cortex and that a breakdown of this balance may contribute to several pathological states, such as epilepsy [239]. However, we also tested other networks similar to those used in this study, but with all pyramidal–pyramidal synapses; the main results did not change meaningfully.

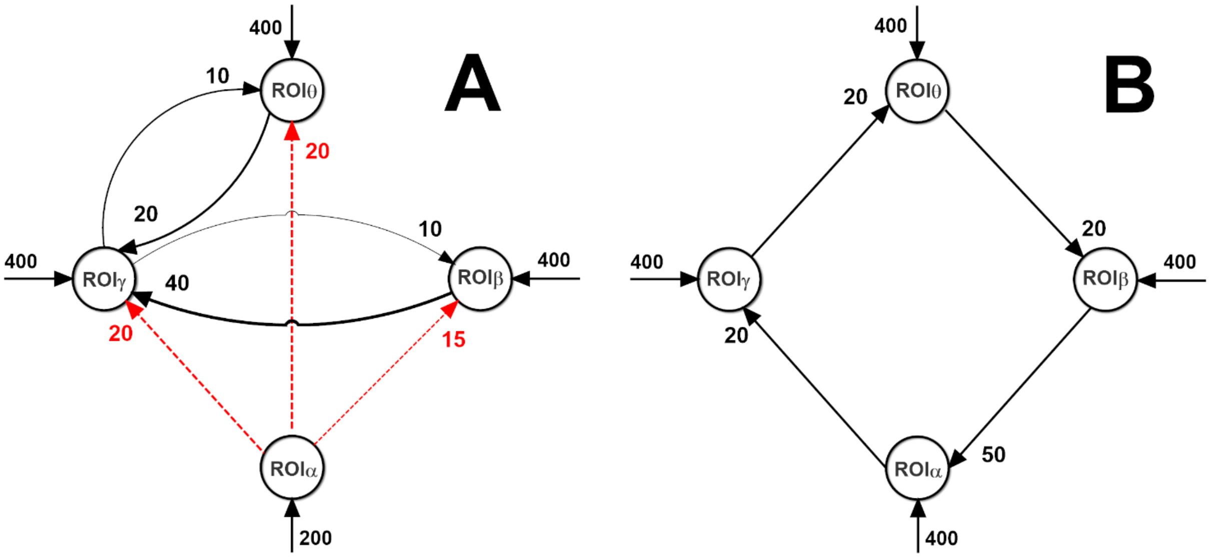


Figure 3.12 Connectivity networks among the four ROIs used in the present work (dashed red lines and continuous black lines denote inhibitory and excitatory connections, respectively). The network in Fig. 3.12**A** simulates a possible physiological connectivity from occipital (or thalamic) regions to motor regions and temporal/frontal regions. This network, by varying the strength of a single connection or the input value to a single ROI, was used to obtain the results shown in Figures 3.16–3.23 using the functional connectivity (FC) metrics based on Granger causality (but see also Fig. 3.15S in Supplementary Material part 2 for the other FC metrics). The simple loop network in Fig. 3.12**B** allows a straightforward analysis of rhythm propagation in a chain of interconnected ROIs; the corresponding results are reported in Supplementary Material part 2 (Figures 3.16S-3.17S).

For each generated network, and for each combination of inputs to the network, ten different simulations were performed using random noise superimposed on the inputs (normal distribution with zero mean value and SD = $\sqrt{5/dt}$, where *dt* is the simulation step). Any group of ten simulations was repeated starting from the same seed to be sure that differences can be ascribed only to the network connections and to the input values, rather than to the particular noise. In the following, the result of each FC estimate is the average of the values obtained in the ten trials.

The set of differential equations (see Supplementary Material part 1) was numerically integrated with the Euler method, with an integration step as low as 10^{-4} s. The duration of each simulation was 11 s, but the first second was excluded from the subsequent computations to avoid the confounding effects of the initial transient phenomena. As discussed in Ursino et al. [137], a 10 s length for the signals ensures a good reliability of the estimates.

The FC estimates were performed using the simulated membrane potentials of pyramidal neurons (quantity v_p, see Equations 3.1 - 3.18 of Chapter 3), as a good approximation of EEG or of mean field potentials for each ROI. To reduce the computational cost, simulated signals were resampled at 100 Hz after low pass-filtering with an antialiasing zero-phase filter (cut-off frequency 50 Hz).

3.2.2.1 Functional Connectivity Estimates

We used eight different bivariate methods to estimate FC: Pearson correlation coefficient, Delayed correlation coefficient, Coherence, Lagged Coherence, Phase Synchronization, Time-Domain Granger Causality, Frequency-domain (spectral) Granger Causality and Transfer Entropy. The mathematical formulation of these estimators is given in Chapter 2 and Section 2.3.1.1.

It is worth noting that in the case of Delayed correlation coefficient, *d* was chosen as the value that maximizes the absolute value in Eq. (2.21). However, the correlation coefficient can assume a positive or negative value. Hence, we used the absolute value to choose the value of *d* for a given connection, and to compute true and false positives in ROC curves. Conversely, we maintained the sign (positive or negative) in other figures to investigate whether this metrics can detect the presence of excitatory or inhibitory connections.

Moreover, all power densities of Coherence formulation (see Eq. 2.22) were computed using the Welch periodogram method [240], with a 0.5 s window (50 samples) and 10 s zero padding (1000 samples) to ensure a spectral resolution as sharp as 0.1 Hz.

3.2.3 Results

3.2.3.1 Power Density Spectra of the Different Regions

First, we assigned parameters to each ROI so that any region can produce an intrinsic rhythm in a different frequency band when excited in the central region of its sigmoidal relationship. Each ROI has been named considering its intrinsic rhythm. The considered bands are θ: 4–8 Hz (ROIθ), α: 8–13 Hz (ROIα), β: 13–26 Hz (ROIβ) and γ: 26–40 Hz (ROIγ). To this aim, we manually modified the parameters representing the synaptic contacts among the populations (C_{ij}) and the reciprocal of time constants (ω_j). All parameter values can be found in *Equations 3.1 - 3.18 of section 3*.

The power spectral densities (PSDs) of all ROIs and the temporal patterns of membrane potentials and spike density for pyramidal neurons, simulated with the connectivity as in Fig. 3.12a, are illustrated in Fig. 3.13 (see also Fig. 3.13S in Supplementary Material 2, where spectrograms are shown).

The temporal patterns show that the rhythms in the model exhibit a complex waveform—i.e., they are not sinusoidal (as the patterns obtained, for instance, with the use of Wilson

Cowan oscillators) and are characterized by intermittent fluctuations. As pointed out by Cole and Voytek [241] and by Jones [242], the use of complex intermittent waveforms is essential to reach a full comprehension of the role and meaning of brain rhythms

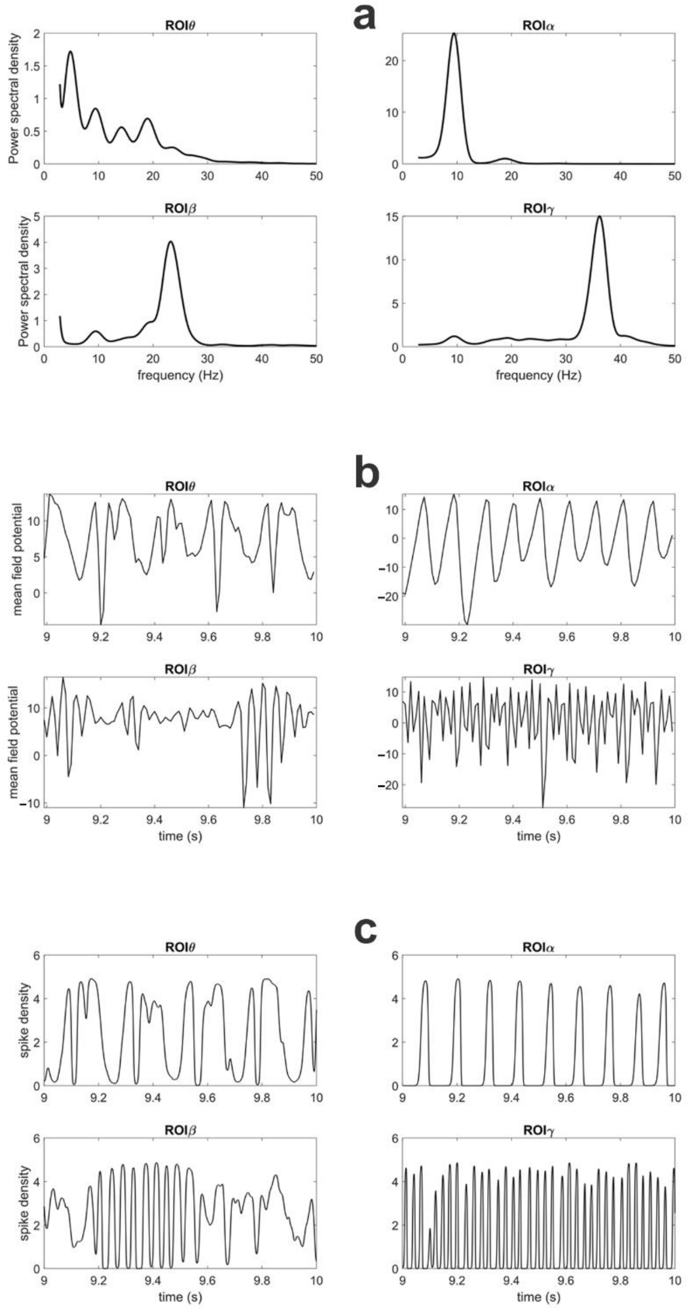


Figure 3.13. Power spectral densities of potential for the pyramidal population in the four different ROIs (panel **a**) simulated with the connectivity as in Fig. 3.12A and all inputs as great as 400. Power spectral densities (PSDs) have been obtained by computing the Welch periodogram on membrane potentials of pyramidal neurons. Parameters for each ROI and noise levels are shown in *Table 3.3S* of *Supplementary Material part 1*. Panels **b** and **c** represent the temporal pattern of membrane potential of pyramidal neurons and their spike densities, respectively, during the last second of the simulation.

3.2.3.2 Analysis of the Different Metrics with Random Network Connectivity

In order to compare the performance of the different estimators, we generated 100 different random networks connecting the four ROIs (with a number of connections ranging between 3 and 9, synaptic weights of 10, 20, 30 or 40, equal probability of excitatory or inhibitory connections, and external input I_p = 400 for each ROI; see Methods section). The performances were assessed by matching the connectivity network obtained via the estimator with the true connectivity network and quantifying this matching via ROC curves and precision–recall curves. This was performed on a binary basis (yes/no connection) by comparing the value estimated by each metrics with a threshold (range 0–0.5) and evaluating the percentage of true positives vs. the percentage of false positives at each threshold. The results are summarized in Fig. 3.14. Finally, Table 3.2 reports the areas under the curves (AUCs) computed starting from the ROCs.

Results show that the Granger estimators (both in the temporal and frequency domains) provide a more reliable description of the connectivity network, with values of the AUCs as high as 0.88. TE and coherence also provide fair results (AUC = 0.77 − 0.78), whereas the performance of the other estimators is lower.

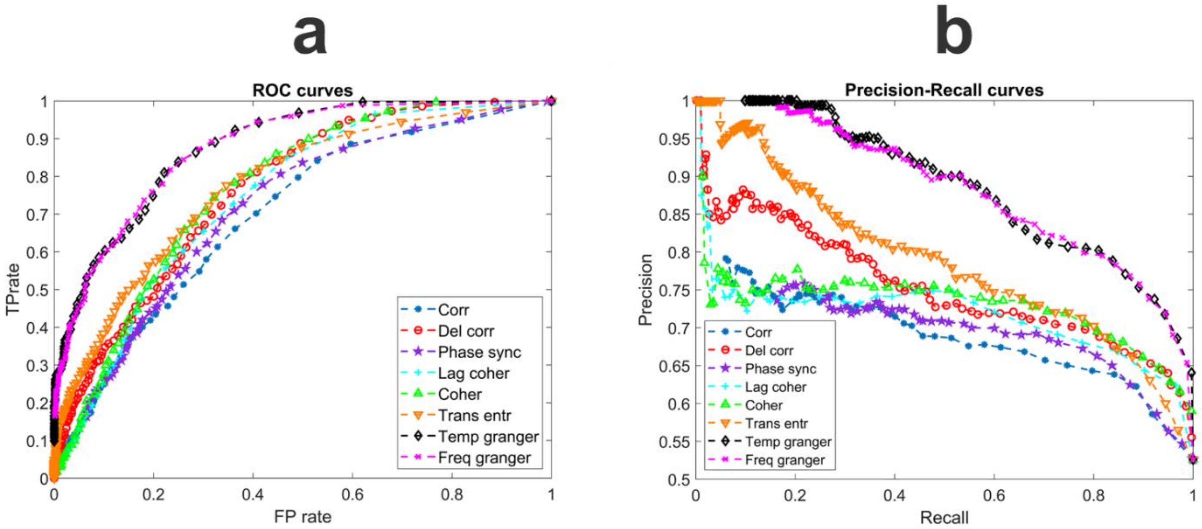


Figure 3.14 - ROC curves (panel **a**) and precision–recall curves (panel **b**) obtained with the different FC estimators using the data obtained from 100 randomly generated connectivity networks among the four ROIs. The area under the ROC curve for each estimator is reported in Table 3.2.

Table 3.2 The different FC estimators.

FC Estimator	AUC
Correlation	0.6987
Delayed Correlation	0.7580
Phase Synchronization	0.7100
Lagged Coherence	0.7465
Coherence	0.7673
Transfer Entropy	0.7753
Temporal Granger	0.8787
Spectral Granger	0.8759

The precision–recall curves prove that our results are valid independently of a possible unbalance in the data set and further confirm that Granger and TE provide better results than the other estimators in the context of the present work. In particular, the precision (i.e., the capacity to reveal a connection only when it really exists) may be quite high with Granger and TE estimators using a high threshold; the delayed correlation exhibits a good precision too. Furthermore, even if the threshold is reduced, the Granger estimates maintain a good precision (higher than 80%), still maintaining a recall as high as 80% (i.e., by limiting the number of false negatives).

The previous analysis, however, is quite limited, since it just compares the topology of the true and estimated connectivity networks in terms of existence/nonexistence of a given connection. A more complete analysis requires the assessment of the connectivity weight and, even more importantly, the role of input changes and nonlinearity. This analysis will be shown in the next subsections, which use the Granger estimators, first on the random net and then on the network in Fig. 3.12A. Results obtained with different estimators and on the net in Fig. 3.12B are shown in *Supplementary Material part 2*.

3.2.3.3 Analysis of the Connectivity Strength

In order to evaluate the capacity of the Granger estimator to detect changes in connectivity strength, we re-examined the results obtained with the 100 random networks generated above, plotting the connection values estimated with the temporal Granger estimator (mean ± SD) vs. the true value (Fig. 3.15). All estimated values show a monotonic dependence on the true value, demonstrating that the estimator is able to catch the alterations in the connectivity strength even in a multivariate condition—i.e., when many connections vary together. On average, the estimator is able to distinguish the absence of a connection (0 value) from a moderate (10–20) or a higher (30–40) connection. However, the SD is quite high, especially at the highest values of the connection strength, indicating that a single estimated value is subject to large variability.

Moreover, from Fig. 3.15 one can observe that the connection $\beta \rightarrow \gamma$ can produce higher values of FC than the other connections. Our impression is that the γ oscillations can be easily modulated by an incoming β rhythm, and this makes the connection $\beta \rightarrow \gamma$ particularly effective in influencing the dynamics of the target population.

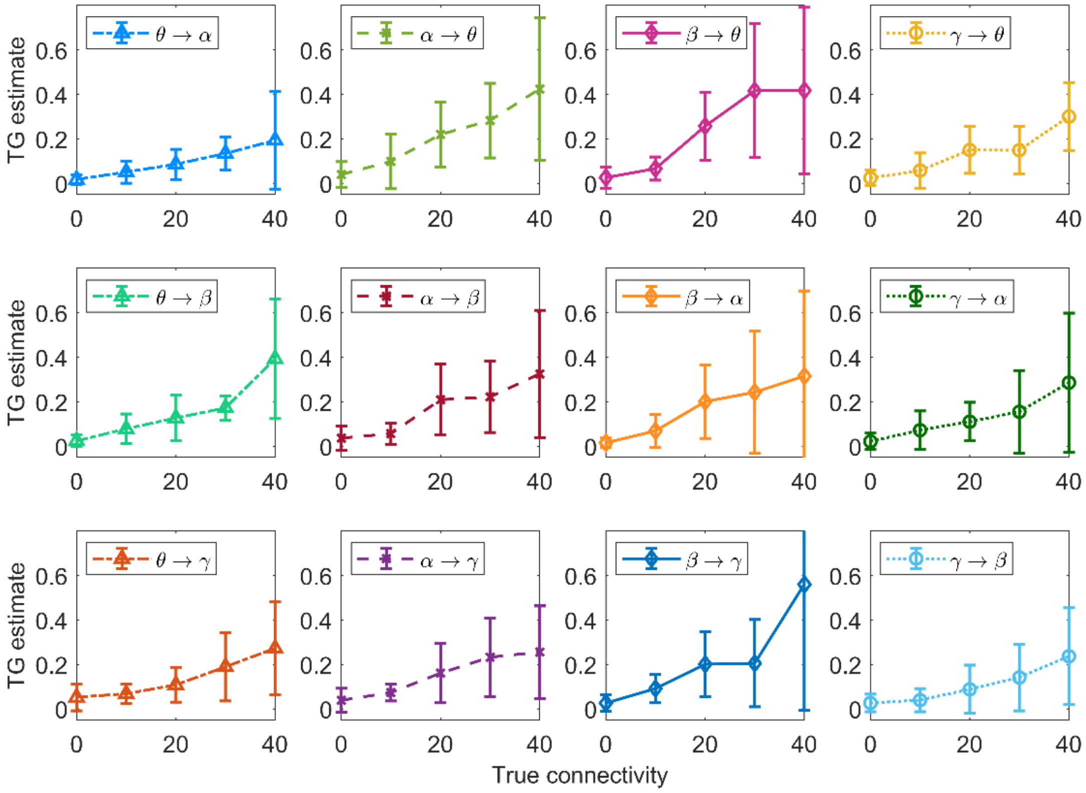


Figure 3.15. Relationship between the connectivity estimated with the temporal Granger causality and the true connectivity, obtained using the data from 100 randomly generated connectivity networks. It is worth noting that, in these nets, connections were randomly generated between 0 and 40 (step 10). Points are mean values at each connection strength, and bars denote standard deviations. The estimator is able to grasp the monotonic increase in connectivity. It is worth noting the large SDs occur especially when the connectivity is high and when a connection emerges from the ROIβ. See discussion.

The previous results show that, in a multivariate condition, just a tendency can be detected, but with a large SD. In the following, we will examine a different condition, i.e., the effect of a single change in connectivity strength, with all other connections maintained at a constant value. This can be of value to understand whether the Granger causality can detect a progressive alteration in one neural pathway (either due to learning or pathological conditions). This analysis was performed starting from the network depicted in panel A of Fig. 3.12. Here, ROIα can represent either an occipital region (which is dominated by an intrinsic α rhythm) or the thalamus, which has been hypothesized to generate an α rhythm and transmit it to other populations [209,243]. The ROIβ can represent motor–premotor areas, where this rhythm becomes evident during motor activation or motor programming tasks [238]. Finally, the ROIγ and ROIθ can represent more fronto-temporal regions (for instance, those involved in working memory), where γ-θ coupling is known to play a pivotal role, or also γ-θ coupling may represent a connectivity between frontal regions and the hippocampus [230,244]. However, this network has been built just as a simple example inspired by present knowledge on brain rhythms. Results obtained with another net (depicted in Fig. 3.12B) are shown in *Supplementary Material part 2* (Figures 3.16S and 3.17S).

In the following, for briefness, connections among two ROIs are indicated by omitting the word ROI and leaving only the Greek symbols denoting the specific ROIs (e.g, θ→γ means connection from ROIθ to ROIγ—i.e., ROIθ→ROIγ).

Fig. 3.16 shows the effect of a progressive change in a single connection strength on the temporal Granger estimates. To this end, each connection strength was individually increased from 0 to 50 (step 10), while all other connections were maintained at the basal value shown in Fig. 3.12A; 10 simulations were performed and averaged for each configuration of the network. Several aspects are of interest.

First, in any case the Granger estimator is able to detect the change in a single connectivity quite well, with a monotonic and almost linear relationship between the estimated values and the connectivity strength. Only the θ→γ and α→β connections show a certain tendency to saturate at the upper values, while the α→θ connection shows a parabolic trend.

Second, in most cases the other estimates are unaffected by the change in one connection and remain quite constant at different values of the sensitivity parameter. We observed just a few exceptions, especially concerning the ROIγ: increases in a connection involving the γ population are associated with the "apparent" decrease in another connection targeting the same ROI (the increase in β→γ is associated with the "apparent" decrease in θ→γ; the increase in θ→γ with the "apparent" decrease in β→γ; the increase in α→γ with the apparent decrease in β→γ). These changes always concern a bisynaptic connections (such as β→γ→θ or α→β→γ). Furthermore, it is interesting to observe that a change in the connection α→β produces some effects also on the connections α→γ and θ→γ, which exhibit a moderate apparent increase. This is likely a consequence of the indirect effect of α on γ via β.

Third, the relationship between the "true" connectivity value and the estimated one exhibits quite a similar slope for all connections, with the exception of the connection γ→θ, which exhibit a higher slope than the others (see the different y-axis in this panel).

Moreover, it is worth noting that the slope of the relationship "estimated connectivity vs. true connectivity" is about half that illustrated in Fig. 3.15. This may be the consequence of the smaller input (=200) to the ROIα used in the network of Fig. 3.12a (the network utilized for the analysis in Fig. 3.16 and following figures) compared to the value used for this input (=400) in the analyses of Figures 3.14 and 3.15 (but see also Fig. 3.14S in Supplementary Material part 2, where we compare the results obtained in the random net with two different inputs to ROIα, and further show how a change in the input affects connectivity).

The latter consideration moves our attention to the role of the inputs reaching the ROIs—i.e., a change in the population working point. This is assessed in the next subsection.

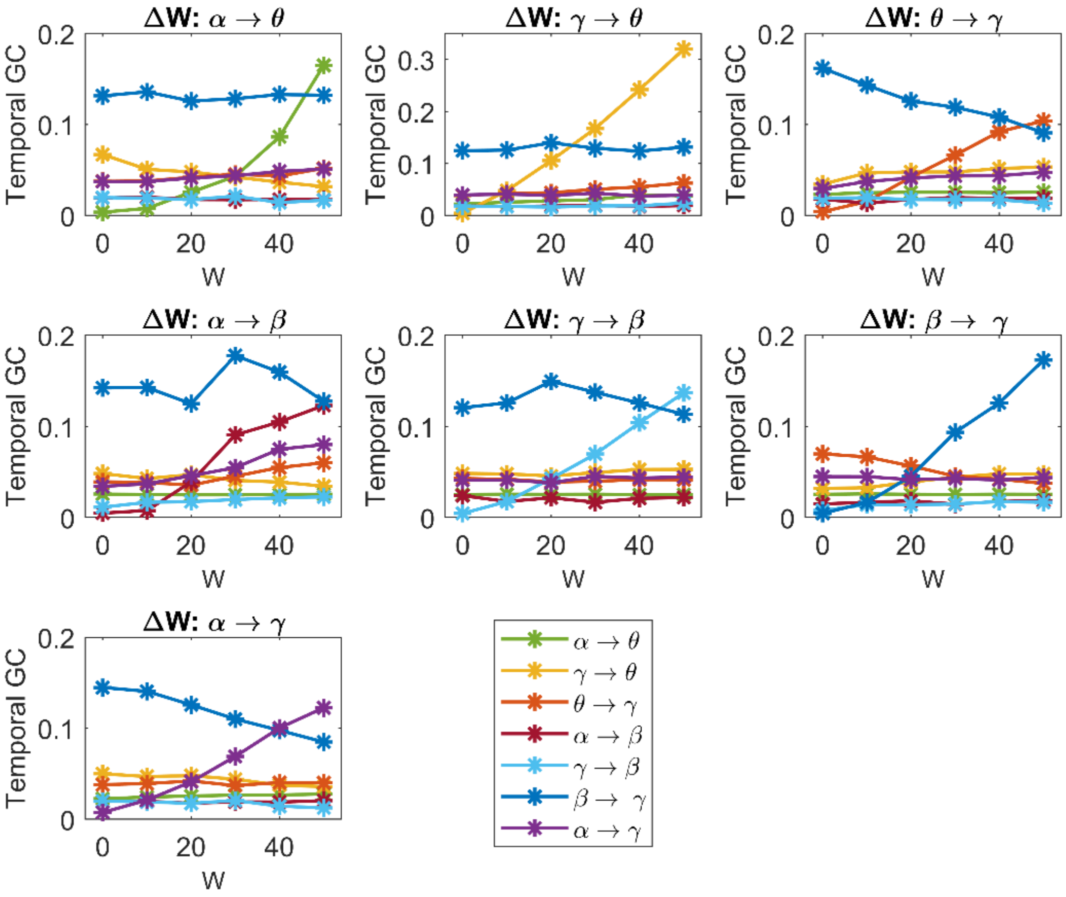


Figure 3.16. Values of connectivity among the ROIs estimated with the temporal Granger estimator, with reference to the network in Fig. 3.12A, when one connection is progressively varied in the x-axis (from 0 to 50, with step of 10), and the other connections are maintained at the basal value as in Fig. 3.12A. It is worth noting that the estimator is able to detect the progressive increase in a single synaptic strength, while the other estimates remain almost constant. As an exception, we observed that the increase in a synapse entering into ROIγ is often associated with a decrease in another synapse entering the same ROI. It is also worth noting the higher sensitivity of the estimator to the connection γ→θ.

3.2.3.4 Effect of the Inputs on the Estimated Connective Strength

In order to investigate the role of a change in the inputs, which in turn modifies the working point of a population along the sigmoidal relationship, we changed the excitation to pyramidal neurons in one ROI, while all other inputs were maintained at the value illustrated in Fig. 3.12A. Figures 3.17, 3.18, 3.19 and 3.20 display the effect of a change in the input to ROIβ, ROIγ, ROIθ and ROIα, on the estimates obtained with the temporal Granger causality. The upper panels show the values of each estimate as a function of the input, while the bottom panels show the corresponding patterns of connectivity, obtained using a threshold as high as 0.015 (taken as optimal from the ROC curve in Fig. 3.14).

As is clear in Figures 3.17–3.20, a change in the input has a significant effect on the estimated connections which enter into or exit from the given ROI. In particular:

(i) Changing the input to ROIβ (Fig. 3.17) causes a dramatic change in the estimated connection β→γ, which exhibits a high value when the input to ROIβ is in the range 300–400, and falls to very low values when the input is 0–100 or 600–800. We ascribe this behavior to the fact that pyramidal neurons in the region ROIβ enters into the bottom or upper saturation zone of the sigmoidal relationship, hence providing a small output signal. Conversely, the entering connections γ→β and α→β exhibit the opposite behavior: they decrease in the central zone (input 300–400) and increase dramatically in the saturation regions. At the same time, the connections involving region ROIγ also change, as illustrated In the right panel of Fig. 3.17.

(ii) Increasing the input to ROIγ (Fig. 3.18) causes a progressive increase in the estimated output connection γ→θ and a dramatic fall in the entering connections β→γ and θ→γ.

(iii) Increasing the input to ROIθ (Fig. 3.19) causes a significant change in the estimated output connection θ→γ. This connection is higher at intermediate levels of the input (300–500) and falls down when the region ROIθ enters into the bottom or upper saturation zones (input 0–100 or 600–800). The opposite pattern is evident as to the entering connection γ→θ, which increases when ROIθ is in the saturation zones and decreases in the central region. Additionally, the entering connection α→θ decreases in the central region but remains low also in the upper saturation region. This global behavior resembles that already described in Fig. 3.17 when changing the input to ROIβ.

(iv) Increasing the input to ROIα (Fig. 3.20) causes an evident increase in all the estimated output connections (α→β, α→γ and α→θ). This is paralleled by a progressive increase in most other connections, including the “spurious” connections β→θ and θ→β, which were set at zero in the original network. Using very high values for the inputs, the network overconnected.

The bottom panels in Figures 3.17–3.20 show examples of connectivity networks, obtained using 0.015 as a threshold. As it is evident, some “true” connections can be lost or other “spurious” connections can emerge as a consequence of the input changes. In particular, in most figures the estimated connections not included in the network (that is, β→θ, θ→β, β→α, γ→α, θ→α) exhibit very low values below the discrimination threshold; however, when increasing the inputs, some of these connections can rise, causing false positive estimations.

Finally, Fig. 3.15S in *Supplementary Material part 2* summarizes the effect of a change in the input to ROIβ (the same as in Fig. 3.17) but evaluated with all estimators in order to better understand the differences between the different metrics. The main considerations developed above are confirmed, although with some differences among the different estimates.

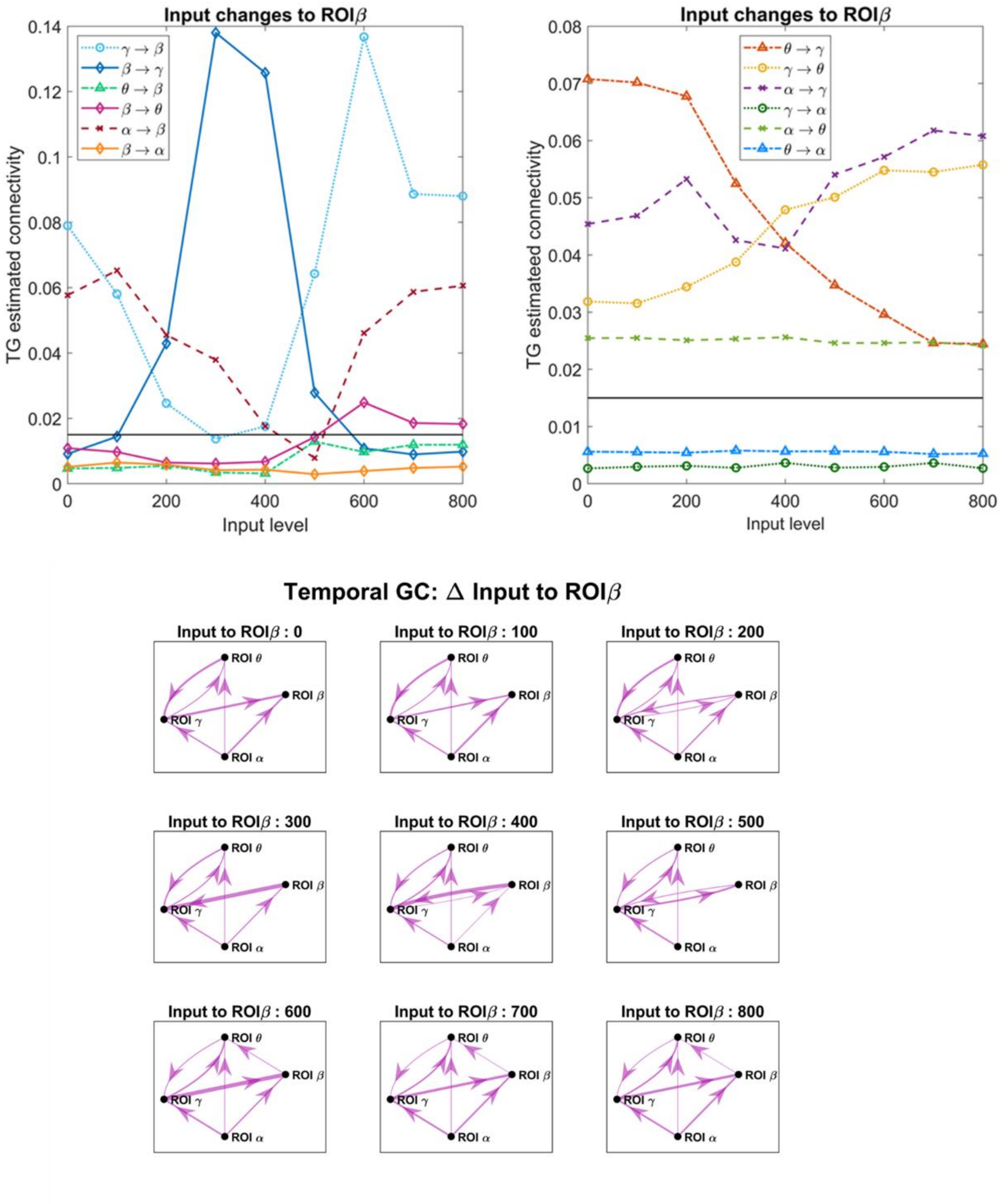


Figure 3.17. Upper panels: Values of connectivity among the ROIs estimated with the temporal Granger causality, with reference to the network in Fig. 3.12A, when the input to pyramidal neurons in ROIβ was progressively varied from 0 to 800, as in the x-axis, and all other inputs and connections were maintained at the basal value, as in Fig. 3.12A. It is worth noting the strong effect that the input change has on the connections which involve the ROIβ (in particular the output connection β→γ and the input connection γ→β). Additionally, the connection between ROIγ and ROIθ is affected. Bottom panels: connectivity graphs obtained from the estimates using a threshold as low as 0.015 (the threshold is depicted as a horizontal black line in the upper panels).

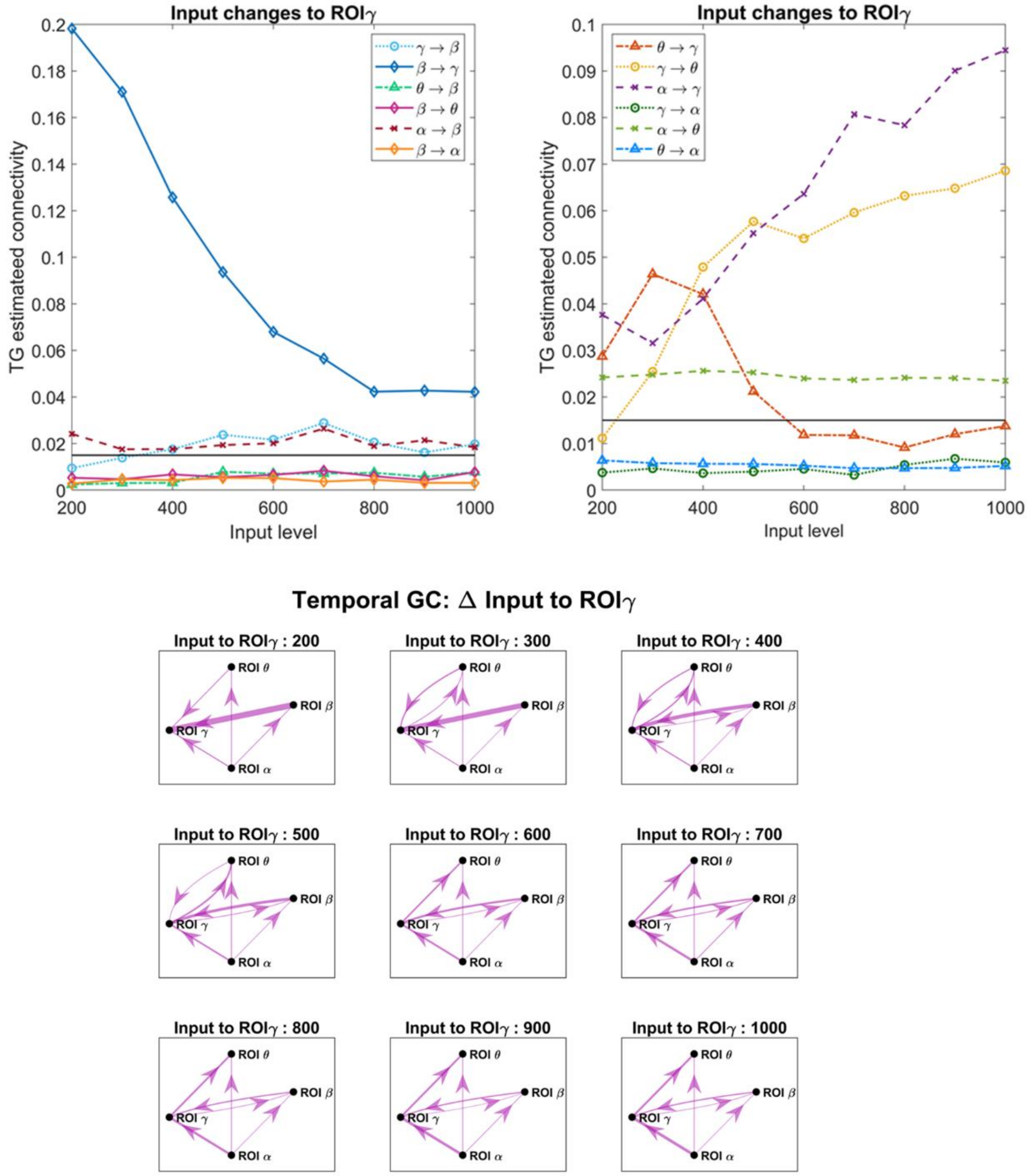


Figure 3.18. Upper panels: Values of connectivity among the ROIs estimated with the temporal Granger causality, with reference to the network in Fig. 3.12A, when the *input* to pyramidal neurons in ROIγ was progressively varied from 200 to 1000, as in the x-axis, and all other inputs and connections were maintained at the basal value, as in Fig. 3.12A. It is worth noting the strong effect that the input change has on several connections which involve the ROIγ. Bottom panels: connectivity graphs obtained from the estimates using a threshold as low as 0.015 (the threshold is depicted as a horizontal black line in the upper panels).

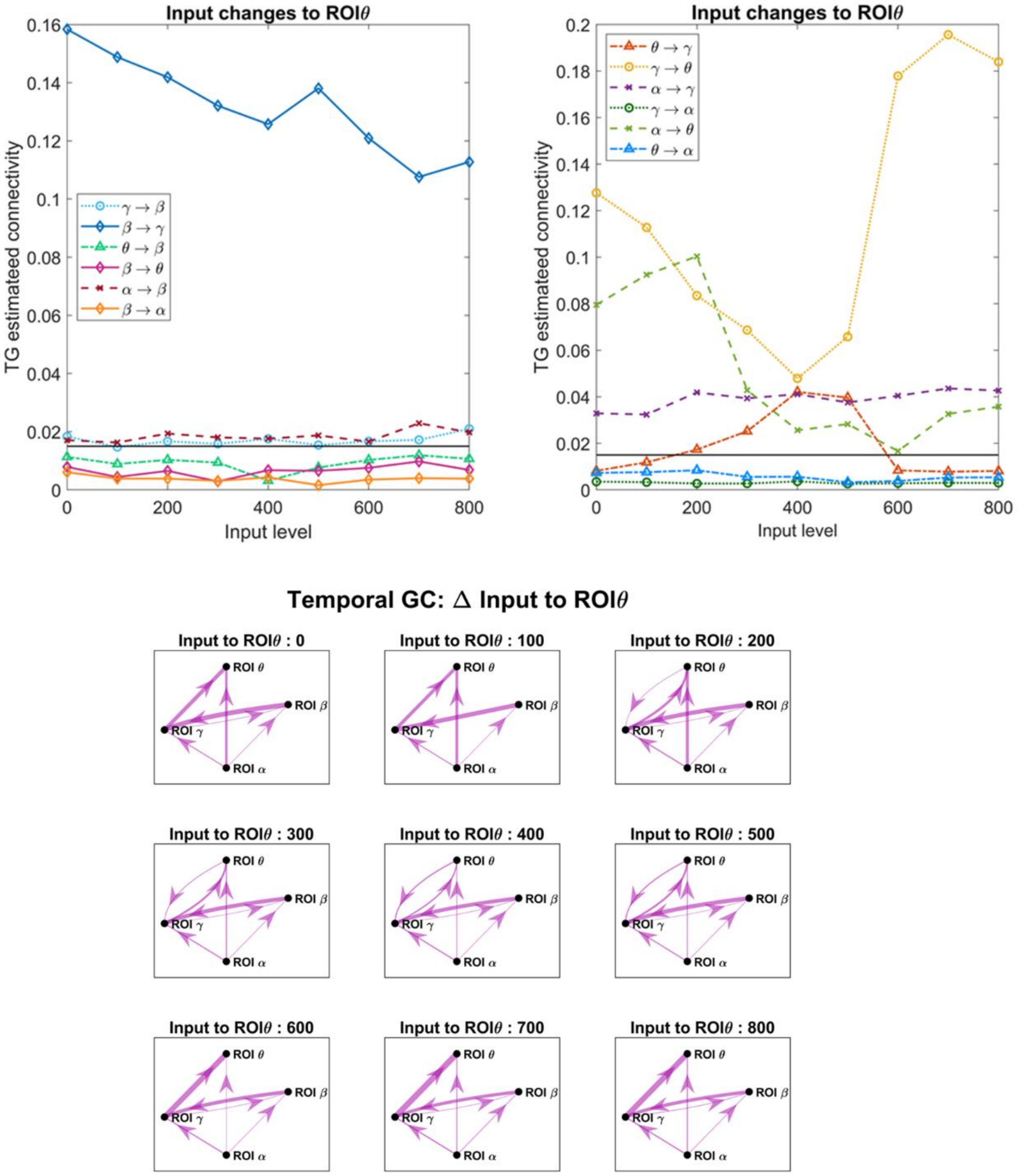


Figure 3.19. Upper panels: Values of connectivity among the ROIs estimated with the temporal Granger causality, with reference to the network in Fig. 3.12A, when the *input* to pyramidal neurons in ROIθ was progressively varied from 0 to 800, as in the x-axis, and all other inputs and connections were maintained at the basal value, as in Fig. 3.12A. It is worth noting the strong effect that the input change has on the connections which involve the ROIθ. Bottom panels: connectivity graphs obtained from the estimates using a threshold as low as 0.015 (the threshold is depicted as a horizontal black line in the upper panels).

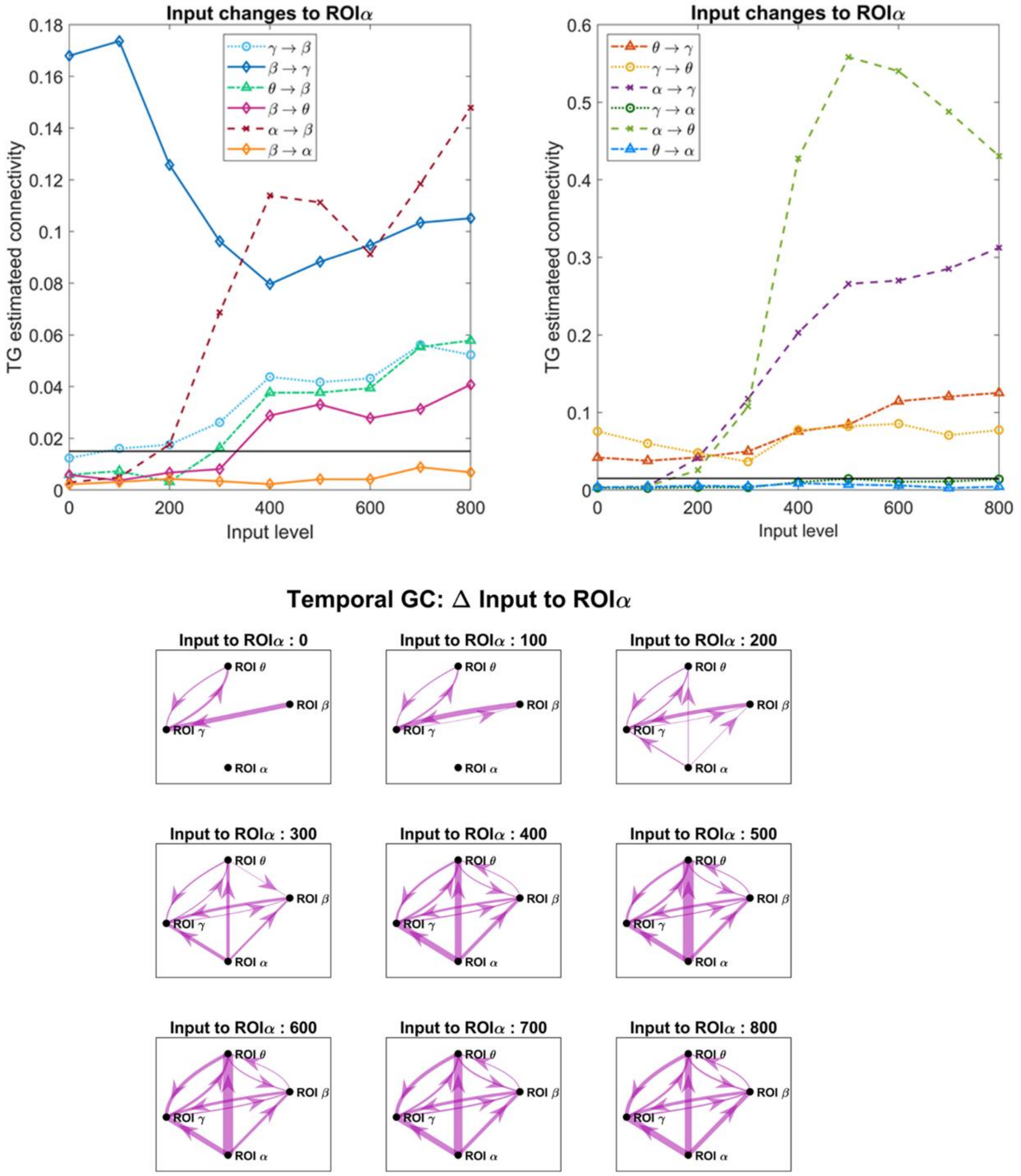


Figure 3.20. Upper panels: Values of connectivity among the ROIs estimated with the temporal Granger causality with reference to the network in Fig. 3.12A, when the *input* to pyramidal neurons in ROIα was progressively varied from 0 to 800, as in the x-axis, and all other inputs and connections were maintained at the basal value, as in Fig. 3.12A. It is worth noting that all estimated connections from ROIα to the other regions increase significantly. Many other connections increase too with the appearance of spurious terms. Bottom panels: connectivity graphs obtained from the estimates using a threshold as low as 0.015 (the threshold is depicted as a horizontal black line in the upper panels). The emergence of spurious connections is evident from these graphs.

3.2.3.5 Analysis in the Frequency Domain

In order to better understand connectivity in the different frequency bands, Figures 3.21–3.23 illustrate a few examples taken from Figures 3.17–3.20, using the Granger estimation in the frequency domain. In these figures, each panel represents the connectivity between a couple of regions.

Fig. 3.21 refers to the connectivity network and input values as in Fig. 3.12A. It is evident (Fig. 3.21 panel A) that the region ROIβ transmits a strong information in the β band to region ROIγ, while the information from ROIγ to ROIβ, located at approximately 35 Hz, is less relevant. A strong coupling is also evident between ROIθ and ROIγ in the respective bands (Fig. 3.21 panel C), whereas the coupling between ROIθ and ROIβ is negligible (panel B; also see the different y-axes in the figures). Finally, panels D, E and F illustrate clearly how the α rhythm is transmitted from ROIα to the other regions, without receiving any relevant rhythm back. We observed the stronger transmission γ→θ, as already underlined above. Basically, Granger in the frequency domain produces similar results as Granger in the temporal domain but adds very useful information on rhythms transmission in different bands.

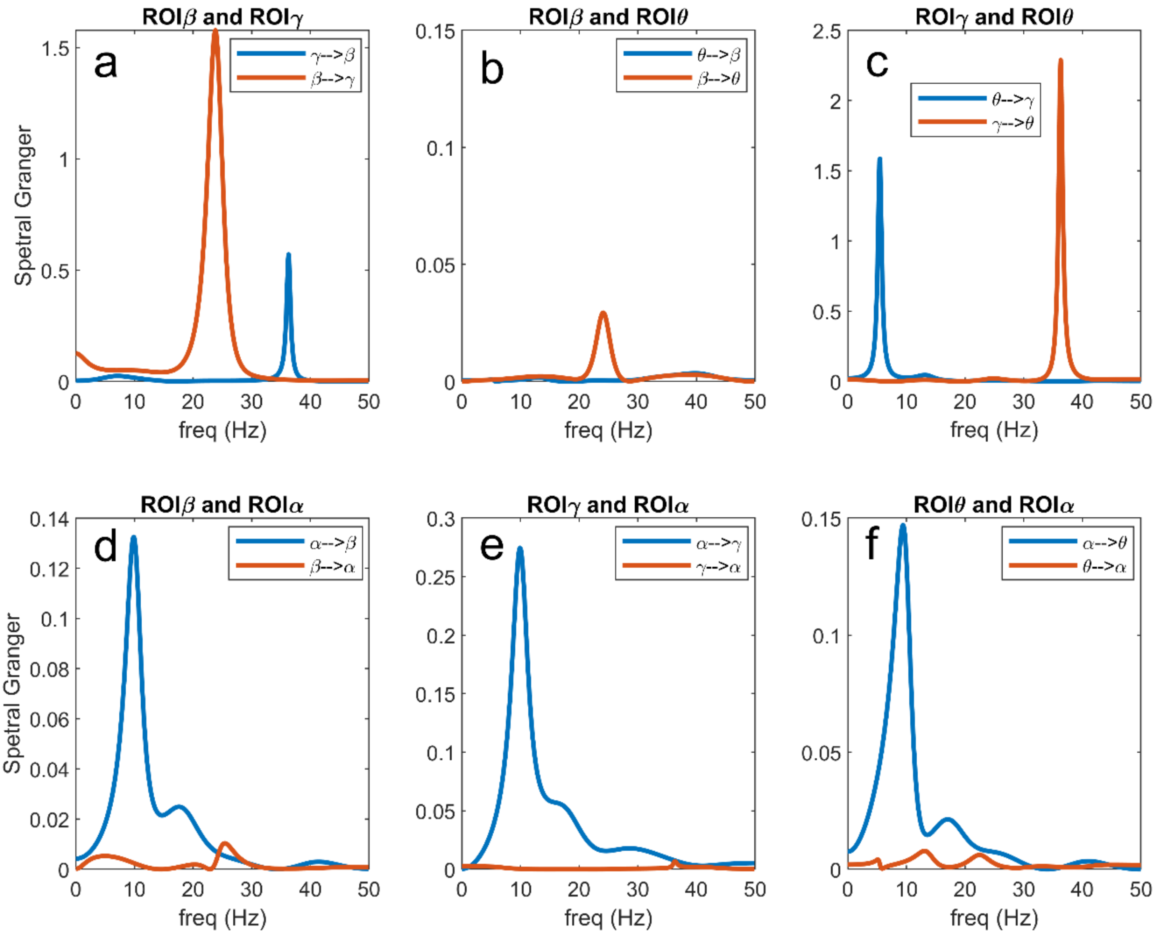


Figure 3.21. Values of connectivity among the ROIs estimated with the spectral Granger causality with reference to the network in Fig. 3.12A and plotted as a function of frequency (note the use of different y-axes to emphasize the different cases). The estimator reproduced the network connectivity quite well: the exchange of rhythms

between ROIβ and ROIγ (panel a) and between ROIγ and ROIθ (panel c) is evident; the coupling between ROIθ and ROIβ is negligible (panel b); and the α rhythm is clearly transmitted from ROIα to the other regions (panels **d,e,f**).

Fig. 3.22 illustrates what is occurring when the input to ROIβ is reduced from 400 to 100 (see also Fig. 3.17). In this condition, ROIβ does not transmit its β rhythm to ROIγ, while the transmission γ→β increases (panel A). Due to a smaller activation of the ROIγ, the rhythm transmitted from ROIγ to ROIθ is also reduced compared with the previous case, while θ→γ increases (panel C). Coupling between ROIθ and ROIβ is still negligible (panel B), but we observed an increased transmission of the α rhythm from ROIα to ROIβ (panel D) and also a small increase in the transmission from ROIα to ROIγ (panel E). Finally, the transmission from ROIα to ROIθ (panel F) is similar as in the previous figure. These results summarize well how an alteration in the input modifies the capacity of a region to transmit its rhythm to others and to receive rhythms from others.

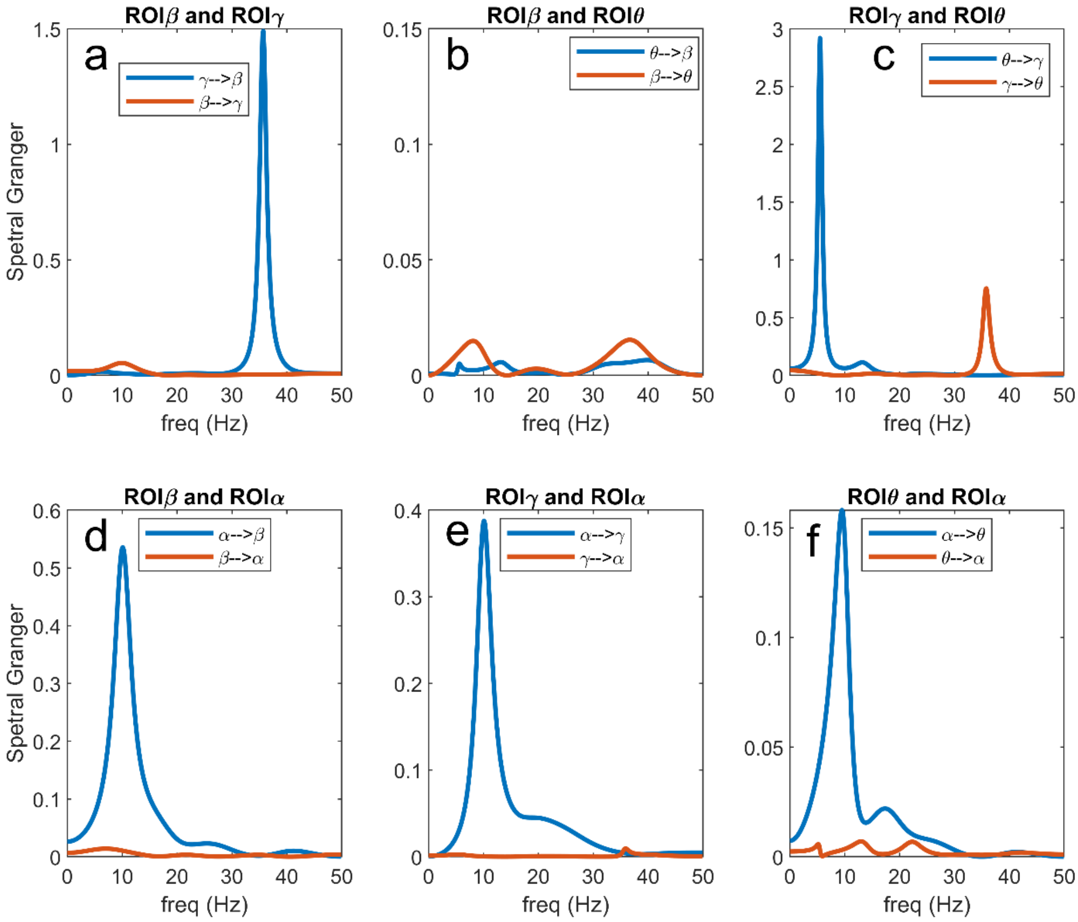


Figure 3.22. Values of connectivity among the ROIs estimated with the spectral Granger causality, with reference to the network in Fig. 3.12A but with the input to ROIβ reduced from 400 down to 100. The values are plotted as a function of frequency (note the use of different y-axes to emphasize the different cases). From the comparison of these panels with those in Fig. 3.21 emerges that: the absence of α rhythm transmitted from ROIβ to ROIγ is evident (panel a); the coupling between ROIθ and ROIβ is negligible (panel b); the transmission γ→θ is significantly reduced, while that θ→γ is increased (panel c); the α rhythm transmission from ROIα to ROIβ (panel

D) and from ROIα to ROIγ is increased (panel E); while the α rhythm transmission from ROIα to ROIθ (panel F) is similar.

Fig. 3.23 shows the effect of an increase in the input to ROIα from 200 to 400. The dramatic increase in the α rhythm propagated from ROIα to the other three ROIs is evident (panels D, E and F). In particular, this rhythm becomes almost sinusoidal, as evident by the sharp peak in the spectra. As a consequence, α rhythm becomes relevant everywhere in the net and is also transmitted between other regions. In particular, ROIβ and ROIγ now exhibit a significant transmission in the α band, while ROIγ almost completely loses its capacity to transit the γ oscillation to ROIβ (panel A). The ROIγ is also able to transmit an α rhythm to ROIθ, but the capacity to transmit its intrinsic γ rhythm is drastically reduced compared with Fig. 3.21 (panel C). Finally, and more importantly, a spurious connectivity appears in panel B, where the regions ROIβ and ROIθ, not physically connected according to the schema in Fig. 3.12A, apparently exchange information in both α and β bands. This agrees with the spurious connectivity illustrated in Fig. 3.20. Further exempla concerning the schema in Fig. 3.12B are illustrated in *Supplementary Material part 2* (see Figures 3.16S and 3.17S).

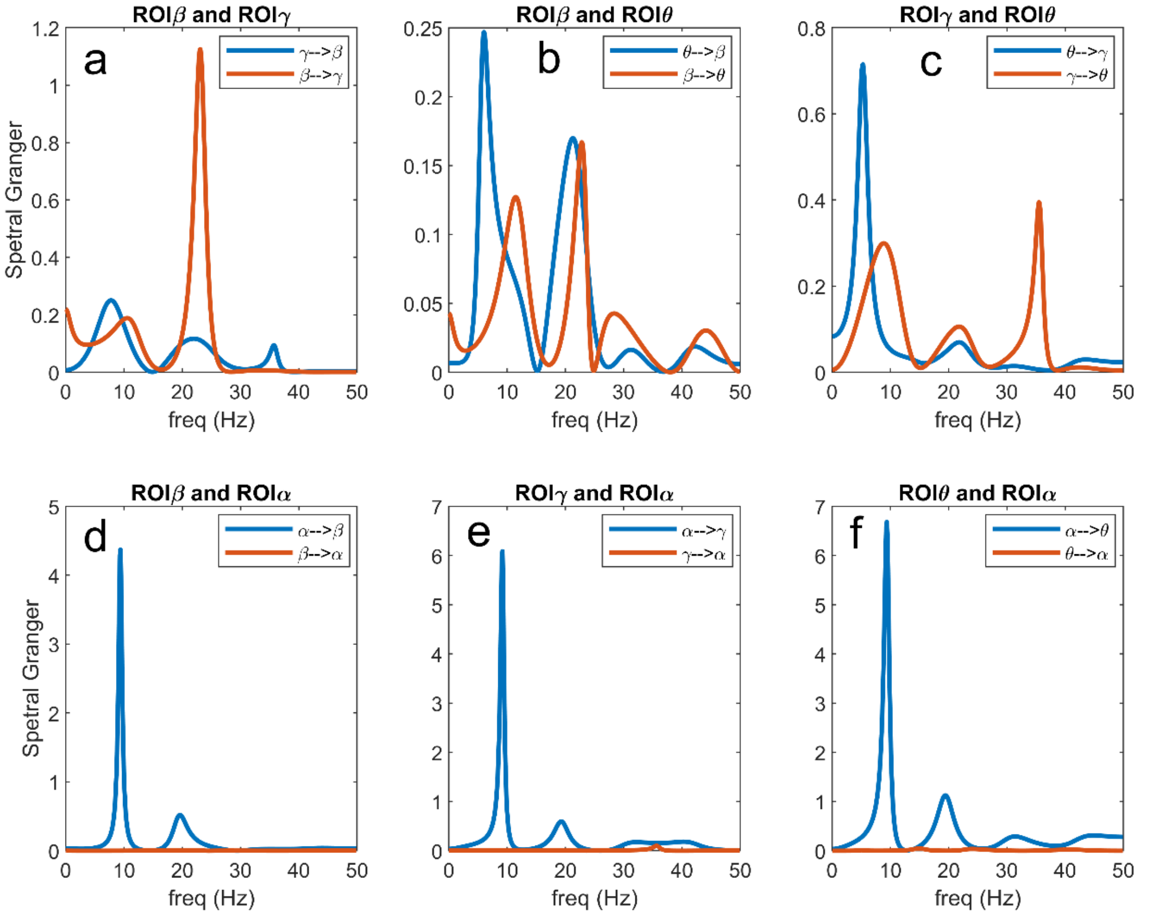


Figure 3.23. Values of connectivity among the ROIs estimated with the spectral Granger causality, with reference to the network in Fig. 3.12A but with the input to ROIα increased from 200 to 400. The values are plotted as a function of frequency (note the use of different y-axes to emphasize the different cases). From the comparison of these panels with those in Fig. 3.21 emerges that: the α rhythm transmitted from ROIα to the other regions becomes almost sinusoidal (high and sharp spectra, panels d,e,f); the rhythm exchange between ROIβ and ROIγ

(panel a) and between ROIθ and ROIγ (panel c) is significantly altered, with the presence of components in the α range; finally, spurious connections appear between ROIβ and ROIθ (panel b), both in the β and α ranges. Comparing this figure with the corresponding bottom panel in Fig. 3.20, the spectral Granger causality provides similar information as the temporal Granger causality, but with emphasis on the frequency bands.

3.2.4 Discussion

Rhythmic oscillations in brain activity are known to play a relevant role in many cognitive tasks, including memory, attention, binding and segmentation of perceptual experience, motor actuation [228–231]. Furthermore, several results suggest that different human cortical regions are characterized by a dominant oscillation, the so-called natural frequency, as demonstrated by single-pulse transcranial magnetic stimulation (s-TMS) [235,237] or by electrocorticogram [245]. In particular, α oscillations (8–12 Hz) dominate over the parieto-occipital cortex, and are also significantly related with attentional modulation [246–248]; low β oscillations (13–20 Hz) are evident over the left superior parietal lobule (BA 7) [235] and the dorsolateral prefrontal cortex [237], while faster frequencies (21–50 Hz) occur over the left premotor cortex and anterior areas [235]. Gamma oscillations are also frequently observed in the prefrontal cortex and in the hippocampus, which are associated with the θ rhythm; in particular, $\gamma - \theta$ coupling is related with working memory tasks and spatial memory tasks [230,244]. Moreover, several studies reveal that the local rhythm, evoked by TMS, spreads toward further connected regions [194,235]. All these data suggest that propagation of rhythms among brain regions is a fundamental instrument to realize complex cognitive functions, leading some authors to postulate the existence of a “system of rhythms” subserving cognition [249].

This “system of rhythms”, of course, needs a network of connections, which permits the transmission of oscillations from one region to another, the possible synchronization/desynchronization among neural activities, and the coordinate exchange of reciprocal information. Indeed, the study of brain connectivity is playing a major role in cognitive neuroscience today. A common method to evaluate these connections, from the perspective of brain rhythms, is to apply methods for estimation of functional connectivity to neuroelectrical signals (such as EEG or MEG measurements which encompass a sufficient temporal dynamics). Although the typical analysis of FC in humans occurs through noninvasive methods, there is also the possibility to perform invasive connectivity studies via ECoG—for instance, during a presurgical evaluation of epilepsy (see [227], for a review). In this case, the objective is to quantify the involvement of the different regions in the triggering of the seizure. Actually, our analysis can also be used to analyze relationships between mean field potentials derived from invasive measurements.

Hence, in order to challenge methods for FC estimation, we need a biologically inspired model able to simulate these rhythms to analyze their reciprocal transmission and the effect of nonlinearity, which play dominant roles in brain dynamics.

Unfortunately, most models used in past years for FC analysis are non-oscillating and often make use of linear equations. Models which make us of spiking neurons are too complex to simulate dynamics of entire brain regions and are suitable for the analysis of neuron cultures in vitro [185]. A few previous studies which investigated FC with neural mass models made use of equations with just two populations (excitatory and inhibitory) or just two interconnected regions. Hence, we think the present model represents an important advancement in the current literature.

The aim of this work was to evaluate the relationship between rhythms transmission among ROIs, network connectivity, and methods for FC estimation, laying emphasis on the possibility to detect the strength of the reciprocal connections and, above all, the effect of nonlinear alterations in neural activity. To this end, we used data simulated with a neural mass model of interconnected populations as a ground-truth. As pointed out by Reid et al. [172], the use of abstract model simulations may offer several advantages compared with more sophisticated models—among others, better intuition of the results, reduced computational costs and the capacity to generalize over several physiological conditions.

The signal characteristics exemplified in the frequency and temporal domains (Fig. 3.13 and Fig. 3.13S in *Supplementary Material part 2*) indicate that the present model, although extremely simplified, can grasp some important aspects of real neurobiological signals. Indeed, although the present simulations have been performed with constant inputs, constant characteristics of the noise and constant parameters (i.e., in a certain sense, with a *stationary* model), the spike density of pyramidal neurons and mean field potentials (i.e., quantity z_p and v_p in the Equations (3.3) and (3.4) of the Equations 3.1 - 3.18 of section 3) exhibit rapid short-living fluctuations. As underlined in [241,242], not only frequency but also nonsinusoidal waveform shapes may play a relevant role in neurophysiological processes and behavior. However, in real signals, this intermittency in rhythmic activity is probably even more accentuated than in our model, due to the typical nonstationarity of all biological systems (i.e., the inputs and parameters are never constant as in our simulations).

The validity of the model can be further assessed and its limits pointed out by looking at the spectral patterns of the different interconnected ROIs.

(a) The spectrum in ROIα exhibits a clear peak at about 10 Hz, with smaller contributions in the β band. There are several neurophysiological regions that can exhibit a similar pattern. This rhythm may originate from the thalamus [250]. Moreover, the spectrum in this α region is similar to mu rhythms, observed in the sensorimotor cortex (see [251] where a clear peak at about 10 Hz is associated with a smaller component in the β range). Similar spectra can also be seen in occipital regions in a relaxed state (e.g., see [252,253]).

(b) The spectrum in the ROIγ exhibits a very large peak, which is difficult to observe in real EEG signals in the scalp, or also in signals reconstructed on the cortex starting from scalp EEG data (for instance, using algorithms for source reconstruction). There are several possible explanations. First, the γ rhythm can be attenuated by low-pass filtering properties of the tissue; hence, its presence in the scalp is strongly reduced. However, an evident γ peak can be observed in local field potentials during invasive measurements when a population is stimulated (for example, see [254,255]). We think that these rhythms are typical of specific brain regions (for instance, limbic regions such as the hippocampus, or sensory regions when excited by external stimuli and involved in the binding information, or frontal regions involved in working memory). Hence, it is important that a portion of the model produces a clear γ to be transmitted to other regions. It is probable that, in a real brain network composed of multiple regions, the effect of this rhythm may be less evident than in our four-region model. However, its role is extremely important, and we need a ROI contributing to it.

(c) The spectrum in the ROIβ is similar to that observable in motor or premotor regions during a motor task [238,256].

(d) The spectrum in the region ROIθ exhibits approximately a 1/f trend with smaller peaks at high frequencies compared with lower frequencies. This is probably the region more representative of the behavior of many cerebral regions, usually observed with non-invasive EEG.

Of course, even more realistic signals can be built starting from the present model—for instance, using a greater number of ROIs, a different combination of weights, or even real LFP signals as inputs to the regions. Moreover, in future works parameters may be fitted to real EEG signals to further improve model simulation of real cases.

With such a biologically inspired model, we then tested several FC estimation techniques in the presence of rhythm transmission.

A preliminary comparison among the different estimation methods showed that, at least for the particular problem under study (i.e., rhythm transmission), the Granger connectivity performs better than the other techniques. Hence, most of the analysis in this work was performed with this estimation method (in the temporal and frequency domains), while some results obtained with the other methods are shown in the Supplementary Material part 2. However, it is important to remark that most estimators require the setting of some parameters. We did not perform an exhaustive search of the best parameter combinations for each estimator, since this is well beyond the aim of the present work, and this aspect may be investigated in future studies.

The poor performance obtained with some estimators, such as the phase synchronization index and the delayed correlation (see Fig. 3.14 and Figures 3.15S and 3.16S in Supplementary Material 2) deserves a comment; this in part contradicts the more encouraging results obtained by us and others in previous works [257]. In particular, in a recent paper we showed that the delayed correlation is able to grasp the sign of a connection (either excitatory or bisynaptic inhibitory) quite carefully when two ROIs transmit the same rhythm with a feedback connection (see Supplementary Material of Section 3.1[137]). Conversely, as shown in Supplementary Material part 2, in the present study we observed that the sign of the delayed correlation is not always related with the nature (excitatory or inhibitory) of the synapses. Similarly, the phase synchronization metrics provide quite poor results in our simulations. We claim that these differences depend on the coupling among oscillations with different frequencies (in the θ, α, β, γ ranges, as actually occurs in the brain), whereas previous studies were especially focused on synchronization of a unique rhythm from one region to another. A different index, involving phase–amplitude coupling [258] may be more appropriate to study connectivity among different rhythms, and may be tested in future work.

However, the main objective of this work was not to compare the performance of different estimators, but rather to critically challenge the concept of FC. Our aim was to assess whether FC can detect not only network topology, but also possible changes in connectivity strength and, above all, how these estimates are affected by nonlinearity in neural dynamics.

We can summarize some interesting indications for MEG/EEG analyses. Fig. 3.15 simulates cases of simultaneous changes in multiple pathways connecting the investigated ROIs, as may occur in several pathological and physiological conditions (for instance, after a stroke, or after training/conditioning paradigms such as multisensory stimulation training or fear conditioning). Our findings indicate that when several pathways change simultaneously, only large variations in a given connection can be detected (for instance, from a high to small value or vice versa; see Fig. 3.15) due to the high SD of the estimates, whereas care must be taken to interpret small connectivity changes, which can be the result of spurious effects. However, a greater reliability of the estimates can be obtained when the results are mediated over large populations of different subjects; in this case, as shown in Fig. 3.15, the average value across many trials can reflect the connectivity in that population quite well. Another interesting point emerges from Fig. 3.16, where only one connection is changed at a time. The results suggest that the FC estimate is quite accurate when only one pathway is changing, while the others maintain a constant level. This may occur in specific stimulation approaches, able to selectively affect only a given neural pathway; in particular, new approaches are emerging (such as cortical–cortical paired associative stimulation via TMS, [259]) that can induce plastic changes of a spatial specific and also functionally specific neural pathway. FC estimates can be reliably applied to these approaches when combined with EEG acquisition (and source reconstruction) to derive quantification of the strength changes induced in the specific neural pathway and quantify the relationship between strengthening of the given pathway and behavior.

A further consideration is the case when a connection has a higher value (as high as 30–40) in our random network. In this case, the estimates exhibit a very high variability (mostly evident in the transmission of the β rhythm, see Fig. 3.15, but also in other rhythms). Our interpretation is that, in this condition, the working point of the target population may be drastically altered by the strong incoming connectivity. In other terms, the target population might be silenced, in the case of strong inhibition, or largely excited, in the case of strong excitation, which may trigger nonlinear phenomena, resulting in a completely different capacity to receive or transmit rhythms. This might explain the enormous variability of the estimated connections. Of course, other multivariate aspects (such as the presence of a common source, mixed inputs, or the presence of multiple pathways) can also affect the estimation in the random net. A future paper can evaluate this phenomenon using multivariate algorithms too, such as the multivariate transfer entropy [182] or the partial directed coherence [116].

