

An information theoretic learning framework based on Renyi's α entropy for brain effective connectivity estimation



Iván De La Pava Panche

Universidad Tecnológica de Pereira

This dissertation is submitted for the degree of
Doctor of Engineering

Supervisor

Andrés Marino Álvarez Meza, PhD.

Co-supervisors

Álvaro Ángel Orozco Gutiérrez, PhD.

Paula Marcela Herrera Gómez, PhD.

Committee

Sebastián Castaño-Candamil, PhD.

Rosairo Iodice Di Vita, PhD.

Mario Andrés Valderrama Manrique, PhD.

Date

November 29, 2021.

To my father, whose belief in me always made me want to do better.

Acknowledgements

First and foremost, I would like to thank my supervisor Andrés Marino Álvarez. Without his constant help and guidance this thesis would have never seen the light of day. Also, he was extremely patient with me throughout all of these years, and for that I'm very grateful.

I thank my co-supervisors Paula Marcela Herrera and Álvaro Ángel Orozco for their valuable insights on several aspects of the work presented in this document.

I also thank the Doctorate in Engineering program and the Automatic Research Group at Universidad Tecnológica de Pereira for their support during my studies.

Finally, I want to thank my family and all the people in my life who stood by me during this process. Their words of encouragement when the going was rough made all the difference in the world.

My doctoral studies were financially supported by the program “Doctorado Nacional en Empresa - Convocatoria 758 de 2016”, funded by Minciencias.

Abstract

The interactions among neural populations distributed across different brain regions are at the core of cognitive and perceptual processing. Therefore, the ability of studying the flow of information within networks of connected neural assemblies is of fundamental importance to understand such processes. In that regard, brain connectivity measures constitute a valuable tool in neuroscience. They allow assessing functional interactions among brain regions through directed or non-directed statistical dependencies estimated from neural time series. Transfer entropy (TE) is one such measure. It is an effective connectivity estimation approach based on information theory concepts and statistical causality premises. It has gained increasing attention in the literature because it can capture purely nonlinear directed interactions, and is model free. That is to say, it does not require an initial hypothesis about the interactions present in the data. These properties make it an especially convenient tool in exploratory analyses. However, like any information-theoretic quantity, TE is defined in terms of probability distributions that in practice need to be estimated from data. A challenging task, whose outcome can significantly affect the results of TE. Also, it lacks a standard spectral representation, so it cannot reveal the local frequency band characteristics of the interactions it detects.

In this thesis, we develop a data driven approach to transfer entropy estimation, based on kernel matrices, that sidesteps the need to obtain probability distributions from the neural time series. Then, we use our TE estimator to propose methodologies that allow capturing directed within-frequency phase to phase interactions and directed cross-frequency phase to amplitude couplings. We begin by expressing TE as a linear combination of marginal and joint Renyi entropies of order α , instead of following the standard definition based on Shannon entropies. Next, we use functionals defined on positive definite kernels matrices that approximate Renyi's α entropy to propose a TE estimator that avoids the intermediate step of probability estimation, termed $TE_{\kappa\alpha}$. It does so by computing TE directly from kernel matrices that, in turn, capture similarity relations among the data. Later, we propose a methodology to estimate TE between single pairs of instantaneous phase time series, termed $TE_{\kappa\alpha}^{\theta}$. Our approach combines the

previously introduced kernel-based TE estimator ($TE_{\kappa\alpha}$) with phase time series obtained by complex filtering the neural signals of interest. Our proposal overcomes the hurdles other single-trial TE estimators face when obtaining TE values from phase time series data, because it does not have to explicitly obtain their probability distributions. Finally, we introduce a novel methodology to estimate directed phase-amplitude interactions through TE, termed $TE_{\kappa\alpha}^{\theta\theta^s}$. Our proposal recasts the problem of detecting directed phase to amplitude couplings as that of estimating directed interactions between phase time series. Therefore, it ultimately allows assessing the cross-frequency interactions of interest through the previously developed $TE_{\kappa\alpha}^{\theta}$ methodology.

Throughout this dissertation, we test the proposed $TE_{\kappa\alpha}$, $TE_{\kappa\alpha}^{\theta}$, $TE_{\kappa\alpha}^{\theta\theta^s}$ approaches on simulated data exhibiting the types of directed interactions each of them is intended to capture. We also test them on electroencephalographic (EEG) data obtained under motor imagery and visuospatial working memory paradigms. For the latter case, we set up classification systems that aim to discriminate between the conditions in each paradigm, and that use as inputs relevant connectivities estimated through the proposed estimators. We then evaluate the classification performances in relation to the connectivity measures used to characterize the EEG signals. For both the simulated and real EEG data experiments, we employ alternative TE estimation approaches and other effective connectivity measures in the state-of-the-art as comparison benchmarks. Regarding the simulated data, obtained results show that the $TE_{\kappa\alpha}$ estimator detects the presence and direction of causal interactions, displaying robustness to varying noise levels and data sizes, as well as to multiple interaction delays in the same connected network. They also show that $TE_{\kappa\alpha}^{\theta}$ and $TE_{\kappa\alpha}^{\theta\theta^s}$ successfully estimate the directed interactions present in the data for the different scenarios tested, with statistically significant results at the target frequencies defined in the simulation models. For the motor imagery data, $TE_{\kappa\alpha}$ and $TE_{\kappa\alpha}^{\theta}$ capture discriminant patterns for the left and right-hand motor imagination tasks, following the temporal structure of the motor imagery paradigm, especially for a sub-group of subjects. On the other hand, the features extracted through $TE_{\kappa\alpha}^{\theta\theta^s}$ do not allow discriminating between the tasks. Finally, concerning the working memory data, the three approaches code discriminant directed connectivities linking the regions that are commonly associated with that cognitive process. Nonetheless, the best results are obtained with the $TE_{\kappa\alpha}^{\theta}$ and $TE_{\kappa\alpha}^{\theta\theta^s}$ -based features, which points to the key role of oscillatory activity in visuospatial working memory. Overall, the results obtained with the proposed TE estimators compare favorably with those of the alternative effective connectivity measures tested in our experiments.

As future lines of work, we discuss the viability of extending our kernel-based TE estimator to multivariate connectivity analysis, the possible effects of alternative parameter selection strategies, as well as the potential to improve the performance of the $\text{TE}_{\kappa\alpha}^{\theta\theta^c}$ approach by selecting kernel functions better suited to the nature of the time series involved in phase-amplitude interactions.

Contents

List of Figures	xii
List of Tables	xv
Notation	xvi
1 Preliminaries	1
1.1 Motivation	1
1.2 Problem statement	4
1.2.1 Probability distribution estimation as an intermediate step in TE computation	6
1.2.2 The lack of a spectral representation for TE	7
1.3 Theoretical background	9
1.3.1 Transfer entropy	9
1.3.2 Granger causality	11
1.3.3 Information theoretic learning from kernel matrices	12
1.4 Literature review on transfer entropy estimation	14
1.4.1 Transfer entropy in the frequency domain	17
1.5 Aims	22
1.5.1 General aim	22
1.5.2 Specific aims	22
1.6 Outline and contributions	23
1.6.1 Kernel-based Renyi's transfer entropy	24
1.6.2 Kernel-based Renyi's phase transfer entropy	24
1.6.3 Kernel-based Renyi's phase transfer entropy for the estimation of directed phase-amplitude interactions	25
1.7 EEG databases	26

1.7.1	Motor imagery	26
1.7.2	Working memory	28
1.8	Thesis structure	31
2	Kernel-based Renyi's transfer entropy	34
2.1	Kernel-based Renyi's transfer entropy	35
2.2	Experiments	36
2.2.1	VAR model	36
2.2.2	Modified linear Kus model	38
2.2.3	EEG data	39
2.2.4	Parameter selection	42
2.3	Results and discussion	43
2.3.1	VAR model	43
2.3.2	Modified linear Kus model	46
2.3.3	EEG data	49
2.3.4	Limitations	57
2.4	Summary	58
3	Kernel-based Renyi's phase transfer entropy	60
3.1	Kernel-based Renyi's phase transfer entropy	61
3.1.1	Phase-based effective connectivity estimation approaches considered in this chapter	61
3.2	Experiments	63
3.2.1	Neural mass models	63
3.2.2	EEG data	66
3.2.3	Parameter selection	67
3.3	Results and discussion	68
3.3.1	Neural mass models	68
3.3.2	EEG data	73
3.3.3	Limitations	82
3.4	Summary	83
4	Kernel-based Renyi's phase transfer entropy for the estimation of directed phase-amplitude interactions	84
4.1	Kernel-based Renyi's phase transfer entropy for the estimation of directed phase-amplitude interactions	85

4.1.1	Transfer entropy for directed phase-amplitude interactions	85
4.1.2	Cross-frequency directionality	85
4.1.3	Phase transfer entropy and directed phase-amplitude interactions	86
4.2	Experiments	88
4.2.1	Simulated phase-amplitude interactions	88
4.2.2	EEG data	90
4.2.3	Parameter selection	91
4.3	Results and discussion	92
4.3.1	Simulated phase-amplitude interactions	92
4.3.2	EEG data	94
4.3.3	Limitations	98
4.4	Summary	99
5	Final Remarks	100
5.1	Conclusions	100
5.2	Future work	103
5.3	Academic products	104
5.3.1	Journal papers	104
5.3.2	Conference papers	105
5.3.3	Conference presentations	105
	Appendix A Kernel methods and Renyi's entropy estimation	106
A.1	Reproducing kernel Hilbert spaces	106
A.1.1	Reproducing kernels	106
A.1.2	Kernel-based learning	107
A.2	Kernel-based estimation of Renyi's entropy	109
	Appendix B Surface Laplacian	113
	Appendix C Permutation testing	115
	Appendix D Kernel-based relevance analysis	117
	Appendix E Cao's criterion	120
	Appendix F Neural mass model equations	122
	References	125

List of Figures

1.1	Venn diagrams showing the linear combination of joint and marginal entropies associated with the definition of TE.	11
1.2	Summary of the main TE estimation methods discussed, and their drawbacks (highlighted in red).	16
1.3	Within-frequency, cross-frequency, and frequency-resolved information transfer.	17
1.4	An oscillation from a neural time series can be described in terms of its frequency, its amplitude envelope, and its phase.	18
1.5	Summary of the approaches used to estimate TE and other effective connectivity measures for within-frequency interactions, and their drawbacks (highlighted in red)	20
1.6	Graphical representation of the most common approach to capture directed phase-amplitude interactions through TE	21
1.7	Summary of the approaches used to estimate TE and other effective connectivity measures for directed phase-amplitude interactions, and their drawbacks (highlighted in red)	22
1.8	Schematic representation of the relationship between the aims of this thesis and the developed methodologies.	23
1.9	(A) Schematic representation of the MI protocol. (B) EEG channel montage used for the acquisition of the MI dataset. (C) Example of a 2 seconds long segmented EEG trial after data pre-processing (without surface Laplacian).	29
1.10	(A) Graphical representation of the change detection task for WM. (B) EEG channels selected from the montage used for the acquisition of the WM database. (C) Example of a 0.7 seconds long segmented EEG trial after data pre-processing (without surface Laplacian).	32

2.1	Schematic representation of our overall classification setup.	41
2.2	Accuracies in the detection of the preferred direction of information flow for synthetic data generated from a unidirectional bivariate autoregressive model of order 3.	44
2.3	Average accuracies in the detection of the preferred direction of information flow for synthetic data, generated from a unidirectional bivariate autoregressive model of order 3 under asymmetric noise level conditions.	45
2.4	Obtained results for the modified linear Kus model.	47
2.5	(A) Connections with statistically significant differences between the MI tasks for time windows 1 to 6 for subject 8. (B) Training and testing classification accuracies per time window for subject 8. (C) Connections with statistically significant differences between the MI tasks for time windows 1 to 6 for subject 9. (D) Training and testing classification accuracies per time window for subject 9.	50
2.6	Normalized average differences in the total information flow coming into each channel for the training set, for all subjects and time windows.	53
2.7	Topoplots of the average node (channel) relevance for two groups of subjects in the MI database.	55
2.8	Highest average classification accuracy for each subject in the WM database (mean \pm standard deviation).	56
2.9	Topoplot of the average nodal relevance for all the subjects in the WM database.	56
3.1	(A) Schematic representation of a neural mass model. (B) 1 s long unidirectionally coupled time series generated by the model. (C) Average power spectrums peaking in the α and lower β frequency bands.	64
3.2	Obtained results for the NMMs data for each of the phase-based effective connectivity estimation approaches considered.	69
3.3	Obtained results for the NMMs data for different values of the α parameter in the proposed $TE_{\kappa\alpha}^{\theta}$ approach.	72
3.4	Average classification accuracies, and their standard deviations, for all subjects in the MI database as a function of the number features selected to train the classifiers.	75

3.5	(A) Highest average classification accuracy for each subject in the MI database during the training-validation stage. (B) Accuracies obtained for each subject during the testing stage.	76
3.6	(A) Two-dimensional representation of the relevance vectors for each subject in the MI database obtained after applying t-SNE on \mathbf{gn} . (B) Groups identified by k-means.	77
3.7	Topoplots of the average node (channel) relevance for each group of clustered subjects and frequency band of interest in the MI database (see Figure 3.6). The arrows represent the most relevant connectivities for each group.	78
3.8	Average classification accuracies, and their standard deviations, for all subjects in the WM database as a function of the number features selected to train the classifiers.	79
3.9	Highest average classification accuracy for each subject in the WM database.	80
3.10	(A) Two-dimensional representation of the relevance vectors for each subject in the WM database obtained after applying t-SNE on \mathbf{gn} . (B) Groups identified by k-means.	80
3.11	Topoplots of the average node (channel) relevance for each group of clustered subjects and frequency band of interest in the WM database (see Figure 3.10). The arrows represent the most relevant connectivities for each group.	81
4.1	(A) Conventional approach to capture directed phase-amplitude interactions through TE. (B) Proposed approach to estimate directed phase-amplitude interactions through phase TE.	87
4.2	Schematic representation of the simulation model used to obtain data with directed phase-amplitude interactions.	89
4.3	$TE_{\kappa\alpha}^{\theta\theta^c}$ results for the simulated data in the case when $SNR = 3$	92
4.4	Results of the permutation tests carried out on the connectivity values estimated using CFD, $TE_{\kappa\alpha}^{\theta\theta^c}$ and $TE_{\kappa\alpha}^{\theta\theta^c}$, for the three noise conditions modeled.	93
4.5	Average classification accuracy for all subjects as a function of the percentage of selected features.	95
4.6	Highest average classification accuracy for each subject in the WM database.	96

4.7	Topoplots of the average nodal (channel) relevance for each frequency band pair tested. The arrows represent the most relevant connections. . .	98
A.1	Schematic representation of the idea behind kernel-based learning	108
A.2	Schematic representation of the elements involved in the data-driven, kernel-based estimation of Renyi's α entropy.	110

List of Tables

2.1	Average training accuracy [%] for the window (w) with the best performance.	51
2.2	Testing accuracy [%] for the window (w) with the best performance. . . .	51
2.3	Average training accuracy for CSP and CSP-related methods.	52
2.4	Average classification accuracies for the WM database for all the effective connectivity measures considered.	55
3.1	MI and WM classification results in terms of the average classification accuracy for all the effective connectivity measures considered.	73
4.1	Average classification accuracies for all subjects in the MI database. . . .	94

Notation

Mathematical notation

Generalities

X	Random variable
\mathbb{R}	Set of the real numbers
\mathbb{N}	Set of the natural numbers
T	Total number of data points

Operators

$\mathbb{E}\{\cdot\}$	Expected value
$H_S(\cdot)$	Shannon entropy
$D_{KL}(\cdot \cdot)$	Kullback-Leibler divergence
$TE(\cdot)$	Transfer entropy
$\text{var}\{\cdot\}$	Variance operator
$GC(\cdot)$	Granger causality
$I(\cdot)$	Amount of information
$H_\alpha(\cdot)$	Renyi's α entropy
$\text{tr}(\cdot)$	Trace of a matrix
$ \cdot _{\mathbb{F}}$	Frobenius norm
\circ	Hadamard product
$TE^\theta(\cdot)$	Phase transfer entropy
$TE^{\theta\varsigma}(\cdot)$	Transfer entropy for directed phase-amplitude interactions
$TE_{\kappa\alpha}(\cdot)$	Kernel-based Renyi's α transfer entropy
$TE_{\kappa\alpha}^\theta(\cdot)$	Kernel-based Renyi's α phase transfer entropy
$TE_{\kappa\alpha}^{\theta\theta\varsigma}(\cdot)$	Kernel-based Renyi's α phase transfer entropy for directed phase-amplitude interactions

Functions

$p(\cdot)$	Probability density/mass function
$\kappa(\cdot, \cdot)$	Kernel function

Vectors and matrices

\mathbf{x}	Real-valued time series
\mathbf{x}_t^d	Time-embbeded time series
$\boldsymbol{\theta}^x$	Instantaneous phase time series
$\boldsymbol{\varsigma}^x$	Instantaneous amplitude time series
\mathbf{K}	A kernel matrix formed by the evaluation of $\kappa(\cdot, \cdot)$
\mathbf{I}	The identity matrix

Abbreviations

TE	Transfer entropy
EEG	Electroencephalogram
BCI	Brain computer interface
ICA	Independent component analysis
GC	Granger causality
DCM	Dynamic causal modeling
GGC	Geweke-Granger causality
PDC	Partial directed coherence
DTF	Directed transfer function
VAR	Vector autoregressive model
CFC	Cross-frequency coupling
PSI	Phase slope index
CFD	Cross-frequency directionality
PAC	Phase amplitude coupling
ITL	Information theoretic learning
CPD	Cumulative probability density function
MI	Motor imagery
WM	Working memory
CKA	Centered kernel alignment
RBF	Radial basis function
ACT	Autocorrelation time
NMM	Neural mass model
SNR	Signal to noise ratio

Chapter 1

Preliminaries

1.1 Motivation

The dynamic interactions among distributed neural populations and brain areas underpin cognitive processing and complex behavior (Engel et al., 2013; Rathee et al., 2017). The concept above is addressed in theoretical neuroscience as the functional integration principle (Sakkalis, 2011). It complements and expands on earlier approaches for studying brain activity, which focused on assessing the degree to which an individual brain area specializes for a particular task (Mišić and Sporns, 2016). The functional integration of cerebral areas can be studied through brain connectivity measures, that are computed as directed or non-directed statistical dependencies between neural time series obtained through various neuroimaging techniques (Bastos and Schoffelen, 2016; Friston, 2011; Gilson et al., 2020; Sakkalis, 2011). Such brain connectivity analyses have revealed that functional interactions among brain areas are involved in multiple cognitive domains including reward processing, cognitive control, memory, attention, language and learning (Da Silva, 2009; Dimitriadis et al., 2016a; Fukuda et al., 2019; Mišić and Sporns, 2016; Wianda and Ross, 2019). Furthermore, anomalous connectivity can be linked to disrupted brain functionality, which leads to neurological disorders (Sridhar et al., 2017). Brain connectivity measures have also found practical applications in several areas, besides their role as fundamental tools in basic neuroscience research. They have been proposed as control signals in brain-computer interfaces (BCI) for motor rehabilitation, and in neuro-feedback systems that aim to achieve behavioral changes through the self-regulation of functional connections; also as biomarkers for neurological disorders such as epilepsy and attention-deficit hyperactivity disorder (Abbas et al., 2021; Duffy et al., 2017; Hassan and Wendling, 2018; Lenartowicz and Loo, 2014; Mišić and Sporns, 2016;

Rathee et al., 2017; Wang et al., 2021a; Wing et al., 2021).

The challenges associated with brain connectivity estimation from neural time series stem from two sources: In the first place, there are challenges associated with the nature of the neuroimaging technique employed to acquire the data under analysis. In particular, in the field of cognitive neuroscience estimating brain connectivity from scalp electroencephalographic (EEG) recordings poses a challenge. EEG signals can provide valuable information about the underlying functional brain networks with a high temporal resolution, which allows tracking their activity over short periods of time, like those of many cognitive tasks (Hassan and Wendling, 2018; Hassan et al., 2015). Also scalp EEG recording systems are non-invasive, relatively easy to use, and are less expensive than other neuroimaging systems (Hassan and Wendling, 2018; Lenartowicz and Loo, 2014). However, EEG signals are quasi-stationary, and stationarity assumptions can be easily violated, particularly during physical and mental activities, which poses a problem for many connectivity estimation strategies (Sakkalis, 2011). An additional issue is the acquisition by the surface electrodes of scalp EEG of mixed activity from more than one brain region, an effect known as volume conduction, which can result in spurious connectivity patterns (Cohen, 2015; Sakkalis, 2011). Moreover, EEG signals display technical and physiological artifacts, such as noise from other electrical sources and cardiac, muscle and eye-related artifacts, that must be taken care of before estimating connectivity (Debener et al., 2012; Mennes et al., 2010). Nonetheless, numerous approaches exist in the literature to deal with the aforementioned challenges, including windowing procedures to address non-stationarity (Cekic et al., 2018), spatial filtering and source reconstruction to attenuate or eliminate volume conduction effects (Cohen, 2015; Hassan and Wendling, 2018; Hassan et al., 2014), and filtering, thresholding, and transformations such as ICA (independent component analysis) for artifact elimination (Mennes et al., 2010).

In the second place, there are hurdles for brain connectivity estimation that arise depending on the type of interaction that the connectivity measures must capture in order to answer a given research question. Whether the target interactions are directed or non-directed, linear or non-linear, time-domain-based or frequency-domain-based, and in the latter case whether they are within-frequency or cross-frequency in nature, will both guide the selection of appropriate connectivity measures and bring about particular estimation issues. For instance, the use of directed connectivity measures based on the concept of Wiener causality requires assuming how far back in the past the interactions might be present (Cekic et al., 2018; Wibral et al., 2013). There can also be scenarios

where little is known about the nature of the neural interactions of interest. In this case, an apt connectivity measure is one suited for exploratory analyses, that is to say, a connectivity measure that is able to capture a range of interactions as wide as possible. Such is the case of transfer entropy (TE), an information theory-based measure of directed interactions between dynamical systems increasingly used in neuroscience (Pinzuti et al., 2020; Schreiber, 2000; Ursino et al., 2020; Vicente et al., 2011; Zhu et al., 2015). However, as detailed in section 1.2, TE has limitations regarding the need for probability distribution estimation as an intermediate step in its computation, and the lack of a standard spectral representation that allows revealing the frequency characteristics of the detected directed interactions.

In a local context, the Automatic Research Group, at Universidad Tecnológica de Pereira, has been working for years on the development of computer-assisted diagnosis systems for neurological disorders through the analysis of EEG signals coupled with machine learning methods. The group has also been working on brain connectivity based visualization systems that represent neural activity during cognitive tasks, and on the development of brain connectivity based neuro-feedback systems for the amelioration of cognitive abilities, particularly working memory, in association with the Instituto de Epilepsia y Parkinson del Eje Cafetero, a leading regional health care institution, through the projects "Sistema de visualización de conectividad cerebral efectiva utilizando representaciones kernel orientado a tareas de memoria de trabajo" and "Estrategia de aprendizaje adaptativo y representación múltiple para el análisis de conectividad cerebral como soporte al mejoramiento cognitivo mediante un sistema de neuro-realimentación", respectively. These projects involve the analysis of brain connectivity during working memory tasks, a cognitive process that implies frequency-dependent directed interactions from the pre-frontal cortex toward temporal and parieto-occipital areas (Dimitriadis et al., 2016b; Johnson et al., 2018), a challenging problem that has given rise to attempts to assess such interactions through TE (Dimitriadis et al., 2016a).

Therefore, from both local and general perspectives, it is necessary to continue developing and improving upon existing measures of brain connectivity, in order to favor the characterization of cognitive processes and pathological brain activity in terms of directed interactions between distributed brain areas.

1.2 Problem statement

The functional interaction of neural assemblies distributed across different brain regions underlies many cognitive and perceptual processes (Bastos and Schoffelen, 2016). Therefore, understanding such processes, and brain function overall, requires identifying the flow of information within networks of connected assemblies, instead of solely focusing on the activity of specific brain regions in isolation (Sakkalis, 2011; Weber et al., 2017). The analysis of the interactions mentioned above is carried out through brain connectivity measures (Friston, 2011). These measures can be subdivided into two categories based on whether they quantify the direction of the neural interactions (Bastos and Schoffelen, 2016; Sakkalis, 2011). On the one hand, non-directed functional connectivity, or simply functional connectivity, aims to capture statistically significant interdependencies among the signals registering the activity of different neural assemblies, without determining their direction. On the other hand, directed connectivity, commonly referred to as effective connectivity, measures the influence that a neural assembly has over another one, establishing statistical causation from their signals, and hence a direction for their interaction. Effective connectivity is of particular importance in neuroscience because a large part of brain activity is endogenous and establishing physical causality among the neural systems supporting that activity is extremely difficult (Vicente et al., 2011). So statistical causality, based on the premise that a cause precedes its effect, becomes a valuable tool to decipher multiple aspects of brain function (Bastos and Schoffelen, 2016; Seth et al., 2015).

In general, effective connectivity is assessed through measures that are either based on a model of the process generating the data, or on approaches based on information theory (Bakhshayesh et al., 2019; Vicente et al., 2011). The former includes methods such as Granger causality (GC) and its variants, and dynamic causal modeling (DCM) (Friston, 2011; Seth et al., 2015); while the latter relies on the concept of information transfer or transfer entropy (TE) (Schreiber, 2000). Even though GC and DCM are widely used in neuroscience, TE has gained increasing attention in the literature (Timme and Lapish, 2018; Ursino et al., 2020), because of the advantages it offers as compared with other effective connectivity measures. Unlike classic GC, TE can capture high order correlations, and it is well suited to detect purely nonlinear interactions in the data, which are believed to be ubiquitous in brain activity (Weber et al., 2017). Although DCM can capture nonlinear interactions too, it requires some *a priori* knowledge on the input of the system and on the target connectivity network, which is not always

available (Vicente et al., 2011); in this sense, TE is model free. As an information-theoretic quantity, TE does not need an initial hypothesis about the interactions present in the data (Timme and Lapish, 2018), so it is a particularly useful tool for exploratory analyses. However, like all other information-theoretic quantities, TE is defined in terms of the probability distributions of the system under study, that in practice need to be estimated from data. Probability estimation is a challenging task, and it can significantly affect the outcome of information theory analyses, including the computation of TE (Giraldo et al., 2015; Timme and Lapish, 2018). Another drawback of TE is that, unlike GC, it lacks a standard spectral representation.

Neural oscillations are observed in the mammalian brain at different temporal and spatial scales (La Tour et al., 2017). These oscillations, and their interactions, have been linked to fundamental cognitive processes such as attention and memory (Da Silva, 2009; Wianda and Ross, 2019), and to information processing at large (Hyafil et al., 2015). If the interactions occur between oscillations of the same frequency (within-frequency), they can be captured through functional and effective connectivity measures, such as coherence, phase-locking value, Geweke-Granger causality statistics (GGC), partial directed coherence (PDC), and directed transfer function (DTF) (Cekic et al., 2018; Sakkalis, 2011). On the other hand, if they occur between oscillations of different frequencies (cross-frequency), they can be studied through cross-frequency coupling (CFC) analysis tools (Aru et al., 2015). Since TE is sensitive to all order correlations, it should be able to detect directed interactions among neural activities of the same spectral content, and couplings among activities at different frequency bands without prior specification (Vicente et al., 2011). However, since it is model-free, TE does not provide information about the type of interactions that it is capturing. In other words, it cannot reveal the local frequency band characteristics of the interactions it detects (Chen et al., 2019), and although several works have proposed strategies to obtain a spectral representation for TE (Besserve et al., 2010; Chen et al., 2019; Dimitriadis et al., 2016a; Lobier et al., 2014; Martínez-Cancino et al., 2020; Pinzuti et al., 2020), the question remains largely open (Pinzuti et al., 2020; Weber et al., 2017).

According to the above, we note that despite its advantages as an effective brain connectivity measure, TE has two limitations to overcome: *i) the need for probability distribution estimation as an intermediate step in TE computation* and *ii) the lack of a spectral representation for TE that allows revealing local frequency band characteristics of directed neural interactions*. Hence, the primary motivation of this thesis is to address these two problems.

1.2.1 Probability distribution estimation as an intermediate step in TE computation

TE is an information-theoretic quantity that estimates the directed interaction, or information flow, between two dynamical systems (Zhu et al., 2015). It was introduced by Schreiber (Schreiber, 2000) as a Wiener-causal measure within the framework of information theory. Therefore, TE is based on the assumption that a time series A causes a time series B if the information of the past of A, alongside the past of B, is better at predicting the future of B than the past of B alone (Cekic et al., 2018). TE quantifies such improvement in prediction by comparing the conditional probability distributions representing the above-described scenarios through a Kullback-Leibler divergence (Schreiber, 2000). Then, the resulting expression is re-written as a sum of Shannon entropies, which are defined in terms of the expected values of a series of joint and marginal probabilities (for details see Section 1.3) (Timme and Lapish, 2018; Vicente et al., 2011). Consequently, practical implementations of TE require estimating such probability distributions from the available data.

Current methods to estimate TE can be divided into two categories: methods that assume a probability distribution for the data, and plug-in estimators. The most prominent of the methods in the first category is the linear estimation method, which assumes Gaussianity (Montalto et al., 2014), although more elaborate approaches, such as assuming local deterministic nonlinear models for the probability distributions, have also been proposed (García and Mujica, 2019). However, by assuming *a priori* probability distributions for the data TE becomes a model-dependent scheme, neutralizing one of TE’s key attributes. For that reason, plug-in estimators are more commonly used in the literature, with the prime examples being binning strategies, usually in combination with symbolization schemes (Dimitriadis et al., 2016a; Montalto et al., 2014), and the local approximation of the probability distributions from nearest neighbor distances (Kraskov et al., 2004; Lindner et al., 2011). The so obtained TE depends on the quality of the estimated distributions and, consequently, on the performance of the plug-in estimator. Since the estimation of probability distributions can by itself be challenging, it would be desirable to be able to compute TE directly from the data, avoiding the intermediate stage of probability estimation. In that sense, recent advances in the area of information-theoretic learning have allowed estimating information-theoretic quantities, such as entropy and mutual information, directly from data through the use of kernel methods, sidestepping probability estimation (Giraldo et al., 2015). These methods are

yet to be adapted for TE estimation.

1.2.2 The lack of a spectral representation for TE

Effective connectivity measures such as GGC, PDC, or DTF can be used to detect directed within-frequency interactions. The former corresponds to the frequency domain representation of GC, and the two latter concepts are closely related to GGC (Cekic et al., 2018). However, since they are derived from VAR (vector autoregressive) estimated quantities, their ability to capture non-linear interactions is unclear (Chen et al., 2019). In the time domain, TE is well suited to analyze such interactions. Nonetheless, as mentioned before, there is no standard spectral representation for TE. The most common approach to estimate TE on a specific frequency band is to simply filter the data, but computing TE from filtered, real-valued time series has been questioned because of the effects of filters on the data's temporal structure, and more importantly, because such approach might not be able to completely isolate frequency-specific interactions (Pinzuti et al., 2020; Weber et al., 2017). There is, however, one alternative approach that has found multiple applications in neuroscience (Hillebrand et al., 2016; Numan et al., 2017; Wang et al., 2021a): phase transfer entropy (Lobier et al., 2014).

Three properties or components characterize a neural oscillation: its amplitude, its frequency, and its phase; the latter referring to the position of a signal within an oscillation cycle (Xie et al., 2021). Oscillation amplitudes are related to neural synchrony expansion in a local assembly, while the relationships between the phases of neural oscillations, such as phase synchronization, are involved in the coordination of anatomically distributed processing (Ahmadi et al., 2020). The concept of phase TE refers to the idea of estimating TE from instantaneous phase time series obtained at a particular frequency (Lobier et al., 2014). The interest of studying phase-based interactions independently from other spectral relationships arises because, from a functional perspective, phase synchronization and amplitude correlations are independent phenomena (Kang et al., 2021). Besides, phase relationships are thought to be linked to information flow within networks of connected neural assemblies (Lobier et al., 2014). Thence the usefulness of phase TE. Especially when considering that metrics such as the GGC, PDC or DTF depend on both amplitude and phase signal components (Lobier et al., 2014), and measures capable of capturing directed phase relationships, such as the phase slope index (PSI), are limited to linear interactions (Jiang et al., 2015; Nolte et al., 2008). Nonetheless, conventional TE estimators are not well suited for circular variables (Lobier et al., 2014) (a variable that is measured on a circle in degrees or radians (Cremers and Klugkist,

2018)), so phase TE estimates are obtained through a binning approach performed over multiple trials simultaneously, in a procedure termed trial collapsing, which bars its use in BCI or other machine learning applications requiring features extracted from single-trial data. Furthermore, recently proposed methods to compute spectrally resolved TE that can obtain single-trial estimates are conceptually different from phase TE (Chen et al., 2019; Pinzuti et al., 2020), as they are not phase-specific.

Regarding the assessment of cross-frequency interactions, which have been hypothesized to be directly related to the integration of distributed information by bridging local and global processes in the brain (Jirsa and Müller, 2013), they are usually carried out through non-directed metrics such as the modulation index, the mean vector length, and variations of the concept of mutual information (Cheng et al., 2018; Malladi et al., 2018; Martínez-Cancino et al., 2019). In general, they are specifically tailored to determine the presence of one out of four types of cross-frequency interactions, namely, amplitude to amplitude, phase to phase, phase to amplitude, and phase to frequency couplings (Jirsa and Müller, 2013). The use of directed metrics is less common, and it is mostly limited to linear quantities, such as the cross-frequency directionality (CFD) (Daume et al., 2017; Jiang et al., 2015; Johnson et al., 2018). However, recent works call for further inclusion of directed connectivity measures in the analysis of CFCs, signaling out TE as a promising candidate (Dimitriadis et al., 2016b; Jiang et al., 2015; La Tour et al., 2017; Martínez-Cancino et al., 2019), especially because GC-based methods have been shown to fail when the signal-to-noise ratios of the interacting signals are different (La Tour et al., 2017), which is often the case in phase to amplitude couplings (PAC), the most widely studied type of CFC (Seymour et al., 2017). Yet the lack of a standard spectral representation for TE also hinders its application in the context of CFCs analysis, with works aiming to capture PAC relying on TE computation from real-valued filtered data (Besserve et al., 2010; Martínez-Cancino et al., 2020; Shi et al., 2019), which is advised against by some authors (Weber et al., 2017).

Therefore, some problems related to TE estimation and its spectral representation remain unsolved. For this reason, there arises the following research question: how to develop a data-driven approach to transfer entropy estimation, that avoids the need to explicitly obtain the probability distribution of the data and reveals directed phase-based interactions at specific frequencies?

1.3 Theoretical background

In this section, we introduce the mathematical formulation of transfer entropy and other relevant related concepts. First, in Section 1.3.1 we formally define transfer entropy. Then, in Section 1.3.2 we do the same for Granger causality. Finally, Section 1.3.3 describes an approach to estimate information theoretic quantities directly from data through kernel matrices.

1.3.1 Transfer entropy

The concept of transfer entropy (TE) was introduced by Schreiber (Schreiber, 2000) as a Wiener-causal measure within the framework of information theory. It estimates the directed interaction, or information flow, between two dynamical systems (Bakhshayesh et al., 2019; Zhu et al., 2015). It is based on the assumption that if a time series A is causal for another time series B, then the history of A should contain information that helps to better predict B as compared with the information contained in the history of B alone (Cekic et al., 2018). It is also based on the information-theoretic concept of Shannon entropy:

$$H_S(X) = \mathbb{E} \{-\log(p(x))\} \approx - \sum_x p(x) \log(p(x)), \quad (1.1)$$

where X is a discrete random variable, $p(\cdot)$ is the probability mass function of X , and $\mathbb{E}\{\cdot\}$ stands for the expected value operator. $H_S(X)$ quantifies the average reduction in uncertainty attained after measuring the values of X . By associating the improvement in prediction power of Wiener's definition of causality with the reduction of uncertainty measured by entropy, Schreiber arrived at the concept of TE (Schreiber, 2000). Formally, TE measures the difference between two conditional probability distributions:

$$\Delta = p(y_{t+1} | \mathbf{y}_t^m, \mathbf{x}_t^n) - p(y_{t+1} | \mathbf{y}_t^m), \quad (1.2)$$

where $\mathbf{x}_t^n \in \mathbb{R}^n$ and $\mathbf{y}_t^m \in \mathbb{R}^m$ are Markov processes, of orders n and m , that approximate two time series $\mathbf{x} = \{x_t\}_{t=1}^T$ and $\mathbf{y} = \{y_t\}_{t=1}^T$, respectively, and $t \in \mathbb{N}$ is a discrete time index, $T \in \mathbb{N}$. This difference is quantified through the Kullback-Leibler divergence ($D_{KL}(p||q) = \sum_x p(x) \log(p(x)/q(x))$) of the conditional probabilities $p(y_{t+1} | \mathbf{y}_t^m, \mathbf{x}_t^n)$ and

$p(y_{t+1}|\mathbf{y}_t^m)$:

$$TE(\mathbf{x} \rightarrow \mathbf{y}) = \sum_{y_{t+1}, \mathbf{y}_t^m, \mathbf{x}_t^n} p(y_{t+1}, \mathbf{y}_t^m, \mathbf{x}_t^n) \log \left(\frac{p(y_{t+1}|\mathbf{y}_t^m, \mathbf{x}_t^n)}{p(y_{t+1}|\mathbf{y}_t^m)} \right). \quad (1.3)$$

Therefore, TE measures whether the probability of a future value of \mathbf{y} increases given the past values of \mathbf{x} and \mathbf{y} , as compared with the probability of that same future value of \mathbf{y} given only the past of \mathbf{y} .

In an attempt to better capture the underlying dynamics of the system that generates the observed data, i.e., the measured values of the random variables contained in the time series, TE is not usually defined directly on the raw data, but on its space state (Vicente et al., 2011). We can reconstruct such state space from the observations through time embedding. The most commonly used embedding procedure in the literature is Takens delay embedding (Takens, 1981). So that for a time series \mathbf{x} its space state is approximated as:

$$\mathbf{x}_t^d = (x(t), x(t - \tau), x(t - 2\tau), \dots, x(t - (d - 1)\tau)), \quad (1.4)$$

where $d, \tau \in \mathbb{N}$ are the embedding dimension and delay, respectively, which can be estimated heuristically or through criteria such as those proposed by Cao or Ragwitz (Lindner et al., 2011). We can now express TE in terms of the embedded data as:

$$TE(\mathbf{x} \rightarrow \mathbf{y}) = \sum_{y_{t+1}, \mathbf{y}_t^{dy}, \mathbf{x}_t^{dx}} p(y_{t+1}, \mathbf{y}_t^{dy}, \mathbf{x}_t^{dx}) \log \left(\frac{p(y_{t+1}|\mathbf{y}_t^{dy}, \mathbf{x}_t^{dx})}{p(y_{t+1}|\mathbf{y}_t^{dy})} \right), \quad (1.5)$$

where $\mathbf{x}_t^{dx}, \mathbf{y}_t^{dy} \in \mathbb{R}^{D \times d}$ are time embedded versions of \mathbf{x} and \mathbf{y} , with $D = T - ((d - 1)\tau)$. To generalize TE to interaction times other than 1, we rewrite Equation 1.5 as:

$$TE(\mathbf{x} \rightarrow \mathbf{y}) = \sum_{y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}} p(y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}) \log \left(\frac{p(y_t|\mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx})}{p(y_t|\mathbf{y}_{t-1}^{dy})} \right), \quad (1.6)$$

where $u \in \mathbb{N}$ represents the interaction delay between the driving and the driven systems. The changes in the time indexing are necessary to guaranty that Wiener's definition of causality is respected (Wibral et al., 2013). Using the definition in Equation 1.1, we can also express Equation 1.6 as a difference of conditional Shannon entropies:

$$TE(\mathbf{x} \rightarrow \mathbf{y}) = H_S(y_t|\mathbf{y}_{t-1}^{dy}) - H_S(y_t|\mathbf{y}_{t-1}^{dy}, \mathbf{x}_t^{dx}), \quad (1.7)$$

or alternatively, as a sum of joint and marginal Shannon entropies (see Figure 1.1 for a graphical representation),

$$TE(\mathbf{x} \rightarrow \mathbf{y}) = H_S(\mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}) - H_S(y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}) + H_S(y_t, \mathbf{y}_{t-1}^{dy}) - H_S(\mathbf{y}_{t-1}^{dy}). \quad (1.8)$$

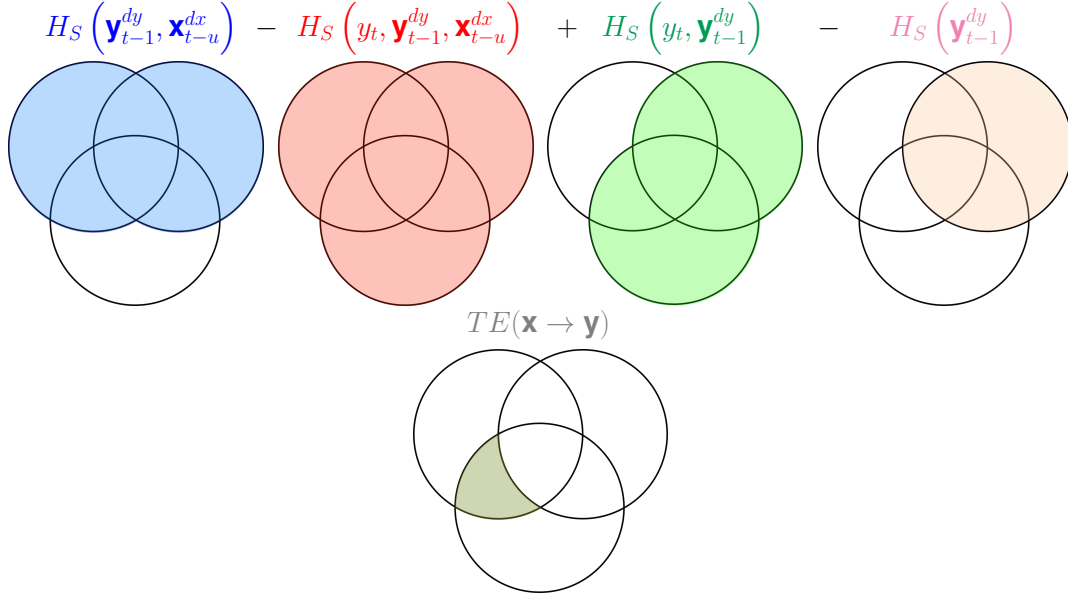


Figure 1.1: Venn diagrams showing the linear combination of joint and marginal entropies associated with the definition of TE, as expressed in Equation 1.8.