The previous consideration moves our analysis towards the most important aspect of this study—i.e., the role of nonlinearity (specifically, the population working point) in connectivity estimation.

Although in this study a thorough analysis on the effect of input changes and nonlinearity has been performed using the Granger methods (both in the temporal and frequency domains, Figures 3.17–3.23, and Fig. 3.14S in *Supplementary Material part 2*), similar results have been obtained with all estimators, with only moderate differences (see Supplementary Material Figures 3.15S and 3.16S). All metrics agree in that they show that a change in the input to a region causes well evident and repeatable alterations in the "apparent" connectivity—i.e., the estimated functional connectivity. This can be ascribed to a dramatic alteration in the capacity to propagate or receive a rhythm, which occurs despite structural connectivity is unchanged. To investigate this phenomenon, we used quite a simple network inspired by biology, where the α rhythm (occipital/thalamic) inhibits the other regions, and the γ rhythm (fronto-temporal) strongly interacts with the β (motor) and θ (hippocampal) populations. However, similar results can be obtained using different nets too. We tested different excitation/inhibition nets (unpublished simulations) confirming this result (an example, concerning a simple chain of interconnected rhythms is shown in *Supplementary Material part 2*, Figures 3.16S and 3.17S). We also tested a network identical to that shown in Fig. 3.12A, but with excitatory connections only, obtaining similar results.

Despite the presence of clear differences between one region and another, some major common points can be drawn: first, the capacity to generate and transmit a rhythm increases when the population of pyramidal neurons is working in the central linear region of the sigmoidal relationship, but this capacity decreases significantly when the population enters into the (upper or lower) saturation regions; second, in the latter condition a ROI becomes much more affected by the incoming rhythms (i.e., the rhythms arriving from other ROIs which send their synapses into). These aspects are especially evident with reference to the β rhythm due to the smaller amplitude of this oscillation, which makes it strongly dependent on the working point in the sigmoidal relationship but visible to all rhythms and all estimators.

Particularly, results obtained on the α rhythm, by varying the input strength, deserve attention (Fig. 3.20). They suggest that the capacity to propagate α from one region to another dramatically changes if the ROIα is excited, and this is reflected in a stronger presence of the α in the other connected regions. We think that this result agrees with the literature and may shed light on some aspects of the brain α. Indeed, many data in the literature show that the α rhythm can exhibit dramatic changes from one mental state to another (this is especially evident in the occipital cortex, but frontal regions can also exhibit significant changes—for instance, during working memory tests). α is the only rhythm that can either increase or decrease compared with basal conditions [248]. Furthermore, it is a marker of stress, fatigue, and can change dramatically with closed or open eyes. Finally, there is large consensus that α is linked to attention and top-down influences, and an increase in α denotes suppressed activity in regions unessential for the specific task (for references on these aspects, see [260–262] as reviews, but also [263,264]). Hence, the fact that α power can strongly change as a function of a top-down mechanism (maybe depending on the thalamus or higher frontal regions) agrees with present knowledge on the role of this rhythm; in this regard, the model may represent a promising instrument for the mechanistic interpretation of this rhythm propagation in various states.

Briefly, although the methods for FC estimation (and, in particular, those based on Granger causality) can be extremely useful to understand how a rhythm is received or transmitted from one ROI to another (see, for instance, the frequency plots in Figures 3.21–3.23), the values do not always reflect the true structural connectivity within the network, being strongly influenced by the particular working conditions in which the net is actually operating. In particular, the presence of a strong rhythm transmitted from a presynaptic region to a postsynaptic region can be due not only to a strong connectivity but can reflect the particular conditions of the receiving ROI, especially when the latter is scarcely excited by other external inputs and becomes more prone to the incoming influences. Similarly, the presence of a poor rhythmic transmission does not necessarily reflect the absence of a causal link but may signify that the source region is scarcely active at that moment, or that the source region reached an upper saturation, which significantly depresses its rhythmic variability. In other words, estimators reflect how much information is actually transmitted in a particular frequency domain, and in a particular operational mode, but this does not always correspond to structural connectivity.

This argument has two main implications for connectivity experimental studies. First, caution must be taken when interpreting the comparison between connectivity networks obtained during different tasks (for instance, in resting conditions and during task execution), since a task may alter the working conditions of some regions. Conversely, most studies making use of FC estimation techniques directly compare the networks obtained during different tasks and try to interpret these differences merely in terms of a change in the underlying connectivity. Indeed, a frequent conclusion in the literature is that a task dramatically alters the structure of the connectome. In our opinion, this conclusion may be questionable if the connectivity network is believed as a real physical structure—i.e., if we are looking for physical synapses linking the regions. Moreover, it is worth noting that this misinterpretation (i.e., assuming any change in the transmitted information as a true change in physical connectivity) is frequently made by using methods for effective connectivity estimation (i.e., methods that consider an underlying model). This is particularly true when these methods make use of a linear model to infer causal parameters from data [199,222,223], since these models neglect nonlinearity.

Of course, we do not exclude the possibility that some physical connections can change quite rapidly, as a function of the particular task (for instance, reflecting receptor binding by neurotransmitters), but we suspect that at least part of task-dependent connectivity changes reflects nonlinear phenomena, as illustrated in the present study, rather than a structural connectivity alteration.

3.2.5 Conclusions

In conclusion, our study critically examines the concept of functional connectivity, using a biologically inspired model as ground-truth, which generates non sinusoidal oscillations in different frequency bands. An important aspect of the study is the emphasis on rhythm propagation and on its dependence on nonlinear phenomena typical of neural systems.

As an innovative emerging concept, we wish to underline that results obtained through FC estimators, reflecting information exchange among ROIs, can provide evidence not only on the network connectivity, as usually carried out in the literature, but also on how a region may be activated or silenced during the given task as a function of other external influences. This may lead to a more complete and innovative analysis of brain functioning. Since different tasks determine different working conditions in neural populations, a more exhaustive analysis of brain functioning should move from a linear to a nonlinear perspective (although the latter is mathematically more complex), and from network connectivity metrics to metrics which incorporate connections, inputs and working conditions altogether.

3.2.6 Supplementary Material part 1

Table 3.3S - Parameters assumed fixed for the four populations

Parameter	**value**	**meaning**
e_0	2.5 Hz	Saturation of the sigmoid
s_0	10 Hz	Center of the sigmoid
r	0.56 mV^{-1}	Slope of the sigmoid
T	10 ms	Delay
Ge	5.17 mV	Synaptic gain of excitatory
Gs	4.45 mV	Synaptic gain of inhibitory slow
Gf	57.1 mV	Synaptic gain of inhibitory fast

Table 3.4S – Parameters values used for the four populations

Parameter	ROI_θ	ROI_α	ROI_β	ROI_γ	**Meaning**
Cep	54	54	54	54	Internal Connectivity Constant
Cpe	54	54	54	54	“
Csp	54	54	54	54	“
Cps	67.5	450	67.5	67.5	“
Cfs	15	10	27	27	“
Cfp	27	35	54	108	“
Cpf	300	300	540	300	“

Cff	10	25	10	10	“
$\boldsymbol{\omega_e}$	75 s^{-1}	66 s^{-1}	68.5 s^{-1}	125 s^{-1}	Reciprocal of time constant of excitatory
$\boldsymbol{\omega_s}$	30 s^{-1}	42 s^{-1}	30 s^{-1}	30 s^{-1}	Reciprocal of time constant of inhibitory slow
$\boldsymbol{\omega_f}$	300 s^{-1}	300 s^{-1}	300 s^{-1}	400 s^{-1}	Reciprocal of time constant of inhibitory fast

3.2.7 Supplementary Material part 2

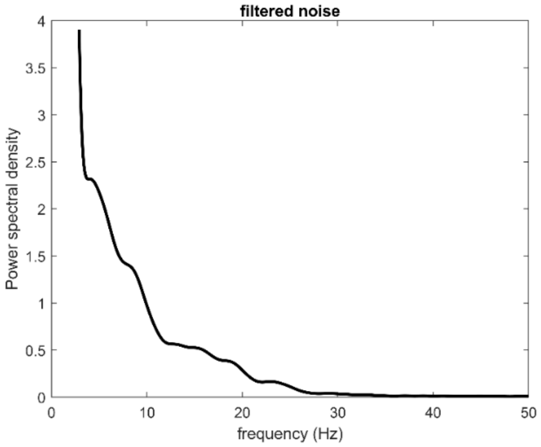


Figure 3.12S – Power density spectrum of the noise, filtered by the dynamics of the glutamatergic synapses and multiplied by the constant *Cpe* (see also Eqs. (3.4), (3.5) and (3.6)).

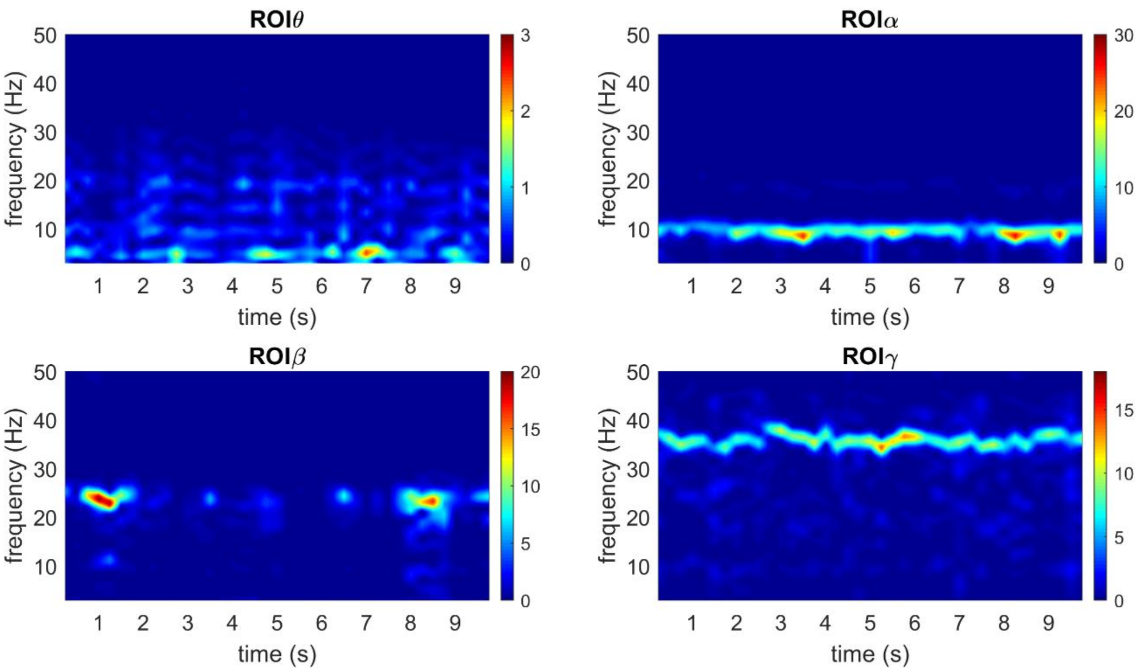


Figure 3.13S – Spectrograms of the potential for the pyramidal population in the four different ROIs simulated with the connectivity as in Fig. 3.12a.

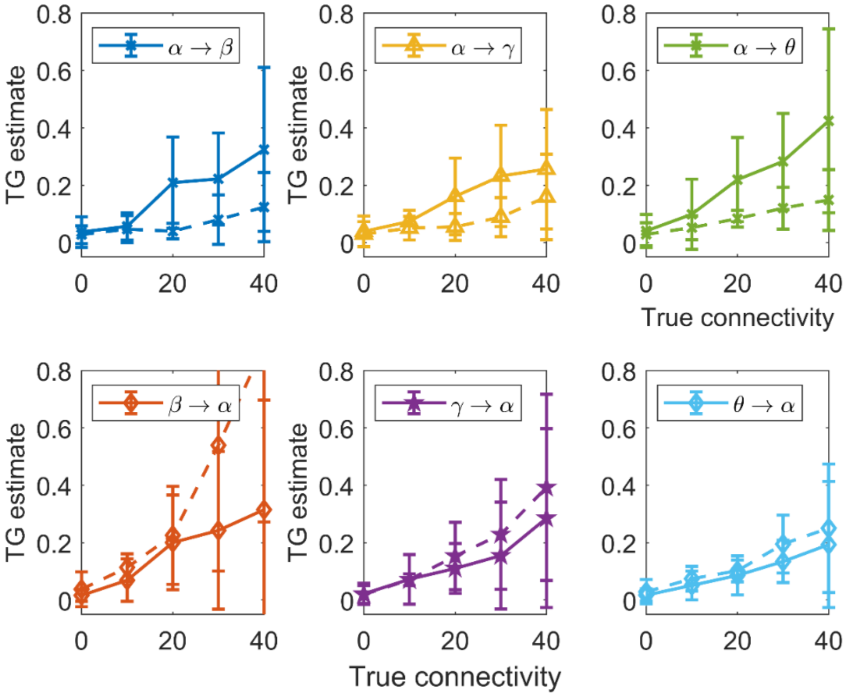


Figure 3.14S– Relationship between the connectivity estimated with the Temporal Granger estimator and the true connectivity, obtained using the data obtained from 100 randomly generated networks. The continuous lines were obtained using all inputs to the ROIs as high as 400 (the same as in the simulations in Fig. 3.14 and 3.15 in the text). The dashed lines have been obtained by reducing the input to ROIα down to 200, while all other inputs are unchanged. Only the connections entering into (upper panels) or exiting from (bottom panels) the ROIα are shown, since the others are not significantly affected. It is worth noting that a decrease in the input to a ROI reduces the estimated values of the output connectivity, and increases the entering estimated connectivity.

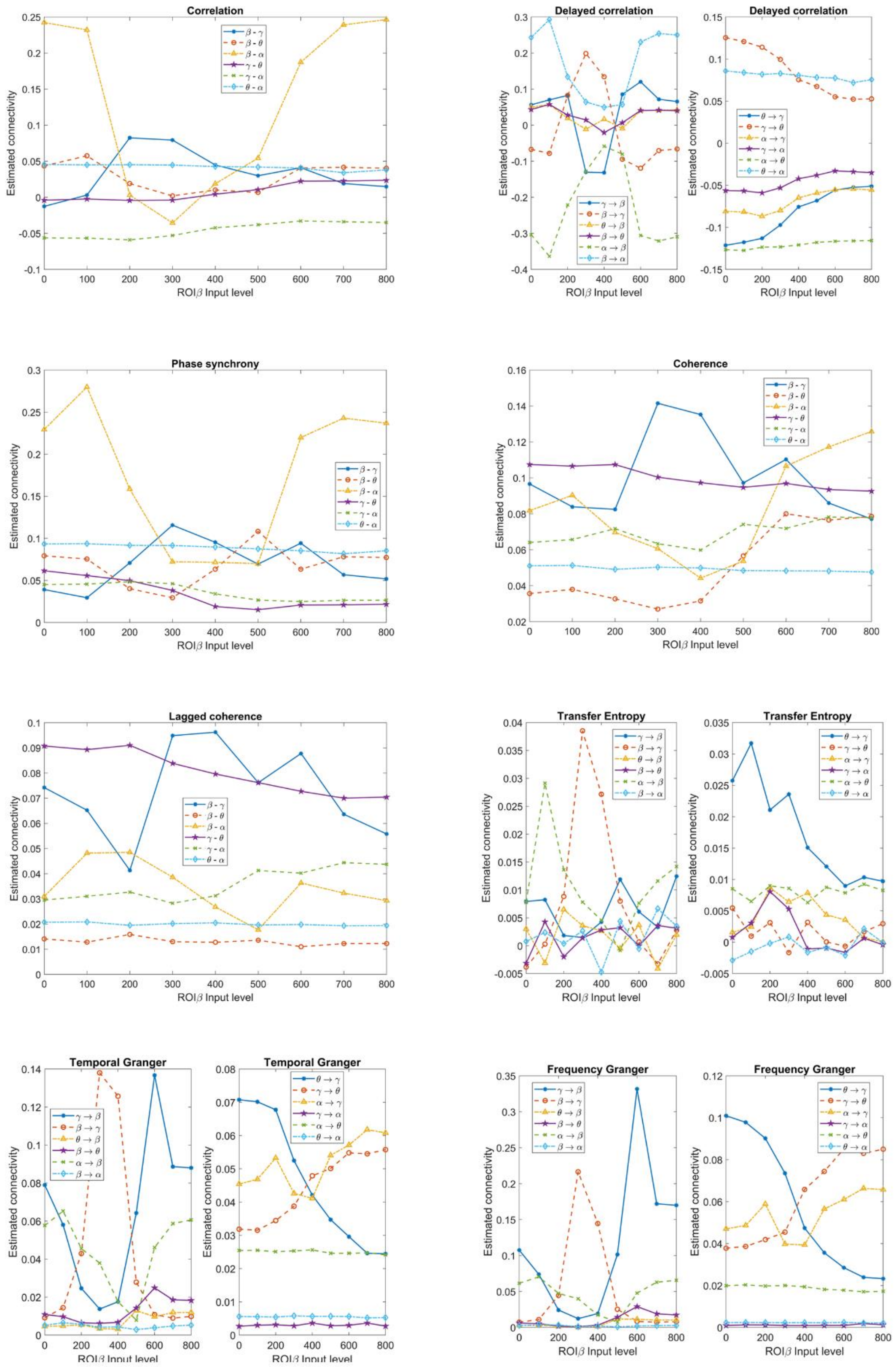


Figure 3.15S - Values of connectivity among the ROIs estimated with the eight different FC estimators, with reference to the network in Fig. 3.12A, when the *input* to pyramidal neurons in ROIβ is progressively varied from 0 to 800, as in the x-axis, and all other inputs and connections are maintained at the basal value as in Fig. 3.12A. It is worth noting the strong effect that the input change has on the connections which affect the ROIβ. This result, although with some differences, is evident with all estimators.

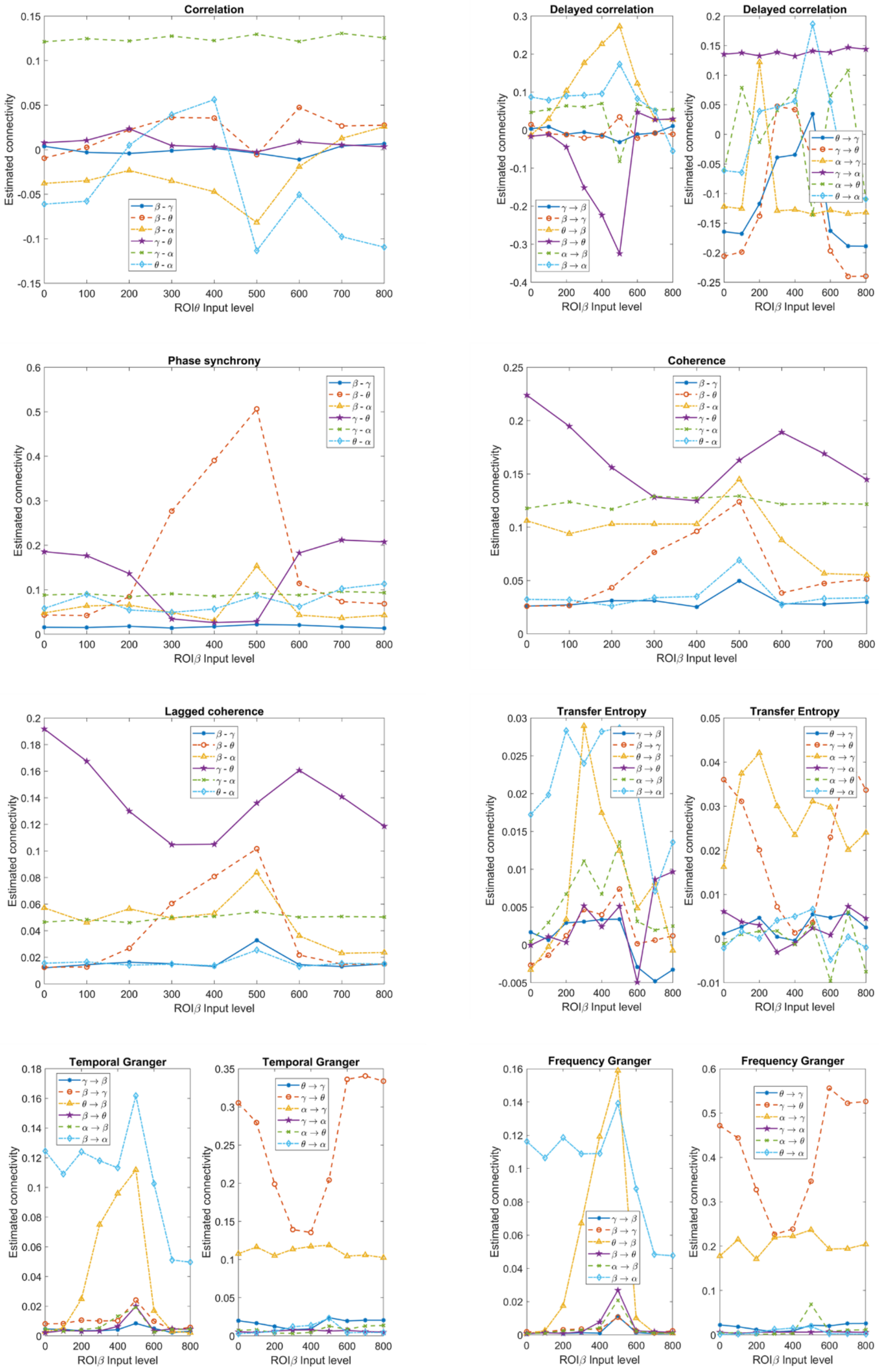


Figure 3.16S - Values of connectivity among the ROIs estimated with the eight different FC estimators, with reference to the network in Fig. 3.12B (i.e., a circular connections), when the *input* to pyramidal neurons in ROIθ is progressively varied from 0 to 800, as in the x-axis, and all other inputs and connections are maintained at the basal value as in Fig. 3.12B. It is worth noting the strong effect that the input change has on the connections which affect the ROIθ. This result, although with some differences, is evident with all estimators.

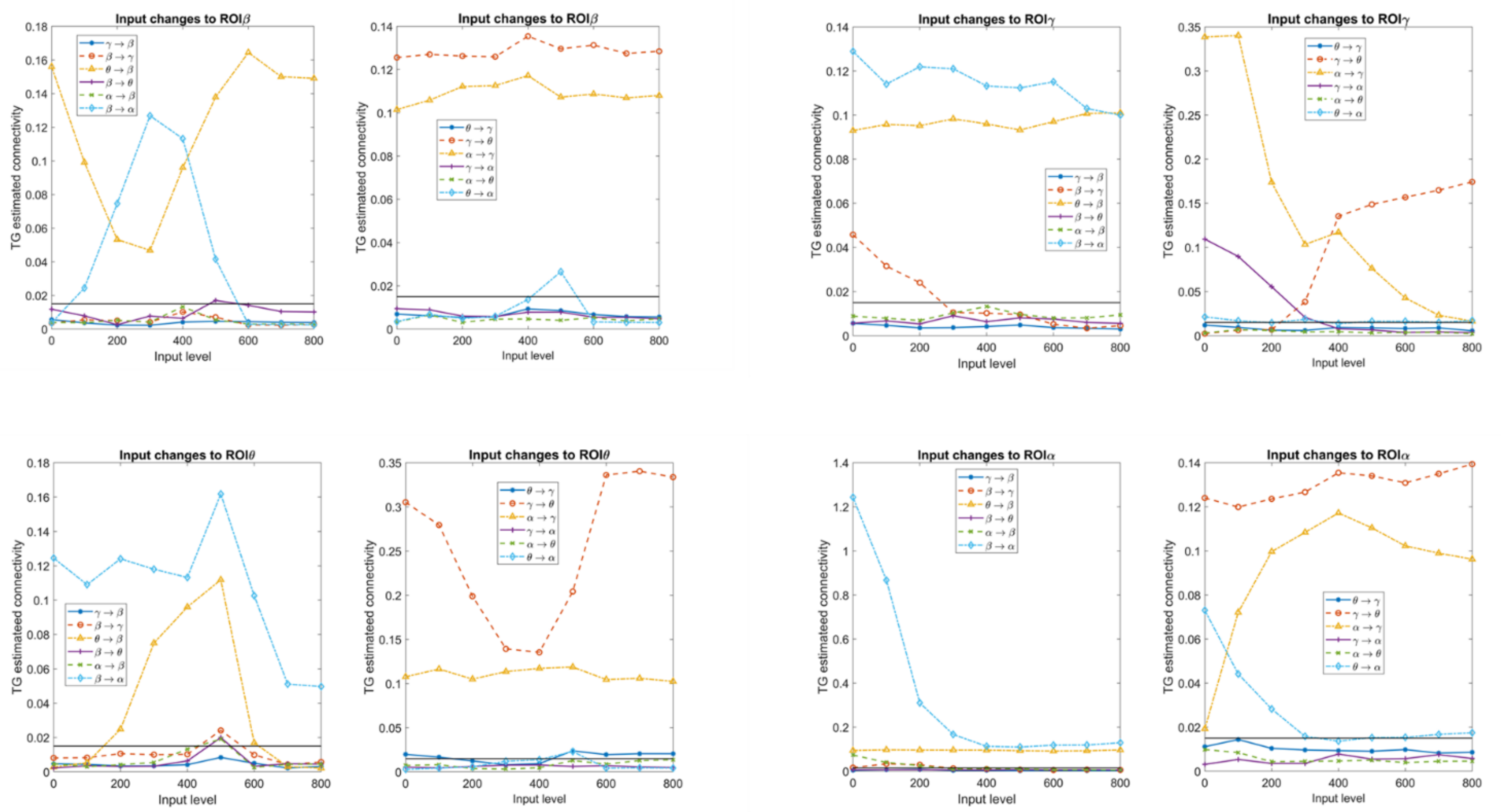


Figure 3.17S - Values of connectivity among the ROIs estimated with the Temporal Granger estimator, with reference to the network in Fig. 3.12B, when the input to pyramidal neurons in ROIα (panel A), ROIβ (panel B), ROIγ (panelC) and ROIθ (panel D) is progressively varied as in the x-axis, and all other inputs and connections are maintained at the basal value as in Fig. 3.12B. It is worth noting the strong influence that the input change has on the connections which enter into and exit from the affected ROI.

3.3 Assessment of Motor Network connectivity of stroke patient through NMMs

The study reported in this chapter refers to the published journal paper entitled "A Novel Method to Assess Motor Cortex Connectivity and Event Related Desynchronization Based on Mass Models", Mauro Ursino[1*], Giulia Ricci[1], Laura Astolfi[2,3], Floriana Pichiorri[3], Manuela Petti[2,3 §] and, Elisa Magosso[1 §], *Brain Sciences* (2021).

In this study, nonlinearities of NMMs were used to simulate the task-dependent connectivity network of a stroke patient during three different conditions: resting condition, movement of the affected and movement of the unaffected hand. Model parameters were fitted to the subject's source-reconstructed EEG data from primary motor areas (M1s), premotor cortices (PMCs) and supplementary motor areas (SMAs). The innovative aspect of this study lies in the ability of the non-linear model to simulate three different conditions by varying the input to the cortical regions, using a single set of connectivity parameters.

***Background**: Knowledge of motor cortex connectivity is of great value in cognitive neuroscience, in order to provide a better understanding of motor organization and its alterations in pathological conditions. Traditional methods provide connectivity estimates which may vary depending on the task. This work aims to propose a new method for motor connectivity assessment based on the hypothesis of a task-independent connectivity network, assuming nonlinear behavior. **Method**: The model considers six cortical regions of interest (ROIs) involved in hand movement. The dynamics of each region is simulated using a neural mass model, which reproduces the oscillatory activity through the interaction among four neural populations. Parameters of the model have been assigned to simulate both power spectral densities and coherences of a patient with left-hemisphere stroke during: resting condition, movement of the affected and movement of the unaffected hand. **Results:** The presented model can simulate the three conditions using a single set of connectivity parameters, assuming that only inputs to the ROIs change from one condition to the other. **Discussion**: The proposed procedure represents an innovative method to assess a brain circuit, which does not rely on a task-dependent connectivity network, and allows brain rhythms and desynchronization to be assessed on a quantitative basis.*

3.3.1 Introduction

Any movement is the result of the interaction among several brain regions, which are mutually interconnected via excitatory and inhibitory links, and whose interplay governs motor preparation and execution. Understanding how these regions work together, and establishing a reliable connectivity network is a matter of intense study in neuroscience today, to improve our comprehension of different aspects of motor behavior. Moreover, it is well known that this motor network is altered in pathological conditions, particularly after stroke, leading to abnormal interactions not only in the lesioned area but also in remote regions [265–267].

A wealth of studies in recent years attempted to quantify the motor network in both healthy subjects and patients, using model-based approaches: in these studies connectivity is estimated starting from an explicit model of causal inference, usually expressed in terms of state space differential equations [222,223,268–270]. Most previous works use a bilinear state space model (see [268]) which incorporates an intrinsic (i.e., task-independent) connectivity matrix, a task-dependent connectivity matrix, explaining the changes in neuronal states during the respective task, and a matrix for the experimental inputs that drive regional activity.

Besides works that investigated motor network in health and disease using causal models, other inferred connectivity by using data-driven approaches, such as correlation, coherence, Granger causality, Direct Transfer Function, Partial Directed Coherence (PDC), Mutual Information or Transfer Entropy (TE)” [271–274]. In general, model-based and data-driven methods can be considered as complementary approaches due to their respective advantages and limitations, and the comparison between their results can provide a deeper understanding of brain functions[16].

A typical result of the studies mentioned above is that the estimated connectivity network changes as a function of the performed task. Generally, during unimanual movements, connectivity towards the contralateral primary motor cortex (M1) is increased, whereas connectivity towards the ipsilateral motor areas is reduced [268,275]. Performing hand movements at higher frequency is associated with a linear increase in connectivity strength [222]. Differences are also evident by comparing motor execution vs. motor imagery for the same task: in these cases, modifications involve both some connectivity weights in the premotor cortices (PMCs) and supplementary motor areas (SMAs), and the inputs to these regions [223,225].

Although such works significantly extended our knowledge of the motor network implicated in movement planning and execution, we call attention to two main limitations that deserve a critical analysis.

First, as specified above, most studies accept the idea that connectivity may dramatically change between one task and another. This point is certainly acceptable if one refers to the amount of correlation or mutual information between two signals. Depending on the particular task, in fact, one area can transmit more or less information to another area, conditioned by the level of activities and non-linear phenomena involved. On the other hand, structural causal connectivity (defined as the existence of anatomical connections physically linking brain regions) cannot exhibit such large variations in a task-dependent fashion, and in a brief time scale. In recent studies [276], using a neural mass model to generate reliable signals in an interconnected network, we demonstrated that the estimated functional connectivity can vary dramatically, even in the presence of the same model network, depending on the presence of non-linear phenomena (such as saturation in neural activity). Our idea is that linear models might overestimate the task-dependent component and underestimate the role of structural links, which remain stable across tasks.

Second, it is well-known that motor execution and motor imagery are based on variations of the neural synchronization in specific brain rhythms, especially in the alpha and beta ranges [238,277]. These rhythms show two main characteristic spatiotemporal patterns during motor processing: a reduction of power in the beta range (event-related desynchronization, ERD) during motor preparation and performance, which can be considered as a correlate of an activated cortical area [278,279]; an increase in power (event-related synchronization, ERS) during motor suppression, characteristic of a deactivated cortical area or inhibited network [280]. In particular, in patients with stroke, the study of ERD and ERS is important not only to characterize their response but also to define potential motor rehabilitation interventions via Brain Computer Interface technology [281].

Clearly, the variations in the neural synchronization - which in turn result in the changes in power in the alpha and beta ranges - are caused by the underlying connectivity between neurons and regions, and determine the consequent motor execution. However, just a few studies focused on the relation between the mechanism underlying this system of rhythms and the motor networks.

Aim of this study is to present a different approach to the problem of model-based connectivity which, although at a preliminary stage (i.e., a proof of concept), can have profound future implications. The distinctive idea here is to ascribe most of the connectivity differences observed between resting state and motor tasks to non-linearity in the neural signal processing, which cannot be grasped by traditional linear models. In other terms, we wish to test the hypothesis that coherence among neuroelectric signals in *different tasks*, as well as ERD and ERS, can be simulated using a *single* connectivity network (i.e. a task-independent network) assuming that the interconnected Regions of Interest (ROIs) exhibit an oscillatory pattern and that these oscillations are non-linearly transmitted from one region to the other, generating neuronal behaviors supporting different tasks.

Once this network has been estimated, our aim is to summarize brain rhythm power changes during different tasks and their causal relationships into a single theoretical framework, to help understanding the possible underlined neural mechanisms.

To reach these objectives and provide a proof-of-concept of this approach for motor connectivity, we simulated a network consisting of six ROIs (M1, PMC and SMA in both hemispheres) connected via excitatory and/or inhibitory links. The neuro-electrical activity in each ROI is simulated using a Neural Mass Model (NMM) developed by the authors in past years [161], able to generate multiple rhythms in the alpha, beta and gamma bands. Parameters of the model are assigned to simulate the power spectral densities and coherences among the six ROIs, obtained from electroencephalographic (EEG) data in a patient with unilateral stroke, both in resting condition, and during movement of the affected and unaffected hand. A stroke patient has been chosen to point out whether the method, besides reproducing ERS, ERD and coherence, can also reveal differences in connectivity and in activation in the affected vs the unaffected hemisphere.

Finally, we compared our results with the motor networks captured by estimators based on data-driven approaches (temporal correlation, coherence, Partial Directed Coherence and Transfer Entropy), with the aim to critically discuss the link between task-related functional networks and the network derived from our method based on the NMM, possibly revealing similarities between complementary approaches as well as aspects grasped by the present method not emerging in the functional networks.

3.3.2 Material and Methods

3.3.2.1 Experimental data: acquisition, processing and connectivity estimates

3.3.2.1.1 Experimental protocol and EEG data measurement

Data subjected to the present analysis were obtained from a previous study [282]. In brief, data were acquired from a stroke patient (male; 62 years; lenticular haemorragia; 1 month since event; left affected hemisphere) enrolled from the rehabilitation hospital ward at Fondazione Santa Lucia, IRCCS, in Rome, Italy. Upon enrolment, the patient was evaluated by means of the European Stroke Scale (ESS=75) [283] and the upper limb section of the Fugl-Meyer Assessment (FMA=40) [284].

The patient was subjected to a screening session during which electroencephalographic (EEG) signals were recorded during the execution of motor tasks. In particular, the patient was asked to execute a simple movement (sustained grasping movement) with the affected and unaffected hand in separate runs. Each run consists of 30 trials, 15 rest and 15 motor trials in randomized order: the task timing was determined by a movement of the cursor on the computer screen, while during rest trials, the patient was only asked to watch the cursor trajectory. EEG was collected from 61 standard positions (according to the extended 10–20 International System), band pass–filtered between 0.1 and 70 Hz, digitized at 200 Hz, and amplified by a commercial EEG system (BrainAmp, Brainproducts GmbH, Germany).

The experimental protocol described above (see [282] for more details) was approved by the local ethics board of the Fondazione Santa Lucia (Prot.CE/AG4-PROG.244-105) and written informed consent was obtained from the patient.

3.3.2.1.2 EEG preprocessing and source reconstruction

In the present study, EEG data were downsampled at 100Hz (with anti-aliasing filter) and bandpass filtered (1–45Hz). Ocular artifacts were removed by Independent Component Analysis (ICA) and residual artifacts (muscular, environmental, etc.) were removed using a semiautomatic procedure, based on the definition of a voltage threshold (±80μV). The preprocessed EEG signals were then segmented, considering the last 4 seconds of each trial as the period of interest. This procedure was performed for all the runs: rest, motor task performed with the affected hand and motor task performed with the unaffected hand.

The estimation of the neuroelectrical activity in the brain source space was obtained by solving the linear inverse problem according to the methods described in previous works [285–287]. The procedure is based on the use of an average geometry head model based on Colin template (T1-weighted MRI volume, Montreal Neurological Institute)[59], composed of around 8000 equivalent current dipoles disposed normally to the cortical surface, and allows estimating over time the signed magnitude of the dipolar moment for each cortical dipole. Then, at each time point, the magnitude derived from the centroid of a particular ROI was used as waveform of the cortical activity in that ROI.

In the present study we focused on 6 ROIs selected on the basis of their involvement in the tasks under investigation: SMA proper left hemisphere (SMAp L), SMA proper right hemisphere (SMAp R), M1 hand left hemisphere (M1h L), M1 hand right hemisphere (M1h R), dorsal premotor cortex left hemisphere (PMD L), dorsal premotor cortex right hemisphere (PMD R).

3.3.2.1.3 Model-free network analysis

In order to gain a preliminary insight into the network structure, we first estimated the connectivity among the 6 ROIs reconstructed in the patient, using four distinct estimators, i.e., the coherence, the temporal correlation, the Partial Directed Coherence and the Transfer Entropy. It is worth noting that the first two estimators estimate undirected links (i.e., the connectivity from node i to node j is equal to the connectivity from j to i) whereas the last two estimators are directional. Each estimator was applied separately to each experimental condition (rest, movement of the affected hand, movement of the unaffected hand).

Both *power spectral densities* of individual signals, and the *coherence* between each couple of signals were evaluated using the Welch's averaged periodogram method, using a Hamming window as long as 0.5 s, 50% overlapping, and zero-padding to have a line spectrum every 0.1 Hz.

Temporal correlation was evaluated by computing the Pearson's linear correlation coefficient between any couple of signals for each trial and then averaged across trials of the same experimental task.

The *Partial Directed Coherence* [116] is a linear spectral quantifier which reveals the existence, the direction and the strength of a functional relationship between any given pair of signals in a multivariate data set. In the present work, we used generalized Partial Directed Coherence (gPDC) [288], that modifies PDC to be scale invariant. The optimal order of the multivariate models was estimated by means of Akaike Information Criterion [289].

Transfer entropy (TE) is a model free implementation of Wiener's principle of observation causality [133]. In this work, TE was estimated using Trentool, a software package implemented as a Matlab toolbox under an open source licence [136]. In particular, we evaluated the so called "high-order" TE, i.e. by considering more than two time bins for the receiving time series and for the sending time series (the number of past bins used is called embedding dimension and is optimized by the software).

It is worth noting that all previous methods estimate a pattern of connectivity for each task, providing values that vary depending on the particular task (in this work, basal resting condition, movement of the affected hand and movement of the unaffected hand). Moreover, coherence and gPDC are spectral estimators, hence they provide one value for each frequency; in these two cases, to obtain a single value for each connection, we summarized the values in the beta range. The three task-dependent networks obtained with each method are shown in the Supplementary Material; in section Results (see section 3.3.3.1), a single network is reported for each method, obtained by averaging the three-task dependent networks, in order to summarize the main aspects of the motor network captured by the given method.

As detailed below (see section 3.3.2.2.2, Fitting procedure), the coherence values were exploited in defining initial guesses for model parameters. The other networks mainly serve for comparison with the connectivity network estimated with the proposed method.

3.3.2.2 The Neural Mass Model and model-based connectivity estimation

Please note that the description and the complete set of equations (Eq. 3.1-3.19) of NMMs used is described at the beginning of section 3 and is represented by Figure 3.1 of Chapter 3.

Our model-based connectivity was evaluated by fitting the outputs of a neural mass model of interconnected ROIs to the experimental data. In particular, we minimized a cost function of the difference between the normalized power spectra of experimental and simulated data, and of the difference between the coherences (see section below). The estimation procedure is designed to provide a unique set of connectivity for the three tasks.

In this work we considered three ROIs in the affected (left L) and unaffected (right R) hemispheres, representing the same ROIs reconstructed from EEG, that is M1h L and M1h R, SMAp L and SMAp R, PMD L and PMD R. We hypothesized that these ROIs can be connected according to the basic scheme shown in Fig. 3.24 (the justification is provided below); the values of these connections were then subjected to the fitting procedure and some of them can go to zero at the end of the fitting.

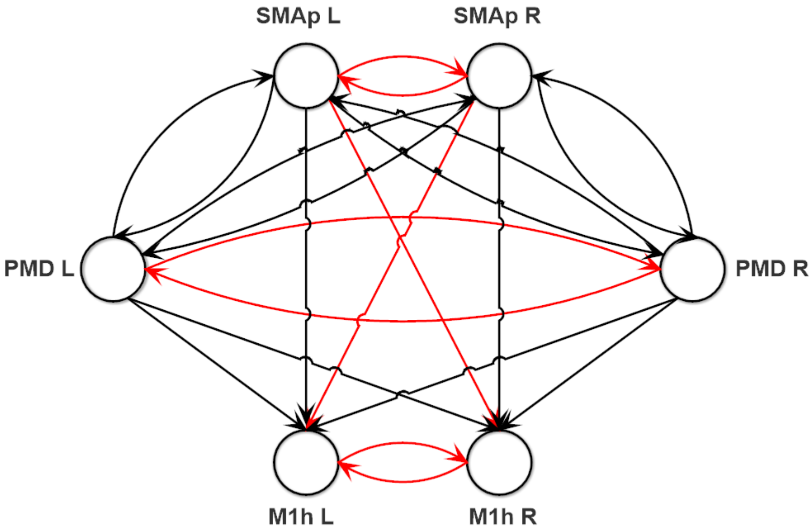


Figure 3.24 – General structure of the connections used to fit the neural mass model to the experimental data. It is evident the presence of feedback connections between the SMAps and the PMDs, and the presence of feedforward connections towards the M1hs. Black lines denote excitatory pyramidal-pyramidal connections, whereas red lines denote inhibitory bi-synaptic connections (pyramidal-fast inhibitory-pyramidal).

Numerical integration of the differential equations was performed with the Euler integration method, with a simulation step as low as 10^{-4} s. We tested the method's accuracy by performing some simulations with a much smaller step (hence longer computation times), without observing significant changes in the results. Each simulation lasted 11 seconds starting from a null initial value of the state variables. The first second of the simulation was then excluded to eliminate the initial transient response. Afterwards, data were passed through a low-pass antialiasing filter, re-sampled at 100 Hz (sampling period 0.01 s) and stored for subsequent processing. In particular, the output signals of the model are the post-synaptic membrane potentials of the pyramidal population in each ROI (i.e. quantity v_p in Eq. (3.4)), which is representative of local mean field potential. Finally, computation of power spectrum density and coherence was applied to these model signals. The computation time required to perform one single simulation (including spectra calculation) on a notebook (i7 last generation CPU) was in the range of about 80 seconds.

3.3.2.2.1 Parameter estimation method

Assumptions on parameters and network topology

We assumed that the following parameters in the model can be assigned on the basis of a fitting procedure between simulated signals and real data:

1) The 8 coefficients C_{ij} between the populations (Eqs. 3.4, 3.8, 3.12 and 3.18): These can reflect the number of internal contacts among populations within a ROI. We assumed that they can be different in the left and right hemisphere, even for the same regions, as a consequence of hand lateralization and, above all, of the Stroke effect. Hence, we have 8x6 = 48 parameters.

2) The reciprocal of time constants ω_e, ω_s, and ω_f (Eqs. 3.2, 3.6, 3.10, 3.14 and 3.16). However, to reduce the number of parameters, we assumed that the same regions in the left and right hemisphere have the same time constants. Hence, we have additional 3x3 = 9 parameters.

3) The synapses connecting the different ROIs (W_j^{hk} in Eq. 3.19), providing the model-based connectivity network, and the external inputs (m_j^k in Eq. 3.19). In order to identify a possible schema of synaptic connections among the ROIs, we started our analysis looking at data on the motor cortex connectivity in the existing literature [266,268,290]. Then, we assumed that:

3a – The same regions in the two hemispheres (i.e. M1h L vs M1h R; SMAp L vs SMAp R and PMD L vs. PMD R) are connected via inhibitory synapses according to the so-called Theory of Inhibition [291]. This theory assumes that inhibition occurs between the same function in the two hemispheres to prevent maladaptive cross talk and to allow a given function to become dominant [292–295]

3b – The connections between regions in the same hemisphere are excitatory. This assumption is strongly supported by previous studies of connectivity in the motor cortex [268,269].

3c - A more difficult problem concerns connections between one area in one hemisphere and a non-homologous area in the other hemisphere. While several studies suggest inhibition [268,269] other suggest that these connections are excitatory, especially when directed toward the performing hand [222,290]. We started from the basic schema depicted in Fig. 3.24, assuming that the connections from one SMAp and the contralateral M1h are inhibitory (this choice reproduces the pattern reported in [268]), whereas the others are excitatory. With just one exception, discussed below, this schema also substantially agrees with the signs obtained from the temporal correlation analysis (See Section 3.3.3.1 and Supplementary Material). In order to test also an alternative hypothesis, the fitting procedure was repeated (alternative fitting procedure), assuming that the synapses from one SMAp and the contralateral M1h are excitatory rather than inhibitory (maintaining unaltered the rest of the schema). Results of the last procedure are reported in Supplementary Material.

3d - In order to reduce the complexity of the fitting problem, we assumed that M1h L and M1h R receive feedforward synapses from the two SMAps and the two PMDs, but do not send feedback synapses back. This is certainly a strong simplification. As a partial justification, we can observe that the connectivity values estimated with a model-based approach in previous works[268] exhibit weaker feedback connections from M1 to SMA and from M1 to PMC than the corresponding feedforward connections. However, the fundamental reason to adopt this simplification is to drastically reduce the complexity of the fitting procedure. In fact, thanks to this assumption, we can first assign parameters to the four regions PMD L, PMD R, SMAp L and SMAp R, connected via reciprocal feedback links, and subsequently to assign parameters for the two M1h regions and the feedforward synapses targeting them. In other terms, the problem is split into two sub-problems, resolved in two separated steps of the fitting procedure (Step 1 and Step 2, see below). This simplification can be removed in future works.

3e - Finally, we assume that the two SMAps and the two PMDs receive an external input (Gaussian noise with a given mean value and assigned variance), impacting on the population of pyramidal neurons. These terms represent all other external sources not included in the model. All input mean values are set to zero in resting conditions, but these values can increase to a positive value during a task execution (left hand or right hand movement) when the regions are further excited, and may contribute to move the working point of the ROIs along their sigmoidal characteristic outside the central linear region. These 8 values (4 mean values during movement of the affected hand and 4 mean values during movement of the unaffected hand) are additional parameters in the fitting procedure. It is worth noticing that the mean inputs to SMAps and PMDs are the only model parameters that assume different values across the three tasks.

All previous hypotheses will be critically discussed in the last section.

In conclusion the total number of estimated parameters for step 1 and step 2 of the fitting procedure are:

Step 1 (PMD L, PMD R, SMAp L, SMAp R): 32 internal constants C_{ij}, 6 time constants ω_{ij}, 4 inhibitory synapses W_f, 8 excitatory synapses W_p, and the 8 input values m_p. Total: 58 parameters.

Step 2 (M1h L and *M1h R):* 16 internal constants C_{ij}, 3 time constants ω_{ij}, 4 inhibitory synapses W_f, 6 feedforward excitatory synapses W_p. Total: 29 parameters (in the alternative fitting procedure (See Supplementary Material), 2 inhibitory synapses W_f and 8 excitatory synapses, W_p, still 29 parameters).

Fitting procedure

In the following we will use the symbols $P_{spe,b}^{ROI}(j2\pi f_k)$, $P_{spe,a}^{ROI}(j2\pi f_k)$, $P_{spe,u}^{ROI}(j2\pi f_k)$ to denote the power spectral densities of the experimental data (subscript *spe*) for a given ROI. The other subscript refers to the basal condition (subscript *b*), movement of the affected hand (subscript *a*) and movement of the unaffected hand (suscript *u*), respectively. f_k is the k-th frequency of the spectrum, computed with the Welch periodogram method. Similarly, we will denote with symbols $P_{\text{mod},b}^{ROI}(j2\pi f_k,\theta)$, $P_{\text{mod},a}^{ROI}(j2\pi f_k,\theta)$, $P_{\text{mod},u}^{ROI}(j2\pi f_k,\theta)$ the spectral densities of the simulated data with reference to the same ROI and at the same frequencies, where θ are the model estimated parameters (i.e. these are the power spectral densities of the post-synaptic potential v_p (Eq. 3.4) of the given ROI in the model). It is worth noting that, in order to allow a direct comparison between experimental and simulated spectra, and to account for ERD, all spectra have been normalized with respect to the maximum of the spectrum in basal condition in the same ROI. By way of example, by denoting with $\tilde{P}_{\text{mod},a}^{ROI}(j2\pi f_k)$ the model spectral density in a ROI during a movement of the affected hand, *without normalization*, we have: $P_{\text{mod},a}^{ROI}(j2\pi f_k)=\dfrac{\tilde{P}_{\text{mod},a}^{ROI}(j2\pi f_k)}{\max\left\{\tilde{P}_{\text{mod},b}^{ROI}(j2\pi f_k)\right\}}$. A similar normalization has been performed for all power spectral densities, both simulated and experimental. Of course, after these normalizations, all spectra in basal conditions have a maximum as large as 1. ERD is evident during hand movement, both of the affected and unaffected hand, by a normalized spectrum having a maximum smaller than 1.

As said above, the fitting procedure has been divided in various steps:

1) Step 0 (preliminary step). First, we estimated a preliminary value to the eight internal parameters C_{ij} in each ROI, and to the reciprocal time constants, ω_j, in order to simulate the power spectral density in basal resting condition in the range 10-30 Hz. All power densities (experimental and simulated) were previously normalized to their maximum. Fitting was achieved by minimizing the following least square cost function

$$F_0(\theta) = \sum_k \left(P^{ROI}_{mod,b}(j2\pi f_k,\theta) - P^{ROI}_{spe,b}(j2\pi f_k)\right)^2 \qquad (3.32)$$

where the sum is extended to all spectral frequencies in the range 10-30 Hz (frequency step Δf = 0.1 Hz), and θ is the vector of *internal parameters* in the given ROI.

In this preliminary step, each ROI is considered separately from the other, i.e., without any connectivity. The only constraint is that we assumed identical ω_j for the homologous ROIs. The estimated parameters are considered as an initial guess for the subsequent steps.

2) Step 1. Subsequently, we estimated all parameters of the two SMAps and of the two PMDs, including their connectivity weights, according to the diagram in Fig. 3.24. The initial guesses of parameters C_{ij} in the four ROIs and of the reciprocal time constants, ω_j, were those obtained in step 0. The initial guesses for the connectivity weights among these four ROIs are described at the end of this section (see below). In this step, we minimized a more complex cost function than in Step 0, to reproduce both the changes in power spectral densities between the three tasks (basal condition, affected hand movement, and unaffected hand movement), and the coherence among the ROIs in basal condition, within the band 10-30 Hz.

By denoting with $C^{ROI1,ROI2}_{spe,b}(j2\pi f_k)$ and $C^{ROI1,ROI2}_{\mathrm{mod},b}(j2\pi f_k,\theta)$ the coherences between the signals in ROI1 and ROI2 (with ROI1 ≠ ROI2) computed in basal conditions from the experimental data and from model simulations, respectively, we have

$$\begin{aligned}F_1(\theta) = &\sum_{tr=b,a,u}\sum_{ROI}\sum_k \left(P^{ROI}_{mod,tr}(j2\pi f_k,\theta) - P^{ROI}_{spe,tr}(j2\pi f_k)\right)^2 + \\ &+ \sum_{\substack{ROI1\ ROI2 \\ ROI1\neq ROI2}}\sum_k \left(C^{ROI1\ ROI2}_{mod,b}(j2\pi f_k,\theta) - C^{ROI1ROI2}_{spe,b}(j2\pi f_k)\right)^2 \\ &+100\sum_{tr=a,u}\sum_{ROI}\left[max\{P^{ROI}_{mod,tr}(j2\pi f_k,\theta)\} - max\{P^{ROI}_{spe,tr}(j2\pi f_k)\}\right]\end{aligned} \qquad (3.33)$$

The first term in the right hand member represents the square difference of all normalized spectra, computed in the four ROIs and in all trials (basal, affected and unaffected); the second term is the square differences of model vs experimental coherences, extended to all couples of the four ROIs, computed only in basal conditions; the third term is the differences between the maxima of the spectra, computed in all ROIs, both in the affected and unaffected trials (indeed, since maxima are equal to 1 in basal conditions, this term is zero in the basal case and is not included in the sum). The multiplicative factor, 100, has been introduced so that the last term in Eq. (3.33) has approximately the same weight in the cost function as the first two terms.

3) Step 2. In this step, we estimated all parameters of the two M1hs, including the feedforward connectivity weights from the SMAps and from the PMDs to the primary motor areas (see Fig. 3.24 again). In this case too, initial guesses of parameters C_{ij} in these two ROIs and of the reciprocal time constants, ω_{ij}, were those obtained in Step 0, while the initial guesses for the connectivity weights followed a procedure similar as in Step 1 (see below). The inputs to the feedforward synapses were the spike densities of the previous regions, computed with the optimal parameters estimated in Step 1. The cost function was similar to that used in Step 1 (with the summations on the ROIs extended only to the two M1hs), reflecting both the changes in power spectral densities between the three tasks (basal condition, affected hand movement, and unaffected hand movement), the coherence among the ROIs in basal condition within the band 10-30 Hz, and the maximum of the normalized spectra in the affected and unaffected conditions.

In particular, it is worth noting that in Step 2 the cost function uses only coherence between M1h L and M1h R. The coherences between the M1hs and the SMAps and between the M1hs and the PMDs were never considered in the fitting procedure (either in basal condition or in affected/unaffected hand movement), for the sake of computational simplicity. Hence, all these coherences are posterior predictions. Furthermore, also coherences between all other ROIs in the two motor tasks (with the affected and unaffected hand) are posterior predictions, since only the coherences in basal condition were used in the cost function.

An initial guess for the connection weights among the four ROIs (PMD L and R, SMA L and R) in Step 1 was given on the basis of the coherence values in the beta band. However, since the fitting procedure generally stops at a local minimum, several different initial guesses were produced by adding noise to the parameters (± 50% of its value), so as to obtain several local minima. Then we chose the best local minimum, and started the procedure again, by adding noise to the parameters so obtained. The procedure was stopped when we arrived at a final simulation that reproduces the spectra and coherence quite well. This was a qualitative choice, based on a visual inspection of the results.

A similar procedure was adopted in Step 2, for what concerns the estimation of the weights entering into the M1h L and M1h R.

All computations were performed using the scientific software environment Matlab (version R2018b Mathworks ©). Minimization was achieved using a direct pattern search algorithm (Matlab command “patternsearch” in the global optimization toolbox), which finds a local solution of the optimization problem without using any information about the gradient of the objective function

3.3.3 Results

3.3.3.1 Model-free connectivity estimation

Fig. 3.25 shows the motor networks obtained with the *coherence* (upper left), the *gPDC* (upper right), the *TE* (bottom left) and the *correlation coefficient* (bottom right), where the thickness and style of lines reflects the connectivity strength. Specifically, each connection was obtained by averaging the values estimated on the three tasks (the three single networks are displayed in the Supplementary Material II), and we show only connections characterized (on average) by values higher than a given threshold (see legend for more details). Results show how connections *within the hemisphere* are strong between the SMAps and the homolateral PMDs and between the SMAps and the homolateral M1hs (especially in the affected side left), but weak between the PMDs and M1hs. Lateral *interhemispheric connections* are strong between SMAP L and SMAP R, and between PMD L and PMD R, but are negligible between the two M1hs. Finally, we can observe the presence of *interhemispheric connections* between the SMAps and the contralateral M1h with all estimators. Moreover, the gPDC also underlines the presence of connections directed from the SMAps towards the contralateral PMDs (which were not evident with coherence). Compared with the others, TE exhibits a greater number of reentrant connections, which were not evident in the previous estimator: in particular, we can observe connections from the M1h L back to both SMAps. As explained above, we neglected these connections originating from the M1hs when performing the connectivity estimation with the NMM. For what concerns the Pearson correlation coefficient, a fundamental difference compared with other estimators is that its value can be positive or negative, hence, this information can provide a discrimination about the presence of excitatory or inhibitory bi-synaptic connections (see also [276]). A problem, however, is that two signals can have either positive or negative correlation depending on the orientation used during source reconstruction. In other terms, all reconstructed signals have an arbitrary sign. To overcome this problem, we introduced two assumptions in the computation of the correlation coefficients, which allow the choice of the signs, but have no effect on the absolute value: i) Two homologous areas in different hemispheres inhibit reciprocally, according to the Brain Inhibition Hypothesis (hence, they should have a negative correlation). ii) Regions in the same hemisphere are connected via excitatory connections (see [268], hence they should have positive correlation). Starting from these assumptions, and by fixing arbitrarily the sign of one signal (for instance PMD R), we were able to provide a sign for all signals in all ROIs. In particular, we fixed the sign to SMAp R and M1 R to have positive correlation with PMD R, and the sign to all regions in the left hemisphere to have negative correlation with the homologous regions in the other hemisphere. All other correlation coefficients were not fixed a priori, and were obtained a posteriori: this provides a new information on the connection type (excitatory or inhibitory) which could not be obtained using TE, coherence or gPDC estimators (Fig. 3.25 right bottom panel).

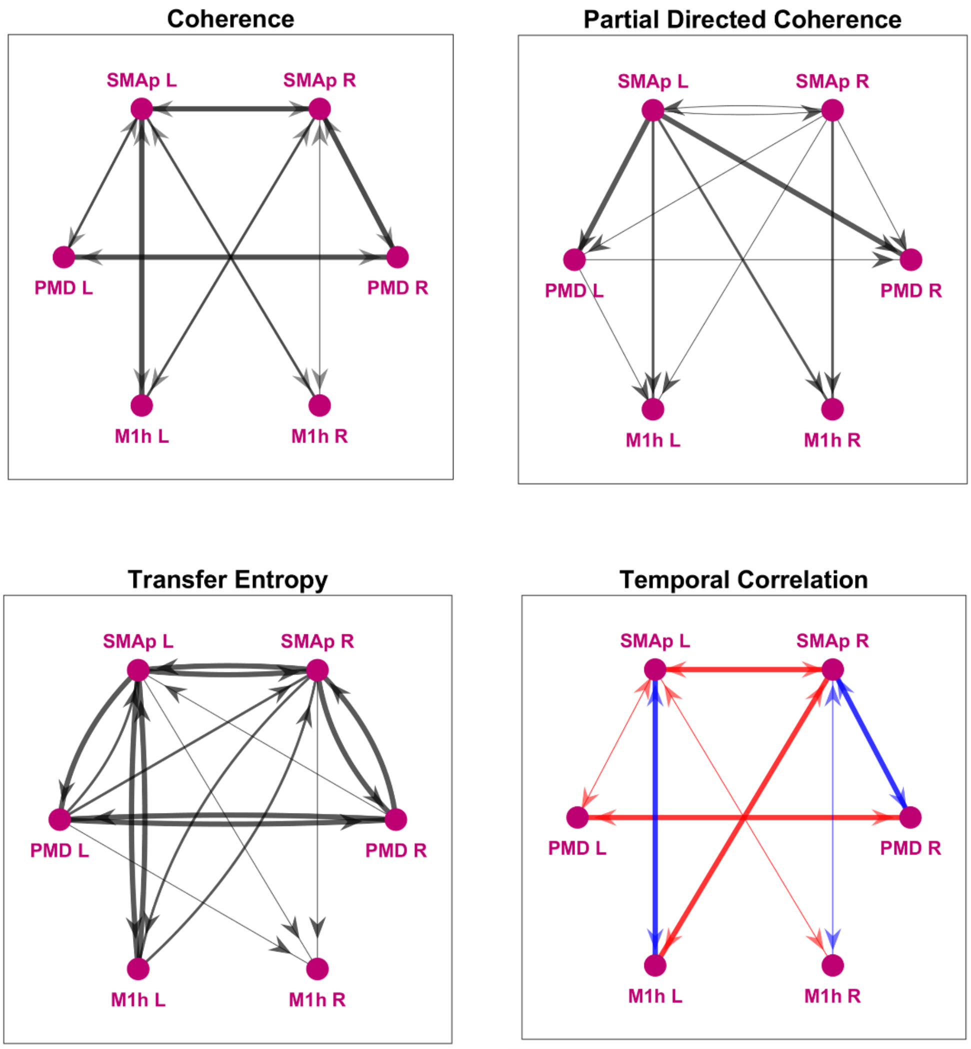


Figure 3.25 – Functional connectivity networks obtained using model-free connectivity estimation methods. All the displayed networks are obtained by averaging the estimates over the three tasks. *Coherence* (upper left panel) is a symmetrical quantity (i.e., it is the same in both directions), therefore all connections are bidirectional: only connections with maximum coherence value (averaged over the tasks) above 0.3 within the beta band are shown. Thick lines denote a maximum coherence greater than 0.5, medium lines a maximum coherence in the range 0.4 – 0.5, and thin lines a maximum coherence in the range 0.3 – 0.4. - *gPDC* (upper right panel) is a directional quantity: only connections with a *g*PDC value higher than 0.01 (averaged over the three trials) within the beta band are shown. Line thickness is proportional to the mean *g*PDC level (thin lines: *g*PDC < 0. 1; medium lines: 0.1 ≤ gPDC < 0.3; thick lines: gPDC > 0.3). *TE* (bottom left panel) is also a directional quantity**:** only connections with TE value (on average) higher than 20% of the maximum TE (maxTE = 0.023) are shown. Line thickness is proportional to the mean TE level (thin lines: TE between 20% and 30% of MaxTE; medium lines: TE between 30% and 50% of MaxTE; thick lines: TE greater than 50% of maxTE). *Pearson's linear correlation* (bottom right panel) is a symmetrical quantity, therefore all connections are bidirectional: only connections with correlation coefficient absolute value above 0.3 are shown. Here, colors specify whether the bidirectional connection is *excitatory* (blue, positive correlation coefficient) or *inhibitory* (red, negative correlation coefficient). It is worth noting that the sign of the correlation never changed from one task to another. Thick lines denote a Pearson correlation coefficient greater than 0.5, medium lines a correlation coefficient in the range 0.4 – 0.5, and thin lines a correlation coefficient in the range 0.3 – 0.4.

3.3.3.2 Parameter estimation with the NMM model.

It is worth noting that the signs of all connections in the right bottom panel of Fig. 3.25 agree with the previous hypotheses (see "Assumptions on parameters and network topology" in Section 3.3.2.2.2), with the only exception of the connection SMAp L – PMD L which turns out weakly negative. However, we decided to maintain the assumption of positive connection within a hemisphere, which seems more physiological, hence we used an excitatory synapse for this connection during the fitting procedure. Furthermore, the results show that the correlations between each SMAp and the contralateral M1h are negative, suggesting the presence of an inhibitory link (see also Fig. 3.26 in [268]).

Starting from the general network structure delineated in Fig. 3.24, we estimated all parameters in order to reproduce the normalized spectral densities and coherences in the beta range (see "Fitting procedure" in Section 3.3.2.2.2). Parameters not subject to fitting can be found in Table 3.3. These values are the same as in previous works [161,194] and identical for all ROIs.

Table 3.3 - Parameters assumed fixed for all tasks

Parameter	**value**	**meaning**
e_0	2.5 Hz	Saturation of the sigmoid
r	0.56 mV^{-1}	Parameter related with the central slope of the sigmoid
T	16.6 ms	delay
G_e	5.17 mV	Synaptic gain excitatory
G_s	4.45 mV	Synaptic gain inhibitory slow
G_f	57.1 mV	Synaptic gain inhibitory fast

Results of Step 1 of the fitting procedure (concerning ROIs SMAp L, SMAp R, PMD L, PMD R) are reported in Table 3.4 and Table 3.5, listing the estimated internal parameters of each ROI, and of the estimated inputs, respectively, while the estimated connectivity parameters between these ROIs are given in the network diagram of Fig. 3.26, upper panel. It is worth noting that a single parameter set (Table 3.4 and Fig. 3.26) is used for the three tasks, i.e., the three tasks differ only for the mean input reaching these ROIs (Table 3.5). Interestingly, the connectivity network resembles the networks estimated with the data-driven methods. Fig. 3.27 shows the normalized power spectral densities in the four ROIs (SMAp L, SMAp R, PMD L, PMD R). The left picture in each panel represents model simulations, while the right picture shows the experimental spectra. The model can simulate the ERD observed in each ROI during the two tasks quite well. Second, ERDs are quite different in the two PMDs and in the SMAps.

A strong desynchronization is evident in both PMDs during the movement of the unaffected hand, especially in the right hemisphere (the one not affected by the stroke). Conversely, desynchronization is less evident during movement of the affected hand, probably as a consequence of stroke. ERD is less evident in the two SMAps, and is just a little stronger during movement of the affected hand.

Table 3.4 - Internal parameters estimated on ROIs: SMAp L, SMAp R, PMD L and PMD R. It is worth noting that ω are the same for the two SMAps and for the two PMDs.

Parameter	SMAp L	SMAp R	PMD L	PMD R	Meaning
ω_e	76.14 s^{-1}	76.14 s^{-1}	62.97 s^{-1}	62.97 s^{-1}	Reciprocal of a time constant
ω_s	33.95 s^{-1}	33.95 s^{-1}	24.07 s^{-1}	24.07 s^{-1}	"
ω_f	336.8 s^{-1}	336.8 s^{-1}	734.9 s^{-1}	734.9 s^{-1}	"
C_{ep}	34.90	5.55	47.41	26.26	Internal connectivity constant
C_{pe}	12.02	5.46	29.04	50.73	"
C_{sp}	13.94	53.58	78.70	227.61	"
C_{ps}	6.92	53.98	68.80	123.99	"
C_{fs}	10.38	5.25	18.52	4.62	"
C_{fp}	45.02	40.91	80.80	55.06	"
C_{pf}	39.06	28.36	34.24	72.65	"
C_{ff}	22.83	5.67	5.44	4.74	"

Table 3.5 - Mean value of the external excitatory inputs estimated on ROIs: SMAp L, SMAp R, PMD L and PMD R during the three tasks (values at rest were set to zero).

Parameter m_j^h	Rest	Movement affected hand	Movement unaffected hand	Meaning
$m_p^{SMAP\ L}$	0	24.66	0	Input mean value to a ROI
$m_p^{SMAP\ R}$	0	190.29	111.87	"
$m_p^{PMD\ L}$	0	277.28	0	"
$m_p^{PMD\ R}$	0	21.09	482.99	"

Table 3.6 - Internal parameters estimated on ROIs: M1h L, M1h R, during the first fitting procedure, i.e. assuming that the connections between each SMAp and the contralateral M1h are inhibitory in type. It is worth noting that ω are the same for the two ROIs.

Parameter	M1h L	M1h R	Meaning
ω_e	60.78 s^{-1}	60.78 s^{-1}	Reciprocal of a time constant
ω_s	68.24 s^{-1}	68.24 s^{-1}	"
ω_f	689.50 s^{-1}	689.50 s^{-1}	"
C_{ep}	176	64	Internal connectivity constant

C_{pe}	63	56	“
C_{sp}	172	329	“
C_{ps}	114	116	“
C_{fs}	20	20	“
C_{fp}	44	204	“
C_{pf}	68	60	“
C_{ff}	36	20	“

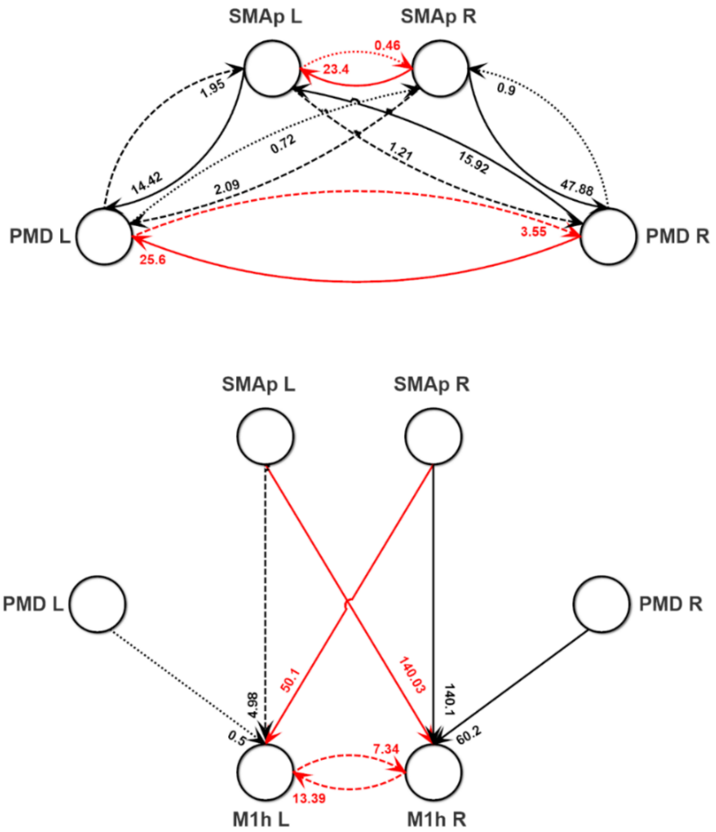


Figure 3.26 – Connectivity strengths obtained by fitting the Neural Mass Model to the normalized power spectra and coherence of the experimental data (see Section 3.3.2.2.2). Since the fitting procedure has been divided in two main steps (Step 1 and Step 2), results of Step 1 (concerning the SMAp L, SMAp R, PMD L, PMD R) are reported in the upper panel, while results of Step 2 (concerning the connection strengths entering into the M1h L and M1h R) are reported in the second panel, although a single network should be considered in the reality. Black lines denote excitatory pyramidal-pyramidal connections, whereas red lines denote inhibitory bi-synaptic connections (pyramidal - fast inhibitory - pyramidal). Continuous lines are used to denote the higher synapses, dashed lines intermediate synapses, and dotted lines the smaller synapses. All remaining synapses are set at zero. The other parameters of the fitting procedure (internal constants within each ROI and inputs to the SMAps and PMDs) can be found in Tables 3.4, 3.5 and 3.6.