1.3.2 Granger causality

Granger Causality (GC), like TE, is a mathematical formalization of the concept of Wiener's causality, one that is widely used in neuroscience to assess effective connectivity (Seth et al., 2015). However, unlike TE, GC is not based on a probabilistic approach. The basic idea behind it is that for two stationary time series $\mathbf{x} = \{x_t\}_{t=1}^T$ and $\mathbf{y} = \{y_t\}_{t=1}^T$, if \mathbf{x} causes \mathbf{y} , then the linear autoregressive model:

$$y_t = \sum_{k=1}^o a_k y_{t-k} + e_t, \quad (1.9)$$

where $o \in \mathbb{N}$ is the model's order and $a_k \in \mathbb{R}$ stands for the model's coefficients, will exhibit larger prediction errors e_t than a model that also includes past observations

of \mathbf{x} ; that is, a linear bivariate autoregressive model of the form:

$$y_t = \sum_{k=1}^o a'_k y_{t-k} + \sum_{k=1}^o b_k x_{t-k} + e'_t. \quad (1.10)$$

where the coefficients $b_k \in \mathbb{R}$. The magnitude of the causal relation from \mathbf{x} to \mathbf{y} can then be quantified by the log ratio of the variances of the residuals or prediction errors (Seth, 2010):

$$\text{GC}(\mathbf{x} \rightarrow \mathbf{y}) = \log \left(\frac{\text{var}\{\mathbf{e}\}}{\text{var}\{\mathbf{e}'\}} \right), \quad (1.11)$$

where $\mathbf{e}, \mathbf{e}' \in \mathbb{R}^{T-o}$ are vectors holding the prediction errors, and $\text{var}\{\cdot\}$ stands for the variance operator. If the past of \mathbf{x} does not improve the prediction of \mathbf{y} , then $\text{var}\{\mathbf{e}\} \approx \text{var}\{\mathbf{e}'\}$ and $\text{GC}(\mathbf{x} \rightarrow \mathbf{y}) \rightarrow 0$, if it does, then $\text{var}\{\mathbf{e}\} \gg \text{var}\{\mathbf{e}'\}$ and $\text{GC}(\mathbf{x} \rightarrow \mathbf{y}) \gg 0$. As defined above, GC is a linear bivariate parametric method that depends on the order o of the autoregressive model. Nonetheless, there are several variations of this basic formulation of GC that aim to capture nonlinear and multivariate relations in the data (Sameshima and Baccala, 2016). As a final remark, it is worth noting that although by definition TE has an advantage over GC by not assuming any *a priori* model for the interaction between the systems under study, the two are linked. As demonstrated in Barnett et al. (2009), they are entirely equivalent for Gaussian variables (up to a factor of 2).

1.3.3 Information theoretic learning from kernel matrices

Information theoretic learning (ITL) is a data-driven learning framework that employs information theoretic quantities as objective functions for supervised and unsupervised learning algorithms (Li and Principe, 2020). However, instead of using the Shannon-based definition of entropy, ITL exploits the properties of a mathematical generalization of such a concept known as Renyi's α -order entropy. As explained before, Shannon entropy is defined as the expected value of the amount of information of the outcomes of a random variable. For a continuous random variable X , and using the linear averaging operator, we have that $H(X) = \mathbb{E}\{I(X)\} = \int p(x)I(x)dx$, where $I(x) = -\log(p(x))$. Nonetheless, the linear mean is only a particular case of the average operator. In general, the expected value associated with a monotonic function $g(x)$, with inverse $g^{-1}(x)$, is $\mathbb{E}\{x\} = g^{-1}(\int p(x)g(x)dx)$. Furthermore, because of the postulate for additivity of independent events the possible choices for $g(x)$ are restricted to only 2 classes: $g(x) = cx$ and $g(x) = c2^{(1-\alpha)x}$, as proven in Rényi et al. (1961). The former gives rise to the linear

mean and therefore to the Shannon entropy, while the latter implies that:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\int p(x)^\alpha dx \right), \quad (1.12)$$

with $\alpha \neq 1$ and $\alpha \geq 0$, which corresponds to Renyi's α entropy (Principe, 2010; Rényi et al., 1961). This parametric family of entropies encompasses the definition of Shannon entropy in the limiting case when $\alpha \rightarrow 1$. In practice one must estimate entropy from discrete data. Given an *i.i.d.* sample of n realizations of a discrete random variable X , $\{x_i\}_{i=1}^n \subset \mathbb{R}^d$, the probability density function of X can be approximated through the Parzen density estimator as $\hat{p}(x) = \frac{1}{n} \sum_{i=1}^n \kappa(x, x_i)$, where $\kappa(\cdot, \cdot) \in \mathbb{R}$ stands for a positive definite kernel function (see Appendix A, Section A.1, for a brief introduction on kernel methods). For the case of $\alpha = 2$, the parzen approximation yields:

$$\hat{H}_2(X) = -\log \left(\frac{1}{n^2} \sum_{i,j=1}^n \kappa(x_i, x_j) \right), \quad (1.13)$$

where the integral in Equation 1.12 has been replaced by a sum. The expression in Equation 1.13 can be rewritten in terms of a Gram matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$ as $\hat{H}_2(X) = -\log \left(\frac{1}{n^2} \text{tr}(\mathbf{K}\mathbf{K}) \right) + C$, where \mathbf{K} holds elements $k_{ij} = \kappa(x_i, x_j)$, $C \in \mathbb{R}^+$ accounts for the normalization factor of the Parzen window, and $\text{tr}(\cdot)$ stands for the matrix trace. From this result we can see that the Frobenius norm of the Gram matrix \mathbf{K} , defined as $\|\mathbf{K}\|_F^2 = \text{tr}(\mathbf{K}\mathbf{K})$, is related to an entropy estimator. In Giraldo et al. (2015) the authors generalize this notion. They extend it to other spectral norms, and introduce an entropy-like quantity with properties that closely resemble those of Renyi's entropy, while avoiding the estimation of probability distributions altogether. Given a Gram matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with elements $a_{ij} = \kappa(x_i, x_j)$, a kernel-based formulation of Renyi's α -order entropy can be defined as:

$$H_\alpha(\mathbf{A}) = \frac{1}{1-\alpha} \log (\text{tr}(\mathbf{A}^\alpha)), \quad (1.14)$$

where it holds that $\text{tr}(\mathbf{A}) = 1$. The power α of \mathbf{A} can be obtained using the spectral theorem (Giraldo et al., 2015). Moreover, under this formulation, the joint entropy is defined as:

$$H_\alpha(\mathbf{A}, \mathbf{B}) = H_\alpha \left(\frac{\mathbf{A} \circ \mathbf{B}}{\text{tr}(\mathbf{A} \circ \mathbf{B})} \right) = \frac{1}{1-\alpha} \log \left(\text{tr} \left(\left(\frac{\mathbf{A} \circ \mathbf{B}}{\text{tr}(\mathbf{A} \circ \mathbf{B})} \right)^\alpha \right) \right), \quad (1.15)$$

where $\mathbf{B} \in \mathbb{R}^{n \times n}$ is a Gram matrix holding the pairwise evaluation of the kernel function $\kappa(\cdot, \cdot)$ on an *i.i.d.* sample of n realizations of a second discrete random variable, and the operator \circ stands for the Hadamard product. The joint entropy in equation (1.15) can be extended to more arguments by computing the Hadamard product of all the corresponding kernel matrices.

For further details on the kernel-based estimation of Renyi's α entropy see Appendix A (Section A.2).

1.4 Literature review on transfer entropy estimation

In practice, one must estimate TE from data. Current methods to estimate TE can be divided into two categories: methods that assume a probability distribution for the data, such as the linear estimator method, and plug-in estimators, such as binning or the nearest neighbor distances approximation.

The linear estimator method works under the assumption that the probability distributions involved in TE are jointly Gaussian (Montalto et al., 2014). Then, the two terms in Equation 1.7 are expressed by means of linear regressions, which leads to the following closed expression for TE:

$$\text{TE}(\mathbf{x} \rightarrow \mathbf{y}) = \frac{1}{2} \log \left(\frac{\text{var}\{\mathbf{e}\}}{\text{var}\{\mathbf{e}'\}} \right), \quad (1.16)$$

where $\mathbf{e}, \mathbf{e}' \in \mathbb{R}^{T-o}$ are vectors holding the prediction errors from the linear regressions (the same as in Equation 1.11). As mentioned before, this formulation is entirely equivalent to GC, except for a scaling factor. In a more elaborate approach, that also assumes a probability distribution for the data, the authors in García and Mujica (2019) rewrite TE in terms of cumulative probability density functions (CPDs) and then estimate the CPDs using deterministic models of the rule generating the data, particularly they chose sigmoid functions for the CPDs, and a Heavyside function as the generating model. However, as the authors in García and Mujica (2019) themselves point out, by assuming *a priori* probability distributions for the data, TE becomes a model-dependent scheme, eliminating one of its advantages over other effective connectivity measures. For that reason, plug-in estimators are preferred in the literature.

In the seminal work of Schreiber (Schreiber, 2000), the probability distributions involved in TE were estimated through a binning approach. In general, this approach consists of a uniform quantization of the time series followed by an approximation of the

probabilities with the frequency of visitation of the quantized states (Montalto et al., 2014). The binning approach for TE estimation has been coupled with symbolic dynamics, a powerful tool for studying complex dynamical systems (Dimitriadis et al., 2012). The infinite number of values that can be attained by a given time series is replaced by a set of symbols through a symbolization scheme. We can then use the relative frequency of the symbols to estimate the joint and conditional probability distributions needed to compute TE (Dimitriadis et al., 2016a). Given the space state reconstruction of a time series \mathbf{x} (see Equation 1.4), we can arrange the elements in \mathbf{x}_t^d according to their amplitude, in ascending order, as follows:

$$x(t - r_1\tau) \leq x(t - r_2\tau) \leq \dots \leq x(t - r_d\tau), \quad (1.17)$$

where $r_1, r_2, \dots, r_d \in \{0, 1, \dots, d-1\}$, in order to obtain a symbolic sequence \mathbf{s}_t^x :

$$\mathbf{x}_t^d \rightarrow \mathbf{s}_t^x \equiv (r_1, r_2, \dots, r_d), \quad (1.18)$$

in what is known as ordinal pattern symbolization (it is also known as the permutation approach for the estimation of information theoretic quantities (Kang et al., 2021)). Finally, we define the symbolic version of TE as:

$$TE(\mathbf{x} \rightarrow \mathbf{y}) = \sum_{\mathbf{s}_{t+1}^y, \mathbf{s}_t^y, \mathbf{s}_{t+1-u}^x} p(\mathbf{s}_{t+1}^y, \mathbf{s}_t^y, \mathbf{s}_{t+1-u}^x) \log \left(\frac{p(\mathbf{s}_{t+1}^y | \mathbf{s}_t^y, \mathbf{s}_{t+1-u}^x)}{p(\mathbf{s}_{t+1}^y | \mathbf{s}_t^y)} \right). \quad (1.19)$$

We can rewrite Equation 1.19 in terms of Shannon entropies, as in Equation 1.8, and estimate the probability functions by counting the occurrences of the discrete symbols instead of quantizing continuous data, as in conventional binning strategies. Other symbolization approaches, such as the neural-gas algorithm-based symbolization, can also be employed (Dimitriadis et al., 2016a).

An alternative approach, widely used in neuroscience studies, is an adaptation for TE of the Kraskov-Stögbauer-Grassberger method for estimating mutual information (Kraskov et al., 2004). The method relies on a local approximation of the probability distributions needed to estimate the entropies from the distances of every data point to its neighbors, within a predefined neighborhood diameter. Also, it deals with the dimensionality differences in the data spaces in Equation 1.8 by fixing the number of neighbors in the highest dimensional space, the one spanned by $(y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx})$, and projecting the distances obtained there to the marginal (and lower dimensional) spaces so

that they serve as neighborhood diameters in those. The Kraskov-Stögbauer-Grassberger estimator for TE is expressed as:

$$TE(\mathbf{x} \rightarrow \mathbf{y}) = \psi(K) + \mathbb{E} \left\{ \psi \left(n_{\mathbf{y}_{t-1}^{dy}} + 1 \right) - \psi \left(n_{y_t \mathbf{y}_{t-1}^{dy}} + 1 \right) - \psi \left(n_{\mathbf{y}_{t-1}^{dy} \mathbf{x}_{t-u}^{dx}} \right) \right\}_t, \quad (1.20)$$

where $\psi(\cdot)$ stands for the digamma function, $K \in \mathbb{N}$ is the selected number of neighbors in the highest dimensional space in Equation 1.8, $\mathbb{E}\{\cdot\}_t$ represents averaging over different time points, and $n \in \mathbb{N}$ is the number of points in the marginal spaces (Lindner et al., 2011; Martínez-Cancino et al., 2020).

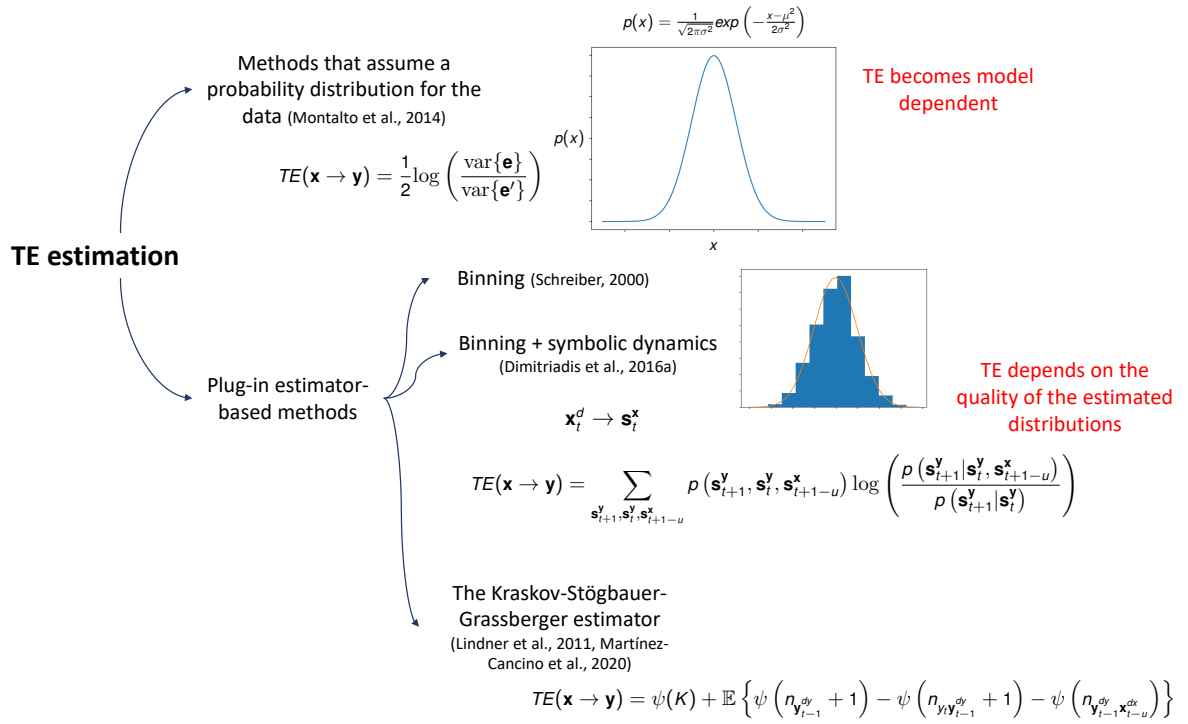


Figure 1.2: Summary of the main TE estimation methods discussed, and their drawbacks (highlighted in red).

The two methods described above, binning and the nearest neighbor distances approximation, rely on the use of plug-in estimators to approximate the probability distributions in the joint and marginal entropies involved in the definition of TE. Therefore, the so obtained TE depends on the quality of the estimated distributions and, consequently, on the performance of the plug-in estimator. Since the estimation of probability distributions can by itself be challenging, it would be desirable to be able to compute TE

directly from the data, avoiding the intermediate stage of probability density estimation, as has been proposed for other information-theoretic quantities (Giraldo et al., 2015). Figure 1.2 summarizes the previous discussion.

1.4.1 Transfer entropy in the frequency domain

Measuring frequency-resolved information transfer, or TE, has been recently shown to correspond to a partial information decomposition problem (Pinzuti et al., 2020). A problem in information theory that has only been partly resolved and whose mathematical formulation is still a subject of research (Pinzuti et al., 2020). Under this framework, most of the current literature about assessing within-frequency and cross-frequency directed interactions through TE, which explicitly assumes one-to-one interactions between narrow band spectral components, ignores many of the mechanisms through which such components may transfer information. Therefore, it addresses only a small part of the problem, as schematized in Figure 1.3. Unfortunately, even the approach put forward in Pinzuti et al. (2020) does not really constitute a frequency-resolved TE estimation method. It is a methodology to determine, through statistical testing, how likely a TE value is once the information at a particular frequency or frequencies in the source or target signals has been destroyed, and according to the authors, it is blind to some possible types of interactions.

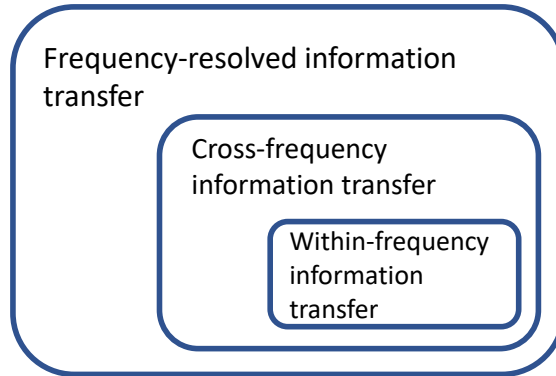


Figure 1.3: Assessing the within-frequency and cross-frequency transfer of information between time series, as commonly understood in the literature, is only a small part of the problem of measuring frequency-resolved information transfer.

In view of the above, and of the research interests of our group, in this thesis we decided to go in the opposite direction. Instead of working on a full information-theoretic analysis that takes into account simultaneously the multiple spectral components that

can interact at both the sender and receiver side of coupled dynamical systems, we focused on addressing some of the hurdles associated with the estimation of TE between specific oscillatory components whose interactions are believed to underlie several cognitive processes. Namely, we centered our attention on directed within-frequency phase to phase interactions (Lobier et al., 2014; Nolte et al., 2008), and on directed cross-frequency phase to amplitude couplings (Jiang et al., 2015; Jirsa and Müller, 2013; La Tour et al., 2017; Seymour et al., 2017).

Phase transfer entropy estimation

The information carried by the phases of neural oscillations is fundamental for the coordination of anatomically distributed processing in the brain (Ahmadi et al., 2020; Xie et al., 2021). The computation of TE between two instantaneous phase time series extracted at a particular frequency (a signal's phase is only physically meaningful when its spectrum is narrow-banded (Wilmer et al., 2012)), while disregarding amplitude information, is known as phase TE (Lobier et al., 2014) (see Figure 1.4).

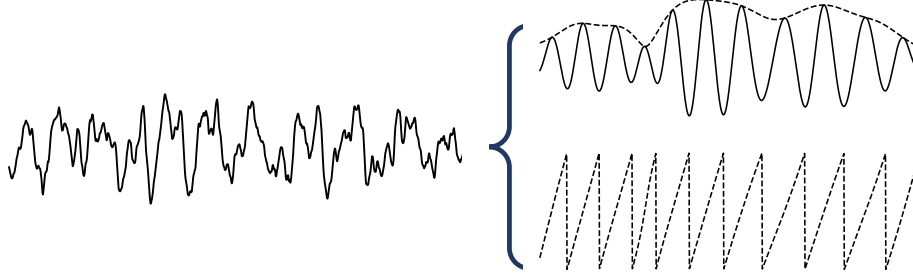


Figure 1.4: An oscillation from a neural time series (left plot) can be described in terms of its frequency, its amplitude envelope (top-right plot), and its phase (bottom-right plot). In phase TE only the information carried by the oscillation's phases is considered.

Formally, in phase TE the time series \mathbf{x} and \mathbf{y} are replaced by instantaneous phase time series $\boldsymbol{\theta}^x(f) \in [-\pi, \pi]_{t=1}^T$ and $\boldsymbol{\theta}^y(f) \in [-\pi, \pi]_{t=1}^T$, obtained from $\mathbf{s}_x = \boldsymbol{\varsigma}^x e^{i\boldsymbol{\theta}^x(f)} \in \mathbb{C}^T$ and $\mathbf{s}_y = \boldsymbol{\varsigma}^y e^{i\boldsymbol{\theta}^y(f)} \in \mathbb{C}^T$, which contain the complex-filtered values of \mathbf{x} and \mathbf{y} at frequency f , respectively, and with $\boldsymbol{\varsigma}^x, \boldsymbol{\varsigma}^y \in \mathbb{R}^T$ the amplitude envelopes of the filtered time series (Lobier et al., 2014). Thus, we have that

$$TE^\theta(\mathbf{x} \rightarrow \mathbf{y}, f) = H_S(\boldsymbol{\theta}_{t-1}^{y,dy}, \boldsymbol{\theta}_{t-u}^{x,dx}) - H_S(\boldsymbol{\theta}_t^y, \boldsymbol{\theta}_{t-1}^{y,dy}, \boldsymbol{\theta}_{t-u}^{x,dx}) + H_S(\boldsymbol{\theta}_t^y, \boldsymbol{\theta}_{t-1}^{y,dy}) - H_S(\boldsymbol{\theta}_{t-1}^y), \quad (1.21)$$

where $\boldsymbol{\theta}_t^{x,dx}$ and $\boldsymbol{\theta}_t^{y,dy}$ are time embedded versions of $\boldsymbol{\theta}^x$ and $\boldsymbol{\theta}^y$. Note that for the sake

of notation simplicity we have dropped the explicit dependency of the phase time series on f .

The concept of phase TE has found multiple applications in neuroscience, such as gaining insight into reduced levels of consciousness by evaluating brain connectivity (Numan et al., 2017), analyzing resting-state networks (Hillebrand et al., 2016), and assessing brain connectivity changes in children diagnosed with attention deficit hyperactivity disorder following neurofeedback training (Ekhlasi et al., 2021; Wang et al., 2021a). It has even been used to detect fluctuations in financial markets data (Yang et al., 2017). Nonetheless, computing the quantity in Equation 1.21 is not straightforward since conventional TE estimators, as most of those previously described, are not well suited for periodic variables. For that reason, in Lobier et al. (2014) phase TE estimates are obtained through a binning approach performed over multiple trials simultaneously, in a procedure termed trial collapsing. However, the so obtained phase TE values cannot be employed in any data analysis strategy that requires features extracted on an independent trial basis, i.e., each trial must be associated with a set of features, such as most machine learning algorithms, or in BCI. In that sense, authors in Ahmadi et al. (2020) applied a binning strategy to estimate single-trial phase TE to set up classification systems for visual attention. Nonetheless, binning estimators for single trial-based estimation of information-theoretic measures exhibit systematic bias (Lobier et al., 2014). Furthermore, other recently proposed approaches that claim to be able to obtain single-trial frequency-specific TE estimates are conceptually different from phase TE, as they are not phase-specific metrics (Chen et al., 2019; Pinzuti et al., 2020). Figure 1.5 summarizes the above discussion about phase TE computation, and places it in the larger context of the approaches used to estimate TE, and other effective connectivity measures, for within-frequency interactions.

Transfer entropy estimation for directed phase-amplitude interactions

The most widely studied instance of CFC is the modulation of the amplitude envelope of high-frequency oscillations by the phase evolution of low-frequency activity, known as phase-amplitude coupling (PAC) (Jirsa and Müller, 2013; La Tour et al., 2017; Seymour et al., 2017). Such phase-amplitude interactions seem to be linked to normal and pathological brain processes in different mammalian species, including humans (Martínez-Cancino et al., 2020), and they have been observed locally and interregionally across a wide range of cognitive tasks (Johnson et al., 2018). Theoretically, PAC allows for information transfer from large-scale brain networks associated with low-frequency os-

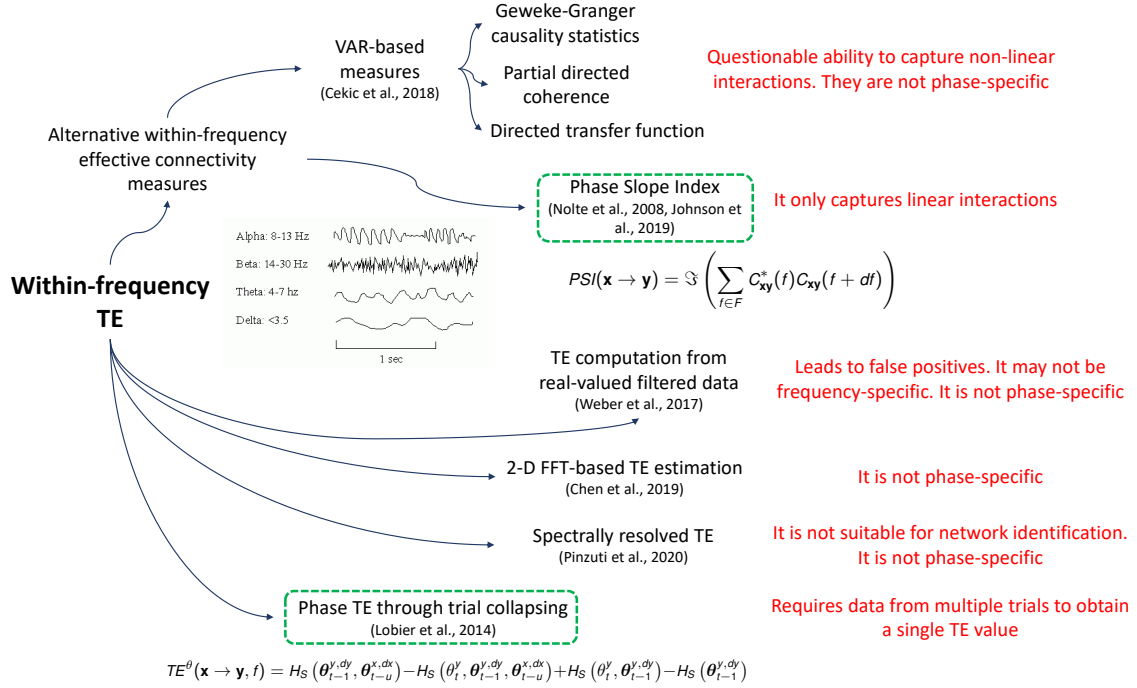


Figure 1.5: Summary of the approaches used to estimate TE and other effective connectivity measures for within-frequency interactions, and their drawbacks (highlighted in red). Phase-specific measures are encircled in green, dashed lines. For further details about the PSI see Section 3.1.1.

cillations to local, fast cortical processing areas exhibiting high-frequency activity (Shi et al., 2018).

Phase-amplitude interactions are commonly assessed from electrophysiological data through metrics of statistical dependency such as the modulation index, the mean vector length, and variations of the concept of mutual information (Cheng et al., 2018; Malladi et al., 2018; Martínez-Cancino et al., 2019). However, these approaches are unable to capture the directionality and delay of phase-amplitude interactions, quantities intrinsic to the concept of information being sent from a neural assembly to another (Jiang et al., 2015; Martínez-Cancino et al., 2020; Pinzuti et al., 2020). A natural solution to this limitation, within the framework of information theory, is to assess PACs using TE (Dimitriadis et al., 2016b; Jiang et al., 2015; La Tour et al., 2017; Martínez-Cancino et al., 2019; Shi et al., 2019).

In that context, most works related to the estimation of directed cross-frequency interactions through TE are focused on the detection of phase to amplitude couplings. The conventional approach to do so consists of two stages (Besserve et al., 2010; Martínez-

Cancino et al., 2020). First, a component extraction stage, which involves complex-filtering, or performing a phase/amplitude decomposition, to extract instantaneous phase and amplitude time series (see Figure 1.6). Then, a TE computation stage that simply consists of estimating the information flow between the aforementioned extracted data.

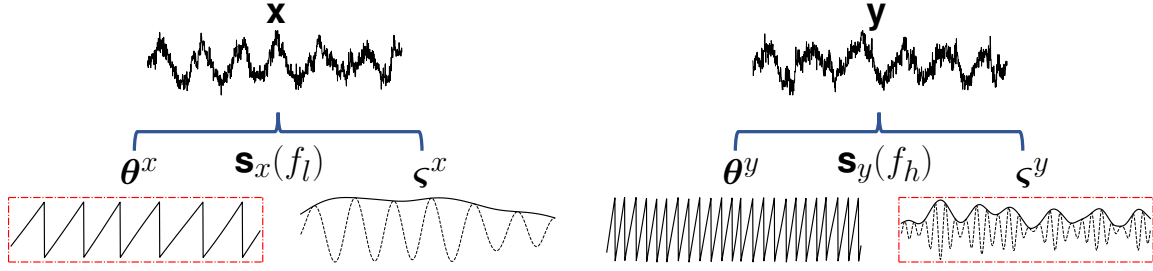


Figure 1.6: Graphical representation of the most common approach to capture directed phase-amplitude interactions through TE. It consists in estimating TE from instantaneous phase and amplitude time series extracted at frequencies f_l and f_h , respectively.

Formally, given two time series \mathbf{x} and \mathbf{y} , in order to estimate the TE from the phase of \mathbf{x} at a frequency f_l (usually a low frequency) to the amplitude envelope of \mathbf{y} at a frequency f_h (commonly a higher frequency than f_l), we obtain complex time series $\mathbf{s}_x(f_l) = \boldsymbol{\zeta}^x e^{i\boldsymbol{\theta}^x} \in \mathbb{C}^T$ and $\mathbf{s}_y(f_h) = \boldsymbol{\zeta}^y e^{i\boldsymbol{\theta}^y} \in \mathbb{C}^T$, which contain the filtered values of \mathbf{x} and \mathbf{y} at f_l and f_h , respectively; where $\boldsymbol{\theta}^x, \boldsymbol{\theta}^y \in [-\pi, \pi]_{t=1}^T$ are instantaneous phase time series, and $\boldsymbol{\zeta}^x, \boldsymbol{\zeta}^y \in \mathbb{R}^T$ are amplitude envelopes (Lobier et al., 2014). Then, we compute the desired TE as:

$$TE^{\theta^x, \zeta^y}(\mathbf{x} \rightarrow \mathbf{y}, f_l, f_h) = H_S(\boldsymbol{\zeta}_{t-1}^{y, dy}, \boldsymbol{\theta}_{t-u}^{x, dx}) - H_S(\boldsymbol{\zeta}_t^y, \boldsymbol{\zeta}_{t-1}^{y, dy}, \boldsymbol{\theta}_{t-u}^{x, dx}) + H_S(\boldsymbol{\zeta}_t^y, \boldsymbol{\zeta}_{t-1}^{y, dy}) - H_S(\boldsymbol{\zeta}_{t-1}^{y, dy}), \quad (1.22)$$

where $\boldsymbol{\theta}_t^{x, dx}$ and $\boldsymbol{\zeta}_t^{y, dy}$ are time embedded versions of $\boldsymbol{\theta}^x$ and $\boldsymbol{\zeta}^y$. In practice, the phase and amplitude time series are input to one of the above-described TE estimation methods, such as a binning approach in Besserve et al. (2010), or a time resolved version of the Kraskov-Stögbauer-Grassberger estimator in Martínez-Cancino et al. (2020).

The conventional approach to estimate directed phase-amplitude interactions through TE, as expressed by Equation 1.22, implies the computation of TE from data of different properties, a phase time series $\boldsymbol{\theta}^x$, which represents a circular variable, and a smooth, continuous real-valued amplitude envelope $\boldsymbol{\zeta}^y$. Thus, as in the case of TE estimation from real-valued narrow-band filtered data, its ability to isolate frequency-specific interactions can be questioned (Pinzuti et al., 2020). Figure 1.7 summarizes the previous discussion about the estimation of TE for directed PAC. It also presents alternative

cross-frequency effective connectivity measures capable of capturing directed phase to amplitude interactions.

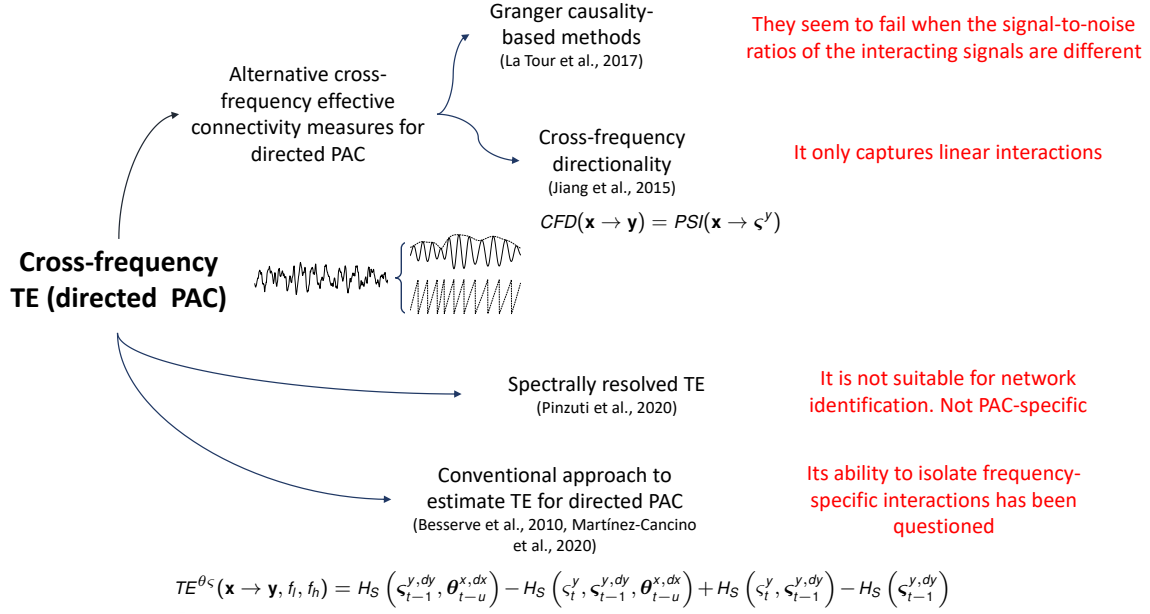


Figure 1.7: Summary of the approaches used to estimate TE and other effective connectivity measures for directed phase-amplitude interactions, and their drawbacks (highlighted in red). For further details about the CFD see Section 4.1.2.

1.5 Aims

1.5.1 General aim

To develop a data-driven approach to transfer entropy estimation that allows revealing directed interactions in the time and frequency domains, while sidestepping the need for probability distribution estimation.

1.5.2 Specific aims

1. To develop a transfer entropy estimation approach that avoids probability estimation by using functionals defined on kernel matrices that approximate Renyi's entropy measures of order α .

2. To develop a spectral representation for transfer entropy that allows revealing directed phase-based within-frequency interactions, using frequency-domain transforms and a kernel-based transfer entropy estimator.
3. To develop a spectral representation for transfer entropy that allows revealing directed phase-amplitude cross-frequency interactions, using frequency-domain transforms and a kernel-based transfer entropy estimator.

1.6 Outline and contributions

In the following, we briefly introduce the main contributions of this thesis. They are summarized in Figure 1.8.

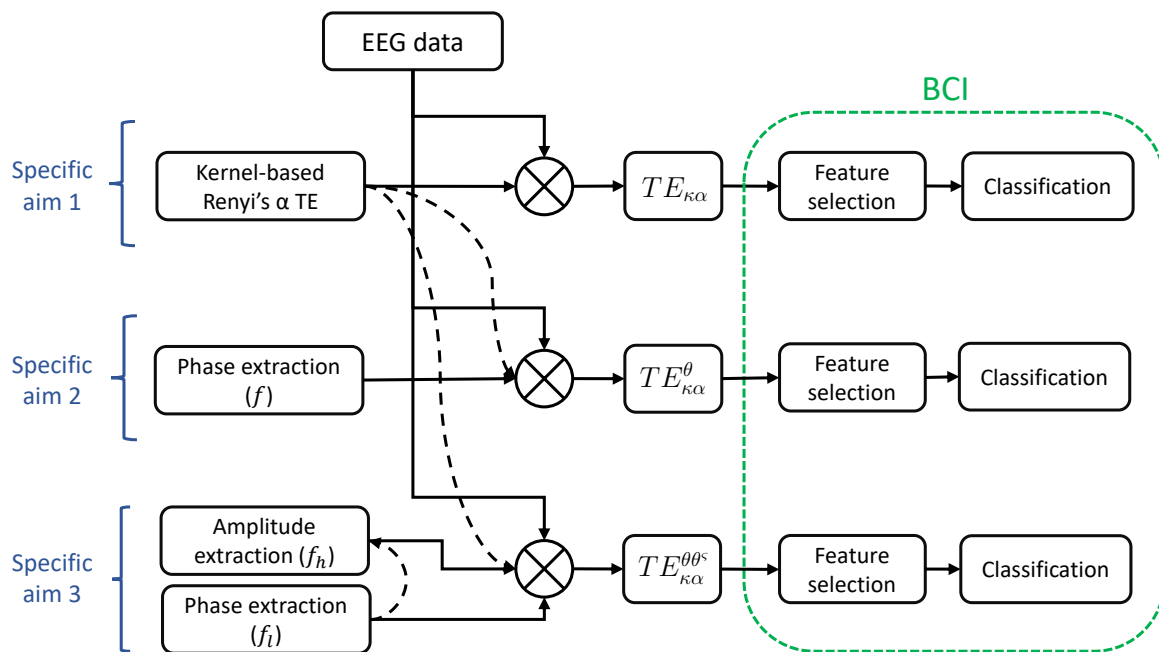


Figure 1.8: Schematic representation of the relationship between the aims of this thesis and the developed methodologies. $\text{TE}_{\kappa\alpha}$: kernel-based Renyi's α transfer entropy, $\text{TE}_{\kappa\alpha}^{\theta}$: Kernel-based Renyi's α phase transfer entropy (phase TE estimation through $\text{TE}_{\kappa\alpha}$), and $\text{TE}_{\kappa\alpha}^{\theta\theta^{\circ}}$: TE estimation for directed phase-amplitude interactions posed as a phase TE problem and obtained through $\text{TE}_{\kappa\alpha}$. In order to test the performance of the proposed approaches on EEG data, as BCI characterization strategies, additional feature selection and classification stages are included.

A Python implementation of the proposed $\text{TE}_{\kappa\alpha}$, $\text{TE}_{\kappa\alpha}^{\theta}$, and $\text{TE}_{\kappa\alpha}^{\theta\theta^s}$ approaches is avail-

able at https://github.com/ide2704/ITL_framework_Renyi_effective_connectivity (accessed on November 28, 2021).