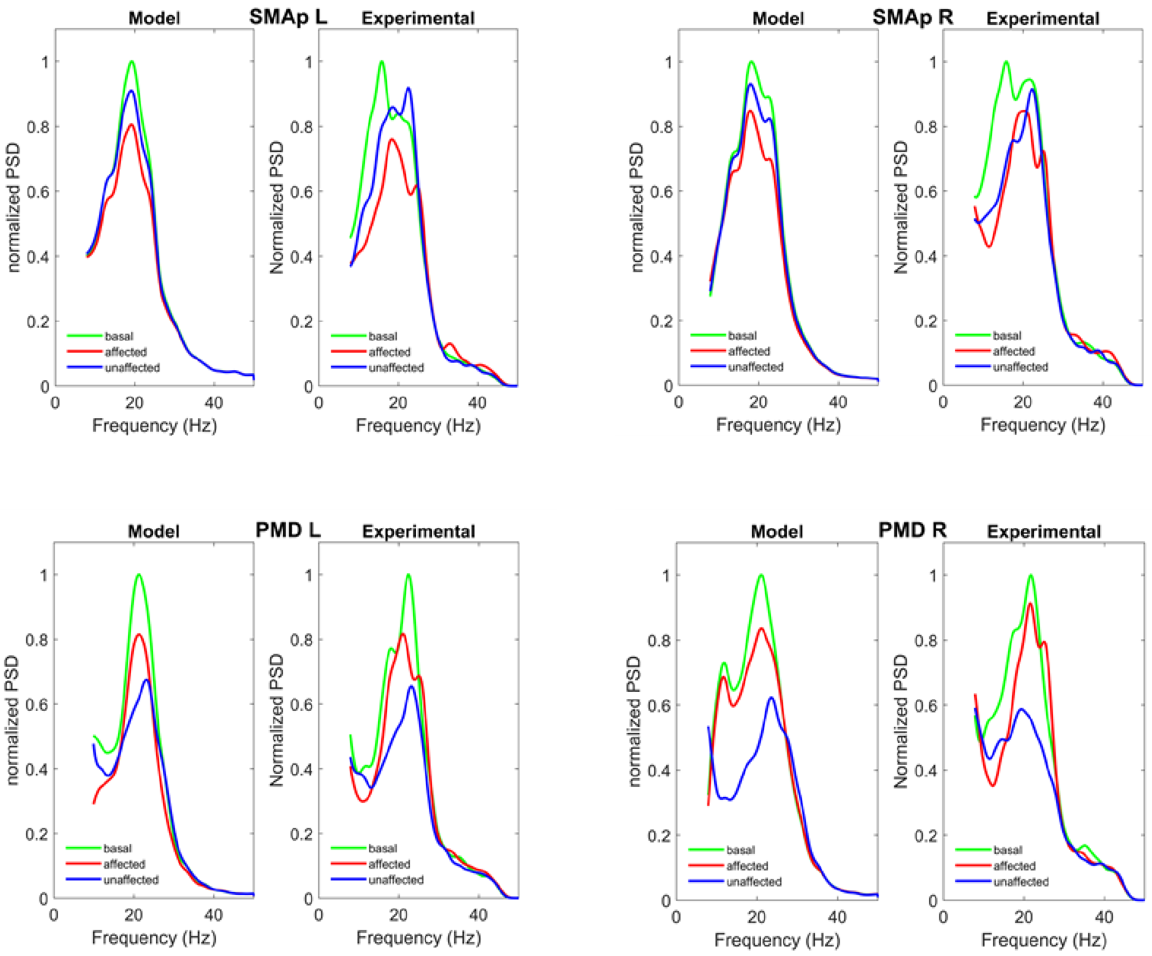


Figure 3.27 – Normalized power spectral densities in the SMAp L (upper left panel), SMAp R (upper right panel), PMD L (bottom left panel) and PMD R (bottom right panel) obtained in basal condition (green lines) and during movement of the affected (red line) and unaffected (blue lines) hands. In each panel, the left part represents model simulation results with optimal parameter values, and the right part the spectra computed from the experimental data. The model can simulate the power spectral density, and ERD quite well in all conditions.

Fig. 3.28 shows a comparison between the coherences among the previous four ROIs predicted by the model, and the experimental ones. As it is clear, the model simulates the coherence quite well in the beta range whereas model coherence falls to zero for frequencies above 30 Hz (gamma range) where a strong coherence can still be observed in the experimental data. This difference will be discussed in the last section.

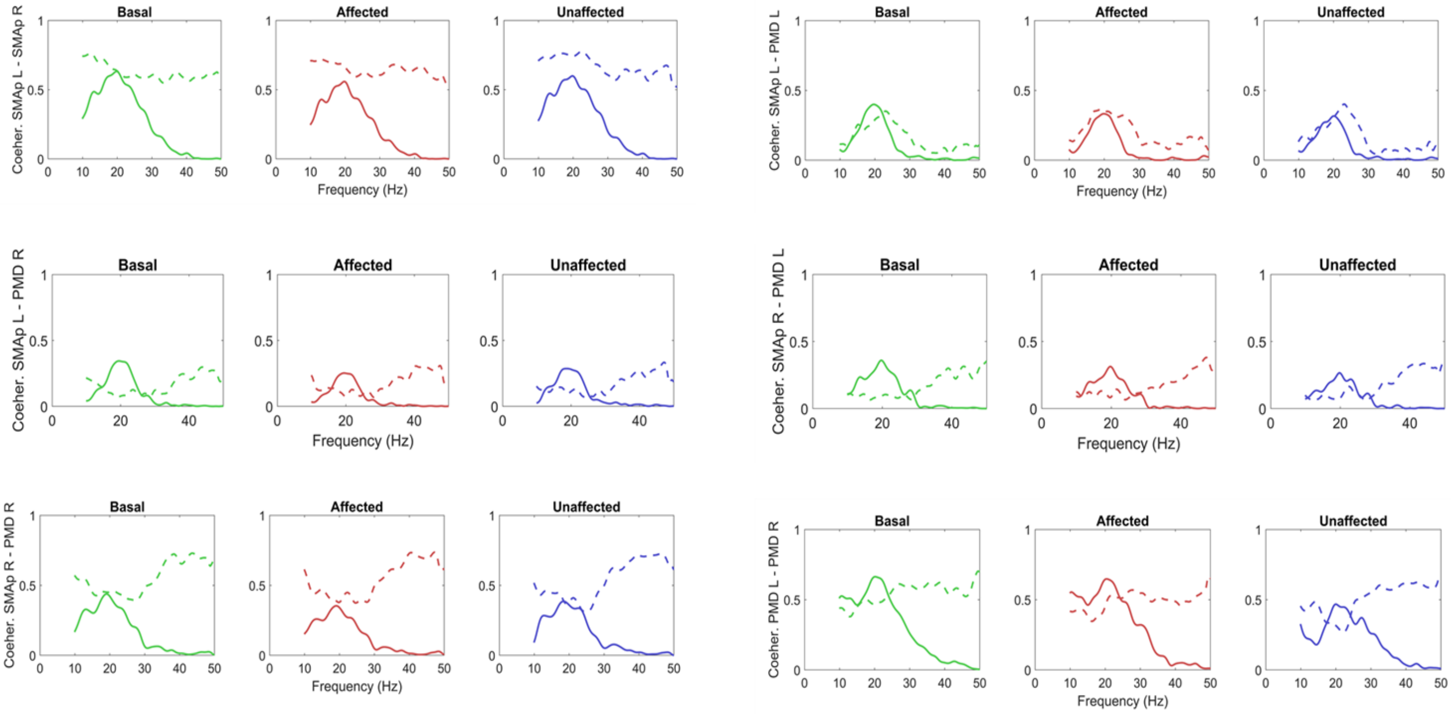


Figure 3.28 – Coherences among the SMAp L, SMAp R, PMD L and PMD R obtained in basal conditions (green lines) and during movement of the affected (red lines) and unaffected (blue lines) hands. The continuous lines represent model simulation results with optimal parameter values, and the dashed lines the values computed from the experimental data. The model can simulate coherences in the range 14-25 Hz rather well in all conditions, but does not incorporate coherence in the gamma range (> 30 Hz). It is worth noting that only the coherences in basal conditions were used in the cost function of the fitting procedure. The others are posterior predictions.

Tab. 3.6 and the lower panel of Fig. 3.26 show the estimated internal parameters of M1h L and M1hR, and the strength of the feedforward connections entering the two ROIs (Step 2 of the fitting procedure), while Fig. 3.29 shows the normalized power spectral densities in the M1h L and M1h R. The model simulates the experimental behavior quite well. We can observe a significant difference in ERD in the affected hemisphere (M1h L) and in the unaffected one (M1h R). M1h L shows a much stronger desynchronization during movement of the affected hand; conversely, the unaffected region, M1h R, shows a comparable ERD during both movements, although with a moderate prevalence during movement of the unaffected hand. This result (discussed below) may suggest that the unaffected region (M1h R) participates more actively to both movements, compared with the M1h L.

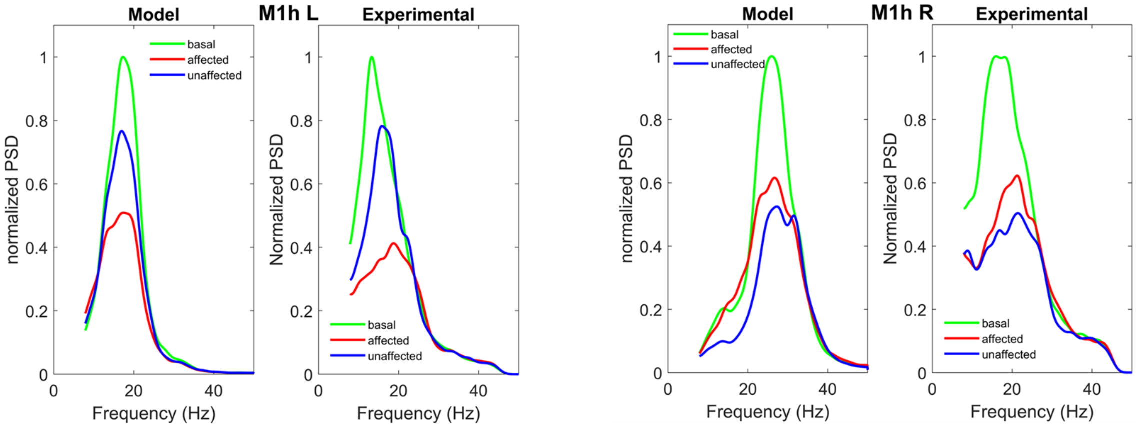


Figure 3.29 - Normalized power spectral densities in the M1h L (left panel) and in the M1h R (right panel), obtained in basal condition (green lines) and during movement of the affected (red lines) and unaffected (blue lines) hands. In each panel, the left figure represents model simulation results with optimal parameter values, and the right figure the spectra computed from the experimental data. The model can simulate the power spectral density, and ERD quite well in all conditions.

Finally, the coherences between the M1h L and the other regions are shown in Fig. 3.30, while the coherences between the M1h R and the other regions in Fig. 3.31. The model can simulate the coherence levels pretty well in the beta range, despite the fact that the coherences between PMCs and M1hs and between SMAps and M1hs were not used at all in the computation of the cost function in Step 2.

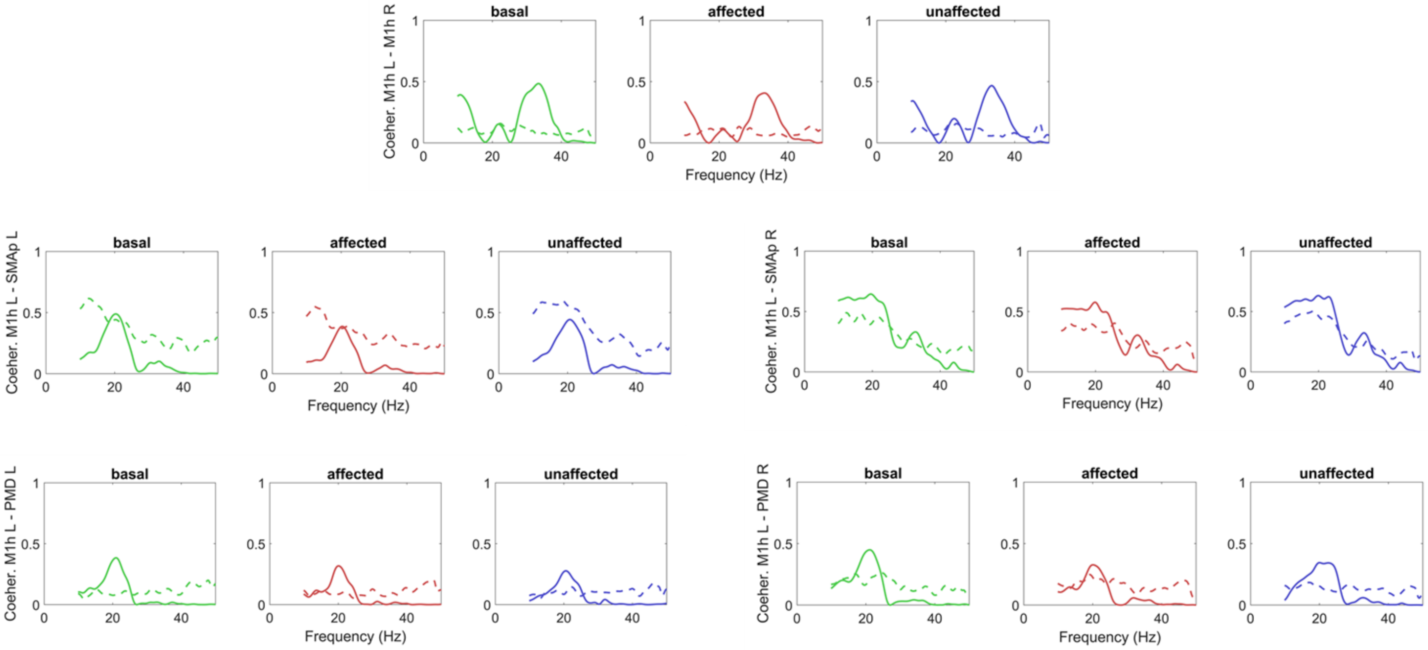


Figure 3.30 - Coherences among the M1h L and all other ROIs obtained in basal conditions (green lines) and during movement of the affected (red lines) and unaffected (blue lines) hands. The continuous lines represent model simulation results with optimal parameter values, and the dashed lines the values computed from the experimental data. It is worth noting that only the coherence M1h L - M1hR in basal conditions was used in the cost function of the fitting procedure. All the others are posterior predictions.

An interesting aspect of the simulation concerns the values of the inputs to the four regions SMAp L, SMAp R, PMD L and PMD R obtained through the fitting procedure (Table 3.5). These values were set to zero in basal conditions, i.e., all ROIs are working in the central linear region. It is worth noting that, during movement of the affected hand, all regions receive a significant input: this is particularly high to the SMAp R (i.e., in the unaffected side) and in the PMD L (in the affected side). Our interpretation is that both sides participate actively to the task. Conversely, during movement of the unaffected hand, only the regions in the unaffected side (SMAp R and PMD R) receive a strong activation.

Finally, as described in the Method section, we repeated the Step 2 of the fitting procedure to test an alternative hypothesis, i.e. assuming that the feedforward connections from each SMAp to the contralateral M1h are excitatory in type (i.e., pyramidal-pyramidal instead of pyramidal-fast inhibitory-pyramidal). The results, reported in the Supplementary Material, show that the model can simulate the ERD in the M1h regions and coherence rather well also assuming excitatory synapses from the contralateral SMAps. However, fitting is worse than in Fig. 3.29.

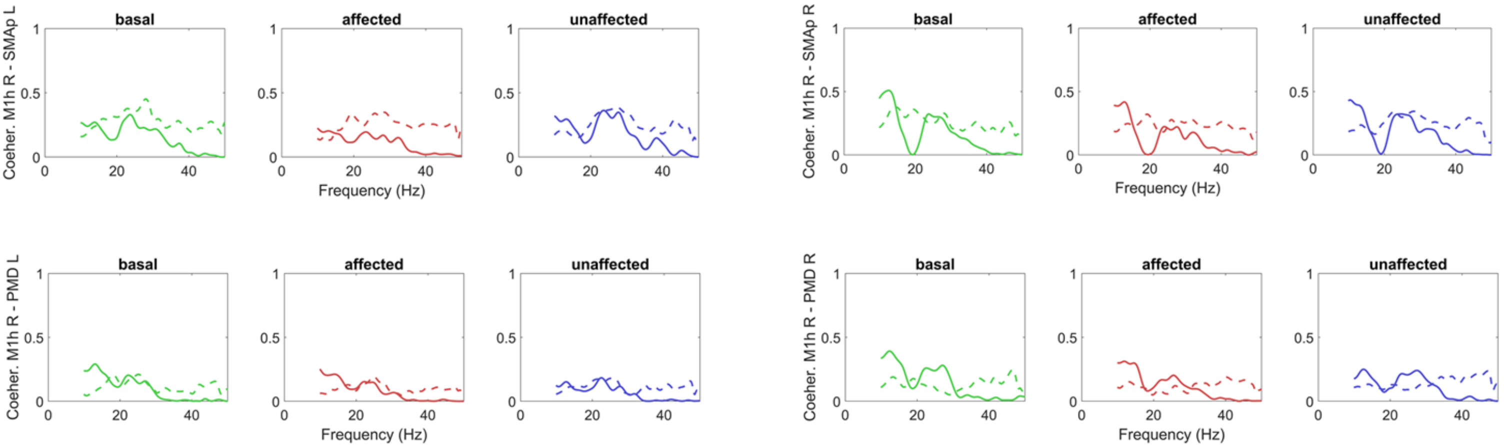


Figure 3.31 - Coherences among the M1h R and all other ROIs obtained in basal conditions (green lines) and during movement of the affected (red lines) and unaffected (blue lines) hands. The continuous lines represent model simulation results with optimal parameter values, and the dashed lines the values computed from the experimental data. It is worth noting that these coherences were not used in the cost function of the fitting procedure, hence are posterior predictions.

3.3.4 Discussion

In this work we present an innovative method to build a connectivity model of the brain motor circuits, using oscillatory networks (i.e., neural mass models). The aim was to investigate the problem of rhythms propagation and power spectral density changes (mainly ERD) within the framework of model-based connectivity. The main new aspect of this study compared with former studies is that differences among tasks are not ascribed to context-dependent changes in connectivity, but rather to the effect of non-linearity. This represents a most parsimonious approach to the problem.

The approach is applied to the study of the power spectral densities and coherence in the beta band in a single subject after stroke, both in resting conditions and during movement of the affected and unaffected hands. Results show that a single set of parameters can mimic all these conditions quite well, and that the values of connectivity obtained are in qualitative agreement with those obtained in former studies. The present study is a proof of concept, applied to a single patient. Indeed, since the connectivity network after stroke may differ significantly among individuals, as a consequence of the locus of the lesion and of time after stroke, each method for connectivity estimation must be applied to single cases. We do not aspire to present a statistic among several subjects here (i.e. a group analysis), but to show how a single fitting procedure actually works.

Results of our study agree quite well with some results appeared in the literature, in which the motor network was assessed in relation to neuroimaging data. First, our main network structure (Fig. 3.26) shares its basic aspects with that obtained by Grefkes et al. [268] on normal individuals. These authors suggest that the most prominent positive influence on intrinsic M1 activity is exerted by the ipsilateral SMA, whereas the intrinsic coupling between PMC and ipsilateral M1 is less pronounced. Moreover, the majority of transcallosal pathways exerts a negative influence on the activity of motor areas in the contralateral hemisphere. Additionally, the interhemispheric interaction between the M1 areas has been studied in humans by means of TMS too [296]. These experiments suggest that both M1 exhibit a mutual inhibitory influence on each other [297,298]. The presence of transcallosal inhibition agrees with the theory of interhemispheric inhibition [299]. In this theory, the capacity of one ROI in a hemisphere to accomplish a specialized task results from effective suppression of the congruent activity in the other hemisphere. This kind of interhemispheric interaction has been observed not only in motor tasks (as in the present study) but also in language and non-spatial visual processing tasks [300].

While the motor network in healthy individuals appears quite symmetrical, a significant asymmetry can be observed in patients after stroke (both in the acute and chronic stages), since neural coupling among areas can be dramatically altered. In particular, a typical finding in stroke connectivity studies [265,290,301,302] is a decreased connectivity in the perilesional area, i.e., a reduction of positive influences from ipsilesional SMA and PMC onto ipsilesional M1, which is observed shortly after the insult and slowly resolves with time. Previous studies showed a correlation between inter-hemispheric coupling in the beta and gamma bands and corticospinal tract integrity [303] as well as between ipsilesional connectivity after a BCI-assisted intervention and the consequent functional recovery in the same bands [282].This finding agrees with the network connectivity shown in Fig. 3.26. Furthermore, in several cases the intrinsic connections C_{ij} are higher in the R hemisphere compared with the L one (Tables 3.4, 3.6 and 3.7).

Moreover, a number of functional neuroimaging studies have shown that, in stroke patients, movements of the affected hand evoke higher and more extended neural activity in cortical brain regions [266,270,304–307]. In particular, unilateral movements of the affected limb are associated with a more bilateral activation pattern in primary motor and premotor areas as compared to neural activity assessed during unilateral hand movements in healthy subjects [301,304,306,307]. The latter results too agree with our model predictions. In fact, our model ascribes the observed differences between the affected and unaffected hemispheres to the following main mechanisms: i) a reduction in the feedforward connections from the premotor areas (both SMA and PSD) to the M1 in the affected side (see Fig. 3.26), in agreement with previous data; ii) a reduction of cross-lateral inhibition from the affected to the unaffected side (see Fig. 3.26), especially evident in the cross-talk between the two SMAs and the two PMDs. The latter mechanism implies a disinhibition of the unaffected side during movement of the paretic limb; iii) a significant asymmetry in the inputs reaching the PMDs and the SMAs during the two hand movements. In particular, just the unaffected hemisphere is strongly stimulated by external excitatory inputs during movement of the unaffected hand (see the third column of Table 3.5), resulting in greater activation in that hemisphere only. Conversely, both sides receive significant stimulation during movement of the affected hand (Table 3.5 second column). This wider input excitation during movement of the affected hand, together with less inhibition from the affected to the unaffected side, results in a significant activation in both hemispheres and in a more bilateral excitation pattern, as suggested by the previous literature. Hence, the areas in the contralesional hemisphere seem to be behaviourally important in the reorganized motor network, to facilitate movements of the affected hand.

The presence of external inputs only to the SMAs and PMCs also agrees with previous studies. Indeed, these areas can receive connections from the visual areas as well as from the parietofrontal system [266]. Finally, the strong input predicted in our study to the PMD R (i.e., in the unaffected side) during movement of the affected hand also agrees with some TMS studies [308,309] showing that interfering with activity of contralesional dorsolateral premotor cortex is associated with a decline in motor performance in stroke patients, but not in controls.

Furthermore, some considerations emerge if we compare the connectivity network obtained with the present model, with the networks obtained using model-free estimation techniques (Fig. 3.25). We can observe that the method we here propose can grasp the main changes reported in the literature between the affected and unaffected hemisphere, both for what concerns a reduction in the connectivity in the affected side, and the reduced transcallosal inhibition from the affected to the unaffected side; these changes clearly emerged in the proposed method but are not equally captured by the other estimators.

There is only an important aspect in Fig. 3.26 which seems at odd with present knowledge and deserves a discussion, i.e., the strong *inhibitory* connection coming from the affected SMA to the unaffected M1. In particular, Rehme et al. [290] observed that the negative coupling from ipsilesional SMA and ipsilesional PMC on contralesional M1 is significantly reduced in Stroke patients. These differences in connectivity may be the consequence of some limitations in the model or in the fitting procedure (see below). However, we think more probable other reasons. A fundamental cause of differences between our model and previous studies is that our model provides a single set of connectivity values, which simulate all tasks together, whereas previous methods based on neuroimaging data make use of a variable connectivity matrix. Finally, discrepancies can also depend on differences in behavior among a rhythmic model, which simulates oscillatory patterns (particularly in the beta band) and models fitted on static neuroimaging data.

The latter consideration moves our analysis to the comparison with the results of more modern techniques, which make use of causal models to simulate EEG/MEG data, with emphasis on power spectral density, brain oscillations and non-linear coupling.

Two main approaches can be found. In the first [310] the model describes phenomenologically the evolution of spectral densities in multivariate time-series including coupling parameters both within and between frequencies, thus taking also non-linear coupling effects into account. This class of models differs significantly from the present since spectra are not simulated with a biologically inspired model and a matrix of parameters is introduced to encode the task dependent influence [311]. Using this approach, Chen et al. [312] studied the human motor system during hand grip, and reached the conclusion that the task-dependent motor network is asymmetric during right hand movements and exhibits strong evidence for nonlinear coupling.

More similar to the present approach, is the recent proposal to use biologically inspired neural mass models within the framework of effective connectivity estimation[313], in order to simulate EEG/MEG data [313–316]. A fundamental difference is that in these models the intrinsic and extrinsic connectivity parameters are affected by the inputs (i.e., are context dependent, in particular see [313]) while the approach we propose here is based on context-independent connectivity and on the effect of non-linearity. Moreover, it is worth-noting that the previous models were not directly applied to the motor network and hand movements, hence we cannot compare their results with ours.

Our study, besides reproducing ERD quite well during different tasks and in different ROIs, provides also some indications on the possible underlying neurobiological mechanisms. Although ERD is a well-known phenomenon, whose first description can be dated back to the mid-seventies [317], its modeling interpretation is still controversial. It is generally thought that event-related desynchronization in the beta range represents increased sensorimotor cortex excitability [318,319]. A core aspect of our model is the use of a sigmoidal relationship to determine the population firing rate in terms of the excitatory potential. Looking at our simulations, we can provide the following neurobiological hypothesis for ERD. When populations work in the central regions of the sigmoidal relationship, they exhibit a good capacity to modify the spiking frequency of individual neurons in response to small changes in the input potentials. This corresponds to a high ability of neurons to synchronize their relative activity, resulting in large collective oscillations. Conversely, when populations work in the upper saturation region (which, as to real spiking neurons, corresponds to the presence of a refractory period) the spiking frequency of individual neurons can only be moderately affected by changes in the input potential. We claim that, in this situation, neurons lose the capacity to synchronize themselves (since synchronization requires the possibility to adapt their reciprocal phase hence to modify their spiking period) thus resulting in a high average activity but with minor oscillatory waves. It is worth noticing, indeed, that in our simulations, beta band ERD in ROIs is in general associated to a decrease in beta-band coherence (see Figs. 3.27-3.31) compared to resting state.

However, as underlined by Byrne et al. [320] we are aware that in this class of NMM models the sigmoidal formulation does not derive from a comparison with a microscopic neuron dynamic, i.e., with a biophysical description of spiking neurons, hence our interpretation is just speculative and requires further study.

Recently, Byrne et al. [320,321] proposed the use of a different kind of neural mass models, which directly incorporates a description of the population synchrony. With this model, the authors simulated the changes on power spectral density of the motor cortex during movement. In particular, they observed that an increase in the excitatory drive causes a decrease in the oscillatory amplitude, i.e., a desynchronization. This result was supported with data obtained with a high dimensional spiking network.

It is worth noting that, although our model and that proposed by Byrne et al. [321] are quite different, they both predict a decrease in oscillation amplitude in response to large drive. Further studies are necessary to compare these two classes of models, and with the behavior of spiking network models.

A further interpretation of ERD, which resembles the present one, was proposed by Grabska-Barwińska et al. [322] using NMM consisting of two excitatory-inhibitory populations in feedback. ERD was mimicked through variations of the external excitation, together with a change in the connection strength between excitatory and inhibitory populations attributed to short-time plasticity. Again, it is worth noting that we do not consider the latter mechanism. More similar to our approach, Mangia et al. [198] simulated ERD with a model consisting of two neural mass models connected together, to reproduce the transmission of information from one cortex to the other. As in our study, cortical activation or deactivation can move the working point in the upper saturation region, causing ERD, or in the linear region, causing ERS.

Finally, we wish to emphasize some limitations of the present study and point out lines for future research.

A first important limitation consists in the difficulty to find a minimum of the cost function, and so in the computational complexity of the fitting procedure. As it is well known, results of a non-linear cost function minimization significantly depend on the initial guess (i.e., on the initial value assigned to the parameters). Unfortunately, the use of global estimation techniques (to reach an absolute minimum instead of a local minimum) would be computationally too cumbersome.

Second, the present model can simulate the power spectral densities and coherences quite well in the beta range, but does not simulate coherences in the gamma band (above 25-30 Hz). There are two possible explanations for this aspect. One possibility is that a significant gamma rhythm is received from other areas not included in the model [194,228]. In alternative, it is possible that a more complex neural mass model, or a different combination of parameters, would allow a better simulation of both gamma and beta coherences together. However, the model used in this paper was built to simulate both beta and gamma rhythms together (see [161]), and makes use of fast inhibitory interneurons often neglected in other NMMs studies.

Hence, we do not think that gamma band limitation can be ascribed to limitation of the NMM used.

Here, we focused on a single patient to show a proof-of-principle of the method we propose. The analysis of many cases, or the longitudinal study of the same subject vs. time, could exploit this new approach to address specific neuroscientific questions, and may be performed in further study, using a better automatization of the fitting procedure.

In conclusion, the present study shows that many different aspects of brain rhythms in the motor network (including power spectra changes, ERD, coherences in the beta range) can be simulated quite well in different tasks, using a single set of parameters for inter-region and intra-region connectivity, without the need to assume a task-dependent change in connectivity weights. Moreover, our study provides a neurobiological interpretation of ERD and ERS, in term of the working point on the sigmoidal relationship.

Our results suggest that, in the patient here analyzed, the contralesional PMD and SMA (PMD R and SMA R) are important in motor reorganization during movement of the affected hand; they receive strong external input and a smaller inhibition from the affected side, and send stronger excitation to the other hemisphere. During movement of the affected hand, ERD occurs both in M1hL and (although to a less extent) in M1hR. This may suggest that M1hR contributes to perform the movement with the affected hand or that the affected hemisphere exerts a reduced inhibition on the healthy one. The observed changes between tasks are due to differences in the external inputs and to non-linear phenomena, but not to task-dependent changes in connectivity. This may represent an important novel aspect to be considered in future studies of brain connectivity.

3.3.5 Supplementary Material

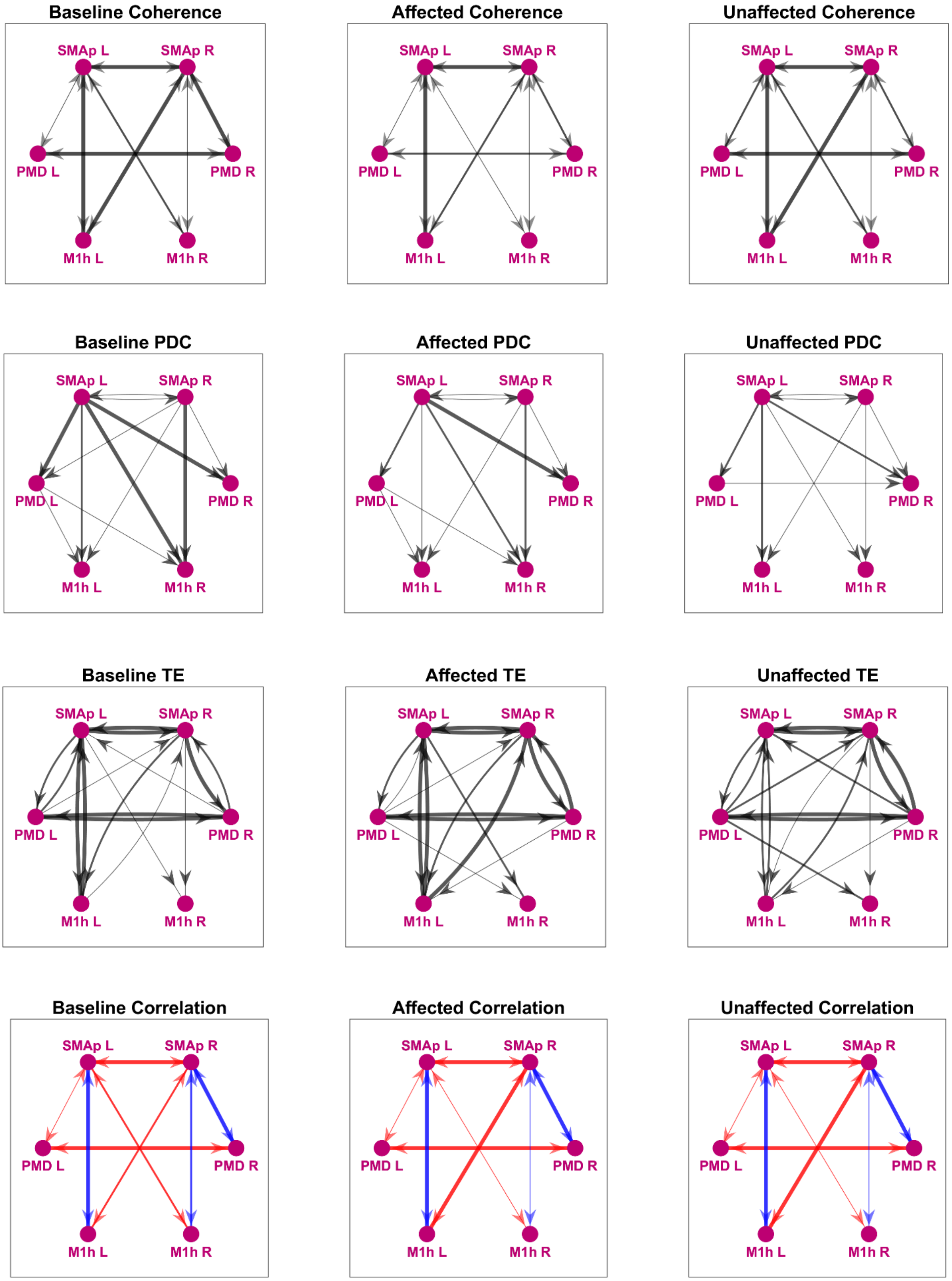


Figure 3.18S - Connectivity networks obtained with the four different data-driven methods (Coherence, Partial Directed Coherence (PDC), Transfer Entropy (TE) and Temporal Correlation), in the three different tasks (baseline,

movement of the affected hand and movement of the unaffected hand). In the temporal correlation networks, colours specify whether the bidirectional connection is *excitatory* (blue, positive correlation coefficient) or *inhibitory* (red, negative correlation coefficient).

As described in the Method section, we repeated the Step 2 of the fitting procedure to test an alternative hypothesis, i.e. assuming that the feedforward connections from each SMAp to the contralateral M1h are excitatory in type (i.e., pyramidal-pyramidal instead of pyramidal-fast inhibitory-pyramidal). The new connectivity strength resulting from the fitting are reported in Fig. 3.19S, while the parameter values can be found in Table 3.5S. The normalized power spectral densities of the M1h L and M1h R are shown in Fig. 3.20S (we do not show coherences for the sake of brevity, but the results are rather similar to those shown in Figs. 3.30 and 3.31 of the main text). As evident in Fig. 3.20S, the model can simulate the ERD in the M1h regions rather well even assuming excitatory synapses from the contra-lateral SMAps. However, the fitting is worse if compared with that in Fig. 3.29 of the main text: in particular, during the movement of the affected hand, the model was not able to simulate an ERD in M1hR as strong as that observed experimentally (the maximum of the normalized spectrum in the model is about 0.75 compared to 0.6 in the experimental data). Furthermore, we can observe that the simulated spectra during the movement of the affected hand are shifted to lower frequencies (around 15 Hz) if compared to the experimental results (around 20 Hz). This is due to an excessive increase of the activity, close to the upper saturation of the sigmoidal relationship, which causes a decrease in the oscillation frequency.

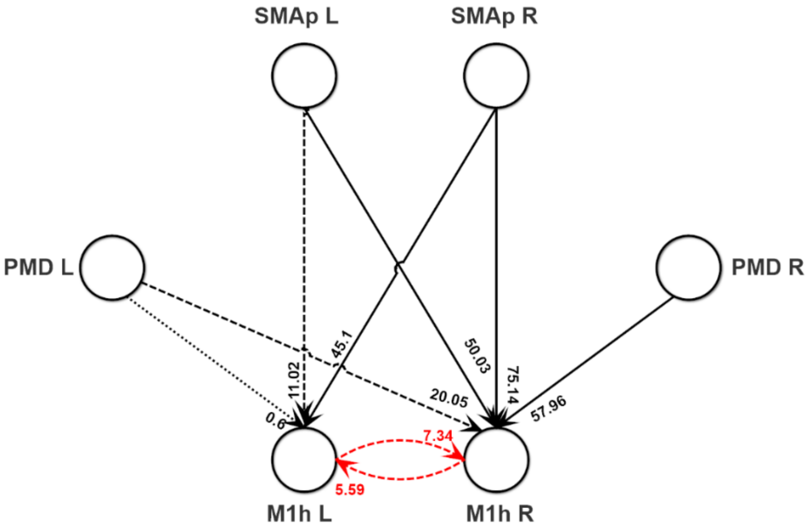


Figure 3.19S - Connectivity strengths obtained by fitting the Neural Mass Model to the normalized power spectra and coherences for M1h L and M1h R, assuming the presence of an *excitatory* connection between each SMAp and the contralateral M1h (see Method section and Supplementary Material). Since the fitting procedure has been divided in two Steps, results of Step 1 (concerning the SMAp L, SMAp R, PMD L, PMD R) are the same as those reported in Figure 3.26 of the main text. Thus, the alternative fitting shown here concerns only Step 2 of the fitting procedure. Black lines denote excitatory pyramidal-pyramidal connections, whereas red lines denote inhibitory bi-synaptic connections (pyramidal-fast inhibitory-pyramidal). Continuous lines are used to denote the higher synapses, dashed lines intermediate synapses, and dotted lines the smaller synapses. All remaining

synapses are set at zero. The other parameters of the fitting procedure (internal constants within each ROI) can be found in Table 3.5S.

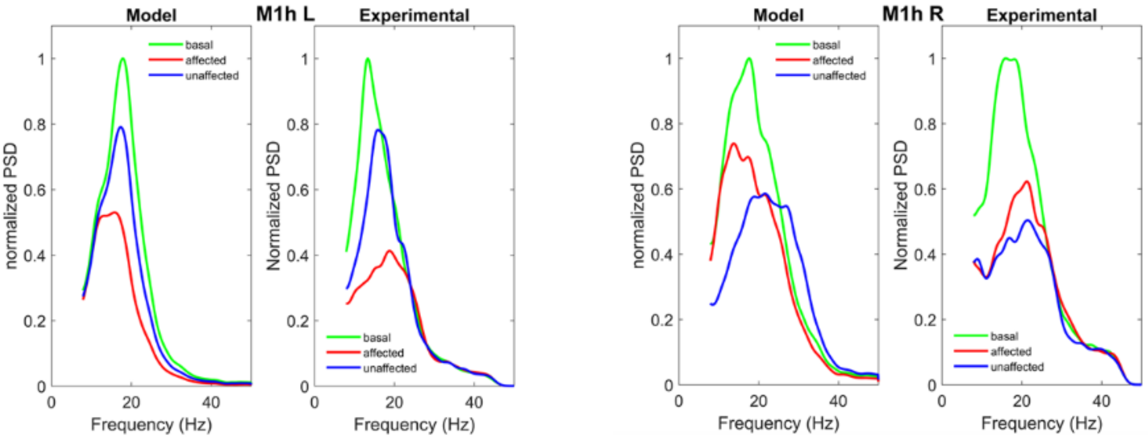


Figure 3.20S - Normalized power spectral densities in the M1h L (left panel) and in the M1h R (right panel), obtained in basal condition (green lines) and during movement of the affected (red lines) and unaffected (blue lines) hands. In each panel, the left part represents model simulation results with optimal parameter values, and the right part the spectra computed from the experimental data. This figure differs from Figure 3.29 in the main text since we used the parameter values shown in Figure 3.19S above, i.e., assuming excitatory connections from each SMAp to the contralateral M1h.

Table 3.5S. internal parameters estimated on ROIs M1h L, M1h R, during the alternative fitting procedure, i.e. assuming that the connections between each SMAp and the contralateral M1h are excitatory in type. It is worth noting that ω are the same for the two ROIs with the same value as in Table 3.6 of the main text.

Parameter	M1h L	M1h R	Meaning
ω_e	60.78 s^{-1}	60.78 s^{-1}	Reciprocal of a time constant
ω_s	68.24 s^{-1}	68.24 s^{-1}	"
ω_f	689.50 s^{-1}	689.50 s^{-1}	"
C_{ep}	118.08	35.53	Internal connectivity constant
C_{pe}	85.22	120.36	"
C_{sp}	5.96	63.54	"
C_{ps}	40.07	44.97	"
C_{fs}	28.07	10.87	"
C_{fp}	41.13	36.65	"
C_{pf}	97.41	142.47	"
C_{ff}	2.00	9.82	"

Table 3.6S. Standard deviation of the error between model and experimental data in the range 10-30 Hz, concerning the normalized power spectral density in the six ROIs, in basal conditions and during

movement of the affected and unaffected hand. These values refer to Figure 3.27 and 3.29. Note that only the M1hR exhibits a significant error in basal condition, to due a shift in the peak frequency of the spectra.

	Basal	Affected	Unaffected
SMAp L	0.1224	0.0767	0.0739
SMAp R	0.0919	0.1127	0.0861
PMD L	0.0806	0.0743	0.0226
PMD R	0.0609	0.1276	0.1317
M1h L	0.2362	0.1979	0.0625
M1h R	0.5381	0.1011	0.1999

Table 3.7S. Standard deviation of the error between model and experimental data, concerning the coherences among the first four ROIs (SMAp L, SMAp R, PMD L, PMD R), in basal conditions and during movement of the affected and unaffected hand. These values refer to Fig. 3.28.

	Basal	Affected	Unaffected
SMAp L - SMAp R	0.1260	0.0971	0.0735
PMD L - PMD R	0.1517	0.1333	0.1601
SMAp L - PMD L	0.1040	0.0710	0.0938
SMAp L - PMD R	0.1345	0.0937	0.0894
SMAp R - PMD L	0.0859	0.0752	0.0592
SMAp R - PMD R	0.1003	0.0991	0.1218

Table 3.8S. Standard deviation of the error between model and experimental data, concerning the coherences among the last two ROIs (M1h L, M1h R) and the other four ROIs (SMAp L, SMAp R, PMD L, PMD R), in basal conditions and during movement of the affected and unaffected hand. These values refer to Figs. 3.30 and 3.31.

	Basal	Affected	Unaffected
M1h L - M1h R	0.1301	0.1011	0.0907
M1h L - SMAp L	0.1652	0.1451	0.1072
M1h L - SMAp R	0.1056	0.1166	0.0905
M1h L - PMD L	0.1199	0.0964	0.0801
M1h L - PMD R	0.1391	0.0947	0.1229
M1h R - SMAp L	0.0924	0.0981	0.0754
M1h R - SMAp R	0.1683	0.1536	0.1314
M1h R - PMD L	0.0946	0.0748	0.0409
M1h R- PMD R	0.0854	0.0575	0.0550

4 Brian connectivity assessment through advanced EEG signal processing

The studies presented in Chapter 3 raised awareness on the limitations of functional connectivity estimators. Two main findings were found that should be kept in mind when applying these metrics to EEG data. First, in nonlinear conditions changes in functional connectivity do not always reflect a true change in connectivity strength but rather a change information transmission between ROIs. Second, in linear conditions Granger Causality in both temporal and spectral domains outperformed the other estimators under analysis, with the sole exception for Transfer Entropy, which showed a similar result in terms of reliability. However, considering the computational burden, Granger Causality proved to be the most cost-effective estimator and was therefore chosen to be applied on electroencephalographic data.

This section shows some applications of Granger Causality as an estimator of directional functional connectivity on experimental EEG data. First, advanced techniques of EEG data processing have been employed to reconstruct cortical sources. Then, cortical sources have been grouped into functionally significant brain regions, according to standardised atlas. At this level, Granger Causality has been computed on three different datasets: *a*) an internal-external attentional task, *b*) a Pavlovian fear conditioning experiment including fear acquisition and reversal, *c*) resting-state of a subclinical population with different autistic traits.

Specifically, in task-dependent studies (*a* and *b*), the role of brain rhythms in cognitive processes has been emphasized by computing Spectral Granger Causality, focusing on theta and alpha frequency bands, which proved to be the most significant rhythms for the tasks under analysis.

Whereas, in the resting-state study (*c*), Temporal Granger Causality has been employed to estimate FC connectivity. Moreover, some of the main centrality measures of Graph Theory have been computed to extract salient features of the brain network under analysis.

4.1 **Alpha and Theta power and spectral connectivity in attentional mechanisms**

The study reported in this chapter refers to the published journal paper entitled “Alpha and theta mechanisms operating in internal-external attention competition”, Elisa Magosso*, Giulia Ricci, Mauro Ursino, *Journal of Integrative Neuroscience* (2021).

In this study, we investigated the modulation of power (at scalp and cortical levels) and connectivity under three different attentional conditions: internal attention, external attention and the attentional competition between the two. The connectivity analysis was performed using Spectral Granger Causality and a subnetwork of selected ROIs has been analysed. The results highlighted the crucial role of theta and alpha brain rhythms in the regulation of attentional mechanisms. Moreover, Supplementary Material section of this study contains a description of the method employed for the estimation of the optimal order (time lag) of the autoregressive model (BVAR).

Background: *Attention is the ability to prioritize a set of information at expense of others and can be internally- or externally-oriented. Alpha and theta oscillations have been extensively implicated in attention. However, it is unclear how these oscillations operate when sensory distractors are presented continuously during task-relevant internal processes, in close-to-real-life conditions. Here, EEG signals from healthy participants were obtained at rest and in three attentional conditions, characterized by the execution of a mental math task (internal attention), presentation of pictures on a monitor (external attention), and task execution under the distracting action of picture presentation (internal-external competition).* ***Method***: *Alpha and theta power were investigated at scalp level and at some cortical regions of interest (ROIs); moreover, functional directed connectivity was estimated via spectral Granger Causality.* ***Results:*** *Results show that frontal midline theta was distinctive of mental task execution and was more prominent during competition compared to internal attention alone, possibly reflecting higher executive control; anterior cingulate cortex appeared as mainly involved and causally connected to distant (temporal/occipital) regions. Alpha power in visual ROIs strongly decreased in external attention alone, while it assumed values close to rest during competition, reflecting reduced visual engagement against distractors; connectivity results suggested that bidirectional alpha influences between frontal and visual regions could contribute to reduce visual interference in internal attention.* ***Discussion***: *This study can help to understand how our brain copes with internal-external attention competition, a condition intrinsic in the human sensory-cognitive interplay, and to elucidate the relationships between brain oscillations and attentional functions/dysfunctions in daily tasks.*

4.1.1 Introduction

In our daily life, the ability to process relevant information and reduce the interfering effect of distracting information is essential to successfully complete any task at hand. This ability is accomplished via attentive processes; indeed, attention acts by prioritizing the processing of a subset of information at the expense of others [323,324].

Attention can be categorized into external and internal attention. Externally-oriented attention is directed towards stimuli in the environment. External attention can be voluntarily driven by task demands in a top-down fashion, e.g. when we focus on a specific spatial location or feature of the sensory stimuli, being this location/feature goal-relevant. External attention can also be involuntarily captured by an object or event, in a bottom-up fashion, even if there is not any goal of attending them. Internally-oriented attention is directed away from external stimulation and towards internal representations and thoughts. Examples of internally directed cognition includes episodic memory retrieval, working memory, planning, mental imagery, mental calculation [323].

The neural mechanisms underlying attention abilities have been the subject of extensive research in the last decades. In particular, electrophysiological research has provided massive support for a functional role of two brain oscillatory rhythms in attentional processes: theta (roughly between 4-8 Hz) and alpha (roughly between 8-13 Hz) rhythms. Indeed, strong associations have been observed between changes in the attentional state and modulations of the power of these oscillations in specific regions as well as modulations of inter-regional synchronization (measuring functional connectivity) within these frequency bands (for a review see [325]).

Regarding theta activity, prominent EEG theta increase has been consistently reported especially over the frontal-midline region (around Fz) in several cognitive tasks [325], with the increase being more pronounced in more demanding conditions. In particular, increase in frontal-midline theta is mostly observable in tasks requiring sustained internally-directed attention, such as working memory tasks [326,327] and mental arithmetic tasks [328–330]; although heterogeneous, these tasks share the need of updating, organizing and holding online the information of multiple items, for their manipulation and retrieval. Besides local frontal theta enhancement, increase in inter-regional theta synchronization has been observed between frontal and temporal and posterior sites in these tasks. Previous investigations have localized the cortical generators of EEG frontal-midline theta in the anterior cingulate cortex (ACC) and adjacent medial prefrontal cortex [326,328,331]. These regions are strongly connected to other cortical areas and theta synchronization is considered as a mechanism through which a frontal supervisory attentional system masters the communication and coordination among the different brain areas involved in these complex tasks [327,331]. An increase in theta activity in medial prefrontal cortex and in temporal and posterior regions was found also in prospective memory tasks (i.e. remembering to execute planned intentions when an appropriate external cue appears); this was associated to attention oriented internally towards the representation of intentions stored in memory and their appropriate retrieval [332].

Regarding alpha activity, much recent research has supplanted the traditional interpretation of alpha activity as just reflecting a cortical idle state, and it is now thought to play a pivotal role in attention, by implementing functional inhibition of task-irrelevant processes that may interfere with the task goals [324,325,333]. Specifically, decrease/increase in alpha power has been associated to cortical excitation/inhibition respectively, based on the adaptive alpha response to task demands. For example, in visual spatial cueing tasks when attention is covertly oriented to one visual hemifield, alpha band oscillations increase over the ipsilateral (unengaged) relatively to the contralateral (engaged) visual system, reflecting inhibition of task-irrelevant visual areas [334–337]. Similarly, when attention is shifted to visual features processed in the ventral visual stream (such as color as opposed to motion), alpha power specifically increases in the task-irrelevant dorsal stream [338]. Furthermore, alpha power in visual areas increases when attention is oriented to other sensory modalities such as somatosensory [339] or auditory [340]. Again, when a prospective memory task is associated to high external attention to detect a difficult visual cue in the environment, alpha power decreases especially in bilateral occipital areas to enhance visual processing [332]. Besides alpha decrease associated to externally-directed attention, many studies report intensification of the alpha rhythm in tasks that require internal attention. In visual working memory tasks, alpha typically increases posteriorly during the retention interval and was positively correlated with memory load (number of memorized items) and task difficulty [341,342], suggesting that it acts to protect the maintenance of relevant information against potential external intrusion. Furthermore, when to-be-maintained items and to-be-ignored items (representing distractors) are simultaneously presented in separated hemifields during the encoding interval, in the retention interval alpha power increases in the hemisphere contralateral to the previously presented distractors [343,344]. Yet, alpha power increases at posterior sites in anticipation of a predictable distractor presented during the retention interval [345]. EEG/MEG studies investigating cortical connectivity suggest that posterior inhibitory alpha modulations are driven by top-down signals from regions of the prefrontal cortex: indeed, long-range alpha influences from these anterior regions towards occipital cortex augment when visual interferences have to be avoided [334,346].

All previous studies have provided huge contribution to the comprehension of the relationships between brain oscillations and attentional control. However, in our opinion there are some issues that have remained overlooked and that might provide further insights into the neural correlates and neuroelectrical manifestations of attention.

First, the role of alpha oscillations in distractors filtering has been mainly explored in conditions when the distractors were *absent,* i.e. they were not longer or not yet presented. That is, the effects on alpha were observed either during the anticipation period before any visual irrelevant/distracting stimulus was actually provided [334,345], or during a retention interval, when all information presented during the previous encoding period was removed [343,344]. Therefore, it is still unclear which is the *online* effect of visual distractors and how local alpha power and alpha connectivity are modulated while visible distractors are interfering with task-relevant cortical processing. Indeed, in this case, at variance with the previous ones, the mechanisms that work to inhibit the distracting input likely compete with an automatic, bottom-up capture of external attention induced by the visual distractors.

Second, most studies investigated either alpha or theta modulation in relation to attentional tasks and only a few have examined both rhythms simultaneously [332,347]. A joint investigation would favor the emergence of possible reciprocal relationship between these two mechanisms. It can also be noted that theta modulation (mainly in frontal midline region) has been especially studied in condition of internally-directed attention, while modifications of theta oscillations inducible by external visual attention have been less explored.

Finally, most of the cited literature uses trial-based experimental paradigms and time-locked analysis of the signals, with analysis windows usually involving a very short period of hundreds of milliseconds. While this procedure may favor the separation of the investigated mechanism from spurious effects, it suffers from a less ecological validity and barely reflect real-life situations.

In order to contribute to the previous points, in this study EEG signals were recorded from healthy subjects in four different conditions with open eyes: i) 5 minutes of resting state, used as basal condition; ii) 5 minutes of external attention, consisting in the presentation of emotionally neutral pictures, with no associated task demand, so that bottom-up external attention is considered here; iii) 5 minutes of internally-directed attention consisting in a mental arithmetic task; iv) 5 minutes of competition between internally-directed and external attention, consisting in the mental arithmetic task performed while simultaneously presenting pictures that acted as distractors. Then, by estimating also cortical activity from the EEG signals via eLORETA, we aimed to address the following questions: What are the alpha and theta power modulations that play a role in these conditions of external (sensory) attention and internal (cognitive) attention respectively, and how are some key cortical regions involved? How is the directional flow of information in the two bands modified among these regions by these two states of attention? How do the previous mechanisms of external and internal attention interact when the two forms of attention are in competition?

4.1.2 Materials and methods

4.1.2.1 Participants

Twenty-four healthy volunteers (13 females), aged 21-27 years, (mean ± std = 24.23 ± 1.63 years) took part to the study. They were recruited among students of the University of Bologna (Italy). Each participant had normal or corrected to normal vision and reported no medical or psychiatric illness. The study was approved by the local bioethical committee of the University of Bologna (file number 29146; year 2019) and written informed consent was obtained from all participants before the beginning of the experiment. All data were analyzed and reported anonymously.

4.1.2.2 Experimental Protocol

The experiment was performed in a controlled laboratory environment. The participants underwent four consecutive experimental sessions, each lasting 5 minutes and separated by a short interval during which the subject, while sitting, could slightly move and speak. The participants comfortably seated facing a computer monitor about 50 cm far; they were instructed to perform the four sessions with eyes open, reducing at minimum eye, head and body movements. First, a *resting state* (R) was recorded for 5 minutes, while the participants stayed relaxed in front of a grey screen displayed on the monitor. This was used as the basal condition. The other three sessions corresponded to the three conditions of attention. The *internal attention session* (IntAtt) consisted in executing a mental arithmetic task throughout the 5-minute period, while a uniform grey screen was displayed on the monitor. The task was a mental serial subtraction in steps of seventeen starting from a given number. This kind of task required intentional orientation of attention to internal processing, and did not rely on any information provided externally. Note that no distracting visual inputs were delivered during this session since only the grey screen was continuously presented to the participant. The *external attention session* (ExtAtt) consisted in presenting to the participant a series of pictures on the monitor during the 5-minute period. Specifically, thirty emotionally neutral pictures extracted from the IAPS (International Affective Picture System) database [348] were used. The pictures were displayed on the monitor one after the other every 10 seconds, in a random order. No specific task was associated to pictures viewing. Therefore, during this session visual external attention of participants was captured by the pictures in a bottom-up fashion, as it may occur in a real-life sensory rich environment. The *session of internal-external attention competition* (IntExtAtt) consisted in the combination of the previous two conditions: during this 5-minute session, participants were required to perform the mental arithmetic task as in the IntAtt condition, while pictures were displayed on the monitor as in the ExtAtt condition. Therefore, in this session, the pictures acted as visual distractors that competed and interfered with the execution of the mental math task. This session simulated a common realistic condition such as when we are engaged in internally-oriented tasks (e.g. problem solution in the classroom, office, etc.) and we need to isolate from the sensory-rich surrounding to avoid the intrusions of task distracting inputs.

The order of the three attentional sessions was randomized across participants. The participants were provided with the instructions for the math task only immediately before the beginning of the corresponding sessions, so they remained unaware of the task until its onset. These instructions also communicated to the participant that he/she would be required to report the final number achieved at the end of the session. The starting number for the mental serial subtraction was 2500 in the first of the two math task sessions (IntAtt or IntExtAtt); the final number reached by the participant in this session was then used as the starting point for the serial subtraction in the subsequent session. Furthermore, participants were not informed in advance of pictures presentation either in case of the ExtAtt or IntExtAtt session. The same 30 pictures were randomly presented during both the two sessions. The selected pictures had normative level of pleasure between 4.5 and 5.5 points on a 9-point scale. Neutral pictures were used to avoid the involvement of emotional factors.

As in previous studies using covert mental arithmetic [328,329,347], it was not possible to control the actual execution of the task during the sessions. However, all participants were aware that not complying with the received instructions would have compromised the study. Furthermore, requiring the participants to report the final reached number likely further motivated them to engage in the task. We also asked each participant at the end of the sessions if he/she actually performed the serial subtractions (otherwise, the recording would have been discarded). A confirmation of task engagement was obtained from all participants; furthermore, all participants reported that performing the task under picture presentation was more demanding.

4.1.2.3 EEG recording and processing

EEG signals were recorded through a Neurowave System (Khymeia, Italy, Brainbox® EEG-1166 amplifier, Braintronics) using an elastic cap with 32 Ag/AgCl scalp electrodes (Fp1, Fp2, AF3, AF4, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, PO3, POz, PO4, O1, O2). The reference electrode was placed on the left earlobe and the ground electrode was located on the forehead. The right earlobe electrode was acquired too, for offline re-referencing. During each experimental session, EEG data were digitized in continuous recording mode at a sample frequency of 128 Hz and 16-bit resolution, and with the inclusion of a hardware notch filter eliminating line noise at 50 Hz. Thus, for each participant, four 5-minutes recordings were acquired. Finally, for each participant and each recording, the thirty-two EEG signals (+ the right earlobe signal) were exported in Matlab (R2019b, The MathWorks Inc., Natick MA) for further analysis. Firstly, data were re-referenced with respect to the average of the two earlobe signals and high-pass filtered at 0.75 Hz to eliminate the DC offset and slow drifts. Then, for each participant the following processing steps were applied.

4.1.2.3.1 Independent Component Analysis (ICA) and artefact removal

First, for each participant, the four EEG recordings were concatenated along the time dimension. Then, to identify and remove artefacts, Independent Component Analysis (ICA) was applied to the concatenated signals using the extended infomax algorithm implemented in the open source Matlab toolbox EEGLAB (https://sccn.ucsd.edu/eeglab/index.php) [349.] To speed up artefactual component identification, we took advantage of the recent EEGLAB plug-in named 'IClabel' that allows for automatic classification of the estimated independent components into 'Brain' ICs (if classified as originating from cortical patches), or artefactual ICs distinguishing between 'Muscle', 'Eye', 'Heart', 'Line Noise', 'Channel Noise','Other' ICs [350]. Outside EEGLAB toolbox, non-Brain IC components were removed based on the automatic classification results, except a few doubtful cases that were subjected to visual inspections (scalp map and time/spectral pattern) to ascertain the presence of artefactual activity before their removal. Overall, an average of 19.8 (SD = 4.5) ICs were removed across participants. Artifact-cleaned EEG signals were then reconstructed by back-projecting the remaining set of non-artefactual ICs. Finally, the so cleaned signals were separated back into four 5-min portions corresponding to the four sessions.

The application of ICA procedure to the concatenated EEG signals ensured that the same ICs were removed from each recorded session, avoiding that differences between the four conditions could emerge because of removal of different ICs.

4.1.2.3.2 Estimation of Individual Alpha-Band Window

In order to sharpen the precision of the spectral analyses, we estimated the Individual Alpha Window (IAW) of each participant, based on previous observations that the alpha band may vary considerably across individuals [351,352]. This was also motivated by a preliminary visual inspection of the Power Spectral Density (PSD) of posterior EEG signals at rest (R): the standard alpha band (8-14 Hz) did not seem appropriate for all participants, in some cases extending beyond in others not completely including the alpha peak. An automatic and objective method for identifying the IAW for each participant was adopted, based on a Matlab algorithm (publicly available) recently proposed [353] and inspired by the manual procedure originally proposed by Klimesh et al[352].

In agreement with the literature [351,352], we grounded the estimation of the IAW only on the signals acquired in the resting session (R). Briefly, the procedure was as follows. For each participant, the PSD of all channels over the resting session was obtained (Welch's periodogram method, Hamming window of 5 seconds, 50% overlap, 10 s zeropadding) and given as input to the algorithm. For each PSD, the algorithm applied a least-squares curve-fitting procedure (via the Savitzky-Golay filter, sample window length set at 27 and polynomial order at 5) in order to obtain a smoothed PSD function and to estimate its first derivative. Based on the first derivative, only channels where a peak (or a split-peak complex) was identified within a putative alpha bandwidth (set at 8-14 Hz) and clearly distinguishable from the 1/f background noise, were kept for the subsequent analysis. The latter consisted in identifying the nearest local minima to the left and right of the peak complex, thus identifying the channel-wise alpha bounds; these were then averaged across the retained channels to obtain the IAW ($f_1 \div f_2$) of the specific participant. Across the 24 participants, we obtained $f_1 = 7.66\,Hz \pm 1.07\,Hz$ (mean $\pm$ std, range = $5.6\,Hz \div 9.7\,Hz$) and $f_2 = 14.54\,Hz \pm 1.14\,Hz$ (mean $\pm$ std, range = $12.8\,Hz \div 16.6\,Hz$). The IAW estimated on each participant was then used to compute the alpha power both at scalp level for each channel (even if the specific channel did not contribute to the IAW definition) and at source level, in each of the four sessions. Furthermore, the lower f_1 bound of the IAW served to define the upper bound of the theta band for each participant, while the lower bound of the theta band was fixed at 4 Hz.

4.1.2.3.3 Alpha and Theta Power Computation at Scalp Level

For each participant, the PSD of each channel over each session, R, IntAtt, ExtAtt, IntExtAtt, was obtained. For each session separately, the alpha power over the IAW and the theta power over the resulting theta band was computed for each channel and the values at the 32 channels were used to realize scalp maps of alpha and theta distribution in the four conditions.

Furthermore, the power was computed at two scalp macro regions, by averaging the alpha and theta powers over frontal electrodes (Fp1, Fp2, AF3, AF4, F3, Fz, F4, FC1, FC2) and parieto-occipital electrodes (P3, Pz, P4, PO3, POz, PO4, O1, O2), obtaining anterior/posterior Theta and anterior/posterior Alpha in each condition. To evidence attentional-dependent changes, the alpha and theta power at regional level in each attentional condition was normalized to the corresponding regional value in the basal resting condition (R).

4.1.2.3.4 Cortical Source Estimate - Alpha and Theta Power Computation at Cortical Voxel Level

Besides an analysis at a scalp level, we were interested in an analysis at cortical source level. To this aim, cortical source activity was reconstructed starting from EEG signals. Specifically, we estimated the intracortical current densities by using the approach eLORETA (exact Low Resolution Electromagnetic Tomography [354]) for solving the inverse problem, as implemented in the LORETA-KEY©® software package. The eLORETA solution space is restricted to the cortical gray matter of a reference brain (MNI 152 template) with a total of 6239 voxels at 5 mm spatial resolution. The eLORETA method is a linear, weighted minimum norm inverse solution; the particular weights used in this solution endow eLORETA with the property of exact localization of test point sources under ideal (noise free) conditions. We used the software LORETA-KEY©® only to compute the transformation (inversion) matrix, say K, starting from the Talairach coordinates of the 32 electrodes; then all subsequent processing steps were implemented with customized code in Matlab. The matrix K has dimension $(3 \cdot 6239) \times 32$, and right-multiplied by the scalp potentials at a given time instant gives the three scalar components of the current density vector at each voxel at that time instant.

For each participant, we reconstructed the three dimensional time series of the current densities at all voxels for each session. For each session separately, the alpha and theta power at each voxel was derived as follows. The PSDs of the three vector components were computed (using the same parameters as at scalp level) and the voxel power in the alpha-band and in the theta-band was taken as the sum of the three corresponding power values computed on the three PSDs. In this computation too, we took into account the individual alpha and theta band for each participant. The power values at the 6239 voxels were used to realize cortical maps of alpha and theta distribution in the four conditions, and for statistical voxel-wise analysis.

4.1.2.3.5 Cortical Regions of Interest (ROI) - Alpha and Theta Power Computation at Cortical ROI Level

We selected some cortical regions of interest (ROIs) to focus both the power analysis and connectivity analysis. The selection was mostly based on a priori considerations about the brain conditions under investigation and on results of previous literature; only in one case the selection was data-driven. These choices are further commented in Section 4.1. Discussion. Two sets of ROIs were selected for the theta-band and alpha-band analysis, with some ROIs common to the two sets. Fig. 4.1 shows the ROIs used for the Theta- and Alpha-band analysis, in the three-dimensional cortical source space adopted for the solution of the EEG inverse problem (Section 4.1.2.3.4). The assignment of a voxel to a specific region is based on information provided by the software LORETA-KEY©®, which specifies the region each voxel belongs to; the Supplementary Material Section 4.1S provides a detailed description of how voxels of the source space were assigned to each ROI.

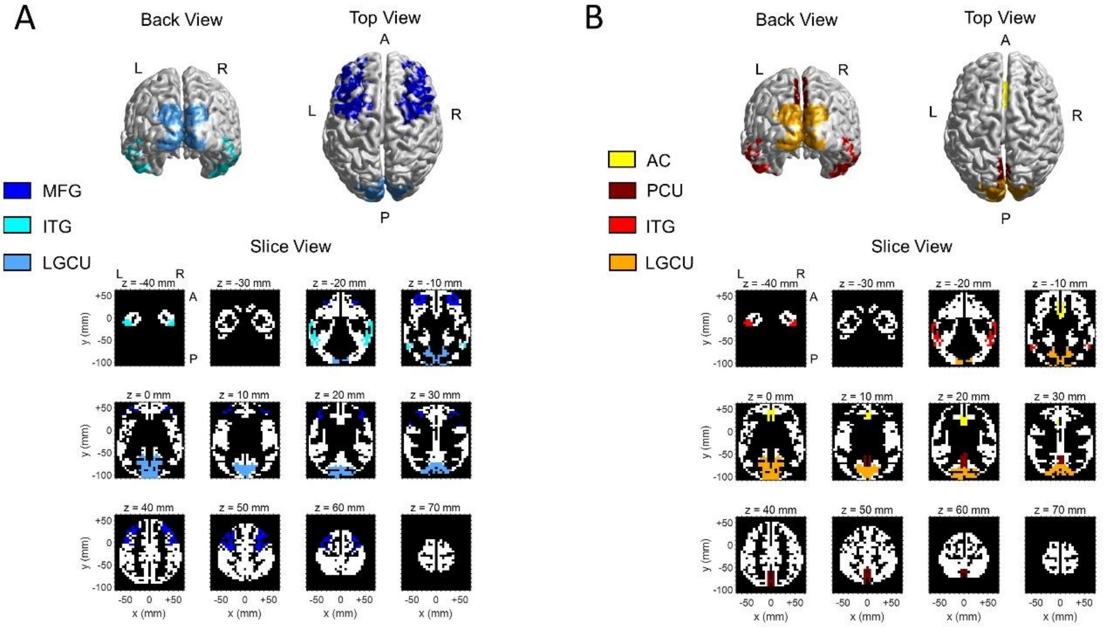


Figure 4.1 – The regions of interest (ROIs) selected for the alpha-band analysis (Panel A) and theta-band analysis (Panel B) at the cortical level. The ROIs are depicted in the cortical source space used by the method (eLORETA) adopted for the solution of the EEG inverse problem. A realistic head model based on the MNI152 template was used, with the solution space restricted to the cortical gray matter and divided into 6239 voxels at 5mm cubic spatial resolution. Both panels show the back view and top view of the surface of the cortical space and the view across transversal slices along the z axis. The x, y, z axes (MNI coordinates in mm) have orientation left to right (x), posterior to anterior (y) and bottom to top (z). The voxels depicted with the same color belong to the same region, as indicated by the legend: note that, for the sake of clarity, the same color was used for homologous regions in the two hemispheres (left and right) and, in case, for the medial portion too (e.g., LGCU). Some ROIs (ITG, LGCU) were analyzed in both bands, and for completeness, they are represented both in panel A and B (but with different colors).

ROIs for Alpha-band Analysis

Lingual Gyrus/Cuneus (LGCU) - These occipital regions are involved in earlier stages of visual processing and perception. They were selected according to previous evidence (see also Section 4.1.1) of alpha inhibition-disinhibition mechanisms operating in these regions, and of their involvement in long-range alpha band synchronizations during attentional tasks [334,346]. We took into account left, right and medial Lingual Gyrus/Cuneus ($LGCU_L$, $LGCU_R$, $LGCU_M$).

Inferior Temporal Gyrus (ITG) – This region is a high-level visual area in the ventral visual stream, and recognized to be involved also in encoding information about scenes [355] (the kind of pictures we used in the experiment). We expected that both alpha power and alpha-band connectivity of this region could be modulated depending on the attentional condition. Left and right ITG ($\mathrm{ITG_L}$ and $\mathrm{ITG_R}$) were considered.

Middle Frontal Gyrus (MFG) – We included this region as a putative area involved in top-down modulation of visual alpha activity. In particular, MFG has been found to be functionally connected to occipital visual areas when internal information has to be maintained, and this long-range interaction involved alpha-band oscillations [346,356,357]. The left and right MFG ($\mathrm{MFG_L}$ and $\mathrm{MFG_R}$) were considered.

ROIs for Theta-band Analysis

Lingual Gyrus/Cuneus (LGCU) and *Inferior Temporal Gyrus* (ITG) – These same 5 ROIs were considered, to investigate whether theta activity in visual areas may be modulated by attentional condition, and to highlight differences in alpha and theta modulation in these areas. This can be also of interest since theta entrainment in visual areas is less investigated than alpha-band modulation.

Anterior Cingulate Cortex (ACC) –This medial area of the frontal cortex was selected since previously identified as the main generator of EEG frontal-midline theta in attentional demand and cognitive monitoring; in particular, frontal theta associated to mental calculation has been localized in ACC [328,329,358].

Precuneus (PCU) – This area, laying on the medial surface of the posterior parietal lobe, was the only area not selected a prior; it was included in the analysis based on initial results at cortical level (see Section 4.1.3) showing that voxels in this area appeared especially involved in theta-band modulation under condition of ExtAtt. This area has been largely linked to memory processes and internally directed functions [359,360]; however, it has been also related to visual functions [361,362] (see also Section 4.1.4 for a critical discussion on this).

Each selected ROIs contains several cortical voxels (see Supplementary Material 4.1S and Supplementary Table 4.1S). To perform the power analysis at the ROIs level, the powers of all voxels within each ROI were averaged, obtaining the theta power and alpha power of each ROI in each session. To evidence attentional-dependent changes, the alpha and theta power of each ROI in each attentional condition was normalized to the corresponding ROI value in the basal resting condition (R).

4.1.2.3.6 Spectral Granger Causality Analysis at Cortical Level

We used the spectrally resolved Granger Causality (GC) analysis to estimate the directional influences, in the Granger sense of predictability [363], between each pair of selected ROIs in the theta and alpha band and to assess their potential modulation as a function of the attentional condition. Considering two time series $x_i(t)$ and $x_j(t)$ representing the signals at site *i* and site *j* respectively, the spectral GC is based on the representation of the system $X(t) = \left[x_i(t) \quad x_j(t)\right]^T$ via a bivariate autoregressive process (of order p) and on the derivation of the spectral representation of the bivariate process by Fourier transforming. From this representation and according to the formulation originally proposed by Geweke [364] (see also [365]), the power spectrum of each time series (let's say $x_i(t)$) can be derived and it can be viewed as composed by an "intrinsic" part and a "causal" part, the latter being the part predicted by the data from the other site (let's say $x_j(t)$). The GC spectrum from j to i at each frequency f ($GC_{j\to i}(f)$) is defined by considering the ratio between the portion of the total power of $x_i(t)$ at f predicted by $x_j(t)$ and the total power of $x_i(t)$ at f [365–367].

The computation of the GC spectrum between two ROIs requires that each ROI is described by a single time series. To derive a single time series representative of each ROI activity, first for each voxel within the ROI, we computed the component of the current density vector perpendicular to the local cortical surface in the head model, and then these scalar components were averaged across all voxels within the ROI. This procedure may be justified considering that each estimated current density vector is mainly representative of the post-synaptic currents at pyramidal neurons dendrites inside a cortical macro-column, and these dendrites are oriented orthogonally to the local cortical surface [368]. Furthermore, in our previous works we used neural mass models to simulate activity of cortical ROIs and their interactions [369–371], where each ROI describes the average behavior over a large population of neurons: we showed that the causal connectivity established in the model between each pair of ROIs is reflected by (and can be estimated from) the post-synaptic activities of the two ROIs.

Based on the previous procedure, a time waveform for each ROI was derived for each of the four sessions. Then, we considered separately the set of 7 ROIs selected for the alpha-band analysis (alpha-set, MFG_L, MFG_R, ITG_L, ITG_R, $LGCU_L$, $LGCU_M$, $LGCU_R$) and the set of 7 ROIs selected for the theta-band analysis (theta set, ACC, PCU, ITG_L, ITG_R, $LGCU_L$, $LGCU_M$, $LGCU_R$). For each session and within each set, the GC spectra were computed for all pairs of ROIs in both directions (of course, since some ROIs overlapped in the two sets, the same GC spectra held). Then, to obtain a single value for each connection, the maximum value of each GC spectrum in the theta band and in the alpha band was extracted, respectively. This was done for each participant by considering the corresponding IAW and theta band. The values of connectivity obtained in each attentional condition were compared with those in basal condition, to assess how causal influences were modulated by attentional demand. The order of the bivariate models to estimate the GC spectra was set at 30. This value was determined by comparing the power spectral densities of the ROIs obtained by the bivariate models and those estimated by the Fourier based method (Welch's method) directly on the ROIs time waveforms: a good compromise was reached at this model order (see Supplementary Material Section 4.2S).

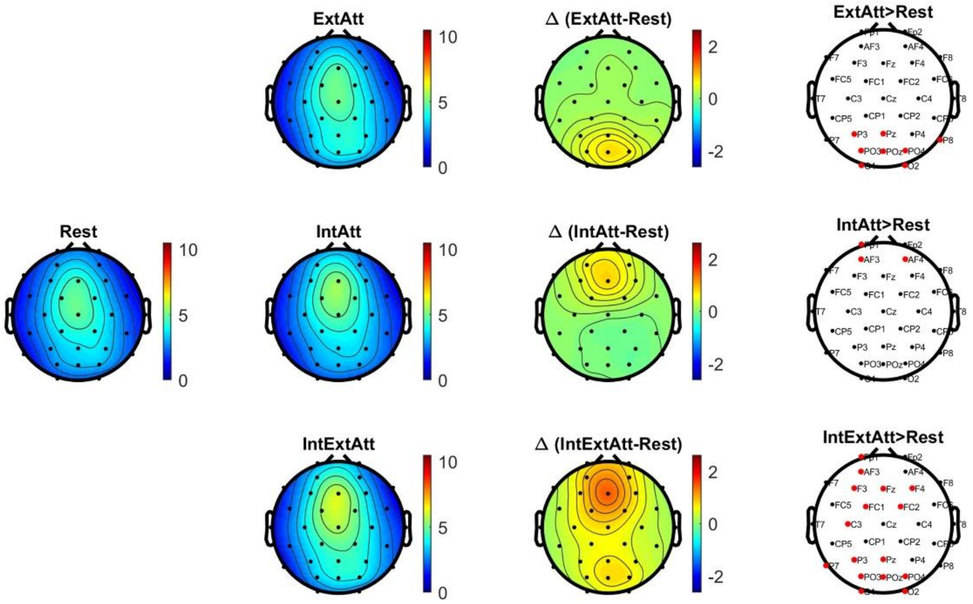


Figure 4.2 – Scalp maps and channel-wise statistical analysis for theta power. The maps in the first and second column represent the theta power (μV^2) averaged across all participants as a function of the experimental session (Rest, IntAtt, ExtAtt, IntExtAtt). The maps in the third column represent the theta power difference (μV^2) between each attentional condition and the rest, averaged across all participants. Each scalp map was obtained by color coding the average theta power value at each electrode position in a 2D circular cartoon head (top view of the head, nose at the top) and using interpolation on a fine 67 × 67 grid. In the cartoon heads of the fourth column, the red markers denote the electrodes that showed statistically significant difference of theta power in the attentional condition compared to rest (p-values < 0.05, one-tailed permutation-based t-test, p-values corrected for multiple comparisons, see procedure (a) in Section 4.1.2.3.7).

4.1.2.3.7 Statistical comparisons

Based on the previous procedures, the power values (at scalp and cortical level) and directed connectivity values were obtained for each of the 24 participants, in the two bands. Then, statistical comparisons were performed between each attentional condition and the baseline condition, both as to power and connectivity values. When the comparison of one attentional condition vs rest involved power maps (both at channel and voxel levels) and connectivity maps, we used the non parametric permutation test for functional neuroimaging [372] that readily deals with the multiple comparison problem of testing at all voxels/channels/connections. When the comparison concerned the variations at the level of an entire single ROI (e.g. the power of a ROI or the overall causal outflow from a ROI) in the attentional conditions vs rest, the non-parametric Wilcoxon signed rank test was used (with Bonferroni correction). The performed comparisons and the applied statistical methods are better specified below.

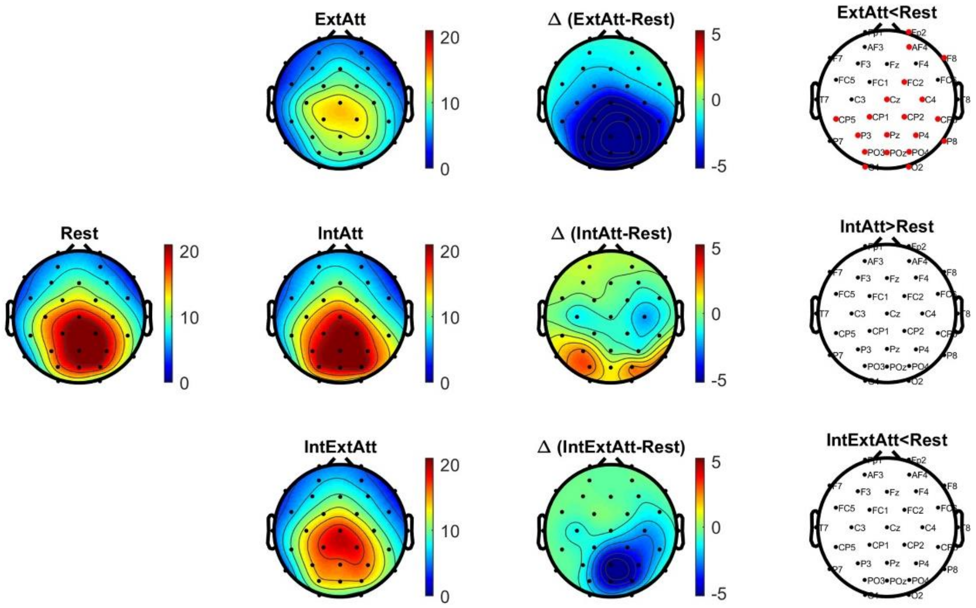


Figure 4.3 – Scalp maps and channel-wise statistical analysis for alpha power. The maps in the first and second column represent the alpha power (μV^2) averaged across all participants as a function of the experimental session (Rest, IntAtt, ExtAtt, IntExtAtt). The maps in the third column represent the alpha power difference (μV^2) between each attentional condition and the rest, averaged across all participants. Each scalp map was obtained as described in Figure 4.2. In the cartoon heads of the fourth column, the red markers denote the electrodes which showed statistically significant difference of alpha power in the attentional condition compared to rest (p-values < 0.05, one-tailed permutation-based t-test, p-values corrected for multiple comparisons, see procedure (a) in Section 4.1.2.3.7).

a) *Channel-wise (scalp level) comparison of alpha power and theta power in attentional condition vs the baseline condition*. For each band and each attentional condition, we statistically evaluated which scalp channel exhibited different power compared to baseline condition. A non-parametric, permutation-based t-test was used. To this aim, the distribution of the t statistic at each channel under the null hypothesis was empirically computed by performing 5000 random permutations of the observed values between the two conditions [372]. The uncorrected p-value at each channel was the proportion of the permutation distribution at least as extreme as the observed t statistic (computed on the non-permuted values). To obtain p-values corrected for multiple comparisons, the permutation distribution of the maximal t statistic was obtained (by collecting at each permutation the maximum of the channel t statistics) and the corrected p-value at each channel was the proportion of the distribution for the maximal statistic at least as extreme as the observed t statistic.

b) *Regional-wise (scalp level) comparison of alpha power and theta power in attentional condition vs the baseline condition*. For each band and each scalp region (anterior/posterior), we statistically evaluated whether the specific attentional condition modified the regional power compared to the baseline. A Wilcoxon signed rank test (the non-parametric equivalent of the parametric paired Student's t-test) was applied to compare each attentional condition to the baseline condition. Correction for multiple comparisons was applied separately for each region and band: corrected p-values were obtained via the Bonferroni correction by multiplying the raw p-values by 3, since three comparisons were made for each region and for each band.

c) *Voxel-wise (cortical level) comparison of alpha power and theta power in attentional condition vs the baseline condition*. For each band and each attentional condition, we statistically evaluated which cortical voxel exhibited different power compared to baseline condition. The same method as in a) (non-parametric permutation-based t-test) was adopted at the voxel level.

d) *ROI-wise (cortical level) comparison of alpha power and theta power in attentional condition vs the baseline condition*. For each ROI and each band of interest related to the ROI, we statistically evaluated whether the specific attentional condition modified the ROI power compared to the baseline. The same method as in b) (Wilcoxon signed rank test) was adopted at the cortical ROI level.

e) *Comparison of causal influences in the alpha and theta band in attentional condition vs the baseline condition*. For each band and each attentional condition, we statistically evaluated which directed causal influences between the selected ROIs were modified compared to the baseline condition. In this case too, the nonparametric permutation-based t-test was used adopting the same procedure as in a) and in c). Furthermore, we summarized some connectivity aspects (e.g. overall flow from one ROI to a set of other ROIs), by averaging across the corresponding connectivity values and we tested whether the specific attentional condition modified this overall connectivity compared to the baseline by using the Wilcoxon signed rank test (same method as in b) and d)).

4.1.3 Results

4.1.3.1 Power analysis at scalp level

Fig. 4.2 shows the scalp maps of theta power in each session, the scalp maps of theta power difference between each attentional condition and rest, and the results of the corresponding channel-wise statistical analysis. Fig. 4.3 displays the same information as to alpha power. According to results in Fig. 4.2, engagement in the mental math task (IntAtt) was associated with frontal midline theta increase. Conversely, in ExtAtt, theta power increased only at parieto-occipital electrodes. In IntExtAtt condition, i.e. when the mental task had to be performed against task-irrelevant pictures, frontal-midline theta showed a strong increase, overcoming that observed in IntAtt condition; furthermore, theta increased at posterior electrodes too. As to alpha power (Fig. 4.3), the ExtAtt induced a dramatic decrease of alpha activity especially at parieto-occipital sites, an effect well expected due to visual stimulation. In IntAtt condition, alpha power tended to increase at the more posterior sites, although significance was not reached at any electrode. Finally, results obtained in IntExtAtt are especially interesting: indeed, the same visual stimuli as in ExtAtt applied during the math task were associated with a much smaller alpha power decrease at posterior sites, with no statistical significance.

A summary of the previous results is represented in Fig. 4.4, which displays the theta and alpha power changes within the two scalp macro-regions (anterior and posterior, see Section 4.1.2.3.3), as a function of the attentional condition. The bar plots emphasized a different pattern of theta activity at the anterior and posterior region. Anterior theta activity exhibited the trend to progressively increase across the three conditions and reached the largest value in the IntExtAtt condition, in agreement with its relation with internal attention. Posteriorly, theta increase occurred to a similar extent in the two conditions involving picture presentation (ExtAtt and IntExtAtt) and was absent in IntAtt condition. Alpha power at the posterior region confirmed a significant decreased in ExtAtt but not in IntExtAtt condition, and a tendency to increase (but with high variability) in the IntAtt condition.

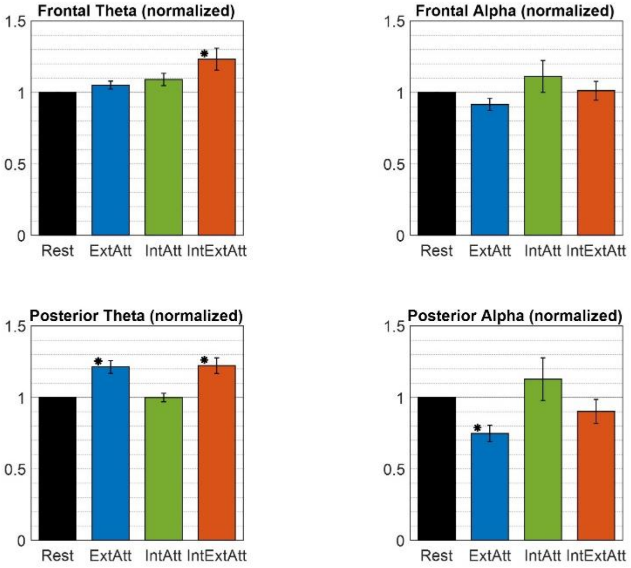


Figure 4.4 – Theta power and alpha power computed at the two scalp macro-regions (anterior and posterior) in each experimental condition, and normalized to the rest condition. Powers at the anterior and posterior regions were obtained via arithmetic average across frontal electrodes (Fp1, Fp2, AF3, AF4, F3, Fz, F4, FC1, FC2) and parieto-occipital electrodes (P3, Pz, P4, PO3, POz, PO4, O1, O2) respectively (see Section 2.3.3). Each bar represents the mean ± SEM across all participants. Within each bar plot, * denotes statistically significant difference between the attentional condition and rest condition ($p<0.05$, Bonferroni corrected for multiple comparisons, see procedure (b) in Section 4.1.2.3.7).

4.1.3.2 Power analysis at cortical level

The comparisons between each attentional condition and rest at the level of voxels are presented in Fig. 4.5 as to theta power and in Fig. 4.6 as to alpha power. For each comparison, the voxel-by-voxel power difference is represented, together with the results of the voxel-wise statistical analysis. Table 4.1 indicates, for each attentional condition, which of the selected ROIs were significantly implicated in power change on the basis of the voxel-wise statistical analysis (see Supplementary Material Section 4.3S and Supplementary Tables 4.2-4.6 for the list of all cortical voxels significantly involved).

Theta power increase in ExtAtt and IntAtt was localized within a limited area of voxels in the posteromedial cortex and in the midline/left frontal cortex, respectively (Fig. 4.5). It is worth noticing that the cluster of statistically significant voxels in IntAtt include also the selected ROI ACC (Table 4.1). Furthermore, as anticipated in Section 4.1.2.3.5, the results in ExtAtt motivated the inclusion of the PCU ROI in our analysis. Indeed, the cluster exhibiting significant theta increase included also voxels belonging to this region (see Table 4.1 and Supplementary Table 4.2), and significant PCU theta increase was confirmed at the level of the entire ROI too (as shown later in Fig. 4.7 and discussed in Section 4.1.4). In IntExtAtt, significant theta increase exhibited a widespread distribution, mostly involving temporal regions (ITG too) and portion of the posteromedial cortex; a few ACC voxels were involved too. As to Alpha power (Fig. 4.6), the ExtAtt condition resulted in significant decrease in almost all voxels in the posterior-parietal, occipital and temporal lobes (including LGCU and ITG ROIs). In IntAtt, although alpha tended to increase at occipital regions, no significant difference was obtained at any voxel. However, the effect of internal attention on alpha power appeared evident when considering the IntExtAtt condition; at variance with ExtAtt condition, here only a small cortical cluster in the parietal cortex exhibited significant decrease.

The previous analysis is useful to obtain a picture of the power changes in each attentional condition at a fine-scale spatial resolution, and reveled that voxels in most of the a priori selected ROIs were significantly implicated. In order to assess the modifications of the overall power in each selected ROI (rather than at single voxel level), Fig. 4.7 and Fig. 4.8 display the modulation of the overall ROI power vs the attentional condition in the theta band and alpha band, respectively. A significant increase of theta activity was seen in ACC in the two conditions involving internal attention (IntAtt and IntExtAtt, Fig. 4.7), larger when internal attention competed with the external one, identifying the ACC as implicated in the frontal-midline theta increase observed at scalp level in these cases (see Fig. 4.2). The other ROIs exhibited a different pattern of modulation: theta increase was associated with conditions involving external attention (ExtAtt and IntExtAtt), especially in posterior regions (except $\mathrm{LGCU_M}$), where it reached statistical significance in both conditions. Analysis in the alpha band (Fig. 4.8) settled that ExtAtt was associated with a strong decrease in the occipital (LGCU) as well as ITG regions, while in IntExtAtt alpha power in these same ROIs did not significantly decrease or decreased only to a much lower extent (see $\mathrm{LGCU_M}$), confirming the tendency of alpha power to increase (although not significantly) in IntAtt.

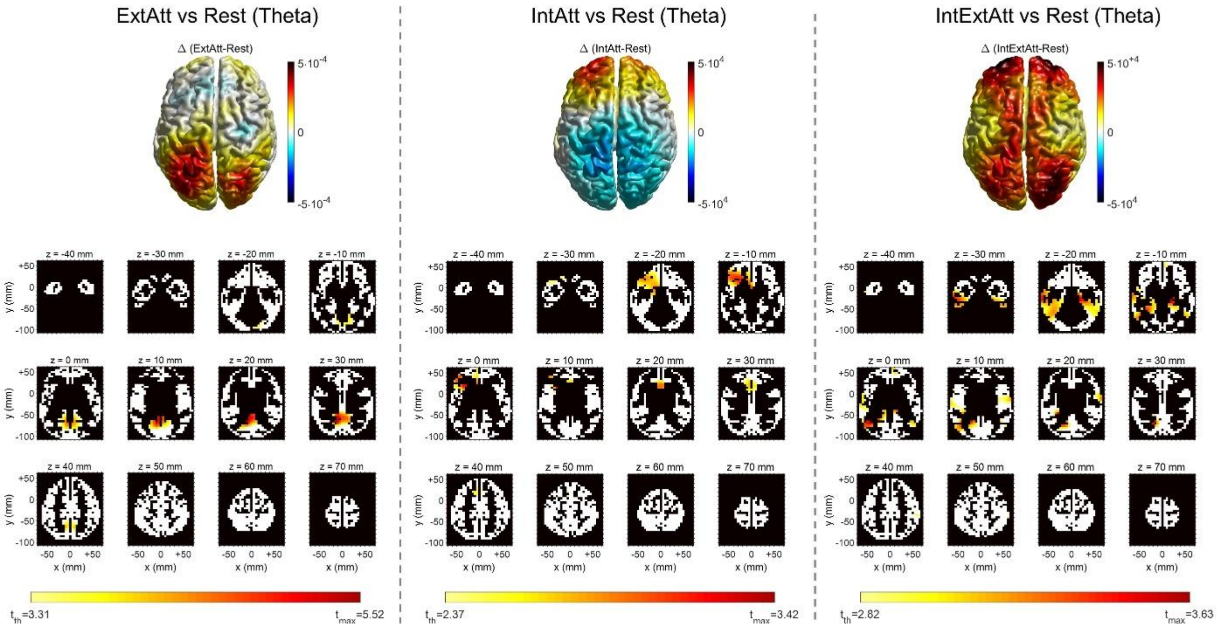


Figure 4.5 – Voxel-wise comparison of theta power between each attentional condition and rest. For each attentional condition, the 3D top view of the cortex shows the theta power difference (($\mu A/mm^2)^2$) at each voxel between the attentional condition and the rest, averaged across all participants. The transversal slices across the z axis (oriented from bottom to top) display the results of the voxel-wise statistical analysis: the colored voxels correspond to significantly higher theta power in the attentional condition compared to rest (p-values < 0.05, one-tailed permutation-based t-test corrected for multiple comparisons, see procedure (c) in Section 4.1.2.3.7), with the color scale corresponding to corrected t-values (t_{th} indicates the critical threshold at 5% probability). It is worth noticing that only the results of the one-tailed statistical analysis (testing attentional condition > rest) are reported, since the test in the other direction did not provide in any significance.

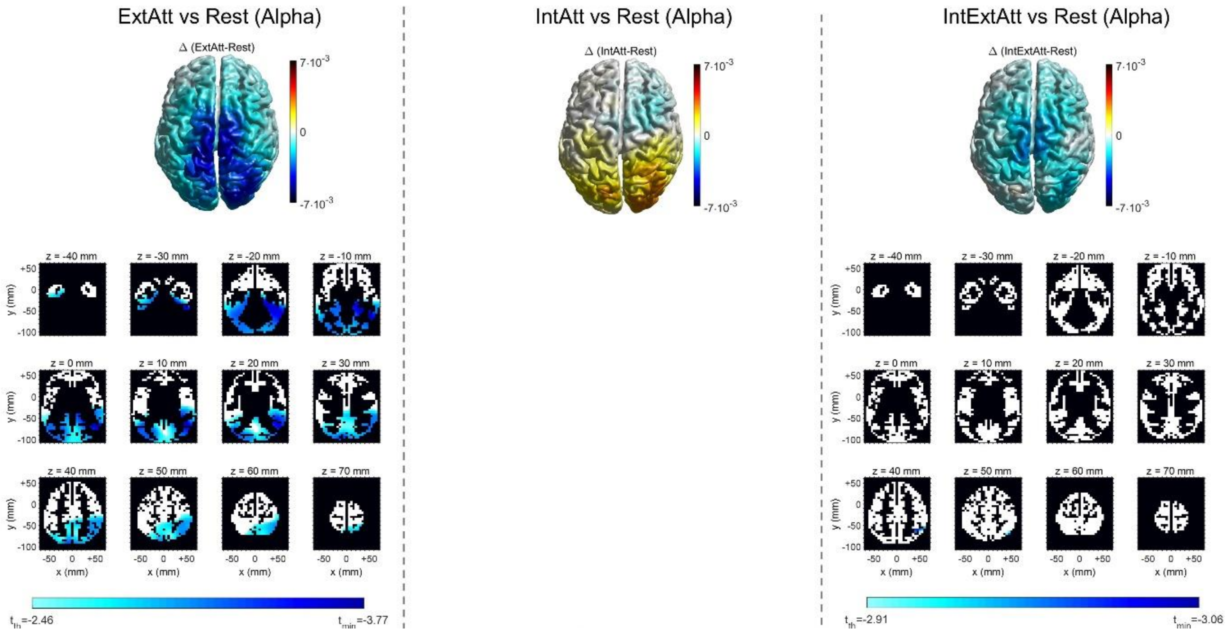


Figure 4.6 – Voxel-wise comparison of alpha power between each attentional condition and rest. For each attentional condition, the 3D top view of the cortex shows the alpha power difference (($\mu A/mm^2)^2$) at each voxel between the attentional condition and the rest, averaged across all participants. The transversal slices across the z axis (oriented from bottom to top) display the results of the voxel-wise statistical analysis: the colored voxels correspond to significantly lower alpha power in the attentional condition compared to rest (p-values < 0.05, one-tailed permutation-based t-test corrected for multiple comparisons, see procedure (c) in Section 4.1.2.3.7), with the color scale corresponding to corrected t-values (t_{th} indicates the critical threshold at 5% probability). For the ExtAtt and IntExtAtt condition, only the results of the one-tailed statistical analysis (testing attentional condition < rest) are reported, since the test in the other direction did not provide in any significance. For the IntAtt condition, one-tailed statistical analyses did not provide any significant results in either direction.

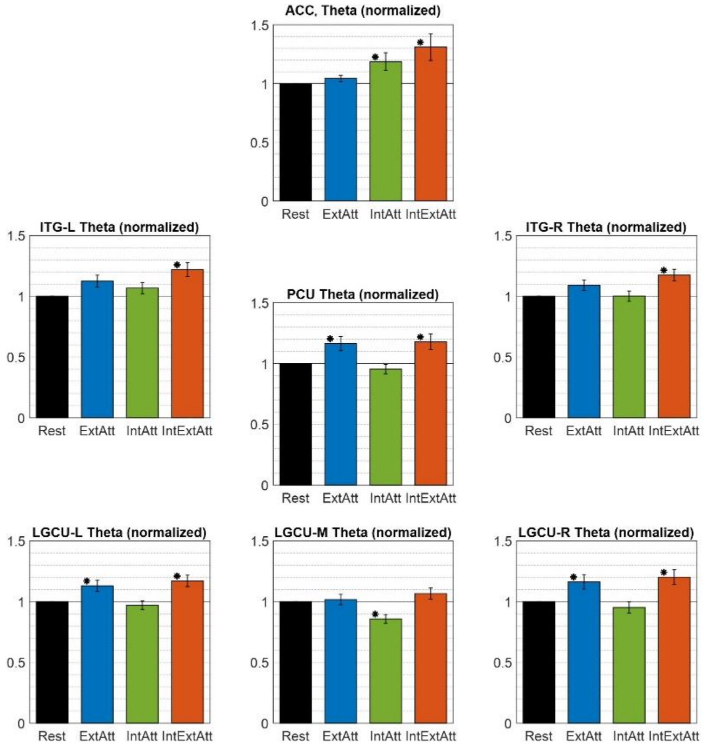


Figure 4.7 – Overall theta power (obtained by averaging across all voxels within the ROI) in each of the seven ROIs selected for the theta-band analysis, computed in each experimental condition and normalized to the rest condition. Each bar represents the mean ± SEM across all participants. Within each bar plot, * denotes statistically significant difference between the attentional condition and rest condition (p<0.05, Bonferroni corrected for multiple comparisons, see procedure (d) in Section 4.1.2.3.7).

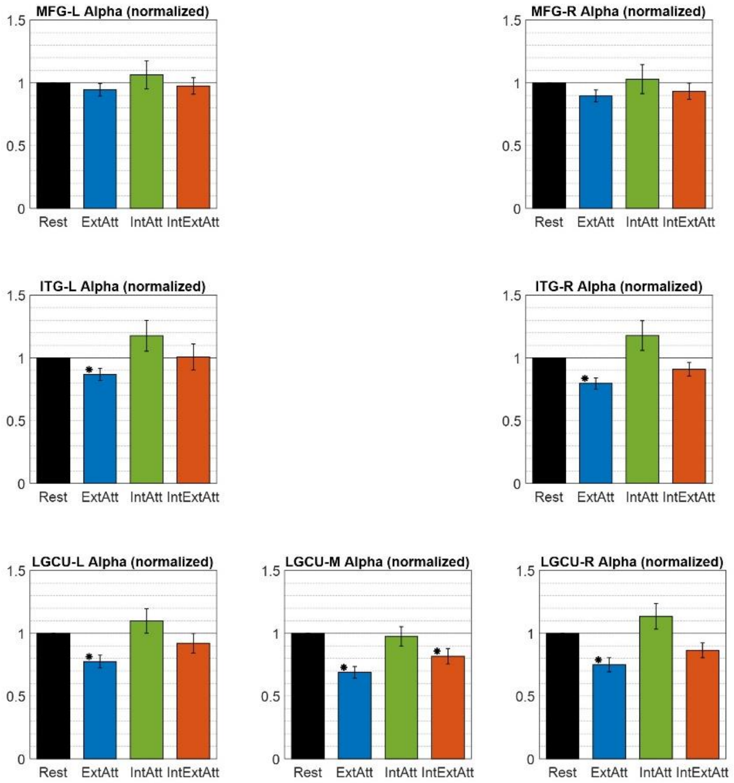


Figure 4.8 – Overall alpha power (obtained by averaging across all voxels within the ROI) in each of the seven ROIs selected for the alpha-band analysis, computed in each experimental condition and normalized to the rest condition. Each bar represents the mean ± SEM across all participants. Within each bar plot, * denotes statistically significant difference between the attentional condition and rest condition ($p<0.05$, Bonferroni corrected for multiple comparisons, see procedure (d) in Section 4.1.2.3.7).

4.1.3.3 Granger Causality Analysis

Figures 4.9 and 4.10 show the results of the GC analysis for the selected ROIs in the theta and alpha band. The colored arrows over the 3D cortical surfaces indicate the single connections that significantly modified ($p<0.05$ uncorrected) compared to rest, with the red and blue indicating an increase and decrease respectively. The bar plots sum up some interesting aspects, reporting the overall causal outflow from a ROI (or a set of homologue ROIs) to a set of other ROIs. The results can be summarized as follows.

Theta-Band (Fig. 4.9) – In ExtAtt, increased theta-band connectivity was localized posteriorly, with the PCU exerting a significantly greater influence towards the visual ROIs compared to rest; no increased influence from ACC was observed. In IntAtt, increased connectivity from ACC emerged, while the influence from PCU exerted a minor role. In IntExtAtt, a sort of summation of the previous two effects occurred, with an overall increase in connectivity among the ROIs, and with ACC and PCU exerting each an overall influence on visual areas (LGCU and ITG) to a similar extent as in IntAtt and ExtAtt respectively (see the bar-plots).

Alpha-Band (Fig. 4.10) – The ExtAtt was characterized by an overall dramatic decrease of causal influence, with a clear bottom-up arrangement. No relevant modifications in top-down connections emerged (see in particular MFG_L, MFG_R→ LGCU). Conversely, IntAtt was associated with an increase of both bottom-up and top-down connectivity between visual areas (ITG, LGCU) and the frontal area, involving mostly the left MFG. Finally, IntExtAtt was characterized by a decrease in bottom-up connectivity, but to a much lower extent than in ExtAtt condition (see in particular the bar plot LGCU→MFG_L, MFG_R), and still by an increase in top-down influence from the left MFG (bar plot MFG_L→LGCU).

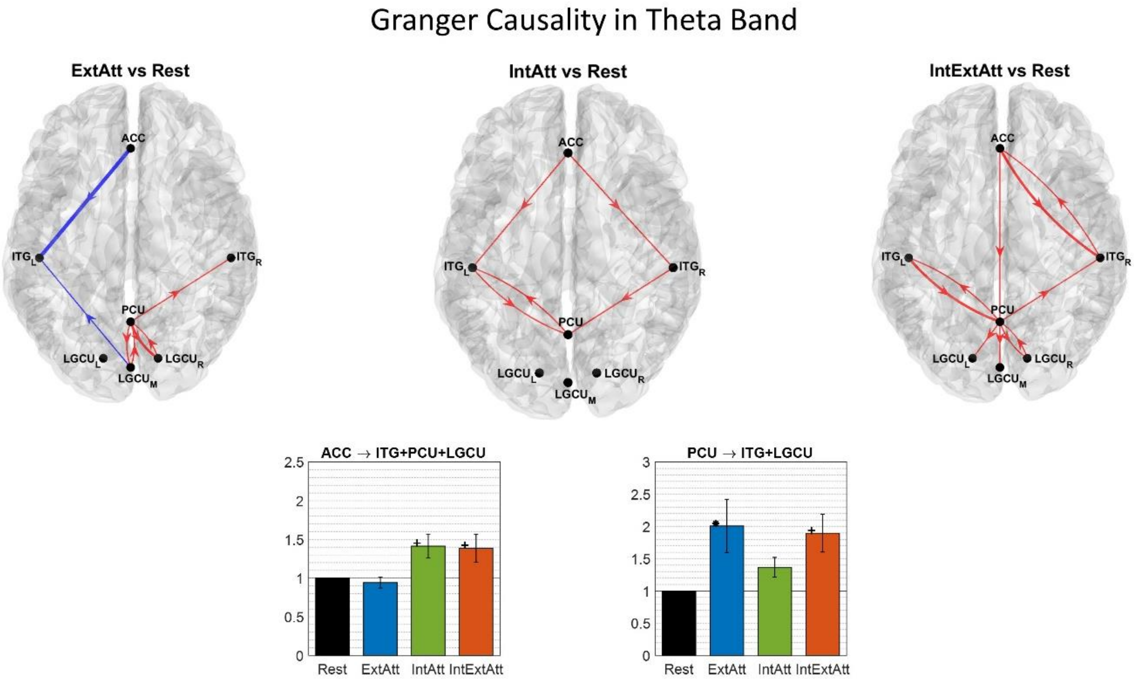


Figure 4.9 – Results of the spectral Granger Causality in the theta band between the selected ROIs. The arrows over the 3D cortical surfaces (top view of the cortex) indicate the connections that increase (red) or decrease (blu) in each attentional condition compared to rest. The displayed arrows correspond to significant connectivity changes at $p<0.05$ (uncorrected p-values, non-parametric permutation based t-test, see procedure (e) in Section 4.1.2.3.7) and the thickness of the line denotes three level of significance, i.e. thinnest line: $0.01 < \text{p-value} < 0.05$; middle line: $0.001 < \text{p-value} < 0.01$; thickest line: $\text{p-value} < 0.001$. The two bar plots display the overall causal influence that emerged from ACC and PCU, respectively, and targeted the set of indicated ROIs (LGCU includes the left, right and medial parts, and ITG includes the left and right part). Each bar represents the mean ± SEM across all participants. Within each bar plot, + denotes statistical comparison vs rest resulting in $p<0.05$ without Bonferroni correction, while * denotes statistical comparison vs rest resulting in $p<0.05$ with Bonferroni correction.

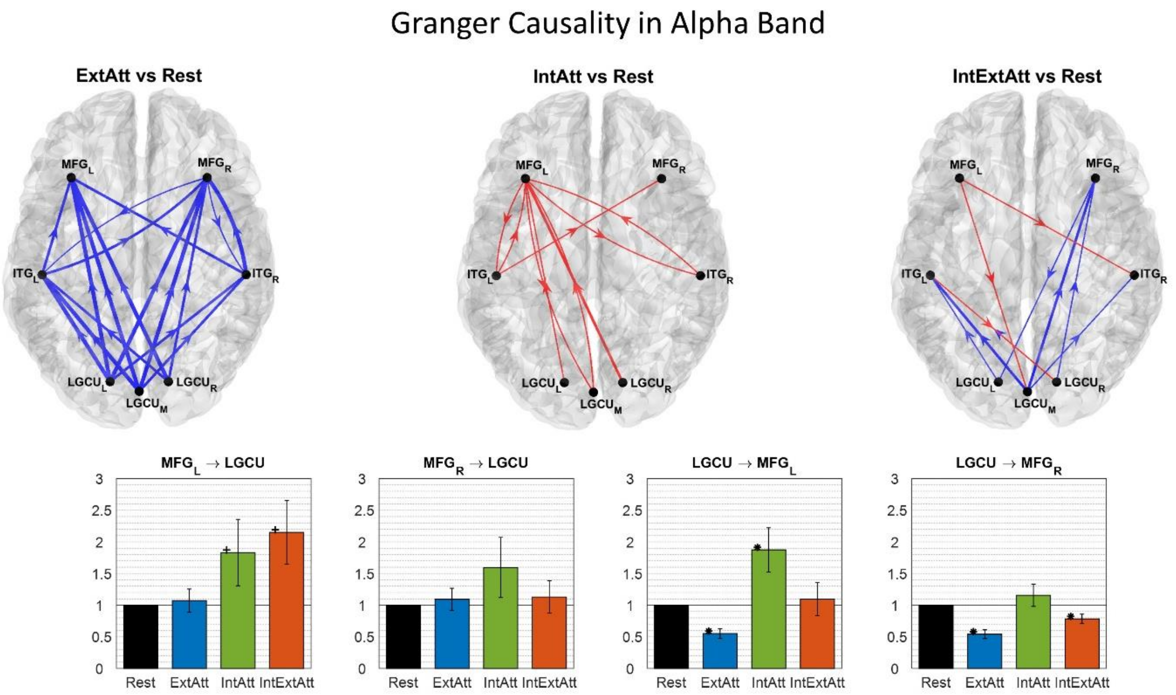


Figure 10 – Results of the spectral Granger Causality in the alpha band between the selected ROIs. The arrows over the 3D cortical surfaces (top view of the cortex) indicate the connections that increase (red) or decrease (blu) in each attentional condition compared to rest. The displayed arrows correspond to connectivity changes at p<0.05 (uncorrected p-values, non-parametric permutation based t-test, see procedure (e) in Section 4.1.2.3.7) and the thickness of the line denotes three level of significance, i.e. thinnest line: 0.01 < p-value < 0.05; middle line: 0.001 < p-value < 0.01; thickest line: p-value < 0.001. The bar plots display the overall causal influence, in both directions, between left and right MFG separately (MFG_L and MFG_R) and total LGCU region (bilateral and medial). Each bar represents the mean ± SEM across all participants. Within each bar plot, + denotes statistical comparison vs rest resulting in p<0.05 without Bonferroni correction, while * denotes statistical comparison vs rest resulting in p<0.05 with Bonferroni correction.

Table 4.1 – Number of voxels statistically significant within each of the selected ROI, and the largest value of the t-statistic observed on the ROI (reported only in case of significance). These comparisons are one-tailed; comparisons in the other direction did not provide any significance

	Theta-Band Analysis					
	ExtAtt>Rest		IntAtt>Rest		IntExtAtt>Rest	
ROI	n. voxels	t-value	n. voxels	t-value	n. voxels	t-value
ACC	-	-	69	3.18	3	3.05
ITG_L	-	-	-	-	61	3.44
ITG_R	-	-	-	-	33	3.12
PCU	34	4.97	-	-	3	2.90
$LGCU_L$	42	4.72	-	-	19	3.25

LGU$_{M}$	29	4.82	-	-	-	-
LGCU$_{R}$	49	4.34	-	-	-	-

	Alpha-Band Analysis					
	ExtAtt<Rest		IntAtt>Rest		IntExtAtt<Rest	
ROI	n. voxels	t-value	n. voxels	t-value	n. voxels	t-value
MFG$_{L}$	-	-	-	-	-	-
MFG$_{R}$	-	-	-	-	-	-
ITG$_{L}$	63	-3.18	-	-	-	-
ITG$_{R}$	37	-3.73	-	-	-	-
LGCU$_{L}$	157	-3.33	-	-	-	-
LGU$_{M}$	130	-3.33	-	-	-	-
LGCU$_{R}$	156	-3.37	-	-	-	-

4.1.4 Discussion

This study investigates how EEG alpha and theta power and functional communication within these bands are implicated in different attentional conditions, involving bottom-up sensory attention only (visual stimulation), cognitive internal attention only (mental math task), and the competition between the two forms of attention. In particular, we were motivated by the following main question: Which are the alpha and theta effects of *online* distractors that are continuously presented during an internal cognition task, and act by capturing external attention?

The electroencephalographic modulations in theta and alpha band were investigated both at scalp and cortical level. It is worth noticing that we tuned alpha and theta rhythms individually by using an objective procedure that emulated the original influential attempt of Klimesh et al. [352] to characterize the individual alpha band. This way we took into account the interindividual variability, enhancing the precision of the analysis.

4.1.4.1 Alpha and theta power changes at scalp level

Some interesting main effects were obtained already at the scalp level (Figures 4.2-4.4). An original result is that theta activity was modulated differently in the anterior and posterior sites by the attentional conditions. At frontal sites theta increase was mainly associated with internal attention, a result well in line with previous studies [327,331]; in particular, increase in scalp frontal midline theta was previously observed during mental arithmetic tasks both similar and different from the one adopted here [328,329,358]. This result at scalp level (together with the source-level results indicating the main involvement of anterior cingulate cortex in frontal theta increase) can represent a posterior validation that the participants were engaged in the math task execution during the two corresponding sessions. Interestingly, frontal theta increase was larger when the math was performed with the presence of visual distractors. This result matches the general observation of higher frontal theta associated with increased task demands, usually implemented by increasing memory load in working memory tasks or by increasing task complexity [326,373]. However, we are not aware of other studies that have investigated and revealed such an effect when the task demand is modulated by the absence vs presence of external distractors rather than by the intrinsic complexity of the internal task (which remained unchanged in IntAtt and IntExtAtt conditions). Here we found that although the internal task remained unchanged, the need to increase the protection of ongoing internal processing from external sensory interference was associated with higher frontal theta; this could reflect the allocation of greater internally-oriented attentional resources or increased error monitoring and conflict resolution.

Theta activity at scalp posterior sites did not follow the same pattern as frontally. Posterior theta increased in conditions involving external attention, i.e. in association with picture presentation. Studies investigating posterior theta in relation to visual attention and visual stimulation are sparser. Most of them reported scalp occipital theta enhancement following or during visual stimulus presentation [374–377], in agreement with our results. Conversely, the effect of visual attention on posterior theta is unclear and controversial results have been reported [375,377]. Our results provided similar levels of posterior theta increase in ExtAtt (when likely more attention was allocated towards visual stimuli) and in IntExtAtt (when likely less attention was directed towards visual stimuli); a speculative hypothesis is that this posterior theta might signal basic visual processing not affected by attention.

At variance with posterior theta, posterior alpha exhibited a clear different modulation in ExtAtt condition vs IntExtAtt condition. Picture presentation alone induced a strong posterior alpha decrease, indicating engagement of the sensory system and enhanced visual processing [333]. When pictures accompanied the mental task, alpha power exhibited a much lower decrease assuming values not statistically different from the rest; this indicates a resistance against the visual interfering information. This effect may result from a push-pull interaction between two opposing mechanisms: bottom-up visual engagement reducing alpha power (as it emerged in ExtAtt) and task-oriented visual inhibition increasing alpha power. Indeed, an average tendency of alpha power to increase occurred in IntAtt condition, although not significant. While this mechanism appeared weak when performing the internal task alone (against a grey screen), its operational role emerged more clearly in the presence of the visual distractors, being able to contrast alpha reduction. In our recent paper [378], which investigated alpha power modulations in a variety of conditions, we found that performing a similar mental task while immersed in a (highly stimulating) virtual reality environment increased alpha power to the same level as in rest condition, whereas the virtual reality immersion alone induced a strong decrease. Here, we confirmed the previous result and proved that even when using simpler and less motivating visual distractors (neutral pictures on a monitor vs virtual reality scenario) the need to isolate from visual engagement induced an increase in alpha oscillations. Our conclusion appears in contrast with a recent study [379] that questions the role of alpha oscillations in inhibition of online distractions. These authors observed that alpha did not increase compared to baseline during the retention interval of a working memory task, in the presence of visual distractors maintained throughout the retention period; furthermore, strong distractors (more similar to the memorized representation) produced larger alpha power decrease than weak distractors. Based on these results, the authors suggested that alpha oscillations were ineffective for inhibition of sustained distraction in the investigated condition. A possible reconciling interpretation may take into account that alpha power during internal vs external competition results from two antagonistic mechanisms, an increase to insulate internal processing and a decrease induced by external stimulation; the former may just reduce but not overcome the latter. This effect could occur in particular when the distractors contain features similar to the internal representation and thus may have strong attention influence.

4.1.4.2 Alpha and theta power changes at source level

Via the analysis at source level, we first investigated the involvement of some key cortical regions in theta and alpha power modulation and then we explored modifications in the pattern of functional connections. Here, source reconstruction was not based on a high-density electrode montage but on a limited number of electrodes (32). Therefore, the results obtained on source activity and connectivity merit additional studies to obtain a more robust validation (see also Section 4.1.4.4).

For the alpha-band analysis, we took into consideration both lower-level (lingual gyrus and cuneus) and higher-level (inferior temporal gyrus) visual regions, based on their functional roles and of previous results in literature (see Section 4.1.1 and Section 4.1.2.3.5). The lingual gyrus and cuneus were considered together (LGCU) in order to reduce the number of ROIs, especially in view of the connectivity analysis, and favor a more straightforward interpretation of the results.

The whole-brain statistical analysis reveals that almost all voxels in the posterior regions are implicated in alpha modulations (Fig. 4.6). More specifically, alpha power modulation in the selected ROIs (Fig. 4.8) paralleled that observed posteriorly at scalp level. In particular, ExtAtt condition was characterized by a significant alpha power decrease in all visual ROIs; notably, the decrease was larger in earlier (LGCU) than superior (ITG) visual ROIs. No significant modulation of the alpha power in the visual ROIs was found in IntExtAtt compared to baseline. Similarly to our results, a recent study [332] reconstructing source activity from magnetoencephalography, found a marked alpha reduction, especially in occipital bilateral regions, during a prospective memory task that required elevated external attention and low internal attention. Conversely, they did not find significant changes of cortical alpha in a task that required lower external monitoring and high internal attention. Again, we interpret the absence of alpha changes in that task as resulting from the balance between the alpha increase to protect the internal representations and the alpha decrease still induced by the external sensory inputs.

Besides the visual ROIs, we also included the middle frontal gyrus in alpha band analysis, since this region, although did not exhibit a significant alpha-power change neither at the voxel-level nor at the overall ROI level, may be implicated in top-down influences in the alpha-band toward the visual ROIs, especially in conditions involving internal attention. Indeed, there is a general agreement that prioritizing or suppressing sensory processing in a goal-oriented manner are controlled by top-down signals from higher-level frontal cortical areas [380,381], although no consensus has emerged yet on the specific sources of these signals. However, implication of the middle frontal gyrus (including the dorsolateral prefrontal cortex) in attention-dependent modulation of visual activity is supported both by fMRI and ERP studies [382,383], and other EEG studies [346,356,357].

For the theta-band analysis, we included not only the visual ROIs (LGCU, ITG) but also two additional ROIs, the Anterior Cingulate Cortex (ACC) and the Precuneus (PCU). The ACC was a priori included as one of the main source of frontal midline theta associated with internally focused attention and cognitively demanding task [326,331], including mental arithmetic [328,329,358]. Our results suggest that the frontal theta increase in our math task originated mainly from the ACC, and from this hub it might potentially extend to other frontal regions (see indeed right top cortical surface in Fig. 4.5 showing higher and broader theta increase in frontal cortex, where significant voxels were found too). Conversely, the Precuneus was selected a posteriori based on the statistical voxel-wise results in ExtAtt condition, which showed a significant involvement of this region (Fig. 4.5 left slice-view maps and Table 4.1). Accordingly, the analysis at overall ROI level showed significant power modulation in PCU, suggesting that PCU could be implicated in theta power and that activity could spread from it to the close cingulate cortex (where significant voxels were found too). However, interpreting the involvement of PCU in ExtAtt is not straightforward and our explanation remains at a highly speculative level. Indeed, the precuneus has been widely associated to internally oriented functions and memory processes (e.g. visual imagery, episodic or semantic memory retrieval) [359,360]. A preliminary interpretation, requiring further support, can be linked to a few studies suggesting that PCU has also visual functions [361,362] and may be implicated in general monitoring of external environment [384]. We can tentatively speculate that, at least in the conditions investigated here, theta increase in PCU and in the other visual regions (especially bilateral LGCU) might represent a basic visual processing uninfluenced by the amount of allocated attention (see the similar level attained in ExtAtt and IntExtAtt in these ROIs, Fig. 4.7). However, this interpretation remains uncertain and definitely requires further inquiries.

4.1.4.3 Functional connectivity in alpha and theta band

When considering internal attention alone, causal theta influences increased from ACC towards temporal and then posterior (PCU) regions (Fig. 4.9), in agreement with previous results showing the importance of fronto-temporal-parietal theta networks in mental arithmetic [329,358]. Conversely, in the ExtAtt condition theta connectivity appeared much more localized and confined posteriorly. Furthermore, the overall theta connectivity increased in IntExtAtt, condition, with the two previous mechanisms operating simultaneously, and apparently not influencing reciprocally.

Mechanisms of alpha-band causal influences seemed to operate in opposite direction in external and internal attention (Fig. 4.10). ExtAtt was dominated by a decrease in alpha-band connections in a bottom-up fashion. On the contrary, IntAtt was characterized by an increasing trend of alpha-band connections in both direction. A few considerations can be drawn. First, the selected frontal ROI (MFG, particularly in the left side) appeared implicated in top-down alpha influences during the internal task. Second, this mechanism was accompanied by a second one, i.e. an increase in bottom-up alpha influences (Fig. 4.10 central top panel). It is worth noting that, besides top-down, also posterior-to-anterior information flow in alpha band has been reported to be involved in internally oriented tasks and conditions [385,386]. This is in line also with our results in IntExtAtt condition where an increase in top-down connectivity was accompanied by a reduced decrease of bottom-up connectivity compared with ExtAtt. Overall, these results might indicate that both top-down and bottom-up communication in alpha band possibly operate to favor insulation of internal processing from ongoing external interference.

4.1.4.4 Limitations and future directions

Some limitations and ideas for further studies are presented.

In our experimental protocol, we used pictures changing every 10 seconds to induce visual attention. The significant and marked decrease of alpha activity in posterior visual regions during ExtAtt condition indicates that visual attention of participants was actually captured by the pictures. However, we cannot totally exclude that mind-wandering occasionally occurred, e.g. at some points between one picture and the next, reducing visual attention. Increasing the rate of picture presentation (e.g. every 5 seconds or less) or watching a video, might increase the effect of external attention capture, with possible impact on alpha power both in ExtAtt and IntExtAtt conditions. This could be explored in future studies. Another point is that the two examined conditions (performing mental arithmetic vs watching pictures) differ, besides the direction of attention, also as to other variables, such as voluntary vs automatic allocation of attention, requiring vs not requiring manipulation of information, demanding vs not demanding task. These variables might have partially contributed to the observed EEG differences in oscillatory activity and connectivity. However, our hypothesis is a posteriori validated by the observation that, according to the existing literature, the obtained changes in alpha and theta mechanisms are mainly modulated by the direction of attention, i.e. are specific of attention orientation rather than being task-specific or being associated to other variables (see the Introduction where we emphasized how similar alpha or theta mechanisms operate in tasks even very different but that share the same direction of attention).

Another important aspect concerns the investigation of other rhythms together with theta and alpha, e.g. gamma rhythm (30-100 Hz). Indeed, enhanced gamma oscillations have been observed during high-level (internally oriented) mental processing (reading, emotion, math task, memory) especially in task-relevant areas [327,328,387], and also increased gamma oscillations in sensory cortices have often been linked with increased sensory attention [325]. Furthermore, theta-gamma coupling, with gamma cycles nested within a theta cycle and phase-locked to it, is considered an efficient scheme for implementing the simultaneous representation and manipulation of multiple items. In relation to our results, adding the exploration of gamma oscillations may help to understand the functional role of the posterior theta in the investigated conditions.

We acknowledge that the accuracy of source reconstruction improves with increasing sensor density and using individualized head models from MRI, while we used 32 electrodes and a head model template (a limitation however common to a great body of literature and even severer in cases when just 19 electrodes are used). However, here we were not interested in an exact cortical localization, and we did not focus on single significant voxels; instead, we considered overall regions of interest, to characterize the average behavior of an area. Therefore, we expect that some blurring and inaccuracies in source reconstruction derived by the adopted techniques may have had a tolerable impact. Furthermore, some reassurances come from results at source level in concordance with the existing literature (as reported previously). Nevertheless, our source-level results, both as to power and connectivity, definitely deserve further studies using high-density electrode montage (e.g. 64 electrodes) and possibly a larger sample size (see also below) to achieve higher reliability. This may be of particular relevance to attest our provisional result showing involvement of theta activity in PCU during visual attention; this result indeed does not find clear interpretation within the existing literature and further validation is required.

Another weakness, that may benefit from a larger number of electrodes and of participants, concerns the low statistical significance of the Granger Causality effects that did not survive multiple comparison correction in most of cases (Fig. 4.9 and Fig. 4.10) and revealed a high inter-individual variability. A certain variability was already visible at power level (especially in alpha activity in IntAtt and IntExtAtt); it is possible that some participants adopted different strategies for keeping internal concentration or learned automatized procedure to perform the task (reducing the effort) or may even be particularly susceptible to external influences. In future studies we can further increase the number of participants to identify a clearer tendency and also perform an analysis at single-subject level to possibly disclose different adopted strategies. This kind of analysis could further benefit from controlling participants' performances in the mental task and/or their arithmetic proficiency, which we did not systematically assess in this study.

4.1.5 Conclusions

In conclusions, we investigated the alpha and theta mechanisms that operate while performing an internal attention task against concurrent visual distractors inducing external attention, in relation to the mechanisms operating when considering each form of attention individually. While theta and alpha are extensively linked to attention, they are less often investigated together, and are scarcely examined in relation to online distractors and during ongoing conditions reflecting real life situations. We found that performing the internal task during external visual interferences was associated with distinctive patterns of power and connectivity in both bands. Frontal midline theta, implicating anterior cingulate cortex and peculiar of internal focus of attention, was further enhanced in presence of visual distractors and was associated with large-scale top-down connectivity. This anterior theta coexisted with posterior (occipital/midline-parietal) theta, which was characterized by localized connectivity and could reflect basic visual processing, although this interpretation remains highly uncertain. Alpha power in visual regions (both temporal and occipital), characterized by a significant decrease in conditions of visual stimulation alone, assumed values closed to rest in the competition state, indicating reduced engagement of the visual system to gate out visual distractors. This appears effected by a bidirectional increase in alpha connectivity between visual regions and frontal regions.

Despite some limitations, this study has made an attempt to provide a comprehensive framework of the alpha and theta oscillatory mechanisms that intervene in the competition internal-external attention, a condition pervasive of the sensory-cognitive interplay in our everyday activities. This attempt can also be of value to reconcile controversial results of attention interpreting them within this framework, and to help understanding the relationships between brain oscillations and attentional functions/dysfunctions in daily life tasks, especially in subject categories more susceptible of external intrusion during cognitive tasks, such as children, older, or subjects suffering from attention-deficit hyperactivity disorder (ADHD).

4.1.6 Supplementary Material

4.1S. Assignment of source space voxels to the selected ROIs

In the LORETA-KEY$^{©®}$ software, each of the 6239 voxels is provided with its x,y,z MNI coordinates, and with the lobe, region and Brodmann area the voxel belongs to. Therefore, the assignment of a voxel to a ROI was based on the label specifying its region (in our study: Lingual Gyrus (LG), Cuneus (CU), Inferior Temporal Gyrus (ITG), Middle Frontal Gyrus (MFG), Anterior Cingulate (ACC), Precuneus (PCU)). The assignment of a voxel to the left, right or medial part of a ROI was based on the x coordinate provided by LORETA-KEY$^{©®}$ software; the x coordinate is directed from left to right, with 5 mm resolution, and assumes negative values for voxel in the left hemisphere and positive values for voxel in the right hemisphere (the origin of MNI coordinate system is located at the anterior commissure). Specifically, we performed the following assignment: voxels belonging to a region (e.g. CU or LG) were assigned to the left part of that ROI (i.e. $LGCU_L$) if they have $x < -5$ mm, to the right part of that ROI ($LGCU_R$) if they have $x > +5$ mm and to the medial part of that ROI ($LGCU_M$) if they have -5 mm $\leq x \leq +5$ mm. Some ROIs do not have the medial part (ITG and MFG). In our study, for the ROIs ACC and PCU we considered only the medial part (hence, we did not use the subscript). Indeed, the Anterior Cingulate region lies mostly medially and only a few left and right voxels remained excluded. On the contrary the overall Precuneus region extends also superiorly on the left and right even beyond the LGCU region; however, we preferred to maintain the selection limited to the medial portion (see Figure 4.1) since the results motivating its selection (voxel-wise statistical analysis contrasting ExtAtt vs Rest in Theta Band, see Figure 4.5) suggested that the medial part was mainly involved. The Supplementary Table 4.1S provides the number of voxels for each investigated ROI.

4.2S. Selection of model order for the bivariate autoregressive (BVAR) models of the data in the computation of spectral Granger Causality

We did not use the Akaike Information Criterion to identify the model order for the AR model in this study, since in some preliminary tests we found that the minimum of this criterion often settled to low values of the model order far to reproduce the data well (a problem already observed previously, see for example[346]). Hence, to select the model order, we compared the power spectral estimates obtained by the BVAR models and those estimated directly from the temporal ROIs waveform via the Welch's method. This was done for different model orders (from 5 to 45 with step of 5) and the mean squared difference between the two estimates was computed (within the range 3-20 Hz) and averaged across the ROIs and participants. We maintained separated the assessment for the frontal-temporal ROIs (ACC, MFG_L, MFG_R, ITG_L, ITG_R) and for the posterior ROIs (PCU, $LGCU_L$, $LGCU_M$, $LGCU_R$) since the error showed a different rate of decrease in the two sets of ROIs, as shown in the Supplementary Figure 4.1. These results show that at model order 30, both errors settled at a lower saturation level. Actually, the error for the posterior ROIs reached the saturation level before (at model order 20) and then fluctuated around it; however, using smaller value for the model order (e.g. 25) caused the AR model to miss the theta peak in the frontal ROIs (see Supplementary Figure 4.2). For completeness, Supplementary Figures from 4.S2 to 4.S10 showed, for each selected ROI, the PSDs estimated directly on the actual ROI signals (Panel A) and the comparison between them and those obtained from the AR models with model order 30 (Panel B). Results in all panels are averaged across the participants.

4.3S. Results of the voxel-wise statistical analysis

Supplementary Tables 4.2S-4.4S refer to the results of the voxel-wise statistical analysis as to the theta band (see Figure 4.5) while Supplementary Tables 4.5S-4.6S as to the alpha band (see Figure 4.6). For each statistical comparison, the Table 4.1S. reports: the regions to which the significant voxels belong (based on the region label provided for each voxel in the software LORETA-KEY$^{©®}$), the number of statistically significant voxel in each region (distinguishing between the left, right, medial subdivision as described in Supplementary Section 4.1S) and the largest value of the observed t-statistic on the whole region (irrespective of the side). It is worth noticing that these comparisons are one-tailed. Comparisons in the other direction did not provide any significance.

Table 4.1S: Number of voxels for each investigated ROI according to the procedure of voxel assignment described in Supplementary section 4.1S.

ROI	***number of voxels***
$LGCU_L$	158
$LGCU_M$	142
$LGCU_R$	157
PCU	122
ITG_L	78
ITG_R	73
MFG_L	237
MFG_R	245
ACC	100

Table 4 2S: Results of the voxel-wise statistical analysis comparing the theta power between ExtAtt vs Rest (ExtAtt > Rest)

Region	**n. voxels: L/M/R**	**t-statics on the ROI**
Cingulate Gyrus	7/21/5	5.06
Cuneus	12/10/5	4.82
Fusiform Gyrus	0/-/6	3.44
Lingual Gyrus	30/19/44	4.62
Middle Occipital Gyrus	2/-/0	3.38
Parahippocampal	4/-/4	4.03
Posterior Cingulate	18/47/15	5.36
Precuneus	34/34/22	5.52

Table 4.3S: Results of the voxel-wise statistical analysis comparing the theta power between IntAtt vs Rest (IntAtt > Rest)

Region	**n. voxels: L/M/R**	**t-statics on the ROI**
Anterior Cingulate	15/69/4	3.18
Cingulate Gyrus	8/21/4	2.69
Inferior Frontal Gyrus	122/-/0	3.18
Insula	17/-/0	3.41
Medial Frontal Gyrus	19/28/1	2.85
Middle Frontal Fyrus	42/-/0	3.04
Orbital Gyrus	6/-/0	2.57
Parahippocampal Gyrus	14/-/0	3.06

Rectal Gyrus	9/3/0	2.76
Subcallosal Gyrus	8/8/0	3.09
Superior Frontal	5/1/0	2.52
Superior Temporal Gyrus	39/-/0	2.84

Table 4.4S: Results of the voxel-wise statistical analysis comparing the theta power between IntExtAtt vs Rest (IntExtAtt > Rest)

Region	**n. voxels: L/M/R**	**t-statics on the ROI**
Anterior Cingulate	0/2/1	3.05
Cingulate Gyrus	2/0/0	3.46
Cuneus	10/0/0	3.25
Fusiform Gyrus	93/-/65	3.39
Inferior Frontal Gyrus	1/-/5	3.02
Inferior Parietal Gyrus	0/-/4	2.91
Inferior Temporal Gyrus	61/-/33	3.44
Insula	10/-/11	3.48
Lingual Gyrus	9/0/0	3.14
Medial Frontal Gyrus	0/7/5	3.06
Middle Occipital	27/-/9	3.48
Middle Temporal Gyrus	99/-/18	3.62
Parahippocampal	77/-/75	3.61
Postcentral Gyrus	2/-/9	3.43
Posterior Cingulate	19/1/0	3.49
Precentral Gyrus	2/-/19	3.27
Precuneus	25/3/0	3.57
Superior Frontal Gyrus	0/2/5	3.02
Superior Temporal Gyrus	54/-/5	3.19

Table 4.5S: Results of the voxel-wise statistical analysis comparing the alpha power between ExtAtt vs Rest (ExtAtt < Rest)

Region	**n. voxels: L/M/R**	**t-statics on the ROI**
Angular Gyrus	15/-/0	-3.15
Cingulate Gyrus	9/41/18	-3.17
Cuneus	89/87/91	-3.33
Fusiform Gyrus	111/-/99	-3.68
Inferior Occipital Gyrus	17/-/19	-3.26
Inferior Parietal Gyrus	7/-/136	-3.35
Inferior Temporal Gyrus	63/-/37	-3.73
Insula	5/-/38	-3.35
Lingual Gyrus	68/43/65	-3.37
Middle Occipital Gyrus	74/-/67	-3.66
Middle Temporal Gyrus	103/-/113	-3.75
Parahippocampal Gyrus	89/-/90	-3.76
Postcentral Gyrus	2/3/119	-3.05
Posterior Cingulate	19/40/19	-3.27
Precuneus	104/108/120	-3.24
Superior Occipital Gyrus	8/-/9	-3.21

Superior Parietal Lobule	30/2/65	-3.1
Superior Temporal Gyrus	39/-/95	-3.72
Supramarginal Gyrus	10/-/24	-3.47

Table 4.6S: Results of the voxel-wise statistical analysis comparing the alpha power between IntExtAtt vs Rest (IntExtAtt < Rest)

Region	n. voxels: L/M/R	t-statics on the ROI
Inferior Parietal Gyrus	0/-/20	-3.06
Superior Parietal Lobule	0/-/2	-3

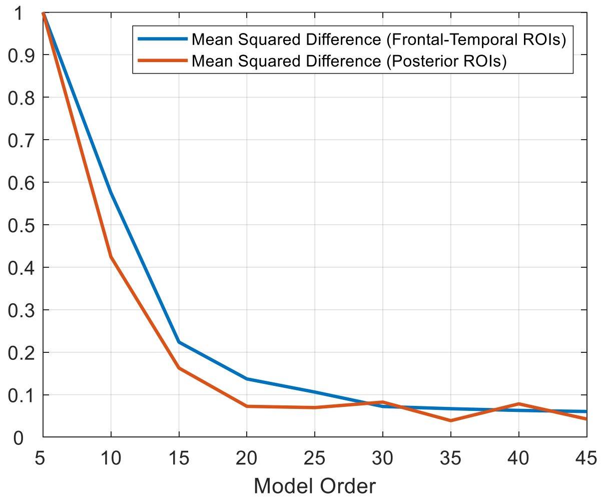


Figure 4.1S – Mean squared difference (error) between the PSDs estimated directly on the actual ROIs data and those obtained by the BVAR models. The model order was increased from 5 to 45 with step of 5. The represented error values were obtained in the frequency range 3-20 Hz and averaged across all ROIs (in the two sets, fronto-temporal and posterior) and participants; furthermore for clarity of visualization, they were normalized with respect to the maximum error value (at model order 5).

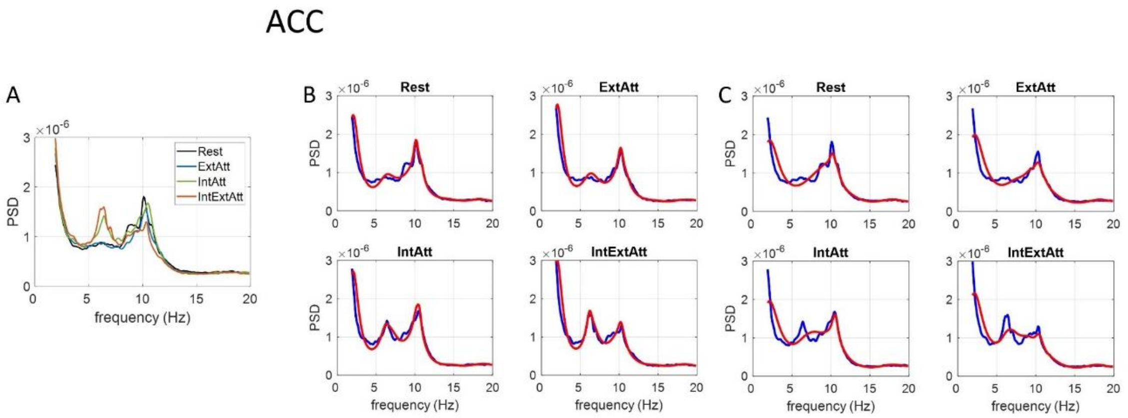


Figure 4.2S – Panel A: PSDs estimated directly on the ROI (ACC) signal in the four conditions (Rest, ExtAtt, IntAtt, IntExtAtt). **Panel B** – Comparison between the PSD estimated directly on the ROI signal (blue lines, the same curves as in Panel A) and the PSD obtained from the AR model of the ROI signal using model order 30 (red lines). **Panel C**–The same as in Panel B but using model order 25, showing that the theta peak was not well reproduced by the AR model. Values are in (μA/mm^2)^2/Hz.

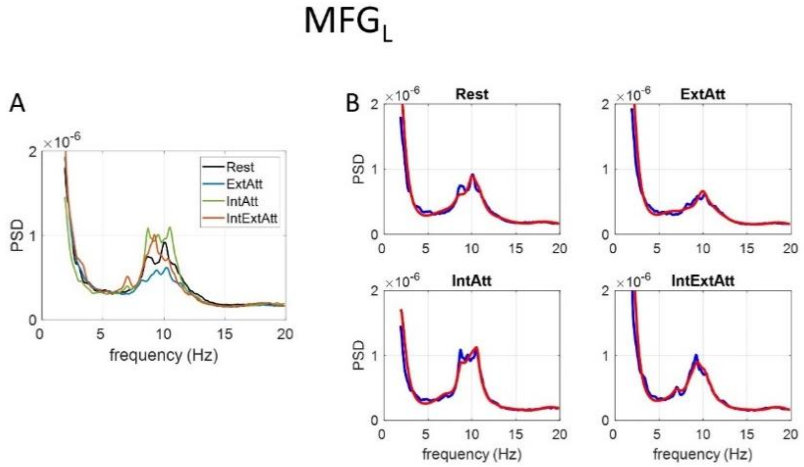


Figure 4.3S - Panel A: PSDs estimated directly on the ROI (MFGL) signal in the four conditions (Rest, ExtAtt, IntAtt, IntExtAtt). **Panel B** – Comparison between the PSD estimated directly on the ROI signal (blue lines, the same curves as in Panel A) and the PSD obtained from the AR model of the ROI signal using model order 30 (red lines). Values are in (µA/mm^2)^2/Hz.

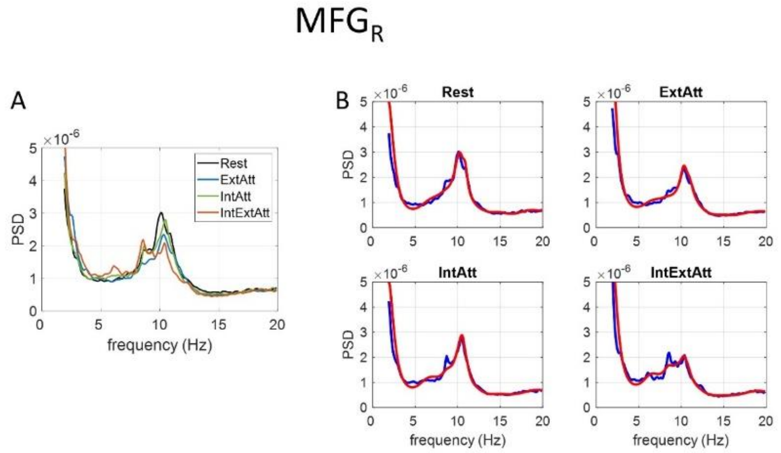


Figure 4.4S - Panel A: PSDs estimated directly on the ROI (MFGR) signal in the four conditions (Rest, ExtAtt, IntAtt, IntExtAtt). **Panel B** – Comparison between the PSD estimated directly on the ROI signal (blue lines, the same curves as in Panel A) and the PSD obtained from the AR model of the ROI signal using model order 30 (red lines). Values are in (µA/mm^2)^2/Hz.

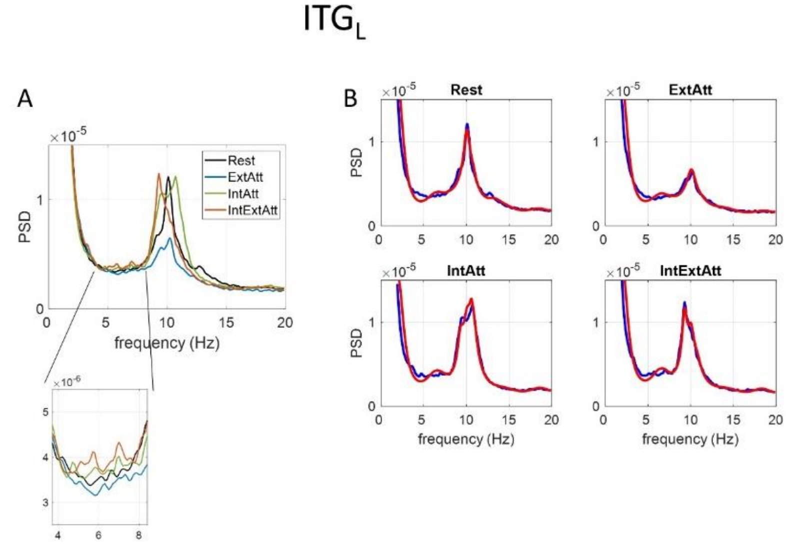


Figure 4.5S - Panel A: PSDs estimated directly on the ROI (ITG$_L$) signal in the four conditions (Rest, ExtAtt, IntAtt, IntExtAtt). The zoom in focuses on the portion roughly corresponding to the theta band. **Panel B** – Comparison between the PSD estimated directly on the ROI signal (blue lines, the same curves as in Panel A) and the PSD obtained from the AR model of the ROI signal using model order 30 (red lines). Values are in (μA/mm^2)^2/Hz.