1.6.1 Kernel-based Renyi’s transfer entropy ($\text{TE}_{\kappa\alpha}$)

Like all information theoretic quantities, TE is defined regarding the probability distributions of the systems under study, which in practice are unknown and must be estimated from data. Commonly used methods for TE estimation rely on a local approximation of the probability distributions from nearest neighbor distances, or on symbolization schemes that then allow the probabilities to be estimated from the symbols’ relative frequencies (Dimitriadis et al., 2016a; Kraskov et al., 2004; Lindner et al., 2011). However, probability estimation is a challenging problem, and avoiding this intermediate step in TE computation is desirable. Our first contribution consists of proposing a novel TE estimator using functionals defined on kernel matrices that approximate Renyi’s entropy measures of order α (De La Pava Panche et al., 2019; Giraldo et al., 2015), after defining TE in terms of Renyi’s entropy instead of the usual Shannon entropy-based definition. Our data-driven approach estimates TE directly from data, sidestepping the need for probability distribution estimation. Also, the proposed estimator encompasses the well-known definition of TE as a sum of Shannon entropies in the limiting case when $\alpha \rightarrow 1$. We test our proposal on a simulation framework consisting of two linear models, based on autoregressive approaches and a linear coupling function, respectively, and on two public EEG databases, obtained under motor imagery (MI) and visual working memory (WM) paradigms (see Section 1.7 for details). For the synthetic data, the proposed kernel-based TE estimation method satisfactorily identifies the causal interactions present in the data. Also, it displays robustness to varying noise levels and data sizes, and to the presence of multiple interaction delays in the same connected network. Obtained results for the EEG data show that our approach codes task-discriminant patterns, with classification performances that compare favorably to the state-of-the-art. This approach is related to the first specific aim, and it is described in Chapter 2.

1.6.2 Kernel-based Renyi’s phase transfer entropy ($\text{TE}_{\kappa\alpha}^{\theta}$)

Phase TE is commonly obtained from probability estimations carried out over data from multiple trials (Lobier et al., 2014), which bars its use as a characterization strategy in BCI. To address that issue, we propose a novel methodology to estimate TE between single pairs of instantaneous phase time series. Our approach combines the kernel-based

TE estimator defined in terms of Renyi's α entropy ($TE_{\kappa\alpha}$) with phase time series obtained by complex filtering the neural signals (De La Pava Panche et al., 2021a). We tested this proposal on simulated coupled data and on the two public EEG databases, recorded under MI and visual WM paradigms, referenced in Section 1.6.1. Attained results demonstrate how the introduced methodology succeeds in detecting the interactions present in the synthetic data, with statistically significant results around the frequencies of interest. It also reflects differences in coupling strength, is robust to realistic noise and signal mixing levels, and captures bidirectional interactions of localized frequency content. Obtained results for the MI and WM databases show that our approach, combined with a relevance analysis strategy (Fernández-Ramírez et al., 2020), codes discriminant spatial and frequency-dependent patterns for the different conditions in each experimental paradigm. This approach is related to the second specific aim, and it is described in Chapter 3.

1.6.3 Kernel-based Renyi's phase transfer entropy for the estimation of directed phase-amplitude interactions ($TE_{\kappa\alpha}^{\theta\theta^c}$)

Phase-amplitude interactions are believed to allow for the transfer of information from large-scale brain networks, oscillating at low frequencies, to local, rapidly oscillating neural assemblies. A promising approach to estimate such interactions is to use TE. The conventional method to do so involves feeding instantaneous phase and amplitude time series, extracted at the target frequencies, to a TE estimator (Besserve et al., 2010; Martínez-Cancino et al., 2020). As our third contribution, we propose recasting the problem of directed phase-amplitude interaction detection as a phase TE estimation problem, under the hypothesis that estimating TE from data of the same nature, i.e. two phase time series, will improve the robustness to common confounding factors that affect connectivity measures, such as the presence of high noise levels (De La Pava Panche et al., 2021b). We implement our proposal using the kernel-based Renyi's α transfer entropy ($TE_{\kappa\alpha}$), which, as pointed out in Section 1.6.2, allows to successfully obtain single-trial phase TE estimates. We tested our approach on synthetic data generated through a simulation model capable of producing time series with directed phase-amplitude interactions between two given frequencies, and on the two EEG databases mentioned in Sections 1.6.1 and 1.6.2. Our proposal detects statistically significant interactions between the simulated signals at the desired frequencies, identifying the correct direction of interaction. Also, it displays higher robustness to noise than the alternative methods

tested. Regarding the results for the EEG databases, our proposal is unable to capture discriminant interactions for the MI data. The same is true for the alternative methods explored. However, attained results for the WM data showed that the proposed approach codes connectivity patterns, based on directed phase-amplitude interactions, that allow discriminating among the different cognitive load levels of the task. This approach is related to the third specific aim, and it is described in Chapter 4.

1.7 EEG databases

In order to test the performance of the proposed $TE_{\kappa\alpha}$, $TE_{\kappa\alpha}^{\theta}$, and $TE_{\kappa\alpha}^{\theta\theta^s}$ approaches as characterization strategies, in the context of BCI, we obtain effective connectivity features from EEG signals recorded under two different cognitive paradigms: the first one consisting of motor imagery (MI) tasks, and the second one of a change detection task designed to study visual working memory (WM). In that context, our objectives are to set up classification systems that allow discriminating between the conditions in each paradigm, using as inputs relevant directed interactions among EEG signals, and then evaluate their performance in relation to the connectivity measures used to characterize the data. To those ends, we employ two publicly available databases: the BCI Competition IV database 2a ¹ and a database from brain activity during visual working memory ². For ease of reference, in this section we describe both databases, as well as the respective pre-processing pipelines applied before estimating brain connectivity.

1.7.1 Motor imagery

Motor imagery (MI) is the process of imagining a motor action without any motor execution. During an MI task, a subject visualizes in their mind an instructed motor action, i.e., to move the right hand, without actually carrying it out. When subjects plan and execute movements, characteristic rhythms in the sensorimotor areas, typically the μ or precentral α rhythm (8–12 Hz) and the β rhythm (13–30 Hz), get activated (Xu et al., 2020). That is to say, MI and motor execution share common sensorimotor areas, and both involve envisioning and executing the same motor plan (García-Murillo et al., 2021). Although, their neural mechanisms seem to have some differences (Matsuo et al., 2021). Assessing and interpreting MI brain dynamics may contribute to applications

¹<http://www.bbc.de/competition/iv/index.html> (accessed on November 28, 2021)

²<https://data.mendeley.com/datasets/j2v7btchdy/2> (accessed on November 28, 2021)

like the evaluation of pathological conditions, the rehabilitation of motor functions, and motor learning and performance (Collazos-Huertas et al., 2020). Particularly, much attention has been paid in the literature to BCI systems that can decode MI-associated task patterns, usually captured through scalp EEG signals, and translate them into commands in order to control external devices (Galindo-Noreña et al., 2020; Xu et al., 2020). One of the main limitations for the widespread use of such systems being that about 15–30% of users display BCI illiteracy, i.e. they do not gain enough control over the interfaces, possibly because subjects with poor control performance do not exhibit discriminative task-related changes over the modulation of sensorimotor rhythms during the interval of MI responses (García-Murillo et al., 2021).

It is worth noting that, in general, brain connectivity-based characterization strategies for MI tasks perform poorly (Rathee et al., 2017), specially when compared with methods such as the common spatial patterns (Elasuty and Eldawlatly, 2015; Gómez et al., 2018; Li et al., 2018). Because of that, our interest in using the BCI Competition IV database 2a to evaluate the behaviour of the proposed TE estimation approaches stems mostly from some characteristics of the database itself. First of all, the MI experimental paradigm has a clear temporal structure, several seconds long. Any task-discriminant directed connectivity values should track the paradigm’s time course, that is to say, they should be observed mainly in the time segment when subjects are performing the MI tasks. Second of all, the database is well known (Tangermann et al., 2012), and multiple works that use it show a consistent result in one regard: there is a subgroup of subjects in the database for which many different characterization strategies lead to high classification performances, while for a different subgroup, they lead to poor results (Elasuty and Eldawlatly, 2015; Galindo-Noreña et al., 2020; Gómez et al., 2018; Li et al., 2018; Liang et al., 2016). It is reasonable to assume that if directed interactions among different regions of the sensorimotor area are involved in MI, then the high performance subgroup of subjects should display, to a degree, similar discriminant connections. Lastly, since the MI-tasks information is associated to the sensorimotor rhythms, the proposed spectral representations for TE at those frequency ranges should hopefully lead to results at least comparable to those obtained through $TE_{\kappa\alpha}$.

BCI Competition IV database 2a

The BCI Competition IV database 2a (Tangermann et al., 2012) comprises EEG data from 9 healthy subjects recorded during an MI paradigm consisting of four different MI tasks, namely, imagining the movement of the left hand, the right hand, both feet,

or the tongue. Each trial of the paradigm starts with a fixation cross displayed on a computer screen, along with a beep. At second 2, a visual cue appears on the screen for a period of 1.25 s (an arrow pointing left, right, down, or up, corresponding to one of the four MI tasks). The cue prompts the subject to perform the indicated MI task until the cross vanishes from the screen at second 6. A representation of the paradigm's time course is shown in Figure 1.9A. Each subject performed 144 trials per MI task. The EEG data are acquired at a sampling rate of 250 Hz, from 22 Ag/AgCl electrodes placed according to the international 10/20 system, as depicted in Figure 1.9B. Next, the data are bandpass-filtered between 0.5 Hz and 100 Hz. A 50 Hz Notch filter is also applied. For each subject, the database contains a training dataset and a testing dataset, obtained following the same experimental paradigm. The former is intended to be used to train the MI task classification system, while the latter should be used to test the performance of the trained system (Gómez et al., 2018; Tangermann et al., 2012).

Data pre-processing In this thesis, we consider a bi-class classification problem involving the left and right hand MI tasks, so we drop the trials associated with the feet and the tongue. Afterward, we also drop the trials marked for rejection in the database itself (Tangermann et al., 2012). Then, we perform a windowing procedure in order to both better capture the temporal dynamics of the MI task, which has several distinct stages, and to favor the stationarity of the EEG signals to be analyzed. We segment each EEG trial into six-time windows 2 seconds long with 50% overlap, using a square window, obtaining six segments of equal length, as schematized in Figure 1.9A. Figure 1.9C displays an example of one of the obtained segments. Finally, we compute the surface Laplacian of each segmented trial using the spherical spline method for source current density estimation (Perrin et al., 1989) (see Appendix B for details). The surface Laplacian reduces the effects of volume conduction by attenuating low spatial frequency activity, and therefore, it also reduces the presence of spurious connections associated with it in connectivity analyses (Cohen, 2015; Rathee et al., 2017).

1.7.2 Working memory

Working memory (WM) is a memory system of limited capacity with the ability to store and manipulate information for a short period of time (Baddeley, 2012; Pavlov and Kotchoubey, 2020). It plays a key role in complex cognitive tasks such as comprehension, reasoning, planning and learning (Johnson et al., 2019; Zhang et al., 2016), as well as in daily activities such as problem solving and decision-making (Dai et al.,

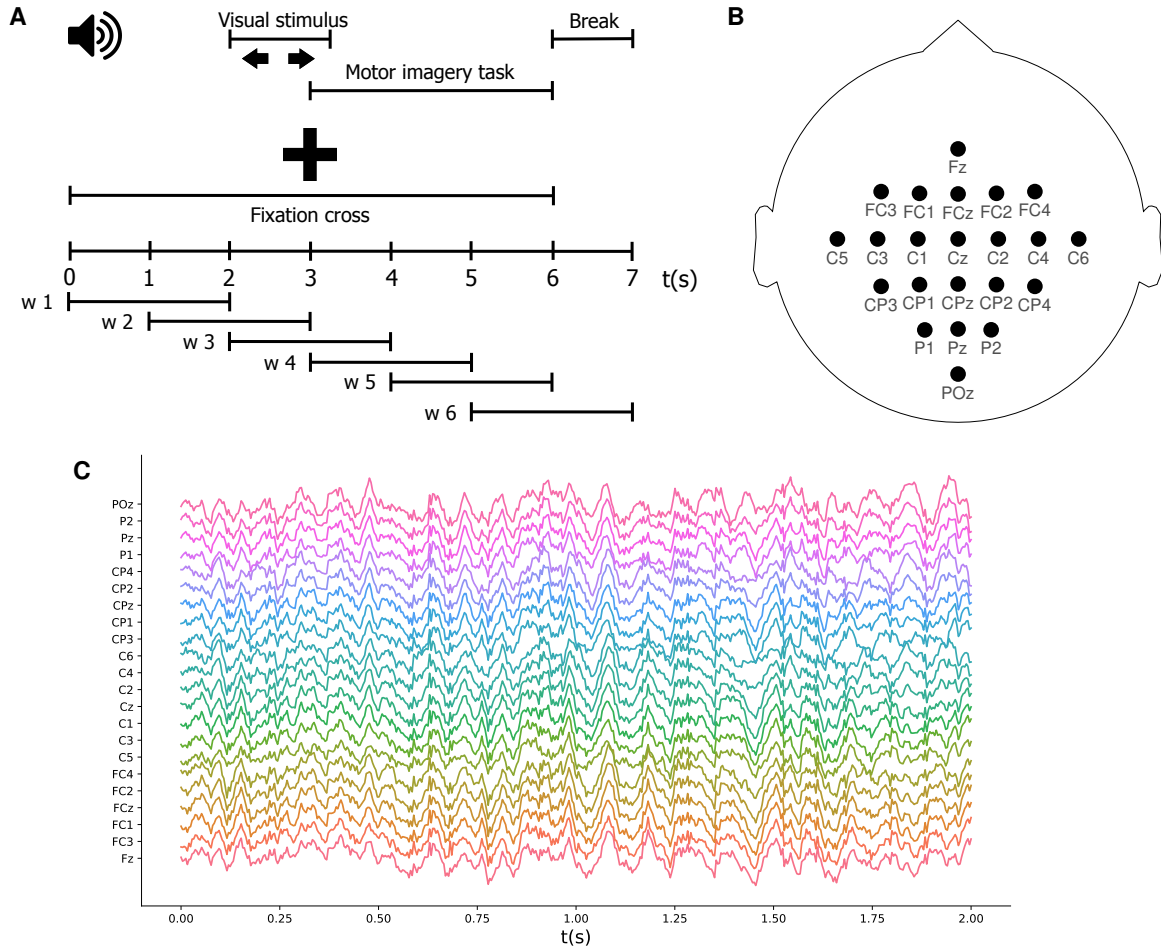


Figure 1.9: (A) Schematic representation of the MI protocol. (B) EEG channel montage used for the acquisition of the MI dataset. (C) Example of a 2 seconds long segmented EEG trial after data pre-processing (without surface Laplacian).

2017). WM consists of three distinct stages of information processing: encoding, maintenance or retention, and retrieval (Johnson et al., 2018), with the retention interval being considered as a defining component of WM, since it differentiates it from other memory types (Pavlov and Kotchoubey, 2020). Furthermore, WM is a cross-modal construct, that is to say, it depends on the stimulus modality (Pavlov and Kotchoubey, 2020). The most widely recognized model of WM (Baddeley, 2012) describes it as a several separate but interacting subsystems: a central component (central executive), two stimuli dependent storage subsystems (the phonological loop and the visuospatial sketchpad), and a system of limited capacity that allows the interaction between the other components (episodic buffer) (Toppi et al., 2018). Moreover, multiple studies have found that neural oscillatory activity in a wide range of frequencies is modulated during

WM (Daume et al., 2017). High frequency activity, in the β (13–30 Hz) and γ (>40 Hz) bands, seem to play a role in encoding, retrieval, and maintenance of the stimulus; while activity in lower frequencies, in the θ (4–7 Hz) and α (8–12 Hz) ranges, especially in frontal areas, is associated with the coordination and integration of different cognitive processes during the execution of WM tasks (Dai et al., 2017). This has led to hypotheses about cross-frequency coupling mechanisms underpinning WM, with oscillatory activity in the central executive component interacting with oscillations at other frequencies in the peripheral storage systems (Dimitriadis et al., 2016a,b). In particular, PAC interactions are thought to play a crucial role in WM (Liang et al., 2021), with bidirectional interactions linking the prefrontal cortex to parieto-occipital and medial temporal regions (Johnson et al., 2018, 2019).

The hypotheses about the underlying mechanisms of WM involving the transfer of information between distant brain areas, possibly through cross-frequency interactions, make it a prime target to test the TE estimation approaches developed in this thesis. Also, the characterization of WM through brain connectivity measures is of special interest to our research group, as mentioned in Section 1.1.

Database from brain activity during visual working memory

The database from brain activity during visual working memory, presented in Villena-González et al. (2020), contains EEG data recorded from twenty-three subjects, with normal or corrected-to-normal vision, and without color-vision deficiency, while performing multiple trials of a change detection task (Vogel and Machizawa, 2004). The task consists of remembering the colors of a set of squares, termed memory array, and then comparing them with the colors of a second set of squares located in the same positions, termed test array. A trial of the task begins with an arrow indicating either the left or the right side of the screen. Then, a memory array appears on the screen for 0.1 s. For every trial, memory arrays are displayed on both hemifields, but the subject must remember only those appearing on the side indicated by the arrow cue. Next, after a retention period lasting 0.9 s, a test array appears. The subject then reports if the colors of all the items in the memory and test arrays match. The task has three levels according to the number of elements in the memory array: low memory load (one square), medium memory load (two squares), and high memory load (four squares). A representation of the above-described experimental paradigm is depicted in Figure 1.10A. The color of one of the squares in the test array differs from its counterpart in the memory array in 50% of the trials. Each subject performed a total of 96 trials, with 32 trials

for each memory load level. The EEG data are acquired at a sampling rate of 2048 Hz, using 64 electrodes (Biosemi ActiveTwo) arranged according to the international 10/20 extended system. Besides the EEG data, the database provides recordings from four EOG channels and two external electrodes located on the left and right mastoids.

Data pre-processing First, we re-reference the data to the average of the mastoid channels. Next, we bandpass-filter the data between 0.01 Hz and 20 Hz using a Butterworth filter of order 2. Afterward, we extract the trial information from the continuous EEG data using a 1.4 s squared window. Each trial segment starts 0.2 s before the presentation of the memory array. Then, we perform a visual inspection of the data and discard two subjects (subjects number 11 and 17) because of the presence of strong artifacts in a very large number of trials. Subjects number 22 and 23 are reassigned as subjects 17 and 11, respectively. After that, we remove ocular artifacts from the EEG data by performing independent component analysis (ICA) on it and then eliminating the components that more closely resemble the information provided by the EOG data (Villena-González et al., 2020). Then, we discard all incorrect trials, i.e., trials for which the subjects incorrectly matched the memory and test arrays. Next, we select 32 out of the 64 channels in the EEG data, as shown in Figure 1.10B. Then, we downsample the data to 1024 Hz, and segment, for each trial, the time window starting 0.3 seconds after the onset of the memory array and ending just before the presentation of the test array (see Figure 1.10A). The 0.7 seconds long segments cover most of the retention interval, the period when the subjects should maintain the stimulus information in their working memories, while leaving out any purely sensory responses elicited immediately after the presentation of the stimulus. An example of segmented data trial is shown in Figure 1.10C. Finally, with the aim of reducing the presence of spurious connections associated with volume conduction effects, we compute the surface Laplacian of each trial.

1.8 Thesis structure

The remainder of this thesis is organized as follows. In Chapter 2, we introduce the kernel-based Renyi's transfer entropy estimator ($TE_{\kappa\alpha}$), which allows obtaining TE estimates while sidestepping the need for probability distribution computation. Chapter 3 develops an approach to obtain single-trial phase TE values, termed kernel-based Renyi's phase transfer entropy ($TE_{\kappa\alpha}^\theta$), that combines the $TE_{\kappa\alpha}$ estimator with instantaneous

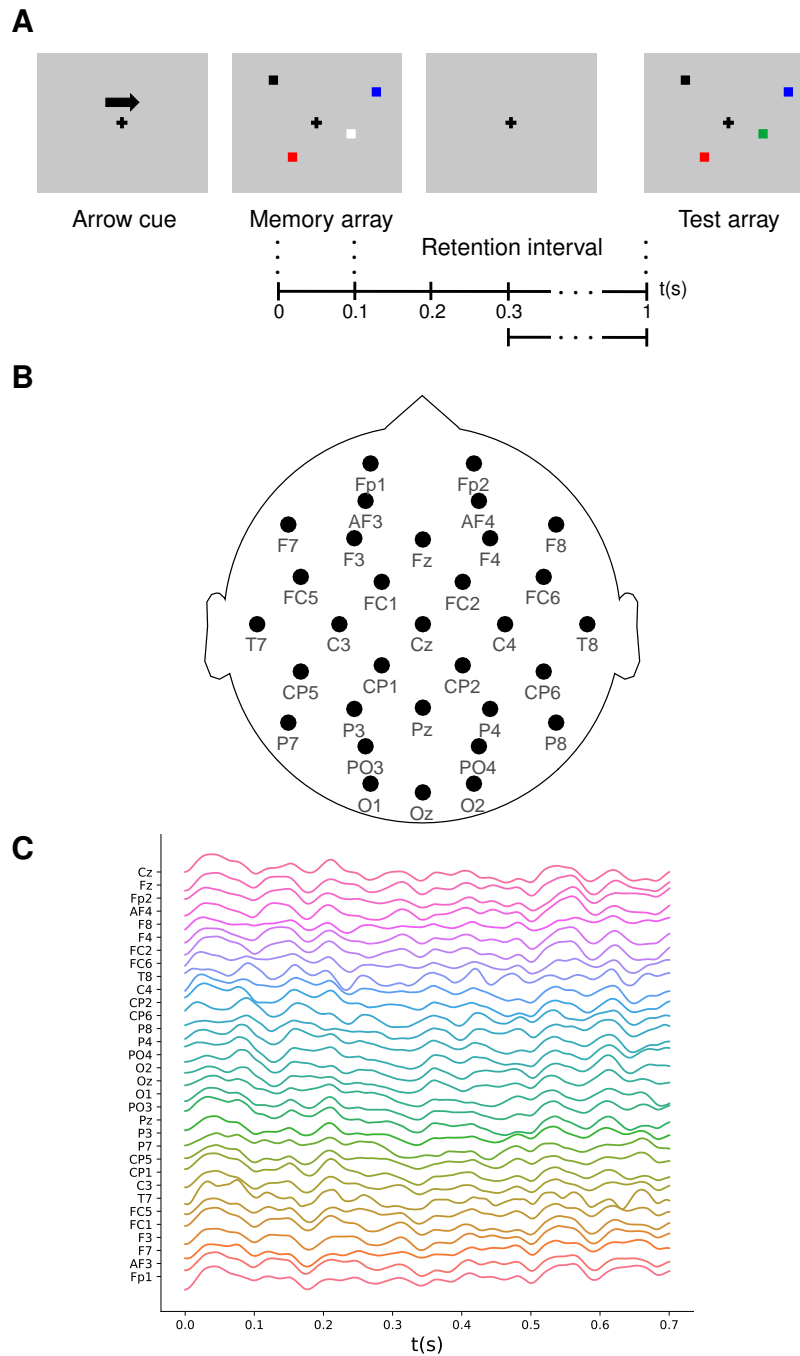


Figure 1.10: (A) Graphical representation of the change detection task for WM. (B) EEG channels selected from the montage used for the acquisition of the WM database. (C) Example of a 0.7 seconds long segmented EEG trial after data pre-processing (without surface Laplacian).

phase times series extracted from complex-filtered data. Chapter 4 describes the use of

the kernel-based Renyi's phase transfer entropy approach for the estimation of directed phase-amplitude interactions ($\text{TE}_{\kappa\alpha}^{\theta\theta^*}$). Finally, in Chapter 5 we present our conclusions, outline some possible venues of future work, and list the academic products associated with this thesis.

Chapter 2

Kernel-based Renyi's transfer entropy

In this chapter, we propose a data-driven TE estimator that sidesteps the need to obtain the probability distributions underlying the data. We begin by expressing TE as a linear combination of Renyi's entropy measures of order α (Principe, 2010; Rényi et al., 1961), instead of using the standard definition in terms of Shannon entropies. Renyi's entropy is a mathematical generalization of the concept of Shannon entropy. It corresponds to a family of entropies that, because of its functional dependence on the parameter α , can emphasize either mean behavior and slowly change features in the data, or rare, uncommon events (Gao et al., 2011; Giraldo et al., 2015). This flexibility gives Renyi's entropy an advantage when it comes to analyzing data from biomedical systems (Liang et al., 2015), and has been exploited in neuroscience studies, for instance, to better characterize the randomness of EEG signals in childhood absence epilepsy (Mammone et al., 2015), and to track EEG changes associated with different anesthesia states (Liang et al., 2015). Renyi's entropy has also been employed as an EEG feature extraction strategy in automatic systems for the diagnosis of epilepsy (Acharya et al., 2015), and for the assessment of cognitive workload (Zarjam et al., 2013). Afterward, we approximate Renyi's entropy through a functional defined on positive definite and infinitely divisible kernels matrices, introduced in Giraldo et al. (2015). The obtained estimator computes TE directly from the kernel matrices that, in turn, capture the similarity relations among data. Also, because of the definition of Renyi's entropy, the proposed approach encompasses the conventional formulation of TE as a sum of Shannon entropies in the limiting case when $\alpha \rightarrow 1$.

In order to test our proposal, we use a simulation framework consisting of two linear

models, based on autoregressive approaches and a linear coupling function, respectively, and on the two EEG databases described in Section 1.7. In particular, we aim to test whether our method fulfills the requirements established in Vicente et al. (2011) for a TE estimator suited for neuroscience data. Namely, it must be robust to moderate levels of noise, it must rely on a limited number of data samples, and it must be reliable when dealing with high dimensional spaces. For the synthetic data, the proposed kernel-based TE estimation method successfully detects the presence and direction of the causal interactions defined in the models. Additionally, it displays robustness to varying noise levels and data sizes, in terms of the available data samples, and to the presence of multiple interaction delays in the same connected network. The results for the MI data show that our approach codes discriminant spatiotemporal patterns for the left and right-hand motor imagination tasks, that are in accordance with the temporal structure of the MI paradigm. Finally, the obtained connectivity values for the WM data also capture information that allows discriminating among the different levels of the task.

2.1 Kernel-based Renyi's transfer entropy

We begin by generalizing the concept of TE from Shannon entropies to Renyi's α -order entropies. Then, we propose a TE estimator using the entropy-like functionals detailed in Section 1.3.3, thus avoiding the intermediate step of probability distribution estimation in the computation of TE from discrete data. Given the state space reconstructions $\mathbf{x}_t^{dx}, \mathbf{y}_t^{dy} \in \mathbb{R}^{D \times d}$ of two time series \mathbf{x} and \mathbf{y} , the flow of information from \mathbf{x} to \mathbf{y} , for an interaction time u , corresponds to the deviation from the following equality: $p(y_t | \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}) = p(y_t | \mathbf{y}_{t-1}^{dy})$. Now, instead of explicitly applying the definition of the Kullback-Leibler divergence, as in the standard derivation of TE, we apply the expected value operator over the logarithm of the probability distributions, yielding:

$$\mathbb{E}_{y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}} \left\{ -\log \left(p(y_t | \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}) \right) \right\} = \mathbb{E}_{y_t, \mathbf{y}_{t-1}^{dy}} \left\{ -\log \left(p(y_t | \mathbf{y}_{t-1}^{dy}) \right) \right\}. \quad (2.1)$$

Using the relations between conditional, joint and marginal probabilities, and rewriting the logarithms of the obtained quotients, we arrive at:

$$\begin{aligned} & \mathbb{E}_{y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}} \left\{ -\log \left(p(y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}) \right) \right\} - \mathbb{E}_{\mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}} \left\{ -\log \left(p(\mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}) \right) \right\} = \\ & \mathbb{E}_{y_t, \mathbf{y}_{t-1}^{dy}} \left\{ -\log \left(p(y_t, \mathbf{y}_{t-1}^{dy}) \right) \right\} - \mathbb{E}_{y_t, \mathbf{y}_{t-1}^{dy}} \left\{ -\log \left(p(\mathbf{y}_{t-1}^{dy}) \right) \right\}. \end{aligned} \quad (2.2)$$

The deviation from the above equality corresponds to transfer entropy, thus:

$$TE(\mathbf{x} \rightarrow \mathbf{y}) = \mathbb{E}_{\mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}} \left\{ -\log \left(p(\mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}) \right) \right\} - \mathbb{E}_{y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}} \left\{ -\log \left(p(y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}) \right) \right\} + \mathbb{E}_{y_t, \mathbf{y}_{t-1}^{dy}} \left\{ -\log \left(p(y_t, \mathbf{y}_{t-1}^{dy}) \right) \right\} - \mathbb{E}_{y_t, \mathbf{y}_{t-1}^{dy}} \left\{ -\log \left(p(\mathbf{y}_{t-1}^{dy}) \right) \right\} \quad (2.3)$$

From the general definition of entropy, $H(x) = \mathbb{E} \{-\log(p(x))\}$, and assuming an expected value associated with the function $g(x) = c2^{(1-\alpha)x}$, we can express TE as a sum of Renyi's α -order entropies:

$$TE_\alpha(\mathbf{x} \rightarrow \mathbf{y}) = H_\alpha(\mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}) - H_\alpha(y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}) + H_\alpha(y_t, \mathbf{y}_{t-1}^{dy}) - H_\alpha(\mathbf{y}_{t-1}^{dy}). \quad (2.4)$$

In the limiting case when $\alpha \rightarrow 1$, Equations 1.8 and 2.4 are equivalent (TE_α yields the well-known TE). Finally, using the kernel-based formulation of Renyi's α -order entropy for marginal and joint probability distributions (Equations 1.14 and 1.15, respectively), we can estimate the TE_α from \mathbf{x} to \mathbf{y} as:

$$TE_{\kappa\alpha}(\mathbf{x} \rightarrow \mathbf{y}) = H_\alpha(\mathbf{K}_{\mathbf{y}_{t-1}^{dy}}, \mathbf{K}_{\mathbf{x}_{t-u}^{dx}}) - H_\alpha(\mathbf{K}_{y_t}, \mathbf{K}_{\mathbf{y}_{t-1}^{dy}}, \mathbf{K}_{\mathbf{x}_{t-u}^{dx}}) + H_\alpha(\mathbf{K}_{y_t}, \mathbf{K}_{\mathbf{y}_{t-1}^{dy}}) - H_\alpha(\mathbf{K}_{\mathbf{y}_{t-1}^{dy}}) \quad (2.5)$$

where the kernel matrices $\mathbf{K}_{y_t}, \mathbf{K}_{\mathbf{y}_{t-1}^{dy}}, \mathbf{K}_{\mathbf{x}_{t-u}^{dx}} \in \mathbb{R}^{(D-u) \times (D-u)}$ hold elements $k_{ij} = \kappa(\mathbf{a}_i, \mathbf{a}_j)$, with $k_{ij}(\cdot, \cdot)$ a positive definite, infinitely divisible kernel function. For matrix \mathbf{K}_{y_t} , $a_i, a_j \in \mathbb{R}$ are the values of the time series \mathbf{y} at times i and j . In the case of matrix $\mathbf{K}_{\mathbf{y}_{t-1}^{dy}}$, the vectors $\mathbf{a}_i, \mathbf{a}_j \in \mathbb{R}^d$ contain the space state reconstruction \mathbf{y}_t^{dy} of \mathbf{y} at times i and j , adjusted according to the time indexing of TE. Likewise for $\mathbf{K}_{\mathbf{x}_{t-u}^{dx}}$.

2.2 Experiments

2.2.1 VAR model

In order to test the ability of the $TE_{\kappa\alpha}$ functional in Equation 2.5 to detect directed interactions under varying noise and data size conditions, we perform two experiments on simulated data. We generate synthetic data from a unidirectional bivariate autoregressive (AR) model of order 3:

$$\mathbf{z}_t = \mathbf{c} + \sum_{i=1}^3 \mathbf{Q}^i \mathbf{z}_{t-i} + \boldsymbol{\varepsilon}_t, \quad (2.6)$$

where $\mathbf{z}_t = (x_t, y_t)^T$ is a vector with the values of the simulated signals, $\mathbf{x} \in \mathbb{R}^l$ and $\mathbf{y} \in \mathbb{R}^l$, at time t , $\boldsymbol{\varepsilon}_t \in \mathbb{R}^2$ is a vector of white noise values at time t , $\mathbf{c} \in \mathbb{R}^2$ is vector of constants, and

$$\mathbf{Q}^i = \begin{pmatrix} q_{11}^i & q_{12}^i \\ q_{21}^i & q_{22}^i \end{pmatrix} \quad ; \quad i = \{1, 2, 3\}, \quad (2.7)$$

holds the model parameters. The directionality of the causal relation between the simulated time series is controlled by setting to 0 either the parameters q_{12}^i , to obtain a causal relation from \mathbf{x} to \mathbf{y} , or q_{21}^i to obtain a causal relation in the opposite direction. The remaining parameters of the model are randomly selected. In order to assess the robustness of our method to different noise conditions, we add noise to the synthetic data as follows:

$$\mathbf{Z}_\eta = (1 - \gamma) \frac{\mathbf{Z}}{\|\mathbf{Z}\|_F} + \gamma \frac{\boldsymbol{\Theta} \boldsymbol{\Xi}}{\|\boldsymbol{\Theta} \boldsymbol{\Xi}\|_F}, \quad (2.8)$$

where $\mathbf{Z} \in \mathbb{R}^{2 \times l}$ is a matrix containing the signals \mathbf{x} and \mathbf{y} , $\|\cdot\|_F$ stands for the Frobenius norm, $\boldsymbol{\Theta} \in \mathbb{R}^{2 \times 3}$ is an instantaneous mixing matrix with random elements, and $\boldsymbol{\Xi} \in \mathbb{R}^{3 \times l}$ is a matrix containing 3 time series generated by 3 independent AR models of order 3 with otherwise random parameters, that represent multiple independent sources of noise and serve to simulate the effects of volume conduction. The parameter γ controls the relative strength of noise and signal (Dimitriadis et al., 2016a). If γ is assigned a scalar value then signals \mathbf{x} and \mathbf{y} will exhibit symmetric noise, that is to say, they will have the same noise level. Alternatively, if γ is assigned a two-dimensional vector value, and the two elements of the vector are different, then the noise levels in \mathbf{x} and \mathbf{y} will be asymmetric (in this case, to be able to use 2.8 we need to perform a column wise stacking of l copies of γ , and replace the scalar multiplication by a Hadamard product). In our first experiment we test both scenarios. First, we assign γ a scalar value that varies in the range from 0 to 1, in steps of 0.1, in order to simulate different symmetric levels of noise for signals of 512 data points. Then, to test the behavior of our TE estimator under asymmetric noise conditions, we assign γ a vector value and vary its two elements so as to form a two-dimensional grid, with each dimension ranging from 0 to 1, in steps of 0.1, for signals with the same number of data points as above. In the second experiment, we evaluate the impact of signal length on our method. To that end, we vary the length l of the noiseless simulated signals between 100 and 1000 data points, in steps of 100 data points. For both experiments, that is to say, for each noise level (in the symmetric and asymmetric cases) and signal length, we estimate the accuracy for 10 realizations of 100 trials each. For each realization, the direction of interaction is

chosen at random. The accuracy is defined in terms of a directionality index:

$$\Delta\lambda = \lambda(\mathbf{x} \rightarrow \mathbf{y}) - \lambda(\mathbf{y} \rightarrow \mathbf{x}), \quad (2.9)$$

where $\lambda(\cdot)$ stands for any of the effective connectivity measures under consideration. $\Delta\lambda$ indicates the preferred direction of information flow. It gets positive values for couplings from \mathbf{x} to \mathbf{y} , and negative values when \mathbf{y} drives \mathbf{x} . We use it to assess whether each effective connectivity measure correctly detects the chosen direction of interaction.

2.2.2 Modified linear Kus model

A method to estimate effective connectivity from multiple channel EEG data should be able to detect causal interactions among multiple signals coming from a connected network. With the aim of testing whether the proposed TE estimator could successfully reveal the presence, or absence, of such interactions in a known network, we use the modified version of the linear Kus model, introduced in [Weber et al. \(2017\)](#). It consists of 5 channels, connected through direct and indirect couplings (for a graphical representation of the model see Figure 2.4A). The input to the model is a time series containing real EEG data that is then contaminated with white Gaussian noise to obtain channel 1. Then, channel 1 is scaled and time-shifted by an interaction delay of 4 time units ($\delta = 4$), and more white Gaussian noise is added, to generate channel 2. Channels 3 and 4 are generated in a similar fashion, while channel 5 consists only of white Gaussian noise. The following set of equations describes all the network interactions present in the model:

$$\begin{aligned} x_1(t) &= \xi(t) + v\eta_1(t) \\ x_2(t) &= 0.4x_1(t-4) + v\eta_2(t) \\ x_3(t) &= 0.4x_2(t-4) + v\eta_3(t) \\ x_4(t) &= 0.4x_2(t-8) + v\eta_4(t) \\ x_5(t) &= v\eta_5(t) \end{aligned} \quad (2.10)$$

where x_j , ξ , and η_j stand for the 5 network channels, the input EEG data, and the added white Gaussian noise at time t , respectively. The parameter v is a scaling factor equivalent to a quarter of the variance of the time series. Additionally, external white Gaussian noise with zero mean and variance equal to v is added to all channels ([Kus et al., 2004](#); [Weber et al., 2017](#)). It is worth noting that the indirect couplings in the

model arise in two different ways. They can be the result of upstream dependences between the network’s channels. For instance, channel 1 generates channel 2, which in turn generates channel 3, giving rise to an indirect coupling between channels 1 and 3. Indirect couplings can also arise from different time shifts applied to one channel in order to generate new channels. Such is the case of the indirect coupling between channels 3 and 4, which are generated by time-shifting channel 2 by 4 and 8 time units, respectively.

For our experimental set-up, we generate 1000 trials of the modified Kus model, divided into 10 realizations. As input to the model, we use EEG data from the BCI Competition IV dataset 2a (for details about this dataset see Section 1.7.1). Namely, we pool together the Fz channels from all subjects and trials in the dataset, and for each realization randomly select 100 of them (without repetition), to be used as inputs to the system. Then, we generate 100 trials of the modified Kus model, each consisting of a 5 channel network. Next, for all pair-wise combinations of channels in each trial, we estimate the directed interactions within the elements of the network using our method, and the other effective connectivity measures under consideration. Afterward, for each realization of 100 trials, we perform a permutation test, based on randomized trial surrogates (see Appendix C for details), to determine which couplings or directed connections within the network are statistically significant at an alpha level of 0.05 (Lindner et al., 2011; Weber et al., 2017). The number of permutations in the test is set to 1000. Finally, in order to assess the overall performance of each method, regarding the detection of the true connections in the modified Kus model, we compare the statistically significant connections per realization with the predefined connections in the network to obtain accuracy, sensitivity, and specificity values.

2.2.3 EEG data

Motor imagery

Feature extraction For each subject, let $\Psi = \{\mathbf{X}_n \in \mathbb{R}^{C \times M}\}_{n=1}^N$ be the EEG set holding N trials of the MI tasks, with $C = 22$ channels, and $M = 1750$ samples. Besides, let $\{l_n\}_{n=1}^N$ be a label set where the n -th element corresponds to the motor imagery task indicated for trial \mathbf{X}_n (it takes the value 1 for right hand motor imagination, and 2 for left hand motor imagination). First, as detailed in Section 1.7.1, we perform a windowing procedure which yields a set of matrices $\{\mathbf{Z}_n^w \in \mathbb{R}^{C \times L}\}_{w=1}^Q$, where $Q = 6$, and $L = 500$. Thus, our goal is to estimate the class label from effective connectivity features extracted

from the segmented EEG trial \mathbf{Z}_n^w . Afterward, we compute the surface Laplacian of each segmented trial (Perrin et al., 1989). Then, for each pairwise combination of channels $\mathbf{z}_c, \mathbf{z}_{c'} \in \mathbb{R}^L$, belonging to the spatially filtered version of \mathbf{Z}_n^w , we estimate the effective connectivity $\lambda(\mathbf{z}_c \rightarrow \mathbf{z}_{c'})$ to build a connectivity matrix $\mathbf{\Lambda} \in \mathbb{R}^{C \times C}$. In the case when $c = c'$, we set $\lambda(\mathbf{z}_c \rightarrow \mathbf{z}_{c'}) = 0$. Next, for time window w and for the N trials of the MI task, we obtain a set of connectivity matrices $\{\mathbf{\Lambda}_n^w \in \mathbb{R}^{C \times C}\}_{n=1}^N$. Then, we normalize each $\mathbf{\Lambda}_n^w$ to the range $[0, 1]$. After that, we apply vector concatenation to $\mathbf{\Lambda}_n^w$ to yield a vector $\phi_n^w \in \mathbb{R}^{1 \times (C \times C)}$. Then, we stack together the N vectors ϕ_n^w , corresponding to each trial, to form a matrix $\Phi^w \in \mathbb{R}^P$, with $P = N \times (C \times C)$. Φ^w holds all directed interactions, estimated through the effective connectivity measure λ , for time window w , for the entire EEG dataset Ψ .