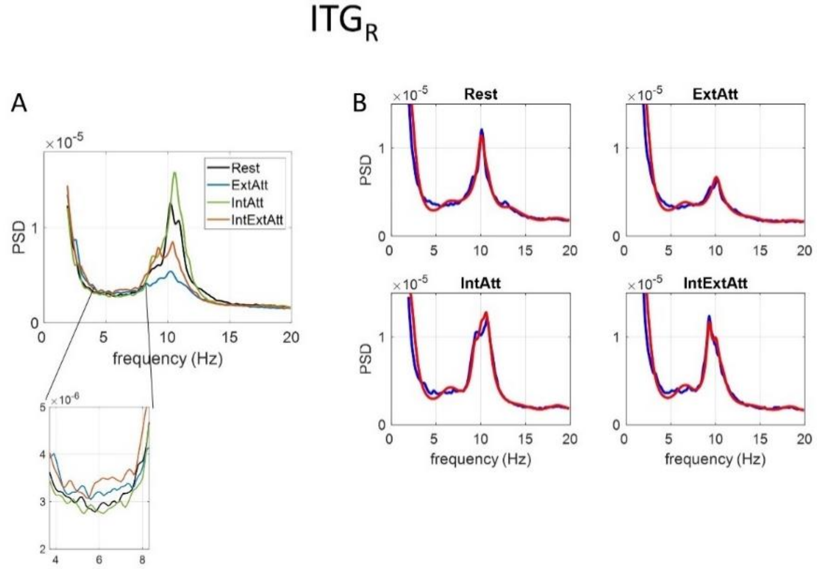


Figure 4.6S - Panel A: PSDs estimated directly on the ROI (ITG$_R$) signal in the four conditions (Rest, ExtAtt, IntAtt, IntExtAtt). The zoom in focuses on the portion roughly corresponding to the theta band. **Panel B** – Comparison between the PSD estimated directly on the ROI signal (blue lines, the same curves as in Panel A) and the PSD obtained from the AR model of the ROI signal using model order 30 (red lines). Values are in (μA/mm^2)^2/Hz.

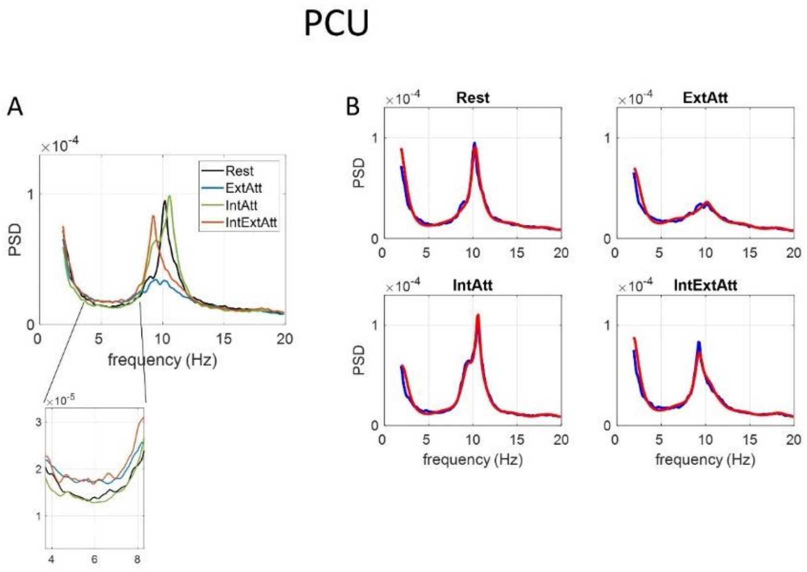


Figure 4.7S - Panel A: PSDs estimated directly on the ROI (PCU) signal in the four conditions (Rest, ExtAtt, IntAtt, IntExtAtt). The zoom in focuses on the portion roughly corresponding to the theta band. **Panel B** – Comparison between the PSD estimated directly on the ROI signal (blue lines, the same curves as in Panel A) and the PSD obtained from the AR model of the ROI signal using model order 30 (red lines). Values are in (μA/mm^2)^2/Hz.

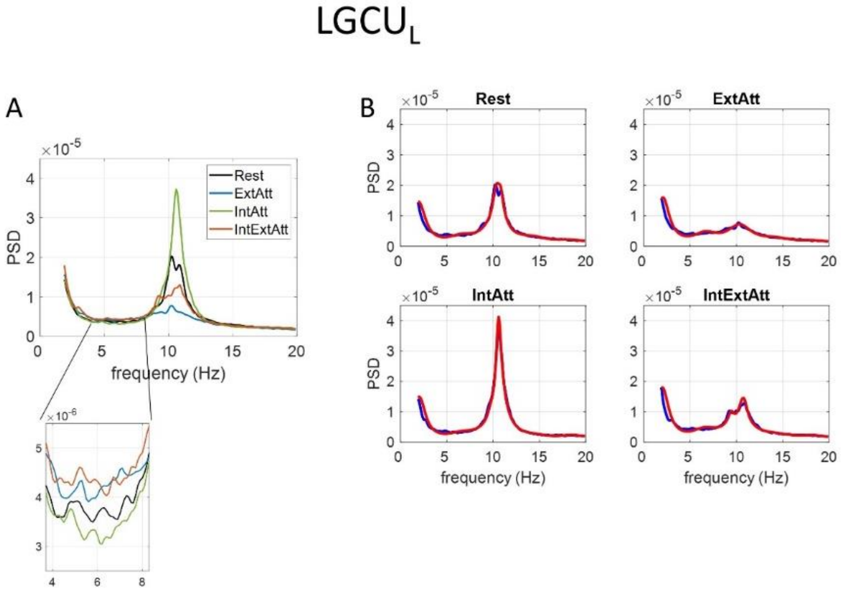


Figure 4.8S - Panel A: PSDs estimated directly on the actual ROI (LGCU$_L$) signal in the four conditions (Rest, ExtAtt, IntAtt, IntExtAtt). The zoom in focuses on the portion roughly corresponding to the theta band. **Panel B** – Comparison between the PSD estimated directly on the ROI signal (blue lines, the same curves as in Panel A) and the PSD obtained from the AR model of the ROI signal using model order 30 (red lines). Values are in (μA/mm^2)^2/Hz.

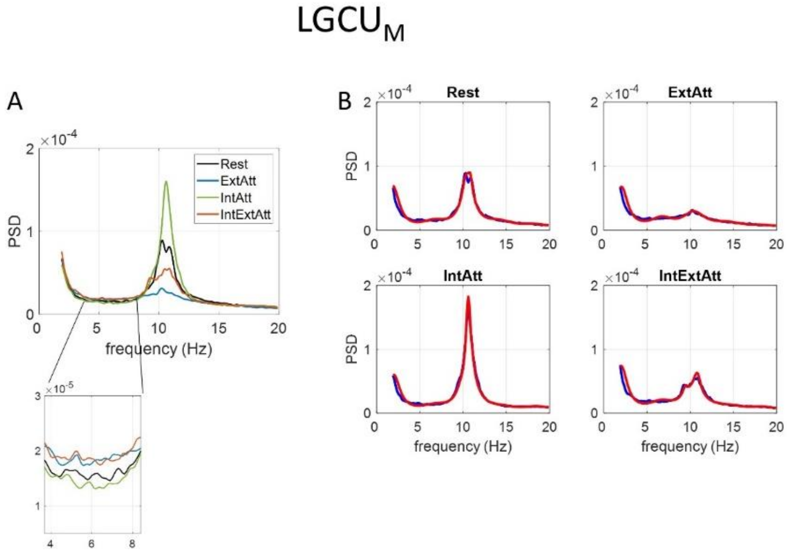


Figure 4.9S - Panel A: PSDs estimated directly on the ROI (LGCU$_M$) signal in the four conditions (Rest, ExtAtt, IntAtt, IntExtAtt). The zoom in focuses on the portion roughly corresponding to the theta band. **Panel B** – Comparison between the PSD estimated directly on the ROI signal (blue lines, the same curves as in Panel A) and the PSD obtained from the AR model of the ROI signal using model order 30 (red lines). Values are in (μA/mm^2)^2/Hz.

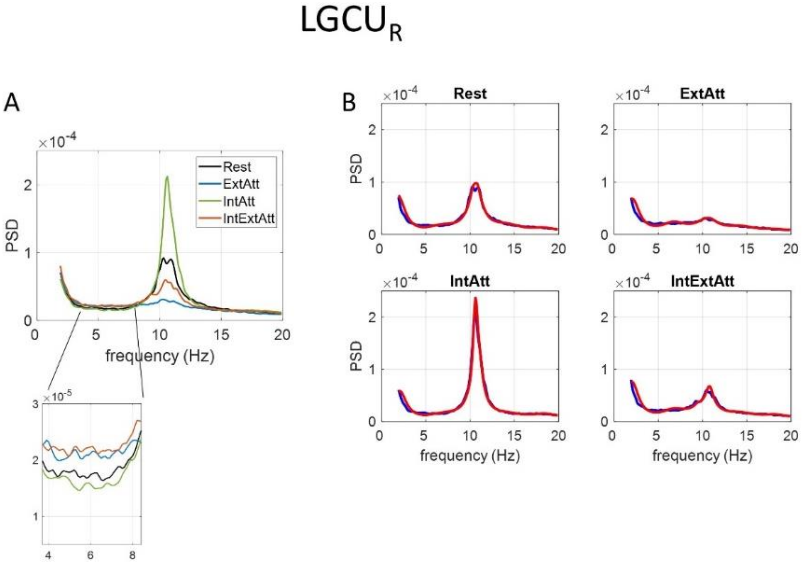


Figure 4.10S - Panel A: PSDs estimated directly on the ROI (LGCU$_R$) signal in the four conditions (Rest, ExtAtt, IntAtt, IntExtAtt). The zoom in focuses on the portion roughly corresponding to the theta band. **Panel B** – Comparison between the PSD estimated directly on the ROI signal (blue lines, the same curves as in Panel A) and the PSD obtained from the AR model of the ROI signal using model order 30 (red lines). Values are in (μA/mm^2)^2/Hz.

4.2 **Brain rhythms power and connectivity modulation in a Pavlovian fear conditioning and reversal task**

The study reported in this chapter refers to a manuscript entitled "Changes in brain rhythms and connectivity tracking Fear Acquisition and Reversal", Gabriele Pirazzini[1], Francesca Starita[2], Giulia Ricci[1], Sara Garofalo[2], Giuseppe di Pellegrino[2], Elisa Magosso[1], Mauro Ursino[1*], submitted to Brain Structure and Function (2022).

In this study, we analysed EEG power and Spectral Granger Causality differences during a protocol of Pavlovian fear conditioning and reversal. The objective was to investigate the task-dependent: a) role of brain rhythms, b) the main regions involved and c) how the rhythms are transmitted in the brain. The main brain rhythms involved in fear acquisition, and its reversal, were found to be in the theta and alpha frequency bands. In this study, functional brain network analysis focused on the connections exiting from the most significant cortical areas, identified through power analysis, in order to investigate how information is transmitted from a generative node to other nodes of the network.

Background: *Fear conditioning is used to investigate the neural bases of threat and anxiety, and to understand their flexible modifications when the environment changes. This study aims to examine the temporal evolution of brain rhythms using electroencephalographic signals recorded in healthy volunteers during a protocol of Pavlovian fear conditioning and reversal.* ***Methods:*** *Power changes and Granger connectivity in theta, alpha, and gamma bands are investigated from neuroelectrical activity reconstructed on the cortex.* ***Results****: Results show a significant increase in theta power in the left (contralateral to electrical shock) portion of the midcingulate cortex during fear acquisition, and a significant decrease in alpha power in a broad network over the left posterior-frontal and parietal cortex. These changes occur since the initial trials for theta power, but require more trials (3/4) to develop for alpha, and are also present during reversal, despite being less pronounced. In both bands, relevant changes in connectivity are mainly evident in the last block of reversal, just when power differences attenuate. No significant changes in the gamma band were detected.* ***Discussion****: We conclude that the increased theta rhythm in the cingulate cortex subserves fear acquisition and is transmitted to other cortical regions via increased functional connectivity allowing a fast theta synchronization, whereas the decrease in alpha power can represent a partial activation of motor and somatosensory areas contralateral to the shock side in the presence of a dangerous stimulus. In addition, changes at the end of reversal may reflect long-term alterations in synapses necessary to reverse the previously acquired contingencies.*

4.2.1 Introduction

Fear conditioning is a paradigm used in neuroscience to study the neurobiological bases of threat and anxiety [388–390]. In this experimental protocol, a neutral stimulus (conditioned stimulus, CS+) is paired with an aversive stimulus (unconditioned stimulus, US) to trigger neural adjustments at the basis of fear acquisition, finally resulting in the expression of a fear response in the presence of the CS+ alone. In some experiments, a second conditioned stimulus (CS-) is unpaired with the US to serve as a control condition.

Results in rodents, primates, and humans provide a precise scenario showing the amygdala, hippocampus, and medial prefrontal cortex (mPFC) involvement in fear acquisition [391–399]. Furthermore, brain oscillations in the theta range (4-8 Hz) play a pivotal role in this process [400]. In rodents enhanced theta synchrony between the amygdala and mPFC has been observed, which differentiated between threat and safety [401]. In rats, behavioral fear expression, like freezing, coincides with internally generated theta oscillations in prefrontal-amygdala circuits [402]. In primates, theta power and coherence in the amygdala and anterior cingulate cortex increase during fear learning, but progressively decline once the association is stabilized [403]. In humans, neuroimaging studies suggest that the recall of conditioned fear involves the anterior midcingulate cortex (AMC), which exhibits more robust activation to fear-conditioned stimuli [404–406]. Source localization from electroencephalographic (EEG) signals reveals that fear-conditioned stimuli evoke significantly more theta activity in the AMC than not fear-conditioned stimuli [93].

Besides theta oscillations increase, a decrease in alpha power (8-14 Hz) is also observable during fear conditioning, especially in the first block of the stimulus train. These changes have been mainly reported in the parietal and occipital channels [407,408], and seem to reflect the valence (i.e., unpleasantness) and salience (i.e., relevance) of the stimulus, rather than fear conditioning per se. Bacigalupo and Luck [409] observed that alpha-band suppression is greater for the CS+ compared to the CS- during fear acquisition, and that this effect is reduced during extinction. Babiloni et al. [410] suggested that greater alpha power reduction occurs during anticipatory processes preceding the integration of painful and motor information, compared with painful stimuli which do not require motor tasks.

However, to understand fear mechanisms, it is also essential to clarify how this response can be flexibly adjusted depending on varying environmental conditions [411–413]. Substantial knowledge of this process is provided by extinction protocols, in which a previously conditioned subject is exposed to CS+ in the absence of the aversive input. Extinction is characterized by a progressive decrease in the response to CS, while deficits in extinction mechanisms are associated with pathological states like post-traumatic stress disorder and anxiety [414–416]. Neuroimaging studies in humans underline the involvement of the ventromedial prefrontal cortex (vmPFC) in extinction [417,418]. EEG studies demonstrate that theta oscillations in the dorsal anterior cingulate cortex are reduced during successful extinction recall, probably involving an interplay between the amygdala and front medial theta activity [400]. A pivotal role in extinction seems related to gamma oscillations (>30 Hz) localized in the vmPFC, since extinguished stimuli evoke greater vmPFC gamma power than not-extinguished stimuli [93].

Reversal learning is an alternative way to study flexibility. This protocol is frequently adopted to analyze the classic reward-based action selection (i.e., decision making) involving the dopaminergic system and basal ganglia plasticity [419]; however, only a few studies have explored reversal during fear conditioning. During reversal learning, a subject must be able to modify the previously learned stimulus-outcome association by inhibiting a previous response in favor of a new one. [420] pointed out fear reversal is more demanding than fear extinction. In fact, during reversal, not only the old aversive stimulus must be extinguished, but also a new stimulus (the old CS-) acquires an aversive value. Hence, using reversal learning, one can examine how a fear response is weakened while another fear response is simultaneously acquired, allowing a concurrent comparison between the two mechanisms and favoring a better understanding of their neurological bases.

There is a large consensus that, during action-selection tasks, reversal learning involves the ventral PFC, especially the orbitofrontal cortex. Increased activation in this area seems to be predominantly associated with unexpected rewards and punishments, thus signaling the need for flexible behavior and playing a fundamental role in action control [421]. However, less is known about the reversal of Pavlovian fear. Using functional neuroimaging in conjunction with a fear-conditioning reversal paradigm, Schiller et al. [420] emphasized the role of the vmPFC, showing that the activity in this region increases during an unexpected safe condition (i.e., during the new CS- in reversal), thus providing a possible reward signal. A similar role for the vmPFC was previously stressed in an fMRI study by Kim et al. [422], suggesting that activity in the orbitofrontal cortex increases not only following a reward but also during a successful avoidance of an aversive outcome. However, an opposite result can be found in Morris et al. [423]. In their study, the right orbitofrontal cortex exhibited increased response during CS+ than CS-, both in the acquisition and reversal phases.

Although changes in rhythm power are well documented during fear conditioning and extinction, as summarized above, we are not aware of studies that examine the variations in these rhythms during a fear reversal paradigm, in particular, by comparing the temporal evolution of theta, alpha, and gamma power when the aversive valence progressively shifts from one stimulus to another. Several aspects need to be further elucidated: i) What is the role of different brain rhythms during reversal? ii) How fast can reversal learning occur? iii) Is the new fearful condition specular to the previous one (i.e., the one acquired before reversal), or do the two fearful conditions exhibit differences in brain activity and rhythms?

The scenario is even more complex if one considers the role of brain connectivity. Hudson et al. [424] studied the brain bases of sustained and acute fear using naturalistic fMRI and showed that fear is associated with profound changes in connectivity. Since conditioning implicates changes in synaptic plasticity not only in the hippocampus and amygdala but above all in the cortex and involves the participation of a system of rhythms in various cortical areas [229,249], it may be of great value to analyze how functional connectivity changes during the acquisition phase and the reversal phase of a fear conditioning paradigm. iv) How does functional connectivity modify during fear acquisition in different brain regions and frequency bands? v) Is connectivity after reversal the specular form of the connectivity obtained in the previous acquisition phase, or does connectivity maintain some reminding of the last state?

Although multiple studies on fear acquisition and extinction have appeared in recent years, we think these questions are not entirely clarified yet. In particular, in the following, we will examine the temporal evolution of the power of brain rhythms and brain connectivity over the course of experimental trials, to fully describe the development of fear during acquisition and its shift from one conditioned stimulus to another during reversal. To this end, we reanalyzed data from a group of healthy participants that completed a fear acquisition and reversal task [425]. We previously reported that changes in theta and alpha power discriminate between threat and safety and correlate with skin conductance response. This study extends those results by examining the changes in theta, alpha and gamma power over the course of experimental trials and brain connectivity obtained from cortical source reconstruction in 76 cortical areas, and estimating Granger connectivity.

4.2.2 Methods

4.2.2.1 Participants.

The experiments took place at the Centre for studies and research in Cognitive Neuroscience (CsrCN) of the University of Bologna. Twenty healthy volunteers were recruited. All participants were right-handed and had normal or corrected to normal vision and reported no medical or psychiatric illness. One participant did not complete the experimental session because of a fainting and was excluded from the dataset. Thus, nineteen participants have completed the study (8 males, mean age=23.48, std=1.85). The study followed the American Psychological Association Ethical Principles of Psychologists and Code of Conduct and the Declaration of Helsinki and was approved by the local Bioethics Committee of the University of Bologna (Protocol number 71559). Each participant signed an informed consent prior to the start of the experiment and all data were analyzed and reported anonymously.

4.2.2.2 Pavlovian fear acquisition and reversal task.

The experiment consists of two phases. During the first phase, the acquisition phase, participants view two different visual stimuli (Japanese hiragana) on screen [426,427]. One hiragana, in the following denoted as Image 1, is used as a control stimulus (CS-), while the other hiragana, denoted as Image 2, acts as a conditioned stimulus (CS+). That is, following 50% of visualizations of the CS+ stimulus, an aversive shock (unconditioned stimulus, US) was administered, whereas no shock ever occurred after the CS- stimulus. During the second phase, the reversal phase, the hiragana are reversed, so that Image 2 acts as the new CS- while Image 1 acts as the new CS+. The unconditioned stimulus (US) consists of a 2ms electrostatic stimulation [428–432] administered in the right wrist (dominant hand) through a Digitimer stimulator (model DS7A, Digitimer Ltd.) via Ag/AgCl pregelatinized electrodes (Friendship Medical, SEAg-S-15000/15x20). The intensity of the shock is calibrated for each subject, via verbal feedback and an ascending level procedure, up to a level defined as 'very annoying and unpleasant, but never painful'.

The task is generated thanks to the OpenSesame software [433]. Each phase of the task (both acquisition and reversal) included two blocks (i.e., we have 4 blocks in total: Acq1, Acq2, Rev1 and Rev2), which were interspersed with a five-minute break. For each block, 40 stimuli were shown on screen (i.e., 40 trials, 20 per CS). Each stimulus was presented for 6 seconds, with an interstimulus time interval ranging from 11 to 14 seconds. In each block, stimuli followed each other in random order, with only constraints on the first two trials (one CS- and one CS+/US) and on the number of consecutive stimuli of the same type, never more than two.

Subjects received no information regarding which visual stimulus would be associated with the shock, only the following instructions at the beginning of each block were provided "You will see two different images, which will appear one at a time on the screen. Occasionally, the image may shock you. Your task is to figure out which image will shock you. Press any key to get started." The CS-US relationship is then learned from scratch with experience. During each block, the subjects' electroencephalographic signals and skin conductance were recorded.

4.2.2.3 Skin conductance response (SCR).

Skin conductance was recorded during each block at 1000 Hz (10Hz low-pass filter, gain switch set to 5) through the BIOPAC MP-150 system (Goleta, CA). Pregelatinized electrodes (BIOPAC EL501) were connected on the palmar surface of the left non-dominant, non-shocked hand. The signal was then digitized and down sampled at 200 Hz using Autonomate software (version 2.8, Green et al., 2014) to detect trough-to-peak SCR values. SCR was considered valid if through-to-peak deflection began between 0.5 s and 4.5 s after Image onset, lasted no longer than 5s, and was greater than 0.02 µS. Skin conductance values from trials that did not meet all of these criteria were set as zero SCR and retained in the analyses [390].

Note that, in the present work, SCR is used for the sole purpose of demonstrating the correct acquisition and reversal of fear in our subject group. The relationship between SCR and brain rhythms (i.e., between the central neural system and autonomic responses) was analyzed in the previous work [425].

4.2.2.4 EEG recording and processing.

EEG signals were collected from all participants through 63 wet Ag/AgCl electrodes. Each signal was referenced to the FCz and grounded to the FPz electrode. For all participants, the EEG signals were amplified by a BrainAmp DC amplifier (Brain Products, Gilching, Germany) and digitized at a sampling rate of 1000 Hz. Data corresponding to the four experimental blocks (Acq1, Acq2, Rev1 and Rev2) were then exported to MATLAB R2021a (MathWorks Inc., Natick MA, USA) and processed offline. Firstly, data were down-sampled at 500 Hz and processed with both a band-pass filter (1-60 Hz) and a notch filter (50 Hz) to remove the irrelevant EEG spectral content and electric coupling interferences. Then, 40 stimulus-locked epochs from 0 (stimulus onset) to 6 seconds (stimulus duration) were extracted from EEG recordings of each block, corresponding to 20 Image 1 trials and 20 Image 2 trials. Moreover, the 10 seconds preceding the onset of the first stimulus were extracted from Acq1 and defined as the participant's baseline signal. Once concatenated the extracted signals along the time dimension (baseline, Acq1, Acq2, Rev1 and Rev2), bad channels were identified by computing the correlation coefficient between each electrode and the others. More precisely, for each EEG electrode we calculated the mean value of the 4 highest (absolute) correlations and marked as bad channels those electrodes whose mean value was <0.4 [434,435]. Then, the remaining good channels were re-referenced to the average of all electrodes, and the reference electrode (FCz) recovered.

Subsequently, in order to remove the artefactual components from EEG data, we performed the Independent Component Analysis (ICA) using the EEGLAB Matlab toolbox (https://sccn.ucsd.edu/eeglab/index.php). Independent Components (ICs) containing artifacts were at first identified through an EEGLAB plugin named 'IClabel', which defines the probability of each extracted IC to be a brain-driven ('Brain') or a non-brain-driven activity ('Muscle', 'Eye', 'Heart', 'Line Noise', 'Channel Noise' and 'Other'). After rejecting the ICs classified as 'Brain' with less than 5% probability, we visually inspected all the remaining components (scalp map, time and spectral activity) and further removed only those showing clear artifactual activity. Finally, artifact-cleaned EEG signals were used to retrieve the previously identified bad channels using the spherical interpolation, and the 64 EEG signals were again re-referenced to the average of all electrodes.

4.2.2.5 Cortical sources reconstruction.

Cortical source activity was reconstructed starting from the 64 artifact-cleaned EEG signals. The estimation of intracortical current densities was performed using the method eLORETA (exact Low Resolution Electromagnetic Tomography, LORETA-KEY$^{©®}$ software package), a functional imaging technique belonging to the family of linear inverse solutions for 3D EEG source distribution modeling. Precisely, the algorithm computes the weighted minimum norm solution, so that the particular weights used in this solution endow eLORETA with the property of exact localization of test point sources under ideal (noise free) conditions [127].

The software employs a template three-layers head model (MNI152 template) comprising the scalp, the outer skull surface, and the inner skull surface and registered to the Talairach human brain atlas. The solution space is restricted to the grey matter of the reference brain, divided into 6239 voxels at 5 mm spatial resolution. The software LORETA-KEY$^{©®}$ was employed to compute the inversion matrix starting from the Talairach coordinates of the 64 electrodes, while all subsequent processing steps were implemented in Matlab.

Since each cortical source is described by a three-dimensional current density vector, by right multiplying the inversion matrix by the 64 EEG signals, we can extract the three scalar components of the current density vector for the 6239 voxels and at each time instant. Then, as the choice of constrained dipole orientations was made, the 3D current densities were projected on the voxels' normal versor obtaining one time series for voxel.

Finally, according to the atlas used by LORETA-KEY$^{©®}$ (76 ROIs, see Table 4.2) voxels were grouped in functionally significant Regions of Interest (ROIs), and the signal representing each ROI was obtained as the mean activity of the voxels belonging to the ROI.

Table 4.2 List of the Regions of Interest (ROIs) in which the cerebral cortex has been divided. Please note that each area includes a left, right and in some cases (10 out of 33, indicated by asterisks) a medial portion, thus totally resulting in 76 regions. Abbreviations used in the article are shown in parentheses

Angular Gyrus (AG)	Lingual Gyrus (LG) *	Precuneus (PCU) *
Anterior Cingulate (AC) *	Medial Frontal Gyrus (MeFG) *	Rectal Gyrus (RG)
Cingulate Gyrus (CG) *	Middle Frontal Gyrus (MFG)	Sub-Gyral (SG)
Cuneus (CU) *	Middle Occipital Gyrus (MOG)	Subcallosal Gyrus (SCG) *
Extra-Nuclear (EN)	Middle Temporal Gyrus (MTG)	Superior Frontal Gyrus (SFG) *
Fusiform Gyrus (FG)	Orbital Gyrus (OG)	Superior Occipital Gyrus (SOG)
Inferior Frontal Gyrus (IFG)	Paracentral Lobule (PCL) *	Superior Parietal Lobule (SPL)
Inferior Occipital Gyrus (IOG)	Parahippocampal Gyrus (PHG)	Superior Temporal Gyrus (STG)
Inferior Parietal Lobule (IPL)	Postcentral Gyrus (PCG)	Supramarginal Gyrus (SMG)
Inferior Temporal Gyrus (ITG)	Posterior Cingulate (PC) *	Transverse Temporal Gyrus (STG)
Insula (IN)	Precentral Gyrus (PG)	Uncus (UN)

4.2.2.6 Cortical Power Computation.

The power spectral density (PSD) was evaluated on the 76 reconstructed cortical ROIs and during each trial using the Welch's periodogram method (Hamming window of 2 seconds, 50% overlap, 10 seconds zero padding).

A power analysis in theta (4-8 Hz), alpha (8-14 Hz) and gamma (30-42 Hz) bands was then performed separately in each experimental block (Acq1, Acq2, Rev1 and Rev2) and for each stimulus (Image 1 and Image 2). In particular, for each participant, the following analyses were performed in each ROI. The power was computed in each frequency band and each trial, and normalized to the baseline condition. This trial-by-trial power signal was then separated between Image 1 and Image 2 (20 trials per image and per block) and used for two computations. 1) For each frequency band and each block, a mean power was computed by averaging the trial-by-trial power over the 20 trials of each image, to obtain one power value for block and stimulus. 2) For each frequency band and each block, the trial-by-trial power was used to compute a moving average signal for each image, considering a window of 3 trials (sliding one trial at a time), in order to evaluate power temporal evolution for each stimulus over the four experimental blocks.

Importantly, the data resulting from mean power computation (computation 1) were used to identify, without any a priori assumption, the ROIs and rhythms implicated in fear conditioning, and thus deserving an in-depth inspection in this work. To this end, for each frequency band we followed two steps. Step 1: we selected all areas that exhibit a statistical significant difference (corrected) in power between Image 1 and Image 2 (i.e., between CS+ and CS-) in at least one of the four experimental blocks, to detect a possible involvement in either acquisition or reversal. We anticipate here that all these corrected significances occur only in the acquisition phase. Step 2: since the focus in this work is on reversal, we further restricted our analysis to those ROIs (among those selected in step 1) that exhibit a statistically significant difference (although not corrected) between Image 1 and Image 2 both during acquisition and reversal. As shown in section Results, this resulted in two regions for the theta rhythm, eleven regions for the alpha rhythm, and no region for the gamma rhythm.

4.2.2.7 Functional Connectivity through frequency-domain Granger Causality.

To further investigate the possible neural mechanisms underlying fear acquisition and reversal, we evaluated the functional connectivity in each considered frequency band among the 76 reconstructed cortical ROIs by using the Spectral Granger Causality estimator, which provides weighted and directional metrics of the causal interactions between ROIs. The connectivity analysis was limited to the theta and alpha bands, since the power analysis in the gamma band provided inconclusive results (see section Results). The Granger Causality is based on the autoregressive (AR) modeling framework and estimates the functional connectivity between ROIs by comparing the prediction ability of two AR models (of a certain order p) on the same process $x_{k,j}$. More precisely, let's consider two time series $x_{k,i}[n]$ and $x_{k,j}[n]$, where *n* is the discrete time ($n = 0, 1, \dots, N-1$), representing the activity at two distinct cortical ROIs (ROI_i and ROI_j) for each participant k ($k = 1, \dots, 19$). The Granger Causality estimator quantifies the causal interaction from ROI_i to ROI_j as the improvement in predictability of $x_{k,j}[n]$ when using a bivariate AR model, based on both past values of $x_{k,j}$ and past values of $x_{k,i}$, compared to a univariate AR model, based only on past values of $x_{k,j}$.

Frequency-domain Granger causality can be formalized starting from the spectral derivation of the bivariate representation of the activity of the two ROIs, $x_{k,j}[n]$ and $x_{k,i}[n]$ via the Fourier Transformation. According to Geweke [436,437] the power spectrum of a time series $x_{k,j}[n]$ can be decomposed into an 'intrinsic' and a 'causal' part, considering the latter predicted by the other time series $x_{k,i}[n]$.

The GC spectrum from i to j ($GC_{i\rightarrow j}(f)$) is defined as the logarithm of the ratio between the total power spectrum of $x_{k,j}[n]$ at frequency f and the difference between the total power spectrum and the 'causal' power predicted by $x_{k,i}[n]$ at the same frequency. Accordingly, at a given frequency *f,* the estimated quantity $GC_{i\rightarrow j}(f)$ is zero when the causal power of $x_{k,i}[n]$ onto $x_{k,j}[n]$ is zero and increases (>0) as the causal power increases. For each participant k, a GC spectrum was computed in each block and for each stimulus by linking 20 trials for each experimental block and stimulus. In all cases, the order p of the AR models was set equal to 30 on the basis of a previous analysis [438,439] which showed that for $p \geq 30$ the estimated values of GC do not change substantially. It is worth noting that spectral GC provides a connectivity matrix ($GC(f)$: 76x75, discarding auto connectivity) for each frequency sample ($n\ sample\ =\ 2501, frequency\ resolution = 0.1\ Hz$). To obtain a single connectivity value representative of the rhythms under analysis, we computed the mean value of GC(f) in the given frequency band. Additionally, for each participant, the theta and alpha connectivity matrices were normalized so that the sum of all connections in each matrix is equal to 100. Then, from these normalized matrices (named *complete* connectivity matrices), *sparse* normalized connectivity matrices were obtained by performing a statistical analysis between the two types of stimuli (Image 1 and Image 2) independently for each experimental block, using the non-parametric permutation t-test (see details below in the section *Statistical analysis*). Hence, for each block and for each frequency band, only connections significantly different (p-value<0.05) between Image 1 and Image 2 among all the 76x75 possible connections were retained in each subject connectivity matrix, while the others were set to zero.

Finally, both the complete and the resulting sparse connectivity matrices were averaged across participants to characterize each block and each stimulus (4 blocks x 2 stimuli) with a connectivity matrix.

We focused our analysis on the set of connections exiting from the most indicative cortical areas, selected following the power analysis, in order to understand how power changes are transmitted from a generative node to others nodes in the network. The resulting functional network is represented through a graph where the involved cortical ROIs are the nodes and the connectivity values are displayed by weighted and directed arrows.

4.2.2.8 Statistical analysis.

For each experimental block, a two-tailed permutation-based t-test for dependent samples between Image 1 and Image 2 was performed on SCR first, and then on the normalized power and connectivity data in each considered frequency band. It should be noted that the two-tailed test was used as we had no a priori hypothesis about how the power and connectivity were varying (increase/decrease) between the two stimuli. The distribution of the t-statistic for each cortical ROI under the null hypothesis was empirically realized by generating 5000 random permutations of the observed values between the 2 stimulus conditions (Monte Carlo method). The uncorrected p-value was the proportion of the permutation distribution greater than or at most equal to the observed t-statistic computed on the non-permuted values. Then, for cortical mean power analysis (i.e. average power over the 20 trials) a correction for multiple comparisons (76 comparisons, one per ROI) was achieved, separately for each block, using the false discovery rate correction (Benjamini Hochberg procedure) [440]. Also, the SCR statistics largely survive correction (4 comparisons, one per block).

Conversely, uncorrected p-values were considered for connectivity and moving average analysis. In the first case, this is justified due to the high number of variables involved (i.e., 76x75 connections), which makes the correction requirement extremely demanding; in the second case due to the strong relationship (temporal and behavioral) that exists between one trial and the subsequent (subjected to a moving average of 3 trials).

4.2.3 Results

4.2.3.1 Skin Conductance Response analysis

First, a statistical analysis was performed between the SCR during the presentation of Image 1 and Image 2 in each block, to assess the correct acquisition and reversal of fear.

For each block (Acq1, Acq2, Rev1 and Rev2), we found a statistically significant difference between images in SCR values. In detail: p_Acq1= $3.9\ 10^{-4}$, p_Acq2 = $3.9\ 10^{-4}$, p_Rev1= $6.0\ 10^{-3}$, and p_Rev2= $3.9\ 10^{-4}$ (uncorrected). Moreover, in all blocks SCR was higher during CS+ than CS- (i.e., during the presentation of Image 2 in Acq1 and Acq2, and Image 1 in Rev1 and Rev2). This analysis confirms the successful acquisition and reversal of fear for the subject group.

4.2.3.2 Cortical sources power analysis.

Data resulting from the mean power analysis (computation 1, see *Cortical Power Computation* section) were used to identify, without any a priori assumptions, the ROIs and rhythms most involved in fear acquisition and reversal. For these regions, we displayed the power mean value across the four blocks and the moving average of the trials. Finally, we further performed functional connectivity analysis.

Theta power – To get a global view, Fig. 4.11 shows the Student's ‘t’ values resulting from the 76 ROI-wise statistical comparison between the two images, carried out on theta mean power. All the four blocks and all the 76 cortical ROIs are depicted. This Figure shows that strong theta differences are especially evident in a portion of the cortex close to the left cingulate gyrus. Furthermore, the power difference is greater for Image 2 (i.e., CS+) than Image 1 (i.e. CS-) during the acquisition phase, and is inverted (i.e., is greater for Image 1, the new CS+) passing from the acquisition to the reversal phase: i.e., theta power is greater during CS+ than CS- both during acquisition and reversal. It is important to note that, as seen in Fig. 4.11, several regions show high absolute ‘t’ values (i.e., both positive and negative), but without reaching the statistical level that survives correction. For this reason, these regions are not considered for the subsequent analyses, but may be of interest for future investigation.

To better summarize the results, Table 4.3 lists all regions which exhibit a statistically significant difference (corrected) in theta power between Image 1 and Image 2 in at least one block of the experiment (step 1 in Method section). These are the left Cingulate Gyrus (CG l), the left Superior Frontal Gyrus (SFG l), the medial Cingulate Gyrus (CG m), and the medial Superior Frontal Gyrus (SFG m). All these corrected statistical differences are evident during the acquisition phase but not during the reversal phase. Moreover, according to the second criterion delineated in the Method section (step 2), two of the previous regions [i.e., the left cingulate gyrus (CG l) and the medial cingulate gyrus (CG m)] exhibit a significant (but uncorrected) statistical difference in the first reversal block, revealing that these regions are implicated not only in acquisition but likely also in reversal. In the following, we will focus the attention on these two regions to better characterize the theta band.

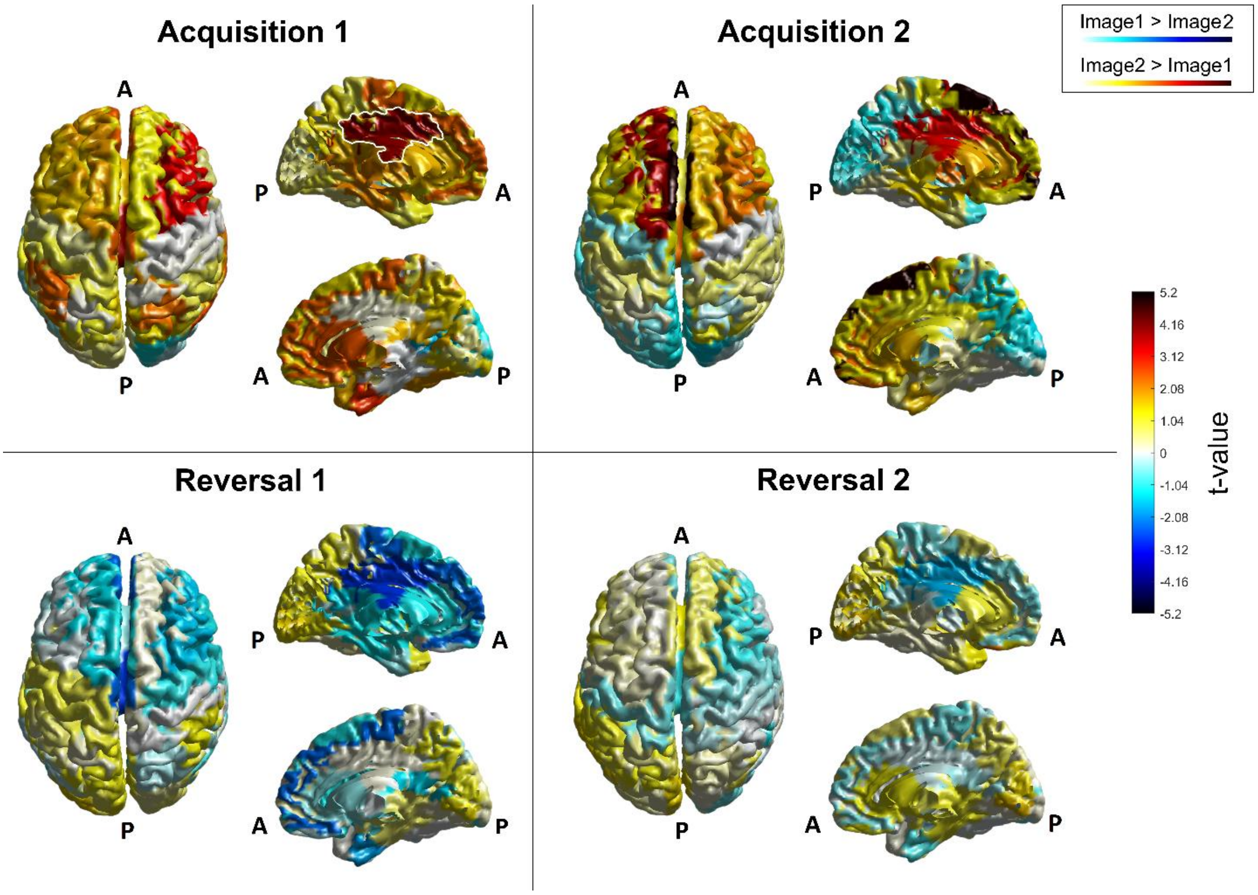


Fig. 4.11 - Student's 't' values, resulting from the statistical comparison between the two images, carried out on theta mean power, in the four blocks and over all the 76 cortical ROIs. In each panel, the left column represents the top view of the cerebral cortex while the right column represents the medial left (top) and the medial right (bottom) view of the cerebral cortex. White outline in the 'Acquisition 1' panel is used to highlight areas that have been selected for further analysis. That is, for the theta rhythm, the left and medial cingulate gyrus (CG l and CG m, visible in the left medial view). Letter 'A' stands for 'Anterior', letter 'P' for 'Posterior'. The color bar corresponds to uncorrected t-values. Positive values (colors tending toward red) indicate higher power during the visualization of Image 2 (CS+ in acquisition, CS- in reversal), while negative values (colors tending toward blue) indicate higher power during the visualization of Image 1 (CS- in acquisition, CS+ in reversal)

Fig. 4.12 shows the block-by-block normalized theta band power of CG l and CG m. It is noticeable that theta power is greater for the CS+ stimulus both during acquisition and reversal (Image 2 in Acq1 and Acq2, Image 1 in Rev1 and Rev2). In agreement with Fig. 4.11, this difference is mostly evident in acquisition and is reduced during reversal. Especially in the second reversal block, the power difference is statistically insignificant even when uncorrected.

Table 4.3 Names and block-by-block corrected p-values of the ROIs that show significant corrected statistical differences in normalized theta power between CS+ and CS- in at least one block. The uncorrected p-values are shown between brackets. NS signifies that no statistically significant difference was observed.

ROIs:	Acquisition 1	Acquisition 2	Reversal 1	Reversal 2
Left Cingulate Gyrus (CG l)	p=0.046 (0.0012)	p=0.010 (3.9*10^$^{-4}$)	NS (p=0.029)	NS (NS)
Left Superior Frontal Gyrus (SFG l)	NS (NS)	p=0.010 (3.9*10^$^{-4}$)	NS (NS)	NS (NS)
Medial Cingulate Gyrus (CG m)	p=0.030 (3.9*10^$^{-4}$)	NS (p=0.021)	NS (p=0.0068)	NS (NS)
Medial Superior Frontal Gyrus (SFG m)	NS (NS)	p=0.010 (3.9*10^$^{-4}$)	NS (NS)	NS (NS)

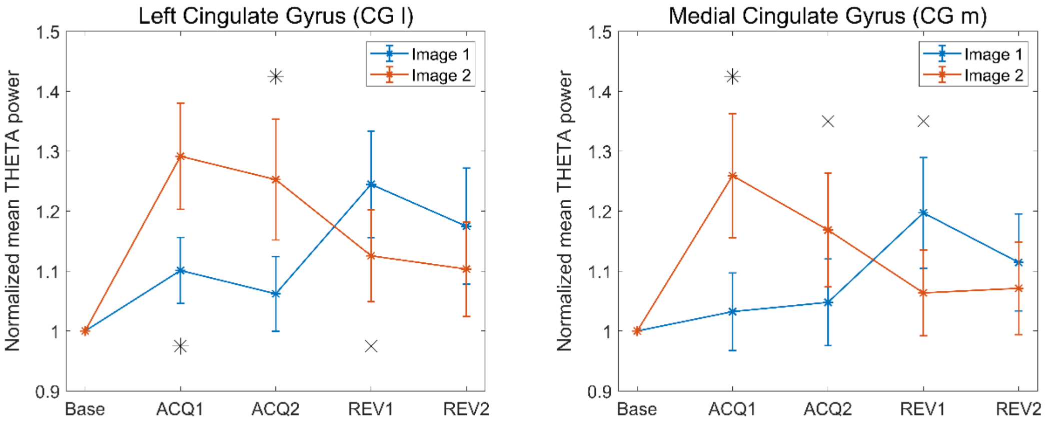
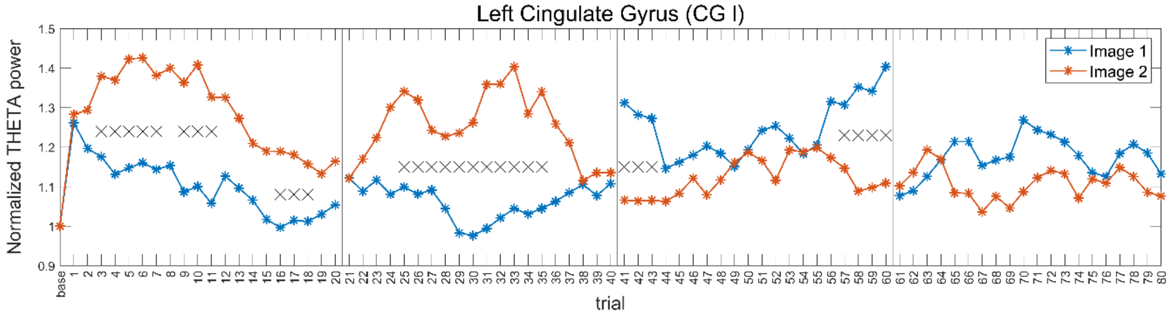


Fig. 4.12 - Normalized mean power in the theta band, for all four blocks and both images, in the left cingulate gyrus (CG l, left column) and in the medial cingulate gyrus (CG m, right column). The power for Image 1 is depicted in blue and the power for Image 2 in red, accompanied in each block by the respective SEM bar. Asterisks indicate presence of corrected statistical significance (p<0.05, false discovery rate corrected) in that particular block while crosses denote the presence of a statistical significance (p<0.05, uncorrected) which does not survive correction for multiple comparisons. It is well evident the power inversion in passing from acquisition to reversal, i.e., theta power is always greater during CS+ (Image 2 in Acq1 and Acq2; Image 1 in Rev1 and Rev2) than CS-

Fig. 4.13 shows the trial by trial moving average (window length = 3 trials) of theta power in the two selected regions, CG l and CG m. The Figure confirms that theta-band power is greater for the CS+ than for the CS- in almost all trials (Image 2 in Acq1 and Acq2, Image 1 in Rev1 and Rev2). The inversion in reversal 1 occurs very quickly, being already evident after the first trial of the moving average. The greatest power difference between the two stimuli is observed in the two acquisition blocks, while in reversal this difference progressively decreases. Indeed, in reversal 2, the power difference between the two images is less marked, as confirmed by the statistical analysis, which does not show any statistically significant difference in any trial of this block. This appears as a result of the drastic fall in theta power during Image 1 (new CS+) at the beginning of the reversal 2.



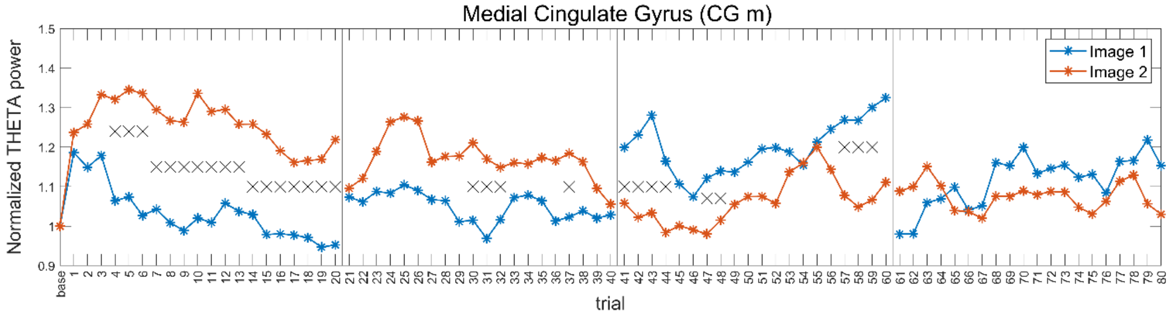


Fig. 4.13 - Moving averages (w=3 trials) of normalized theta power, trial by trial, for the four blocks. The top row shows the normalized power of the left cingulate gyrus (CG l), the bottom row the normalized power of the medial cingulate gyrus (CG m). The two images are shown in the same color (blue for Image 1, red for Image 2). Crosses indicate the presence of statistical significance (p<0.05, uncorrected) between the power of the two images for that particular trial. Vertical lines are used to delineate the four different blocks. It is evident that theta power is always greater during CS+ (Image 2 in Acq1 and Acq2; Image 1 in Rev1 and Rev2), and that inversion occurs already after the first reversal trial of the moving average.

Alpha power – Fig. 4.14 shows the Student's ‘t’ values resulting from the 76 ROI-wise statistical comparison between the two images, carried out on alpha mean power. All the four blocks and all the 76 cortical ROIs are depicted. This Figure shows that significant alpha differences are evident in a large portion of the parietal cortex in the left hemisphere (i.e., contralateral to the shocked hand) and these changes are also evident (although less marked) during reversal. Alpha power is smaller for Image 2 (i.e., CS+) during the acquisition phase, and becomes smaller for Image 1 (i.e. the new CS+) during reversal. Also in this case, other areas exhibit stronger statistical differences which however do not survive correction.

Table 4.4 shows all regions which exhibit a significant statistical difference (corrected) in alpha power between Image 1 and Image 2 in at least one block of the experiment (step 1 in Method section). These differences occur during acquisition only. Moreover, according to the second criterion (step 2), among these, eleven areas show at least one statistical significance (uncorrected) in the reversal phase of the experiment.

For the sake of brevity, in the following figures, we will focus attention on six regions [right cingulate gyrus (CG l), left cingulate gyrus (CG r), left inferior parietal lobule (IPL l), left precentral gyrus (PG l), left and medial paracentral lobules (PCL l and PCL m)], which show a clear power difference between images, a clear inversion during reversal, and are limbic, motor or somatosensory areas already reported in the literature as belonging to the so-called "Fear Network" [397,424,441,442].

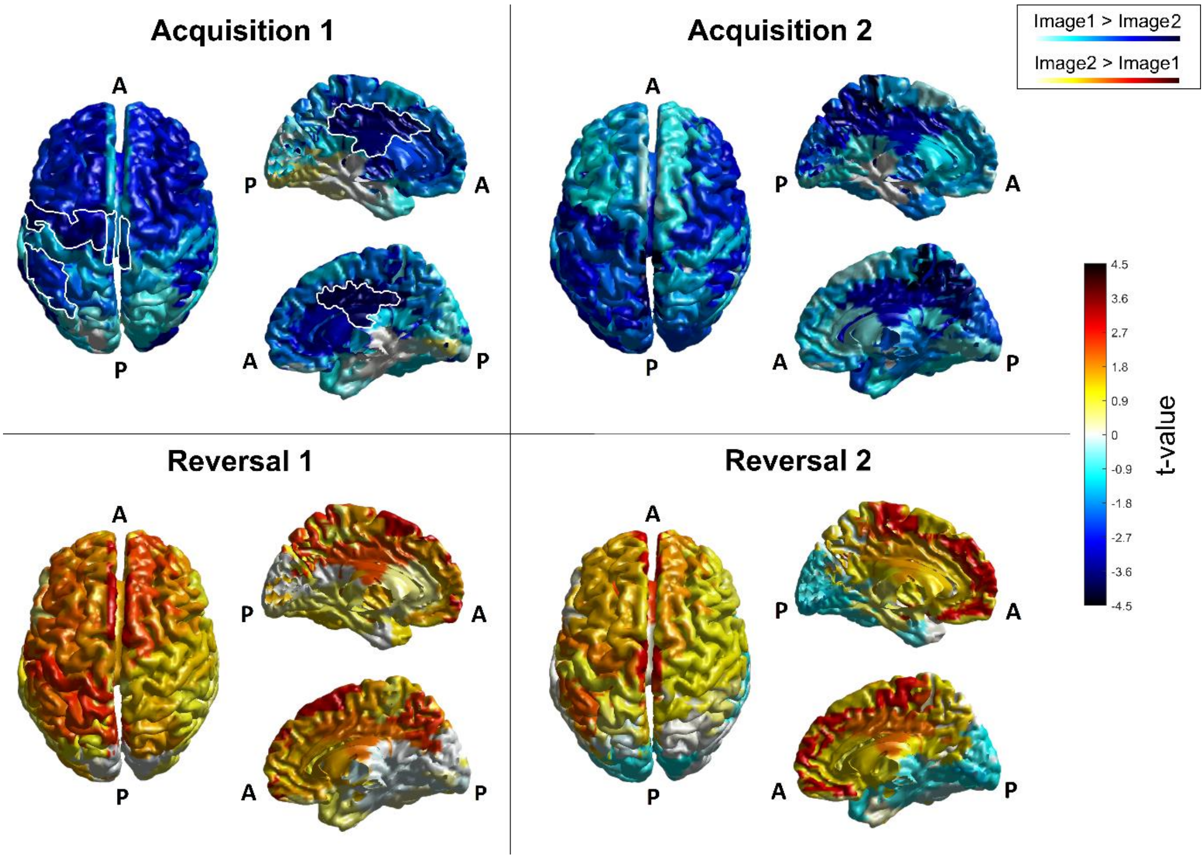


Fig. 4.14 - Student's ‘t’ values, resulting from the statistical comparison between the two images, carried out on alpha mean power, in the four blocks and over all the 76 cortical ROIs. In each panel, the left column represents the top view of the cerebral cortex while the right column represents the medial left (top) and the medial right (bottom) view of the cerebral cortex. White outlines in the 'Acquisition 1' panel are used to highlight areas that have been selected for further analysis. That is, for the alpha rhythm, the left inferior parietal lobule, the left and medial paracentral lobule, the left precentral gyrus (IPL l, PCL l, PCL m and PG l, visible in the top view) and the left and right cingulate gyrus (CG l and CG r, visible in the left and the right medial view). Letter ‘A’ stands for ‘Anterior’, letter ‘P’ for ‘Posterior’. The color bar corresponds to uncorrected t-values. Negative values (colors tending toward blue) indicate higher power during the visualization of Image 1 (CS- in acquisition, CS+ in reversal), while positive values (colors tending toward red) indicate higher power during the visualization of Image 2 (CS+ in acquisition, CS- in reversal).

Table 4.4 Names and block-by-block corrected p-values of the ROIs that show significant differences between normalized alpha power between CS+ and CS- in at least one block. The uncorrected p-values are shown between brackets. NS signifies that no statistically significant difference was observed.

ROIs:	Acquisition 1	Acquisition 2	Reversal 1	Reversal 2
Right Anterior Cingulate (AC r)	p=0.30 (0.060)	NS (NS)	NS (NS)	NS (NS)
Right Cingulate Gyrus (CG r)	p=0.023 (0.0012)	p=0.037 (0.0072)	NS (NS)	(p=0.031)
Right Extra-Nuclear (EN r)	p=0.046 (0.012)	NS (NS)	NS (NS)	NS (NS)

Right Inferior Frontal Gyrus (IFG r)	(p=0.042)	p=0.035 (0.0040)	NS (NS)	NS (NS)
Right Inferior Parietal Lobule (IPL r)	p=0.043 (0.010)	p=0.030 (0.0024)	NS (NS)	NS (NS)
Right Middle Frontal Gyrus (MFG r)	p=0.024 (0.0020)	(p=0.029)	NS (NS)	NS (NS)
Right Middle Occipital Gyrus (MOGr)	p=0.024 (0.0032)	(p=0.035)	NS (NS)	NS (NS)
Right Posterior Cingulate (PC r)	p=0.043 (0.011)	p=0.035 (0.0060)	NS (NS)	NS (NS)
Right Superior Frontal Gyrus (SFG r)	p=0.043 (0.0010)	NS (NS)	(p=0.021)	NS (NS)
Right Transverse Temporal Gyrus (TTG r)	p=0.030 (0.0064)	(p=0.046)	NS (NS)	NS (NS)
Left Angular Gyrus (AG l)	NS (NS)	p=0.044 (0.0092)	NS (NS)	NS (NS)
Left Cingulate Gyrus (CG l)	p=0.024 (0.0020)	p=0.035 (0.0052)	(p=0.026)	NS (NS)
Left Extra-Nuclear (EN l)	p=0.023 (0.0012)	(p=0.014)	(p=0.016)	NS (NS)
Left Inferior Parietal Lobule (IPL l)	(p=0.020)	p=0.035 (0.0060)	(p=0.022)	(p=0.022)
Left Insula (IN l)	p=0.024 (0.024)	NS (NS)	NS (NS)	NS (NS)
Left Medial Frontal Gyrus (MeFG l)	p=0.023 (8.0*10^$^{-4}$)	NS (NS)	NS (NS)	(p=0.0028)
Left Middle Occipital Gyrus (MOG l)	NS (NS)	p=0.010 (3.9*10^$^{-4}$)	NS (NS)	NS (NS)
Left Middle Temporal Gyrus (MTG l)	p=0.024 (0.0032)	p=0.035 (0.0060)	NS (NS)	NS (NS)
Left Paracentral Lobule (PCL l)	p=0.024 (0.0032)	p=0.024 (0.0016)	NS (NS)	(p=0.047)
Left Precentral Gyrus (PG l)	p=0.023 (7.9*10^$^{-4}$)	p=0.010 (3.9*10^$^{-4}$)	(p=0.038)	(p=0.048)
Left Precuneus (PCU l)	NS (NS)	p=0.035 (0.0064)	NS (NS)	NS (NS)
Left Superior Frontal Gyrus (SFG l)	p=0.030 (0.0060)	NS (NS)	NS (NS)	NS (NS)
Left Supramarginal Gyrus (SMG l)	p=0.028 (0.0044)	(p=0.015)	NS (NS)	NS (NS)
Left Transverse Temporal Gyrus (TTG l)	p=0.028 (0.0040)	p=0.035 (0.0044)	NS (NS)	NS (NS)
Medial Anterior Cingulate (AC m)	p=0.046 (0.013)	(p=0.036)	NS (NS)	(p=0.023)
Medial Cingulate Gyrus (CG m)	NS (NS)	p=0.030 (0.0028)	NS (NS)	NS (NS)
Medial Paracentral Lobule (PCL m)	p=0.030 (0.0060)	p=0.023 (0.0012)	NS (NS)	(p=0.022)
Medial Precuneus (PCU m)	NS (NS)	p=0.010 (3.9*10^$^{-4}$)	(p=0.024)	NS (NS)

Normalized alpha-power differences in the remaining five regions [right superior frontal gyrus (SFG r), left extra-nuclear (EN l), left medial frontal gyrus (MeFG l), medial anterior cingulate (AC m) and medial precuneus (PCU m)], which nonetheless show similar trends, are reported for completeness in the Supplementary Material Fig. 4.11S.

Fig. 4.15 shows the block-by-block patterns of alpha power for the six selected regions. It is evident that the alpha power is higher in the CS- (i.e., during the presentation of Image 1 in Acq1 and Acq2, Image 2 in Rev1 and Rev2) than in CS+. This difference is especially evident and statistically significant during the acquisition phase but is also present during the reversal phase (although with reduced statistical significance due to the greater inter-subject variability). In addition, the alpha power in the CG r and the CG l shows the tendency to increase for both stimuli, block by block.

Fig. 4.16 shows the alpha power moving average for two regions, CG l and PG l. Among all, these two regions have been selected since they exhibit the greatest number of significant (uncorrected) trials in the moving average signal. In fact, in almost all trials, alpha power is higher in the CS- (i.e., during the presentation of Image 1 in Acq1 and Acq2, Image 2 in Rev1 and Rev2) than in the CS+, with a clear inversion occurring in the reversal phase. This difference emerges after 3-4 steps of the moving average signal (as evident during Acq1 and Rev1). However, in Rev2, the power difference between the two images is reduced compared with the previous blocks. Indeed, fewer trials exhibit statistically significant differences in this last block. Finally, an abrupt fall in alpha-power is always evident from one block to the next, probably reflecting a more stressful or attentive condition at the beginning of each new block. A similar trend, not shown for brevity and reported in the Supplementary Material (Fig 4.12S-2.14S), is evident also for the other nine selected regions.

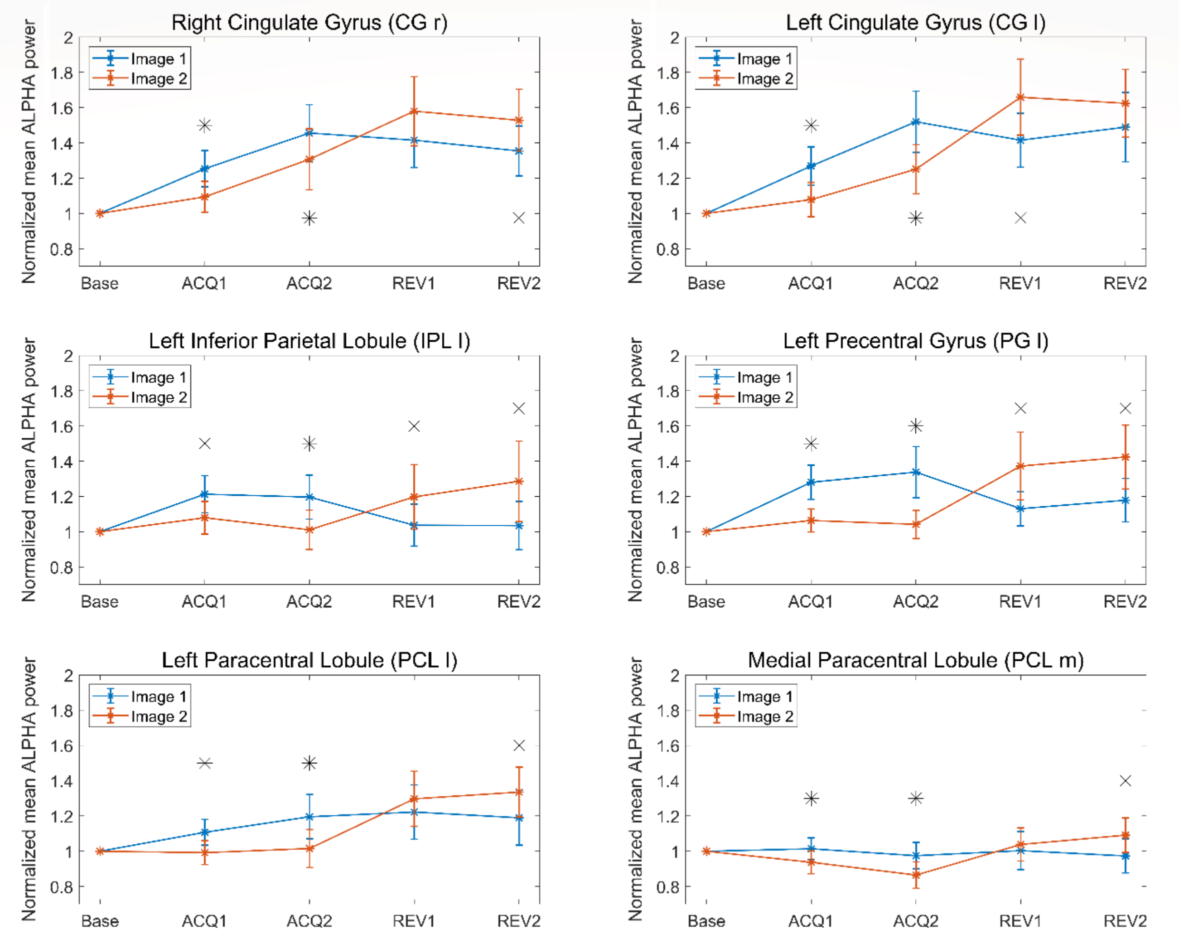


Fig. 4.15 - Normalized mean power in the alpha band, for all four blocks and both images, in the right and left cingulate gyrus (CG r and CG l, top row), left inferior parietal lobule (IPL l, middle row, left column), left precentral gyrus (PG l, middle row, right column), left and medial paracentral lobule (PCL l and PCL m, bottom row). Results for Image 1 are depicted in blue and those for Image 2 in red, accompanied in each block by the respective SEM bar. Asterisks indicate the presence of corrected statistical significance (p<0.05, false discovery rate correction) in the specific block, while crosses denote the presence of a statistical significance (p<0.05, uncorrected) which does not survive correction for multiple comparisons. It is well evident the power inversion in passing from acquisition to reversal, i.e., alpha power is always greater during CS- (Image 1 in Acq1 and Acq2; Image 2 in Rev1 and Rev2) than CS+.

Gamma power – Although some significant differences were found in gamma power between Image 1 and Image 2 in some regions and some blocks (more specifically, in Acq1: left anterior cingulate, left cuneus, left orbital gyrus, left subcallosal gyrus; in Acq2: right paracentral lobule, right precentral gyrus, left angular gyrus, left medial frontal gyrus, left superior occipital gyrus, medial precuneus; in Rev1: right inferior occipital gyrus, left cingulate gyrus, medial cingulate gyrus; in Rev2: medial cuneus), these differences never survived the statistical correction (step 1 in Method section). For this reason, we did not further analyze gamma power changes or perform connectivity analysis in the gamma band.

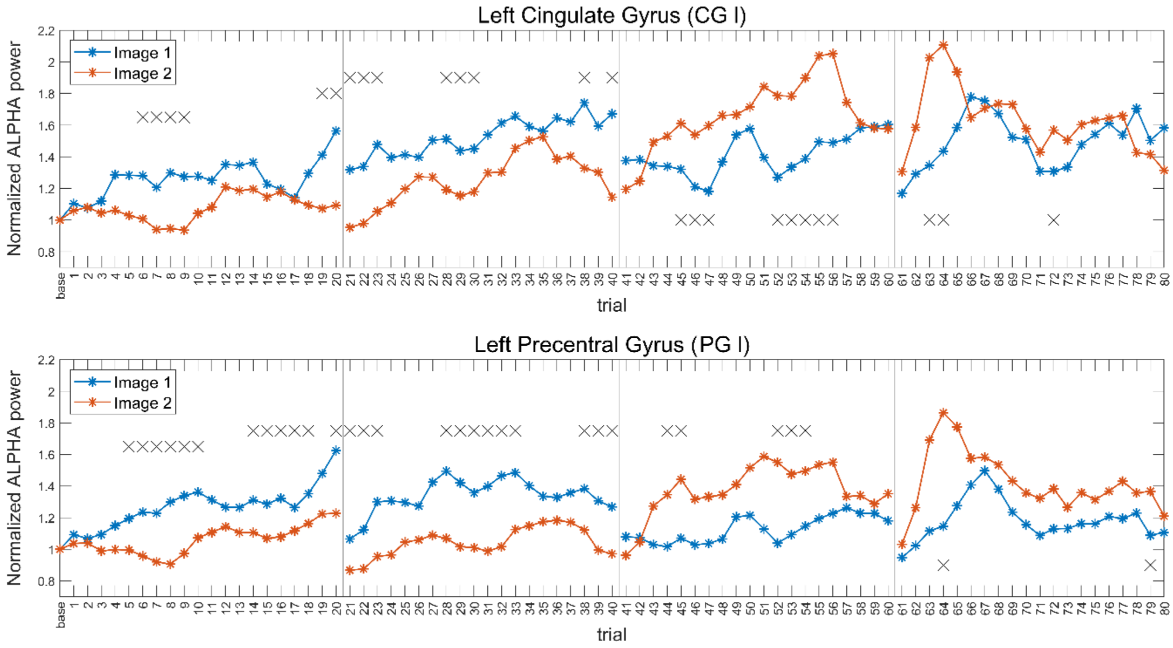


Fig. 4.16 - Moving averages (w=3 trials) of normalized alpha power, trial by trial, for the four blocks. The top row shows the power of the left cingulate gyrus (CG l), the bottom row the power of the left precentral gyrus (PG l). The two images are shown in the same color (blue for Image 1, red for Image 2). The crosses indicates the presence of statistical significance (p<0.05, uncorrected) between the power of the two images for that particular trial. Vertical lines are used to delineate the four different blocks. It is evident that alpha power is always greater during CS- (Image 1 in Acq1 and Acq2; Image 2 in Rev1 and Rev2), and that inversion occurs within three-four reversal trials of the moving average.