Feature selection and classification After characterizing the EEG data, we set up our subject dependent MI task classification system. As mentioned before, the classification is carried out separately for each time interval or time window w . Also, since the MI database has training and testing datasets, we divide our classification system into a training-validation stage and a testing stage. For the training-validation stage, we first specify a cross-validation scheme of 10 iterations. For each iteration, 70% of the trials of the training dataset are randomly assigned to a training set and the remaining 30% to a validation set. Then, we use CKA (see Appendix D for details) over the connectivity features obtained from the training set to generate a relevance vector $\boldsymbol{\rho} \in [0, 1]^P$ (P equals the number of features in Φ^w). Then, we use $\boldsymbol{\rho}$ to rank Φ . Next, we select a varying percentage of the ranked features, from 5% to 100% in 5% steps, and input them to the classification algorithm. The features associated with the highest values of $\boldsymbol{\rho}$ are input first, and as the percentage of features increases those associated with lower values of $\boldsymbol{\rho}$ are progressively included. In this work, we use a support vector classifier (SVC) with an RBF kernel (Pedregosa et al., 2011). All classification parameters, including the percentage of discriminant features, are tuned at this stage through a grid search. We select the parameters according to the classification accuracy, aiming to improve the system's performance. Then, for the testing stage, we train an SVC using the connectivity features from all trials in the training dataset as well as the parameters found in the previous stage. Lastly, we quantify the performance of the trained system in terms of the classification accuracy, obtained after predicting the MI task class labels of the testing dataset from its connectivity features.

Working memory

Feature extraction Let $\Psi = \{\mathbf{X}_n \in \mathbb{R}^{C \times M}\}_{n=1}^N$ be an EEG set holding N trials of the WM dataset, recorded from a single subject, where $C = 32$ stands for the number of channels and $M = 717$ corresponds to the number of samples. In addition, let $\{l_n\}_{n=1}^N$ be a set whose n -th element is the label associated with trial \mathbf{X}_n . Thus, l_n can take the values of 1, 2, and 3 corresponding to low, medium, and high memory loads. Our goal is to estimate the class label from relevant effective connectivity features extracted from \mathbf{X}_n . Again, let $\lambda(\mathbf{x}_c \rightarrow \mathbf{x}_{c'})$ be a measure of effective connectivity between channels $\mathbf{x}_c, \mathbf{x}_{c'} \in \mathbb{R}^M$. By computing $\lambda(\mathbf{x}_c \rightarrow \mathbf{x}_{c'})$ for each pairwise combination of channels in \mathbf{X}_n we obtain a connectivity matrix $\mathbf{\Lambda} \in \mathbb{R}^{C \times C}$. In the case when $c = c'$, we set $\lambda(\mathbf{x}_c \rightarrow \mathbf{x}_{c'}) = 0$. Then, we normalize $\mathbf{\Lambda}$ to the range $[0, 1]$. After performing this procedure for the N trials, we get set of connectivity matrices $\{\mathbf{\Lambda}_n \in \mathbb{R}^{C \times C}\}_{n=1}^N$. Then, we apply vector concatenation to $\mathbf{\Lambda}_n$ to yield a vector $\phi_n \in \mathbb{R}^{1 \times (C \times C)}$. Next, we stack the N vectors ϕ_n , corresponding to each trial, to obtain a matrix $\Phi \in \mathbb{R}^P$, with $P = N \times (C \times C)$, that characterizes the EEG set Ψ in terms of directed interactions estimated through λ . A graphical representation of the above-described steps, as well as of our overall classification setup, is depicted in Figure 2.1.

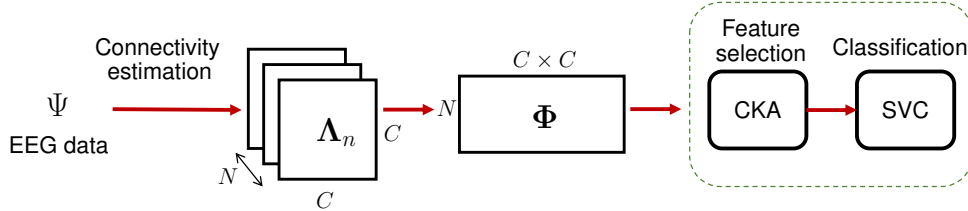


Figure 2.1: Schematic representation of our overall classification setup. Notice that for the MI dataset Φ varies according to the time window w .

Feature selection and classification The classification system we set up for the WM data closely resembles the one previously detailed for the MI data, with three changes. First, the WM database consists of one set of data for each subject, instead of two, so there is only a training-validation stage. Second, given the reduced number of trials available for each memory load level, each of the 10 iterations of the cross-validation scheme follows an 80%-20% split for the training and validation sets (instead of a 70%-30% split). Third, since the results provided by CKA are not stable for the low number of trials available from each subject (less than 30 trials per class, on average),

we opted to add an auxiliary cross-validation step, with the same characteristics as the one described above, and use it to estimate a single relevance vector $\bar{\mathbf{q}}$, obtained as the average of the relevance vectors of each data split. Then, we use $\bar{\mathbf{q}}$ to perform feature selection in every iteration of the main cross-validation scheme.

2.2.4 Parameter selection

We performed all the experiments mentioned above for two connectivity measures, namely TE and GC. For TE, we tested three different estimation strategies: the Kraskov-Stögbauer-Grassberger method (TE_{KSG}), the symbolic version of TE based on ordinal pattern symbolization (TE_{Sym}), and the proposed kernel-based Renyi's Transfer Entropy ($\text{TE}_{\kappa\alpha}$). For the latter, we explored the following values of the α parameter: $\alpha = \{1.01, 2, 3, 5, 10\}$, with special emphasis on $\alpha = 1.01$ and $\alpha = 2$, using as kernel function the Gaussian radial basis function or Gaussian RBF kernel with Euclidean distance (Liu et al., 2011):

$$\kappa(\mathbf{a}_i, \mathbf{a}_j) = \exp\left(-\frac{d^2(\mathbf{a}_i, \mathbf{a}_j)}{2\sigma^2}\right), \quad (2.11)$$

where $d^2(\cdot, \cdot)$ is the distance operator. We used in-house Python and Matlab implementations of the algorithms for GC, TE_{Sym} , and $\text{TE}_{\kappa\alpha}$; while for TE_{KSG} we used the implementation provided by the open access toolbox TRENTOOL, a TE estimation and analysis toolbox for Matlab (Lindner et al., 2011).

Regarding the selection of parameters involved in the different effective connectivity estimation methods, we proceeded as follows: For the TE methods, the embedding delay τ was set to 1 autocorrelation time (ACT) (Vicente et al., 2011). The embedding dimension d and the interaction delay u were set in an experiment-dependent fashion, in most cases after a heuristic search intended to maximize performance. For all experiments, except for the WM data, d was set to 3 after heuristic searches in the range $d = \{1, 2, \dots, 10\}$. For the VAR model experiment and the MI tasks experiment u was set to 1, after heuristic searches in the ranges $u = \{1, 2, 3\}$ and $u = \{1, 2, \dots, 100\}$, respectively. While for the Kus model experiment, u was set to 4, because that is the most common delay present in the model's network. For the WM data, d was selected using Cao's criterion (Cao, 1997; Lindner et al., 2011) from the same range as the other experiments (see Appendix E for details), and u was set to 120, a value close to the average interaction times found in Dimitriadis et al. (2016a) for a visuospatial WM task.

The number of neighbors K , and the Theiler correction window in TRENTOOL's implementation of the TE_{KSG} algorithm were left at their default values of 4 and 1 ACT, respectively (Lindner et al., 2011). The bandwidth σ in the RBF kernel introduced in Equation 2.11, for the proposed $TE_{\kappa\alpha}$ method, was set in each case as the median distance of the data (Schölkopf et al., 2002). The order of the autoregressive model o for GC was set to 3 for all experiments. In the case of the VAR model experiment $o = 3$ was chosen to coincide with the order of the data generation model, while for the Kus model and the EEG data it was the result of heuristic searches in the range $o = \{1, 3, 5, 7, 9\}$. Finally, the two values of the parameter α emphasized throughout the experiments were selected with the following rationale: as $\alpha \rightarrow 1$ Renyi's entropy tends to Shannon's entropy, so a value of $\alpha = 1.01$ should allow for a better comparison with Shannon's entropy-based TE estimation strategies. Also, for Renyi's entropy a value of $\alpha = 2$ is considered to be neutral to weighting (Giraldo et al., 2015), i.e. it does not emphasize or penalize rare events, which makes $\alpha = 2$ a convenient choice when there is no previous knowledge about the values of the α parameter better suited for a particular application.

2.3 Results and discussion

2.3.1 VAR model

The experiments described in Section 2.2.1 test whether the effective connectivity measures under consideration correctly detect the direction of interaction between two time series, under varying noise and data size conditions. Figures 2.2 and 2.3 present the results of such experiments. Figures 2.2A and 2.2C show the obtained average accuracies regarding the detection of the preferred direction of information flow as the scalar γ parameter in Equation 2.8, and thus the amount of symmetric noise added to the simulated signals, increases from 0 to 1. For all the methods tested the performance peaks for low noise levels and progressively falls as the noise level increases. At $\gamma = 1$ the average accuracies reach values of around 50%, which reflects the fact that for $\gamma = 1$ noise completely replaces the signals generated by the VAR model, and therefore no causal interaction is present. Figures 2.2B and 2.2D show the average accuracies obtained with the effective connectivity measures tested as the number of data points of the VAR signals increases. In all cases, the performance is lowest for the lowest number of data points considered (100), and increases as the simulated signals become lengthier.

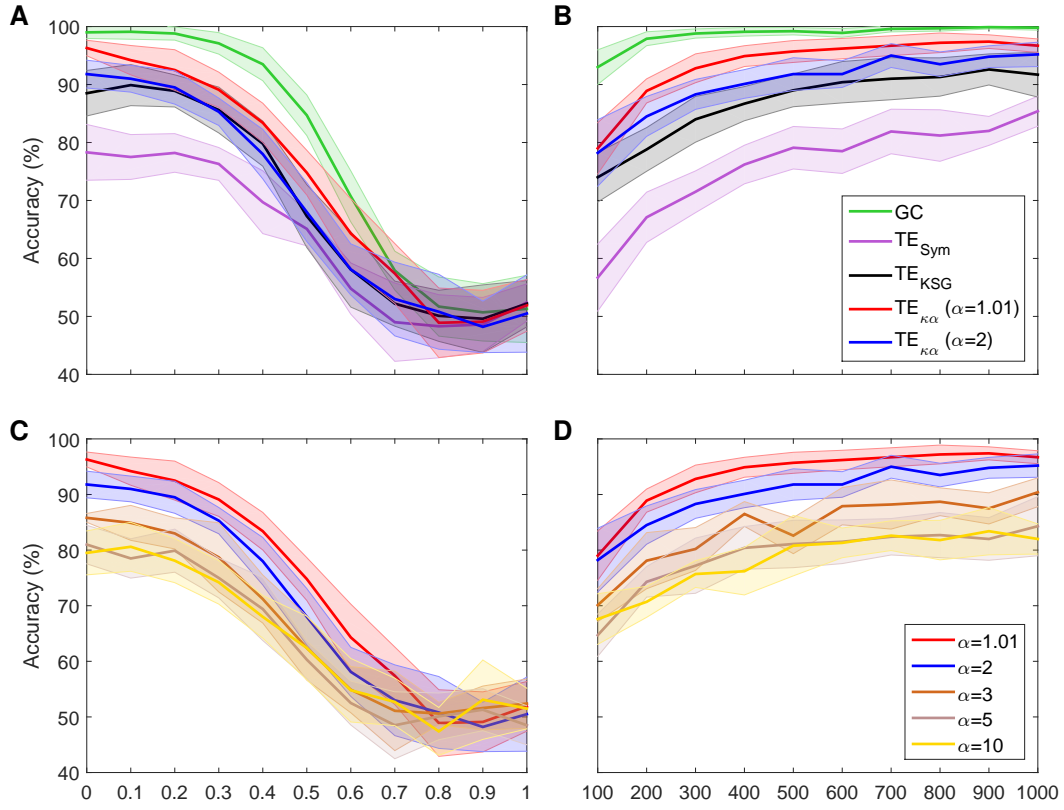


Figure 2.2: Accuracies in the detection of the preferred direction of information flow for synthetic data generated from a unidirectional bivariate autoregressive model of order 3: (A) for varying symmetric noise levels (as a function of the parameter γ), (B) for a varying number of data points. Figures (C) and (D) are analogous to (A) and (B), respectively, but show the accuracies for the $\text{TE}_{\kappa\alpha}$ method for different values of the parameter α .

This behavior is explained by the fact that a larger number of data points allows for a better estimation of the entropies (or their associated probability distributions) needed to compute TE, and for a better adjustment of the AR models in GC.

Figure 2.3 presents the average accuracies obtained with the effective connectivity measures studied under asymmetric noise conditions, in which the noise level varied independently for the driving and driven time series. This case is particularly interesting because asymmetries in the data, like different signal-to-noise ratios, different overall power or spectral details, and other asymmetries that can arise from volume conduction, have the potential to affect causality estimates (Haufe et al., 2013). Notice that for $\text{TE}_{\kappa\alpha}$ we only explore two values of α , $\alpha = \{1.01, 2\}$, since they yielded the highest accuracies under symmetric noise conditions. In general, as the noise in any of the two time

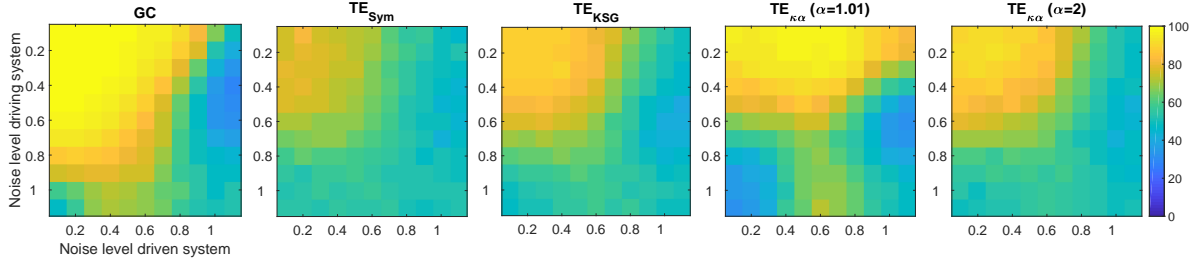


Figure 2.3: Average accuracies in the detection of the preferred direction of information flow for synthetic data, generated from a unidirectional bivariate autoregressive model of order 3 under asymmetric noise level conditions. The vertical axis displays the noise level for the driving time series, while the horizontal axis does so for the driven time series.

series increases, the accuracy in the detection of the preferred direction of information flow decreases. However, some of the methods tested produced spurious results when the noise levels differed, consistently estimating an incorrect interaction direction. This issue is not present in the results displayed in Figure 2.2A for symmetric noise. Particularly, GC failed when the noise level was moderate for the driving time series and high for driven time series. Under those conditions GC estimates had an accuracy of around 30%, which means that for 70% of the simulated time series in that scenario GC estimated an incorrect direction of interaction. $TE_{\kappa\alpha}$ for $\alpha = 1.01$ also failed under the noise asymmetry conditions described above. Additionally, it failed when the noise level was high in the driving time series and low in the driven time series. On the flip side, it was more robust when the noise levels were reversed, that is, a low noise level in the driving time series and a high noise level in the driven time series. Our TE estimation method for $\alpha = 2$ and the other approaches for TE estimation tested were not as affected by asymmetric noise.

For both VAR model experiments, GC outperforms TE, regardless of the TE estimation method. This result is not surprising, since the simulated data were generated using an AR model, and such models are at the core of the definition of GC. Furthermore, since the interactions present in the simulated data are purely linear, a linear method, such as GC, is better suited to capture them than TE (Vicente et al., 2011). However, despite being outperformed by GC, within the proposed simulation framework, TE does reveal the direction of interaction of the data with high accuracy, albeit with marked estimation method dependent differences. Specifically, $TE_{\kappa\alpha}$ exhibits the best performance of the TE estimation methods under study (for $\alpha = \{1.01, 2\}$, it degrades for higher values of α). In particular, for $\alpha = 1.01$, it almost matches GC for the ideal

conditions tested (a noiseless scenario, and a large number of data points). Interestingly, GC and $TE_{\kappa\alpha}$, for $\alpha = 1.01$, were the two methods most affected by asymmetric noise. Overall, within the tested simulation framework, our method fulfills two of the necessary conditions for a TE estimator apt for neuroscience applications (Vicente et al., 2011). Namely, it is robust to moderate levels of noise, represented in this case by a superposition of the signals of interest with those coming from unknown sources. This factor is at play in most noninvasive electrophysiological measurements such as EEG, which, to a large extent, contain unknown superpositions of many sources (Dimitriadis et al., 2016a). Also, our estimator requires a smaller number of data samples to successfully determine the direction of interaction between a pair of signals, as compared with other TE estimators. The former is relevant because neuronal dynamics usually unfolds in periods of a few hundred milliseconds, which restricts the number of samples available to uncover any interaction of interest (Vicente et al., 2011). Additionally, the use of windowing to offset the effects of the non-stationarity of EEG signals further limits the number of data samples available to estimate TE (Cekic et al., 2018).

2.3.2 Modified linear Kus model

Figure 2.4A shows a graphical representation of the 5 channel network constituting the modified linear Kus model. The solid and dashed lines represent the direct and indirect couplings present in the network, respectively; while the arrowheads indicate the direction of the causal relations introduced in the network by the time shifts δ . Figure 2.4B translates the network in Figure 2.4A to a binary class matrix representation. The positive class groups the direct and indirect connections among the network's channels. It is represented by the yellow elements, and their position, in the 5×5 connectivity matrix. On the other hand, the negative class is depicted in blue and represents non-existing interactions in the network. For instance, channel 1 drives channel 2; therefore element $(1, 2)$ belongs to the positive class; but since the opposite is not true, element $(2, 1)$ belongs to the negative class. Notice that all connections to and from channel 5 belong to the negative class. That is because channel 5 consists only of white Gaussian noise and is not coupled to the rest of the network.

In this work, the Kus model experiment is intended to evaluate if our method can detect causal interactions among multiple signals. Unlike the VAR-model experiment, in which we were solely interested in determining the correct direction of the model's causal interactions, this experiment also requires determining whether such interactions exist

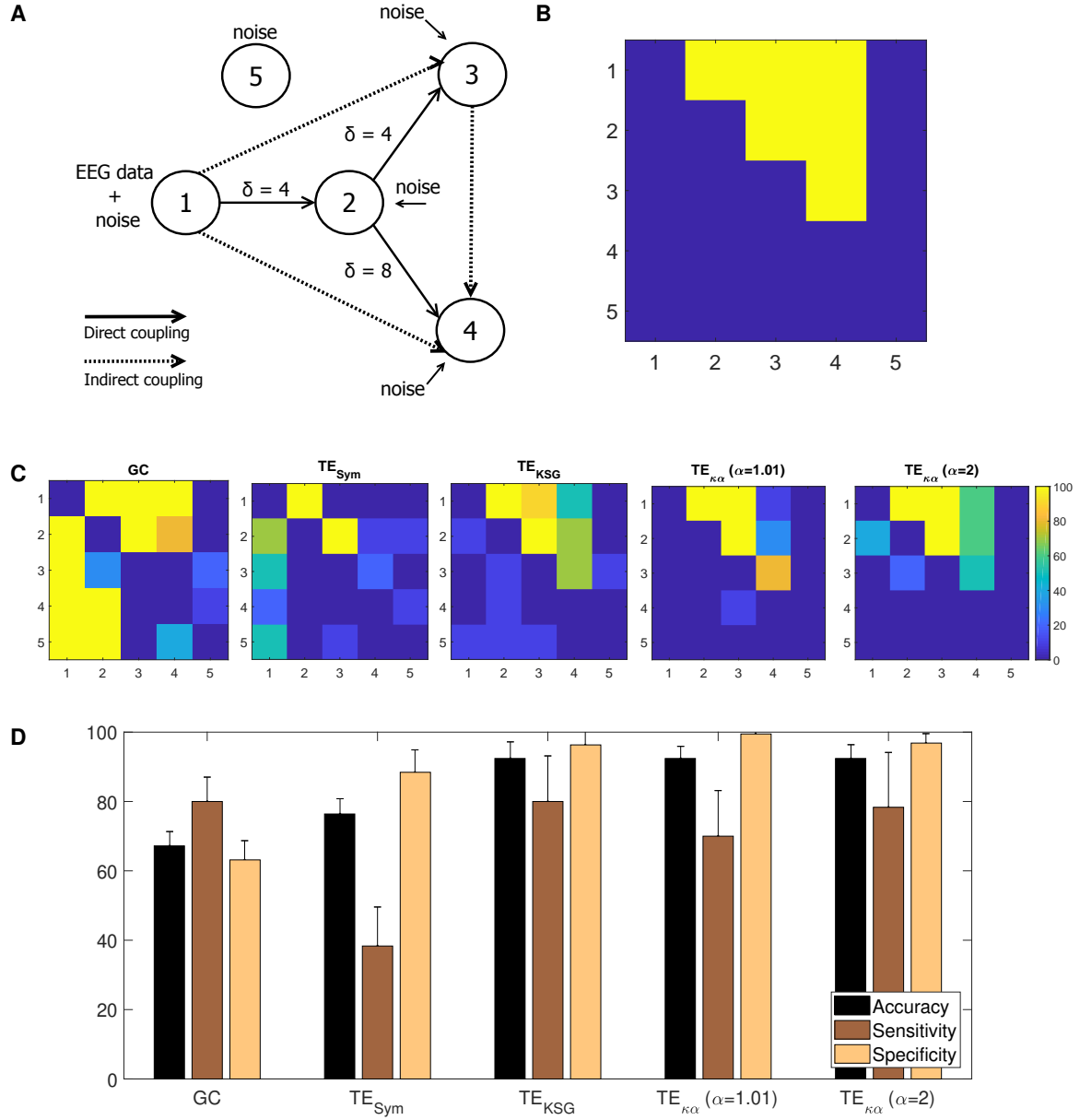


Figure 2.4: (A) Modified Kus model network coupling scheme. (B) Binary class matrix representation of the direct and indirect couplings in the modified Kus model network. (C) From left to right, statistically significant couplings according to a permutation test using trial randomized surrogates for GC, TE_{Sym} , TE_{KSG} , $TE_{\kappa\alpha} (\alpha = 1.01)$, and $TE_{\kappa\alpha} (\alpha = 2)$. (D) Accuracy, sensitivity and specificity values obtained after comparing the statistically significant couplings shown in **C** with the matrix representation of the Kus model network presented in **A**.

at all for any pair of signals within the model. To that end, we performed a permutation test, based on randomized surrogate trials, over the connectivity estimations, obtained

with the methods studied, for each combination of channels (Lindner et al., 2011; Weber et al., 2017).

Figure 2.4C shows, from left to right, the percentage of statistically significant couplings in the 10 realizations of the experiment, according to the permutation test, for GC, TE_{Sym} , TE_{KSG} , $TE_{\kappa\alpha}$ ($\alpha = 1$), and $TE_{\kappa\alpha}$ ($\alpha = 2$). A visual inspection of Figure 2.4C reveals that the proposed $TE_{\kappa\alpha}$ method and the TE_{KSG} method display the best performances. Namely, on average, for the 10 realizations of the experiment, the connectivity values estimated through those methods allow to better determine the actual connections present in the Kus model network. Therefore, their maps of statistically significant couplings more closely resemble the actual Kus model connectivity matrix (Figure 2.4B). Note that the TE estimators tested correctly detect both the presence and direction of the direct connections in the network for every realization, given that the time shift δ of the connection in question matches the chosen interaction delay u . That is, the interactions from channel 1 to channel 2, and from channel 2 to channel 3, for which $\delta = 4$, are successfully revealed. However, the direct connection between channels 2 and 4, for which $\delta = 8$, proves more elusive. The TE_{KSG} method obtains statistically significant results for that specific coupling in 70% of the 10 realizations, while the $TE_{\kappa\alpha}$ method does so for 60% of them. Interestingly, our method always detects the indirect connection from channel 1 to 3, despite an accumulated time shift of 8 time units. In addition, the proposed $TE_{\kappa\alpha}$ method ($\alpha = 1.01$) detects the indirect connection between channels 3 and 4 in more than 80% of the realizations. For the remaining connections, performance degrades for all the TE methods, probably as a result of both larger accumulated time shifts and the increasing amount of noise present in the network. It is also worth noting that our method does not point to the presence of directed interactions involving channel 5 for any realization.

Finally, by comparing the statistically significant couplings per realization with the binary class matrix representation of the Kus model network, we obtained accuracy, sensitivity, and specificity values for each of the effective connectivity estimation approaches tested. Figure 2.4D presents these results. The highest accuracies are achieved by the $TE_{\kappa\alpha}$ and TE_{KSG} methods. Therefore, the proposed $TE_{\kappa\alpha}$ method matches the performance of the TE_{KSG} algorithm regarding the detection of unknown causal interactions within a network from multi-channel data. Furthermore, the shown specificity values reflect the small number of false positives obtained with said methods. Along with the results observed in Figure 2.4C, this indicates that our approach seems to be well suited to detect the couplings among the signals of a connected network with several interac-

tion delays, while at the same time successfully identifying the pairs of non-interacting signals.

2.3.3 EEG data

Motor imagery

The MI tasks performed during the acquisition of the BCI IV database have a clear temporal structure, as depicted in Figure 1.9A. It follows that any characterization of the ensuing brain activity must reflect this structure. That is, since the visual cue indicating the MI task to be executed during a particular trial is presented to the subject at second 2, any information extracted from the EEG signals before that moment should not exhibit any discriminative power between tasks. Furthermore, since the subjects performed the MI task from seconds 3 to 6, this is the time period when the features extracted from the EEG signals of different tasks are expected to diverge. Since we aimed to test the ability of the proposed TE estimation method to elucidate the directed interactions among EEG signals during the MI tasks, and to determine whether those directed interactions allow discriminating between tasks, we can establish the compliance with the above-described temporal constraints as a necessary condition to achieve those aims.

Figures 2.5(A) and 2.5(C) depict 10% of the directed connections estimated with the $TE_{\kappa\alpha}$ method ($\alpha = 2$), discriminated by time window, that present statistically significant differences between the left and right hand MI tasks for subjects 8 and 9, respectively. Such differences were assessed for each connection by applying a two-sample Kolmogorov-Smirnov hypothesis test to the connectivity data for the training dataset, after separating them in function of their associated class labels, and imposing a significance level of 0.01. We found few or no connections with statistically significant differences between conditions for time windows 1 and 2, which span from seconds 0 to 2, and 1 to 3, respectively. Then, for windows 3, 4 and 5 numerous connections to and from the centro-parietal area exhibit statistically significant task-dependent differences. Finally, the number of such connections decreases sharply for window 6, which covers seconds 5 to 7, and includes the break period after the MI task. Therefore, our method reveals directed interactions between EEG signals that present statistically significant differences between the right and left hand MI tasks, according to the temporal evolution of the MI protocol. Since the proposed classification system exploits the differences in the directed connections of each MI task to discriminate between them, its performance

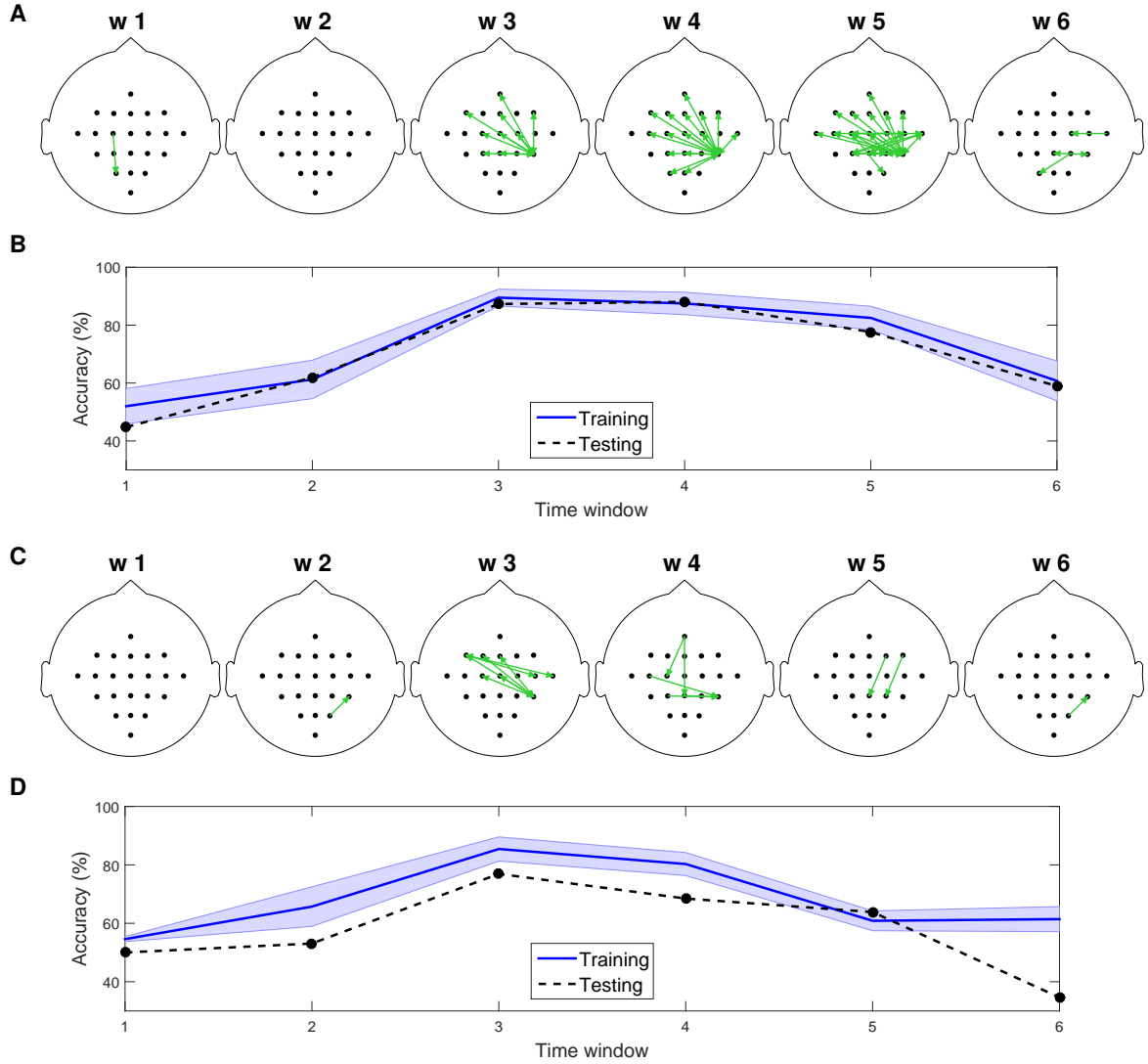


Figure 2.5: (A) Connections with statistically significant differences between the MI tasks for time windows 1 to 6 for subject 8. (B) Training and testing classification accuracies per time window for subject 8. (C) Connections with statistically significant differences between the MI tasks for time windows 1 to 6 for subject 9. (D) Training and testing classification accuracies per time window for subject 9. For visualization purposes, only 10% of the statistically significant connections, those with the smallest p-values, are depicted in (A) and (C).

should also be conditioned by the same temporal constraints. Figures 2.5(B) and 2.5(D) display the training and testing classification accuracies, per time window for subjects 8 and 9, respectively. As expected, the classification system achieved its highest performances for the time windows during which the MI task was being executed by the subjects. Here we must point out the fact that the results in Figure 2.5 differ slightly

from the ones we published in [De La Pava Panche et al. \(2019\)](#). The same holds true for all the MI classification results in the present chapter. The reason is that in this thesis document we opted to use the same feature selection and classification strategies for all EEG databases and connectivity measures, in order to facilitate comparisons and help the document’s readability. The feature selection strategy we employed in [De La Pava Panche et al. \(2019\)](#) is not suitable for multiclass data, and such is the case for the WM database. Therefore, we proceeded as described in Section 2.2.3 instead.

Table 2.1: Average training accuracy [%] for the window (**w**) with the best performance.

Subject	$\mathbf{TE}_{\kappa\alpha} (\alpha = 2)$		$\mathbf{TE}_{\kappa\alpha} (\alpha = 1.01)$		\mathbf{TE}_{KSG}		\mathbf{TE}_{Sym}		\mathbf{GC}	
	acc	(w)	acc	(w)	acc	(w)	acc	(w)	acc	(w)
s 01	70.7 ± 5.2	(3)	77.1 ± 5.8	(3)	58.8 ± 8.0	(3)	65.2 ± 5.9	(3)	78.3 ± 5.4	(3)
s 02	57.1 ± 6.6	(6)	62.0 ± 7.7	(6)	60.5 ± 5.2	(4)	57.8 ± 4.8	(3)	57.1 ± 4.0	(3)
s 03	76.7 ± 5.5	(3)	75.0 ± 5.2	(4)	78.3 ± 6.8	(4)	86.2 ± 5.3	(4)	79.3 ± 5.6	(4)
s 04	60.8 ± 3.3	(2)	60.8 ± 6.0	(3)	60.8 ± 6.0	(4)	59.5 ± 4.6	(1)	55.6 ± 4.3	(3)
s 05	69.0 ± 6.0	(3)	66.2 ± 7.4	(3)	67.2 ± 7.3	(3)	64.6 ± 4.4	(4)	66.4 ± 10.2	(3)
s 06	60.0 ± 5.9	(6)	62.4 ± 5.1	(4)	59.7 ± 9.4	(4)	58.3 ± 6.6	(2)	60.6 ± 4.0	(6)
s 07	75.5 ± 8.1	(3)	74.0 ± 5.6	(3)	63.3 ± 7.6	(3)	62.3 ± 5.9	(6)	72.5 ± 5.2	(3)
s 08	89.5 ± 2.9	(3)	77.3 ± 5.8	(4)	72.3 ± 5.6	(4)	76.5 ± 5.7	(4)	90.0 ± 3.5	(4)
s 09	85.4 ± 4.1	(3)	71.7 ± 2.7	(3)	68.0 ± 7.2	(4)	80.6 ± 6.7	(3)	78.6 ± 6.0	(3)
AVG	71.6 ± 5.3		69.6 ± 5.7		65.4 ± 7.0		67.9 ± 5.5		70.9 ± 5.4	

Table 2.2: Testing accuracy [%] for the window (**w**) with the best performance.

Subject	$\mathbf{TE}_{\kappa\alpha} (\alpha = 2)$		$\mathbf{TE}_{\kappa\alpha} (\alpha = 1.01)$		\mathbf{TE}_{KSG}		\mathbf{TE}_{Sym}		\mathbf{GC}	
	acc	(w)	acc	(w)	acc	(w)	acc	(w)	acc	(w)
s 01	67.4	(4)	66.0	(4)	57.4	(4)	66.0	(4)	67.4	(3)
s 02	57.0	(6)	57.7	(6)	55.6	(2)	52.8	(3)	57.7	(6)
s 03	82.5	(4)	73.7	(4)	67.2	(4)	82.5	(4)	71.5	(5)
s 04	58.6	(4)	59.5	(3)	56.0	(5)	57.8	(5)	56.9	(5)
s 05	51.9	(1)	53.3	(4)	55.6	(6)	51.9	(3)	48.9	(3)
s 06	59.3	(1)	65.7	(4)	63.0	(1)	61.1	(3)	58.3	(4)
s 07	65.0	(2)	67.1	(2)	61.4	(3)	62.1	(4)	50.7	(1)
s 08	88.1	(4)	68.7	(4)	66.4	(4)	83.6	(4)	82.1	(4)
s 09	76.9	(3)	74.6	(3)	58.5	(3)	70.8	(3)	51.5	(3)
AVG	67.4 ± 11.8		65.2 ± 6.7		60.1 ± 4.3		65.4 ± 11.0		60.6 ± 10.4	

Tables 2.1 and 2.2 present the highest accuracies achieved by the proposed classification system, for all subjects, and each of the effective connectivity methods studied. During the training-validation stage, the classifiers based on GC features and features extracted with $\mathbf{TE}_{\kappa\alpha}$, for $\alpha = 2$, exhibited the highest average performances. However, during the testing stage the average performance of the GC-based classifier drops more

than that of the $TE_{\kappa\alpha}$ -based classifier, which implies that the latter generalizes better to new data. This points to a more stable identification of discriminant directed interactions across trials by our method as compared to other effective connectivity estimation approaches. Also, note that, in general, the classifiers attain their best performances for the time windows corresponding to the execution of the MI task. Here, we must highlight the fact that the accuracies presented in Tables 2.1 and 2.2 fall short of those obtained with feature extraction strategies other than effective connectivity analyses, such as common spatial patterns and its variations (Li et al., 2018; Zhang et al., 2018), as shown in Table 2.3:

Table 2.3: Average training accuracy for CSP and CSP-related methods.

Subject	CSP (Zhang et al., 2018)	CSP + LSTSVM (Li et al., 2018)	TSGSP (Zhang et al., 2018)
s 01	76.9	88.2	87.0
s 02	57.1	64.6	64.7
s 03	91.6	94.4	93.8
s 04	65.5	66.0	74.3
s 05	54.8	76.4	90.4
s 06	58.3	67.4	63.9
s 07	73.4	75.0	91.4
s 08	92.8	88.2	95.8
s 09	67.8	88.9	81.3
AVG	70.9 ± 14.1	78.8 ± 11.4	82.5 ± 12.2

This underperformance of connectivity-based analysis for MI tasks discrimination has been linked to the difficulties of measuring local or short-range connectivities, such as those expected to appear among different zones of the motor areas during MI tasks, due to volume conduction effects (Rathee et al., 2017). Interestingly, the results obtained with the classifiers based on features extracted with our method, and with the other effective connectivity measures studied, tend to coincide with those of classifiers based on alternative characterization strategies, in terms of the ranking of the performances per subject; that is, subjects 8, 9 or 3 present the highest performances, while subjects 2, 4, 5 or 6 exhibit the lowest ones (Elasuty and Eldawlatly, 2015; Gómez et al., 2018; Li et al., 2018; Liang et al., 2016; Zhang et al., 2018).

In order to gain insight into the large differences in classification performance observed for the different subjects, we computed the average differences in the total information flow coming into each channel, estimated through the proposed $TE_{\kappa\alpha}$ method ($\alpha = 2$), for all subjects and time windows. Namely, for each trial, we obtained the total information flow coming into a particular channel as the sum of all directed interactions

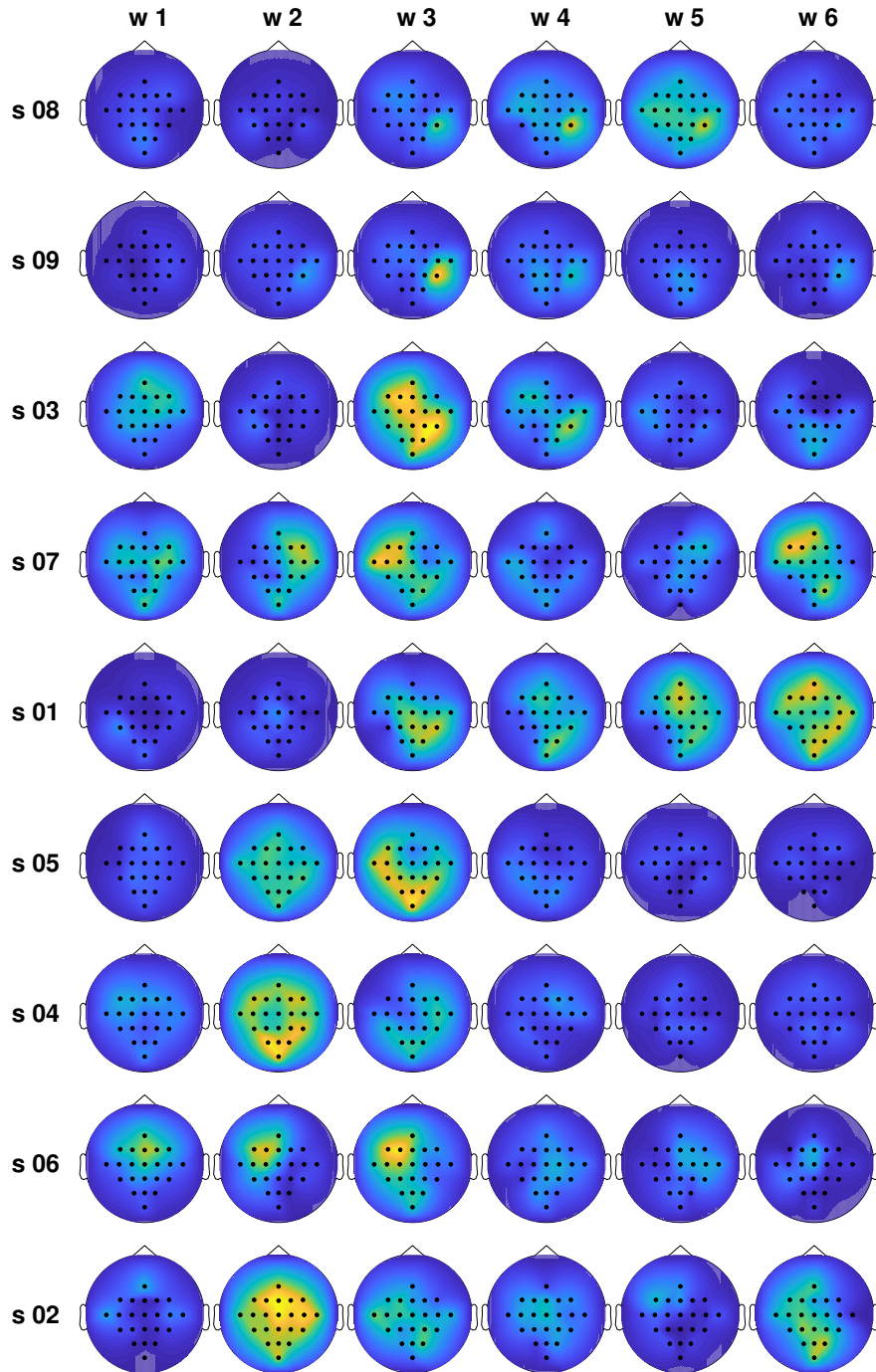


Figure 2.6: Normalized average differences in the total information flow coming into each channel for the training set, for all subjects and time windows. Large differences are coded in yellow, while small differences are presented in blue.

targeting that channel, then averaged that magnitude across all trials of the same MI task, and finally subtracted the averages of the left and right MI tasks. Figure 2.6 shows

the obtained results. The subjects are organized in descending order according to the classification accuracies presented in Table 2.1. For the subjects at the top of the plot, we observed a clear temporal evolution, with small variations between the information flow of both tasks for time windows 1 and 2, and large localized differences during the time windows corresponding to MI execution. Furthermore, we observe a trend regarding the spatial location of the information flow differences. For the top 3 subjects, particularly for time windows 3 and 4, they are centered around the centro-parietal region, specifically channel CP4. For the subjects at the bottom of the plot, the same temporal and spatial patterns are not present. We can also gain this type of information from the relevance vectors \mathbf{q} obtained by applying CKA to the training data during the feature selection stage of our classification system. Figure 2.7 shows the average nodal relevance, i.e. the sum of the relevance values corresponding to all directed interactions targeting and originating from a particular channel, and the most relevant connectivities for two groups of subjects, for the time segment ranging from 3 s to 5 s ($TE_{\kappa\alpha}$, $\alpha = 2$). The first group (G.I) comprises the subjects with low classification performances (subjects 2, 6, 4, and 5), while the second group (G.II) gathers the subjects that displayed higher classification accuracies (subjects 8, 9, 3, 7 and 1). Again, for the subjects belonging to G.II the most relevant connectivities, those that allow to discriminate between right and left hand motor imagination, arise and target nodes located around the centro-parietal region. The same isn't true for the subjects in G.I.

It is worth noting that we have focused our analyses on the differences in the obtained effective connectivities for the left and right MI tasks, instead of analyzing the connectivities that arise for each task as compared with the resting state (Gong et al., 2018). Bearing this in mind, and considering the physiological interpretation of MI which states that motor imagination mainly activates motor representations in the premotor cortex and the parietal area (Héту et al., 2013), we can argue that it is the differences in the information flow to and from the right parietal cortex, during the activation associated with MI, which allowed us to discriminate between tasks for a subgroup of subjects.

Working memory

The change detection task performed by the subjects during the acquisition of the WM database introduced in Villena-González et al. (2020), and detailed in Section 1.7.2, aims to study how the characteristics of brain activity are modulated by changes in memory loads, as induced by WM tasks with different difficulty levels (Dai et al., 2017). This topic is closely related to the concept of WM capacity, the amount of information that

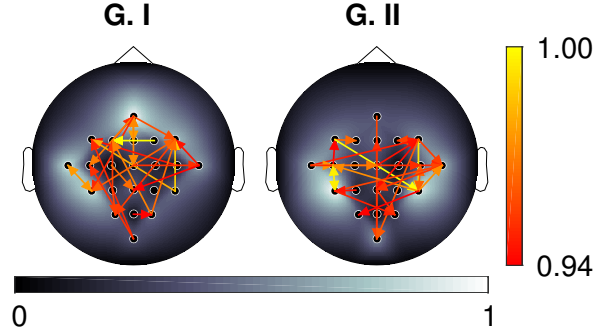


Figure 2.7: Topoplots of the average node (channel) relevance for two groups of subjects in the MI database (G.I: subjects 2, 6, 4, 5; G.II: subjects 8, 9, 3, 7, 1). The arrows represent the most relevant connectivities for each group. For visualization purposes, only 6% of the connections, those with the highest average relevance values per group, are depicted.

can be maintained and manipulated in WM (Zhang et al., 2016). WM capacity is in turn linked to important abilities, including non-verbal reasoning, control of attention, among others; and it can be altered in people with psychiatric disorders (Constantinidis and Klingberg, 2016). Our goal was to evaluate whether the directed connectivity values estimated through the proposed $TE_{\kappa\alpha}$ method captured information useful to discriminate among the low, medium, and high memory loads corresponding to the different the levels of the change detection task.

Figure 2.8 shows the highest classification accuracy obtained for each subject in the WM database according to the different connectivity measures studied. Table 2.4 summarizes this results in terms of the average accuracies obtained for all subjects. The TE_{KSG} and $TE_{\kappa\alpha}$ -based classification systems exhibited the highest performances, with accuracies well above what would be expected from chance in a three class classification task. In the latter case, setting the α parameter to 2 yielded better results than setting it to 1.01. Notice that although the obtained accuracies for the subjects differ, when the same connectivity measure was used to characterize the data such differences are not as stark as the ones observed for the MI data.

Table 2.4: Average classification accuracies for the WM database for all the effective connectivity measures considered.

	$TE_{\kappa\alpha} (\alpha = 2)$	$TE_{\kappa\alpha} (\alpha = 1.01)$	TE_{KSG}	TE_{Sym}	GC
AVG acc	68.7 ± 4.8	63.5 ± 5.8	80.2 ± 6.0	49.9 ± 4.5	53.0 ± 7.4

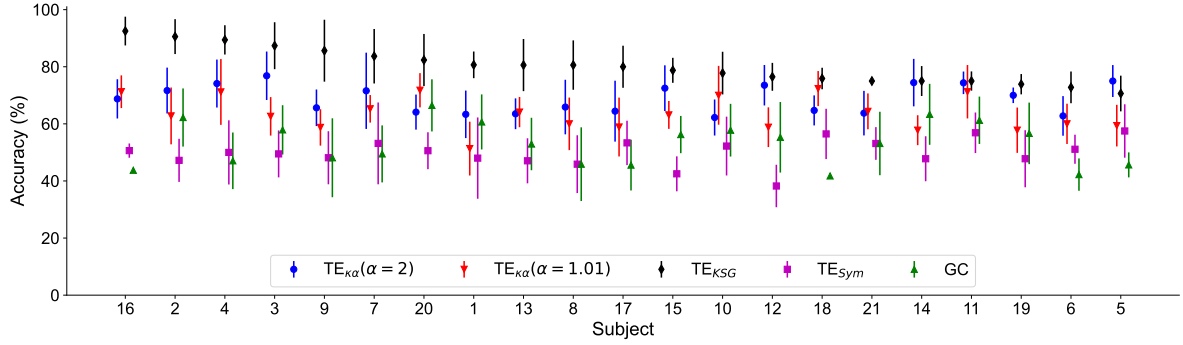


Figure 2.8: Highest average classification accuracy for each subject in the WM database (mean \pm standard deviation). The subjects are ordered from highest to lowest accuracy for the TE_{KSG} -based classifier, which displayed the best overall performance.

At this point, we must highlight the fact that the auxiliary cross-validation step introduced for feature selection in the case of the WM data, aiming to obtain stable CKA results for the reduced number of available trials, leads to data leakage. This is because, ultimately, it requires all the available data to estimate $\bar{\rho}$, which can inflate performance evaluations, such as the accuracy results previously described. However, since the same strategy was implemented for all classification systems and connectivity measures, comparisons among them remain valid, and the relative differences in performance are still informative.

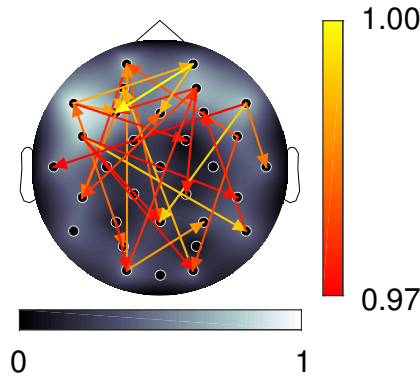


Figure 2.9: Topoplot of the average nodal relevance for all the subjects in the WM database. The arrows represent the most relevant connectivities. For visualization purposes only the most relevant connections are depicted (top 3%).

Figure 2.9 displays the most relevant connectivities (average for all subjects), according to the relevance vector $\bar{\rho}$ for the $TE_{\kappa\alpha}$ method ($\alpha = 2$). The background topoplot shows the average nodal relevance, that is to say, the relevance of the total information

flow of every channel. In general, the highest nodal relevance is displayed for interactions involving the frontal and pre-frontal regions, followed by temporal and parieto-occipital areas. Regarding the connections with the highest class discrimination capability, we observe long-range interactions involving mostly the regions previously listed, many of them linking frontal and parieto-occipital nodes. There are also relevant, shorter range connections among frontal and pre-frontal channels. These results coincide with those of studies that identified the presence of fronto-parietal and fronto-temporal interactions during cognitive task that activated visuospatial WM (Dimitriadis et al., 2016a; Johnson et al., 2018). However, as we explained in Section 1.7.2, the current understanding of the underlying mechanisms of WM entails interactions among neural oscillations originating from different brain regions. As will become evident in the next chapters, the WM data is better characterized by effective connectivity estimation approaches that capture frequency-specific information.

The above results, and those of Sections 2.3.1 and 2.3.2, show that the proposed $TE_{\kappa\alpha}$ method is apt for TE estimation from neuroscience data. Regarding the requirements outlined in the chapter’s introduction, we have shown that our TE estimator is robust to moderate levels of noise and performs satisfactorily under data size constraints. The third requirement, concerning the reliability of the estimator when dealing with high-dimensional spaces, is readily taken care of by the intrinsic capacity of kernels to deal with such spaces (Schölkopf et al., 2002). Nonetheless, our approach also has shortcomings, which we will discuss in the following.

2.3.4 Limitations

First, we must note that the exponentiation operation in 1.14, central to the kernel-based approximation of Renyi’s entropy, makes our TE estimator ill-suited for the analysis of long time series (i.e. time series with several thousands of data points) due to the increase in computational cost. This is especially true for non-integer values of α . Furthermore, our approach also exhibits limitations inherent to the concept of TE (Vicente et al., 2011). Namely, the definition of causality underlying TE is observational, so unobserved common causes cannot be analyzed. This shortcoming encompasses the different delay driving problem. Given three variables, this problem occurs when the first variable drives the two remaining variables but each with a different delay, giving rise to an indirect casualty relation between the second and the third variables that cannot be identified as spurious in bivariate connectivity analyses (Cekic et al., 2018). This is the case, for instance, of some of the interactions present in the modified linear Kus model

(see Sections 2.2.2 and 2.3.2). Additionally, the fact that TE is model-free implies that while TE provides information about the directed or causal interactions among data, it does not give any further insight into the nature of those interactions. Furthermore, TE assumes at most weak non-stationarities in the data, so strong non-stationarities pose a challenge for its estimation; although progress has been made in that regard (Wollstadt et al., 2014). Finally, by using Renyi's entropy measures of order α to define TE, instead of Shanon's entropy, we gain flexibility regarding the characteristics of the data we wish to highlight, by having at our disposal an entire parametric family of entropies. As observed in our results, the choice of the parameter α indeed influences the performance of the $TE_{\kappa\alpha}$ estimator. It becomes more or less successful at uncovering the interactions of interest as a function of α . The flip side of this flexibility is that in practice α becomes one more parameter to select. In general, the choice of α should be associated with the task goal (Principe, 2010). For Renyi's entropy, a large α emphasizes slowly changing features (Giraldo et al., 2015). Particularly, $\alpha > 2$ characterizes mean behaviour, while $\alpha < 2$ emphasizes rare events or multiple modalities, and $\alpha = 2$ is neutral to weighting.

2.4 Summary

In this chapter, we proposed a new TE estimator based on Renyi's entropy of order α , which we approximate through positive definite kernel matrices. Our data-driven method, termed $TE_{\kappa\alpha}$, sidesteps the probability distribution estimation stage involved in the computation of TE from discrete data, thus avoiding the challenges associated with it. We tested the performance of our method on two different synthetic datasets, and on two EEG-databases obtained under MI and WM paradigms. We compared it with that of state-of-the-art methods for TE estimation, as well as with that of GC, another commonly used brain effective connectivity measure. Our results show that the proposed TE estimator successfully detects the presence and direction of Wiener-causal interactions between a pair of signals, exhibiting robustness to varying noise levels and number of available data samples, and to the presence of multiple interaction delays within a connected network. Furthermore, our method revealed discriminant spatiotemporal patterns for the MI data, that are consistent across the top performing subjects, and which follow the temporal constraints imposed by the MI experimental paradigm. Regarding the results obtained for the WM data, our proposal coded discriminant directed connectivities linking the regions that the literature usually associates with visuospatial WM. For all the performance evaluation metrics employed, the proposed kernel-based

TE estimation method was competitive with the state-of-the-art.

Chapter 3

Kernel-based Renyi's phase transfer entropy

In this chapter, we propose a novel methodology to estimate TE between single pairs of instantaneous phase time series. Our approach combines the kernel-based TE estimator introduced in Chapter 2, with phase time series obtained by convolving neural signals with a Morlet Wavelet. We hypothesize that our estimator could overcome the hurdles other single-trial TE estimators face when obtaining TE values from this type of data, since it would not have to explicitly obtain probability distributions from circular variables (Lobier et al., 2014).

In order to test our proposal we employ a simulation model, and the EEG databases described in Section 1.7. The simulated data are obtained from neural mass models (NMM), mathematical models of neural mechanisms that generate time series with oscillatory behavior similar to that of electrophysiological signals (David and Friston, 2003). Our results for such data show that the proposed kernel-based phase TE estimation approach successfully detects the direction of interaction imposed by the model. Indeed, it detects statistically significant connections in the frequency bands of interest, even for weak couplings and narrowband bidirectional interactions. It also displays robustness to realistic levels of noise and signal mixing. Regarding the EEG data, attained classification results demonstrate that our approach is competitive compared to real-valued and phase-based directed connectivity measures. Thus, this proposal extends the approach described in Chapter 2 by introducing a measure that captures directed interactions between the phases of oscillations at specific frequencies. Unlike alternative approaches in the literature, it can be obtained from single trial data, which allows it to be potentially used as a characterization strategy in BCI applications. In addition, the results obtained

for the EEG data show that our approach, coupled with a CKA-based relevance analysis (see Appendix D), largely outperforms the real-valued kernel-based transfer entropy ($TE_{\kappa\alpha}$) as characterization strategy for cognitive tasks such as working memory.

3.1 Kernel-based Renyi's phase transfer entropy

As described in Section 1.4.1, in phase TE the time series \mathbf{x} and \mathbf{y} are replaced by instantaneous phase time series $\boldsymbol{\theta}^x(f) \in [-\pi, \pi]_{t=1}^T$ and $\boldsymbol{\theta}^y(f) \in [-\pi, \pi]_{t=1}^T$, obtained by complex-filtering \mathbf{x} and \mathbf{y} at frequency f (Lobier et al., 2014). Thus, phase TE is defined as

$$TE^\theta(\mathbf{x} \rightarrow \mathbf{y}, f) = H_S(\boldsymbol{\theta}_{t-1}^{y,dy}, \boldsymbol{\theta}_{t-u}^{x,dx}) - H_S(\boldsymbol{\theta}_t^y, \boldsymbol{\theta}_{t-1}^{y,dy}, \boldsymbol{\theta}_{t-u}^{x,dx}) + H_S(\boldsymbol{\theta}_t^y, \boldsymbol{\theta}_{t-1}^{y,dy}) - H_S(\boldsymbol{\theta}_{t-1}^{y,dy}),$$

where $\boldsymbol{\theta}_t^{x,dx}, \boldsymbol{\theta}_t^{y,dy} \in [-\pi, \pi]^{D \times d}$ are time embedded versions of $\boldsymbol{\theta}^x$ and $\boldsymbol{\theta}^y$, and the explicit dependency of the phase time series on f has been dropped.

Here, we hypothesize that the $TE_{\kappa\alpha}$ estimator, having previously displayed robustness to common issues that affect connectivity analyses (De La Pava Panche et al., 2019), could overcome many of the problems associated with single-trial phase TE estimation. Hence, we propose a kernel-based Renyi's phase TE estimator defined as:

$$\begin{aligned} TE_{\kappa\alpha}^\theta(\mathbf{x} \rightarrow \mathbf{y}, f) = & H_\alpha(\mathbf{K}_{\boldsymbol{\theta}_{t-1}^{y,dy}}, \mathbf{K}_{\boldsymbol{\theta}_{t-u}^{x,dx}}) - H_\alpha(\mathbf{K}_{\boldsymbol{\theta}_t^y}, \mathbf{K}_{\boldsymbol{\theta}_{t-1}^{y,dy}}, \mathbf{K}_{\boldsymbol{\theta}_{t-u}^{x,dx}}) \\ & + H_\alpha(\mathbf{K}_{\boldsymbol{\theta}_t^y}, \mathbf{K}_{\boldsymbol{\theta}_{t-1}^{y,dy}}) - H_\alpha(\mathbf{K}_{\boldsymbol{\theta}_{t-1}^{y,dy}}), \end{aligned} \quad (3.1)$$

where the kernel matrices $\mathbf{K}_{\boldsymbol{\theta}_t^y}, \mathbf{K}_{\boldsymbol{\theta}_{t-1}^{y,dy}}, \mathbf{K}_{\boldsymbol{\theta}_{t-u}^{x,dx}} \in \mathbb{R}^{(D-u) \times (D-u)}$ hold elements analogous to those of matrices $\mathbf{K}_{y_t}, \mathbf{K}_{y_{t-1}^{dy}},$ and $\mathbf{K}_{x_{t-u}^{dx}}$ in Equation 2.5, while replacing the time series \mathbf{x} and \mathbf{y} for their instantaneous phase time series $\boldsymbol{\theta}^x$ and $\boldsymbol{\theta}^y$ at frequency f , respectively.

3.1.1 Phase-based effective connectivity estimation approaches considered in this chapter

Phase transfer entropy

We obtain phase TE values through three different estimators that allow computing TE from individual signal pairs. First, the proposed kernel-based Renyi's phase TE estimator ($TE_{\kappa\alpha}^\theta$), defined in Equation (3.1). Second, the Kraskov-Stögbauer-Grassberger TE estimator (TE_{KSG}^θ), which relies on a local approximation of the probability distribu-

tions needed to estimate the entropies in TE from the distances of every data point to its neighbors (Kraskov et al., 2004) (see Equation 1.20). And third, symbolic TE (TE_{Sym}^θ) (Dimitriadis et al., 2016a), as defined in Equation 1.19.

In all cases, θ^x and θ^y are obtained by convolving the real-valued time series with a Morlet wavelet, defined as

$$h(t, f) = \exp(-t^2/2\xi_t^2)\exp(i2\pi ft), \quad (3.2)$$

where f stands for the filter frequency, $\xi_t = m/2\pi f$ is the time domain standard deviation of the wavelet, and m defines the compromise between time and frequency resolution (Lobier et al., 2014).

Phase slope index

The phase slope index (PSI) is an effective brain connectivity measure that assesses the direction of coupling between two oscillatory signals of similar frequencies (Jiang et al., 2015). Given two time series $\mathbf{x} = \{x_t\}_{t=1}^T$ and $\mathbf{y} = \{y_t\}_{t=1}^T$, the PSI is defined as the slope of the phase of the cross-spectra between \mathbf{x} and \mathbf{y} :

$$PSI(\mathbf{x} \rightarrow \mathbf{y}) = \Im \left(\sum_{f \in F} C_{\mathbf{xy}}^*(f) C_{\mathbf{xy}}(f + df) \right), \quad (3.3)$$

where $C_{\mathbf{xy}}(f) = S_{\mathbf{xy}} / \sqrt{S_{\mathbf{xx}} S_{\mathbf{yy}}}$ is the complex coherence, $S_{\mathbf{xy}} \in \mathbb{C}$ is the cross-spectrum between \mathbf{x} and \mathbf{y} , $S_{\mathbf{xx}}, S_{\mathbf{yy}} \in \mathbb{C}$ are the auto-spectrums of \mathbf{x} and \mathbf{y} , $df \in \mathbb{R}^+$ is the frequency resolution, F stands for the set of frequencies over which the slope is summed, and \Im indicates selecting only the imaginary part of the sum (Nolte et al., 2008). If the PSI, as defined in Equation (3.3), is positive, then there is directed interaction from \mathbf{x} to \mathbf{y} in F . Conversely, if the PSI is negative, the directed interaction goes from \mathbf{y} to \mathbf{x} . Note that by definition the PSI is an antisymmetric measure: $PSI(\mathbf{x} \rightarrow \mathbf{y}) = -PSI(\mathbf{y} \rightarrow \mathbf{x})$.

Granger causality

We also characterize the simulated and EEG data using Granger causality (GC) (see Section 1.3.2). In analogy to the concept of phase TE, we define $GC^\theta(\mathbf{x} \rightarrow \mathbf{y}, f) = GC(\theta^x \rightarrow \theta^y)$, where θ^x and θ^y are instantaneous phase time series obtained by filtering \mathbf{x} and \mathbf{y} at frequency f , as a measure within the framework of GC that captures phase-based interactions.

3.2 Experiments

3.2.1 Neural mass models

Neural mass models (NMM) are biologically plausible mathematical descriptions of neural mechanisms (David and Friston, 2003). They represent the electrical activity of neural populations at a macroscopic level through a set of stochastic differential equations (David et al., 2004) (see Appendix F for details). NMMs allow generating mildly nonlinear time series with properties that resemble the oscillatory dynamics of electrophysiological signals, such as EEG, and how they change as a result of coupling between different cortical areas. Therefore, NMMs are useful to study the behavior of brain connectivity measures that aim to capture such interactions (Chen et al., 2019; David et al., 2004; Lobier et al., 2014; Ursino et al., 2020). Figure 3.1A shows a schematic representation of an NMM with two interacting cortical areas from which two signals, \mathbf{x} and \mathbf{y} (see Figure 3.1B), can be obtained. The parameters C_{12} and C_{21} are known as coupling coefficients, and they determine the strength of the coupling from Area 1 to Area 2, and from Area 2 to Area 1, respectively. The parameter ν represents the interaction lag between the two areas, while \mathbf{p}_1 and \mathbf{p}_2 are external inputs coming from other cortical regions.

In this work, we use NMMs to generate interacting time series with known oscillatory properties in order to test the performance of the proposed phase TE estimator. In particular, we test our proposal in terms of its ability to detect directed interactions for different levels of coupling strength, under the presence of noise and signal mixing, and for bidirectional narrowband couplings. We proceed as follows: first, we set the model parameters describing Areas 1 and 2 as in David et al. (2004), so as to generate signals with power spectrums peaking in the α (8 Hz–12 Hz) and lower β bands (12–20 Hz), as depicted in Figure 3.1C. Then, in order to generate unidirectionally coupled signals, with interactions from \mathbf{x} to \mathbf{y} , we set the parameter C_{21} to 0 for all simulations. Also, the parameter ν is set to 20 ms, and the extrinsic inputs \mathbf{p}_1 and \mathbf{p}_2 are modeled as Gaussian noise (David et al., 2004). Afterward, we generate 50 pairs (trials) of 3 s long signals, using a simulation time step of 1 ms, equivalent to a sampling frequency of 1000 Hz, for each condition in the three scenarios detailed in Sections 3.2.1–3.2.1. Next, we select a 2 s long segment from the signals, from 0.5 s to 2.5 s, and downsample them to 250 Hz. Then, we compute connectivity estimates for the simulated data in the frequency range between 2 Hz and 60 Hz, in steps of 2 Hz. After that, we obtain net

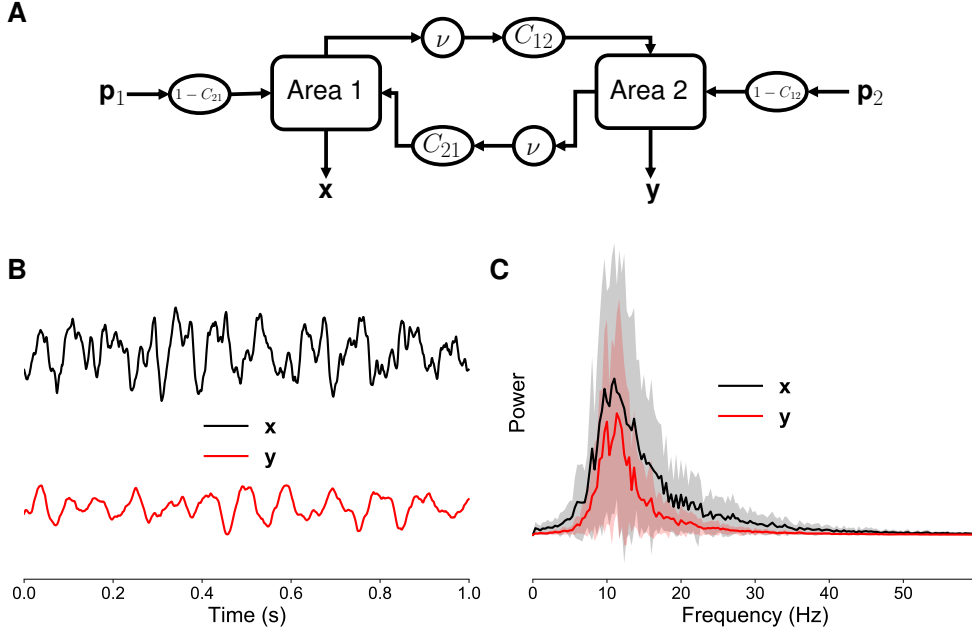


Figure 3.1: (A) Schematic representation of a neural mass model. (B) 1 s long unidirectionally coupled time series generated by the model. (C) Average power spectra peaking in the α and lower β frequency bands.

connectivity values, defined as

$$\Delta\lambda(\mathbf{x}, \mathbf{y}, f) = \lambda(\mathbf{x} \rightarrow \mathbf{y}, f) - \lambda(\mathbf{y} \rightarrow \mathbf{x}, f), \quad (3.4)$$

where λ stands for any of the phase-based effective connectivity measures studied, except for the PSI, in which case all subsequent analyses are performed directly on the PSI values. Lastly, for each condition in the three scenarios and at each frequency evaluated, we perform permutation tests based on randomized surrogate trials (Lindner et al., 2011; Weber et al., 2017) to determine which net couplings or directed connections are statistically significant. The permutation test employed uses the trial structure of the data to generate surrogate datasets for the null hypothesis (absence of directed interactions). It does so by shuffling the data from different trials (see Appendix C). The significance level for the tests was set to 3.3×10^{-4} after applying the Bonferroni correction to an initial alpha level of 0.01 in order to account for 30 independent tests, one for each evaluated frequency per condition.

Coupling strength

In order to test the ability of our phase TE estimation method to detect phase-based directed interactions of varying intensity, we modify the coupling strength between the simulated signals, \mathbf{x} and \mathbf{y} , by varying the parameter C_{12} in the range $\{0, 0.2, 0.5, 0.8\}$, with 0 indicating the absence of coupling and 0.8 a strong interaction between the two signals.

Noise and signal mixing

To assess the robustness of our proposal to realistic levels of noise and signal mixing, we do the following: we generate a noise time series $\boldsymbol{\eta}$, with the same power spectrum of \mathbf{x} , through the methodology proposed in Lobier et al. (2014). Then, we add \mathbf{x} and $\boldsymbol{\eta}$ to generate a noisy version of \mathbf{x} , $\mathbf{x}_\eta = \mathbf{x} + 10^{-\frac{\text{SNR}}{20}} \boldsymbol{\eta}$, where SNR is the signal to noise ratio. Likewise for \mathbf{y} . Then, we mix \mathbf{x}_η and \mathbf{y}_η to simulate one of the effects of volume conduction, by doing $\mathbf{x}_\eta^w = \left(1 - \frac{w}{2}\right) \mathbf{x}_\eta + \left(\frac{w}{2}\right) \mathbf{y}_\eta$, and $\mathbf{y}_\eta^w = \left(1 - \frac{w}{2}\right) \mathbf{y}_\eta + \left(\frac{w}{2}\right) \mathbf{x}_\eta$, with w the mixing strength. We set the parameters SNR and w to 3 and 0.25 respectively, based on the results obtained in Lobier et al. (2014) for realistic values of noise and mixing for EEG signals. The coupling coefficient C_{12} is held constant at a value of 0.5 to simulate couplings of medium strength.

Narrowband bidirectional interactions

In this experiment, we aim to evaluate how our proposal deals with bidirectional interactions of localized frequency content. Particularly, we want to assess its performance for signals \mathbf{x} and \mathbf{y} containing a directed interaction from \mathbf{x} to \mathbf{y} at 10 Hz and an interaction in the opposite direction, from \mathbf{y} to \mathbf{x} , at 40 Hz. To generate such signals, first, we modify the model parameters of Area 2 so that it produces a signal \mathbf{y} with a power spectrum peaking in the γ band (David and Friston, 2003). The power spectrum of \mathbf{x} remains as before. The coupling coefficient C_{12} is again held constant at a value of 0.5. The change in the parameters of Area 2 leads to strong directed interactions from \mathbf{x} to \mathbf{y} around 10 Hz and 40 Hz. Then, we use a Morlet wavelet (Equation 3.2) to filter both \mathbf{x} and \mathbf{y} at those frequencies (10 Hz and 40 Hz). The obtained real-valued narrowband time series are then combined as follows: $\mathbf{x}_* = \mathbf{x}_{10 \text{ Hz}} + \mathbf{y}_{40 \text{ Hz}}$ and $\mathbf{y}_* = \mathbf{y}_{10 \text{ Hz}} + \mathbf{x}_{40 \text{ Hz}}$. Next, \mathbf{x}_* and \mathbf{y}_* are added to broadband noise generated following the same approach described for the previous experiment, with an SNR of 6.

3.2.2 EEG data

In order to test the performance of our phase TE estimator in the context of BCI, we obtain effective connectivity features from EEG signals recorded under two different cognitive paradigms: the first one consisting of motor imagery (MI) tasks and the second one of a change detection task designed to study working memory (WM) (see Section 1.7). Our aims are to set up classification systems that allow discriminating between the conditions in each paradigm, using as inputs relevant directed interactions among EEG signals and then evaluate their performance in relation to the connectivity measures used to train them.

Classification setup

Feature extraction Let $\Psi = \{\mathbf{X}_n \in \mathbb{R}^{C \times M}\}_{n=1}^N$ be an EEG set holding N trials from either an MI or a WM dataset, recorded from a single subject, where C stands for the number of channels and M corresponds to the number of samples. For MI, $C = 22$ and $M = 500$ (we only select a 2 s long time window stretching from second 3 to second 5), while for WM, $C = 32$ and $M = 717$. In addition, let $\{l_n\}_{n=1}^N$ be a set whose n -th element is the label associated with trial \mathbf{X}_n . For the MI database l_n can take the values of 1 and 2, corresponding to right hand and left hand motor imagination, respectively. Similarly, for the WM database, l_n can take the values of 1, 2, and 3 corresponding to low, medium, and high memory loads. In both cases, our goal is to estimate the class label from relevant effective connectivity features extracted from \mathbf{X}_n . Because of the results obtained for the simulated data (see Section 3.3.1 for details), here we consider features from only three approaches for phase-based effective connectivity estimation, namely, $\text{TE}_{\kappa\alpha}^\theta$, GC^θ , and PSI. Additionally, we also characterize the data through the real-valued versions of $\text{TE}_{\kappa\alpha}$ and GC.

For the real-valued effective connectivity measures considered, we do the following: let $\lambda(\mathbf{x}_c \rightarrow \mathbf{x}_{c'})$ be a measure of effective connectivity between channels $\mathbf{x}_c, \mathbf{x}_{c'} \in \mathbb{R}^M$. By computing $\lambda(\mathbf{x}_c \rightarrow \mathbf{x}_{c'})$ for each pairwise combination of channels in \mathbf{X}_n we obtain a connectivity matrix $\mathbf{\Lambda} \in \mathbb{R}^{C \times C}$. In the case when $c = c'$, we set $\lambda(\mathbf{x}_c \rightarrow \mathbf{x}_{c'}) = 0$. Then, we normalize $\mathbf{\Lambda}$ to the range $[0, 1]$. After performing the above procedure for the N trials, we get set of connectivity matrices $\{\mathbf{\Lambda}_n \in \mathbb{R}^{C \times C}\}_{n=1}^N$. Then, we apply vector concatenation to $\mathbf{\Lambda}_n$ to yield a vector $\phi_n \in \mathbb{R}^{1 \times (C \times C)}$. Next, we stack the N vectors ϕ_n , corresponding to each trial, to obtain a matrix $\Phi \in \mathbb{R}^{N \times (C \times C)}$ holding all directed interactions, estimated through λ , for the EEG set Ψ .

For the phase-based effective connectivity measures of interest, we proceed in a similar fashion: let $\lambda^\theta(\mathbf{x}_c \rightarrow \mathbf{x}_{c'}, f)$ be a measure of phase-based effective connectivity between channels $\mathbf{x}_c, \mathbf{x}_{c'}$ at frequency f . By computing $\lambda^\theta(\mathbf{x}_c \rightarrow \mathbf{x}_{c'}, f)$ for each pair-wise combination of channels in \mathbf{X}_n we obtain a connectivity matrix $\mathbf{\Lambda}(f) \in \mathbb{R}^{C \times C}$ (when $c = c'$, we set $\lambda^\theta(\mathbf{x}_c \rightarrow \mathbf{x}_{c'}, f) = 0$). For the MI database, we vary the values of f in the range from 8 Hz to 18 Hz, in 2 Hz steps, since activity in that frequency range has been associated with MI tasks (Collazos-Huertas et al., 2020). Then, we define two bandwidths of interest $\Delta f \in \{\alpha \in [8 - 12], \beta_l \in [14 - 18]\}$ Hz. Afterward, we average the matrices $\mathbf{\Lambda}(f)$ within each bandwidth, normalize the resulting matrices to the range $[0, 1]$, and stack them together, so that for each trial we have a connectivity matrix $\mathbf{\Lambda}' \in \mathbb{R}^{C \times C \times 2}$. Therefore, for the N trials, we get set of connectivity matrices $\{\mathbf{\Lambda}'_n \in \mathbb{R}^{C \times C \times 2}\}_{n=1}^N$. Then, we apply vector concatenation to $\mathbf{\Lambda}'_n$ to yield a vector $\phi_n \in \mathbb{R}^{1 \times (C \times C \times 2)}$. After that, we stack the N vectors ϕ_n in order to obtain a single matrix $\Phi \in \mathbb{R}^{N \times (C \times C \times 2)}$ characterizing Ψ for the MI data. For the WM we follow the same steps, only that in this case we vary the values of f in the range from 4 Hz to 18 Hz, in 2 Hz steps, since oscillatory activity at those frequencies has been shown to play a role in the interactions between different brain regions during WM (Johnson et al., 2018, 2019). Next, we define three bandwidths of interest $\Delta f \in \{\theta \in [4 - 6], \alpha \in [8 - 12], \beta_l \in [14 - 18]\}$ Hz, which leads to a connectivity matrix $\mathbf{\Lambda}' \in \mathbb{R}^{C \times C \times 3}$ for each trial and ultimately to a matrix $\Phi \in \mathbb{R}^{N \times (C \times C \times 3)}$ characterizing Ψ for the WM data. Note that since the PSI is an antisymmetric connectivity measure, we only use the upper triangular part of the connectivity matrix associated with each trial to build Φ .

Feature selection and classification After characterizing the EEG data, either through real-valued or phase-based effective connectivity measures, we set up subject-dependent classification systems for the MI and WM databases following the same steps described in Section 2.2.3 for each database.

3.2.3 Parameter selection

We used in-house Python implementations of the algorithms for all the connectivity measures studied, except for TE_{KSG}^θ . In that case, we used the implementation provided by the open access toolbox TRENTOOL, a TE estimation and analysis toolbox for Matlab (Lindner et al., 2011).

Regarding the selection of parameters involved in the different effective connectivity

estimation methods, we proceeded as follows: For the TE methods, we estimated all parameters from the real-valued time series, i.e., before extracting the phase time series. The embedding delay τ was set to 1 autocorrelation time (ACT), as proposed in [Vicente et al. \(2011\)](#). The embedding dimension d was selected from the range $d = \{1, 2, \dots, 10\}$ using Cao's criterion ([Cao, 1997](#); [Lindner et al., 2011](#)) (see Appendix E). Note that for any signal pair, the embedding parameters selected are those of the driven or target time series, i.e., to estimate $\text{TE}(\mathbf{x} \rightarrow \mathbf{y})$ we use for both time series the embedding parameters found for \mathbf{y} . The interaction delay u was set as the value generating the largest TE from ranges that varied depending on the experiment: $u = \{1, 2, \dots, 10\}$ for the NMMs, $u = \{1, 4, \dots, 25\}$ for the MI data, and $u = \{50, 60, \dots, 250\}$ for the WM data. Note that the meaning of u in terms of the time delay of the directed interaction between the driving and driven systems is associated with the sampling frequency, e.g., $u = \{1, 2, \dots, 10\}$ for data sampled at 250 Hz translates to a time range between 4 ms and 40 ms. For $\text{TE}_{\kappa\alpha}^\theta$, in most experiments we select a value of $\alpha = 2$, which is neutral to weighting, a convenient choice when there is no previous knowledge about the values of the α parameter better suited for a particular application ([De La Pava Panche et al., 2019](#); [Giraldo et al., 2015](#)). In addition, as kernel function, we employ an RBF kernel with Euclidean distance (see Equation (2.11)). The bandwidth σ was set in each case as the median distance of the data ([Schölkopf et al., 2002](#)). For TE_{KSG}^θ the Theiler correction window and the number of neighbors were left at their default values in TRENTOOL, 4 and 1 ACT, respectively ([Lindner et al., 2011](#)). For the GC methods the order of the autoregressive models o was selected from the range $o = \{1, 3, \dots, 9\}$ using Akaike information criterion ([Akaike, 1974](#); [Gong et al., 2018](#)). Furthermore, in order to estimate the PSI we employed a sliding window 5 frequency bins long (3 bins long for the WM data), centered on the frequency of interest. Finally, for all the connectivity methods involving the extraction of phase time series through Morlet wavelets, we varied the parameter m (see Equation (3.2)) from 3 to 10 in a logarithmic scale, according to the selected frequency of the filter.