4.2.3.3 Functional connectivity analysis.

The previous analysis revealed the presence of several regions that exhibit significant differences between Image 1 and Image 2 power. In this section we investigate, through Granger connectivity, how information is transferred from these regions toward other regions in the brain (i.e., how this increased or decreased power is transferred). Connectivity in the theta band is illustrated considering the connections that emerge from the two regions, CG l and CG m, since both display significant connectivity differences between Image 1 and Image 2. Regarding the alpha rhythm, only connectivity from two areas (CG l and PG l) is shown. In fact, connectivity was evaluated also from the remaining areas (CG r, IPL l, PCL l, and PCL m) but statistical differences between Image 1 and Image 2 were unclear and did not provide any evident network topology. In each of the following Figures (4.17-4.20), the first column displays the connections that are stronger during the processing of Image 1 (Image1>Image2: blue arrows), whereas the second column shows the connections that are stronger during the processing of Image 2 (Image2>Image1: red arrows). In the first set of plots, we show the connections which exhibit the largest differences (in absolute value) between Image 1 and

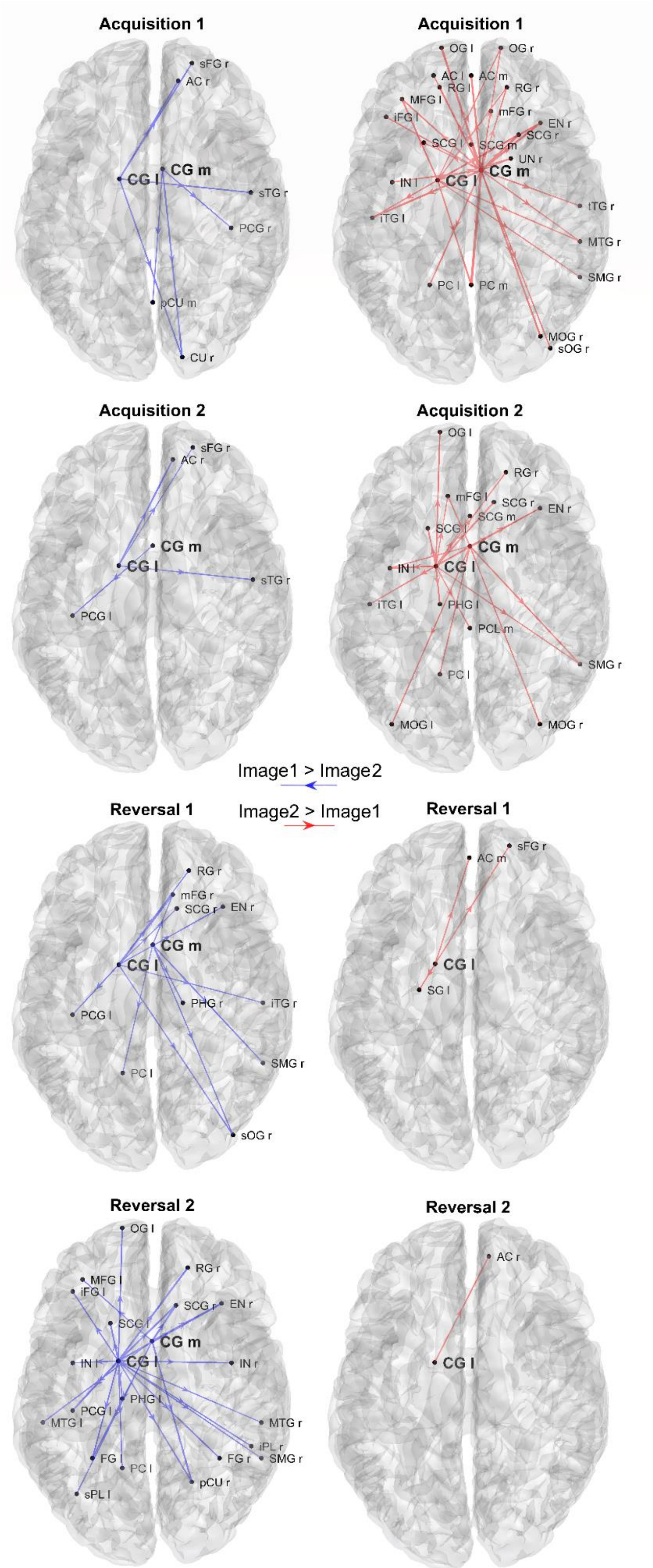
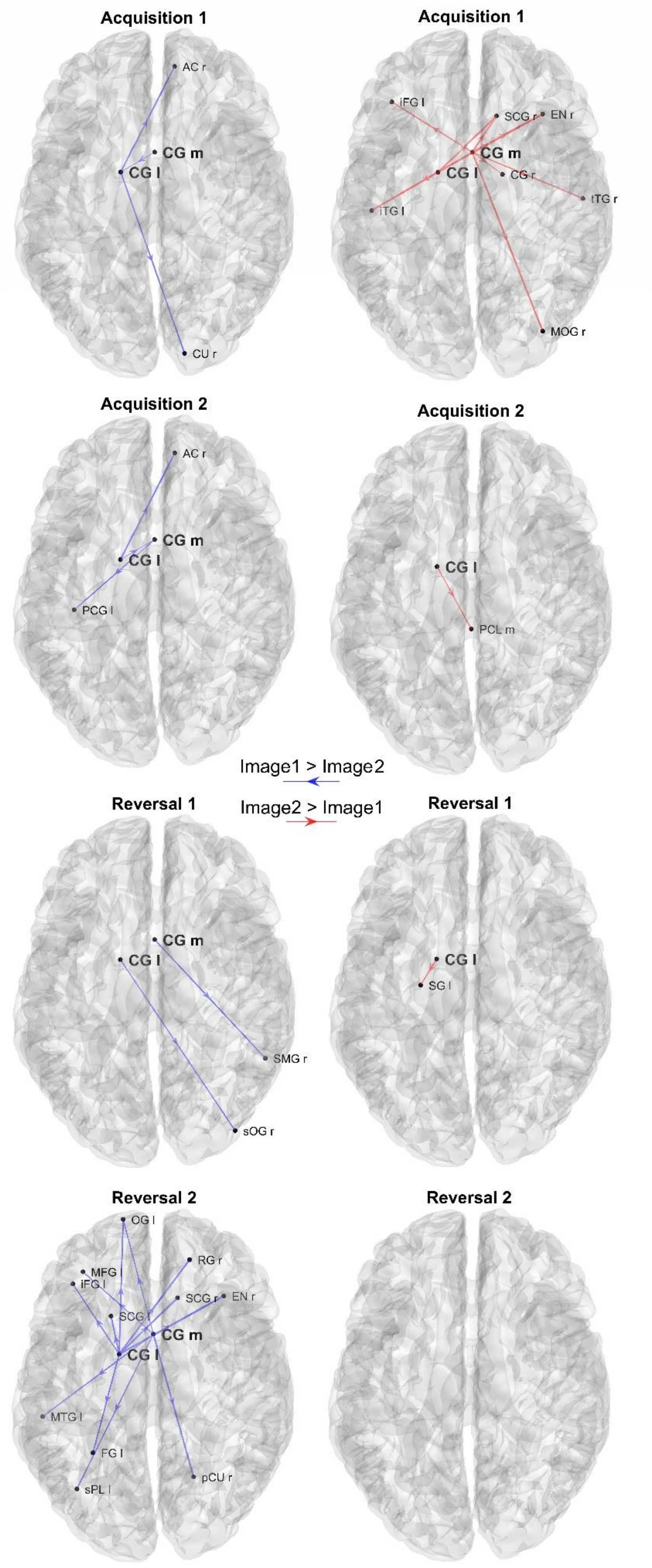


Image 2 in each block, independently of the statistical difference (Figures 4.17, 4.19), to show a general trend.

Since all connectivity matrices are normalized to 100, and we have a total of 76x75 connections, the mean value of the connections is 0.0175. We chose to plot all connection differences that overcome 1/4 of the mean value for theta and 1/6 for alpha band. These different thresholds were chosen because the differences turned out to be higher in the theta than in the alpha range. In the second set of plots, only connections that exhibit a significant statistical difference between Image 1 and Image 2 ($p<0.05$, uncorrected) are displayed (Figures 4.18, 4.20).

Fig. 4.17 - Plots of the strongest differences in connectivity between Image 1 and Image 2, calculated in theta band, block by block. Only connection differences with absolute value greater than 1/4 of the mean are displayed, exiting from the two selected regions CG m and CG l. The left column shows the connections that are greater during the processing of the Image 1 (CS- in acquisition, CS+ in reversal; blue directional arrows), whereas the right column shows connections that are greater during processing of Image 2 (CS+ in acquisition, CS- in reversal; red directional arrows). In all blocks the connectivity is stronger during CS+ [i.e., during the presentation of Image 2 in Acq1 and Acq2 (right column), and Image 1 in Rev1 and Rev2 (left column)], with a clear inversion occurring from acquisition to reversal.



Theta connectivity – Fig. 4.17 shows the strongest differences in absolute value (greater than 0.0043 = 1/4 of the mean) concerning the connections that exit from the two areas CG l and CG m. In both ROIs and all blocks, the connectivity is stronger during CS+ (i.e., during the presentation of Image 2 in Acq1 and Acq2, Image 1 in Rev1 and Rev2), thus mimicking changes in power.

An inversion in outgoing connectivity is evident from acquisition to reversal in both regions. Moreover, this difference is more pronounced in the second reversal block than in the first (at odd with the pattern of theta power, whose difference between Images was more evident in reversal 1 than in reversal 2). In Fig. 4.18, theta-band connectivity is displayed showing only those connections from the two selected ROIs which are significantly different between Image 1 and Image 2 (sparse connectivity matrices).

Fig. 4.18 - Plots of the significant differences in connectivity between Image 1 and Image 2, calculated in the theta band, block by block. Only the connections exiting from the two selected regions CG m and CG l and which are significantly different between Image 1 and Image 2 in each block ($p<0.05$, uncorrected) are displayed (sparse matrices). The left column shows the connections that are greater during processing of the Image 1 (CS- in acquisition, CS+ in reversal; blue directional arrows), whereas the right

column shows connections that are greater during processing of Image 2 CS+ in acquisition, CS- in reversal; red directional arrows). It is worth noting that differences are evident during acquisition 1 and reversal 2 only, with a clear inversion of the connectivity and a strong impact especially in reversal 2.

As in the non-sparse representation (Fig. 4.17), a higher number of connections from CG l and CG m can be observed during CS+ (i.e., during the presentation of Image 2 in Acq1 and Acq2, Image 1 in Rev1 and Rev2) than CS-, but this difference is evident only in the first acquisition block and in the second reversal block. Finally, the effect in reversal 2 (Image 1 (new CS+) > Image 2 (new CS-), bottom row) is even more pronounced than in acquisition 1.

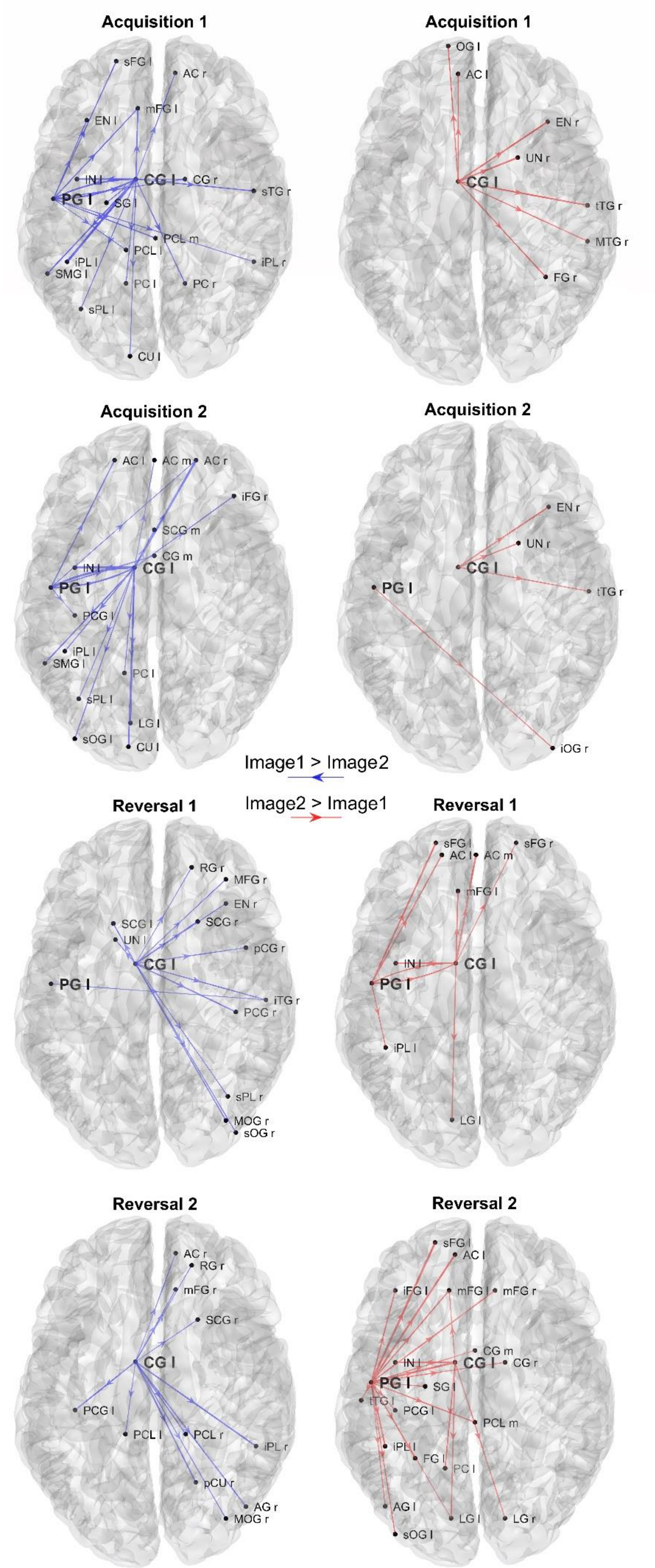


Fig. 4.19 - Plots of the strongest differences in connectivity between Image 1 and Image 2, calculated in alpha band, block by block. Only the connection differences with absolute value greater than 1/6 of the mean are displayed, coming from the two selected regions PG l and CG l. The left column shows the connections that are greater during processing of the Image 1 (CS- in acquisition, CS+ in reversal; blue directional arrows), whereas the right column shows connections that are greater during processing of Image 2 (CS+ in acquisition, CS- in reversal; red directional arrows). In all blocks the connectivity is stronger during CS- [i.e., during the presentation of Image 1 in

Acq1 and Acq2 (left column), and Image 2 in Rev1 and Rev2 (right column)], with a clear inversion occurring from acquisition to reversal

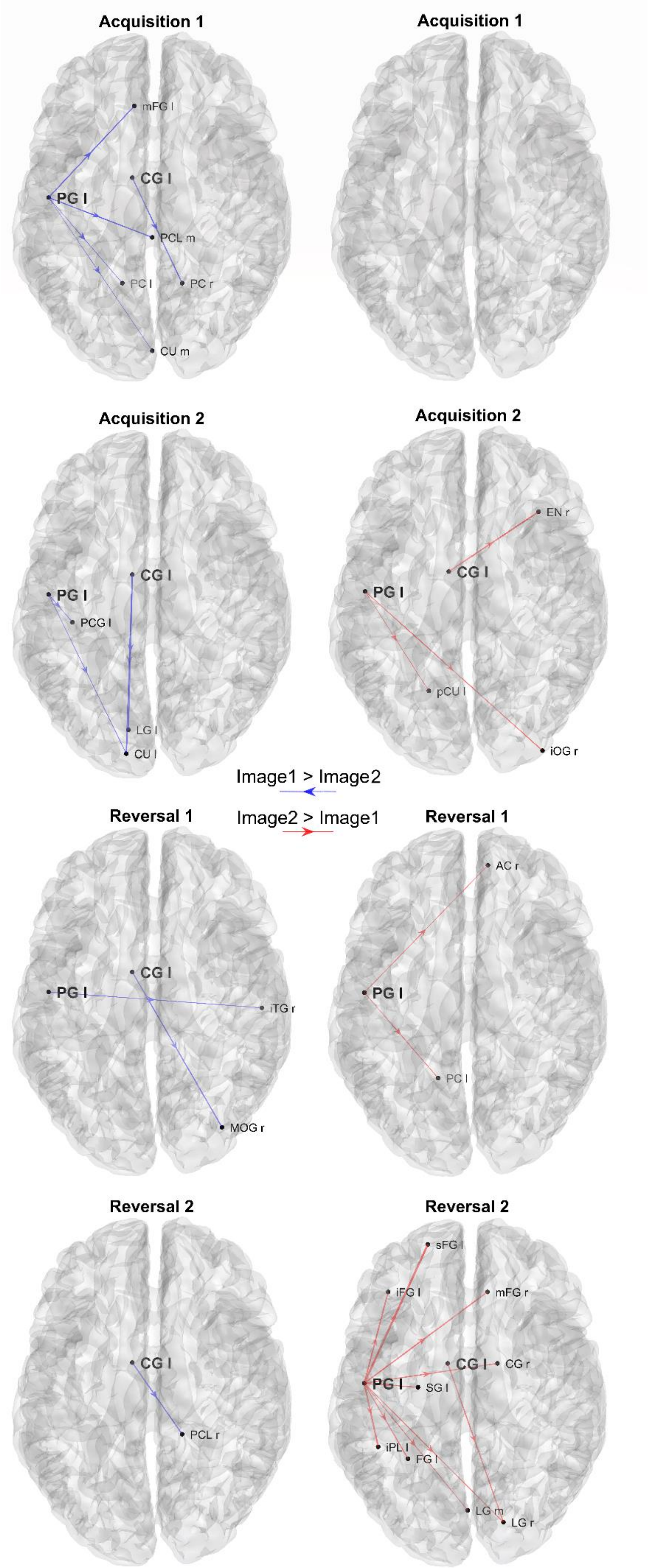


Fig. 4.20 Plots of the significant differences in connectivity between Image 1 and Image 2, calculated in alpha band, block by block. Only the connections coming from the two selected regions PG l and CG l and which are significantly different between Image 1 and Image 2 in each block (p<0.05, uncorrected) are displayed (sparse matrices). The left column shows the connections that are greater during processing of the Image 1 (CS- in acquisition, CS+ in reversal; blue directional arrows), whereas the right column shows connections that are greater during processing of Image 2 (CS+ in acquisition, CS- in reversal; red directional arrows). It is worth noting that differences in connectivity are evident only concerning PG l during acquisition 1 and reversal 2, with a clear inversion of the connectivity and a strong impact especially in reversal.

Alpha connectivity – As evident in Fig. 4.19 (which shows the connections with an absolute value higher than 0.0029 = 1/6 of the mean, emerging from the two selected ROIs), both in the acquisition and reversal, connections arising from the PG l are greater during the CS- stimulus (Image 1 during Acq1 and Acq2, and Image 2 during Rev1 and Rev2), thus mimicking the same behavior as the alpha power. Conversely, the CG l shows a more complex behavior, with some connections being higher during Image 1 and other during Image 2. All figures show an explicit inversion of connections in the acquisition-reversal transition.

In Fig. 4.20, the same connectivity is displayed, but showing only significantly different connections between Image 1 and Image 2 (i.e., using sparse connectivity matrices). In this graph, a clear difference in connections is still evident in PG l, with connections stronger in CS- (i.e., during the presentation of Image 1 in Acq1 and Acq2, Image 2 in Rev1 and Rev2) than in CS+. However, this effect is mainly limited to the first block (acquisition 1) and becomes even more evident in the final block (reversal 2). Conversely, the differences in connections emerging from CG l are scarcely noticeable when the sparse matrix is used.

4.2.4 Discussion

The objective of this work was to investigate the mechanisms of fear acquisition and reversal in healthy human volunteers, laying particular emphasis on the contribution of brain rhythms. To this end, we used high-density scalp EEG and SCR measurements during acquisition and reversal of Pavlovian fear conditioning. Even though fear conditioning has been the subject of many studies in recent years, our work introduces some aspects of novelty: first, we compared the pattern of brain rhythms during acquisition and reversal in all cortical ROIs, to point out similarities and differences between the two phases; second, we looked at the effect of time on fear learning, to point out in which phases rhythms play a pivotal role, and in which phases their role is less evident; third, we analyzed changes in output connectivity from the regions of interest and in the frequency bands more robustly implicated in the fear learning response. Our results confirm several aspects of the literature and introduce new elements that can help clarify the involvement of brain rhythms in Pavlovian fear conditioning.

Theta rhythm - A first significant result concerns the role of the theta rhythm in fear acquisition. Our data show an increase in theta power in the left-mid cingulate cortex in response to the CS+ stimuli. This role is particularly marked during the first block of acquisition (see also Table 4.3) but remains evident (although less pronounced) during the second acquisition block and the first reversal block. Noticeably, in the second reversal block, although CS+ theta power is still higher than the CS- theta power, the difference becomes less substantial, and the role of theta rhythm progressively attenuates. The same pattern is confirmed by the trial-by-trial analysis using the moving average signal, which provides additional attractive cues. Indeed, at the beginning of the first acquisition block, theta power in the cingulate cortex increases abruptly both during CS+ and CS-, probably signaling an alert phase of the experiment. Then, after just 1-2 trials, theta power differentiated between CS+ and CS-. It is worth noting that this difference becomes maximally evident during the mid-period of the acquisition phase and then progressively declines toward the end of acquisition 2. The rapidity of the theta response is confirmed by looking at the reversal phase: just 1-2 shock-associated trials are sufficient to significantly increase theta power in response to the new aversive image.

These patterns, taken together, suggest that theta oscillations in the cingulate cortex signal the presence of a new aversive event, and this pattern is already evident during the first trials of the learning phase. However, these results also underline some differences between acquisition and reversal. Although the theta power difference between CS+ and CS- develops promptly, it remains weaker during the reversal phase than in the previous acquisition phase, especially during the second reversal period. Overall, it seems that the theta rhythm signals the novelty of the aversive event and then declines when the association has been established, with this decline especially evident in the second reversal period, making power difference during reversal less manifest than during acquisition. This result agrees with a recent finding by Taub et al. [403]: the authors suggest that the increase in theta power during aversive conditioning is correlated with the magnitude of conditioned responses but declines once the association is stabilized. Similarly, Ridderbusch et al. [443] observed a temporary increase in neural activation in the anterior cingulate cortex after re-exposure to the US after extinction training and suggested that this is associated with exploratory behavior, signaling changes in US-expectancy and arousal ratings.

Several studies underline the implication of the anterior (mid or dorsal) cingulate cortex in fear acquisition [93,405,444–446]. These results substantially agree with ours. Our study adopted the subdivision among ROIs illustrated in Table 4.2, according to the atlas used by LORETA-KEY©®. Using this atlas, we found significant theta power differences in the cingulate gyrus (left and medial). According to this atlas, the cingulate gyrus includes, among the others, the posterior portions of the Brodmann regions 24 and 32, which are traditionally ascribed to the ACC. However, it is worth noting that the Atlas also includes two other "cingulate" regions named "anterior cingulate" and "posterior cingulate" (see Table 4.2). In particular, the region called "anterior cingulate" includes the most anterior portions of areas 24 and 32. This subdivision agrees with the functional description of the cingulate gyrus proposed by Vogt et al. [447]. The author states that "the greatest number of "fear" activations occur in the anterior part of the midcingulate cortex MCC and not in ACC". The first roughly corresponds to the posterior portions of the Brodmann areas 32 and 24, i.e. to the CG region used in the present atlas.

Many other results in rodents, primates, and humans underline the impact of theta oscillations in fear learning. Synchronization at theta frequencies is suggested to characterize activity in amygdala-hippocampal pathways associated with the consolidation of fear memory [448] and to represent a general mechanism of fear learning across species [449]. A shared hypothesis is that the theta rhythm develops in the amygdala and hippocampus limbic system and is then transmitted to the ventromedial prefrontal cortex and to the anterior midcingulate cortex to synchronize ACC activity, and to transfer error signal information to support memory formation [450]. Indeed, the anterior midcingulate cortex receives afferents from the amygdala [451,452]. Furthermore, synchronized frontomedial theta oscillations are a potential mechanism to support memory communication between brain regions [400].

As to the last point (i.e., theta transmission and synchronization), an original significant result in our study concerns the pattern of connections emerging from the previous two regions (CG l and CG m). We observed that Granger connectivity in the theta range is stronger during CS+ than during CS-, and these differences are also evident during reversal. A possible interpretation is that the increased theta power in these regions is then transmitted to other areas of the brain, thus producing a generalized theta synchronization, subserving the retrieval of fear responses, or a general process of adaptive control of an unpleasant event (see also [93,453]). Interestingly, if attention is focused only on connections statistically different between CS+ and CS- (not only to the absolute differences), the increase in connectivity during CS+ is especially evident during the first acquisition and second reversal blocks. Hence, a puzzling phenomenon is that theta power differences between CS+ and CS- decline in reversal 2, whereas connectivity differences become more evident in the same block. Thus, functional connectivity does not simply reflect a change in power of the theta rhythm transmitted to other regions but may also depend on an effective alteration of synapses, especially in the last portion of the experiment. Further studies are needed to clarify this crucial point.

Alpha rhythm – A significant observation emerging from our data is that alpha power changes are less localized than the changes in theta power and involve a more extensive network mainly located in portions of the posterior frontal cortex and parietal lobes, with a predominance in the left hemisphere. Some of these zones (the precentral gyrus, paracentral lobule and inferior parietal lobule) are implicated in motor and sensory innervation. As expected, alpha power is smaller during CS+ than CS- in all these regions, reflecting a condition of greater arousal. However, it is worth noting that alpha power is higher than baseline during all phases of the experiment, probably since the initial period of the experiment (before any trial) was felt as the most stressful condition for the participants, possibly due to uncertainty of what will happen next. Alpha power differences are more evident in the acquisition phase than in the reversal phase, and, in the cingulate gyrus, alpha power exhibits a progressive increase during the experiment, suggesting increasing relaxation of the participants over the course of the experiment.

The pattern of alpha power changes during fear conditioning was investigated in detail by Chien et al. [407]. The authors observed a significant alpha event-related desynchronization (ERD; i.e. a decrease in power) at parietal and occipital channels, hence over sensory structures related to (visual) CS processing. These changes were especially evident in the early phase of the stimulus train, reflecting a difference between the early and late stages of acquisition. By comparing their results with SCR data, the authors concluded that alpha power changes mainly reflect the valence and salience of the stimulus, i.e., the ability of CS to capture attention and motivate behavior. However, at odd with our results, the authors did not find significant differences in alpha power between CS+ and CS-. Differences between our results and those by Chien et al. can be explained by thinking that these authors mainly focused on the magnitude of alpha ERD, which is maximal in the occipital regions, implicated in the visual processing of the external stimuli. Conversely, we focused on statistical differences between

CS+ and CS-, concentrated in parietal and posterior frontal regions, i.e., in the zones mainly involved in tactile and motor processing.

Since the precentral gyrus is primarily involved in motor processing, a decreased alpha during CS+ in this zone may reflect greater motor activation in preparation for an escape (for instance, preparation of movement of the right arm where the shock is delivered). Indeed, alpha ERD reflects the gradual release of inhibition associated with the emergence of a task-response. In contrast, an increase in alpha oscillations (event related synchronization, ERS) is observed with the CS-. Alpha ERS is commonly ascribed to idling or suppressing activity in task-irrelevant sites [454]. Hence, our result supports the idea that alpha power changes observed in parietal and posterior frontal zones primarily reflect a preparative response to an action (during CS+) or a partial idling (during CS-) of the same activity.

Finally, it is worth noting that the time response of this alpha pattern is slower than that of the theta response: as evident looking at the moving average, alpha ERS during CS- requires 3-4 trials to develop. Another interesting aspect is that alpha power exhibits a drastic fall (ERD) at the beginning of any new block, reflecting greater attention/arousal due to the unfamiliar new conditions. Then alpha power progressively increases (especially in CS-), reducing the response in motor areas.

The connectivity pattern in the alpha band further underlines the pivotal role played by the left precentral gyrus. Stronger outflow connectivity is evident in this area during CS- than during CS+, reflecting the higher alpha power transmitted towards other occipital, parietal and frontal regions. It is worth noting that the reverse of this connectivity pattern is relatively slow, being maximally evident during reversal 2 than during reversal 1. This phenomenon is similar to what has already been observed for the theta connectivity from the left and medial cingulate cortex. In other terms, connectivity changes mature more slowly during reversal than during acquisition, becoming fully evident in the second reversal block. This pattern probably reflects synaptic changes necessary to overcome a previous pattern of connectivity developed during the acquisition phase. Indeed, as shown in recent modeling studies [126,455] functional connectivity mainly reflects the amount of information transmitted from one region to another: the latter can depend both on the power in the source region and on the strength of the effective connectivity linking the two regions.

Unexpectedly, the pattern of alpha-band connectivity emerging from the left cingulate cortex apparently contradicts the pattern of alpha power: in fact, many of these connections are higher during CS+ than during CS-, i.e., in conditions of ERD. We do not have a definitive explanation for this pattern. However, we suspect that these seemingly anomalous connectivity patterns reflect non-linear phenomena and are strongly affected by changes in theta power (which, as demonstrated above, are significant in the left cingulate cortex and are higher in CS+ than CS-). In previous papers [126,455,456] using a neural mass model as ground truth and comparing the actual connectivity values in the model with those obtained with methods for functional connectivity assessment, we demonstrated that non-linear phenomena play a significant role in connectivity estimation, resulting in possible interference between frequency bands and alterations in the connectivity values.

Gamma rhythm – There is a consensus in the literature that gamma power is implicated in inhibiting a previously acquired fear response [457–459]. In agreement with Mueller et al. (2014)[93], in humans extinguished vs. non-extinguished stimuli evoked an increased gamma power localized in the vmPFC. The role of the vmPFC in extinction is further supported by neuroimaging studies [460]. However, Schiller et al. [420] pointed out that reversal is a more complex process than extinction. Using fMRI, these authors observed that, during reversal, the activity in the vmPFC signals the presence of a safe stimulus (hence the new CS- in reversal, previously CS+ during acquisition), which can be interpreted as an unexpected reward.

According to the studies mentioned above, a significant gamma activity in the vmPFC was expected in reversal; however, in our research, we were unable to find any corrected statistical difference in gamma between CS+ and CS- during any phase of the experiment. For this reason, gamma activity was not further analyzed.

4.2.5 Conclusions

The results obtained in this study confirm several observations of previous studies and add new aspects. i) Increase in theta rhythm power occurs in the mid portion of the cingulate cortex during CS+ and is associated with an increase in outflow connectivity. This may reflect a rhythm from the amygdala and hippocampus, which is then transmitted to other cortical regions allowing a fast theta synchronization, as supported by our Granger causality analysis. Theta synchronization may play a pivotal role during the acquisition of fear conditioning. ii) Alpha power ERD during CS+ and alpha power ERS during CS- occur mainly in the left posterior frontal and parietal cortex, with the most substantial evidence in the left precentral gyrus. These two phenomena may reflect an excitation of these motor areas (movement preparation) in case of an aversive stimulus and a progressive inhibition of these areas in case of a safe stimulus, respectively. iii) The dynamics of theta power changes appear faster than those of the alpha rhythm, reflecting a trial-by-trial basis. iv) All the previous phenomena are present during acquisition and reversal, but differences between CS+ and CS- are less prominent in the reversal phases. This may be due to the difficulty of overcoming a previously acquired memory. v) Changes in power are associated with increased Granger connectivity emerging from the areas involved. Unexpectedly, these connectivity changes are also strongly evident in the second reversal block when power differences are attenuated. This phenomenon may reflect changes in real connectivity instead of simple changes in oscillation power and requires further study.

4.2.6 Supplementary Material

Fig. 11S shows the normalized mean alpha power graphs for the five regions not shown in *Cortical sources power analysis – Alpha*, in the Results section.

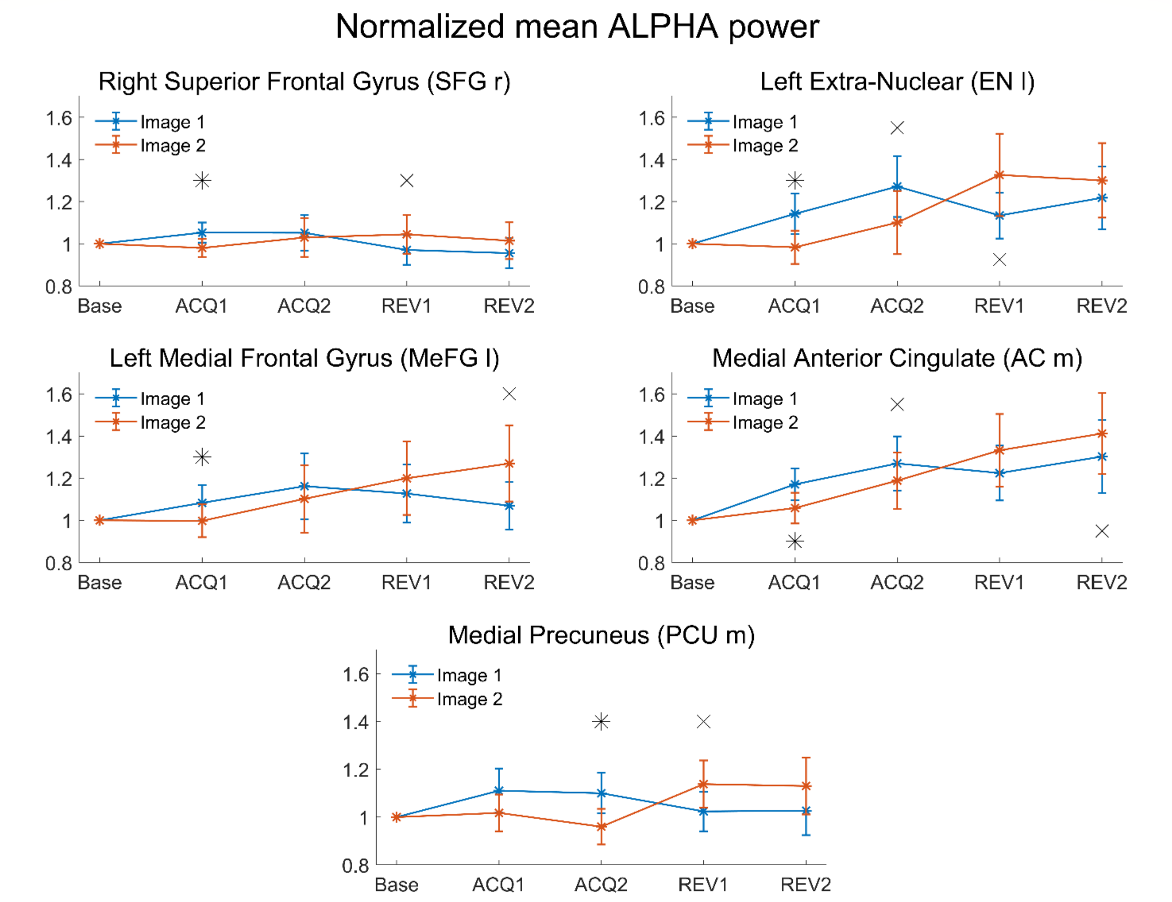


Fig.11S - Normalized mean power in the alpha band, for all four blocks and both images, in the right superior frontal gyrus (SFG r), left extra-nuclear (EN l), left medial frontal gyrus (MeFG l), medial anterior cingulate (AC m) and medial precuneus (PCU m). Results for Image 1 are depicted in blue and those for Image 2 in red, accompanied in each block by the respective SEM bar. Asterisks indicate the presence of corrected statistical significance (p<0.05, false discovery rate correction) in the specific block, while crosses denote the presence of a statistical significance (p<0.05, uncorrected) which does not survive correction for multiple comparisons. It is well evident the power inversion in passing from acquisition to reversal, i.e., alpha power is always greater during CS- (Image 1 in Acq1 and Acq2; Image 2 in Rev1 and Rev2) than CS+.

Below, Fig. 12S-14S show the normalized alpha power moving average graphs for the nine regions not presented in **Cortical sources power analysis –** ***Alpha***, in the Results section.

Please note how the following images do not show the results of the statistical analysis (crosses at particular trials, see main text). Nevertheless, the statistical analysis was still performed and the total number of significant (uncorrected) trials is lower for all of the following areas than for the areas shown in the main text (CG l and PG l).

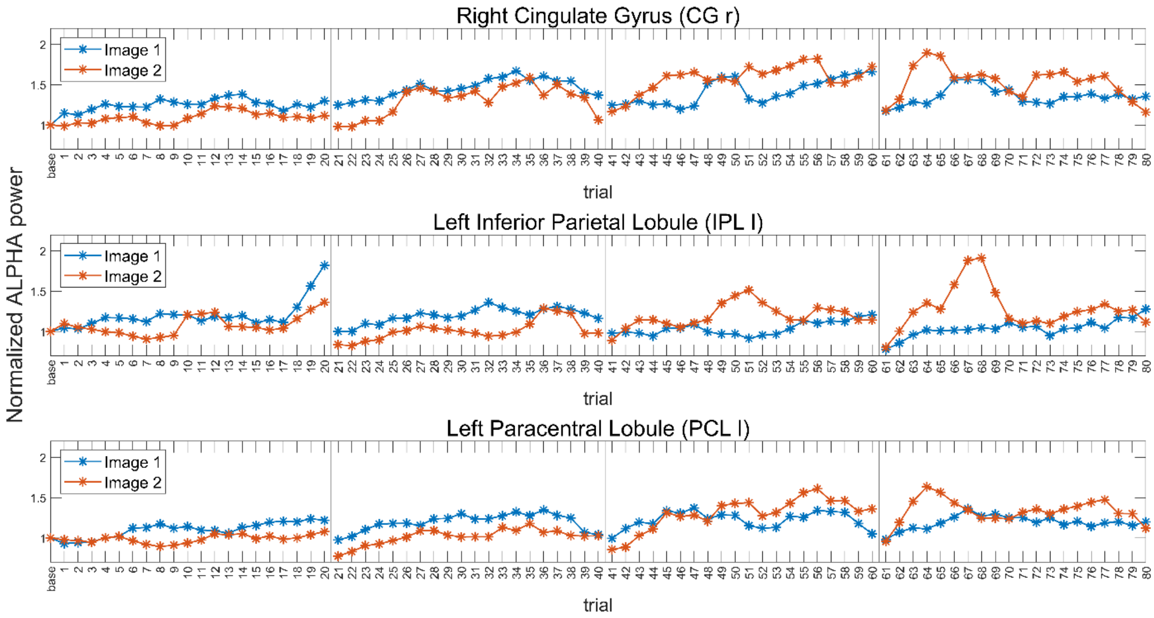


Fig. 12S - Moving averages (w=3 trials) of normalized alpha power, trial by trial, for the four blocks. The top row shows the normalized power of the right cingulate gyrus (CG r), the central row the normalized power of the left inferior parietal lobule (IPL l), and the bottom row the normalized power of the left paracentral lobule (PCL l). The two images are shown in the same color (blue for Image 1, red for Image 2). Vertical lines are used to delineate the four different blocks.

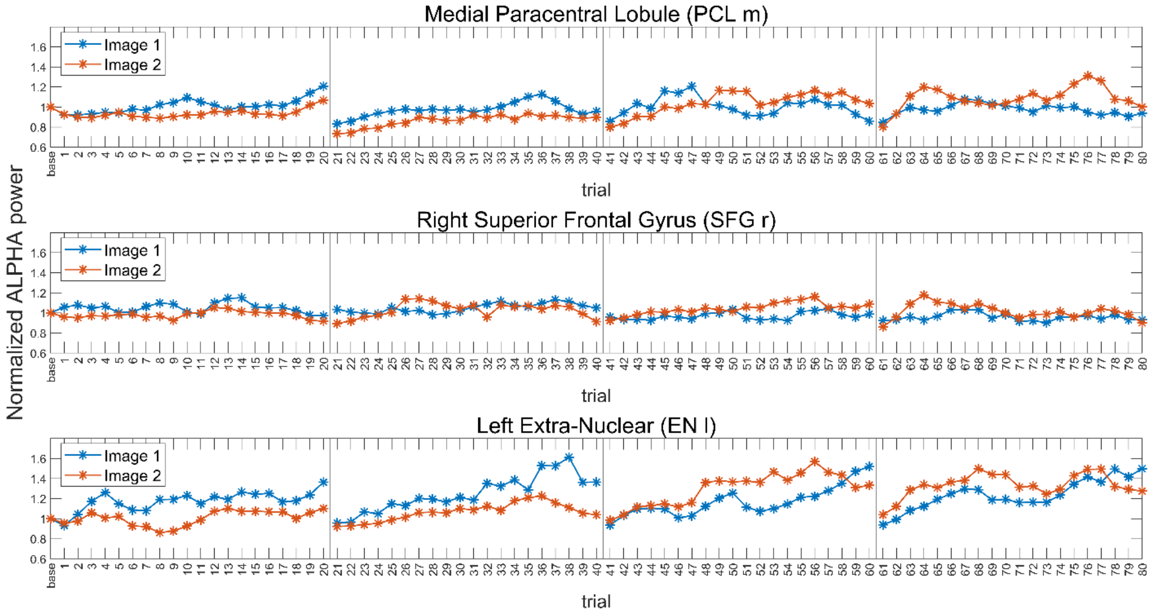


Fig. 13S - Moving averages (w=3 trials) of normalized alpha power, trial by trial, for the four blocks. The top row shows the normalized power of the medial paracentral lobule (PCL m), the central row the normalized power of the right superior frontal gyrus (SFG r), and the bottom row the normalized power of the left extra-nuclear (EN l). The two images are shown in the same color (blue for Image 1, red for Image 2). Vertical lines are used to delineate the four different blocks.

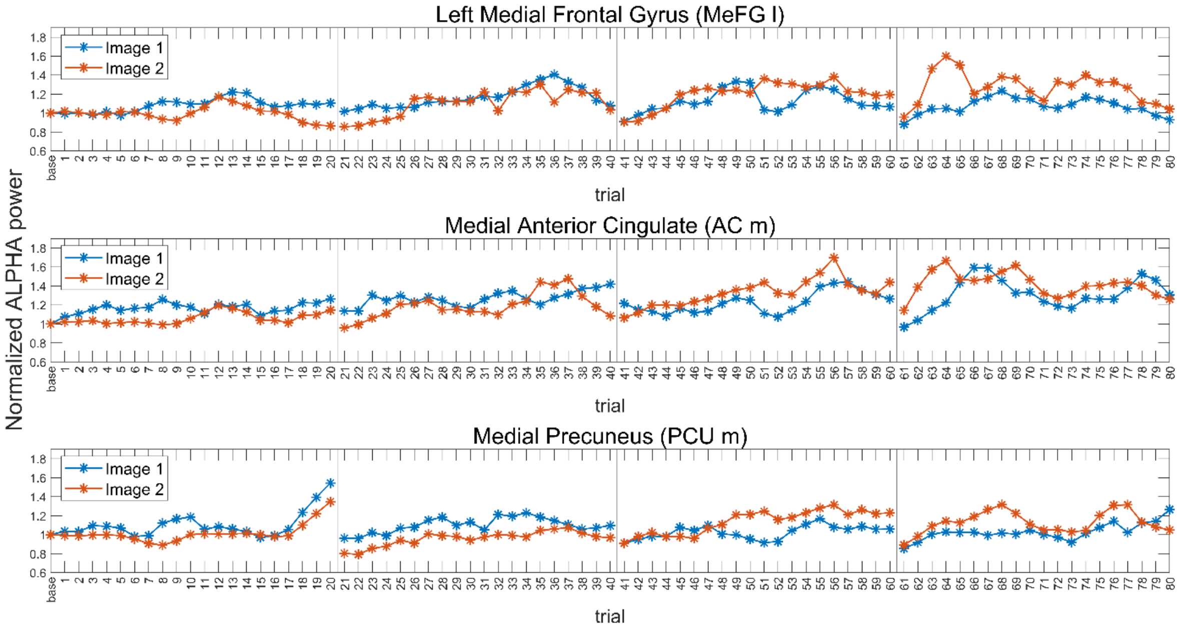


Fig. 14S - Moving averages (w=3 trials) of normalized alpha power, trial by trial, for the four blocks. The top row shows the normalized power of the left medial frontal gyrus (MeFG l), the central row the normalized power of the medial anterior cingulate (AC m), and the bottom row the normalized power of the medial precuneus (PCU m). The two images are shown in the same color (blue for Image 1, red for Image 2). Vertical lines are used to delineate the four different blocks.

4.3 Resting-state bottom-up connectivity in individuals with high autistic traits

The study reported in this chapter refers to the published journal paper entitled "Bottom-up vs. top-down connectivity imbalance in individuals with high-autistic traits: An electroencephalographic study", Mauro Ursino[1*], Michele Serra[1], Luca Tarasi[2], Giulia Ricci[1], Elisa Magosso[1], Vincenzo Romei[1,3], *Frontiers in Systems Neuroscience* (2022).

In this study, we investigated differences in directed connectivity using EEG resting-state recordings in individuals with low and high autistic traits. The connectivity network analysis was performed using Temporal Granger Causality and some centrality indices taken from graph theory: *in degree*, *out degree*, *authority*, and *hubness*. These measures were chosen since they preserve the information on the direction of the connection, which is of great relevance in autism spectrum disorder (ASD).

Background*: Brain connectivity is often altered in autism spectrum disorder (ASD). However, there is little consensus on the nature of these alterations, with studies pointing to either increased or decreased connectivity strength across the broad autism spectrum. An important confound in the interpretation of these contradictory results is the lack of information about the directionality of the tested connections. Here, we aimed at disambiguating these confounds by measuring differences in directed connectivity using EEG resting-state recordings in individuals with low and high autistic traits.* ***Methods:*** *Brain connectivity was estimated using temporal Granger Causality applied to cortical signals reconstructed from EEG. Between-group differences were summarized using centrality indices taken from graph theory (in degree, out degree, authority, and hubness).* ***Results:*** *Results demonstrate that individuals with higher autistic traits exhibited a significant increase in authority and in degree in frontal regions involved in high-level mechanisms (emotional regulation, decision-making, and social cognition), suggesting that anterior areas mostly receive information from more posterior areas. Moreover, the same individuals exhibited a significant increase in the hubness and out degree over occipital regions (especially the left and right pericalcarine regions, where the primary visual cortex is located), suggesting that these areas mostly send information to more anterior regions.* ***Discussion:*** *Hubness and authority appeared to be more sensitive indices than the in degree and out degree. The observed brain connectivity differences suggest that, in individual with higher autistic traits, bottom-up signaling overcomes top-down channeled flow. This imbalance may contribute to some behavioral alterations observed in ASD.*

4.3.1 Introduction

Autism is a complex neurodevelopmental condition characterized by several behavioral peculiarities, involving avoidance of social interactions, reduced communication, and restricted interests (see the American Psychiatric Association [APA] (2022)). The biological origin of this condition is a subject of active research, in an effort to understand its fundamental neural mechanisms. In this regard, a current perspective is that autistic traits could be explained by modifications in brain network characteristics, especially in the connectivity among brain areas underlying perception, social cognition, language, and executive functions [461].

Indeed, many recent studies have reported that individuals within the autism spectrum disorder (ASD) exhibit altered brain connectivity compared to typically developing individuals. However, literature reports are often inconsistent (see review papers by [462–464]). The traditional point of view, predominantly supported by studies using structural and functional MRI, hypothesizes that autism is characterized by long-range underconnectivity, potentially combined with local overconnectivity [465–467]. Conversely, there have been several more recent studies, using EEG and MEG, in which the hypoconnectivity hypothesis could not be confirmed in ASD. Rather, several studies pointed to hyperconnectivity among specific brain areas, especially between thalamic and sensory regions [468] or between the extrastriatal cortex, frontal and temporal regions [469–471]. Finally, a third line of evidence points towards the existence of a more subtle mixture of hypo- and hyper-connectivity, suggesting the presence of multiple mechanisms [461,472–474].

Some of these differences, of course, can derive from methodological issues. Connectivity is an elusive concept that can be dramatically affected by the measurement technique adopted (for instance, fMRI vs. EEG/MEG), by the particular task involved (vs. resting state analysis), and perhaps more importantly, by the specific measure employed to estimate the connection strength (e.g., functional, effective or anatomical connectivity, directed or undirected measures, bivariate or multivariate). Indeed, most connectivity measures in literature are not-directional and hence are inadequate to discover differences in lateralization or in top-down vs. bottom-up information processing [475].

In particular, it is well-known that cognitive functions are characterized by a complex balance between integration, involving the coordination among several brain areas, and segregation, involving specialized computations in local areas. According to the predictive coding theory [476], the brain continually generates models of the world by integrating data coming from sensory input with information from memory. Sensory perception is thus the result of a combination between present data from the external world (usually carried by feedforward bottom-up connectivity) and past or prior knowledge (mainly conveyed through feedback, top-down connections); hence, an equilibrium between these directional connectivity patterns is necessary to adaptatively integrate stimuli-driven and internally-driven representations, preventing their segregation or excessive bias towards one or the other.

Recent hypotheses [477,478] assume that ASD individuals exhibit an impaired predictive coding, characterized by an imbalance between these two processing streams, i.e., dominant bottom-up processing and relatively weaker top-down influences compared with control individuals. This signifies that people in the autistic spectrum would pose much more emphasis on present sensory stimuli and somewhat less weight on contextual information. This imbalance, in turn, may result in poor social adaptation and insufficient appropriateness to social requirements [479]. Results that support this point of view include a reduced susceptibility to illusions and top-down expectations [480,481] and increased local (vs. global) processing in individuals within the autism spectrum [482,483] leading to a more stimulus- and detail-driven perceptual style.

The aforementioned alterations in predictive coding may be caused by altered brain connectivity, especially concerning top-down vs. bottom-up circuitry [484]. Additionally, alterations in connectivity patterns may involve a different transmission of brain rhythms and an impaired wave synchronization, which plays a pivotal role in several cognitive tasks, including attention, information selection, working memory, and emotion [485,486].

Finally, increasing evidence both at the genetic and behavioral levels demonstrates that autism does not represent a dichotomy condition (i.e., one ON/OFF in type) but is best described as a spectrum of manifestations ranging from clinical forms to trait-like expressions within the general population [483,487,488] that share a peculiar cognitive style that distinguishes them from the rest of the clinical and nonclinical population [484].

Following these ideas, in a recent paper [439], we investigated whether the patterns of brain connectivity, estimated with Granger causality from EEG source reconstruction, exhibit differences in two nonclinical groups classified as low or high on autistic traits. Preliminary results suggested that connectivity along the fronto-posterior axis is sensitive to the magnitude of the autistic features and that a prevalence of ascending connections characterized participants with higher autistic traits.

The present study aims to further extend the previous work on a larger cohort allowing for an improved connectivity analysis by implementing measures taken from the graph theory. In particular, new aspects of the present study concern: i) the use of a larger data set; ii) a preliminary analysis at the lobe level; iii) the use of more sophisticated indices taken from the graph theory, such as hubness and authority; iv) the use of a more sophisticate statistical analysis (i.e., the use of sparse connectivity matrices) to better point out differences in connectivity between the two groups.

Particularly, graph theory represents a powerful tool able to summarize complex networks consisting of hundreds of edges, using a few parameters with a clear geometrical meaning. Recently, this theory has been applied with increasing success as an integrative approach, able to evaluate the complex networks that mediate brain cognitive processes [229,489–491]. In particular, since our attention here is primarily devoted to the presence of differences in the direction of connections (ascending vs. descending, lateralization, etc.), we focused our analysis on the *in degree* and *out degree*, defined as the sum of connection strengths entering or leaving a given node. Furthermore, we also tested whether two analogous but more specialized measures of centrality, *hubness* and *authority*, can provide additional information to better characterize directionality. The hub's index of a node is the weighted sum of the authority's indices of all its successors; hence, this measure summarizes the capacity of a node to send information to other critical, authoritative nodes. The authority's index of a node is the weighted sum of the hub's indices of all its predecessors and summarizes the capacity of a node to receive essential information from hubs. Here, we investigate whether differences in these measures, and the pattern of *out* and *in* connections from the dominant nodes, can reveal a difference in the network's topology, and alterations in information processing, as a function of the autistic trait.

4.3.2 Materials and methods

4.3.2.1 Participants

Forty participants (23 female; age range 21–30, mean age = 24.1, SD = 2.4), with no neurocognitive or psychiatric disorders, took part in the study. All participants signed a written informed consent before taking part in the study, conducted according to the Declaration of Helsinki and approved by the Bioethics Committee of the University of Bologna. All participants completed the Autism-Spectrum Quotient test (AQ)[487]. The mean AQ score was 16.1 ± 6.6. The AQ is a self-report widely used to measure autistic traits in the general population. It provides a global score, with higher values indicating higher levels of autistic traits. We used the original scoring methods converting each item into a dichotomous response (agree/disagree) and assigning the response a binary code (0/1). In the present study, the total score of the AQ was considered, and the Italian version of the AQ was adopted [492]. The participants were divided into two groups, depending on their AQ score being below or above a given cutoff, with the cutoff set to 17, since this value corresponds to the average AQ score in the non-clinical population [493]. In the following, we will refer to the two groups of participants as Low AQ score Group (N = 21) and High AQ score Group (N = 19).

4.3.2.2 EEG acquisition and preprocessing

Participants comfortably sat in a room with dimmed lights. EEG was recorded at rest for two minutes while participants kept their eyes closed. A set of 64 electrodes was mounted according to the international 10–10 system. EEG was measured with respect to a vertex reference (Cz), and all impedances were kept below 10 kΩ. EEG signals were acquired at a rate of 1000 Hz. EEG was processed offline with custom MATLAB scripts (version R2020b) and the EEGLAB toolbox [494]. The EEG recording was filtered offline in the 0.5-70 Hz band. The signals were visually inspected, and noisy channels were spherically interpolated. An average of 0.05 ± 0.15 channels were interpolated. The recording was then re-referenced to the average of all electrodes. Subsequently, we applied the Independent Component Analysis (ICA), an effective method largely employed to remove EEG artifacts. In particular, we removed the EEG recording segments corrupted by noise through visual inspection and then we removed all the independent components containing artifacts clearly distinguishable by means of visual inspection from brain-related components. An average of 3 ± 3.7 independent components were removed for each participant.

4.3.2.3 Cortical Sources Reconstruction and ROIs definition

Since we were interested in connectivity analysis, cortical source activity was reconstructed from pre-processed EEG signals. To this aim, intracortical current densities were estimated using the Matlab toolbox Brainstorm [495]. Firstly, to solve the forward problem, a template head model based on realistic anatomical information (ICBM 152 MNI template) was used. The model consists of three layers representing the scalp, the outer skull surface, and the inner skull surface, and includes the cortical source space discretized into 15002 vertices. The forward problem was solved in OpenMEEG software [496] via the Boundary Element Method.

sLORETA (standardized Low-Resolution Electromagnetic Tomography) algorithm was used for cortical sources estimation. sLORETA is a functional imaging technique belonging to the family of linear inverse solutions for 3D EEG distributed source modeling [497]. Specifically, this method computes a weighted minimum norm solution, where localization inference is based on standardized values of the current density estimates. The solution provided is instantaneous, distributed, discrete, linear with the property of zero dipole-localization error under ideal (noise-free) conditions. Constrained dipole orientations were chosen for sources estimation, modeling each dipole as oriented perpendicularly to the cortical surface. Hence, for each participant, we reconstructed the resting-state time series of standardized current densities at all 15002 cortical vertices.

Then, the cortical vertices were grouped into cortical regions according to the Desikan-Killiany atlas [498] provided in Brainstorm, which defines 68 regions of interest (ROIs). The activities of all vertices belonging to a particular ROI were averaged at each time point, obtaining a single time series representative of the activity of that cortical ROI. It is worth noticing that, by considering the average behavior at the ROIs level, it was possible to mitigate some possible inaccuracies in source reconstruction at single vertex level, due to the use of a template head model for all participants (instead of subject-specific head models). Table 4.5 lists the 68 Desikan-Killiany ROIs and provide the mapping of individual ROIs to each lobe.

Table 4.5 – The approximate mapping of the 'Desikan-Killiany' ROIs to the lobes. The Desikan-Killiany atlas comprises 34 ROIs in each hemisphere (see Fig. 2.2). The mapping proposed by FreeSurfer (https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation) was used as a reference. The only difference between our mapping and the reference resides in the mapping of the insula, which was not ascribed to any lobe in FreeSurfer. We assigned the insula to the parietal lobe.

ROI	Label	Lobe	ROI	Label	Lobe
Banks of Sup. Temp. Sulcus	BK	Temporal	**Parahippocampal**	PH	Temporal
Caudal Anterior Cingulate	cAC	Frontal	**Pars Opercularis**	pOP	Frontal
Caudal Middle Frontal	cMF	Frontal	**Pars Orbitalis**	pOR	Frontal
Cuneus	CU	Occipital	**Pars Triangularis**	pTR	Frontal
Entorhinal	EN	Temporal	**Pericalcarine**	PCL	Occipital
Frontal Pole	FP	Frontal	**Postcentral**	POC	Parietal
Fusiform	FU	Temporal	**Posterior Cingulate**	PCG	Parietal
Inferior Parietal	IP	Parietal	**Precentral**	PRC	Frontal
Inferior Temporal	IT	Temporal	**Precuneus**	PCU	Parietal
Insula	IN	Parietal	**Rostral Anterior Cingulate**	rAC	Frontal
Isthmus Cingulate	IST	Parietal	**Rostral Middle Frontal**	rMF	Frontal
Lateral Occipital	LO	Occipital	**Superior Frontal**	SF	Frontal
Lateral Orbitofrontal	lOF	Frontal	**Superior Parietal**	SP	Parietal
Lingual	LG	Occipital	**Superior Temporal**	ST	Temporal
Medial Orbitofrontal	mOF	Frontal	**Supramarginal**	SMG	Parietal
Middle Temporal	MT	Temporal	**Temporal Pole**	RP	Temporal
Paracentral	PAC	Frontal	**Transverse Temporal**	TT	Temporal

4.3.2.4 Granger Causality Analysis

Once the time waveform in each cortical ROI was estimated (as described above), for each participant k ($k = 1, \dots ,40$) we evaluated the connectivity among the ROIs. To this aim, we adopted Granger Causality (GC)[18,114,117,499], which provides directional metrics of functional connectivity, and is based on the autoregressive (AR) modeling framework as described in the following.

Let's indicate with $x_{k,i}[n]$ and $x_{k,j}[n]$ two temporal series representing the activity of two distinct cortical ROIs (ROI_i and ROI_j) for participant k, where n is the discrete time index. The Granger Causality quantifies the causal interaction from ROI_i to ROI_j as the improvement in predictability of $x_{k,j}[n]$ at time sample n when using a bivariate AR representation, including both past values of $x_{k,j}$ and past values of $x_{k,i}$, compared to a univariate AR representation, including only past values of $x_{k,j}$. Mathematically, the following two equations hold for the univariate and bivariate AR model, respectively

$$x_{k,j}[n] = \sum_{m=1}^{p} a_{k,j}[m]\, x_{k,j}[n-m]] + \eta_{k,j}[n] \qquad 4.1$$

$$x_{k,j}[n] = \sum_{m=1}^{p} b_{k,j}[m]\, x_{k,j}[n-m] + \sum_{m=1}^{p} c_{k,ji}[m]\, x_{k,i}[n-m] + \varepsilon_{k,j}[n] \qquad 4.2$$

Index m represents the time lag (in time samples), and p (model order) defines the maximum time lag, i.e., the maximum number of lagged observations included in the models. Thus, in Eq. 4.3.1, the current value of $x_{k,j}$ (at time sample n) is predicted in terms of its own p past values (at time samples $n-1, n-2, \dots, n-p$), while in Eq. 4.3.2 prediction is made also in terms of the p past values of $x_{k,i}$. a, b, c are the model's coefficients (dependent on time lag), and the time series $\eta_{k,j}[n]$ and $\varepsilon_{k,j}[n]$ represent the prediction error of the univariate and bivariate AR model, respectively. The prediction error variance quantifies the model's prediction capability based on past samples: the lower the variance, the better the model's prediction. The GC from $x_{k,i}$ to $x_{k,j}$ is defined as the logarithm of the ratio between the variances of the two prediction errors, i.e.

$$GC_{k,ROI_i \to ROI_j} = \ln \frac{var\{\eta_{k,j}[n]\}}{var\{\varepsilon_{k,j}[n]\}} \qquad 4.3$$

The measure in Eq. 4.3.3 is always positive: the larger its value, the larger the improvement in $x_{k,j}[n]$ prediction when using information from the past of $x_{k,i}$ together with the past of $x_{k,j}$, and this is interpreted as a stronger causal influence from ROI_i to ROI_j. Similarly, Granger Causality from $x_{k,j}$ to $x_{k,i}$, $GC_{k,ROI_j \to ROI_i}$, is computed via the same procedure, building the AR models for the time series $x_{k,i}$.

For each participant k, we computed the two directed measures of GC for each pair of ROIs, overall obtaining 68 x 68 connectivity values (with all auto-loops equal to zero). In all cases, the order p of the AR models was set equal to 20, corresponding to 20 ms time span at 1000 Hz sampling rate (as in our data); thus, in this study, the functional interactions between nodes were evaluated within 20 ms time delay. This value for parameter p was determined based on a preliminary analysis where we tested different values for the order of the model, obtaining that GC results did not change substantially for $p \geq 20$.

4.3.2.5 Indices Derived from Graph Theory

As previously reported by other authors [500,501] the connectivity between the ROIs of a brain network can be described as a weighted graph, where the magnitude of the connectivity between two ROIs is represented as the weight of an edge, whilst the ROIs connected by the edge are the nodes of the graph. A most remarkable consequence of the adoption of this representation for the brain network is the introduction of several concepts and measures from Graph Theory, which allows us to achieve a better understanding of the network's topology [489–491]. For this study, we focused on centrality indices that take into account the direction of connections, specifically *authority*, *hubness*, *in degree*, and *out degree* centralities. These indices, which will be detailed in the following, were specifically selected for their focus on the ROIs' inputs and outputs, which we hypothesized could offer confirmatory evidence of connectivity patterns previously observed in individuals with low and high autistic traits [439].

4.3.2.6 The Graph

A graph is the mathematical abstraction of the relationships between some entities. The entities connected in a relationship are called "nodes" of the graph and are often represented graphically in the form of points. These nodes are connected by edges. While the simplest form of a graph is undirected (i.e., the edges do not have orientation), the graph we use to describe a brain network is a weighted directed graph (or digraph), i.e., it has oriented edges, each one with a weight representing the strength of the connection.

To obtain the graphs, for each participant the connectivity matrix was normalized so that its elements provided a sum of 100 (i.e., each connectivity value was divided by the total sum of connections and multiplied by 100). Furthermore, the normalized 68 x 68 matrices (which we will be calling "complete" matrices for clarity) were turned into 68 x 68 sparse matrices by removing (i.e., setting to zero) any connection that was not significantly different between the High and Low AQ score Groups. In particular, a two-tailed Monte-Carlo testing was applied (5000 permutations) and, based on its results, not significant connections were defined as having an uncorrected p-value greater than 0.05.

Forty graphs (one per participant) were obtained both for the complete normalized and the sparse matrices. For each of these graphs, centrality indices were then computed. Although a preliminary investigation was performed on the complete matrices, our analysis is mainly focused on sparse matrices since by excluding "similar" connections we expect to better capture differences in the connectivity patterns and in graph indices between the two groups.

4.3.2.7 Centrality Indices

Graph theory defines a multitude of indices and coefficients that allow describing the topology of a network from different points of view. Centrality indices are part of these. They measure the importance of a particular node in the network. The four centrality indices considered in this study (*in degree*, *out degree*, *authority*, *hubness*) quantify the importance of a node as a source or a sink for the edges. In the following, we will first introduce the *in degree* and *out degree* centralities; then, *authority* and *hubness* will be described, stressing on how they differ from *in degree* and *out degree*. *In degree* is the sum of the weights of the edges entering into a node, while *out degree* is the sum of the weights of the edges exiting from a node. As a result of their direct dependence on the strength of input and output connections, *in degree* and *out degree* provide an immediate description of the nodes most involved in the transmission (*out degree*) and reception (*in degree*) of information. The mathematical formulation of *in degree* and *out degree* is given in Section 2.4.2 of Chapter 2 and is described by Eq. (2.35-2.36).

Authority and *hubness* centralities include a more refined concept compared to *in degree* and *out degree* centralities and have a distinctive feature of strict interdependence. Their mathematical formulation is the following one.

Authority is proportional to the sum of the weights of edges entering a node, multiplied by the *hubness* of the node the edge originates from; *hubness* is proportional to the sum of the weights of edges exiting from a node, multiplied by the *authority* of the node the edge points to. The mathematical formulation of *in degree* and *out degree* is given in Section 2.4.2 of Chapter 2 and is described by Eq. (2.37-2.38).

These indices were computed using the function provided by the Matlab's libraries contained in the Category "Graph and network algorithms" (Matlab R2021a), particularly the command digraph/centrality. This function sets both α and β equal to 1 and calculates *authority* and *hubness* via an iterative procedure.

Similar to *in degree* and *out degree*, *hubness* and *authority* provide a measure of which nodes of the network are primarily involved in the transmission (*hubness*) and reception (*authority*) of information, but they also mutually account for the centrality of the receiving and sending nodes. In particular, since these two centrality indices point to each other (i.e., to compute *authority*, we use *hubness*, and vice versa), they imply that strong connections exist between nodes with high *authority* and nodes with high *hubness*, and these indices may be useful to further emphasize any existing directionality in the connectivity pattern.

4.3.2.8 Connectivity Analysis

For each participant, starting from either the complete normalized or the sparse 68 x 68 matrix, the four centrality indices were computed at each of the 68 ROIs. Additionally, we computed the average complete and sparse connectivity matrix in the Low AQ score Group and in the High AQ score Group, and then their difference.

Initially, we performed an analysis at the level of macro regions (englobing several ROIs) rather than at single ROI level. To this aim, we considered 8 regions corresponding to brain lobes (frontal, parietal, temporal, and occipital lobes, both left and right). Specifically, for each participant, the 68 x 68 connectivity matrix was transformed into an 8 x 8 connectivity matrix; the elements of the 8 x 8 matrix were filled in with the mean value of all the connections going from one lobe to another. The elements of the 8 x 8 matrices were subsequently tested for statistical significance across the two groups of participants, by applying a two-tailed t-test (significance level 0.05, no correction), resulting in 64 comparisons. Furthermore, the 8 x 8 difference matrix was computed, by subtracting the 8 x 8 mean connectivity matrix of the Low AQ score Group from the 8 x 8 mean connectivity matrix of the High AQ score Group. Thus, the elements of the difference matrix greater than 0 represented stronger connectivity for the High AQ score Group, while elements of the difference matrix less than 0 represented stronger connectivity for the Low AQ score Group.

Then, a more detailed analysis was performed at the level of each ROI.

In the case of the complete connectivity matrix, we identified the ROIs which exhibit a significant correlation between the centrality indices (in particular authority and hubness) and the AQ score. The p-value is computed by transforming the correlation to create a t -statistic having N-2 degrees of freedom, where N is the number of data points.

In the case of the sparse matrix, for each centrality index, we identified the ROIs that exhibited a significant difference between the two groups. ROI's significance was defined as a Bonferroni-corrected p-value less or equal to 0.05 where the p-value was obtained via Monte-Carlo testing.

Then, both in case of the complete and sparse matrix, once the significant ROIs were identified for each index, the connectivity differences between the Low and High AQ Score Group were plotted for the significant ROIs only, separately for each index (in particular in case of the *authority* index and *hubness* index); this serves to evidence differences between the two groups in the pattern of connections entering into *authority* nodes and exiting from *hub* nodes.

4.3.3 Results

4.3.3.1 Analysis on the complete connectivity matrix

4.3.3.1.1 Lobes' Analysis

Using the complete connectivity matrix, the connection difference between the two groups does not reach a significativity level. Hence the following results can only be considered just as a preliminary exploratory analysis, and connection differences can be only regarded as indicative of a main flow pattern in the two groups. The results are illustrated in Fig. 4.21, where we show only the connection differences with $|t| > 1$ (which corresponds to a $p < 0.15$ in the case of a one-tailed student t-test). Higher blue lines denote connectivity higher in the Low AQ score Group (left panel), and red lines connectivity higher in the High AQ score Group (right panel). Results show that left to right connections (i.e., entering into the right temporal lobe) were higher in the Low AQ score Group; conversely, connectivity was mainly bottom-up (i.e., entering into the frontal lobes) in the High AQ Score Group.

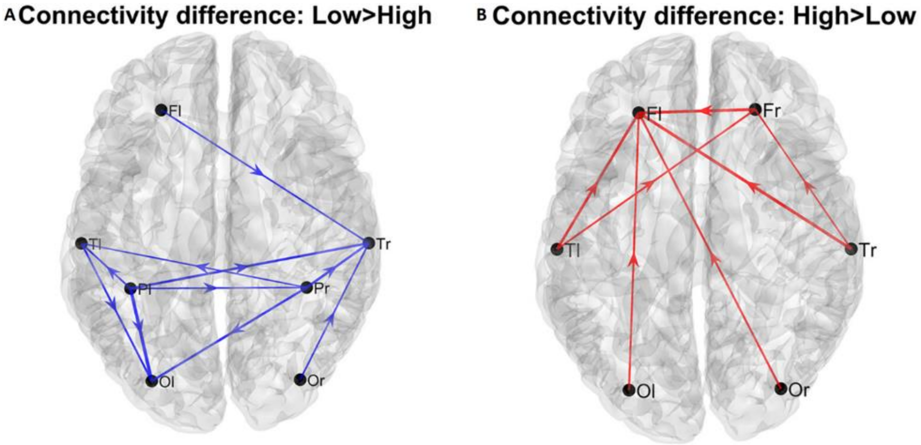

A Connectivity difference: Low>High
Fl
Tl
Tr
Pl
Pr
Ol
Or
B Connectivity difference: High>Low
Fl
Fr
Tl
Tr
Ol
Or

Figure 4.21 – Patterns of the main connection differences linking the four lobes (Frontal left and right, Fl and Fr, Temporal left and right, Tl and Tr, Parietal left and right, Pl and Pr, Occipital left and right, Ol and Or). The left panel (A) describes connections differences which are higher in the Low AQ score Group, while the right panel (B) describes connections differences which are higher in the High AQ score Group. Only connections differences with $|t|$> 1 (student t-test) are plotted.

4.3.3.1.2 Analysis on the individual ROIs

For what concerns authority, seven regions (EN r, IST l, IST r, LO r, PH r, ST r, and SMG l) exhibited a significant correlation between the AQ score and authority (see Fig. 4.22, upper panels). It is worth-noting that, in all these ROIs, correlation was negative signifying that authority increased in subjects with smaller autistic traits.

For what concerns hubness, only two regions (PCL l and ST r) exhibited a significant correlation between the AQ score and hubness; in both cases, the correlation was positive, signifying that hubness increased with the autistic traits (see Fig. 4.22 bottom panels).

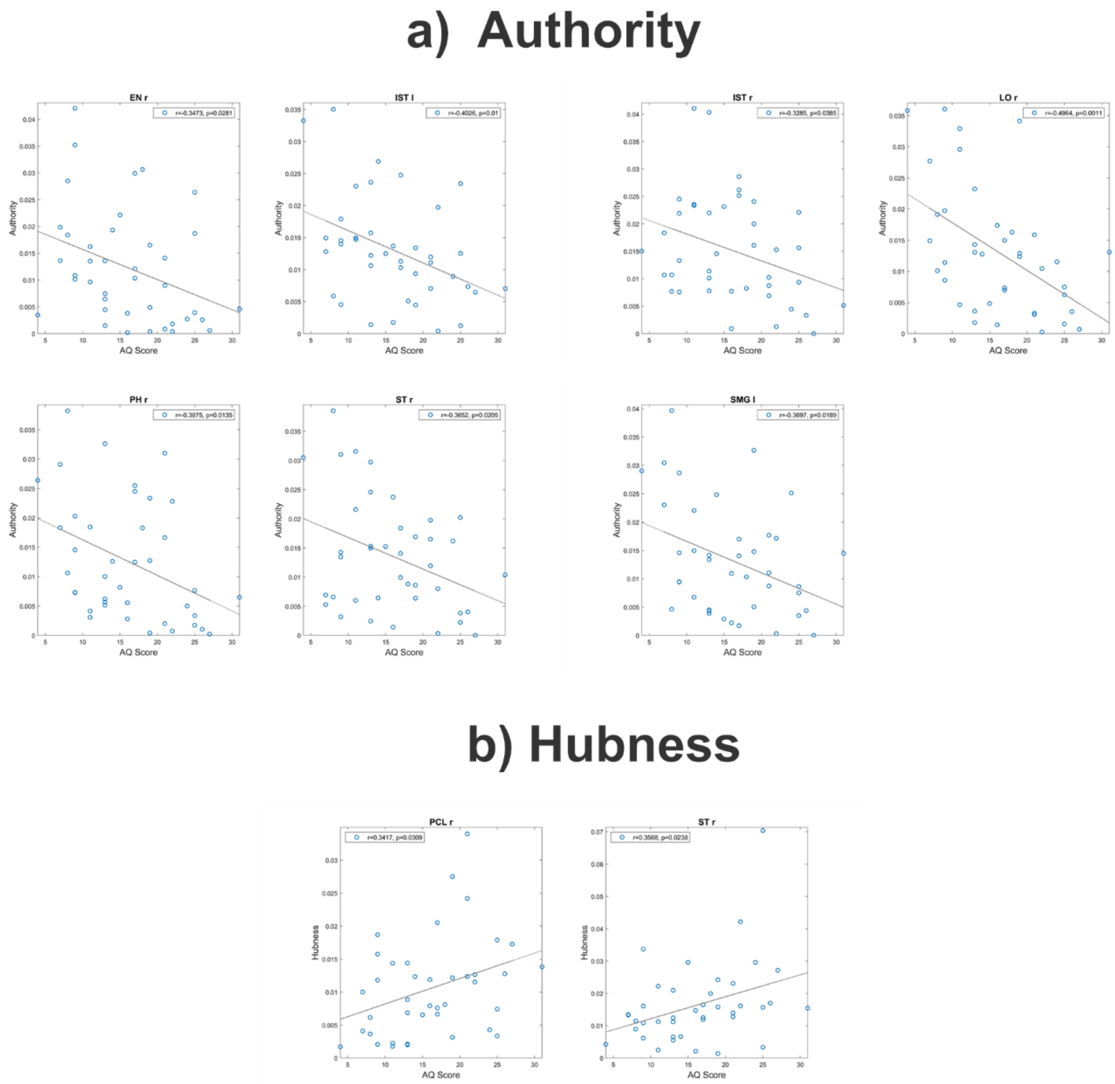


Figure 4.22 – Correlation between the authority and the AQ score [upper panel (**a**)] and correlation between the hubness and the autistic score [bottom panel (**b**)] for all ROIs which exhibit a significant p-value (uncorrected) for the correlation. These correlations have been computed on the complete normalized connectivity matrix. It is worth noting that the correlation is negative for the authority, denoting a more significant input flow in the Low AQ score Group, while correlation is positive for the hubness, denoting a more significant output flow for the High AQ score Group.