3.3 Results and discussion

3.3.1 Neural mass models

The experiments described in Section 3.2.1 are intended to assess whether the phase-based connectivity measures considered in this study correctly detect the direction of

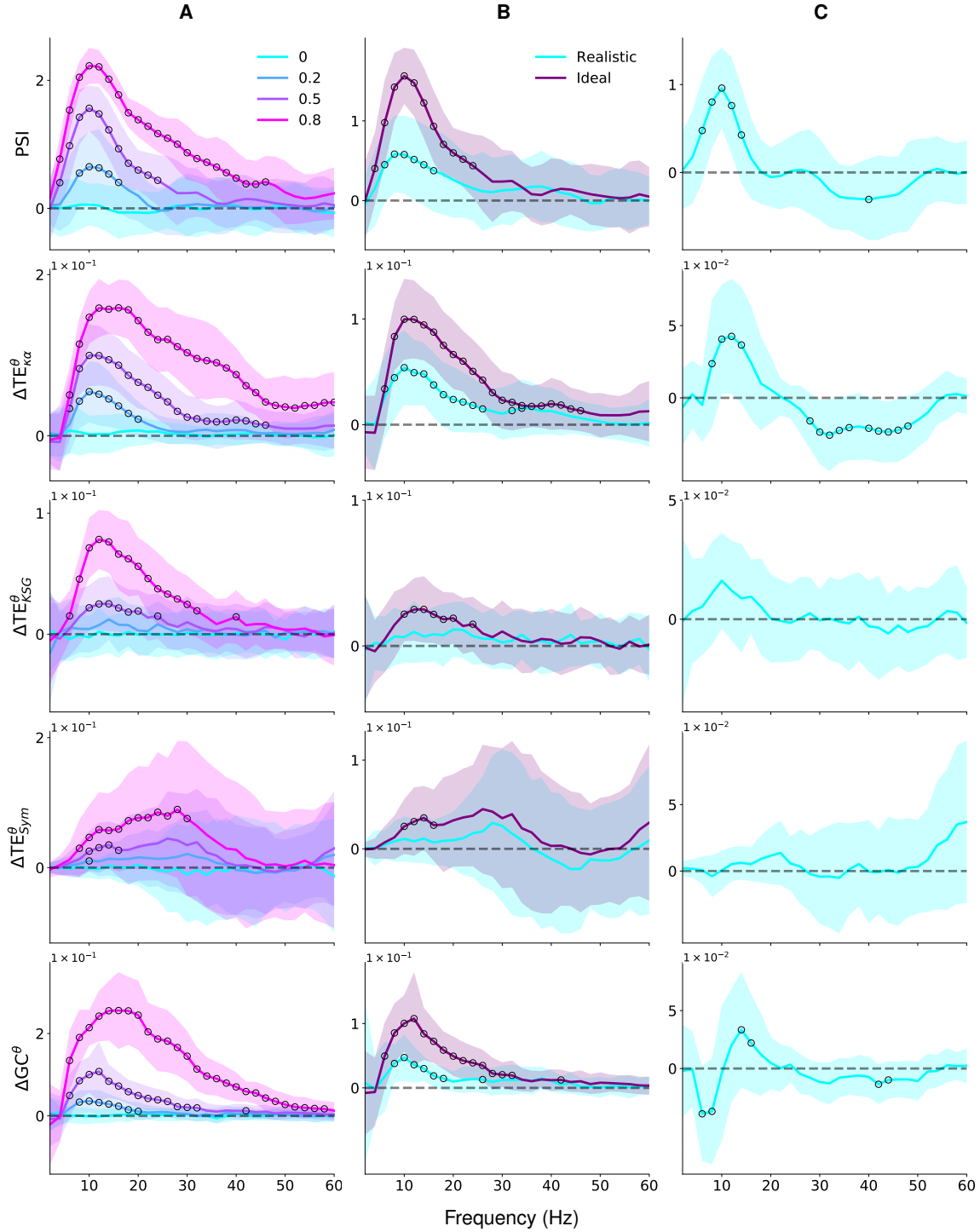


Figure 3.2: Obtained results for the experiments performed using simulated data from NMMs. Column (A) shows the average connectivity values obtained for different levels of coupling strength. Column (B) presents the average connectivity values estimated for ideal signals and for signals contaminated with noise and signal mixing. Column (C) displays the average connectivity values obtained for bidirectional narrowband couplings. The rows correspond to each of the net phase-based effective connectivity estimation approaches considered for the aforementioned experiments. Circled values indicate statistically significant results.

interaction between two time series of known oscillatory properties. Figure 3.2 presents the results obtained from such experiments. Namely, column **A** shows the connectivity values obtained for different levels of coupling strength, column **B** compares the connectivities estimated for ideal signals with those of signals contaminated with noise and mixing, and column **C** displays the results obtained for bidirectional narrowband couplings. The rows in Figure 3.2 correspond to each of the phase-based connectivity measures studied. The first row contains average PSI values computed on the frequency range between 2 Hz and 60 Hz, while rows two to five display average net connectivity values for $TE_{\kappa\alpha}^\theta$, TE_{KSG}^θ , TE_{Sym}^θ , and GC^θ , respectively. Circled values indicate statistically significant connectivities at a particular frequency, according to a permutation test based on randomized surrogate trials. The test identifies connectivity values that are, on average, significantly different from those expected for that connectivity measure applied to non-interacting signals. For the three experiments involving simulated data from NMMs, we use the PSI as a comparison standard, since it is a robust and well-established measure of linear directed interactions defined in terms of phase relations (Jiang et al., 2015; Nolte et al., 2008). Therefore, it is suited to analyze the coupled, mildly nonlinear time series generated by NMMs.

Regarding the first experiment, which modifies the coupling strength between the simulated signals, the obtained results (Figure 3.2, column **A**) show that all the measures studied satisfactorily detect the coupling direction of the simulated data. Note that since we set the NMMs to generate unidirectional interactions from \mathbf{x} to \mathbf{y} , and because of the way we defined $\Delta\lambda$, then all net connectivity values for the simulated coupled signals should be positive. The same is true for the $PSI(\mathbf{x} \rightarrow \mathbf{y})$. On the other hand, only the PSI , $TE_{\kappa\alpha}^\theta$, and GC^θ fulfill the criteria for an overall description of the phase-based interactions present in the data. First, we observe higher net connectivity values at higher coupling strengths, that is to say, stronger interactions lead to larger connectivity estimates. Second, for each coupling strength, there are higher net connectivity values around the frequencies corresponding to the main oscillatory components of the time series generated by the NMMs, in this case, oscillations in the range between 8 Hz and 20 Hz. Third, there are statistically significant results for all the coupling strengths explored, except for non-interacting time series ($C_{12} = 0$). TE_{KSG}^θ does not capture statistically significant interactions for a coupling coefficient value of 0.2, indicating a lower sensitivity to weak couplings. While TE_{Sym}^θ exhibits a very distorted connectivity profile when compared with the PSI . In addition, it has much larger standard deviations for all the coupling strengths considered.

The second experiment assesses the robustness of our proposal to realistic levels of noise and signal mixing, two sources of signal degradation that can lead to spurious connectivity results. In electrophysiological signals, such as EEG, signal mixing arises as a result of field spread, while noise is the result of technical and physiological artifacts (Debener et al., 2012; Mennes et al., 2010; Sakkalis, 2011). The results in Figure 3.2, column **B**, show that PSI , $\text{TE}_{\kappa\alpha}^\theta$, and GC^θ capture statistically significant interactions in the frequencies of interest for both the ideal (no noise or signal mixing) and realistic conditions. The smaller connectivity values for the data contaminated with noise and signal mixing, as compared with the ideal signals, are mostly explained by the reduction in asymmetry between the driving and driven signals caused by mixing (Lobier et al., 2014). On the contrary, we observe that neither TE_{Sym}^θ nor TE_{KSG}^θ produce statistically significant results under the realistic scenario, indicating that those estimators are less robust to signal degradation.

The third experiment aims to evaluate how our proposal deals with bidirectional interactions of localized frequency content. Because of our experimental setup, the obtained results should exhibit a positive deflection around 10 Hz in order to capture the directed interaction from \mathbf{x} to \mathbf{y} and a negative deflection around 40 Hz to represent the directed interaction from \mathbf{y} to \mathbf{x} . Figure 3.2, column **C**, shows that both PSI and $\text{TE}_{\kappa\alpha}^\theta$ successfully detect the change in the direction of interaction in localized frequency bands, with statistically significant connectivity values around the frequencies of interest. However, under this scenario, $\text{TE}_{\kappa\alpha}^\theta$ is less frequency specific for high-frequency interactions than the PSI , with statistically significant connections present for a large range of frequency values around 40 Hz. This is probably due to the filtering step involved in the estimation of $\text{TE}_{\kappa\alpha}^\theta$, while PSI is directly defined on the data spectra. Additionally, TE_{KSG}^θ and TE_{Sym}^θ fail to produce any significant results, while GC^θ shows a statistically significant, non-existing coupling from \mathbf{y} to \mathbf{x} for frequencies under 10 Hz. Note that, ultimately, the permutation test indicates whether the connectivity values obtained are unlikely to be the result of chance and not whether they correctly capture the directed interactions present in the data. In this case, the statistically significant results mean that GC^θ consistently found a directed interaction from \mathbf{y} to \mathbf{x} in the range mentioned before.

The results discussed above indicate that the proposed phase TE estimator is able to detect directed interactions between time series resembling electrophysiological data for different levels of coupling strength, under the presence of noise and signal mixing, and for bidirectional narrowband couplings. Furthermore, they show that it is competi-

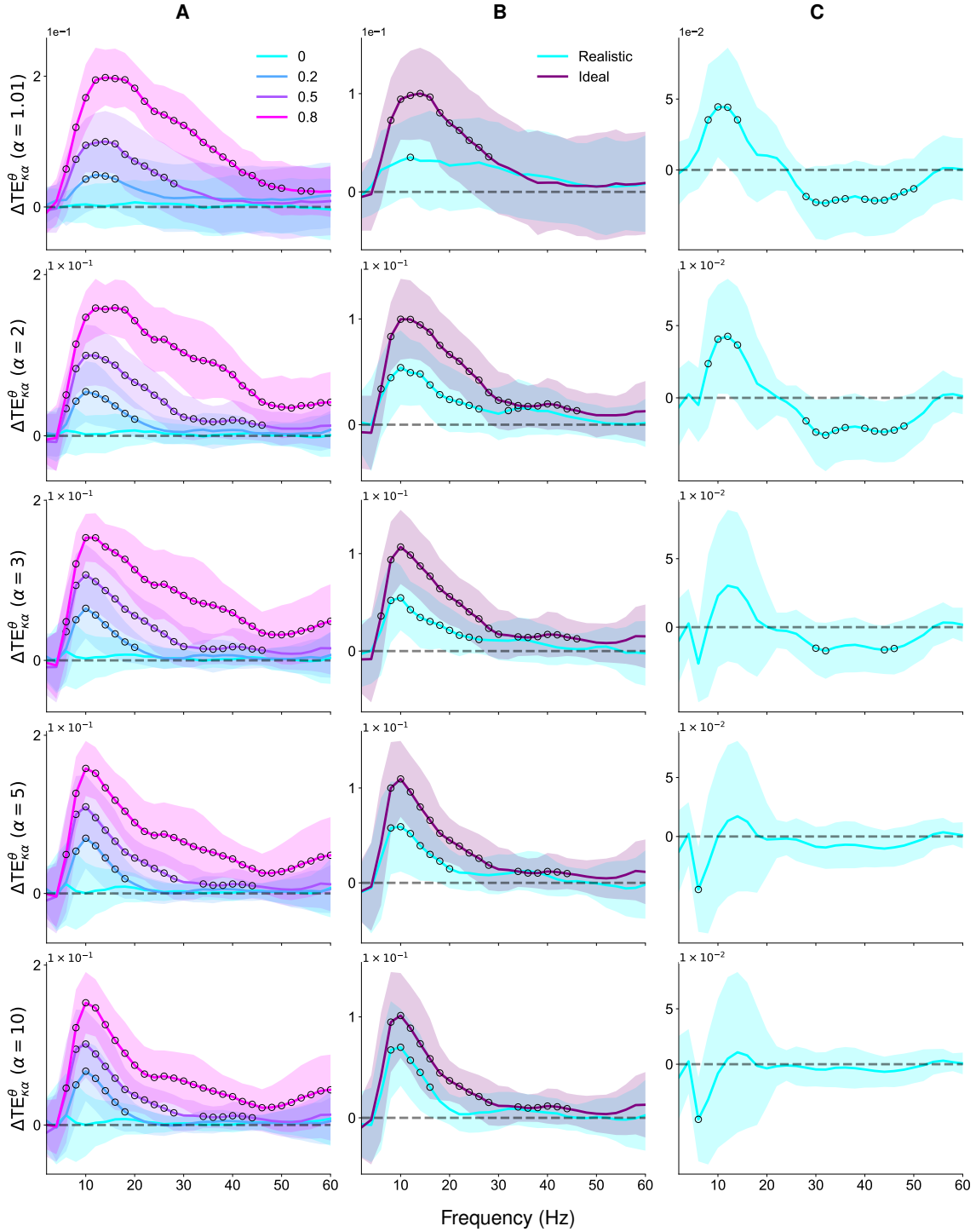


Figure 3.3: Obtained results for the experiments performed using simulated data from NMMs for different values of the α parameter in the proposed $TE_{\kappa\alpha}^\theta$ approach. Column (A) shows the average connectivity values obtained for different levels of coupling strength. Column (B) presents the average connectivity values estimated for ideal signals and for signals contaminated with noise and signal mixing. Column (C) displays the average connectivity values obtained for bidirectional narrowband couplings. The rows correspond to different values of the α parameter in the $TE_{\kappa\alpha}^\theta$ approach. Circled values indicate statistically significant results.

tive with well-established approaches for phase-based net connectivity estimation, such as PSI, in the case of weakly nonlinear signals. Lastly, our results also show that commonly used single-trial TE estimators, such TE_{KSG} and TE_{Sym} , are ill-suited to measure directed interactions between instantaneous phase time series.

Finally, Figure 3.3 presents results analogous to those of Figure 3.2, except that its rows correspond to different values of the α parameter in the proposed $TE_{\kappa\alpha}^\theta$ approach. Most of the discussion previously presented for the case when $\alpha = 2$ holds for the other values of α explored. However, we observe variations in two aspects. First, the connectivity profiles become more peaky around the frequencies of interest as α increases, even in the presence of noise and signal mixing. Second, for $\alpha > 2$ the estimated $TE_{\kappa\alpha}^\theta$ connectivities stop capturing bidirectional interactions of localized frequency content correctly. Because of the latter, and considering the higher computational cost of estimating $TE_{\kappa\alpha}^\theta$ for non-integer values of α (see Section 2.3.4), we set $\alpha = 2$ for all subsequent experiments.

3.3.2 EEG data

Table 3.1: MI and WM classification results in terms of the average classification accuracy for all the effective connectivity measures considered.

	Motor Imagery (acc %)		Working Memory (acc %)
	Cross-Validation	Testing	Cross-Validation
GC	64.3 ± 11.7	57.1 ± 11.0	53.0 ± 7.4
$TE_{\kappa\alpha}$	65.5 ± 11.4	62.8 ± 11.7	67.5 ± 4.2
PSI	62.4 ± 7.8	58.8 ± 8.3	75.2 ± 5.2
GC^θ	67.0 ± 11.9	63.5 ± 14.4	74.5 ± 4.4
$TE_{\kappa\alpha}^\theta$	70.4 ± 12.5	69.0 ± 14.8	93.0 ± 5.9

Table 3.1 presents the average accuracies achieved by the proposed classification systems for both the MI and WM databases, for each effective connectivity method studied. For the MI database, in the training-validation stage, the classifier based on $TE_{\kappa\alpha}^\theta$ features exhibited the highest average performance, closely followed by the one based on GC^θ . In the testing stage, we observe the same overall accuracy ranking, although a smaller drop in the classification accuracy occurs for $TE_{\kappa\alpha}^\theta$ than for GC^θ , which points to a better generalization capacity by the system trained using features

extracted through phase TE. For the WM database, the classifier trained from $TE_{\kappa\alpha}^\theta$ features also displays the highest average accuracy. However, in this case, there is a large gap in performance between the $TE_{\kappa\alpha}^\theta$ -based classification system and the closest results from an alternative approach. Furthermore, the results in Table 3.1 show a consistent improvement in performance between the classifiers that use real-valued TE estimates and those that are trained from phase TE values. They also show relatively low accuracies for the classifiers trained using PSI features. We believe the latter can be explained by two factors. First, by definition, the PSI is unable to explicitly detect bidirectional interactions. It measures connectivity in terms of lead/lag relations, which leads to ambiguity regarding the meaning of PSI values close to zero, since they can be the result of either the lack of interaction or evenly balanced bidirectional connections. If the relevant information to discriminate among the conditions of a cognitive paradigm is related to the bidirectionality of interactions, such as those present in WM (Johnson et al., 2018, 2019), then the PSI might not be an adequate characterization strategy. Secondly, the PSI, like GC, is a linear measure; its performance degrades for strongly nonlinear phase relationships.

In the sections below, we detail and further discuss the results obtained for each database.

Motor imagery

Figure 3.4 depicts the average classification accuracy for all subjects in MI database as a function of the number of selected features during the training-validation stage, for $TE_{\kappa\alpha}$ and $TE_{\kappa\alpha}^\theta$. These results show there is a small improvement in the ability to discriminate between the MI tasks when using features extracted through phase TE, as compared with real-valued TE. In addition, they reveal that the CKA-based feature selection strategy successfully identified the most relevant connections for MI task classification. That is to say, the classification system has a stable performance even for a very reduced number of connectivity features. This is fundamental for any practical BCI application that intends to use phase TE as a characterization strategy, since estimating single-trial phase TE is computationally expensive (Lobier et al., 2014). Therefore, it is important to reduce as much as possible the number of channel pair connectivity features required to achieve peak classification performance. Additionally, it is important to highlight that while classification accuracies in Figure 3.4, and in Table 3.1, are in the same range of those obtained through other connectivity-based characterization approaches (De La Pava Panche et al., 2019; García-Murillo et al., 2021), they are far below those obtained

from methods such as common spatial patterns (Elasuty and Eldawlatly, 2015; Gómez et al., 2018; Li et al., 2018). A possible explanation is that bivariate TE might be more robust at describing long-range interactions rather than local ones (Ursino et al., 2020), like those arising from MI-related activity, centered on the sensorimotor area. In addition, the differences with the results in Section 2.3.3, where we used $TE_{\kappa\alpha}$ to characterize the same database, lay mostly in the fact that in this chapter we select and analyze one 2 s long time window covering the period right after the end of the visual cue, while in Section 2.3.3 we report results from multiple overlapping windows covering the entirety of the task. Lastly, the large standard deviations from the average accuracies in Figure 3.4 point to disparate performances for different subjects.

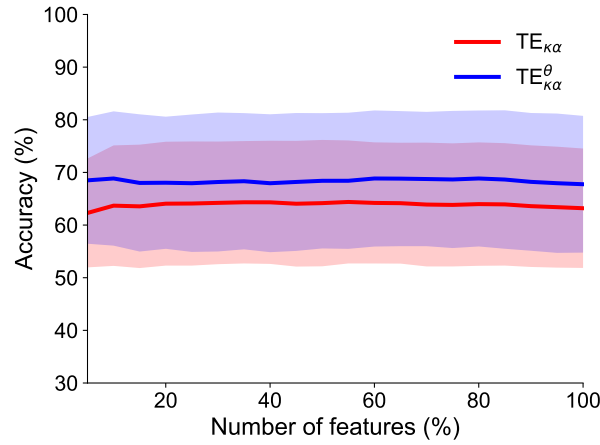


Figure 3.4: Average classification accuracies, and their standard deviations, for all subjects in the MI database as a function of the number features selected to train the classifiers.

Figure 3.5A shows the highest average classification accuracy per subject for $TE_{\kappa\alpha}^{\theta}$, GC^{θ} and PSI, during the training-validation stage. The subjects are ordered from highest to lowest performance. The analogous information for the testing stage is presented in Figure 3.5B. In both stages, the $TE_{\kappa\alpha}^{\theta}$ -based classifier performs slightly better than those based on alternative connectivity estimation strategies in most subjects. In addition, as inferred from Figure 3.4, there are large variations in performance for the different subjects in the database, consistent across the two classification stages. We previously observed this behavior in the results presented in Section 2.3.3. It has also been reported elsewhere (De La Pava Panche et al., 2019; Elasuty and Eldawlatly, 2015; Gómez et al., 2018; Li et al., 2018; Liang et al., 2016).

In order to gain insight into the observed performance differences, in the case of $TE_{\kappa\alpha}^{\theta}$, we exploited the second advantage provided by the CKA-based relevance analysis

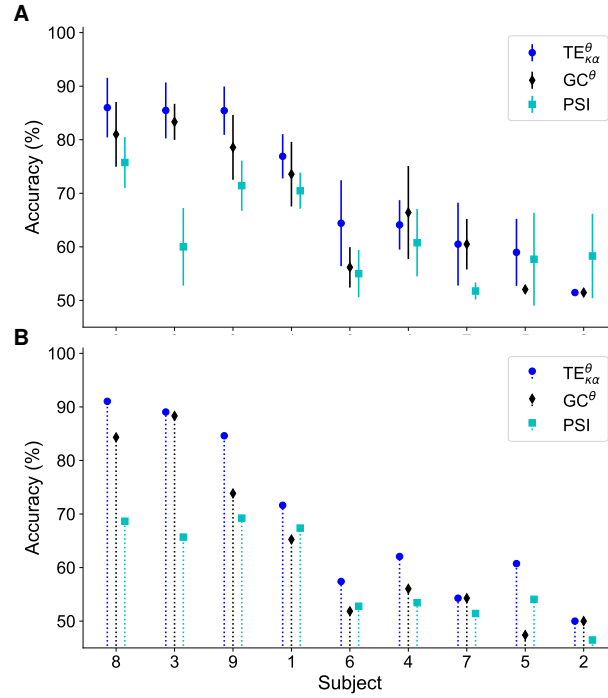


Figure 3.5: **(A)** Highest average classification accuracy for each subject in the MI database during the training-validation stage. **(B)** Accuracies obtained for each subject during the testing stage. The subjects are ordered from highest to lowest performance according to the accuracies obtained for the $TE_{\kappa\alpha}^{\theta}$ -based classifier in the training-validation stage.

(see Appendix D). The relevance vector index $\boldsymbol{\varrho}$ not only allows us to perform feature selection but also provides a one-to-one relevance mapping to each connectivity feature. That is to say, we can reconstruct normalized relevance connectivity matrices by properly reshaping $\boldsymbol{\varrho}$, so as to visualize the connectivity pairs and frequency ranges that are discriminant for the task of interest. In that line, we followed the approach proposed in [García-Murillo et al. \(2021\)](#) to interpret the relevance information by clustering the subjects according to common relevance patterns.

First, for each subject and frequency band of interest, we obtained a relevance vector $\boldsymbol{\varrho}_{n,\Delta f} \in \mathbb{R}^C$ whose elements were associated with each node (EEG channel) in the data by computing the relevance of the total information flow of every node. Such magnitude was defined as the sum of the relevance values $\boldsymbol{\varrho}$, obtained from all data in the training dataset, corresponding to all directed interactions targeting and originating from a particular node. Then, we concatenated the vectors $\boldsymbol{\varrho}_{n,\Delta f} \in \mathbb{R}^C$ for all frequency bands to obtain a single relevance vector $\boldsymbol{\varrho}_n \in \mathbb{R}^{2C}$. Next, we reduced the dimension of the relevance vectors $\boldsymbol{\varrho}_n$ of each subject through t-Distributed Stochastic Neighbor

Embedding (t-SNE), which preserves the spatial relationships existing in the initial higher-dimensional space (Linderman and Steinerberger, 2019). Figure 3.6A shows the obtained two-dimensional representation of the relevance vectors for each subject in the MI database, colored according to their respective classification accuracy. Note that the distribution of the subjects in the plot is related to their classification accuracies. This indicates that shared relevance patterns are related to the obtained classification results, meaning that subjects with similar \mathbf{q}_n had close performances. Then, we grouped the subjects into two clusters using the k-means algorithm. The number of clusters was selected by visual inspection of the t-SNE results. Figure 3.6B displays the two groups, termed G. I and G. II. The $\text{TE}_{\kappa\alpha}^\theta$ -based classifier has average accuracies of 0.59 ± 0.05 and 0.80 ± 0.09 for the subjects in G. I and G. II, respectively.

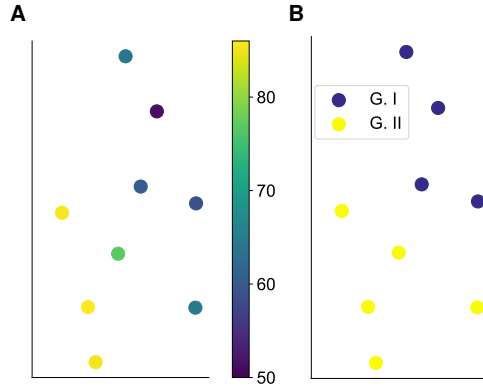


Figure 3.6: (A) Two-dimensional representation of the relevance vectors for each subject in the MI database obtained after applying t-SNE on \mathbf{q}_n . (B) Groups identified by k-means. For the $\text{TE}_{\kappa\alpha}^\theta$ -based classifier the subjects grouped in G. I have an average accuracy of 0.59 ± 0.05 , while those in G. II have an average accuracy of 0.80 ± 0.09 .

Finally, Figure 3.7 shows the average nodal relevance, as defined by \mathbf{q}_n , and the most relevant connectivities for each group, discriminated by frequency band. For G. I we observe high node relevance mostly in the α band in right fronto-central, left-central, and centro-parietal regions. The most relevant connections in the α band tend to originate or target fronto-central nodes, while the ones in the β_l band favor parietal and centro-parietal areas. For G. II, the node relevance is concentrated around the right centro-parietal region, particularly channel CP4, for both frequency bands. The most relevant connections in the α band involve short-range interactions mainly between centro-parietal and central regions. The most relevant connections in the β_l band, which display higher values than those of α , originate from CP3 and CP4 and target central and fronto-central nodes. Since the G. II includes all the subjects with

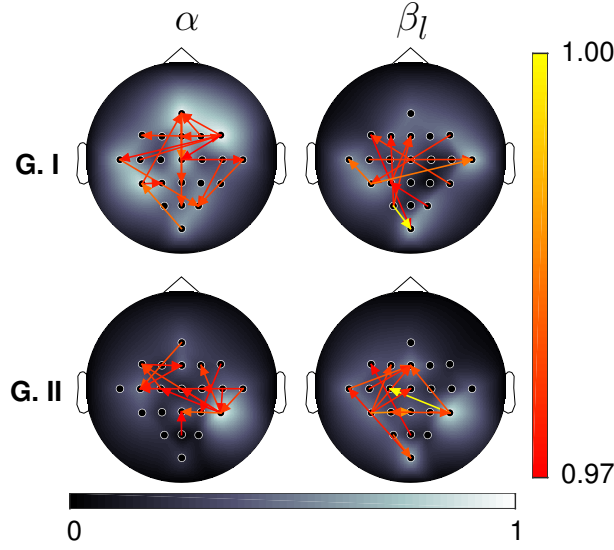


Figure 3.7: Topoplots of the average node (channel) relevance for each group of clustered subjects and frequency band of interest in the MI database (see Figure 3.6). The arrows represent the most relevant connectivities for each group. For visualization purposes, only 3% of the connections, those with the highest average relevance values per group, are depicted.

good classification performances, we can conclude that the information that allows to satisfactorily classify the left and right hand MI tasks from $TE_{\kappa\alpha}^\theta$ features corresponds mostly to the incoming and outgoing information flow coded in the phases of the oscillatory activity in the centro-parietal region. These results are in line, in terms of spatial location, with those we found in [De La Pava Panche et al. \(2019\)](#), and with physiological interpretations that argue that MI activates motor representations in the parietal area and the premotor cortex ([Héту et al., 2013](#)).

Working memory

Figure 3.8 presents the average classification accuracy for all subjects in the WM database as a function of the number of selected features, for $TE_{\kappa\alpha}$ and $TE_{\kappa\alpha}^\theta$. The results show that the classifier trained from phase TE features markedly outperforms the one trained using real-valued TE estimates, as long as the appropriate percentage of features is selected. This difference might be attributed to the hypothesized phase-based nature of directed interactions during WM tasks ([Dimitriadis et al., 2016a](#); [Johnson et al., 2018](#)), which would be better captured by phase TE. Furthermore, both accuracy curves highlight the importance of feature selection, since they show a steep performance degradation as more features are used to train the classifiers. In this case, the CKA-based

relevance analysis not only allows reducing the number of features needed to successfully classify the three cognitive load levels present in the WM data but also prevents the classifiers from being confounded by connections that do not hold relevant information to discriminate between the target conditions.

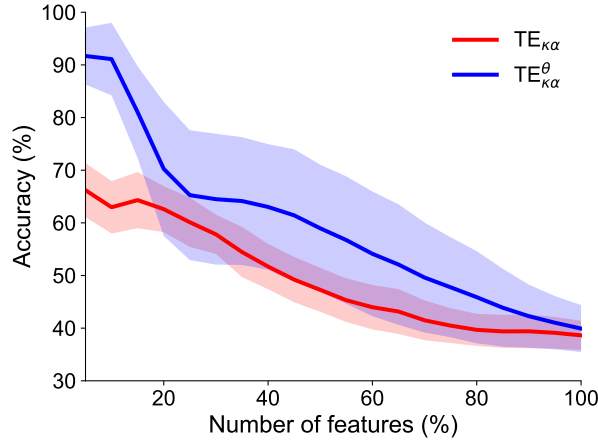


Figure 3.8: Average classification accuracies, and their standard deviations, for all subjects in the WM database as a function of the number features selected to train the classifiers.

Figure 3.9 depicts the highest average classification accuracy per subject for $\text{TE}_{\kappa\alpha}^{\theta}$, GC^{θ} and PSI. The subjects are ordered from highest to lowest performance. Unlike the results obtained for the MI database, we do not observe an underperforming group of subjects, especially after considering the fact that for the WM database the classifiers must discriminate among three classes instead of two. On the other hand, in this case, the $\text{TE}_{\kappa\alpha}^{\theta}$ -based classifier largely outperforms those based on alternative connectivity estimation strategies in most subjects. Here, as in Section 2.3.3, we must point out that the auxiliary cross-validation step introduced for feature selection leads to data leakage, since it requires all the available data to estimate $\bar{\varrho}$, which renders it a nonviable approach for practical BCI implementations and can inflate performance measures. Nonetheless, we implemented the same strategy for all classification systems involving the WM database. Therefore, we can carry out valid comparisons among the classifiers trained with features extracted through the different connectivity measures explored.

In order to elucidate the pairwise connectivities, and their corresponding frequency bands, that allow the $\text{TE}_{\kappa\alpha}^{\theta}$ -based classification system to successfully discriminate among different memory loads, we proceeded as described in Section 3.3.2, and from $\bar{\varrho}$ obtained a node relevance vector $\bar{\varrho}_n \in \mathbb{R}^{3C}$. Then, we applied t-SNE on $\bar{\varrho}_n$. Figure 3.10A shows

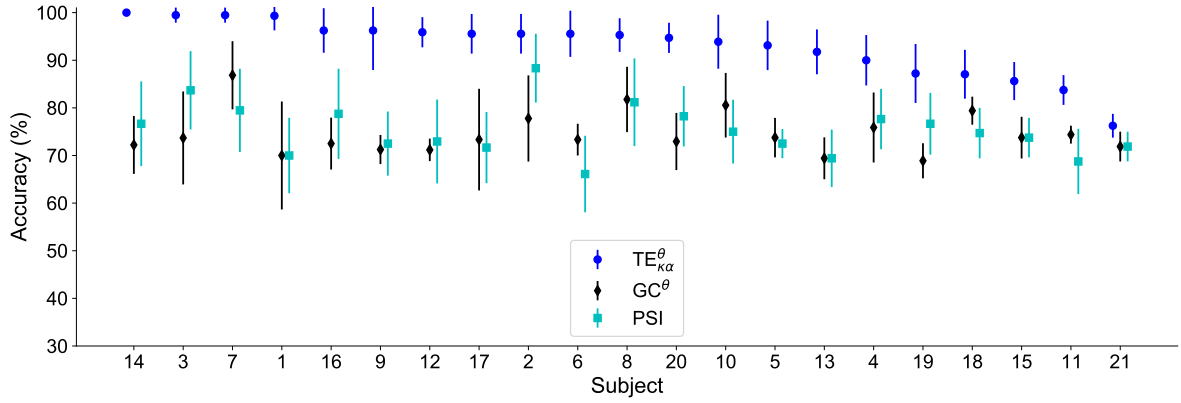


Figure 3.9: Highest average classification accuracy for each subject in the WM database. The subjects are ordered from highest to lowest performance according to the accuracies obtained for the $TE_{\kappa\alpha}^{\theta}$ -based classifier.

the obtained two-dimensional representation of the relevance vectors for each subject in the WM database. Unlike the results observed before for the MI database, there is not a clear association between the subject distribution on the plot and their classification accuracies. Nonetheless, Figure 3.10A shows the presence of well-defined groups sharing similar relevance patterns. As before, we grouped the subjects into clusters using the k-means algorithm. The number of clusters was selected as three by visual inspection of the t-SNE results. Figure 3.10B displays the three groups, termed G. I, G. II, and G. III. The $TE_{\kappa\alpha}^{\theta}$ -based classifier has average accuracies of 0.94 ± 0.04 , 0.92 ± 0.08 , and 0.93 ± 0.08 for the subjects in G. I, G. II, and G. III, respectively.

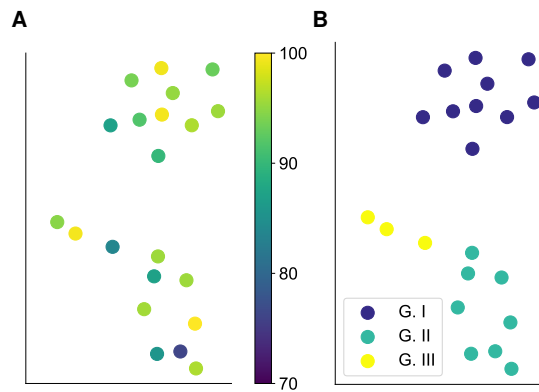


Figure 3.10: (A) Two-dimensional representation of the relevance vectors for each subject in the WM database obtained after applying t-SNE on ϱn . (B) Groups identified by k-means. For the $TE_{\kappa\alpha}^{\theta}$ -based classifier the subjects grouped in G. I, have an average accuracy of 0.94 ± 0.04 , while those in G. II and G. III have average accuracies of 0.92 ± 0.08 and 0.93 ± 0.08 , respectively.

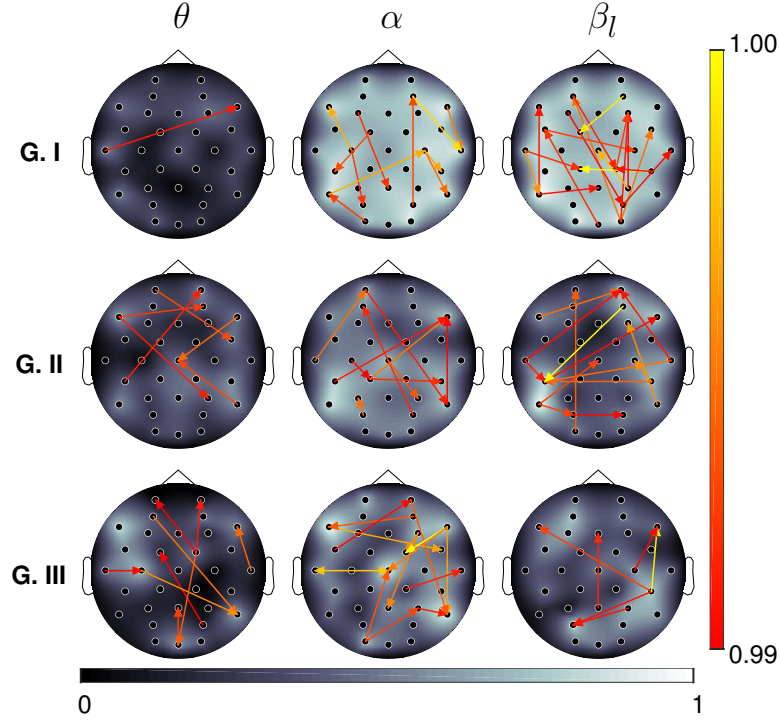


Figure 3.11: Topoplots of the average node (channel) relevance for each group of clustered subjects and frequency band of interest in the WM database (see Figure 3.10). The arrows represent the most relevant connectivities for each group. For visualization purposes, only the 1% of the connections, those with the highest average relevance values per group, are depicted.

Lastly, Figure 3.11 shows the average nodal relevance, as defined by ϱ_n , and the most relevant connectivities for each group, discriminated by frequency band. For G. I we observe widespread high node relevance in both the α and β_l bands and low node relevance in the θ band. Most relevant connections are present in the β_l band with many connections originating in the parieto-occipital region and targeting frontal and centro-frontal areas. For G. II and G. III node relevance is more evenly distributed across the three frequency bands considered. Spatially, it is more prominent around some pre-frontal, frontal, centro-parietal, and parietal nodes. In terms of the most relevant connections, we observe long-range contralateral interactions involving mostly the regions previously listed, as well as some connections to and from temporal areas. Therefore, we argue that the information flow between frontal, parietal, and temporal regions, coded in the phases of oscillatory activity in the θ , α , and β_l bands, is what allowed us to discriminate among different memory loads from $TE_{\kappa\alpha}^\theta$ features. These results agree with several studies that identify fronto-parietal and fronto-temporal neural circuits operating in frequency ranges spanning from θ to β as key during the activation

of working memory (Dimitriadis et al., 2016a; Johnson et al., 2018, 2019).

3.3.3 Limitations

In this study, we employed Morlet wavelets as filters for instantaneous phase extraction prior to phase TE estimation, as proposed in Lobier et al. (2014). However, as discussed by the authors in Lobier et al. (2014), the choice of filter can influence the behavior of phase TE. This is an aspect we have yet to explore for our proposal. In the same line, in Weber et al. (2017) the authors showed, using the Kraskov-Stögbauer-Grassberger TE estimator on real-valued filtered signals, that filtering and downsampling are deleterious for TE, since they can lead to altered time delays and hide certain causal interactions. Furthermore, from a conceptual perspective, while filtering dampens spectral power, it does not always remove the information contained in specific frequencies (Pinzuti et al., 2020). This would hinder the isolation of frequency specific interactions in TE estimates from real-valued filtered data, the most common approach to obtain spectrally resolved TE values. Whether those effects are also present in the case of phase TE is yet to be analyzed; however, as pointed out in Pinzuti et al. (2020), phase TE is conceptually different from spectrally resolved TE. Additionally, the results obtained with our phase TE estimator for the NMM data closely follow those obtained with the PSI, a measure that does not rely on data filtering, which points to a certain degree of robustness to the negative effects that might be associated with phase extraction through complex filtering. A related issue is the possible effects on our proposal of the preprocessing pipelines employed on the EEG data, which involve spectral and spatial filtering. Regarding the former, we have not studied its effects in this work; while for the latter, surface Laplacian positively impacted the discrimination capability of the connectivity features obtained from the different measures considered.

In addition, we selected the autocorrelation time and Cao's criterion to obtain the embedding parameters for all the TE estimation methods. More complex approaches such as time-delayed mutual information and Ragwitz criterion may yield better results (Lindner et al., 2011). However, since our motivation was to propose a single-trial phase TE estimator suited as characterization method for BCI applications, the choice of simple parameter estimation methods is justified. As a matter of fact, a practical implementation of a phase TE-based BCI system would likely require further simplifications regarding parameter estimation, in order to facilitate the computation of phase TE in real time. Furthermore, our proposed phase TE estimator inherits most of the limitations of $TE_{\kappa\alpha}$, described in Section 2.3.4.

3.4 Summary

In this chapter, we proposed a single-trial phase TE estimator. Our method combines a kernel-based TE estimation approach, which defines effectivity connectivity as a linear combination of Renyi's entropy measures of order α , with instantaneous phase time series extracted from the data under analysis. We tested the performance of our proposal on synthetic data generated through NMMs and on two EEG databases obtained under MI and WM paradigms. We compared it with commonly used single-trial TE estimators, applied to phase time series, and the PSI and GC. Our results showed that the proposed phase TE estimator successfully detected the direction of interaction between individual pairs of signals, capturing the differences in coupling strength and displaying statistically significant results around the frequencies corresponding to the main oscillatory components present in the data. It also succeeded in detecting bidirectional interactions of localized frequency content and was robust to realistic noise and signal mixing levels. Moreover, our method, coupled with a CKA-based relevance analysis, revealed discriminant spatial and frequency-dependent patterns for both the MI and WM databases, leading to improved classification performance compared with approaches based on real-valued TE estimation. In all our experiments, the proposed single-trial kernel-based phase TE estimator was competitive with the comparison methods previously listed in terms of the performance assessment metrics employed.

Chapter 4

Kernel-based Renyi's phase transfer entropy for the estimation of directed phase-amplitude interactions

In this chapter, we reframe the problem of estimating directed phase-amplitude interactions through TE as a phase TE problem. That is to say, as the computation of TE between two instantaneous phase time series. We do so by borrowing the underlying premise of the cross-frequency directionality (CFD), a linear connectivity measure capable of estimating the direction of PAC ([Jiang et al., 2015](#)), along with the use of the phase TE estimation methodology developed in Chapter 3. We hypothesize that for directed PAC the proposed approach can correctly identify the interacting frequencies, as well as the direction of interaction, while being robust to common factors that degrade the performance of connectivity estimation methods such as the presence of high levels of noise and volume conduction effects ([Bastos and Schoffelen, 2016](#)).

In order to test our proposal, we use a simulation model that allows generating synthetic data with unidirectionally phase-amplitude couplings at two target frequencies ([Jiang et al., 2015](#)), as well as the MI and WM databases described in Section 1.7. Obtained results for the simulated data show that the proposed approach successfully captures statistically significant phase-amplitude interactions, correctly identifying the direction of interaction and the target frequencies under noisy and signal mixing conditions. Furthermore, the results for the WM data reveal that our proposal captures discriminant phase-amplitude connectivity patterns that allow identifying the cognitive

load associated with a trial of the change-detection task. On the other hand, the proposed approach is unable to capture discriminant interactions for the MI data.

4.1 Kernel-based Renyi's phase transfer entropy for the estimation of directed phase-amplitude interactions

4.1.1 Transfer entropy for directed phase-amplitude interactions

As described in Section 1.4.1, given two time series \mathbf{x} and \mathbf{y} , in order to estimate the TE from the phase of \mathbf{x} at a frequency f_l to the amplitude envelope of \mathbf{y} at a frequency f_h , we first obtain complex time series $\mathbf{s}_x(f_l) = \boldsymbol{\varsigma}^x e^{i\boldsymbol{\theta}^x} \in \mathbb{C}^T$ and $\mathbf{s}_y(f_h) = \boldsymbol{\varsigma}^y e^{i\boldsymbol{\theta}^y} \in \mathbb{C}^T$, which contain the filtered values of \mathbf{x} and \mathbf{y} at f_l and f_h , respectively; where $\boldsymbol{\theta}^x, \boldsymbol{\theta}^y \in [-\pi, \pi]_{t=1}^T$ are instantaneous phase time series, and $\boldsymbol{\varsigma}^x, \boldsymbol{\varsigma}^y \in \mathbb{R}^T$ are amplitude envelopes (Lobier et al., 2014). Then, we compute the desired TE as:

$$TE^{\theta_\varsigma}(\mathbf{x} \rightarrow \mathbf{y}, f_l, f_h) = H_S(\boldsymbol{\varsigma}_{t-1}^{y,dy}, \boldsymbol{\theta}_{t-u}^{x,dx}) - H_S(\boldsymbol{\varsigma}_t^y, \boldsymbol{\varsigma}_{t-1}^{y,dy}, \boldsymbol{\theta}_{t-u}^{x,dx}) + H_S(\boldsymbol{\varsigma}_t^y, \boldsymbol{\varsigma}_{t-1}^{y,dy}) - H_S(\boldsymbol{\varsigma}_{t-1}^{y,dy}),$$

where $\boldsymbol{\theta}_t^{x,dx}$ and $\boldsymbol{\varsigma}_t^{y,dy}$ are time embedded versions of $\boldsymbol{\theta}^x$ and $\boldsymbol{\varsigma}^y$. This approach is schematized in Figure 4.1A.

4.1.2 Cross-frequency directionality

The cross-frequency directionality (CFD) estimates the direction of interaction between the phase of low frequency (f_l) oscillations and the amplitude of faster, higher frequency (f_h) oscillations (Jiang et al., 2015). It is based on the phase slope index (PSI), which measures the coupling directionality between two oscillatory signals of similar frequencies (Nolte et al., 2008). Given two time series \mathbf{x} and \mathbf{y} , the CFD from \mathbf{x} to \mathbf{y} is computed as the PSI between \mathbf{x} , the time series containing the slow oscillations of interest, and the amplitude envelope of \mathbf{y} at f_h ($\boldsymbol{\varsigma}^y$). Thus:

$$CFD(\mathbf{x} \rightarrow \mathbf{y}, f_h) = PSI(\mathbf{x} \rightarrow \boldsymbol{\varsigma}^y) = \Im \left(\sum_{f \in F} C_{\mathbf{x}\boldsymbol{\varsigma}}^*(f) C_{\mathbf{x}\boldsymbol{\varsigma}}(f + df) \right), \quad (4.1)$$

where $C_{\mathbf{x}\varsigma} = S_{\mathbf{x}\varsigma}/\sqrt{S_{\mathbf{x}\mathbf{x}}S_{\varsigma\varsigma}}$ is the complex coherence, $S_{\mathbf{x}\varsigma} \in \mathbb{C}$ is the cross-spectrum between \mathbf{x} and ς^y , $S_{\mathbf{x}\mathbf{x}}, S_{\varsigma\varsigma} \in \mathbb{C}$ are the auto-spectrums of \mathbf{x} and ς^y , $df \in \mathbb{R}^+$ corresponds to the frequency resolution, F indicates the frequency range over which the slope is summed, and \Im stands for the fact that only the imaginary part of the sum is selected (Nolte et al., 2008). Therefore, the CDF estimates the slope of the phase difference, between the phases of \mathbf{x} and ς^y , as a function of frequency. That is to say, it translates the problem of estimating the direction of phase-amplitude interactions to estimating the interaction between phases.

4.1.3 Phase transfer entropy and directed phase-amplitude interactions

We begin by noting that the conventional approach to estimate directed phase-amplitude interactions through TE implies the computation of TE from data of different properties, a phase time series θ^x , which represents a circular variable, and a smooth, continuous amplitude envelope ς^y . In this work, we reformulate the problem of directed phase-amplitude interaction detection using TE as a phase TE estimation task (Lobier et al., 2014). We do so by applying the same idea behind the CFD: obtaining the directionality of interaction between phase and amplitude time series can be redefined as estimating the direction of interaction between two phase time series. We hypothesize that such change can improve the robustness of phase-amplitude TE estimates to signal degradation by noise and volume conduction effects.

Given two time series \mathbf{x} and \mathbf{y} , we want to estimate the TE from the phase of \mathbf{x} at a frequency f_l to the amplitude envelope of \mathbf{y} at a frequency f_h . As before, $\theta^x \in [-\pi, \pi]_{t=1}^T$ and $\varsigma^y \in \mathbb{R}^T$ are the corresponding phase and amplitude time series, obtained at the adequate frequencies. However, before TE computation, we obtain a complex representation of ς^y at f_l , $\mathbf{s}_\varsigma(f_l) = \varsigma^\varsigma e^{i\theta^\varsigma} \in \mathbb{C}^T$, where $\theta^\varsigma \in [-\pi, \pi]_{t=1}^T$ is an instantaneous phase time series, and $\varsigma^\varsigma \in \mathbb{R}^T$ is an amplitude envelope (See Figure 4.1B). Next, we define:

$$TE^{\theta^\varsigma}(\mathbf{x} \rightarrow \mathbf{y}, f_l, f_h) = H_S(\theta_{t-1}^{\varsigma, d\varsigma}, \theta_{t-u}^{x, dx}) - H_S(\theta_t^\varsigma, \theta_{t-1}^{\varsigma, d\varsigma}, \theta_{t-u}^{x, dx}) + H_S(\theta_t^\varsigma, \theta_{t-1}^{\varsigma, d\varsigma}) - H_S(\theta_{t-1}^{\varsigma, d\varsigma}), \quad (4.2)$$

where $\theta_t^{x, dx}$ and $\theta_t^{\varsigma, d\varsigma}$ are time embedded versions of θ^x and θ^ς .

The quantity in Equation 4.2 indicates the estimation of TE from two-phase time series extracted at the same frequency (f_l). Thus, it corresponds to the definition of

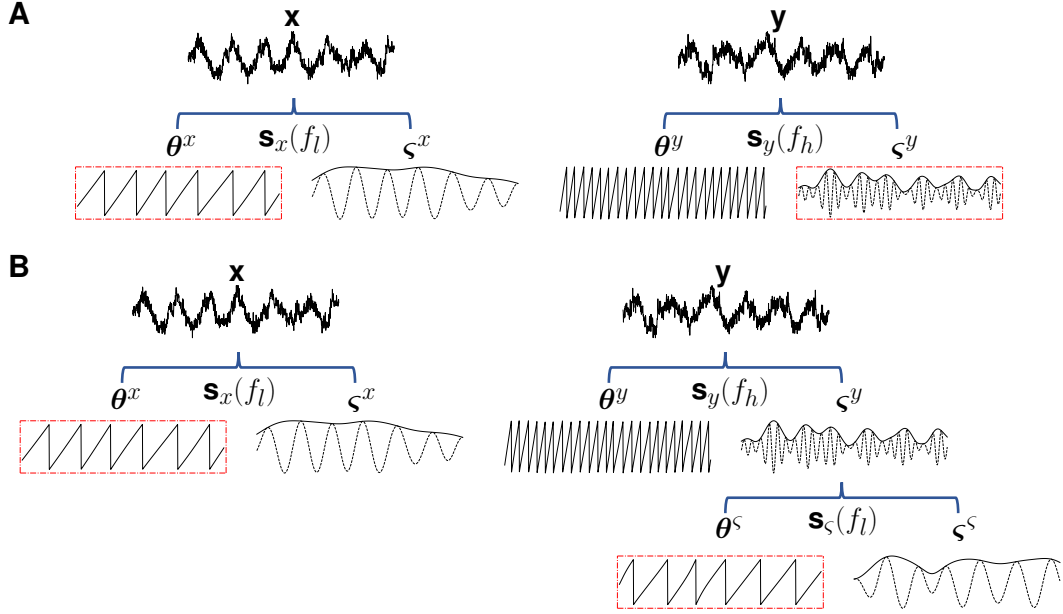


Figure 4.1: (A) The conventional approach to capture directed phase-amplitude interactions through TE consists in estimating TE from instantaneous phase and amplitude time series extracted at frequencies f_l and f_h , respectively. (B) Our approach implies a further step comprising the extraction of the phase of the amplitude time series at frequency f_l , in order to reformulate the problem as a phase TE estimation task.

phase TE. Using the approach developed in Chapter 3 for phase TE estimation, we can redefine Equation 4.2 as:

$$\begin{aligned}
 TE_{\kappa\alpha}^{\theta\theta^\varsigma}(\mathbf{x} \rightarrow \mathbf{y}, f_l, f_h) &= H_\alpha \left(\mathbf{K}_{\theta_{t-1}^{\varsigma, d\varsigma}}, \mathbf{K}_{\theta_{t-u}^{x, dx}} \right) - H_\alpha \left(\mathbf{K}_{\theta_t^\varsigma}, \mathbf{K}_{\theta_{t-1}^{\varsigma, d\varsigma}}, \mathbf{K}_{\theta_{t-u}^{x, dx}} \right) \\
 &\quad + H_\alpha \left(\mathbf{K}_{\theta_t^\varsigma}, \mathbf{K}_{\theta_{t-1}^{\varsigma, d\varsigma}} \right) - H_\alpha \left(\mathbf{K}_{\theta_{t-1}^{\varsigma, d\varsigma}} \right), \quad (4.3)
 \end{aligned}$$

where $H_\alpha(\cdot)$ stands for the kernel-based formulation of Renyi's α entropy introduced in Giraldo et al. (2015), $\mathbf{K}_{\theta_t^\varsigma}, \mathbf{K}_{\theta_{t-1}^{\varsigma, d\varsigma}}, \mathbf{K}_{\theta_{t-u}^{x, dx}} \in \mathbb{R}^{(D-u) \times (D-u)}$ are kernel matrices holding elements $k_{ij} = \kappa(\mathbf{a}_i, \mathbf{a}_j)$, with $\kappa(\cdot, \cdot) \in \mathbb{R}$ a positive definite and infinitely divisible kernel function. For $\mathbf{K}_{\theta_t^\varsigma}$, $a_i, a_j \in \mathbb{R}$ contain the values of the time series θ^ς at times i and j . While for $\mathbf{K}_{\theta_{t-1}^{\varsigma, d\varsigma}}$ and $\mathbf{K}_{\theta_{t-u}^{x, dx}}$ the vectors $\mathbf{a}_i, \mathbf{a}_j \in \mathbb{R}^d$ correspond to the time embedded versions of θ^ς and θ^x , $\theta_t^{\varsigma, d\varsigma}$ and $\theta_t^{x, dx}$, respectively, at times i and j , in accordance with the time indexing of TE.

4.2 Experiments

4.2.1 Simulated phase-amplitude interactions

Simulation model

In order to evaluate the performance of our proposal, we generate simulated time series using a modified version of the PAC modeling strategy introduced in [Jiang et al. \(2015\)](#). The model simulates a directed interaction from the phase of a time series $\mathbf{x} \in \mathbb{R}^N$, at a low frequency f_l , to the amplitude of a time series $\mathbf{y} \in \mathbb{R}^N$, at a high frequency f_h . We implement the model as follows: first, we build d time series' segments \mathbf{x}'_i corresponding to a period of a sinusoidal signal:

$$\mathbf{x}'_i = A_i(\sin(2\pi f_i \mathbf{t}_i + 1.5\pi) + 1), \quad (4.4)$$

where $f_i = 1/T_i$, $\mathbf{t}_i = \{0, dt, 2dt, \dots, T_i\}$, and $dt = 1/f_s$, with f_s the sampling frequency in Hz; as shown in Figure 4.2A. For each segment i , the amplitude A_i and the period T_i are drawn from Gaussian distributions with means $A = 1$ and $T = 1/f_l$, and standard deviations of $0.1A$ and $0.2T$, respectively. Then, we concatenate the d segments to obtain a continuous signal of varying amplitude, as depicted in Figure 4.2B, $\mathbf{x}' \in \mathbb{R}^N = [\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_d]$, and with a power spectrum peaking at around f_l . Next, we generate a signal oscillating at f_h , whose amplitude is a function of \mathbf{x}' , by defining:

$$\mathbf{y}' = \zeta \left(1 - \frac{1}{1 + \exp(-a(\mathbf{x}' - c))} \right) (\sin(2\pi f_h \mathbf{t}) + 1), \quad (4.5)$$

where $\mathbf{t} = \{0, dt, 2dt, \dots, (N-1)dt\}$, $\zeta = \sqrt{f_l/f_h}$, $c = 0.6$, and $a = 10$ (see Figure 4.2C). The constant c represents a threshold value, so that when $\mathbf{x}' < c$ the amplitude of \mathbf{y}' increases; a controls the steepness of that increase. Next, we impose a directionality of interaction from \mathbf{x}' to \mathbf{y}' by time-shifting \mathbf{y}' by Δt seconds $\mathbf{y}'_{\Delta t} = \mathbf{y}'(t + \Delta t)$. Afterward, we construct two pairs of auxiliary signals following the steps described above. From one pair, we select the signal with low frequency components, \mathbf{x}'' . From the remaining pair, we select the signal oscillating at the highest frequency, \mathbf{y}'' . Finally, we define $\mathbf{x} = \mathbf{x}' + \mathbf{y}''$, and $\mathbf{y} = \mathbf{x}'' + \mathbf{y}'_{\Delta t}$, so that, by design, \mathbf{x} and \mathbf{y} have closely resembling power spectra (see Figures 4.2D-4.2F).

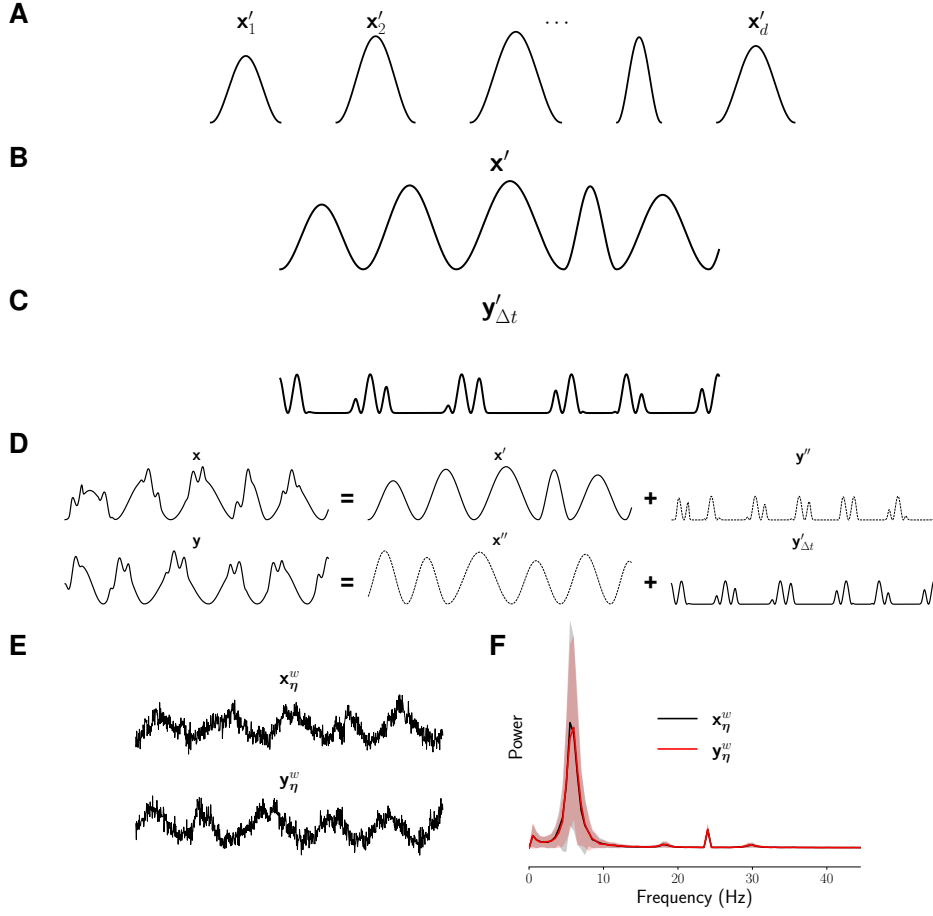


Figure 4.2: Schematic representation of the simulation model used to obtain data with directed phase-amplitude interactions. (A) Sinusoidal signal segments of varying amplitude and period (x'_i). (B) A continuous, low frequency signal (x') is obtained after concatenating the sinusoidal signal segments. (C) A high frequency signal (y') modulated by x' is computed through Equation 4.5. Then y' is time-shifted by Δt seconds ($y'_{\Delta t}$). (D) x' and $y'_{\Delta t}$ are combined with a pair of auxiliary, non-interacting signals x'' and y'' to generate signals x and y . (E) Signals x and y are contaminated with noise and linearly mixed. (F) The resulting signals, x_{η}^w and y_{η}^w , have very similar power spectra.

Experimental setup

We simulate 100 pairs (trials) of 2 s long signals, sampled at 1000 Hz, with directed phase-amplitude interactions from the θ band ($f_l = 6$ Hz) to the β band ($f_h = 24$ Hz), with a time shift of 20 ms. We detrend and normalized the simulated signals, before contaminating them with normalized white (η) and pink (η_p) noise, as follows:

$$\mathbf{x}_{\eta} = \mathbf{x} + 10^{-\frac{\text{SNR}}{20}} (0.6\eta^{\mathbf{x}} + 0.4\eta_p^{\mathbf{x}}),$$

$$\mathbf{y}_\eta = \mathbf{y} + 10^{-\frac{\text{SNR}}{20}} (0.6\boldsymbol{\eta}^y + 0.4\boldsymbol{\eta}_p^y),$$

where the parameter SNR controls the signal to noise ratio. We vary it to simulate low (SNR = 6), moderate (SNR = 3), and high (SNR = 1) noise conditions. For each scenario, we also mix the noisy signals, \mathbf{x}_η and \mathbf{y}_η , aiming to reproduce the effects of volume conduction, by defining $\mathbf{x}_\eta^w = \left(1 - \frac{w}{2}\right) \mathbf{x}_\eta + \left(\frac{w}{2}\right) \mathbf{y}_\eta$, and $\mathbf{y}_\eta^w = \left(1 - \frac{w}{2}\right) \mathbf{y}_\eta + \left(\frac{w}{2}\right) \mathbf{x}_\eta$, with $w = 0.25$ the mixing strength. Then, we downsample \mathbf{x}_η^w and \mathbf{y}_η^w to 250 Hz, and estimate the directed phase-amplitude interactions present in the data for a square frequency grid ranging from 3 Hz to 45 Hz, in 3 Hz steps, using the three approaches described in Section 4.1. Finally, in order to determine whether the estimated interactions are statistically significant, we perform permutation tests based on randomized surrogate trials (Lindner et al., 2011; Weber et al., 2017) for each condition and frequency pair evaluated (see Appendix C). The Bonferroni-corrected significance level for the tests is set to 4.4×10^{-5} .

4.2.2 EEG data

As in previous chapters, our aim is to set up subject-dependent classification systems (one classifier per subject) that allow discriminating between the different conditions in the MI and WM paradigms. Here we use as inputs relevant directed phase-amplitude interactions captured through the proposed $\text{TE}_{\kappa\alpha}^{\theta\theta\varsigma}$ approach, as well as through the CFD, and the conventional methodology to estimate directed PAC through TE, implemented using our kernel-based approach ($\text{TE}_{\kappa\alpha}^{\theta\varsigma}$).

Classification Setup

Feature Extraction Let $\Psi = \{\mathbf{X}_n \in \mathbb{R}^{C \times M}\}_{n=1}^N$ be an EEG set holding N trials from either an MI or a WM dataset, recorded from a single subject, where C stands for the number of channels and M corresponds to the number of samples. For MI, $C = 22$ and $M = 500$ (we only select a 2 s long time window stretching from second 3 to second 5), while for WM, $C = 32$ and $M = 717$. In addition, let $\{l_n\}_{n=1}^N$ be a set whose n -th element is the label associated with trial \mathbf{X}_n . For the MI database l_n can take the values of 1 and 2, corresponding to right hand and left hand motor imagination, respectively. Similarly, for the WM database, l_n can take the values of 1, 2, and 3 corresponding to low, medium, and high memory loads. We want predict the label l_n from TE or CFD features that capture the directed phase-amplitude interactions present in \mathbf{X}_n .

Let $\lambda(\mathbf{x}_c \rightarrow \mathbf{x}_{c'}, f_c, f_{c'})$ be a directed connectivity measure between the phase of channel \mathbf{x}_c at frequency f_c , and the amplitude envelope of $\mathbf{x}_{c'}$ at frequency $f_{c'}$, as defined in Equation 4.3. For all pairwise combinations of channels in \mathbf{X}_n , computing $\lambda(\mathbf{x}_c \rightarrow \mathbf{x}_{c'}, f_c, f_{c'})$ yields a connectivity matrix $\mathbf{\Lambda}(f_c, f_{c'}) \in \mathbb{R}^{C \times C}$. When $c = c'$, then $\lambda(\mathbf{x}_c \rightarrow \mathbf{x}_{c'}, f_c, f_{c'}) = 0$. The values of $f_c, f_{c'}$ vary in the range from 4 Hz to 18 Hz, in 2 Hz steps for both MI and WM data (Collazos-Huertas et al., 2020; Johnson et al., 2018). Note that for the MI data this time we have also included activity in the 4 Hz to 6 Hz range, because it is commonly involved in PAC interactions (Aru et al., 2015). Then, three bandwidths are defined $\Delta f \in \{\theta \in [4 - 6], \alpha \in [8 - 12], \beta_l \in [14 - 18]\}$ Hz, and the matrices $\mathbf{\Lambda}(f_c, f_{c'})$ are averaged within each pairwise combination of bandwidths ($\theta - \alpha$, $\theta - \beta_l$, and so on). After that, each of the averaged matrices are normalized to the range $[0, 1]$, and stacked together. Therefore, each trial is characterized by a connectivity matrix $\mathbf{\Lambda}' \in \mathbb{R}^{C \times C \times 6}$. For the N trials, there is a set of connectivity matrices $\{\mathbf{\Lambda}'_n \in \mathbb{R}^{C \times C \times 6}\}_{n=1}^N$. Then, applying vector concatenation to $\mathbf{\Lambda}'_n$ a vector $\phi_n \in \mathbb{R}^{1 \times (C \times C \times 6)}$ is obtained. Finally, the N vectors ϕ_n are stacked together in order to obtain a single bi-dimensional matrix $\Phi \in \mathbb{R}^{N \times P}$, $P = C \times C \times 6$, that characterizes Ψ in terms of directed phase-amplitude measures.

Feature Selection and Classification After characterizing the EEG data through the directed PAC estimation methods considered, we set up subject-dependent classification systems for the MI and WM databases following the same steps described in Section 2.2.3 for each database.

4.2.3 Parameter selection

For the TE methods, all parameters were estimated before extracting the phase and amplitude time series, that is to say, from the initial real-valued time series data. All TE parameters were selected as detailed in Section 3.2.3, except the parameter u for the simulated data, which was chosen from the range $u = \{4, 8, \dots, 40\}$ as the delay producing the largest TE. For CFD estimation, we used a sliding window 5 frequency bins long. Furthermore, for all measures the required phase and amplitude decompositions were carried out by convolving the real-valued data with a Morlet wavelet (see Equation 3.2). The parameter m in the wavelet was also tuned as described in Section 3.2.3. Finally, all connectivity values were obtained through in-house implementations of the algorithms for the different measures studied.

4.3 Results and discussion

4.3.1 Simulated phase-amplitude interactions

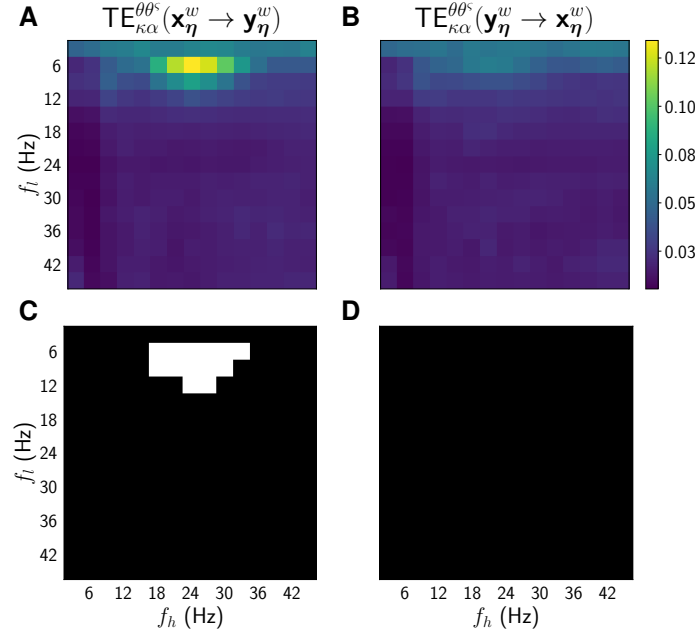


Figure 4.3: $TE_{\kappa\alpha}^{\theta\theta^s}$ results for the simulated data in the case when SNR = 3. (A) Average $TE_{\kappa\alpha}^{\theta\theta^s}(\mathbf{x}_\eta^w \rightarrow \mathbf{y}_\eta^w)$ values. (B) average $TE_{\kappa\alpha}^{\theta\theta^s}(\mathbf{y}_\eta^w \rightarrow \mathbf{x}_\eta^w)$ values. (C) Results of the permutation test performed on the $TE_{\kappa\alpha}^{\theta\theta^s}(\mathbf{x}_\eta^w \rightarrow \mathbf{y}_\eta^w)$ values. Statistically significant connectivities are indicated in white. (D) Results of the permutation test performed on the $TE_{\kappa\alpha}^{\theta\theta^s}(\mathbf{y}_\eta^w \rightarrow \mathbf{x}_\eta^w)$ values.

Figure 4.3 presents the results obtained through the proposed $TE_{\kappa\alpha}^{\theta\theta^s}$ approach for the simulated data in the case when SNR = 3. Figure 4.3A shows the average values obtained for $TE_{\kappa\alpha}^{\theta\theta^s}(\mathbf{x}_\eta^w \rightarrow \mathbf{y}_\eta^w, f_l, f_h)$, with f_l and f_h varying in the range from 3 Hz to 45 Hz, in 3 Hz steps. That is to say, it shows TE values computed following our proposal assuming that the phases of the oscillations in \mathbf{x}_η^w drive the amplitude envelopes of the oscillations in \mathbf{y}_η^w . Likewise, Figure 4.3B shows the obtained average $TE_{\kappa\alpha}^{\theta\theta^s}(\mathbf{y}_\eta^w \rightarrow \mathbf{x}_\eta^w, f_l, f_h)$ values, where the underlying assumption is that phases of the oscillations in \mathbf{y}_η^w are causal to the amplitude envelopes of the oscillations in \mathbf{x}_η^w . Figures 4.3C and 4.3D display the results returned by the permutation tests carried out over the TE data estimated for all the simulated trials, and whose average values are displayed in Figures 4.3A and 4.3B, respectively (statistically significant results are indicated in white). The results show that the proposed approach captures strong and statistically significant phase-amplitude

directed interactions around the target frequencies used to generate the simulated data. Also, statistically significant results were obtained only when TE was estimated from the phase of \mathbf{x}_η^w in the θ band, at around 6 Hz, to the amplitude of \mathbf{y}_η^w in the β band, at around 24 Hz, and not when estimated assuming causality in the opposite direction.

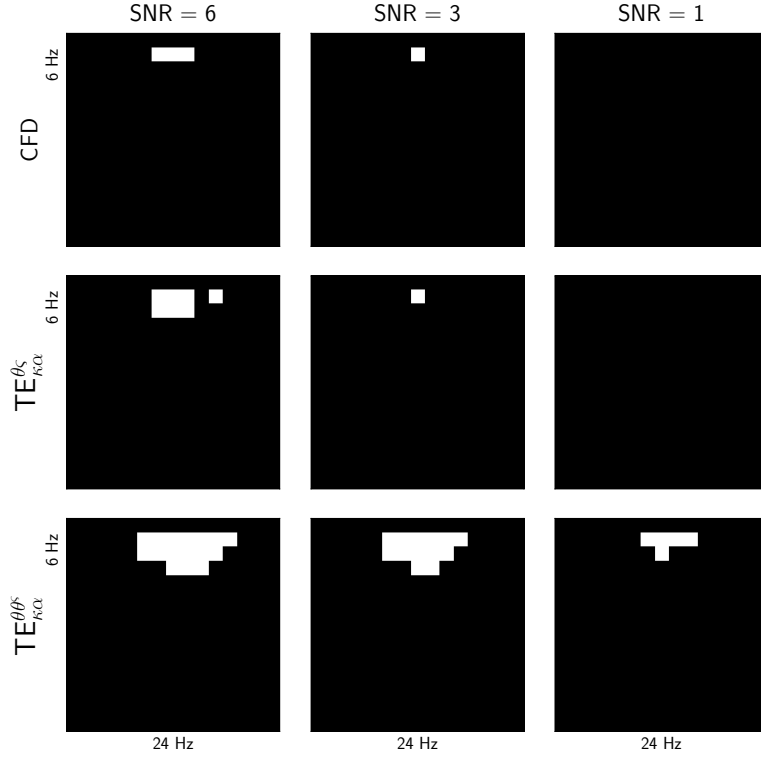


Figure 4.4: Results of the permutation tests carried out on the connectivity values estimated using CFD, $\text{TE}_{\kappa\alpha}^{\theta\zeta}$ and $\text{TE}_{\kappa\alpha}^{\theta\theta\zeta}$, for the three noise conditions modeled ($\text{SNR} = 6, 3, 1$), assuming interactions in the simulated data from the phase of \mathbf{x}_η^w to the amplitude of \mathbf{y}_η^w . Statistically significant results are shown in white.

Figure 4.4 shows the results of the permutation tests carried out on the connectivity values estimated for all trials from the simulated data using, from top to bottom, CFD, $\text{TE}_{\kappa\alpha}^{\theta\zeta}$ (conventional approach to estimate directed phase-amplitude through TE, using our kernel-based TE estimator), and $\text{TE}_{\kappa\alpha}^{\theta\theta\zeta}$, under the three noise conditions modeled, from left to right, low ($\text{SNR} = 6$), moderate ($\text{SNR} = 3$) and high ($\text{SNR} = 1$) noise levels. In this case, all connectivity measures were obtained assuming the correct direction of causality, from \mathbf{x}_η^w to \mathbf{y}_η^w . As before, statistically significant results are displayed in white. For the low and moderate noise levels the three connectivity estimation methods successfully capture statistically significant phase-amplitude directed interactions around the target frequencies. Note that the CFD and $\text{TE}_{\kappa\alpha}^{\theta\zeta}$ display significant results on narrower

frequency ranges around the interacting frequencies actually present in the data than $TE_{\kappa\alpha}^{\theta\theta\varsigma}$. This is likely to be associated with the additional filtering stages involved in the computation of $TE_{\kappa\alpha}^{\theta\theta\varsigma}$, and in the case of CFD, its high frequency specificity is probably linked to its robustness against false positives (Jiang et al., 2015). However, unlike $TE_{\kappa\alpha}^{\theta\theta\varsigma}$, both CFD and $TE_{\kappa\alpha}^{\theta\varsigma}$ failed to capture any statistically significant interactions for the high noise scenario. Therefore, our proposal exhibits higher robustness to the noise present in the data. Thus, from the results presented in Figures 4.3 and 4.4, we can argue that the proposed $TE_{\kappa\alpha}^{\theta\theta\varsigma}$ approach allows uncovering directed phase-amplitude interactions, capturing their strength and direction, even in the presence of high levels of noise and a confounding factor such as signal mixing due to volume conduction.

4.3.2 EEG data

Motor imagery

Table 4.1 presents the average classification accuracies obtained for all subjects in the MI database, after characterizing the EEG data through the three directed PAC estimation methods considered. Our results show that none of the approaches tested coded discriminant connectivity features for the right and left MI tasks. Having in mind that better classification results were obtained through the proposed $TE_{\kappa\alpha}$ and $TE_{\kappa\alpha}^{\theta}$ approaches, that we have used the same feature selection and classification strategies for all the MI-related experiments throughout this dissertation, and that the accuracies displayed in Table 4.1 are in the range of what would be expected from random guessing, we can argue that directed phase amplitude interactions for oscillations between 4 Hz and 20 Hz do not carry discriminant information for the MI tasks of interest. It is worth noting that we also carried out tests disregarding any interactions involving the θ band, so as to focus only on the α and lower β bands, which the literature associates to MI (Collazos-Huertas et al., 2020), but the obtained results were equally poor.

Table 4.1: Average classification accuracies for all subjects in the MI database.

	Cross-Validation	Testing
CFD	54.3 ± 3.1	51.0 ± 3.5
$TE_{\kappa\alpha}^{\theta\varsigma}$	56.8 ± 4.5	51.7 ± 2.5
$TE_{\kappa\alpha}^{\theta\theta\varsigma}$	56.7 ± 10.2	50.4 ± 4.0

Working memory

Phase-amplitude interactions are thought to play a crucial role in WM (Liang et al., 2021), with current hypothesis pointing to bidirectional interactions between the θ , α , and β bands (particularly around 13.5 Hz to 16 Hz for the latter), linking the prefrontal cortex to parieto-occipital and medial temporal regions during the activation of WM (Johnson et al., 2018, 2019). In this chapter, we built subject-dependent classification systems based on directed PAC features obtained from EEG data recorded during the retention interval of a visuospatial working memory task. The goal of the classifiers is to assign a particular cognitive load (number of elements in the memory array) to a trial characterized through directed PAC.

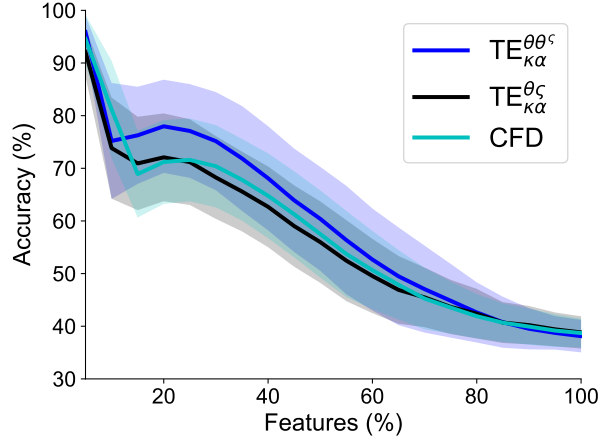


Figure 4.5: Average classification accuracy for all subjects as a function of the percentage of selected features.

Figure 4.5 presents the classification accuracy for all subjects in the WM database as a function of the percentage of features used to train the classifiers. The best results for all the directed PAC estimation approaches tested are obtained when 5% of the features are selected, those with the highest relevance values according to the CKA-based relevance vector $\bar{\mathbf{q}}$ (see Appendix D). The $\text{TE}_{\kappa\alpha}^{\theta\theta^c}$ -based classifier achieves an average classification accuracy of $95.9 \pm 3.1\%$ for the three classes corresponding to each cognitive load level in the change detection task, while the $\text{TE}_{\kappa\alpha}^{\theta^c}$ and CFD-based classifiers achieve accuracies of $94.6 \pm 4.6\%$ and $92.1 \pm 4.7\%$, respectively.

Figure 4.6 shows the highest accuracy obtained for each subject in the WM database. Although the subjects differ in performance, for all of them the proposed classification systems exhibit accuracies well above what would be expected from chance in a three

class classification task. These results show that our $TE_{\kappa\alpha}^{\theta\theta^c}$ characterization strategy, as well as the other approaches tested, successfully captured discriminant directed phase-amplitude interactions that are elicited during the change detection task. Moreover, only a small fraction of those interactions are required to discriminate between the task's levels. In fact, employing a higher percentage of features led to a pronounced decrease in classification performance, as shown in Figure 4.5. This phenomenon (also observed in Figure 3.8) can be explained because all to all channel connectivity analyses lead to datasets with a large number of features, which in turn leads to a well-known problem in machine learning: the curse of dimensionality. The larger the dimensionality of the data the higher the chance that most training instances are far away from each other, and that new instances will also be far away from those used to train the machine learning system, which makes it difficult to make good predictions (Géron, 2019). Besides, many of the obtained connectivity features may not provide useful information to discriminate between the conditions of the cognitive paradigm of interest (De La Pava Panche et al., 2019), only adding noise and complexity to the classification stage. Adequate feature selection reduces the dimensionality of the data by getting rid of non-relevant features (connectivity values), which can, as in our case, help classification performance.

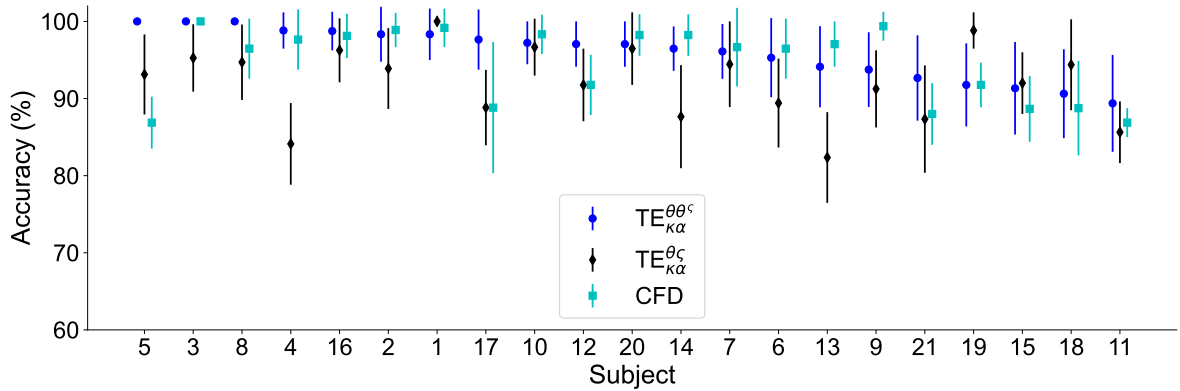


Figure 4.6: Highest average classification accuracy for each subject in the WM database. The subjects are ordered from highest to lowest performance, according to the results obtained for the $TE_{\kappa\alpha}^{\theta\theta^c}$ -based classifier.

At this point, it is worth stating once again that the vector $\bar{\varrho}$, used to rank the features, is obtained as the average of the relevance vectors stemming from the different folds of an auxiliary cross-validation scheme, outside of the cross-validation process where classification is carried out. Therefore, our classification setup suffers from data leakage, and, as a consequence, the above-described results are probably higher than what could

be obtained from a classification system suitable for real world applications. Nonetheless, our results still show that relatively few connectivity values describing directed phase-amplitude interactions suffice to successfully discriminate among brain activity elicited during the different levels of the change detection task. Furthermore, the vector $\bar{\varrho}$ in itself provides valuable insights regarding the directed phase-amplitude interactions that arise while the subjects perform the task. It does so by assigning a relevance value ϱ_p to each column, or feature, in the characterizing matrix Φ . In the case of our $\text{TE}_{\kappa\alpha}^{\theta\theta^*}$ analysis, ϱ_p indicates whether a particular connectivity between two channels, for a specific frequency band pair ($\theta - \alpha$, $\theta - \beta_l$, etc), allows discriminating between the different cognitive loads of the task. A feature's discriminant capability is tantamount to its variation across classes. That is to say, the most relevant connectivities in $\bar{\varrho}$ are those that change consistently across trials as a function of the cognitive load, which points to the involvement of those directed phase-amplitude interactions in the underlying working memory systems being activated during the task.

Figure 4.7 displays the most relevant connectivities, on average, according to the relevance vector $\bar{\varrho}$, discriminated by frequency band pair. The average was taken over all subjects after a t-SNE-based analysis, analogous to that of Section 3.3.2, did not reveal any evident relevance clusters. The background topoplots show the average nodal relevance, which corresponds to the relevance of the total information flow of every node. That is to say, the sum of all the ϱ_p in $\bar{\varrho}$ associated to all directed interactions originating an targeting a specific node or EEG channel. We note that, in general, there is high nodal relevance for interactions where the phases of oscillations in the θ band drive the amplitudes of oscillations in the α and β_l bands. However, the highest nodal relevance is achieved by the phase-amplitude interactions from α to β_l band activity. As expected, the relevant connectivities are distributed in a similar fashion frequency-wise. Spatially, they tend to involve long range interactions linking frontal, temporal and parietal regions. In particular, for the phase-amplitude interactions from α to β_l , there are multiple bidirectional connections between channels on frontal and prefrontal areas and channels on parietal and parieto-occipital areas. Also there are several long range relevant connections targeting the right temporal region. These results coincide with those we obtained in Chapter 3 from within-frequency phase-based connectivity values, as well as with studies that identified the presence of fronto-parietal and fronto-temporal interactions during cognitive task that activate visuospatial working memory (Dimitriadis et al., 2016a; Johnson et al., 2018, 2019).