The left panel in Fig. 4.23 shows the main connections differences entering into the seven regions (EN r, IST l, IST r, LO r, PH r, ST r, and SMG l) whose authority was significantly correlated with the AQ score. The right panel shows the main connection differences exiting from the two regions (PCL l and ST r) whose hubness was significantly correlated with AQ score. Blue lines denote higher connectivity for the Low AQ score Group, red lines higher connectivity for the High AQ score Group. Since we are working with a complete connection matrix, only connection differences above a given threshold (threshold = 0.015) are plotted to simplify the figure. In particular, since all connectivity matrices are normalized to 100, and we have a total number of 68_67 connections, the average value of each connection is 0.021. The previous threshold approximately corresponds to the difference between one connection increased 33% above the mean value, and another connection reduced by 33% below the mean value (i.e., 66% of the mean). The Figure shows that the majority of connections entering the authority regions were stronger in the Low AQ score Group (as expected from the previous analysis), and these connections were mainly top-down in type (especially entering into the LO r) and left to right (especially entering into the EN r and the ST r). Conversely, the majority of connections exiting from the two hubs, PCL l and ST r, were stronger in the High AQ score Group (as expected from the previous analysis), with a bottom-up connectivity, especially emerging from the PCL l, and right to left from ST r. These results are coherent with those at lobe level displayed in Fig. 4.21.

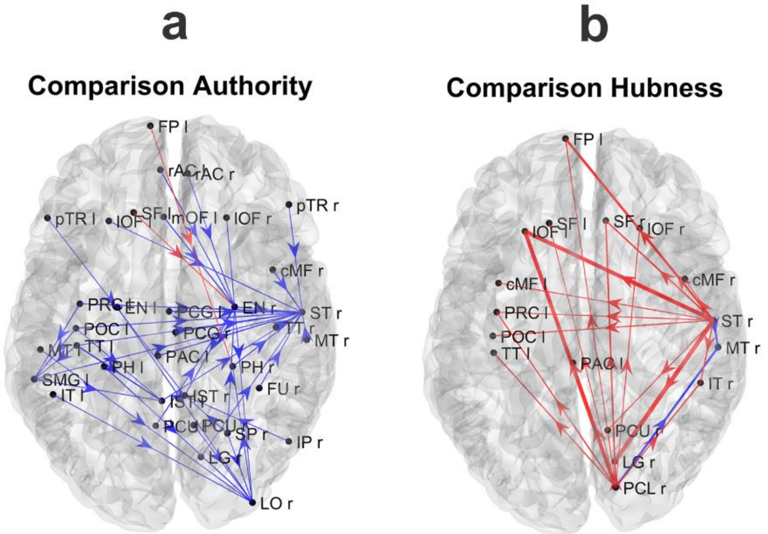


Figure 4.23 – Patterns of the main connection difference which exit from the ROIs with a significant correlation between authority and the AQ score [left panel (**a**)] and which enters into the ROIs with a significant correlation between the hubness and the AQ score [right panel (**b**)]. Blue lines denote correlation differences that are higher in the Low AQ score group, and red lines connections which are higher in the High AQ score group. Only connection differences higher than 0.015 on the complete connectivity matrix have been plotted. Three levels of thickness are adopted, with a larger thickness indicating a larger connectivity difference.

4.3.3.2 Analysis on the sparse connectivity matrix

The previous analysis, accomplished on the overall normalized connectivity matrix, pointed out the presence of some authority nodes especially involved in top-down connectivity for the low-autistic trait population, and some hubness nodes characterized by bottom-up connectivity for the high-autistic trait population. The difficulty in the use of a complete connectivity matrix, however, derives from the presence of many connections with no clear statistical difference between the two groups. This is reflected in the poor statistical significance of the connection difference and, for what concerns the correlation, in a p value that, although significant, cannot survive the statistical correction. This means that the previous results can be considered as a mere hypothesis generated from data, requiring further more complete validation.

For this reason, in order to better unmask differences, in the following a different analysis is presented, by focusing attention only on the connections which exhibited a significant statistical difference in the two groups. Hence, as described in the Method section, we consider sparse connectivity matrices. This kind of analysis has the benefit of revealing a greater number of regions with statistical differences in connection flow.

4.3.3.2.1 Lobes' Analysis

Fig. 4.24 shows the centrality indices (*in degree*, *out degree*, *authority*, *hubness*) computed at the level of the four lobes (frontal, parietal, temporal and occipital) from the sparse matrix. The asterisks denote statistically significant differences between the two groups. As it is evident from the left panels, High AQ score individuals exhibited a statistically significant increase in the connections entering into the frontal regions, and this difference was even more marked if *authority* was used as a centrality measure instead of the *in degree*. Conversely, Low AQ score individuals exhibited more significant connections entering into the temporal regions; even in this case, the significance increased if the *authority* measure was used. For what concerns the connections emerging from regions (right panels), High AQ score individuals exhibited more significant connections emerging from the occipital regions, whereas Low AQ score individuals showed a higher significance in the parietal regions. For both emerging connection outcomes, the significance was more evident if *hubness*, instead of the *out degree* measure, was used.

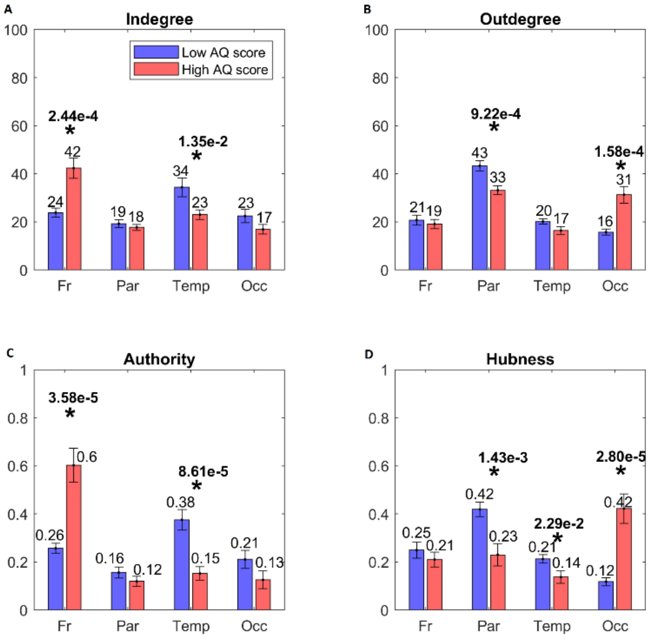


Figure 4.24 – Bar plots representing the centrality indices [in degree: panel (**a**), out degree: panel (**b**), authority: panel (**c**), hubness: panel (**d**)] for the four lobes of the brain, i.e., Frontal (Fr), Parietal (Par), Temporal (Temp), and Occipital (Occ) in each group of participants (red bars for the High AQ score Group, blue bars for the Low AQ score Group). Each bar shows the index value (mean ± SEM) for the specific area in the specific group of participants. As per definition, the sum of the authority values and the sum of the hubness values across all areas provide a total of 1, while the sum of the in degree values and the sum of out degree values across all areas is

equal to 100. The asterisks indicate the presence of a statistically significant difference between the two groups ($p < 0.05$, uncorrected).

In order to further investigate the results arising from the above histograms, Fig. 4.25 represents the statistically significant connections (i.e., those which exhibited significant differences between the two groups) linking the eight lobes of the brain; in this case, the homologous regions in the left and right hemisphere were considered separately. The upper panel displays the p value of all significant connections using a color scale, while the bottom panel shows the connection differences (in red the connections which were significantly stronger in High AQ score individuals, in blue the connections significantly stronger in Low AQ score individuals). The results confirm those reported in Fig. 4.24, showing that, in the High AQ score Group, significantly stronger connections were mainly directed from the occipital toward the frontal regions. The pattern in the Low AQ score Group showed significantly stronger connections emerging from the left parietal lobe, directed toward the right parietal, left temporal and left occipital regions.

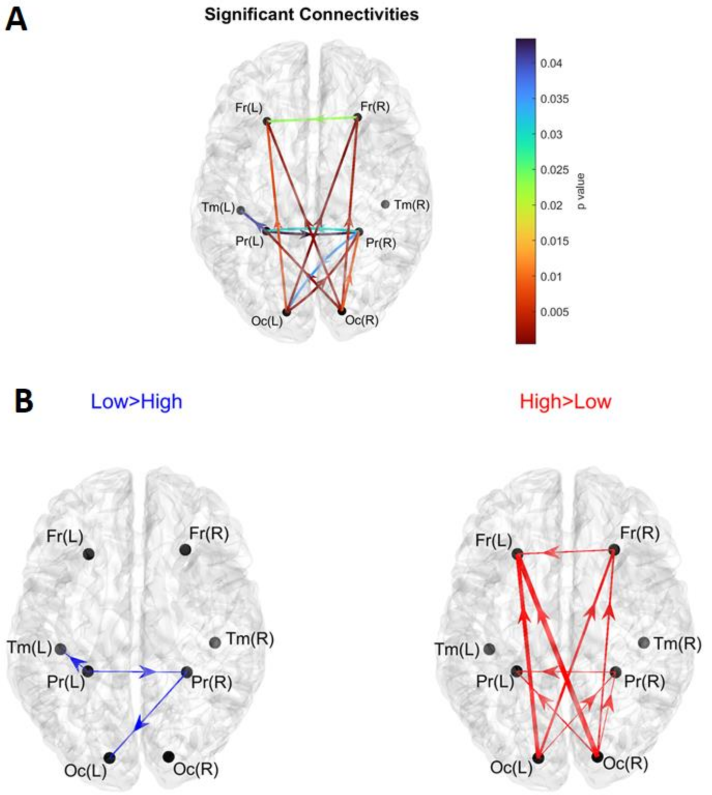


Figure 4.25 – Representation of the connections linking the eight lobes of the brain, Frontal (F), Parietal (P), Temporal (T), and Occipital (O), considering separately the right (r) and left (l) hemispheres. Only the connections that exhibited a statistically significant difference between the two groups ($p < 0.05$, uncorrected) are represented. The upper panel (A) shows the p-values of the significantly different connections. The lower panels (B) represent the differences in connectivity strength: the blue diagram (Low > High) shows the connection differences for those connections that resulted significantly stronger in the Low AQ score Group compared to the High AQ score Group; the red diagram (High > Low) shows the connection differences for those connections that resulted significantly stronger in the High AQ score Group compared to the Low AQ score Group. The thickness of each link varies according to the value of the connection difference. Three levels of thickness are adopted, with a larger thickness indicating a larger connectivity difference.

4.3.3.2.2 Analysis of the individual ROIs

Fig. 4.26 shows the positions of the ROIs which exhibited a significant difference (Bonferroni corrected) in the *in degree* (upper panels) and in the *authority* (bottom panels) indices between the two groups. The right upper panel evidences that in the High AQ score Group the *in degree* index was significantly higher (compared to the other group) especially in the frontal ROIs. This pattern was even more evident when *authority* index was used (bottom right panel). Conversely, the Low AQ score Group did not exhibit any appreciable increase in the *in degree* index, while some regions in the temporal, parietal and frontal lobes exhibited an increased *authority* without a clear topological organization.

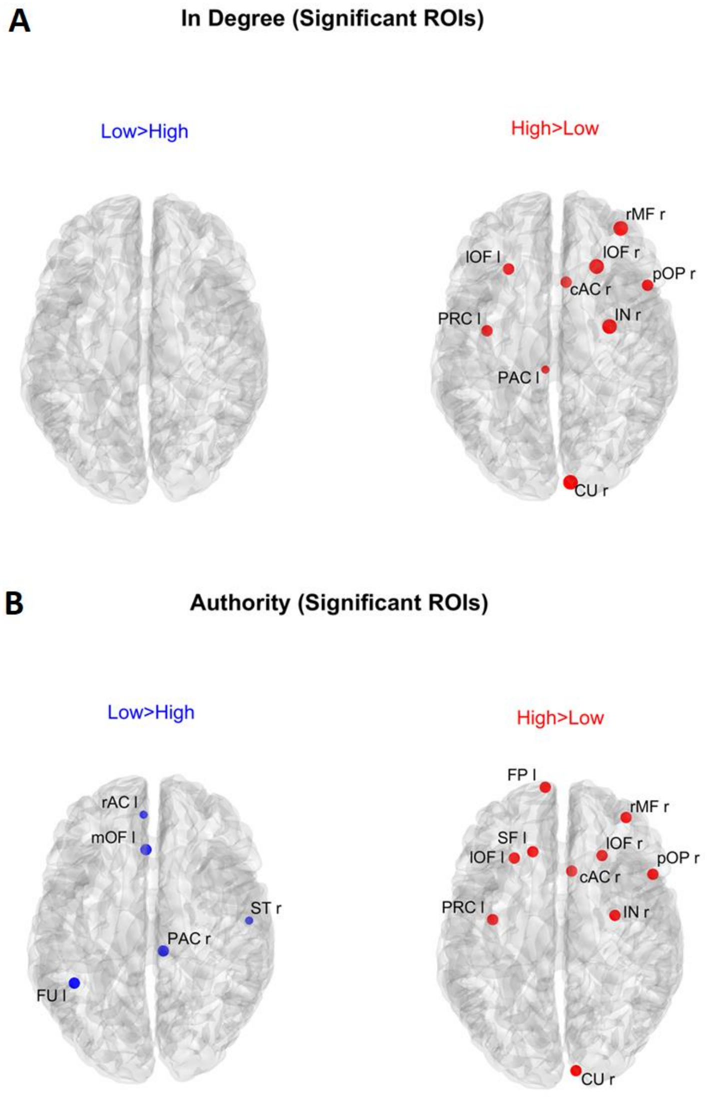


Figure 4.26 - Positions of the ROIs which exhibited a significant difference in the in degree index [upper panels (A)] or in the authority index [lower panels (B)] between the two groups (p-value < 0.05, Bonferroni corrected). The left panels in blue (Low > High) display the ROIs having significantly higher centrality index in the Low AQ score Group compared to the High AQ score Group. The right panels in red (High > Low) display the ROIs having significantly higher centrality index in the High AQ score Group compared to the Low AQ score Group. The significant ROIs are shown as simple dots and represent regions to which important information enters. Three levels of dots' size have been adopted: the larger the dot size, the more significant the centrality difference. For

the panels where no dots appear over the brain map (i.e., in degree for Low > High), the constraint of significance was not satisfied by any of the 68 ROIs.

In order to gain a deeper understanding of the previous patterns (limited to *authority* only), Fig. 4.27 shows the connection differences entering into all ROIs with significantly higher *authority* in either group. In the High AQ score Group, these connections mainly linked the two occipital regions PCL (right and left) toward frontal regions: particularly evident were the connections entering the two lOF (left and right), and the right rMF. Thus, a clear bottom-up pattern of connections emerged, supporting the results in Fig. 4.25. Conversely, in Low AQ score individuals the pattern of connections entering into nodes with higher *authority* was less structured, showing connections directed to frontal (PAC r), right temporal (ST r) and left temporal (FU l) regions.

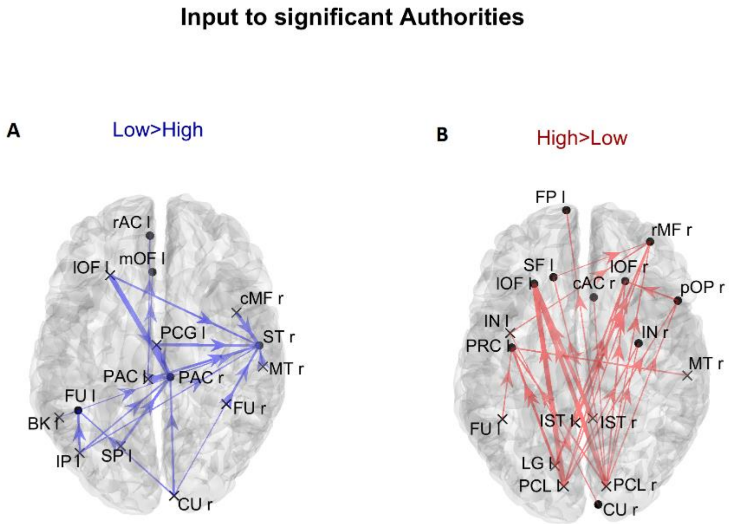


Figure 4.27 – Representation of the connection differences entering into the ROIs which exhibited significant differences of authority between the two groups. The left panel in blue [Low > High, panel (**A**)] displays the connection differences entering into the "Low > High" authority ROIs (the ROIs shown in the left lower panel in Figure 4.26), for connections higher in the Low compared to the High AQ score Group. The right panel in red [High > Low, panel (**B**)] displays the connection differences entering into the "High > Low" authority ROIs (the ROIs shown in the right lower panel in Figure 4.26), for connections higher in the High compared to the Low AQ score Group. The plotted connections run from a generic output ROI (marked with a cross) toward the ROIs with significantly different authorities (marked with a dot). The thickness of each link varies according to the value of the connection difference. Three levels of thickness are adopted, with a higher thickness indicating a larger connectivity difference.

Fig. 4.28 shows the positions of the cortical ROIs that exhibited a significant difference (Bonferroni corrected) in the *out degree* (upper panels) and *hubness* (bottom panels) indices between the two groups. As shown in the right panels, in the High AQ score Group, both the above-mentioned centrality measures were significantly higher (compared to the other group) in the occipital PCL regions of both hemispheres and in the occipital left LG region. Moreover, some frontal regions also exhibited increased *hubness*, a result apparently in contradiction with previous figures. However, as will be clarified when discussing Fig. 4.29 below, connections originating from these hubs were less significant than those originating from the occipital regions. The Low AQ score Group exhibited an appreciable increase in the *hubness* of parietal and temporal regions, especially in the left hemisphere, whereas no significant increase emerged from the *out degree* index. It is interesting to note that also an occipital region (the CU right) exhibited an increased *hubness* in the Low AQ score Group.

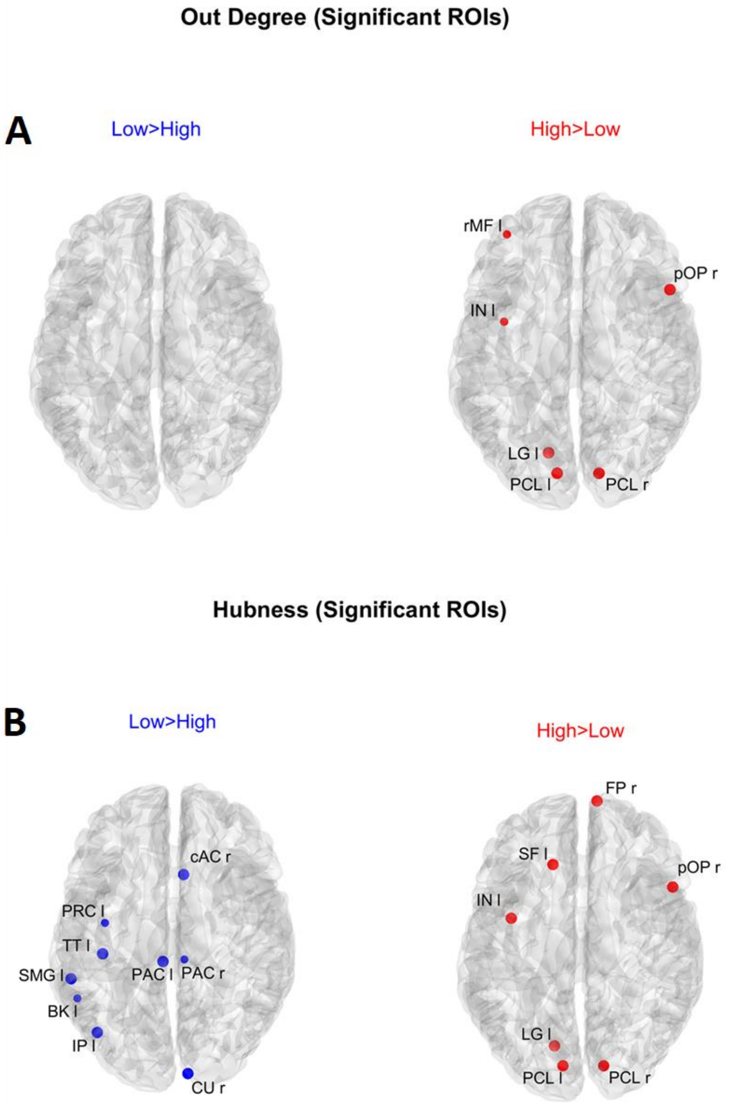


Figure 4.28 Positions of the ROIs which exhibited a significant difference in the out degree index [upper panels (**A**)] or in the hubness index [lower panels (**B**)] between the two groups (p-value < 0.05, Bonferroni corrected). The left panels in blue (Low > High) display the ROIs having significantly higher centrality index in the Low AQ score Group compared to the High AQ score Group. The right panels in red (High > Low) display the ROIs having significantly higher centrality index in the High AQ score Group compared to the Low AQ score Group. The significant ROIs are shown as simple dots and represent regions from which important information originates. Three levels of dots' size have been adopted: the larger the dot size, the more significant the centrality difference. For the panels where no dots appear over the brain map (i.e., out degree for Low > High), the constraint of significance was not satisfied by any of the 68 ROIs.

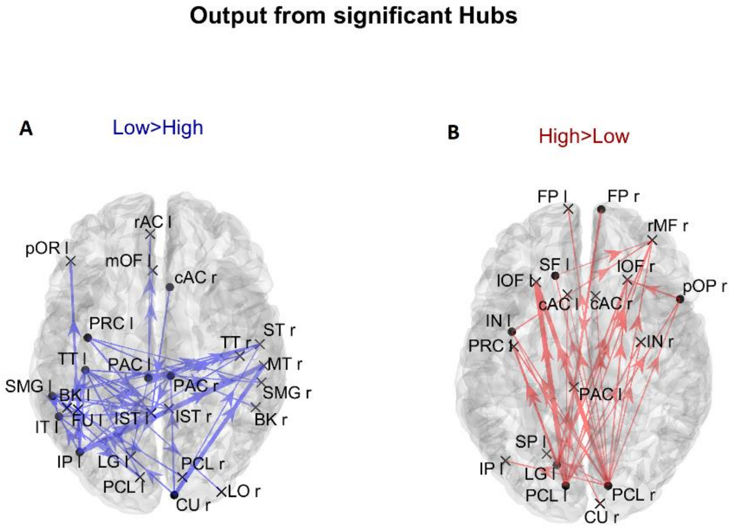


Figure 4.29 – Representation of the connection differences exiting from the ROIs which exhibited significant differences of hubness between the two groups. The left panel in blue [Low > High (**A**)] displays the connection differences exiting from the "Low > High" hubness ROIs (the ROIs shown in the left lower panel in Figure 4.28), for connections higher in the Low compared to the High AQ score Group. The right panel in red [High > Low (**B**)] displays the connection differences exiting from the "High > Low" hubness ROIs (the ROIs shown in the right lower panel in Figure 4.28), for connections higher in the High compared to the Low AQ score Group. The plotted connections run from the ROIs with significant hubness (marked with a dot) toward generic input ROIs (marked with a cross). The thickness of each link varies according to the value of the connection difference. Three levels of thickness are adopted, with a higher thickness indicating a larger connectivity difference.

The results illustrated in Fig. 4.28 are further clarified in Fig. 4.29, which shows the connection differences exiting from the nodes with significant higher *hubness* in either group. Once again, a clear bottom-up pattern is evident in the High AQ score Group. It is worth noting that, in this group of individuals, the fronto-parietal regions with increased *hubness* (i.e., the SF l, FP r, pOP r and IN l) generated only weak output connections (when compared to the other group). These were sufficient to make the *hubness* of these ROIs significantly higher, without altering the general bottom-up pattern of the overall circuitry. In fact, much stronger connections exited from the two PCL regions, defining a clear bottom-up trend. The pattern of connections originating from significant hubs in the Low AQ score Group were mainly directed from temporal and parietal left regions to the right ones, with some connections also directed downwards to the occipital nodes. As anticipated above, also the right CU exhibited a clear bottom-up function in this group, while, in agreement with Fig. 4.27, the right temporal regions received most of the significant connectivity originating from the hubs. It is worth noting that connections toward frontal regions were less significant in this group.

4.3.4 Discussion

The present paper analyzes the differences in brain connectivity between two groups of non-clinical individuals who differ in the degree of autistic traits (low vs. high), as classified based on the Autistic Quotient [487] score. Results have two main important aspects of interest.

First, we confirm that autistic traits can be observed within a wide spectrum encompassing both clinical and non-clinical populations. Specifically, the degree of autistic traits clearly differs in the non-clinical population between low and high AQ scores. Second, we show that these differences can be quantified as alterations in brain connectivity. In particular, we show that Granger Causality, computed from neuroelectric signals reconstructed in the cortex [499,502,503], together with indices taken from the Graph Theory [489–491], can represent a valuable tool to characterize differences in brain networks and deepen our analysis of the neurobiological bases of brain disorders. Further, we confirm a previous hypothesis [439,484] that individuals with higher autistic traits are characterized by more evident bottom-up mechanisms for processing sensory information.

A critical point may be the selection of the threshold used to discriminate between the two classes. Despite the inherent arbitrariness of the choice, we used as a discriminative threshold the average AQ score obtained in a nonclinical population from the large-sample work of Ruzich et al. (2015)[493], and this seems the most natural choice. Moreover, using this value, the present population of 40 subjects is subdivided in 19 and 21 subjects, i.e., the threshold we chose is quite proximal to the median of the considered population. It is worth noting that similar approaches of partitioning the sample around a threshold have been used previously in the literature [504].

In the following, we will first analyze methodological issues, then the neurophysiological significance of the obtained results will be explored. Finally, limitations of the present study will be analyzed.

4.3.4.1 Granger Causality

In this work, we have chosen temporal Granger causality as a tool to reconstruct brain connectivity from EEG data. This measure mathematically represents the impact that knowledge of an upstream signal can have on the prediction of a downstream temporal signal. Thus, it represents a causal directed index of connectivity. Indeed, Granger Causality is widely employed in neuroscience today [499,502,503,505]. Moreover, in a recent paper, using artificial signals produced by a neurocomputational model as ground truth, we demonstrated that the Granger Causality overcame other functional connectivity estimators in terms of accuracy and reproducibility [126]. This method has evident computational advantages compared with other suitable methods (such as Transfer Entropy, see [137]).

The analysis was initially performed (see Section 4.3.3.1) on the complete normalized connectivity matrix, to show the main characteristics of the Granger flow in the two groups. Then, to improve the significance of the results, we considered only connections which exhibited a significant statistical difference between the two populations, thus working with a sparse matrix (i.e., all connections which did not show statistically significant differences between the two groups were set at zero). In other terms, the graphs in Section 4.3.3.2 do not represent the overall connectivity patterns, but rather highlight the differences between the two populations. The connectivity matrices so obtained were then used to compute some indices taken from Graph Theory.

4.3.4.2 Graph Theory

Several studies using Graph Theory in ASD have appeared in recent years: most of them suggest that ASD individuals exhibit alterations in modularity (i.e., densely connected modules that are more segregated), in global efficiency (i.e., average path length required to go from one node to another), in betweenness (the capacity of a node to connect to other nodes) or in connection density [506–510]. EEG and MEG connectivity studies using graph analysis generally report autism to be associated with sub-optimal network properties (less clustering, larger characteristic path, and architecture less typical of small-world networks) [511–517]. This, in turn, results in a less optimal balance between local specialization (segregation) and global integration [518]. Although of particular significance, we think that these indices do not consider the fundamental problem of directionality in the processing pathway and the different importance that bottom-up and top-down connectivity plays in several brain processing.

Accordingly, an essential novelty of the present study concerns the use of some specific centrality indices (*in degree*, *out degree*, and above all, *hubness* and *authority*) to characterize group differences in network directionality. The basic idea is that the directionality of the processing streams plays a major role in determining group differences (at least for what concerns autistic traits), rather than other indices like betweenness, path length, or clustering, more frequently adopted in the characterization of brain networks. In particular, by considering macro-regions and sparse connectivity matrices, these indices provided highly significant statistical differences and provided a precise scenario to distinguish the two groups.

4.3.4.3 Connectivity among macro-areas (lobes)

The connectivity analysis was performed at two levels. First, we concentrated on the connectivity among macro-regions (lobes) of the cortex, the frontal, parietal, temporal, and occipital zones, to discover the main traits of connectivity differences.

This analysis confirms the result of a previous preliminary study [439], i.e., individuals with higher autistic traits exhibit stronger outgoing connections from the occipital regions and stronger incoming connections toward frontal areas (i.e., bottom-up) compared to those observed in individuals with lower autistic traits. In addition to confirm the results of our previous study, as a new significant result of the present study we propose that two other centrality measures, i.e., *hubness* and *authority*, allow for a finer discrimination of connectivity directionality. The reason for this improvement will be critically analyzed in the next section. If these two measures are used, significant statistical differences can be observed to characterize the directionality of the connections in High AQ score vs. the Low AQ score individuals. In particular, using sparse matrices statistically significant differences were evident between the *hubness* of the occipital regions in the two classes, with much stronger *hubness* for individuals with high autistic traits. Looking at *authority*, a significant increase in the *authority* of the frontal region was observed in the group with higher autistic traits.

The same patterns were confirmed by computing (from the sparse matrices) the connectivity among the macro-regions and plotting only those which exhibited a significant statistical difference. As shown in Fig. 4.25, increased bottom-up connectivity from occipital to frontal regions was evident in individuals with high autistic traits.

4.3.4.4 Connectivity among individual ROIs

Besides connectivity analysis at lobe level, we performed connectivity analysis at single ROI level. To this aim, *centrality indices* were computed by considering all the 68 ROIs in the Desikan-Killiany atlas. It is interesting that the results obtained on the overall connectivity matrix and on the sparse matrix provide similar indications, emphasizing the presence of bottom-up connections in the high-score group and top-down connections in the low-score group. However, analysis performed on the overall connectivity matrix did not reach a significant level, whereas a greater significance was obtained from sparse matrices. For this reason, in the following we will mainly refer to the results of sparse matrix.

An important result of our study is that hubness and authority provided more significant differences compared with in degree and out degree, respectively; hence we suggest that these indices should be used to characterize the flow in a network of multiple ROIs. In particular, by comparing in degree vs. authority in Fig. 4.26 we can observe that the results are quite similar for what concerns the High AQ score Group (authority produces just one more significant frontal node compared with in degree), whereas significant differences can be observed in the Low AQ score Group (no significant node is evident if in degree is used, compared with five nodes using authority). Consequently, authority allowed the detection of a clear left to right connectivity in the Low AQ score Group. Similarly, only moderate differences can be observed using hubness vs. out degree in the High AQ score Group (Fig. 4.28, hubness detects two additional regions in the frontal cortex, allowing a better analysis of top-down influences). Also in this case, hubness provided a significant improvement compared with the out degree in the Low AQ score Group (nine significant ROIs are detected by hubness, mainly located in left and medial parietal and temporal regions, vs. no significant region by the out-degree). These differences suggest that the overall graph is more complex in the Low AQ score Group compared with the High AQ score one, requiring more sophisticate indices for detecting the flow of transmitted information.

To understand why authority and hubness are more powerful compared with in degree and out degree, we remind that authority not only takes into account the number and strength of the connections entering a node but also weights these connections by the *hubness* of the upstream nodes. Similarly, *hubness* does not only take into account the number and strength of connections exiting from a node but also weights these connections by the *authority* of the downstream nodes. Of course, these measures need to be computed together via recursive formulas, as illustrated in Eqs. 2.37-2.38. Briefly, the importance of the information exiting from a node (or the importance of the information entering into a node) is not simply the sum of its output connections (or the sum of the input connections), but also depends on the role played by the sending nodes (or by the receiving nodes). For instance, a connectivity of value 0.04 reaching an almost completely isolated node (one which does not send information to others nodes in the network) can be scarcely important compared with a connection of value 0.02, which reaches a crucial node. Hubness is able to quantify this difference compared with a simple sum of outgoing connectivity. Similarly, authority is more able to summarize the effective significance of the incoming flow compared with the simple sum of entering connections.

Using these indices, we then mapped the stronger connections that exited from ROIs with higher *hubness* and entered into the ROIs with greater *authority*. These results computed on each ROI extend the lobe analysis to several aspects: i) The main hubs for High AQ score individuals were located in the left and right PCL regions. A pattern of bottom-up connections emerging from these two regions seems to be the dominant feature that characterizes this group. Left and right PCL are the ROIs in which the primary visual cortex is located. These areas handle the transmission of incoming visual inputs from the thalamus to higher-order processing regions. The enhanced bottom-up signaling arising from this site resembles the pattern observed in individuals with clinical form of autism characterized by hyper-engagement of sensory regions [519,520] that could underpin the sensory and visuospatial peculiarities typically observed in ASD [482,521]. ii) The leading *authorities* for High AQ score individuals were located in the frontal and prefrontal regions, particularly in the left and right lOF. These two ROIs encapsulate frontal sites involved in high-level mechanisms such as emotional regulation, decision-making and social cognition [421]. Crucially, these domains tend to be altered in ASD individuals. Excessive information inflow in brain areas related to emotional and social processing could be implicated in the difficulty to manage complex and multifaceted social interactions typically observed in this spectrum. This could also explain why ASD individuals tend to prefer less social-demanding environments as they are linked to a lower risk of over-stimulation. iii) The previous connections were distributed bilaterally, from both PCLs to both homolateral and contralateral frontal hemispheres. iv) Conversely, the pattern of connectivity in Low AQ score individuals exhibited a broader and less defined distribution, involving several connections in the temporal, parietal, and occipital lobes, with hubs mainly located in the left hemisphere and a direction from left to right. This suggests that the pattern of inter-areas communication in low-AQ individuals is more distributed and varied and not rigidly channeled into narrow pathways.

We remind, however, that these connectivity patterns reflect *differences* between the two groups, hence a relative role in one population vs. the other, not the absolute impact that connections have on the overall brain network. In other words, it is possible that some strong connections did not appear in our graph since they were equally relevant in both populations, hence without significant difference (this is the reason why the overall connectivity matrix provides less significant results). Moreover, we remind that trials were performed at rest. Thus, the examined connectivity reflects differences in a resting state.

In general, the present results support the findings obtained in our previous study on a smaller population [439], even though the exact position of the ROIs representing the increased bottom-up connectivity is not identical. In our previous study, we observed increased connectivity from the right PCL and the left LG (instead of the left PCL as found here). Still, these differences can be explained based on minor variances in source reconstruction and grouping among proximal voxels. Moreover, in our previous study, the bottom-up connectivity in High-AQ score individuals was especially evident in the right hemisphere (particularly toward the right MFr, a region that still plays e significant role among the authorities in the present study). In contrast, this connectivity seems to be more bilaterally distributed in the current results.

These results support the idea that the brain network in individuals with higher autistic traits vs. individuals with lower autistic traits is not characterized by a general reduction in connectivity (as hypothesized in some theorizations) but rather that mixed patterns of under- and over-connectivity can be appreciated. Over-connectivity is evident in the fronto-posterior axis, involving bottom-up influences, whereas hypoconnectivity involves many tempo-parietal regions, especially in the left hemisphere.

4.3.4.5 Neurophysiological meaning

Several hypotheses on brain connectivity in ASD have been formulated in past years, with apparently contradictory outcomes: while some authors hypothesized more robust connectivity in ASD, others reported reduced connectivity (see Introduction). These contradictions, however, can be reconciled by thinking that differences between controls and individuals within the autistic spectrum can especially reflect a directionality in the connections rather than the number and total strength of edges in the overall network. Furthermore, a mixed pattern of increased connectivity among some regions and decreased among others probably characterizes the autistic brain. Directionality in the connectivity patterns, in turn, may reflect a hierarchical organization of the processing stream, with bottom-up connections (especially from the occipital towards the frontal lobes) involved in sensory processing and top-down connections reflecting context modulation, and prior knowledge, planning, and attention. This connectivity organization agrees with the so-called predictive coding theory, which assumes that environmental and internal signals are joined together to form a unified model of reality. In particular, the predictive coding theory of ASD [478,484] hypothesizes that ASD people do not form accurate predictions of the external environment since sensory information supersedes the internal expectation. Our results support this theory, showing that differences in bottom-up connectivity (hence, in the impact that sensory input can have on the global internal model) are stronger in individuals with higher autistic traits, even within a population of healthy individuals.

4.3.4.6 *Limitations of the present study*

A limitation of the present study may be the limited sample size (19 vs 21 participants). Actually, this number is in line with (and in many cases higher than) the sample employed in published works that use similar experimental procedures and investigate similar phenomena (see [522,523]). However, the complexity of the analysis performed and, in particular, the study accomplished on the complete connectivity matrix, reveal the necessity of a larger number of participants to achieve statistically more solid results. Hence, future studies on a large cohort can allow a more detailed comprehension of the problem.

In this study, we did not include participants with a diagnosis of ASD, hence we cannot be confident that the present results would stand up also in a clinical population. However, the results obtained go exactly in the direction hypothesized by theoretical and empirical work on connectivity features in clinical ASD. Moreover, substantial behavioral [504], genetic [488] and neural [524] evidence suggests that ASD is a continuum of conditions ranging from trait-like expression to the diagnosed clinical form of autism. Of course, additional studies on a clinical population are required to definitely support the present initial results and definitely validate the hypothesis of a continuous spectrum ranging from normality to ASD.

An interesting point concerns the relationship between the Granger connectivity, evaluated in this study, and the structural connectivity (i.e., the physical traits that connect brain regions, generally estimated by diffusion-weighted imaging). Some studies (e.g.,[525]) have shown that there is significant overlap between neuroanatomical connections and correlations of functional brain signals. Conversely, other recent studies of our group, using neural mass models as a ground-truth, showed that in some conditions the two aspects may differ, as a consequence of non-linear phenomena [126,137,456]. Hence, it is still unclear how the brain network interacts during specific tasks or at rest, accounting for all structural and functional aspects in terms of causality, given the many nonlinear dynamics that characterize brain functioning. Moreover, the present results show some connections crossing the midline. Regarding this point, although the connections traveling through the corpus callosum typically connect homotypic areas, a substantial number of traits connecting heterotypic areas in the two cerebral hemispheres have been observed (e.g. [526]). Of course, without structural data, it remains difficult for the current study to formulate more precise hypotheses about this issue.

Finally, in the present study we have observed differences in bottom-up and top-down connectivity in the two groups. Works in the literature emphasize that these connections can be implicated in sensory processing, especially in multisensory conditions [527] or after sensory deprivation [528]. Furthermore, several studies suggest that atypical sensory processing is a common characteristic of ASD and that sensory traits have important implications in the developmental phase of this pathology [529,530]. The present experiments were performed in a resting condition, so it would be difficult to make strong inferences about sensory processing from the current data. Further studies, examining the response to sensory stimuli, are required to test whether these neural signatures of autistic traits (more bottom-up processing in high AQ score, more top-down processing in low AQ score) have an impact at the behavioral level, for example to explain the observed differences in sensory profile.

5 Discussion

Detecting changes in functional connectivity (FC) is crucial for improving the comprehension of brain functioning in both healthy and pathological individuals. However, despite its relevance, it still poses one of the greatest challenges in computational neuroscience. Indeed, a complete understanding of the reliability of the wide range of existing connectivity estimation techniques is still lacking. The first part of this PhD work aimed to fill this gap by testing the performance of the main FC measures under controlled conditions. Then, once these aspects had been investigated, the second objective of the thesis was to estimate brain connectivity on experimental electroencephalographic (EEG) data.

The concept of functional connectivity was critically examined, using biologically inspired Neural Mass Models (NMMs), which simulates EEG oscillations in different frequency bands and can be interconnected, generating a ground-truth network of cortical regions. Such models are simplifications of the real neurophysiology and may not capture all the complexity and variability of the underlying neural dynamics. However, they can provide a realistic representation of the complex brain functioning, allowing the simulation of brain networks, as well as the assessment of the performances of different functional connectivity estimators under controlled conditions and without the confounding factors that may arise from empirical data. Therefore, NMMs were employed to test the reliability of FC measures in capturing the connectivity strength imposed by the model.

A first important aspect emerged by comparing the performances of eight different estimators, namely *Pearson correlation coefficient*, *Delayed correlation coefficient, Coherence, Lagged Coherence, Phase Synchronization, Temporal Granger Causality, Spectral Granger Causality and Transfer Entropy*. The outcome of the analysis, conducted under linear conditions, showed that Temporal and Spectral Granger Causality outperforms the other FC measures, followed by Transfer Entropy, proving to be the most reliable estimation method.

Furthermore, by testing Granger causality and Transfer Entropy under linear and non-linear conditions, an innovative concept emerged that needs to be underlined. Since functional connectivity estimators reflect the exchange of information between Regions of Interest (ROIs), the results obtained from such measurements are influenced by the working conditions of the cortical regions in the network. Indeed, while under linear conditions FC measurements reflect the true network connectivity, under nonlinear conditions they are affected by the activation/deactivation effect of cortical regions driven by external inputs, failing to capture the true connectivity strength imposed by the model. Therefore, when using FC estimators, one should bear in mind the meaning of these metrics, which reflect a change in information transmission rather than a real change in synaptic strength as often reported in the literature.

The significance of external influences and non-linear phenomena was also emphasised by simulating a motor network of a stroke patient during: a) rest, b) movement of the affected hand and c) movement of the unaffected hand. In this case, task-dependent changes were achieved by using a single set of fixed connectivity parameters, only by varying working point of the ROIs on the sigmoidal relationship (by varying the external input to the ROIs).

Since Granger causality in both temporal and spectral domains proved to be the most reliable method, it was chosen to estimate the directed functional connectivity on experimental EEG data.

Indeed, in the second part of this PhD work, advanced EEG processing techniques have been employed in order to reconstruct cortical sources and shed light on: 1) task-dependent changes in brain connectivity within the same population, 2) resting-state network alterations between two different populations.

The first aspect (1) was investigated in two different datasets: an internal-external attention competition task, and a Pavlovian fear conditioning and reversal experiment. In both studies, task-dependent differences have been investigated by highlighting the role of brain rhythms. Indeed, power and spectral Granger connectivity changes have been assessed focusing on the most significant frequency bands, that showed to play a functional role in the aforementioned tasks. In both studies, results show that theta and alpha brain rhythms provide crucial information on the regulatory mechanisms underlying the task-related cognitive processes.

Indeed, in the internal-external attention competition task results showed that frontal midline theta is distinctive of mental task execution and is more prominent during attentional competition, which may reflect higher executive control; moreover, anterior cingulate cortex showed an increased theta-band connectivity with distant regions, such as temporal and occipital. Whereas, alpha power in visual brain regions strongly decreased in external attention alone, while it assumed values close to rest during attentional competition, reflecting reduced visual engagement against distractors; alpha-band connectivity results suggested that bidirectional connections between frontal and visual regions could contribute to reduce visual interference in internal attention.

In the Pavlovian fear acquisition and reversal protocol, results showed that increased theta rhythm in the cingulate cortex probably subserve fear acquisition, and is transmitted to other cortical regions via increased functional connectivity, allowing a fast theta synchronization. Whereas, alpha power showed a decrease that may represent a partial activation of motor and somatosensory areas contralateral to the shock side in the presence of a dangerous stimulus. Furthermore, connectivity changes that appeared in both frequency bands at the end of reversal may reflect long-term alterations in synapses, necessary to reverse the previously acquired contingencies.

Concerning the second aspect (b), the analysis of brain connectivity alterations in resting-state were investigated in a non-clinical population with different autistic traits. In this case, temporal Granger Causality was computed and directional indices of centrality (*out dregree, in degree, hubness* and *authority*) from Graph Theory were extracted to characterize the network. Results evidenced that individuals with higher autistic traits are characterized by a more robust bottom-up mechanisms, typical of sensory information processing, compared to subjects with lower autistic traits. This imbalance may contribute to some behavioral alterations observed in Autism Spectrum Disorders (ASD).

Overall, directed functional connectivity measures, together with brain network indices, represent valuable neuromarkers to characterize brain functioning in task execution and to discriminate between different classes of individuals. Remarkably, these methods can provide deeper insight into brain functioning in healthy and pathological conditions, and may be used, in perspective, to design future innovative diagnostic methods as well as therapeutic interventions.

As described above, the present research brings some innovations and improvements in the field of systems neuroscience. However, it also suffers from some limitations that can be addressed in the near future.

First, autoregressive (AR) model techniques, such as Granger Causality, requires estimation of the optimal model order - i.e., the number of past observations (time-steps). Early approaches often set the model order based on prior knowledge or in ad hoc ways. The use of different orders may result in different conclusions, further complicating the interpretation of Granger causality. If the order is too low, Granger causal connections at longer lags will be missed, while overfitting may occur if the order is too high. Hence, regularization-based approaches are often used for estimating the optimal model order from data[118].

Second, the Granger Causality formulation employed in this thesis makes use of a bivariate (BVAR) approach. In case of multivariate networks like the brain this may lead to some spurious connections. In contrast, multivariate (MVAR) approaches consider information of all brain regions when estimating the interaction term, thus enabling the distinction between direct and indirect interactions. This is significant since elements in a multivariate system may function cooperatively or competitively, or interact generally in a more complex fashion than traditional bivariate analysis can capture. However, if there are dependencies on unknown (exogenous) or unrecorded (latent) variables, then it will in general be impossible to entirely delate spurious effect on causal inference.

Third, as revealed by studies with NMMs, Granger Causality (and Transfer Entropy) dramatically reduces its ability to capture true connectivity strength under nonlinear conditions. However, it is well-known that neural mechanisms are characterised by nonlinear behaviour. A way to overcome this limitation could be through nonlinear Dynamic Causal Modeling, which are neurophysiological models that identify the effective connectivity of a network. The latter class of models is hypothesis driven and can be formulated in terms of NMMs, defined through a set of nonlinear equations (i.e., nonlinear state-space models). These equations model the dynamics of hidden states in the nodes of a probabilistic graphical model, where conditional dependencies are parameterised in terms of directed effective connectivity. Models' parameters are then estimated by fitting the experimental data, yet raising serious optimisation problems.

Fourth, in the context of this thesis, only centrality measures of brain network topology have been taken into account that allow the directionality aspect of connectivity to be emphasised. Consequently, all measures of graph theory that describe the segregation or integration tendency of the network have been disregarded. However, aspects such as small-worldness are of great interest when describing brain networks. This feature supports efficient information segregation and integration with low energy and wiring costs, and it is well suited for complex brain dynamics[147]. Indeed, many brain network studies grounded on EEG and connectivity estimates indicates that the brain holds an optimal small-world scheme of communication during its normal and healthy functioning. Furthermore, recent studies have shown that such brain topology undergoes changes during ageing[38,144], as well as in the case of brain disorders[30], demonstrating the importance of graph theory and network analysis in the understanding of brain functioning in health and disease.

Therefore, future investigations will focus on: a) algorithm implementation for the automatic assessment of the optimal VAR order; b) BVAR and MVAR Granger Causality comparison NMMs to underline the strengths of the multivariate techniques; c) effective connectivity estimation through nonlinear DCM in order to account for neural nonlinearities; d) network analysis extension to segregation, integration and small-worldness metrics of graph theory.

Finally, although not reported in this thesis, preliminary analyses are currently ongoing in which graph theory indices, including those mentioned in (d), are being employed on EEG data not only to investigate some brain disorders of great neurological interest, such as schizotypy and epilepsy, but also to assess network topology in the presence and absence of dreams during sleep.

6 References

1. Nazarova, M. & Blagovechtchenski, E. Modern brain mapping–what do we map nowadays? *Frontiers in psychiatry* **6**, 89 (2015).

2. van den Heuvel, M. P. & Sporns, O. Network hubs in the human brain. *Trends in Cognitive Sciences* **17**, 683–696 (2013).

3. Hebb, D. O. The first stage of perception: growth of the assembly. *The Organization of Behavior* **4**, 60–78 (1949).

4. Lopes da Silva, F. EEG and MEG: Relevance to Neuroscience. *Neuron* **80**, 1112–1128 (2013).

5. Tononi, G., Sporns, O. & Edelman, G. M. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5033–5037 (1994).

6. Park, H.-J. & Friston, K. Structural and Functional Brain Networks: From Connections to Cognition. *Science* **342**, 1238411 (2013).

7. Luppi, A. I. *et al.* Consciousness-specific dynamic interactions of brain integration and functional diversity. *Nat Commun* **10**, 4616 (2019).

8. Chen, B., Ciria, L. F., Hu, C. & Ivanov, P. Ch. Ensemble of coupling forms and networks among brain rhythms as function of states and cognition. *Commun Biol* **5**, 82 (2022).

9. Minati, L., Varotto, G., D'Incerti, L., Panzica, F. & Chan, D. From brain topography to brain topology: relevance of graph theory to functional neuroscience. *NeuroReport* **24**, 536–543 (2013).

10. Friston, K. J. Functional and effective connectivity in neuroimaging: A synthesis. *Hum. Brain Mapp.* **2**, 56–78 (1994).

11. Bullmore, E. & Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* **10**, 186–198 (2009).

12. Bassett, D. S. & Sporns, O. Network neuroscience. *Nat Neurosci* **20**, 353–364 (2017).

13. Ursino, M., Cona, F. & Zavaglia, M. The generation of rhythms within a cortical region: Analysis of a neural mass model. *NeuroImage* **52**, 1080–1094 (2010).

14. Adey, W. R., Walter, D. O. & Hendrix, C. E. Computer techniques in correlation and spectral analyses of cerebral slow waves during discriminative behavior. *Experimental Neurology* **3**, 501–524 (1961).

15. Horwitz, B. The elusive concept of brain connectivity. *Neuroimage* **19**, 466–470 (2003).

16. Reid, A. T. *et al.* Advancing functional connectivity research from association to causation. *Nat Neurosci* **22**, 1751–1760 (2019).

17. Wiener, N. The theory of prediction. *Modern mathematics for engineers* (1956).

18. Granger, C. W. J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* **37**, 424–438 (1969).

19. Friston, K. J. Functional and Effective Connectivity: A Review. *Brain Connectivity* **1**, 13–36 (2011).

20. Seth, A. K., Barrett, A. B. & Barnett, L. Granger Causality Analysis in Neuroscience and Neuroimaging. *Journal of Neuroscience* **35**, 3293–3297 (2015).

21. Friston, K. J., Harrison, L. & Penny, W. Dynamic causal modelling. *NeuroImage* **19**, 1273–1302 (2003).

22. Wang, H. E. *et al.* A systematic framework for functional connectivity measures. *Front. Neurosci.* **8**, (2014).

23. Sporns, O., Tononi, G. & Edelman, G. M. Connectivity and complexity: the relationship between neuroanatomy and brain dynamics. *Neural networks* **13**, 909–922 (2000).

24. Medaglia, J. D., Lynall, M.-E. & Bassett, D. S. Cognitive Network Neuroscience. *Journal of Cognitive Neuroscience* **27**, 1471–1491 (2015).

25. Sporns, O. Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience* **17**, 652–660 (2014).

26. Petersen, S. E. & Sporns, O. Brain Networks and Cognitive Architectures. *Neuron* **88**, 207–219 (2015).

27. DeSalvo, M. N., Douw, L., Takaya, S., Liu, H. & Stufflebeam, S. M. Task-dependent reorganization of functional connectivity networks during visual semantic decision making. *Brain and behavior* **4**, 877–885 (2014).

28. Davison, E. N. *et al.* Brain Network Adaptability across Task States. *PLOS Computational Biology* **11**, 1–14 (2015).

29. Cole, M. W. *et al.* Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat Neurosci* **16**, 1348–1355 (2013).

30. Dai, Z. & He, Y. Disrupted structural and functional brain connectomes in mild cognitive impairment and Alzheimer's disease. *Neuroscience Bulletin* **30**, 217–232 (2014).

31. Uhlhaas, P. J. & Singer, W. Neural Synchrony in Brain Disorders: Relevance for Cognitive Dysfunctions and Pathophysiology. *Neuron* **52**, 155–168 (2006).

32. Van Mierlo, P. *et al.* Functional brain connectivity from EEG in epilepsy: Seizure prediction and epileptogenic focus localization. *Progress in neurobiology* **121**, 19–35 (2014).

33. Tahedl, M., Levine, S. M., Greenlee, M. W., Weissert, R. & Schwarzbach, J. V. Functional connectivity in multiple sclerosis: recent findings and future directions. *Frontiers in neurology* **9**, 828 (2018).

34. Hull, J. V. *et al.* Resting-state functional connectivity in autism spectrum disorders: a review. *Frontiers in psychiatry* **7**, 205 (2017).

35. Konrad, K. & Eickhoff, S. B. Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder. *Human brain mapping* **31**, 904–916 (2010).

36. Wang, X.-J. Neurophysiological and Computational Principles of Cortical Rhythms in Cognition. *Physiological Reviews* **90**, 1195–1268 (2010).

37. Lynall, M.-E. *et al.* Functional connectivity and brain networks in schizophrenia. *Journal of Neuroscience* **30**, 9477–9487 (2010).

38. Gao, L. & Wu, T. The study of brain functional connectivity in Parkinson’s disease. *Translational neurodegeneration* **5**, 1–7 (2016).

39. Fasiello, E. *et al.* Functional connectivity changes in insomnia disorder: A systematic review. *Sleep Medicine Reviews* **61**, 101569 (2022).

40. Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000).

41. Wendel, K. *et al.* EEG/MEG Source Imaging: Methods, Challenges, and Open Issues. *Computational Intelligence and Neuroscience* **2009**, 1–12 (2009).

42. Ahlfors, S. P., Han, J., Belliveau, J. W. & Hämäläinen, M. S. Sensitivity of MEG and EEG to Source Orientation. *Brain Topogr* **23**, 227–232 (2010).

43. Malmivuo, J., Suihko, V. & Eskola, H. Sensitivity distributions of EEG and MEG measurements. *IEEE Transactions on Biomedical Engineering* **44**, 196–208 (1997).

44. Waldert, S. *et al.* Hand movement direction decoded from MEG and EEG. *Journal of neuroscience* **28**, 1000–1008 (2008).

45. Gevins, A., Leong, H., Smith, M. E., Le, J. & Du, R. Mapping cognitive brain function with modern high-resolution electroencephalography. *Trends in neurosciences* **18**, 429–436 (1995).

46. Michel, C. M. *et al.* EEG source imaging. *Clinical neurophysiology* **115**, 2195–2222 (2004).

47. Pascual-Marqui, R. D. Review of methods for solving the EEG inverse problem. *International journal of bioelectromagnetism* **1**, 75–86 (1999).

48. Bell, A. J. & Sejnowski, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural computation* **7**, 1129–1159 (1995).

49. Darvas, F., Pantazis, D., Kucukaltun-Yildirim, E. & Leahy, R. M. Mapping human brain function with MEG and EEG: methods and validation. *NeuroImage* **23**, S289–S299 (2004).

50. Brazier, M. A. The electrical fields at the surface of the head during sleep. *Electroencephalography & Clinical Neurophysiology* (1949).

51. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: a user-friendly application for MEG/EEG analysis. *Computational intelligence and neuroscience* **2011**, (2011).

52. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience* **2011**, (2011).

53. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods* **134**, 9–21 (2004).

54. Gramfort, A. *et al.* MNE software for processing MEG and EEG data. *Neuroimage* **86**, 446–460 (2014).

55. Zorzos, I., Kakkos, I., Ventouras, E. M. & Matsopoulos, G. K. Advances in Electrical Source Imaging: A Review of the Current Approaches, Applications and Challenges. *Signals* **2**, 378–391 (2021).

56. Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J. & Lounasmaa, O. V. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics* **65**, 413 (1993).

57. Michel, C. M. & He, B. EEG source localization. *Handbook of clinical neurology* **160**, 85–101 (2019).

58. Huiskamp, G., Vroeijenstijn, M., van Dijk, R., Wieneke, G. & van Huffelen, A. C. The need for correct realistic geometry in the inverse EEG problem. *IEEE Transactions on Biomedical Engineering* **46**, 1281–1287 (1999).

59. Holmes, C. J. *et al.* Enhancement of MR images using registration for signal averaging. *Journal of computer assisted tomography* **22**, 324–333 (1998).

60. Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P. & Lancaster, J. A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage* **2**, 89–101 (1995).

61. Geddes, L. A. & Baker, L. E. The specific resistance of biological material—a compendium of data for the biomedical engineer and physiologist. *Medical and biological engineering* **5**, 271–293 (1967).

62. Gramfort, A., Papadopoulo, T., Olivi, E. & Clerc, M. OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomedical engineering online* **9**, 1–20 (2010).

63. Baillet, S., Mosher, J. C. & Leahy, R. M. Electromagnetic brain mapping. *IEEE Signal Process. Mag.* **18**, 14–30 (2001).

64. Grech, R. *et al.* Review on solving the inverse problem in EEG source analysis. *J NeuroEngineering Rehabil* **5**, 25 (2008).

65. S. Supek & C. J. Aine. Simulation studies of multiple dipole neuromagnetic source localization: model order and limits of source resolution. *IEEE Transactions on Biomedical Engineering* **40**, 529–540 (1993).

66. Mosher, J. C. & Leahy, R. M. Recursive MUSIC: a framework for EEG and MEG source localization. *IEEE Transactions on Biomedical Engineering* **45**, 1342–1354 (1998).

67. Dale, A. M. & Sereno, M. I. Improved Localizadon of Cortical Activity by Combining EEG and MEG with MRI Cortical Surface Reconstruction: A Linear Approach. *Journal of Cognitive Neuroscience* **5**, 162–176 (1993).

68. Hauk, O., Wakeman, D. G. & Henson, R. Comparison of noise-normalized minimum norm estimates for MEG analysis using multiple resolution metrics. *Neuroimage* **54**, 1966–1974 (2011).

69. Pascual-Marqui, R. D., Michel, C. M. & Lehmann, D. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of Psychophysiology* **18**, 49–65 (1994).

70. Pascual-Marqui, R. D. Standardized low resolution brain electromagnetic. *Clinical Pharmacology* (2002).

71. Pascual-Marqui, R. D. *et al.* Assessing interactions in the brain with exact low-resolution electromagnetic tomography. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **369**, 3768–3784 (2011).

72. Tikhonov, A. N., Arsenin, V. J., Arsenin, V. I. & Arsenin, V. Y. *Solutions of ill-posed problems*. (Vh Winston, 1977).

73. Jatoi, M. A., Kamel, N., Malik, A. S. & Faye, I. EEG based brain source localization comparison of sLORETA and eLORETA. *Australas Phys Eng Sci Med* **37**, 713–721 (2014).

74. Knösche, T. R., Gräser, M. & Anwander, A. Prior knowledge on cortex organization in the reconstruction of source current densities from EEG. *NeuroImage* **67**, 7–24 (2013).

75. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).

76. Birle, C. *et al.* Cognitive function: holarchy or holacracy? *Neurol Sci* **42**, 89–99 (2021).

77. Ward, L. M. Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences* **7**, 553–559 (2003).

78. Koepsell, K., Wang, X., Hirsch, J. & Sommer, F. Exploring the function of neural oscillations in early sensory systems. *Frontiers in Neuroscience* **3**, (2010).

79. Tao, Z. Neural oscillations and information flow associated with synaptic plasticity. (2011).

80. Linkenkaer-Hansen, K., Nikouline, V. V., Palva, J. M. & Ilmoniemi, R. J. Long-Range Temporal Correlations and Scaling Behavior in Human Brain Oscillations. *J. Neurosci.* **21**, 1370–1377 (2001).

81. Herrmann, C. S., Strüber, D., Helfrich, R. F. & Engel, A. K. EEG oscillations: From correlation to causality. *International Journal of Psychophysiology* **103**, 12–21 (2016).

82. Singer, W. Synchronization of Cortical Activity and its Putative Role in Information Processing and Learning. *Annu. Rev. Physiol.* **55**, 349–374 (1993).

83. Mizuseki, K., Sirota, A., Pastalkova, E. & Buzsáki, G. Theta Oscillations Provide Temporal Windows for Local Circuit Computation in the Entorhinal-Hippocampal Loop. *Neuron* **64**, 267–280 (2009).

84. Amzica, F. & Lopes da Silva, F. H. 20C2Cellular Substrates of Brain Rhythms. in *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields* (eds. Schomer, D. L., Lopes da Silva, F. H., Schomer, D. L. & Lopes da Silva, F. H.) 0 (Oxford University Press, 2017). doi:10.1093/med/9780190228484.003.0002.

85. Jacob, M. S., Roach, B. J., Sargent, K. S., Mathalon, D. H. & Ford, J. M. Aperiodic measures of neural excitability are associated with anticorrelated hemodynamic networks at rest: A combined EEG-fMRI study. *NeuroImage* **245**, 118705 (2021).

86. Başar-Eroglu, C., Başar, E., Demiralp, T. & Schürmann, M. P300-response: possible psychophysiological correlates in delta and theta frequency channels. A review. *International Journal of Psychophysiology* **13**, 161–179 (1992).

87. Harmony, T. The functional significance of delta oscillations in cognitive processing. *Frontiers in Integrative Neuroscience* **7**, (2013).

88. Klimesch, W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews* **29**, 169–195 (1999).

89. Mitchell, D. J., McNaughton, N., Flanagan, D. & Kirk, I. J. Frontal-midline theta from the perspective of hippocampal "theta". *Progress in Neurobiology* **86**, 156–185 (2008).

90. Clayton, M. S., Yeung, N. & Cohen Kadosh, R. The roles of cortical oscillations in sustained attention. *Trends in Cognitive Sciences* **19**, 188–195 (2015).

91. Hsieh, L.-T. & Ranganath, C. Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. *NeuroImage* **85**, 721–729 (2014).

92. Mizuhara, H., Wang, L.-Q., Kobayashi, K. & Yamaguchi, Y. A long-range cortical network emerging with theta oscillation in a mental task. *NeuroReport* **15**, (2004).

93. Mueller, E. M., Panitz, C., Hermann, C. & Pizzagalli, D. A. Prefrontal oscillations during recall of conditioned and extinguished fear in humans. *J Neurosci* **34**, 7059–7066 (2014).

94. Frey, J. N., Ruhnau, P. & Weisz, N. Not so different after all: The same oscillatory processes support different types of attention. *Brain Research* **1626**, 183–197 (2015).

95. Klimesch, W. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in cognitive sciences* **16**, 606–617 (2012).

96. Schürmann, M. & Başar, E. Functional aspects of alpha oscillations in the EEG. *International Journal of Psychophysiology* **39**, 151–158 (2001).

97. Cona, G. *et al.* Theta and alpha oscillations as signatures of internal and external attention to delayed intentions: A magnetoencephalography (MEG) study. *NeuroImage* **205**, 116295 (2020).

98. Magosso, E., De Crescenzio, F., Ricci, G., Piastra, S. & Ursino, M. EEG alpha power is modulated by attentional changes during cognitive tasks and virtual reality immersion. *Computational intelligence and neuroscience* **2019**, (2019).

99. Ricci, G., De Crescenzio, F., Santhosh, S., Magosso, E. & Ursino, M. Relationship between electroencephalographic data and comfort perception captured in a Virtual Reality design environment of an aircraft cabin. *Scientific Reports* **12**, 10938 (2022).

100. Neuper, C. & Pfurtscheller, G. Evidence for distinct beta resonance frequencies in human EEG related to specific sensorimotor cortical areas. *Clinical Neurophysiology* **112**, 2084–2097 (2001).

101. Wendiggensen, P. *et al.* Processing of embedded response plans is modulated by an interplay of frontoparietal theta and beta activity. *Journal of neurophysiology* **128**, 543–555 (2022).

102. Kilavik, B. E., Zaepffel, M., Brovelli, A., MacKay, W. A. & Riehle, A. The ups and downs of beta oscillations in sensorimotor cortex. *Experimental Neurology* **245**, 15–26 (2013).

103. Engel, A. K. & Fries, P. Beta-band oscillations—signalling the status quo? *Current Opinion in Neurobiology* **20**, 156–165 (2010).

104. Merker, B. Cortical gamma oscillations: the functional key is activation, not cognition. *Neuroscience & Biobehavioral Reviews* **37**, 401–417 (2013).

105. Womelsdorf, T., Fries, P., Mitra, P. P. & Desimone, R. Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature* **439**, 733–736 (2006).

106. Henrie, J. A. & Shapley, R. LFP power spectra in V1 cortex: the graded effect of stimulus contrast. *Journal of neurophysiology* **94**, 479–490 (2005).

107. Fries, P., Reynolds, J. H., Rorie, A. E. & Desimone, R. Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention. *Science* **291**, 1560–1563 (2001).

108. Liu, J. & Newsome, W. T. Local Field Potential in Cortical Area MT: Stimulus Tuning and Behavioral Correlations. *J. Neurosci.* **26**, 7779 (2006).

109. Barbas, H. & Rempel-Clower, N. Cortical structure predicts the pattern of corticocortical connections. *Cerebral cortex (New York, NY: 1991)* **7**, 635–646 (1997).

110. Sakkalis, V. Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Computers in Biology and Medicine* **41**, 1110–1117 (2011).

111. Sauseng, P. & Klimesch, W. What does phase information of oscillatory brain activity tell us about cognitive processes? *Neuroscience & Biobehavioral Reviews* **32**, 1001–1013 (2008).

112. Di Russo, F., Martínez, A., Sereno, M. I., Pitzalis, S. & Hillyard, S. A. Cortical sources of the early components of the visual evoked potential. *Human brain mapping* **15**, 95–111 (2002).

113. Bullmore, E. & Sporns, O. The economy of brain network organization. *Nat Rev Neurosci* **13**, 336–349 (2012).

114. Ding, M., Chen, Y. & Bressler, S. L. 17 Granger causality: basic theory and application to neuroscience. *Handbook of time series analysis: recent theoretical developments and applications* **437**, (2006).

115. C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379–423 (1948).

116. Baccalá, L. A. & Sameshima, K. Partial directed coherence: a new concept in neural structure determination. *Biol Cybern* **84**, 463–474 (2001).

117. Geweke, J. Measurement of Linear Dependence and Feedback Between Multiple Time Series. *Journal of the American Statistical Association* **77**, 304–313 (1982).

118. Shojaie, A. & Michailidis, G. Discovering graphical Granger causality using the truncating lasso penalty. *Bioinformatics* **26**, i517–i523 (2010).

119. Kus, R., Kaminski, M. & Blinowska, K. J. Determination of EEG Activity Propagation: Pair-Wise Versus Multichannel Estimate. *IEEE Trans. Biomed. Eng.* **51**, 1501–1510 (2004).

120. Blinowska, K. J., Kuś, R. & Kamiński, M. Granger causality and information flow in multivariate processes. *Phys. Rev. E* **70**, 050902 (2004).

121. Greenblatt, R. E., Pflieger, M. E. & Ossadtchi, A. E. Connectivity measures applied to human brain electrophysiological data. *Journal of Neuroscience Methods* **207**, 1–16 (2012).

122. Quyen, M. L. V., Martinerie, J., Adam, C. & Varela, F. J. Nonlinear analyses of interictal EEG map the brain interdependences in human focal epilepsy. *Physica D: Nonlinear Phenomena* **127**, 250–266 (1999).

123. Kraskov, A., Stögbauer, H. & Grassberger, P. Estimating mutual information. *Phys. Rev. E* **69**, 066138 (2004).

124. Barnett, L., Barrett, A. B. & Seth, A. K. Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables. *Phys. Rev. Lett.* **103**, 238701 (2009).

125. Lizier, J. T. & Prokopenko, M. Differentiating information transfer and causal effect. *The European Physical Journal B* **73**, 605–615 (2010).

126. Ricci, G., Magosso, E. & Ursino, M. The Relationship between Oscillations in Brain Regions and Functional Connectivity: A Critical Analysis with the Aid of Neural Mass Models. *Brain Sciences* **11**, 487 (2021).

127. Pascual-Marqui, R. D. *et al.* Assessing interactions in the brain with exact low-resolution electromagnetic tomography. *Philos Trans A Math Phys Eng Sci* **369**, 3768–3784 (2011).

128. Wendling, F., Ansari-Asl, K., Bartolomei, F. & Senhadji, L. From EEG signals to brain connectivity: a model-based evaluation of interdependence measures. *J. Neurosci. Methods* **183**, 9–18 (2009).

129. Sun, J., Li, Z. & Tong, S. Inferring functional neural connectivity with phase synchronization analysis: a review of methodology. *Comput Math Methods Med* **2012**, 239210 (2012).

130. Granger, C. W. J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* **37**, 424–438 (1969).

131. Ding, M., Chen, Y. & Bressler, S. L. Granger Causality: Basic Theory and Application to Neuroscience. in *Handbook of Time Series Analysis* 437–460 (John Wiley & Sons, Ltd, 2006). doi:10.1002/9783527609970.ch17.

132. Chicharro, D. On the spectral formulation of Granger causality. *Biol Cybern* **105**, 331–347 (2011).

133. Schreiber, null. Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000).

134. Wibral, M. *et al.* Measuring information-transfer delays. *PLoS ONE* **8**, e55809 (2013).

135. Vicente, R., Wibral, M., Lindner, M. & Pipa, G. Transfer entropy--a model-free measure of effective connectivity for the neurosciences. *J Comput Neurosci* **30**, 45–67 (2011).

136. Lindner, M., Vicente, R., Priesemann, V. & Wibral, M. TRENTOOL: a Matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC Neurosci* **12**, 119 (2011).

137. Ursino, M., Ricci, G. & Magosso, E. TRANSFER ENTROPY AS A MEASURE OF BRAIN CONNECTIVITY: A CRITICAL ANALYSIS WITH THE HELP OF NEURAL MASS MODELS. *Front. Comput. Neurosci.* **14**, (2020).

138. Farahani, F. V., Karwowski, W. & Lighthall, N. R. Application of Graph Theory for Identifying Connectivity Patterns in Human Brain Networks: A Systematic Review. *Front. Neurosci.* **13**, 585 (2019).

139. Fries, P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences* **9**, 474–480 (2005).

140. Sporns, O., Tononi, G. & Kötter, R. The Human Connectome: A Structural Description of the Human Brain: e42. *PLoS Computational Biology* **1**, (2005).

141. Liao, X., Vasilakos, A. V. & He, Y. Small-world human brain networks: Perspectives and challenges. *Neuroscience & Biobehavioral Reviews* **77**, 286–300 (2017).

142. Avena-Koenigsberger, A., Misic, B. & Sporns, O. Communication dynamics in complex brain networks. *Nat Rev Neurosci* **19**, 17–33 (2018).

143. Liang, X., Zou, Q., He, Y. & Yang, Y. Topologically Reorganized Connectivity Architecture of Default-Mode, Executive-Control, and Salience Networks across Working Memory Task Loads. *Cerebral Cortex* **26**, 1501–1511 (2016).

144. Collin, G. & van den Heuvel, M. P. The Ontogeny of the Human Connectome: Development and Dynamic Changes of Brain Connectivity Across the Life Span. *Neuroscientist* **19**, 616–628 (2013).

145. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **52**, 1059–1069 (2010).

146. Fagiolo, G. Clustering in complex directed networks. *Phys. Rev. E* **76**, 026107 (2007).

147. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).

148. Hagmann, P. *et al.* Mapping the Structural Core of Human Cerebral Cortex: e159. *PLoS Biology* **6**, e159 (2008).

149. Braun, U. *et al.* Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proceedings of the National Academy of Sciences* **112**, 11678–11683 (2015).

150. Xia, M. & He, Y. Magnetic resonance imaging and graph theoretical analysis of complex brain networks in neuropsychiatric disorders. *Brain connectivity* **1**, 349–365 (2011).

151. Maex, R. & De Schutter, E. Mechanism of spontaneous and self-sustained oscillations in networks connected through axo-axonal gap junctions. *European Journal of Neuroscience* **25**, 3347–3358 (2007).

152. Breakspear, M. Dynamic models of large-scale brain activity. *Nat Neurosci* **20**, 340–352 (2017).

153. Wilson, H. R. & Cowan, J. D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal* **12**, 1–24 (1972).

154. Lopes da Silva, F. H., Hoeks, A., Smits, H. & Zetterberg, L. H. Model of brain rhythmic activity. *Kybernetik* **15**, 27–37 (1974).

155. Freeman, W. J. Models of the dynamics of neural populations. *Electroencephalography and clinical neurophysiology. Supplement* 9–18 (1978).

156. Jansen, B. H. & Rit, V. G. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological cybernetics* **73**, 357–366 (1995).

157. Sotero, R. C., Trujillo-Barreto, N. J., Iturria-Medina, Y., Carbonell, F. & Jimenez, J. C. Realistically coupled neural mass models can generate EEG rhythms. *Neural computation* **19**, 478–512 (2007).

158. Wendling, F., Bartolomei, F., Bellanger, J. J. & Chauvel, P. Epileptic fast activity can be explained by a model of impaired GABAergic dendritic inhibition: Epileptic activity explained by dendritic dis-inhibition. *European Journal of Neuroscience* **15**, 1499–1508 (2002).

159. Ursino, M., Zavaglia, M., Astolfi, L. & Babiloni, F. Use of a neural mass model for the analysis of effective connectivity among cortical regions based on high resolution EEG recordings. *Biological Cybernetics* **96**, 351–365 (2007).

160. Cona, F., Zavaglia, M., Astolfi, L., Babiloni, F. & Ursino, M. Changes in EEG power spectral density and cortical connectivity in healthy and tetraplegic patients during a motor imagery task. *Computational intelligence and neuroscience* **2009**, (2009).

161. Ursino, M., Cona, F. & Zavaglia, M. The generation of rhythms within a cortical region: analysis of a neural mass model. *Neuroimage* **52**, 1080–1094 (2010).

162. Friston, K. Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS Biol.* **7**, e33 (2009).

163. Horwitz, B. The elusive concept of brain connectivity. *Neuroimage* **19**, 466–470 (2003).

164. van den Heuvel, M. P. & Hulshoff Pol, H. E. Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur Neuropsychopharmacol* **20**, 519–534 (2010).

165. Rossini, P. M. *et al.* Methods for analysis of brain connectivity: An IFCN-sponsored review. *Clin Neurophysiol* **130**, 1833–1858 (2019).

166. Sakkalis, V. Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Comput. Biol. Med.* **41**, 1110–1117 (2011).

167. Koenig, T., Studer, D., Hubl, D., Melie, L. & Strik, W. K. Brain connectivity at different time-scales measured with EEG. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **360**, 1015–1023 (2005).

168. Astolfi, L. *et al.* Imaging functional brain connectivity patterns from high-resolution EEG and fMRI via graph theory. *Psychophysiology* **44**, 880–893 (2007).

169. Friston, K., Moran, R. & Seth, A. K. Analysing connectivity with Granger causality and dynamic causal modelling. *Curr. Opin. Neurobiol.* **23**, 172–178 (2013).

170. Valdes-Sosa, P. A., Roebroeck, A., Daunizeau, J. & Friston, K. Effective connectivity: influence, causality and biophysical modeling. *Neuroimage* **58**, 339–361 (2011).

171. Penny, W. D., Stephan, K. E., Mechelli, A. & Friston, K. J. Comparing dynamic causal models. *Neuroimage* **22**, 1157–1172 (2004).

172. Reid, A. T. *et al.* Advancing functional connectivity research from association to causation. *Nat. Neurosci.* **22**, 1751–1760 (2019).

173. Bastos, A. M. & Schoffelen, J.-M. A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. *Front Syst Neurosci* **9**, 175 (2015).

174. Wang, H. E. *et al.* A systematic framework for functional connectivity measures. *Front Neurosci* **8**, 405 (2014).

175. Wiener N. The theory of prediction. in *Modern mathematics for the engineer.* 165–190 (McGraw-Hill, 1956).

176. Granger, C. W. J. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 227–238 (1980).

177. Bajaj, S., Adhikari, B. M., Friston, K. J. & Dhamala, M. Bridging the Gap: Dynamic Causal Modeling and Granger Causality Analysis of Resting State Functional Magnetic Resonance Imaging. *Brain Connect* **6**, 652–661 (2016).

178. Lobier, M., Siebenhühner, F., Palva, S. & Palva, J. M. Phase transfer entropy: a novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. *Neuroimage* **85 Pt 2**, 853–872 (2014).

179. Schindler, K., Palus, M., Vejmelka, M. & Bhattacharya, J. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports* **441**, 1–46 (2007).

180. Wollstadt, P., Martínez-Zarzuela, M., Vicente, R., Díaz-Pernas, F. J. & Wibral, M. Efficient transfer entropy analysis of non-stationary neural time series. *PLoS ONE* **9**, e102833 (2014).

181. Wibral, M., Vicente, R. & Lindner, M. Transfer Entropy in Neuroscience. in *Directed Information Measures in Neuroscience* (eds. Wibral, M., Vicente, R. & Lizier, J. T.) 3–36 (Springer Berlin Heidelberg, 2014). doi:10.1007/978-3-642-54474-3_1.

182. Montalto, A., Faes, L. & Marinazzo, D. MuTE: a MATLAB toolbox to compare established and novel estimators of the multivariate transfer entropy. *PLoS ONE* **9**, e109462 (2014).

183. Lizier, J. T. & Prokopenko, M. Differentiating information transfer and causal effect. *Eur. Phys. J. B* **73**, 605–615 (2010).

184. Ito, S. *et al.* Extending transfer entropy improves identification of effective connectivity in a spiking cortical network model. *PLoS ONE* **6**, e27431 (2011).

185. Garofalo, M., Nieus, T., Massobrio, P. & Martinoia, S. Evaluation of the performance of information theory-based methods and cross-correlation to estimate the functional connectivity in cortical networks. *PLoS ONE* **4**, e6482 (2009).

186. Orlandi, J. G., Stetter, O., Soriano, J., Geisel, T. & Battaglia, D. Transfer entropy reconstruction and labeling of neuronal connections from simulated calcium imaging. *PLoS ONE* **9**, e98842 (2014).

187. Timme, N. M. & Lapish, C. A Tutorial for Information Theory in Neuroscience. *eNeuro* **5**, (2018).

188. David, O., Cosmelli, D. & Friston, K. J. Evaluation of different measures of functional connectivity using a neural mass model. *Neuroimage* **21**, 659–673 (2004).

189. Ansari-Asl, K., Senhadji, L., Bellanger, J.-J. & Wendling, F. Quantitative evaluation of linear and nonlinear methods characterizing interdependencies between brain signals. *Phys Rev E Stat Nonlin Soft Matter Phys* **74**, 031916 (2006).

190. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).

191. David, O., Harrison, L. & Friston, K. J. Modelling event-related responses in the brain. *Neuroimage* **25**, 756–770 (2005).

192. Zavaglia, M., Astolfi, L., Babiloni, F. & Ursino, M. The effect of connectivity on EEG rhythms, power spectral density and coherence among coupled neural populations: analysis with a neural mass model. *IEEE Trans Biomed Eng* **55**, 69–77 (2008).

193. Zavaglia, M., Cona, F. & Ursino, M. A neural mass model to simulate different rhythms in a cortical region. *Comput Intell Neurosci* 456140 (2010) doi:10.1155/2010/456140.

194. Cona, F., Zavaglia, M., Massimini, M., Rosanova, M. & Ursino, M. A neural mass model of interconnected regions simulates rhythm propagation observed via TMS-EEG. *Neuroimage* **57**, 1045–1058 (2011).

195. David, O. & Friston, K. J. A neural mass model for MEG/EEG: coupling and neuronal dynamics. *Neuroimage* **20**, 1743–1755 (2003).

196. Nichols, J. M. *et al.* Detecting nonlinearity in structural systems using the transfer entropy. *Phys Rev E Stat Nonlin Soft Matter Phys* **72**, 046217 (2005).

197. Dynamical Systems and Turbulence. in *Lecture Notes in Mathematics. Detecting Strange Attractors in Turbulence* vol. 898 366–381 (Springer, 1980).

198. Mangia, A. L., Ursino, M., Lannocca, M. & Cappello, A. Transcallosal Inhibition during Motor Imagery: Analysis of a Neural Mass Model. *Front Comput Neurosci* **11**, 57 (2017).

199. Grefkes, C., Eickhoff, S. B., Nowak, D. A., Dafotakis, M. & Fink, G. R. Dynamic intra- and interhemispheric interactions during unilateral and bilateral hand movements assessed with fMRI and DCM. *Neuroimage* **41**, 1382–1394 (2008).

200. Pool, E.-M. *et al.* Network dynamics engaged in the modulation of motor behavior in stroke patients. *Hum Brain Mapp* **39**, 1078–1092 (2018).

201. Sporns, O. The human connectome: a complex network. *Ann. N. Y. Acad. Sci.* **1224**, 109–125 (2011).

202. Sporns, O. Network attributes for segregation and integration in the human brain. *Curr. Opin. Neurobiol.* **23**, 162–171 (2013).

203. Izhikevich, E. M. Polychronization: computation with spikes. *Neural Comput* **18**, 245–282 (2006).

204. Goodman, D. & Brette, R. Brian: a simulator for spiking neural networks in python. *Front Neuroinform* **2**, 5 (2008).

205. Wendling, F., Bartolomei, F., Bellanger, J. J. & Chauvel, P. Epileptic fast activity can be explained by a model of impaired GABAergic dendritic inhibition. *Eur. J. Neurosci.* **15**, 1499–1508 (2002).

206. Moran, R. J. *et al.* Bayesian estimation of synaptic physiology from the spectral responses of neural masses. *Neuroimage* **42**, 272–284 (2008).

207. Bhattacharya, B. S., Coyle, D. & Maguire, L. P. A thalamo-cortico-thalamic neural mass model to study alpha rhythms in Alzheimer’s disease. *Neural Netw* **24**, 631–645 (2011).

208. Sotero, R. C., Trujillo-Barreto, N. J., Iturria-Medina, Y., Carbonell, F. & Jimenez, J. C. Realistically coupled neural mass models can generate EEG rhythms. *Neural Comput* **19**, 478–512 (2007).

209. Cona, F., Lacanna, M. & Ursino, M. A thalamo-cortical neural mass model for the simulation of brain rhythms during sleep. *J Comput Neurosci* **37**, 125–148 (2014).

210. Cona, F. & Ursino, M. A neural mass model of place cell activity: theta phase precession, replay and imagination of never experienced paths. *J Comput Neurosci* **38**, 105–127 (2015).

211. Harmah, D. J. *et al.* Measuring the Non-linear Directed Information Flow in Schizophrenia by Multivariate Transfer Entropy. *Front Comput Neurosci* **13**, 85 (2019).

212. Shimono, M. & Beggs, J. M. Functional Clusters, Hubs, and Communities in the Cortical Microconnectome. *Cereb. Cortex* **25**, 3743–3757 (2015).

213. Olejarczyk, E., Marzetti, L., Pizzella, V. & Zappasodi, F. Comparison of connectivity analyses for resting state EEG data. *J Neural Eng* **14**, 036017 (2017).

214. Song, S., Sjöström, P. J., Reigl, M., Nelson, S. & Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3**, e68 (2005).

215. Wollstadt, P. *et al.* Breakdown of local information processing may underlie isoflurane anesthesia effects. *PLoS Comput. Biol.* **13**, e1005511 (2017).

216. Honey, C. J., Kötter, R., Breakspear, M. & Sporns, O. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10240–10245 (2007).

217. Vinck, M., van Wingerden, M., Womelsdorf, T., Fries, P. & Pennartz, C. M. A. The pairwise phase consistency: a bias-free measure of rhythmic neuronal synchronization. *Neuroimage* **51**, 112–122 (2010).

218. Chu, C. J. *et al.* Emergence of stable functional networks in long-term human electroencephalography. *J. Neurosci.* **32**, 2703–2713 (2012).

219. Bonita, J. D. *et al.* Time domain measures of inter-channel EEG correlations: a comparison of linear, nonparametric and nonlinear measures. *Cogn Neurodyn* **8**, 1–15 (2014).

220. Fraschini, M. *et al.* The effect of epoch length on estimated EEG functional connectivity and brain network organisation. *J Neural Eng* **13**, 036015 (2016).

221. Bastos, A. M. & Schoffelen, J.-M. A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. *Front Syst Neurosci* **9**, 175 (2015).

222. Pool, E.-M., Rehme, A. K., Fink, G. R., Eickhoff, S. B. & Grefkes, C. Network dynamics engaged in the modulation of motor behavior in healthy subjects. *Neuroimage* **82**, 68–76 (2013).

223. Bajaj, S., Butler, A. J., Drake, D. & Dhamala, M. Brain effective connectivity during motor-imagery and execution following stroke and rehabilitation. *Neuroimage Clin* **8**, 572–582 (2015).

224. Bönstrup, M., Schulz, R., Feldheim, J., Hummel, F. C. & Gerloff, C. Dynamic causal modelling of EEG and fMRI to characterize network architectures in a simple motor task. *Neuroimage* **124**, 498–508 (2016).

225. Kim, Y. K., Park, E., Lee, A., Im, C.-H. & Kim, Y.-H. Changes in network connectivity during motor imagery and execution. *PLoS ONE* **13**, e0190715 (2018).

226. Larsen, L. H. *et al.* Modulation of task-related cortical connectivity in the acute and subacute phase after stroke. *Eur. J. Neurosci.* **47**, 1024–1032 (2018).

227. He, B. *et al.* Electrophysiological Brain Connectivity: Theory and Implementation. *IEEE Trans Biomed Eng* (2019) doi:10.1109/TBME.2019.2913928.

228. Fries, P. Rhythms for Cognition: Communication through Coherence. *Neuron* **88**, 220–235 (2015).

229. Wang, X.-J. Neurophysiological and Computational Principles of Cortical Rhythms in Cognition. *Physiol Rev* **90**, 1195–1268 (2010).

230. Roux, F. & Uhlhaas, P. J. Working memory and neural oscillations: α-γ versus θ-γ codes for distinct WM information? *Trends Cogn Sci* **18**, 16–25 (2014).

231. Uhlhaas, P. J. *et al.* Neural synchrony in cortical networks: history, concept and current status. *Front Integr Neurosci* **3**, 17 (2009).

232. Feige, B. *et al.* Cortical and Subcortical Correlates of Electroencephalographic Alpha Rhythm Modulation. *Journal of Neurophysiology* **93**, 2864–2872 (2005).

233. Laufs, H. *et al.* EEG-correlated fMRI of human alpha activity. *Neuroimage* **19**, 1463–1476 (2003).

234. Mantini, D., Perrucci, M. G., Gratta, C. D., Romani, G. L. & Corbetta, M. Electrophysiological signatures of resting state networks in the human brain. *PNAS* **104**, 13170–13175 (2007).

235. Rosanova, M. *et al.* Natural frequencies of human corticothalamic circuits. *J Neurosci* **29**, 7679–7685 (2009).

236. Thut, G. *et al.* Rhythmic TMS Causes Local Entrainment of Natural Oscillatory Signatures. *Current Biology* **21**, 1176–1185 (2011).

237. Vallesi, A. *et al.* Natural oscillation frequencies in the two lateral prefrontal cortices induced by Transcranial Magnetic Stimulation. *NeuroImage* **227**, 117655 (2021).

238. Khanna, P. & Carmena, J. M. Neural oscillations: beta band activity across motor networks. *Curr. Opin. Neurobiol.* **32**, 60–67 (2015).

239. Dehghani, N. *et al.* Dynamic Balance of Excitation and Inhibition in Human and Monkey Neocortex. *Sci Rep* **6**, 23176 (2016).

240. Welch, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics* **15**, 70–73 (1967).

241. Cole, S. R. & Voytek, B. Brain Oscillations and the Importance of Waveform Shape. *Trends Cogn Sci* **21**, 137–149 (2017).

242. Jones, S. R. When brain rhythms aren’t ‘rhythmic’: implication for their mechanisms and meaning. *Curr Opin Neurobiol* **40**, 72–80 (2016).

243. Lopes da Silva, F. H., Vos, J. E., Mooibroek, J. & Van Rotterdam, A. Relative contributions of intracortical and thalamo-cortical processes in the generation of alpha rhythms, revealed by partial coherence analysis. *Electroencephalogr Clin Neurophysiol* **50**, 449–456 (1980).

244. Lisman, J. E. & Jensen, O. The θ-γ neural code. *Neuron* **77**, 1002–1016 (2013).

245. Groppe, D. M. *et al.* Dominant frequencies of resting human brain activity as measured by the electrocorticogram. *Neuroimage* **79**, 223–233 (2013).

246. Magosso, E., De Crescenzio, F., Ricci, G., Piastra, S. & Ursino, M. EEG Alpha Power Is Modulated by Attentional Changes during Cognitive Tasks and Virtual Reality Immersion. *Comput Intell Neurosci* **2019**, 7051079 (2019).

247. Magosso, E., Ricci, G. & Ursino, M. Modulation of brain alpha rhythm and heart rate variability by attention-related mechanisms. *AIMS Neurosci* **6**, 1–24 (2019).

248. Klimesch, W. α-band oscillations, attention, and controlled access to stored information. *Trends Cogn Sci* **16**, 606–617 (2012).

249. Buzsaki, G. *Rhythms of the Brain*. (Oxford University Press, 2006).

250. Schreckenberger, M. *et al.* The thalamus as the generator and modulator of EEG alpha rhythm: a combined PET/EEG study with lorazepam challenge in humans. *Neuroimage* **22**, 637–644 (2004).

251. Yin, S., Liu, Y. & Ding, M. Amplitude of Sensorimotor Mu Rhythm Is Correlated with BOLD from Multiple Brain Regions: A Simultaneous EEG-fMRI Study. *Front Hum Neurosci* **10**, 364 (2016).

252. Halgren, M. *et al.* The generation and propagation of the human alpha rhythm. *Proc Natl Acad Sci U S A* **116**, 23772–23782 (2019).

253. Wan, L. *et al.* From eyes-closed to eyes-open: Role of cholinergic projections in EC-to-EO alpha reactivity revealed by combining EEG and MRI. *Hum Brain Mapp* **40**, 566–577 (2019).

254. Jia, X., Smith, M. A. & Kohn, A. Stimulus selectivity and spatial coherence of gamma components of the local field potential. *J Neurosci* **31**, 9390–9403 (2011).

255. Berens, P., Keliris, G. A., Ecker, A. S., Logothetis, N. K. & Tolias, A. S. Feature selectivity of the gamma-band of the local field potential in primate primary visual cortex. *Front Neurosci* **2**, 199–207 (2008).

256. Khanna, P. & Carmena, J. M. Beta band oscillations in motor cortex reflect neural population signals that delay movement onset. *eLife* **6**,.

257. Ursino, M., Ricci, G. & Magosso, E. Transfer entropy as a measure of brain connectivity: a critical analysis with the help of neural mass models. *Frontiers in computational neuroscience* **14**, 45 (2020).

258. Tort, A. B. L., Komorowski, R., Eichenbaum, H. & Kopell, N. Measuring Phase-Amplitude Coupling Between Neuronal Oscillations of Different Frequencies. *Journal of Neurophysiology* **104**, 1195–1210 (2010).

259. Chiappini, E., Silvanto, J., Hibbard, P. B., Avenanti, A. & Romei, V. Strengthening functionally specific neural pathways with transcranial brain stimulation. *Curr Biol* **28**, R735–R736 (2018).

260. Clayton, M. S., Yeung, N. & Cohen Kadosh, R. The many characters of visual alpha oscillations. *Eur J Neurosci* **48**, 2498–2508 (2018).

261. Frey, J. N., Ruhnau, P. & Weisz, N. Not so different after all: The same oscillatory processes support different types of attention. *Brain Res* **1626**, 183–197 (2015).

262. Palva, S. & Palva, J. M. New vistas for alpha-frequency band oscillations. *Trends Neurosci* **30**, 150–158 (2007).

263. Popov, T. *et al.* Cross-frequency interactions between frontal theta and posterior alpha control mechanisms foster working memory. *Neuroimage* **181**, 728–733 (2018).

264. Wang, C., Rajagovindan, R., Han, S.-M. & Ding, M. Top-Down Control of Visual Alpha Oscillations: Sources of Control Signals and Their Mechanisms of Action. *Front Hum Neurosci* **10**, 15 (2016).

265. Desowska, A. & Turner, D. L. Dynamics of brain connectivity after stroke. *Rev Neurosci* **30**, 605–623 (2019).

266. Pool, E.-M. *et al.* Network dynamics engaged in the modulation of motor behavior in stroke patients. *Hum Brain Mapp* **39**, 1078–1092 (2018).

267. Volz, L. J. *et al.* Motor cortex excitability and connectivity in chronic stroke: a multimodal model of functional reorganization. *Brain Struct Funct* **220**, 1093–1107 (2015).

268. Grefkes, C., Eickhoff, S. B., Nowak, D. A., Dafotakis, M. & Fink, G. R. Dynamic intra- and interhemispheric interactions during unilateral and bilateral hand movements assessed with fMRI and DCM. *Neuroimage* **41**, 1382–1394 (2008).

269. Grefkes, C. & Fink, G. R. Reorganization of cerebral networks after stroke: new insights from neuroimaging with connectivity approaches. *Brain* **134**, 1264–1276 (2011).

270. Rehme, A. K., Fink, G. R., von Cramon, D. Y. & Grefkes, C. The role of the contralesional motor cortex for motor recovery in the early days after stroke assessed with longitudinal FMRI. *Cereb. Cortex* **21**, 756–768 (2011).

271. de Vico Fallani, F. *et al.* Evaluation of the brain network organization from EEG signals: a preliminary evidence in stroke patient. *Anat Rec (Hoboken)* **292**, 2023–2031 (2009).

272. Gerloff, C. *et al.* Multimodal imaging of brain reorganization in motor areas of the contralesional hemisphere of well recovered patients after capsular stroke. *Brain* **129**, 791–808 (2006).

273. Larivière, S., Ward, N. S. & Boudrias, M.-H. Disrupted functional network integrity and flexibility after stroke: Relation to motor impairments. *Neuroimage Clin* **19**, 883–891 (2018).

274. Pool, E.-M., Rehme, A. K., Eickhoff, S. B., Fink, G. R. & Grefkes, C. Functional resting-state connectivity of the human motor network: differences between right- and left-handers. *Neuroimage* **109**, 298–306 (2015).

275. Pool, E.-M., Rehme, A. K., Fink, G. R., Eickhoff, S. B. & Grefkes, C. Handedness and effective connectivity of the motor system. *Neuroimage* **99**, 451–460 (2014).

276. Ursino, M., Ricci, G. & Magosso, E. Transfer Entropy as a Measure of Brain Connectivity: A Critical Analysis With the Help of Neural Mass Models. *Front Comput Neurosci* **14**, 45 (2020).

277. Athanasiou, A. *et al.* Investigating the Role of Alpha and Beta Rhythms in Functional Motor Networks. *Neuroscience* **378**, 54–70 (2018).

278. Neuper, C., Wörtz, M. & Pfurtscheller, G. ERD/ERS patterns reflecting sensorimotor activation and deactivation. *Prog. Brain Res.* **159**, 211–222 (2006).

279. Neuper, C. & Pfurtscheller, G. Event-related dynamics of cortical rhythms: frequency-specific features and functional correlates. *Int J Psychophysiol* **43**, 41–58 (2001).

280. Heinrichs-Graham, E. & Wilson, T. W. Is an absolute level of cortical beta suppression required for proper movement? Magnetoencephalographic evidence from healthy aging. *Neuroimage* **134**, 514–521 (2016).

281. Kaiser, V. *et al.* Relationship between electrical brain responses to motor imagery and motor impairment in stroke. *Stroke* **43**, 2735–2740 (2012).

282. Pichiorri, F. *et al.* Brain-computer interface boosts motor imagery practice during stroke recovery. *Ann. Neurol.* **77**, 851–865 (2015).

283. Hantson, L. *et al.* The European Stroke Scale. *Stroke* **25**, 2215–2219 (1994).

284. Gladstone, D. J., Danells, C. J. & Black, S. E. The fugl-meyer assessment of motor recovery after stroke: a critical review of its measurement properties. *Neurorehabil Neural Repair* **16**, 232–240 (2002).

285. Toppi, J. *et al.* Investigating the effects of a sensorimotor rhythm-based BCI training on the cortical activity elicited by mental imagery. *Journal of Neural Engineering* **11**, 035010 (2014).

286. Astolfi, L. *et al.* Comparison of different cortical connectivity estimators for high-resolution EEG recordings. *Hum Brain Mapp* **28**, 143–157 (2007).

287. Babiloni, F. *et al.* Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function. *Neuroimage* **24**, 118–131 (2005).

288. Baccala, L. A., Sameshima, K. & Takahashi, D. Y. Generalized Partial Directed Coherence. in 163–166 (IEEE, 2007). doi:10.1109/ICDSP.2007.4288544.

289. Takahashi, D. Y., Baccalà, L. A. & Sameshima, K. Connectivity Inference between Neural Structures via Partial Directed Coherence. *Journal of Applied Statistics* **34**, 1259–1273 (2007).

290. Rehme, A. K., Eickhoff, S. B., Wang, L. E., Fink, G. R. & Grefkes, C. Dynamic causal modeling of cortical activity from the acute to the chronic stage after stroke. *Neuroimage* **55**, 1147–1158 (2011).

291. Hellige, J. B., Taylor, A. K. & Eng, T. L. Interhemispheric interaction when both hemispheres have access to the same stimulus information. *J Exp Psychol Hum Percept Perform* **15**, 711–722 (1989).

292. Adam, R. & Güntürkün, O. When one hemisphere takes control: metacontrol in pigeons (Columba livia). *PLoS ONE* **4**, e5307 (2009).

293. Bloom, J. S. & Hynd, G. W. The role of the corpus callosum in interhemispheric transfer of information: excitation or inhibition? *Neuropsychol Rev* **15**, 59–71 (2005).

294. Kinsbourne, M. Hemispheric specialization and the growth of human understanding. *Am Psychol* **37**, 411–420 (1982).

295. Welcome, S. E. & Chiarello, C. How dynamic is interhemispheric interaction? Effects of task switching on the across-hemisphere advantage. *Brain Cogn* **67**, 69–75 (2008).

296. Siebner, H. R., Peller, M. & Lee, L. Applications of combined TMS-PET studies in clinical and basic research. *Suppl Clin Neurophysiol* **56**, 63–72 (2003).

297. Ferbert, A. *et al.* Interhemispheric inhibition of the human motor cortex. *J. Physiol. (Lond.)* **453**, 525–546 (1992).

298. Wassermann, E. M., Fuhr, P., Cohen, L. G. & Hallett, M. Effects of transcranial magnetic stimulation on ipsilateral muscles. *Neurology* **41**, 1795–1799 (1991).

299. Kinsbourne, M. The cerebral basis of lateral asymmetries in attention. *Acta Psychol (Amst)* **33**, 193–201 (1970).

300. Sack, A. T., Camprodon, J. A., Pascual-Leone, A. & Goebel, R. The dynamics of interhemispheric compensatory processes in mental imagery. *Science* **308**, 702–704 (2005).

301. Grefkes, C. *et al.* Cortical connectivity after subcortical stroke assessed with functional magnetic resonance imaging. *Ann. Neurol.* **63**, 236–246 (2008).

302. Sharma, N., Baron, J.-C. & Rowe, J. B. Motor imagery after stroke: relating outcome to motor network connectivity. *Ann. Neurol.* **66**, 604–616 (2009).

303. Pichiorri, F. *et al.* An EEG index of sensorimotor interhemispheric coupling after unilateral stroke: clinical and neurophysiological study. *Eur. J. Neurosci.* **47**, 158–163 (2018).

304. Chollet, F. *et al.* The functional anatomy of motor recovery after stroke in humans: a study with positron emission tomography. *Ann. Neurol.* **29**, 63–71 (1991).

305. Diekhoff-Krebs, S. *et al.* Interindividual differences in motor network connectivity and behavioral response to iTBS in stroke patients. *Neuroimage Clin* **15**, 559–571 (2017).

306. Ward, N. S., Brown, M. M., Thompson, A. J. & Frackowiak, R. S. J. Neural correlates of motor recovery after stroke: a longitudinal fMRI study. *Brain* **126**, 2476–2496 (2003).

307. Weiller, C., Chollet, F., Friston, K. J., Wise, R. J. & Frackowiak, R. S. Functional reorganization of the brain in recovery from striatocapsular infarction in man. *Ann. Neurol.* **31**, 463–472 (1992).

308. Fridman, E. A. *et al.* Reorganization of the human ipsilesional premotor cortex after stroke. *Brain* **127**, 747–758 (2004).

309. Johansen-Berg, H. *et al.* The role of ipsilateral premotor cortex in hand movement after stroke. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14518–14523 (2002).

310. Chen, C. C., Kiebel, S. J. & Friston, K. J. Dynamic causal modelling of induced responses. *Neuroimage* **41**, 1293–1312 (2008).

311. Chen, C.-C. *et al.* A dynamic causal model for evoked and induced responses. *Neuroimage* **59**, 340–348 (2012).

312. Chen, C.-C. *et al.* Nonlinear coupling in the human motor system. *J. Neurosci.* **30**, 8393–8399 (2010).

313. Friston, K. J. *et al.* Dynamic causal modelling revisited. *Neuroimage* **199**, 730–744 (2019).

314. Kiebel, S. J., Garrido, M. I. & Friston, K. J. Dynamic causal modelling of evoked responses: the role of intrinsic connections. *Neuroimage* **36**, 332–345 (2007).

315. Pinotsis, D. A., Loonis, R., Bastos, A. M., Miller, E. K. & Friston, K. J. Bayesian Modelling of Induced Responses and Neuronal Rhythms. *Brain Topogr* **32**, 569–582 (2019).

316. van Wijk, B. C. M., Cagnan, H., Litvak, V., Kühn, A. A. & Friston, K. J. Generic dynamic causal modelling: An illustrative application to Parkinson’s disease. *Neuroimage* **181**, 818–830 (2018).

317. Pfurtscheller, G. & Aranibar, A. Event-related cortical desynchronization detected by power measurements of scalp EEG. *Electroencephalogr Clin Neurophysiol* **42**, 817–826 (1977).

318. Rau, C., Plewnia, C., Hummel, F. & Gerloff, C. Event-related desynchronization and excitability of the ipsilateral motor cortex during simple self-paced finger movements. *Clin Neurophysiol* **114**, 1819–1826 (2003).

319. Takemi, M., Masakado, Y., Liu, M. & Ushiba, J. Event-related desynchronization reflects downregulation of intracortical inhibition in human primary motor cortex. *J. Neurophysiol.* **110**, 1158–1166 (2013).

320. Byrne, Á., O’Dea, R. D., Forrester, M., Ross, J. & Coombes, S. Next-generation neural mass and field modeling. *J Neurophysiol* **123**, 726–742 (2020).

321. Byrne, Á., Brookes, M. J. & Coombes, S. A mean field model for movement induced changes in the beta rhythm. *J Comput Neurosci* **43**, 143–158 (2017).

322. Grabska-Barwińska, A. & Zygierewicz, J. A model of event-related EEG synchronization changes in beta and gamma frequency bands. *J. Theor. Biol.* **238**, 901–913 (2006).

323. Chun, M. M., Golomb, J. D. & Turk-Browne, N. B. A taxonomy of external and internal attention. *Annu Rev Psychol* **62**, 73–101 (2011).

324. Frey, J. N., Ruhnau, P. & Weisz, N. Not so different after all: The same oscillatory processes support different types of attention. *Brain Res* **1626**, 183–197 (2015).

325. Clayton, M. S., Yeung, N. & Cohen Kadosh, R. The roles of cortical oscillations in sustained attention. *Trends Cogn Sci* **19**, 188–195 (2015).

326. Hsieh, L.-T. & Ranganath, C. Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. *Neuroimage* **85 Pt 2**, 721–729 (2014).

327. Sauseng, P., Griesmayr, B., Freunberger, R. & Klimesch, W. Control mechanisms in working memory: a possible function of EEG theta oscillations. *Neurosci Biobehav Rev* **34**, 1015–1022 (2010).

328. Ishii, R. *et al.* Frontal midline theta rhythm and gamma power changes during focused attention on mental calculation: an MEG beamformer analysis. *Front Hum Neurosci* **8**, 406 (2014).

329. Mizuhara, H., Wang, L.-Q., Kobayashi, K. & Yamaguchi, Y. A long-range cortical network emerging with theta oscillation in a mental task. *Neuroreport* **15**, 1233–1238 (2004).

330. Mizuhara, H. & Yamaguchi, Y. Human cortical circuits for central executive function emerge by theta phase synchronization. *Neuroimage* **36**, 232–244 (2007).

331. Cavanagh, J. F. & Frank, M. J. Frontal theta as a mechanism for cognitive control. *Trends Cogn Sci* **18**, 414–421 (2014).

332. Cona, G. *et al.* Theta and alpha oscillations as signatures of internal and external attention to delayed intentions: A magnetoencephalography (MEG) study. *Neuroimage* **205**, 116295 (2020).

333. Klimesch, W. α-band oscillations, attention, and controlled access to stored information. *Trends Cogn Sci* **16**, 606–617 (2012).

334. Doesburg, S. M., Bedo, N. & Ward, L. M. Top-down alpha oscillatory network interactions during visuospatial attention orienting. *Neuroimage* **132**, 512–519 (2016).

335. Rihs, T. A., Michel, C. M. & Thut, G. Mechanisms of selective inhibition in visual spatial attention are indexed by alpha-band EEG synchronization. *Eur J Neurosci* **25**, 603–610 (2007).

336. Thut, G., Nietzel, A., Brandt, S. A. & Pascual-Leone, A. Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J Neurosci* **26**, 9494–9502 (2006).

337. Worden, M. S., Foxe, J. J., Wang, N. & Simpson, G. V. Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *J Neurosci* **20**, RC63 (2000).

338. Snyder, A. C. & Foxe, J. J. Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: a high-density electrical mapping study. *J Neurosci* **30**, 4024–4032 (2010).

339. Anderson, K. L. & Ding, M. Attentional modulation of the somatosensory mu rhythm. *Neuroscience* **180**, 165–180 (2011).

340. Foxe, J. J., Simpson, G. V. & Ahlfors, S. P. Parieto-occipital approximately 10 Hz activity reflects anticipatory state of visual attention mechanisms. *Neuroreport* **9**, 3929–3933 (1998).

341. Busch, N. A. & Herrmann, C. S. Object-load and feature-load modulate EEG in a short-term memory task. *Neuroreport* **14**, 1721–1724 (2003).

342. Jensen, O., Gelfand, J., Kounios, J. & Lisman, J. E. Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task. *Cereb Cortex* **12**, 877–882 (2002).

343. Sauseng, P. *et al.* Brain oscillatory substrates of visual short-term memory capacity. *Curr Biol* **19**, 1846–1852 (2009).

344. Vissers, M. E., van Driel, J. & Slagter, H. A. Proactive, but Not Reactive, Distractor Filtering Relies on Local Modulation of Alpha Oscillatory Activity. *J Cogn Neurosci* **28**, 1964–1979 (2016).

345. Bonnefond, M. & Jensen, O. Alpha oscillations serve to protect working memory maintenance against anticipated distracters. *Curr Biol* **22**, 1969–1974 (2012).

346. Wang, C., Rajagovindan, R., Han, S.-M. & Ding, M. Top-Down Control of Visual Alpha Oscillations: Sources of Control Signals and Their Mechanisms of Action. *Front Hum Neurosci* **10**, 15 (2016).

347. Kitaura, Y. *et al.* Functional localization and effective connectivity of cortical theta and alpha oscillatory activity during an attention task. *Clin Neurophysiol Pract* **2**, 193–200 (2017).

348. Bradley, M. M. & Lang, P. J. The International Affective Picture System (IAPS) in the study of emotion and attention. in *Handbook of emotion elicitation and assessment* 29–46 (Oxford University Press, 2007).

349. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* **134**, 9–21 (2004).

350. Pion-Tonachini, L., Kreutz-Delgado, K. & Makeig, S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage* **198**, 181–197 (2019).

351. Klimesch, W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res Brain Res Rev* **29**, 169–195 (1999).

352. Klimesch, W., Schimke, H., Ladurner, G. & Pfurtscheller, G. Alpha frequency and memory performance. *Journal of Psychophysiology* **4**, 381–390 (1990).

353. Corcoran, A. W., Alday, P. M., Schlesewsky, M. & Bornkessel-Schlesewsky, I. Toward a reliable, automated method of individual alpha frequency (IAF) quantification. *Psychophysiology* **55**, e13064 (2018).

354. Pascual-Marqui, R. D. *et al.* Assessing interactions in the brain with exact low-resolution electromagnetic tomography. *Philos Trans A Math Phys Eng Sci* **369**, 3768–3784 (2011).

355. Conway, B. R. The Organization and Operation of Inferior Temporal Cortex. *Annu Rev Vis Sci* **4**, 381–402 (2018).

356. Sato, J. *et al.* Alpha keeps it together: Alpha oscillatory synchrony underlies working memory maintenance in young children. *Dev Cogn Neurosci* **34**, 114–123 (2018).

357. Zanto, T. P., Rubens, M. T., Thangavel, A. & Gazzaley, A. Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nat Neurosci* **14**, 656–661 (2011).

358. Sammer, G. *et al.* Relationship between regional hemodynamic activity and simultaneously recorded EEG-theta associated with mental arithmetic-induced workload. *Hum Brain Mapp* **28**, 793–803 (2007).

359. Cavanna, A. E. & Trimble, M. R. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* **129**, 564–583 (2006).

360. Fuentemilla, L., Barnes, G. R., Düzel, E. & Levine, B. Theta oscillations orchestrate medial temporal lobe and neocortex in remembering autobiographical memories. *NeuroImage* **85**, 730–737 (2014).

361. Costigan, A. G. *et al.* Neurochemical correlates of scene processing in the precuneus/posterior cingulate cortex: A multimodal fMRI and 1 H-MRS study. *Hum Brain Mapp* **40**, 2884–2898 (2019).

362. Pflugshaupt, T. *et al.* Bottom-up Visual Integration in the Medial Parietal Lobe. *Cereb Cortex* **26**, 943–949 (2016).

363. Granger, C. W. J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* **37**, 424–438 (1969).

364. Geweke, J. Measurement of Linear Dependence and Feedback Between Multiple Time Series. *Journal of the American Statistical Association* **77**, 304–313 (1982).

365. Ding, M., Chen, Y. & Bressler, S. L. Granger Causality: Basic Theory and Application to Neuroscience. *arXiv:q-bio/0608035* (2006).

366. Brovelli, A. *et al.* Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. *Proc Natl Acad Sci U S A* **101**, 9849–9854 (2004).

367. Chicharro, D. On the spectral formulation of Granger causality. *Biol Cybern* **105**, 331–347 (2011).

368. Buzsáki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nat Rev Neurosci* **13**, 407–420 (2012).

369. Cona, F., Zavaglia, M., Massimini, M., Rosanova, M. & Ursino, M. A neural mass model of interconnected regions simulates rhythm propagation observed via TMS-EEG. *Neuroimage* **57**, 1045–1058 (2011).

370. Ursino, M., Ricci, G. & Magosso, E. Transfer Entropy as a Measure of Brain Connectivity: A Critical Analysis With the Help of Neural Mass Models. *Front Comput Neurosci* **14**, 45 (2020).

371. Ursino, M. & Zavaglia, M. Modeling analysis of the relationship between EEG rhythms and connectivity among different neural populations. *J Integr Neurosci* **6**, 597–623 (2007).

372. Nichols, T. E. & Holmes, A. P. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp* **15**, 1–25 (2002).

373. Sauseng, P., Hoppe, J., Klimesch, W., Gerloff, C. & Hummel, F. C. Dissociation of sustained attention from central executive functions: local activity and interregional connectivity in the theta range. *Eur J Neurosci* **25**, 587–593 (2007).

374. Bastiaansen, M. C. M., Posthuma, D., Groot, P. F. C. & de Geus, E. J. C. Event-related alpha and theta responses in a visuo-spatial working memory task. *Clin Neurophysiol* **113**, 1882–1893 (2002).

375. Han, H.-B., Lee, K. E. & Choi, J. H. Functional Dissociation of θ Oscillations in the Frontal and Visual Cortices and Their Long-Range Network during Sustained Attention. *eNeuro* **6**, (2019).

376. Harris, A. M., Dux, P. E., Jones, C. N. & Mattingley, J. B. Distinct roles of theta and alpha oscillations in the involuntary capture of goal-directed attention. *Neuroimage* **152**, 171–183 (2017).

377. Kawasaki, M. & Yamaguchi, Y. Effects of subjective preference of colors on attention-related occipital theta oscillations. *Neuroimage* **59**, 808–814 (2012).

378. Magosso, E., De Crescenzio, F., Ricci, G., Piastra, S. & Ursino, M. EEG Alpha Power Is Modulated by Attentional Changes during Cognitive Tasks and Virtual Reality Immersion. *Comput Intell Neurosci* **2019**, 7051079 (2019).

379. Schroeder, S. C. Y., Ball, F. & Busch, N. A. The role of alpha oscillations in distractor inhibition during memory retention. *Eur J Neurosci* **48**, 2516–2526 (2018).

380. Dixon, M. L., Fox, K. C. R. & Christoff, K. A framework for understanding the relationship between externally and internally directed cognition. *Neuropsychologia* **62**, 321–330 (2014).

381. Gazzaley, A. & Nobre, A. C. Top-down modulation: bridging selective attention and working memory. *Trends Cogn Sci* **16**, 129–135 (2012).

382. Clapp, W. C., Rubens, M. T. & Gazzaley, A. Mechanisms of working memory disruption by external interference. *Cereb Cortex* **20**, 859–872 (2010).

383. Gazzaley, A. *et al.* Functional interactions between prefrontal and visual association cortex contribute to top-down modulation of visual processing. *Cereb Cortex* **17 Suppl 1**, i125-135 (2007).

384. Raichle, M. E. *et al.* A default mode of brain function. *PNAS* **98**, 676–682 (2001).

385. Hillebrand, A. *et al.* Direction of information flow in large-scale resting-state networks is frequency-dependent. *Proc Natl Acad Sci U S A* **113**, 3867–3872 (2016).

386. Johnson, E. L. *et al.* Bidirectional Frontoparietal Oscillatory Systems Support Working Memory. *Curr Biol* **27**, 1829-1835.e4 (2017).

387. Luo, Q. *et al.* Theta band activity in response to emotional expressions and its relationship with gamma band activity as revealed by MEG and advanced beamformer source imaging. *Front Hum Neurosci* **7**, 940 (2013).

388. Duits, P. *et al.* Updated Meta-Analysis of Classical Fear Conditioning in the Anxiety Disorders. *Depression and Anxiety* **32**, 239–253 (2015).

389. Starita, F., Làdavas, E. & di Pellegrino, G. Reduced anticipation of negative emotional events in alexithymia. *Sci Rep* **6**, 27664 (2016).

390. Starita, F., Kroes, M. C. W., Davachi, L., Phelps, E. A. & Dunsmoor, J. E. Threat learning promotes generalization of episodic memory. *J Exp Psychol Gen* **148**, 1426–1434 (2019).

391. Bechara, A. *et al.* Double Dissociation of Conditioning and Declarative Knowledge Relative to the Amygdala and Hippocampus in Humans. *Science* **269**, 1115–1118 (1995).

392. Bechara, A., Damasio, H., Damasio, A. R. & Lee, G. P. Different Contributions of the Human Amygdala and Ventromedial Prefrontal Cortex to Decision-Making. *J. Neurosci.* **19**, 5473–5481 (1999).

393. LaBar, K. S., LeDoux, J. E., Spencer, D. D. & Phelps, E. A. Impaired fear conditioning following unilateral temporal lobectomy in humans. *J. Neurosci.* **15**, 6846–6855 (1995).

394. LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E. & Phelps, E. A. Human Amygdala Activation during Conditioned Fear Acquisition and Extinction: a Mixed-Trial fMRI Study. *Neuron* **20**, 937–945 (1998).

395. Knight, D. C., Nguyen, H. T. & Bandettini, P. A. The role of the human amygdala in the production of conditioned fear responses. *NeuroImage* **26**, 1193–1200 (2005).

396. Cheng, D. T., Knight, D. C., Smith, C. N. & Helmstetter, F. J. Human amygdala activity during the expression of fear responses. *Behavioral Neuroscience* **120**, 1187–1195 (2006).

397. Fullana, M. A. *et al.* Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Mol Psychiatry* **21**, 500–508 (2016).

398. Bertini, C. *et al.* Fear-specific enhancement of tactile perception is disrupted after amygdala lesion. *Journal of Neuropsychology* **14**, 165–182 (2020).

399. Battaglia, S., Garofalo, S., Pellegrino, G. di & Starita, F. Revaluing the Role of vmPFC in the Acquisition of Pavlovian Threat Conditioning in Humans. *J. Neurosci.* **40**, 8491–8500 (2020).

400. Sperl, M. F. J. *et al.* Fear Extinction Recall Modulates Human Frontomedial Theta and Amygdala Activity. *Cereb Cortex* **29**, 701–715 (2019).

401. Likhtik, E., Stujenske, J. M., Topiwala, M. A., Harris, A. Z. & Gordon, J. A. Prefrontal entrainment of amygdala activity signals safety in learned fear and innate anxiety. *Nat Neurosci* **17**, 106–113 (2014).

402. Karalis, N. *et al.* 4-Hz oscillations synchronize prefrontal-amygdala circuits during fear behavior. *Nat Neurosci* **19**, 605–612 (2016).

403. Taub, A. H., Perets, R., Kahana, E. & Paz, R. Oscillations Synchronize Amygdala-to-Prefrontal Primate Circuits during Aversive Learning. *Neuron* **97**, 291-298.e3 (2018).

404. Knight, D. C., Cheng, D. T., Smith, C. N., Stein, E. A. & Helmstetter, F. J. Neural substrates mediating human delay and trace fear conditioning. *J Neurosci* **24**, 218–228 (2004).

405. Milad, M. R. *et al.* A role for the human dorsal anterior cingulate cortex in fear expression. *Biol Psychiatry* **62**, 1191–1194 (2007).

406. Phelps, E. A., Delgado, M. R., Nearing, K. I. & LeDoux, J. E. Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* **43**, 897–905 (2004).

407. Chien, J. H. *et al.* Oscillatory EEG activity induced by conditioning stimuli during fear conditioning reflects Salience and Valence of these stimuli more than Expectancy. *Neuroscience* **346**, 81–93 (2017).

408. Yin, S. *et al.* Fear conditioning prompts sparser representations of conditioned threat in primary visual cortex. *Soc Cogn Affect Neurosci* **15**, 950–964 (2020).

409. Bacigalupo, F. & Luck, S. J. Alpha-Band EEG Suppression as a Neural Marker of Sustained Attentional Engagement to Conditioned Threat Stimuli. *Soc Cogn Affect Neurosci* nsac029 (2022) doi:10.1093/scan/nsac029.

410. Babiloni, C. *et al.* Cortical alpha rhythms are related to the anticipation of sensorimotor interaction between painful stimuli and movements: a high-resolution EEG study. *J Pain* **9**, 902–911 (2008).

411. Garofalo, S., Maier, M. E. & di Pellegrino, G. Mediofrontal negativity signals unexpected omission of aversive events. *Sci Rep* **4**, 4816 (2014).

412. Garofalo, S., Timmermann, C., Battaglia, S., Maier, M. E. & di Pellegrino, G. Mediofrontal Negativity Signals Unexpected Timing of Salient Outcomes. *Journal of Cognitive Neuroscience* **29**, 718–727 (2017).

413. Magosso, E., Forcelli, V., Garofalo, S., di Pellegrino, G. & Ursino, M. Event-related brain potential signaling unexpected timing of feedback: A source localization analysis. *Annu Int Conf IEEE Eng Med Biol Soc* **2015**, 618–621 (2015).

414. Çalışkan, G. & Stork, O. Hippocampal network oscillations at the interplay between innate anxiety and learned fear. *Psychopharmacology (Berl)* **236**, 321–338 (2019).

415. Marin, M.-F. *et al.* Skin Conductance Responses and Neural Activations During Fear Conditioning and Extinction Recall Across Anxiety Disorders. *JAMA Psychiatry* **74**, 622–631 (2017).

416. Trenado, C., Pedroarena-Leal, N., Cif, L., Nitsche, M. & Ruge, D. Neural Oscillatory Correlates for Conditioning and Extinction of Fear. *Biomedicines* **6**, E49 (2018).

417. Etkin, A., Egner, T. & Kalisch, R. Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn Sci* **15**, 85–93 (2011).

418. Milad, M. R. & Quirk, G. J. Fear extinction as a model for translational neuroscience: ten years of progress. *Annu Rev Psychol* **63**, 129–151 (2012).

419. Schirru, M., Véronneau-Veilleux, F., Nekka, F. & Ursino, M. Phasic Dopamine Changes and Hebbian Mechanisms during Probabilistic Reversal Learning in Striatal Circuits: A Computational Study. *Int J Mol Sci* **23**, 3452 (2022).

420. Schiller, D., Levy, I., Niv, Y., LeDoux, J. E. & Phelps, E. A. From fear to safety and back: reversal of fear in the human brain. *J Neurosci* **28**, 11517–11525 (2008).

421. Rolls, E. T. The functions of the orbitofrontal cortex. *Brain and Cognition* **55**, 11–29 (2004).

422. Kim, H., Shimojo, S. & O'Doherty, J. P. Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol* **4**, e233 (2006).

423. Morris, J. S. & Dolan, R. J. Dissociable amygdala and orbitofrontal responses during reversal fear conditioning. *Neuroimage* **22**, 372–380 (2004).

424. Hudson, M. *et al.* Dissociable neural systems for unconditioned acute and sustained fear. *Neuroimage* **216**, 116522 (2020).

425. Starita, F. *et al.* Theta and alpha power track the acquisition and reversal of threat predictions and correlate with skin conductance response. *Psychophysiology* (in press).

426. Starita, F. & di Pellegrino, G. Alexithymia and the Reduced Ability to Represent the Value of Aversively Motivated Actions. *Frontiers in Psychology* **9**, (2018).

427. Starita, F., Pietrelli, M., Bertini, C. & di Pellegrino, G. Aberrant reward prediction error during Pavlovian appetitive learning in alexithymia. *Social Cognitive and Affective Neuroscience* **14**, 1119–1129 (2019).

428. Effting, M. & Kindt, M. Contextual control of human fear associations in a renewal paradigm. *Behaviour Research and Therapy* **45**, 2002–2018 (2007).

429. Krypotos, A.-M., Effting, M., Arnaudova, I., Kindt, M. & Beckers, T. Avoided by Association: Acquisition, Extinction, and Renewal of Avoidance Tendencies Toward Conditioned Fear Stimuli. *Clinical Psychological Science* **2**, 336–343 (2014).

430. Krypotos, A.-M., Effting, M., Kindt, M. & Beckers, T. Avoidance learning: a review of theoretical models and recent developments. *Frontiers in Behavioral Neuroscience* **9**, (2015).

431. Starita, F., Garofalo, S., Dalbagno, D., Degni, L. A. E. & di Pellegrino, G. Pavlovian threat learning shapes the kinematics of action. *Frontiers in Psychology* **13**, (2022).

432. Stemerding, L. E., van Ast, V. A., Gerlicher, A. M. V. & Kindt, M. Pupil dilation and skin conductance as measures of prediction error in aversive learning. *Behaviour Research and Therapy* **157**, 104164 (2022).

433. Mathôt, S., Schreij, D. & Theeuwes, J. OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behav Res Methods* **44**, 314–324 (2012).

434. Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M. & Robbins, K. A. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front Neuroinform* **9**, 16 (2015).

435. da Cruz, J. R., Chicherov, V., Herzog, M. H. & Figueiredo, P. An automatic pre-processing pipeline for EEG analysis (APP) based on robust statistics. *Clin Neurophysiol* **129**, 1427–1437 (2018).

436. Geweke, J. Measurement of Linear Dependence and Feedback between Multiple Time Series. *Journal of the American Statistical Association* **77**, 304–313 (1982).

437. Geweke, J. F. Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association* **79**, 907–915 (1984).

438. Magosso, E., Ricci, G. & Ursino, M. Alpha and theta mechanisms operating in internal-external attention competition. *J Integr Neurosci* **20**, 1–19 (2021).

439. Tarasi, L., Magosso, E., Ricci, G., Ursino, M. & Romei, V. The Directionality of Fronto-Posterior Brain Connectivity Is Associated with the Degree of Individual Autistic Traits. *Brain Sciences* **11**, 1443 (2021).

440. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995).

441. Lai, C.-H. Fear Network Model in Panic Disorder: The Past and the Future. *Psychiatry Investig* **16**, 16–26 (2019).

442. Tovote, P., Fadok, J. P. & Lüthi, A. Neuronal circuits for fear and anxiety. *Nat Rev Neurosci* **16**, 317–331 (2015).

443. Ridderbusch, I. C. *et al.* Neural adaptation of cingulate and insular activity during delayed fear extinction: A replicable pattern across assessment sites and repeated measurements. *Neuroimage* **237**, 118157 (2021).

444. Bierwirth, P., Sperl, M. F. J., Antov, M. I. & Stockhorst, U. Prefrontal Theta Oscillations Are Modulated by Estradiol Status During Fear Recall and Extinction Recall. *Biol Psychiatry Cogn Neurosci Neuroimaging* **6**, 1071–1080 (2021).

445. Feng, P., Feng, T., Chen, Z. & Lei, X. Memory consolidation of fear conditioning: bi-stable amygdala connectivity with dorsal anterior cingulate and medial prefrontal cortex. *Soc Cogn Affect Neurosci* **9**, 1730–1737 (2014).

446. Toyoda, H. *et al.* Interplay of amygdala and cingulate plasticity in emotional fear. *Neural Plast* **2011**, 813749 (2011).

447. Vogt, B. *Cingulate Neurobiology and Disease*. (OUP Oxford, 2009).

448. Pape, H.-C., Narayanan, R. T., Smid, J., Stork, O. & Seidenbecher, T. Theta activity in neurons and networks of the amygdala related to long-term fear memory. *Hippocampus* **15**, 874–880 (2005).

449. Chen, S. *et al.* Theta oscillations synchronize human medial prefrontal cortex and amygdala during fear learning. *Sci Adv* **7**, eabf4198 (2021).

450. Verbeke, P., Ergo, K., De Loof, E. & Verguts, T. Learning to Synchronize: Midfrontal Theta Dynamics during Rule Switching. *J Neurosci* **41**, 1516–1528 (2021).

451. Vogt, B. A. Pain and emotion interactions in subregions of the cingulate gyrus. *Nat Rev Neurosci* **6**, 533–544 (2005).

452. Vogt, B. A. & Pandya, D. N. Cingulate cortex of the rhesus monkey: II. Cortical afferents. *J Comp Neurol* **262**, 271–289 (1987).

453. Shackman, A. J. *et al.* The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat Rev Neurosci* **12**, 154–167 (2011).

454. Klimesch, W., Sauseng, P. & Hanslmayr, S. EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res Rev* **53**, 63–88 (2007).

455. Ursino, M., Ricci, G. & Magosso, E. Transfer Entropy as a Measure of Brain Connectivity: A Critical Analysis With the Help of Neural Mass Models. *Front Comput Neurosci* **14**, 45 (2020).

456. Ursino, M. *et al.* A Novel Method to Assess Motor Cortex Connectivity and Event Related Desynchronization Based on Mass Models. *Brain Sci* **11**, 1479 (2021).

457. Courtin, J., Karalis, N., Gonzalez-Campo, C., Wurtz, H. & Herry, C. Persistence of amygdala gamma oscillations during extinction learning predicts spontaneous fear recovery. *Neurobiol Learn Mem* **113**, 82–89 (2014).

458. Fenton, G. E., Halliday, D. M., Mason, R., Bredy, T. W. & Stevenson, C. W. Sex differences in learned fear expression and extinction involve altered gamma oscillations in medial prefrontal cortex. *Neurobiol Learn Mem* **135**, 66–72 (2016).

459. Stujenske, J. M., Likhtik, E., Topiwala, M. A. & Gordon, J. A. Fear and safety engage competing patterns of theta-gamma coupling in the basolateral amygdala. *Neuron* **83**, 919–933 (2014).

460. Milad, M. R. *et al.* Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biol Psychiatry* **62**, 446–454 (2007).

461. Kana, R. K., Uddin, L. Q., Kenet, T., Chugani, D. & Müller, R.-A. Brain connectivity in autism. *Frontiers in Human Neuroscience* **8**, (2014).

462. Maximo, J. O., Cadena, E. J. & Kana, R. K. The implications of brain connectivity in the neuropsychology of autism. *Neuropsychol Rev* **24**, 16–31 (2014).

463. Mohammad-Rezazadeh, I., Frohlich, J., Loo, S. K. & Jeste, S. S. Brain connectivity in autism spectrum disorder. *Curr Opin Neurol* **29**, 137–147 (2016).

464. Carroll, L. *et al.* Autism Spectrum Disorders: Multiple Routes to, and Multiple Consequences of, Abnormal Synaptic Function and Connectivity. *Neuroscientist* **27**, 10–29 (2021).

465. Just, M. A., Keller, T. A., Malave, V. L., Kana, R. K. & Varma, S. Autism as a neural systems disorder: a theory of frontal-posterior underconnectivity. *Neurosci Biobehav Rev* **36**, 1292–1313 (2012).

466. Abrams, D. A. *et al.* Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proc Natl Acad Sci U S A* **110**, 12060–12065 (2013).

467. Delbruck, E., Yang, M., Yassine, A. & Grossman, E. D. Functional connectivity in ASD: Atypical pathways in brain networks supporting action observation and joint attention. *Brain Research* **1706**, 157–165 (2019).

468. Nair, A., Treiber, J. M., Shukla, D. K., Shih, P. & Müller, R.-A. Impaired thalamocortical connectivity in autism spectrum disorder: a study of functional and anatomical connectivity. *Brain* **136**, 1942–1955 (2013).

469. Murphy, E. R., Foss-Feig, J., Kenworthy, L., Gaillard, W. D. & Vaidya, C. J. Atypical Functional Connectivity of the Amygdala in Childhood Autism Spectrum Disorders during Spontaneous Attention to Eye-Gaze. *Autism Res Treat* **2012**, 652408 (2012).

470. Uddin, L. Q. *et al.* Salience network-based classification and prediction of symptom severity in children with autism. *JAMA Psychiatry* **70**, 869–879 (2013).

471. Fu, Z. *et al.* Transient increased thalamic-sensory connectivity and decreased whole-brain dynamism in autism. *Neuroimage* **190**, 191–204 (2019).

472. Di Martino, A. *et al.* Aberrant striatal functional connectivity in children with autism. *Biol Psychiatry* **69**, 847–856 (2011).

473. Lynch, C. J. *et al.* Default mode network in childhood autism: posteromedial cortex heterogeneity and relationship with social deficits. *Biol Psychiatry* **74**, 212–219 (2013).

474. Abbott, A. E. *et al.* Repetitive behaviors in autism are linked to imbalance of corticostriatal connectivity: a functional connectivity MRI study. *Soc Cogn Affect Neurosci* **13**, 32–42 (2018).

475. O'Reilly, C., Lewis, J. D. & Elsabbagh, M. Is functional brain connectivity atypical in autism? A systematic review of EEG and MEG studies. *PLOS ONE* **12**, e0175870 (2017).

476. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* **36**, 181–204 (2013).

477. Pellicano, E. & Burr, D. When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences* **16**, 504–510 (2012).

478. Van de Cruys, S. *et al.* Precise minds in uncertain worlds: predictive coding in autism. *Psychol Rev* **121**, 649–675 (2014).

479. Sinha, P. *et al.* Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences* **111**, 15220–15225 (2014).

480. Skewes, J. C., Jegindø, E.-M. & Gebauer, L. Perceptual inference and autistic traits. *Autism* **19**, 301–307 (2015).

481. Crespi, B. & Dinsdale, N. Autism and psychosis as diametrical disorders of embodiment. *Evolution, Medicine, and Public Health* **2019**, 121–138 (2019).

482. Mottron, L., Dawson, M., Soulières, I., Hubert, B. & Burack, J. Enhanced Perceptual Functioning in Autism: An Update, and Eight Principles of Autistic Perception. *J Autism Dev Disord* **36**, 27–43 (2006).

483. Cribb, S. J., Olaithe, M., Di Lorenzo, R., Dunlop, P. D. & Maybery, M. T. Embedded Figures Test Performance in the Broader Autism Phenotype: A Meta-analysis. *J Autism Dev Disord* **46**, 2924–2939 (2016).

484. Tarasi, L. *et al.* Predictive waves in the autism-schizophrenia continuum: A novel biobehavioral model. *Neuroscience & Biobehavioral Reviews* **132**, 1–22 (2022).

485. Basar-Eroglu, C. *et al.* Working memory related gamma oscillations in schizophrenia patients. *International Journal of Psychophysiology* **64**, 39–45 (2007).

486. Clayton, M. S., Yeung, N. & Cohen Kadosh, R. The roles of cortical oscillations in sustained attention. *Trends in Cognitive Sciences* **19**, 188–195 (2015).

487. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord* **31**, 5–17 (2001).

488. Bralten, J. *et al.* Autism spectrum disorders and autistic traits share genetics and biology. *Mol Psychiatry* **23**, 1205–1212 (2018).

489. van Wijk, B. C. M., Stam, C. J. & Daffertshofer, A. Comparing brain networks of different size and connectivity density using graph theory. *PLoS One* **5**, e13701 (2010).

490. Minati, L., Varotto, G., D'Incerti, L., Panzica, F. & Chan, D. From brain topography to brain topology: relevance of graph theory to functional neuroscience. *Neuroreport* **24**, 536–543 (2013).

491. Farahani, F. V., Karwowski, W. & Lighthall, N. R. Application of Graph Theory for Identifying Connectivity Patterns in Human Brain Networks: A Systematic Review. *Front Neurosci* **13**, 585 (2019).

492. Ruta, L., Mazzone, D., Mazzone, L., Wheelwright, S. & Baron-Cohen, S. The Autism-Spectrum Quotient—Italian Version: A Cross-Cultural Confirmation of the Broader Autism Phenotype. *J Autism Dev Disord* **42**, 625–633 (2012).

493. Ruzich, E. *et al.* Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular Autism* **6**, 2 (2015).

494. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* **134**, 9–21 (2004).

495. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience* **2011**, e879716 (2011).

496. Gramfort, A., Papadopoulo, T., Olivi, E. & Clerc, M. OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomed Eng Online* **9**, 45 (2010).

497. Pascual-Marqui, R. D. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find Exp Clin Pharmacol* **24 Suppl D**, 5–12 (2002).

498. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).

499. Stokes, P. A. & Purdon, P. L. A study of problems encountered in Granger causality analysis from a neuroscience perspective. *Proc Natl Acad Sci U S A* **114**, E7063–E7072 (2017).

500. Deshpande, G., LaConte, S., James, G. A., Peltier, S. & Hu, X. Multivariate Granger causality analysis of fMRI data. *Hum Brain Mapp* **30**, 1361–1373 (2009).

501. Sporns, O. Graph theory methods: applications in brain networks. *Dialogues Clin Neurosci* **20**, 111–121 (2018).

502. Deshpande, G. & Hu, X. Investigating effective brain connectivity from fMRI data: past findings and current issues with reference to Granger causality analysis. *Brain Connect* **2**, 235–245 (2012).

503. Cekic, S., Grandjean, D. & Renaud, O. Time, frequency, and time-varying Granger-causality measures in neuroscience. *Stat Med* **37**, 1910–1931 (2018).

504. Alink, A. & Charest, I. Clinically relevant autistic traits predict greater reliance on detail for image recognition. *Sci Rep* **10**, 14239 (2020).

505. Seth, A. K., Barrett, A. B. & Barnett, L. Granger causality analysis in neuroscience and neuroimaging. *J Neurosci* **35**, 3293–3297 (2015).

506. Rudie, J. D. *et al.* Altered functional and structural brain network organization in autism. *Neuroimage Clin* **2**, 79–94 (2012).

507. Redcay, E. *et al.* Intrinsic functional network organization in high-functioning adolescents with autism spectrum disorder. *Front Hum Neurosci* **7**, 573 (2013).

508. You, X. *et al.* Atypical modulation of distant functional connectivity by cognitive state in children with Autism Spectrum Disorders. *Front Hum Neurosci* **7**, 482 (2013).

509. Keown, C. L. *et al.* Network organization is globally atypical in autism: A graph theory study of intrinsic functional connectivity. *Biol Psychiatry Cogn Neurosci Neuroimaging* **2**, 66–75 (2017).

510. Chen, L. *et al.* Changes in the topological organization of the default mode network in autism spectrum disorder. *Brain Imaging Behav* **15**, 1058–1067 (2021).

511. Barttfeld, P. *et al.* A big-world network in ASD: dynamical connectivity analysis reflects a deficit in long-range connections and an excess of short-range connections. *Neuropsychologia* **49**, 254–263 (2011).

512. Tsiaras, V. *et al.* Extracting biomarkers of autism from MEG resting-state functional connectivity networks. *Comput Biol Med* **41**, 1166–1177 (2011).

513. Boersma, M. *et al.* Disrupted functional brain networks in autistic toddlers. *Brain Connect* **3**, 41–49 (2013).

514. Peters, J. M. *et al.* Brain functional networks in syndromic and non-syndromic autism: a graph theoretical study of EEG connectivity. *BMC Med* **11**, 54 (2013).

515. Leung, R. C., Ye, A. X., Wong, S. M., Taylor, M. J. & Doesburg, S. M. Reduced beta connectivity during emotional face processing in adolescents with autism. *Mol Autism* **5**, 51 (2014).

516. Takahashi, T. *et al.* Band-specific atypical functional connectivity pattern in childhood autism spectrum disorder. *Clin Neurophysiol* **128**, 1457–1465 (2017).

517. Soma, D. *et al.* Atypical Resting State Functional Neural Network in Children With Autism Spectrum Disorder: Graph Theory Approach. *Front Psychiatry* **12**, 790234 (2021).

518. Sporns, O. & Zwi, J. D. The small world of the cerebral cortex. *Neuroinformatics* **2**, 145–162 (2004).

519. Jao Keehn, R. J. *et al.* Impaired downregulation of visual cortex during auditory processing is associated with autism symptomatology in children and adolescents with autism spectrum disorder. *Autism Research* **10**, 130–143 (2017).

520. Jao Keehn, R. J. *et al.* Atypical Local and Distal Patterns of Occipito-frontal Functional Connectivity are Related to Symptom Severity in Autism. *Cereb Cortex* **29**, 3319–3330 (2019).

521. Samson, F., Mottron, L., Soulières, I. & Zeffiro, T. A. Enhanced visual functioning in autism: An ALE meta-analysis. *Human Brain Mapping* **33**, 1553–1581 (2012).

522. Carter Leno, V., Tomlinson, S. B., Chang, S.-A. A., Naples, A. J. & McPartland, J. C. Resting-state alpha power is selectively associated with autistic traits reflecting behavioral rigidity. *Sci Rep* **8**, 11982 (2018).

523. Harris, C. *et al.* Unique features of stimulus-based probabilistic reversal learning. *Behav Neurosci* **135**, 550–570 (2021).

524. Aykan, S. *et al.* Right Anterior Theta Hypersynchrony as a Quantitative Measure Associated with Autistic Traits and K-Cl Cotransporter KCC2 Polymorphism. *J Autism Dev Disord* **52**, 61–72 (2022).

525. Hermundstad, A. M. *et al.* Structural foundations of resting-state and task-based functional connectivity in the human brain. *Proc Natl Acad Sci U S A* **110**, 6169–6174 (2013).

526. De Benedictis, A. *et al.* New insights in the homotopic and heterotopic connectivity of the frontal portion of the human corpus callosum revealed by microdissection and diffusion tractography. *Human Brain Mapping* **37**, 4718–4735 (2016).

527. Choi, I., Lee, J.-Y. & Lee, S.-H. Bottom-up and top-down modulation of multisensory integration. *Curr Opin Neurobiol* **52**, 115–122 (2018).

528. Yusuf, P. A. *et al.* Deficient Recurrent Cortical Processing in Congenital Deafness. *Front Syst Neurosci* **16**, 806142 (2022).

529. Marco, E. J., Hinkley, L. B. N., Hill, S. S. & Nagarajan, S. S. Sensory Processing in Autism: A Review of Neurophysiologic Findings. *Pediatr Res* **69**, 48R-54R (2011).

530. Robertson, C. E. & Baron-Cohen, S. Sensory perception in autism. *Nat Rev Neurosci* **18**, 671–684 (2017).