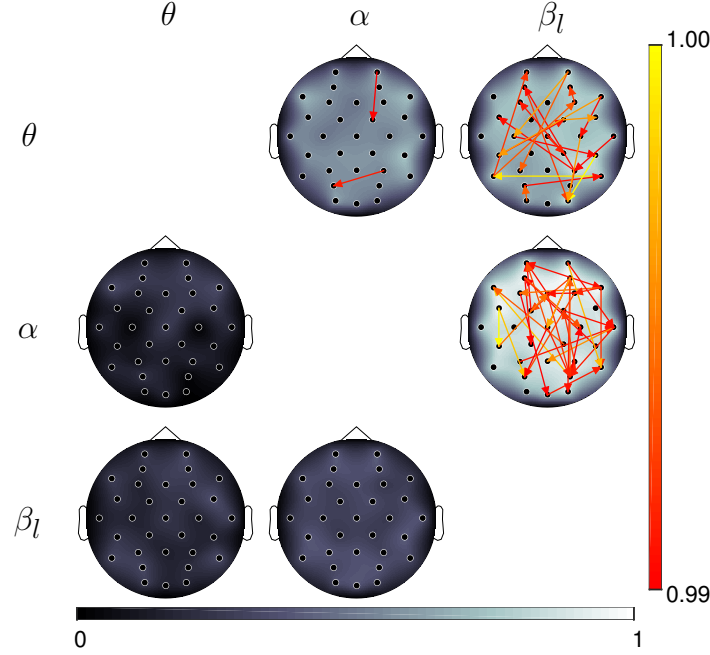


Figure 4.7: Topoplots of the average nodal (channel) relevance for each frequency band pair tested. The arrows represent the most relevant connections. For visualization purposes, only the connections with the highest average relevance values are depicted (1% of all connections). The rows indicate the frequency band driving the phase-amplitude interactions, while the columns correspond to the frequency band of the driven oscillations (e.g. the topoplot at the top right corner shows the nodal relevance, and the relevant connections, obtained when $TE_{\kappa\alpha}^{\theta\theta^s}$ was computed assuming directed interactions from the phases of oscillations in the θ band to the amplitude envelope of oscillations in the β_l band).

4.3.3 Limitations

The results obtained for the simulated data show that our proposal is, unlike the other approaches tested, able to detect statistically significant directed phase-amplitude interactions in data with high levels of noise. However, it tends to be less frequency specific than the CFD and the conventional approach for the estimation of directed phase-amplitude interactions through TE. Therefore, our method is well suited for the analysis of noisy data, but care should be taken with results under more ideal conditions. Also, the proposed approach relies on the kernel-based phase TE estimation method introduced in Chapter 3, and therefore, it suffers from the same limitations, especially regarding the selection of the many parameters involved in TE computation. Here, we employed relatively simple parameter selection strategies, but it is possible that more elaborate approaches (Zhou et al., 2018) may lead to better results. Additionally,

our kernel-based TE estimator assumes stationary or weakly non-stationary data, and cannot tell apart direct interactions from those originating from unobserved common causes (De La Pava Panche et al., 2019). Finally, our proposal is strictly limited to the detection of directed phase-amplitude interactions. It is unable to capture any of the other types of cross-frequency interactions that arise in oscillatory neural activity (phase-phase coupling, phase-frequency coupling, etc) (La Tour et al., 2017). It does not intend to be a full information-theoretic analysis that takes into account simultaneously the multiple rhythms that can interact at both the sender and receiver side of coupled dynamical systems (Pinzuti et al., 2020).

4.4 Summary

In this chapter, we proposed a novel approach to estimate directed phase-amplitude interactions through TE. The central idea behind our proposal is to recast the problem of detecting directed PAC as that of estimating directed interactions between phase time series. Doing so allowed us to employ a single-trial kernel-based phase TE estimator to assess the cross-frequency interactions of interest. We tested the performance of our proposal on synthetic data containing directed phase-amplitude interactions and on EEG databases obtained under MI and WM paradigms. The obtained results for the synthetic data showed that our approach successfully detected the direction of interaction and the interacting frequencies while being more robust to noise than alternative methods. Additionally, for the WM data, our proposal revealed discriminant directed phase-amplitude interactions associated with the different cognitive loads of the task. However, it did not capture discriminant phase-amplitude interactions for the MI data.

Chapter 5

Final Remarks

5.1 Conclusions

In this dissertation we have shown that the TE between pairs of time series can be successfully estimated through a data-driven approach based on a kernel formulation of Renyi's α entropy, thus avoiding the need to obtain the probability distribution of the data. Directed phase to phase within frequency interactions, and directed phase to amplitude cross-frequency couplings can also be satisfactorily assessed through such an approach, provided the data has previously undergone phase-amplitude decomposition at the specific frequency or frequencies of interest. Furthermore, the TE estimation proposals developed in this document allow uncovering directed bivariate interactions among data degraded by factors such as the presence of noise or signal mixing, as evidenced by the attained results for simulated data, which are challenging problems for successful connectivity estimation commonly encountered in electrophysiological signals. Indeed, our proposals coded, with varying performance levels, discriminant connectivity patterns for real EEG data obtained under motor imagery and visuospatial working memory paradigms. All the above points to the viability of using our kernel-based approach to TE estimation as a tool to characterize and gain insights from EEG data obtained under different cognitive tasks. In the following, we support these overall conclusions by highlighting the main results from our work:

- We introduced a new approach to estimate TE between pairs of time series. We arrived at the proposed kernel-based Renyi's α transfer entropy ($TE_{\kappa\alpha}$) estimator by expressing TE as a linear combination of joint and marginal Renyi entropies, and then replacing the elements of such linear combination by functionals defined

- on positive definite kernel matrices. Our matrix-based formulation of TE bypasses the need to explicitly obtain the probability distribution of the data. We tested our approach using synthetic time series generated through two simulation models: a vector-autoregressive model that produces data with known directed interactions, and a linear coupling function that models a network of several interacting channels. We also tested the ability of our proposal to estimate discriminant directed connections from EEG data obtained under MI and WM paradigms. In that sense, we aimed to classify a data trial, characterized through $TE_{\kappa\alpha}$, as belonging to the different classes in each database through subject-dependent classification systems coupled with kernel-based relevance analysis. For the synthetic data, results showed that our proposal successfully detected the direction of the interactions in the data, that it was robust to moderate levels of noise, and that it worked well for different data sizes. Also, when coupled with a trial-based permutation test, it revealed the presence, or absence, of directed interactions in a connected network with multiple time delays. For the MI data, the proposed estimator coded discriminant interactions for a subgroup of subjects that allowed separating left hand from right hand trials, with the discriminant interactions appearing on the appropriate time windows. Results for the WM data showed long range discriminant connections for the three different cognitive load levels in the data, linking mostly frontal and parieto-occipital regions. Also, discriminant shorter range connections among frontal and pre-frontal channels were identified.
- We proposed a methodology to obtain phase TE estimates from single trial data, a task for which commonly used TE estimators are ill-suited. Our approach, termed kernel-based Renyi's phase transfer entropy ($TE_{\kappa\alpha}^\theta$), combines phase time series, obtained from the data of interest through complex filtering, with our $TE_{\kappa\alpha}$ estimator. We tested our proposal using simulated data from NMMs, which allow generating time series with properties that resemble the oscillatory dynamics of electrophysiological signals. We used the NMM data containing directed interactions at known frequencies to explore the response of our proposal under scenarios of varying coupling strengths, of noise and signal mixing, and in terms of its ability to detect narrow-band bidirectional interactions. Like for $TE_{\kappa\alpha}$, we also tested it on the MI and WM EEG databases. Obtained results for the simulated data showed that the proposed $TE_{\kappa\alpha}^\theta$ approach captured statistically significant phase-based interactions at the frequencies of interest for all the scenarios considered. In the same experiments other phase TE estimation approaches failed to achieve positive

results. For the MI data, the $TE_{\kappa\alpha}^\theta$ -based analysis provided results similar to those of $TE_{\kappa\alpha}$, regarding the obtained classification performances and the distribution of the connectivities relevant to discriminate between the MI tasks, additionally showing that such interactions arise in the α and β_l bands. Finally for the WM data, characterizing the signals through $TE_{\kappa\alpha}^\theta$ led to improved classification accuracies, as compared with $TE_{\kappa\alpha}$, with discriminant directed connectivities found among frontal, parietal, and temporal regions, coded in the phases of oscillatory activity in the θ , α and β_l bands.

- We put forward a methodology that estimates directed phase-amplitude interactions through kernel-based Renyi's phase TE ($TE_{\kappa\alpha}^{\theta\theta^\varsigma}$). The key of our proposal is recasting the problem of computing TE for phase-amplitude interactions as a phase TE estimation problem, by extracting the phase of the target amplitude time series at a specific frequency, exploiting the underlying idea of the CFD (Jiang et al., 2015). In order to test the proposed approach, we employed synthetic data obtained from a model that simulates directed phase-amplitude interactions at specific frequencies. Furthermore, we also tested it on the MI and WM EEG databases, following the same strategy described for the $TE_{\kappa\alpha}$ method. According to the attained results, our proposal correctly detected the direction of phase-amplitude interactions, with statistically significant connectivity values at the frequencies of interest. Also, it was more robust to noise than the alternative approaches tested. Regarding the results for the MI data, the $TE_{\kappa\alpha}^{\theta\theta^\varsigma}$ -based characterization did not capture discriminant interactions, and neither did the comparison methods considered. On the other hand, for the WM data, the classifiers based on $TE_{\kappa\alpha}^{\theta\theta^\varsigma}$ features performed on par with those trained using features extracted through $TE_{\kappa\alpha}^\theta$. This behavior was also observed for the other directed PAC estimation methods explored. Additionally, the relevance analysis carried out on the $TE_{\kappa\alpha}^{\theta\theta^\varsigma}$ features revealed discriminant interactions for the WM data involving, in general, directed connections among the same regions mentioned in the items above, with the phases of oscillations in the θ band driving the amplitudes of oscillations in the α and β_l bands, and specially, phase-amplitude interactions from α to β_l band activity.

5.2 Future work

We have presented a data-driven framework to TE estimation that allows revealing directed interactions in the time and frequency domains, while sidestepping the need for probability distribution estimation. However, as explained in the limitation sections of Chapters 2 to 4, there are still many issues that can be addressed to improve the performance of our proposals, or to build upon them. In particular, the following aspects could be of interest for future work:

- As pointed out in Section 2.3.4, bivariate TE is, by definition, blind to the effects of third variables, which leads to issues such as the different delay driving problem. Given three time series $\mathbf{x} = \{x_t\}_{t=1}^T$, $\mathbf{y} = \{y_t\}_{t=1}^T$, and $\mathbf{z} = \{z_t\}_{t=1}^T$, with $t \in \mathbb{N}$ a discrete time index, and $T \in \mathbb{N}$, an approach to account for the influence of \mathbf{z} in the information transfer from \mathbf{x} to \mathbf{y} is to condition TE on \mathbf{z} ,

$$TE(\mathbf{x} \rightarrow \mathbf{y} | \mathbf{z}) = \sum_{y_{t+1}, \mathbf{y}_t^m, \mathbf{x}_t^n, \mathbf{z}_t^o} p(y_{t+1}, \mathbf{y}_t^m, \mathbf{x}_t^n, \mathbf{z}_t^o) \log \left(\frac{p(y_{t+1} | \mathbf{y}_t^m, \mathbf{x}_t^n, \mathbf{z}_t^o)}{p(y_{t+1} | \mathbf{y}_t^m, \mathbf{z}_t^o)} \right). \quad (5.1)$$

Extending 5.1 to more possibly interacting systems (and their corresponding sampled time series) is theoretically possible under stationarity assumptions (Montalto et al., 2014). However, in practice, the growing dimensionality of the data as more time series are included poses an estimation challenge. It would be interesting to explore the performance of our kernel-based TE estimator in the above-described multivariate setting.

- One of the downsides of TE estimation approaches, including ours, is the need to adjust multiple parameters. In this thesis, we chose to follow simple and commonly-used parameter estimation strategies. It is possible that the performance of our proposals could be improved by a more careful parameter selection process. The same logic applies to the selection of the complex-filters in $TE_{\kappa\alpha}^\theta$ and $TE_{\kappa\alpha}^{\theta\theta^s}$. It would be specially useful if such selection procedures could also be automatized.
- In Chapter 4, we reframed the problem of estimating directed phase-amplitude interactions through TE as a phase TE problem, under the hypothesis that avoiding the estimation of TE from time series with markedly different behaviors (phase and amplitude time series), could be beneficial. However, our kernel-based formulation of TE allows for another possible solution to that problem: using different kernels for different types of input data. One could, for instance, use a Gaussian RBF

kernel for the terms involving the amplitude time series, and a periodic kernel, such as

$$\kappa(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) = \exp \left(-\frac{2\sin^2(\pi d^2(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)/p)}{l^2} \right), \quad (5.2)$$

where $d(\cdot, \cdot)$ stands for the Euclidean distance operator, $p, l \in \mathbb{R}^+$ are the kernel periodicity and length scale, respectively; for the phase time series.

- Although clinical application questions were not raised in this work, the proposed approaches and the obtained results could have clinical applicability in experimental models with patients affected by specific mental disorders. For instance, they could be used to assess connectivity changes for different cognitive load levels in working memory tasks in patients diagnosed with schizophrenia, since alterations in such cognitive process are well documented in this population group (Karls-godt et al., 2007; Schlösser et al., 2003). One could hypothesize that patients with schizophrenia will exhibit higher cognitive loads, and their associated connectivity changes, with low-complexity working memory tasks compared to a control group. Alternatively, an absence of objective changes in connectivity could be observed regardless of the degree of difficulty of the working memory task to which the patient is subjected. A third possible hypothesis would be to evaluate dynamic changes in connectivity, obtained at different clinical moments to assess whether there is any correlation with acute clinical decompensation or with the severity of symptoms in general. More generally, one could hypothesize that our approaches to TE estimation could serve as a biomarkers of the clinical status of patients with mental illness who have alterations in working memory, particularly schizophrenia, but also in other pathologies such as attention deficit disorder with or without hyperactivity (Roman-Urrestarazu et al., 2016), depressive episodes of moderate to severe intensity (Wang et al., 2021b), or bipolar disorders (Townsend et al., 2010).

5.3 Academic products

5.3.1 Journal papers

1. De La Pava Panche, I., Gómez-Orozco, V., Álvarez-Meza, A., Cárdenas-Peña, D., & Orozco-Gutiérrez, Á. (2021). *Estimating Directed Phase-Amplitude Interactions from EEG Data through Kernel-Based Phase Transfer Entropy*. Applied Sciences, 11(21), 9803. (Q2-A1).

2. De La Pava Panche, I., Álvarez-Meza, A., Herrera Gómez, P. M., Cárdenas-Peña, D., Ríos Patiño, J. I., & Orozco-Gutiérrez, Á. (2021). *Kernel-Based Phase Transfer Entropy with Enhanced Feature Relevance Analysis for Brain Computer Interfaces*. Applied Sciences, 11(15), 6689. (Q2-A1).
3. De La Pava Panche, I., Alvarez-Meza, A. M., & Orozco-Gutierrez, A. (2019). *A data-driven measure of effective connectivity based on Renyi's α -entropy*. Frontiers in neuroscience, 13, 1277. (Q1-A1).

5.3.2 Conference papers

The conference papers listed below are not directly related to the contents of this thesis. However, they are the result of the student's first steps into the field of EEG analysis techniques in general, and brain connectivity measures in particular, as part of his doctoral studies.

1. De La Pava, I., Álvarez, A., Herrera, P., Castellanos-Dominguez, G., & Orozco, A. (2018, November). *Gender Effects on an EEG-Based Emotion Level Classification System*. In Iberoamerican Congress on Pattern Recognition (pp. 810-819). Springer, Cham. (International conference).
2. De La Pava, I., Álvarez-Meza, A., & Orozco, A. A. (2017, November). *Emotion assessment by variability-based ranking of coherence features from EEG*. In Iberoamerican Congress on Pattern Recognition (pp. 203-211). Springer, Cham. (International Conference)

5.3.3 Conference presentations

1. An oral presentation titled "*A Novel phase-based transfer entropy estimator from a kernel approach*", given at the 14th International Conference on Brain Informatics (BI2021), which was held virtually from the 17th to the 19th of September, 2021.

Appendix A

Kernel methods and Renyi's α entropy estimation

The contents of this appendix are largely based on references (Liu et al., 2011) and (Giraldo et al., 2015).

A.1 Reproducing kernel Hilbert spaces

A Hilbert space is a linear, complete, and normed space endowed with an inner product. A reproducing kernel Hilbert space (RKHS) is a special Hilbert space associated with a kernel κ such that it reproduces (via an inner product) each function f in the space (Liu et al., 2011). The RKHS framework is commonly applied in machine learning (Géron, 2019), where it provides a way of simplifying the computation of kernel-based methods thanks to the "kernel trick", and also in information-theoretic learning (ITL), where kernel-based probability density estimation has a central role in obtaining information-theoretic quantities such as entropy.

A.1.1 Reproducing kernels

Let \mathcal{H}_κ be a Hilbert space of real-valued functions defined on a set \mathcal{X} , equipped with an inner product $\langle \cdot, \cdot \rangle$ and a real-valued bivariate function $\kappa(x_i, x_j)$ on $\mathcal{X} \times \mathcal{X}$. Then the function $\kappa(x_j, x_j)$ is said to be nonnegative definite (positive semi-definite) if for any finite point set $\{x_1, x_2, \dots, x_n\} \subset \mathcal{X}$ and for any not all zero corresponding real numbers

$$\{\alpha_1, \alpha_2, \dots, \alpha_n\} \subset \mathbb{R},$$

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \kappa(x_i, x_j) \geq 0. \quad (\text{A.1})$$

Because of the following theorem, any nonnegative definite bivariate function $\kappa(x_i, x_j)$ is a reproducing kernel.

Moore-Aronszajn's theorem *Given any nonnegative definite function $\kappa(x_i, x_j)$, there exists a uniquely determined (possibly infinite-dimensional) Hilbert space \mathcal{H}_κ consisting of functions on \mathcal{X} such that*

$$\begin{aligned} \text{(I)} \quad & \forall x \in \mathcal{X}, \quad \kappa(\cdot, x) \in \mathcal{H}, \\ \text{(II)} \quad & \forall x \in \mathcal{X}, \quad \forall f \in \mathcal{H}, \quad f(x) = \langle f, \kappa(\cdot, x) \rangle_{\mathcal{H}_\kappa}. \end{aligned} \quad (\text{A.2})$$

By property (I) each point in the input space is mapped onto a function in the RKHS defined by the selected kernel. Property (II) is called the reproducing property of $\kappa(x_i, x_j)$ in \mathcal{H}_κ . It allows defining a nonlinear mapping from the input space to an RKHS as $\phi(x) = \kappa(\cdot, x)$, which leads to

$$\langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}_\kappa} = \langle \kappa(\cdot, x_i), \kappa(\cdot, x_j) \rangle = \kappa(x_i, x_j), \quad (\text{A.3})$$

and thus $\phi(x) = \kappa(\cdot, x)$ defines the Hilbert space associated with the kernel. The kernel property of Equation A.3, is normally called the kernel trick. Moreover, the similarity between functions in the RKHS is also totally defined by the kernel because it defines the inner product of functions (Liu et al., 2011).

A.1.2 Kernel-based learning

The idea behind kernel-based learning consists in mapping the input data $\{x_1, x_2, \dots, x_n\} \subset \mathcal{X}$ (where \mathcal{X} is usually \mathbb{R}^d) into a another space \mathcal{H}_κ , potentially of a much higher dimension, through a non-linear mapping ϕ . Then, the initial learning problem in \mathcal{X} is solved in \mathcal{H}_κ instead, by working with the mapped features $\{\phi(x_1), \phi(x_2), \dots, \phi(x_n)\} \subset \mathcal{H}_\kappa$. Since \mathcal{H}_κ is high-dimensional, a linear learning algorithm applied to the mapped data has the potential to solve arbitrarily nonlinear problems in the input space (provided \mathcal{H}_κ is sufficiently rich). This idea is schematized in Figure A.1. The linear algorithm is implicitly executed in the kernel feature space but thanks to the kernel trick (Equation A.3), the calculations are actually done in the input space, provided the quantities

of interest in the algorithm can be expressed by inner products. The existence of the nonlinear mapping ϕ is ensured by Mercer's theorem.

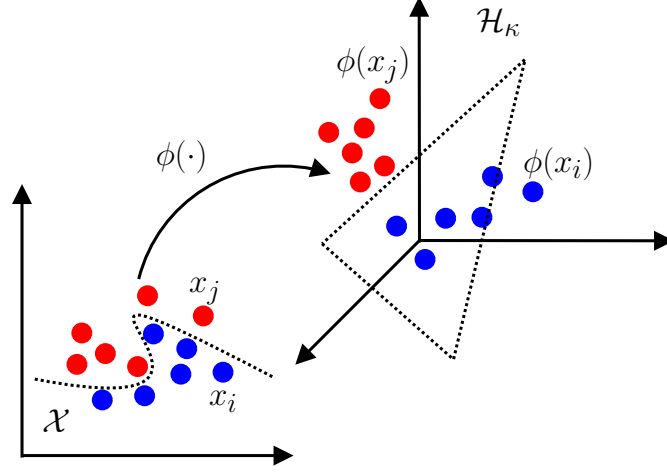


Figure A.1: Schematic representation of the idea behind kernel-based learning. The input data is mapped to a higher dimensional space, where a nonlinear problem in the input space could potentially be solved by a linear learning algorithm.

Mercer's theorem Consider a symmetric kernel function $\kappa \in L_\infty(\mathcal{X} \times \mathcal{X})$. If κ is the kernel of a positive integral operator in $L_2(\mathcal{X})$, and \mathcal{X} is a compact subset of \mathbb{R}^d then

$$\forall \psi \in L_2(\mathcal{X}) : \int_{\mathcal{X}} \kappa(x_i, x_j) \psi(x_i) \psi(x_j) dx dy \geq 0. \quad (\text{A.4})$$

Let $\phi_i \in L_2(\mathcal{X})$ be orthonormal eigenfunctions of the above operator and $\lambda_i > 0$ their corresponding eigenvalues. Then

$$\kappa(x_i, x_j) = \sum_{i=1}^{N_F} \lambda_i \phi_i(x_i) \phi_i(x_j) \quad (\text{A.5})$$

holds for $N_F < \infty$ or $N_F = \infty$. In the latter case the series converges absolutely and uniformly for almost all x_i and x_j in \mathcal{X} . Because of Mercer's theorem, one can use κ as a kernel because ϕ is guaranteed to exist, even if one does not know explicitly what ϕ is (Géron, 2019).

A kernel that satisfies Equation A.5 is known as a Mercer kernel. One of the most

commonly used Mercer kernels is the Gaussian radial basis function (RBF):

$$\kappa(\mathbf{x}_i, \mathbf{x}_j; \sigma) = \exp\left(-\frac{d^2(\mathbf{x}_i, \mathbf{x}_j)}{2\sigma^2}\right) \quad (\text{A.6})$$

where the kernel bandwidth σ acts as a scale parameter. The Gaussian RBF kernel is widely used because of its universal approximating capability, i.e. it allows approximating any continuous function up to a desired accuracy, and mathematical tractability (Hammer and Gersmann, 2003; Liu et al., 2011). Furthermore, it can be shown that for the Gaussian RBF kernel, ϕ maps each training instance to an infinite-dimensional space (Géron, 2019).

A.2 Kernel-based estimation of Renyi's α entropy

ITL, as discussed in Section 1.3.3, is a data-driven learning framework that employs information theoretic quantities as objective functions for learning algorithms (Li and Principe, 2020). In particular, it exploits the properties of Renyi's α -order entropy (Principe, 2010):

$$H_\alpha(X) = \frac{1}{1-\alpha} \log\left(\int_{\mathcal{X}} p(x)^\alpha dx\right),$$

where p is the probability density function (or the probability mass function, in which case the integral is taken as a sum) of the random variable X with support \mathcal{X} .

Given an *i.i.d.* (independent and identically distributed) sample of n realizations of X , $\{x_i\}_{i=1}^n \subset \mathbb{R}^d$, in order to estimate Renyi's α entropy, for $\alpha = 2$, one can use Parzen density estimation to approximate p as $\hat{p}(x) = \frac{1}{n} \sum_{i=1}^n \kappa(x, x_i)$, where $\kappa(\cdot, \cdot) \in \mathbb{R}$ stands for a positive definite kernel function. Then, the obtained probabilities are plugged-in into the definition of Renyi's α entropy, which yields the following estimator:

$$\hat{H}_2(X) = -\log\left(\frac{1}{n^2} \sum_{i,j=1}^n \kappa(x_i, x_j)\right). \quad (\text{A.7})$$

The expression above can be rewritten in terms of Gram matrices. A Gram matrix contains all pairwise evaluations of a positive definite function G on $\{x_i\}_{i=1}^n$, that is, it consists of elements $g_{i,j} = G(x_i, x_j)$ for all $i, j = 1, \dots, n$. Therefore, the Parzen-based entropy estimator in Equation A.7 can be defined in terms of the Gram matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$ as

$$\hat{H}_2(X) = -\log\left(\frac{1}{n^2} \text{tr}(\mathbf{K}\mathbf{K})\right) + C, \quad (\text{A.8})$$

where \mathbf{K} holds elements $k_{ij} = \kappa(x_i, x_j)$, $C \in \mathbb{R}^+$ accounts for the normalization factor of the Parzen window, and $\text{tr}(\cdot)$ stands for the matrix trace.

From Equation A.8 we can see that the Frobenius norm of \mathbf{K} , defined as $\|\mathbf{K}\|_F^2 = \text{tr}(\mathbf{K}\mathbf{K})$, is related to an entropy estimator. Thereby, the Gram matrix \mathbf{K} establishes a connection between information theoretic concepts and kernel methods. In a wider sense, this means that the intermediate step of probability density estimation in the computation of Renyi's α entropy may be bypassed by directly applying a positive definite kernel to the data, as schematized in Figure A.2.

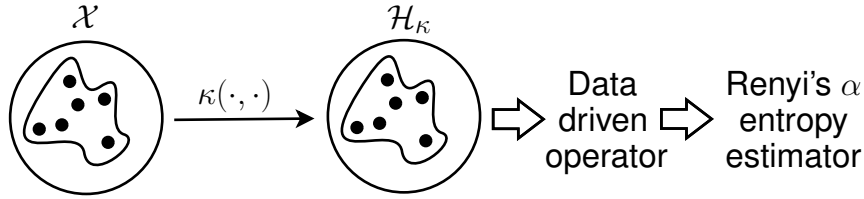


Figure A.2: Schematic representation of the elements involved in the data-driven, kernel-based estimation of Renyi's α entropy.

In Giraldo et al. (2015) the authors employ this notion to introduce entropy functionals that are defined on positive definite matrices and that satisfy similar axioms to those of Renyi's definition of entropy (Rényi et al., 1961). Given a positive definite Gram matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with elements $a_{ij} = \kappa(x_i, x_j)$, and $\text{tr}(\mathbf{A}) = 1$, they propose a kernel-based formulation of entropy expressed as

$$H_\alpha(\mathbf{A}) = \frac{1}{1-\alpha} \log(\text{tr}(\mathbf{A}^\alpha)). \quad (\text{A.9})$$

Such extension of Renyi's α entropy relies on the following theorem to define matrix functions from scalar continuous functions:

Spectral decomposition theorem *Let $D \subset \mathbb{C}$ be a given set and let $\mathcal{N}_n(D) := \{\mathbf{A} \in \mathbf{M}_n : \mathbf{A} \text{ is normal and } \sigma(\mathbf{A}) \in D\}$, where $\sigma(\mathbf{A})$ is the spectrum of \mathbf{A} , and \mathbf{M}_n is the set of all real valued matrices of size $n \times n$. If $f(x)$ is a continuous scalar-valued function on D , then the matrix function*

$$f(\mathbf{A}) = U \begin{pmatrix} f(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f(\lambda_n) \end{pmatrix} U^* \quad (\text{A.10})$$

is continuous on $\mathcal{N}_n(D)$, where $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$, $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$, and $\mathbf{U} \in \mathbf{M}_n$ is unitary.

The spectral theorem implies that if \mathbf{A} is a positive definite matrix with spectrum $\sigma(\mathbf{A})$, and $f(x)$ is a continuous real function defined for all $x \in \sigma(\mathbf{A}) \subseteq \mathbb{R}$. Then, the matrix function $f(\mathbf{A})$ is defined as $\sum_{i \in \mathcal{I}} f(\lambda_i) \mathbf{u}_i \mathbf{u}_i^T$, where $\{\lambda_i\} \in \sigma(\mathbf{A})$ are the eigenvalues of \mathbf{A} and $\{\mathbf{u}_i\}$ the corresponding eigenvectors. In [Giraldo et al. \(2015\)](#), the authors use it to prove that the functional in Equation A.9 satisfies the following set of conditions:

- $H_\alpha(\mathbf{PAP}^*) = H_\alpha(\mathbf{A})$ for any orthonormal matrix $\mathbf{P} \in \mathbf{M}_n$.
- $H_\alpha(p\mathbf{A})$ is a continuous function for $0 < p \leq 1$.
- $H_\alpha\left(\frac{1}{n}\mathbf{I}\right) = \log(n)$, where \mathbf{I} is the identity matrix.
- $H_\alpha(\mathbf{A} \otimes \mathbf{B}) = H_\alpha(\mathbf{A}) + H_\alpha(\mathbf{B})$.
- If $\mathbf{AB} = \mathbf{BA} = \mathbf{0}$; then for the function $g(x) = 2^{(\alpha-1)x}$, for $\alpha \neq 1$ and $\alpha \geq 0$, we have that $H_\alpha(t\mathbf{A} + (1-t)\mathbf{B}) = g^{-1}(tg(H_\alpha(\mathbf{A})) + (1-t)g(H_\alpha(\mathbf{B})))$,

where $\mathbf{B} \in \mathbb{R}^{n \times n}$ is a second positive definite matrix, and $\text{tr}(\mathbf{B}) = 1$. The authors also prove that for $\alpha > 1$

$$H_\alpha(\mathbf{A}) \leq H_\alpha\left(\frac{1}{n}\mathbf{I}\right) \quad (\text{A.11})$$

Furthermore, they use Hadamard products to extend their matrix-based definition of entropy to convey a joint representation of two random variables X and Y , while ensuring that the proposed joint entropy is compatible with the entropies of its components (for instance, the joint entropy should always be larger than the individual entropies). Given two Gram matrices \mathbf{A} and \mathbf{B} holding elements $a_{ij} = \kappa(x_i, x_j)$ and $b_{ij} = \kappa(y_i, y_j)$, with $\kappa_1 : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ and $\kappa_2 : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ positive definite kernels, and $\{z_i = (x_i, y_i)\}_{i=1}^n$ a sample of n pairs representing two sets measurements $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ obtained from the same realization, the joint entropy is defined as:

$$H_\alpha(\mathbf{A}, \mathbf{B}) = H_\alpha\left(\frac{\mathbf{A} \circ \mathbf{B}}{\text{tr}(\mathbf{A} \circ \mathbf{B})}\right) = \frac{1}{1-\alpha} \log\left(\text{tr}\left(\left(\frac{\mathbf{A} \circ \mathbf{B}}{\text{tr}(\mathbf{A} \circ \mathbf{B})}\right)^\alpha\right)\right), \quad (\text{A.12})$$

where the operator \circ stands for the Hadamard product. The authors show that for the expression in Equation A.12 the following inequalities hold, as long as $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{B}) = 1$,

the entries of \mathbf{A} and \mathbf{B} are nonnegative, and $a_{ii} = b_{ii} = 1/n$:

$$H_\alpha \left(\frac{\mathbf{A} \circ \mathbf{B}}{\text{tr}(\mathbf{A} \circ \mathbf{B})} \right) \geq H_\alpha(\mathbf{B}) \quad (\text{A.13})$$

$$H_\alpha \left(\frac{\mathbf{A} \circ \mathbf{B}}{\text{tr}(\mathbf{A} \circ \mathbf{B})} \right) \leq H_\alpha(\mathbf{A}) + H_\alpha(\mathbf{B}) \quad (\text{A.14})$$

Moreover, under this formulation, it is also possible to define conditional entropy and mutual information, provided the additional constraint that the kernels be infinitely divisible. Namely, the conditional entropy can be expressed as:

$$H_\alpha(\mathbf{A}|\mathbf{B}) = H_\alpha(\mathbf{A}, \mathbf{B}) - H_\alpha(\mathbf{B}), \quad (\text{A.15})$$

while the mutual information can be written as:

$$I_\alpha(\mathbf{A}; \mathbf{B}) = H_\alpha(\mathbf{A}) + H_\alpha(\mathbf{B}) - H_\alpha(\mathbf{A}, \mathbf{B}). \quad (\text{A.16})$$

Appendix B

Surface Laplacian

The surface Laplacian of electric potentials measured on the scalp is an estimate of current density entering, or exiting, the scalp through a local region of the skull (Nunez et al., 2006). It is defined as the second spatial derivative of the scalp potential ($\nabla^2 \phi_S$). In the context of EEG analysis, the surface Laplacian serves as a spatial filter that reduces components of low spatial frequency (electrical activity that is broadly distributed, that is to say, activity present simultaneously in several electrodes), which can be associated, for instance, to volume conduction effects (Cohen, 2014). It is this property, its ability to attenuate volume conduction, which makes the surface Laplacian particularly useful as a pre-processing step in electrode-level connectivity analysis (Cohen, 2015).

Spherical spline method for current density approximation: One way to estimate the surface Laplacian is the spherical spline method for source current density estimation, first proposed in Perrin et al. (1989). This approach consists on projecting the electrodes positions onto a sphere, interpolating their corresponding electric potentials on the new coordinates, where the resulting expressions are defined in terms of Legendre Polynomials P_n , and then exploiting the fact that $\nabla^2 P_n = -(2n + 1)P_n$. In practice, for an EEG montage with C channels one proceeds as follows: the first step is obtaining two weighting matrices \mathbf{G} , $\mathbf{H} \in \mathbb{R}^{C \times C}$ holding elements

$$g(i, j) = \frac{1}{4\pi} \sum_{n=1}^o \frac{(2n + 1)P_n(\cos(\mathbf{e}_i, \mathbf{e}_j))}{(n(n + 1))^m}, \quad (\text{B.1})$$

and

$$h(i, j) = \frac{1}{4\pi} \sum_{n=1}^o \frac{-(2n + 1)P_n(\cos(\mathbf{e}_i, \mathbf{e}_j))}{(n(n + 1))^{m-1}}, \quad (\text{B.2})$$

where P_n stands for a Legendre Polynomial of order n , o is the highest polynomial order considered (usually ranges from 7 to 10), $m \in \mathbb{Z}^+$ is smoothness constant (commonly given values from 2 to 6), $\cos(\mathbf{e}_i, \mathbf{e}_j) = 1 - \|\mathbf{e}_i - \mathbf{e}_j\|^2/2$, and $\mathbf{e}_i, \mathbf{e}_j \in [-1, 1]^3$ correspond to the Cartesian coordinates of electrodes i and j , respectively, normalized to a unit-radius sphere. Then, the Laplacian for the C electrodes at time t , $\mathbf{l}_t \in \mathbb{R}^{C \times 1}$, is computed as

$$\mathbf{l}_t = (\boldsymbol{\varpi} H^T)^T, \quad (\text{B.3})$$

with the vector $\boldsymbol{\varpi} \in \mathbb{R}^{1 \times C}$ being defined as

$$\boldsymbol{\varpi} = \mathbf{d}'_t - \frac{\sum \mathbf{d}'_t}{\sum \sum \mathbf{G}_s^{-1}} \sum \mathbf{G}_s^{-1}, \quad (\text{B.4})$$

and,

$$\mathbf{d}'_t = \mathbf{d}_t^T \mathbf{G}_s^{-1}, \quad (\text{B.5})$$

where $\mathbf{d}_t \in \mathbb{R}^{C \times 1}$ holds the values of the C EEG channels at time t , and $\mathbf{G}_s = \mathbf{G} + \mu \mathbf{I}$ is a smoothed version of \mathbf{G} , $\mathbf{I} \in \mathbb{R}^{C \times C}$ is the identity matrix, and μ is a parameter usually ranging from 10^{-6} to 10^{-5} (Cohen, 2014).

Appendix C

Permutation testing

A permutation test is a non-parametrical statistical significance test (Maris and Oostenveld, 2007). It is used to evaluate whether two groups of data are exchangeable (null hypothesis). In the context of TE analyses, permutation tests using trial randomized surrogates have been proposed to address the fact that absolute TE values have limited meaning (Lindner et al., 2011), since information theoretic estimators usually exhibit some level of bias for finite datasets. The overall idea is to test whether a set of TE values, measuring the directed interactions present for a pair of time series from multiple trials of an experiment, are statistically different from a second set of TE values, obtained from non-interacting surrogate data generated by shuffling the target time series between trials (Weber et al., 2017).

Given $\Omega_x = \{\mathbf{x} \in \mathbb{R}^T\}_{n=1}^N$, a set containing N trials of source time series \mathbf{x} ; $\Omega_y = \{\mathbf{y} \in \mathbb{R}^T\}_{n=1}^N$, a set of N trials of the corresponding target time series \mathbf{y} ; and $\lambda(\mathbf{x} \rightarrow \mathbf{y})$, an effective connectivity measure (TE, GC, etc); the permutation test consists of the following steps:

- Compute $\lambda(\mathbf{x} \rightarrow \mathbf{y})$ for the N data trials, so as to obtain a connectivity vector $\boldsymbol{\lambda} = [\lambda(\mathbf{x}_1 \rightarrow \mathbf{y}_1), \lambda(\mathbf{x}_2 \rightarrow \mathbf{y}_2), \dots, \lambda(\mathbf{x}_N \rightarrow \mathbf{y}_N)] \in \mathbb{R}^N$.
- Generate the surrogate data by shuffling the set of target time series between trials, and then compute $\lambda(\mathbf{x} \rightarrow \mathbf{y})$. The result is a connectivity vector $\boldsymbol{\lambda}_{sh} = [\lambda(\mathbf{x}_1 \rightarrow \mathbf{y}_2), \lambda(\mathbf{x}_2 \rightarrow \mathbf{y}_3), \dots, \lambda(\mathbf{x}_N \rightarrow \mathbf{y}_1)] \in \mathbb{R}^N$, where the trials have been shifted by one position.
- Calculate the mean difference between $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}_{sh}$ ($v = \mathbb{E}\{\boldsymbol{\lambda} - \boldsymbol{\lambda}_{sh}\}$).
- Randomly swap (exchange) the elements of $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}_{sh}$, to obtain auxiliary connec-

tivity vectors $\boldsymbol{\lambda}'$ and $\boldsymbol{\lambda}'_{sh}$, and then calculate the mean difference between them ($v' = \mathbb{E}\{\boldsymbol{\lambda}' - \boldsymbol{\lambda}'_{sh}\}$).

- Repeat the above step P_r times (number of permutations) to generate a vector $\boldsymbol{v}' = [v'_1, v'_2, \dots, v'_{P_r}] \in \mathbb{R}^{P_r}$, which holds the distribution of the average differences of the permuted results. In this thesis, the number of permutations was set to 10000, unless otherwise stated.
- If v falls above the 99 percentile of the distribution in \boldsymbol{v}' (v is larger than 99% of the elements of \boldsymbol{v}'), the null hypothesis is rejected, and the connectivity values in $\boldsymbol{\lambda}$ are considered to be statistically significant at an alpha level of 1%. For other significance thresholds the percentage of elements of \boldsymbol{v}' that should be smaller than v is modified accordingly.

Note that the test implicitly assumes that the connectivity values are larger for interacting data than for non-interacting data, so care should be taken when applying it to measures such as the PSI or net connectivities. In those cases, the test should be carried out over the absolute values of the estimated measures.

Appendix D

Kernel-based relevance analysis

When characterizing EEG data through effective brain connectivity measures, it is common practice to carry out all to all channel analyses, which results in a large number of features (compared to the number of trials or observations available). This is especially true in the case when the selected measure also involves spectral information (De La Pava Panche et al., 2021b). Furthermore, many of the estimated connectivity features may not be useful to discriminate between the conditions of the cognitive paradigm of interest (De La Pava Panche et al., 2021a). This can add noise and complexity to any subsequent analysis stage, and potentially lead to a well-known problem in machine learning: the curse of dimensionality (Géron, 2019). This hurdle can be addressed by identifying the set of pairwise channel connectivities that are relevant to discriminate between specific conditions, which, additionally, would lead to a clearer neurophysiological interpretation of the obtained results (Ahmadi et al., 2020). To that end, in this thesis we employ a relevance analysis strategy based on centered kernel alignment.

Centered kernel alignment (CKA) allows quantifying the similarity between two sample spaces by comparing two characterizing kernel functions (Cortes et al., 2012). First, assume we have a feature matrix $\Phi \in \mathbb{R}^{N \times P}$, and a corresponding vector of labels $\mathbf{l} \in \mathbb{Z}^N$, with N the number of observations and P the number of features. For the case of connectivity-based EEG analysis, each element in Φ holds a connectivity value for a pair of channels, with each row of Φ containing multiple connectivity values (features) estimated for a single trial or observation. The corresponding element in \mathbf{l} holds a label identifying the condition associated to that trial. Next, we define two kernel matrices $\mathbf{K}_\Phi \in \mathbb{R}^{N \times N}$ and $\mathbf{K}_\mathbf{l} \in \mathbb{R}^{N \times N}$. The first matrix holds elements $k_{ij}^\Phi = \kappa_\Phi(\phi_i, \phi_j)$ with $\phi_i, \phi_j \in \mathbb{R}^P$ row vectors belonging to Φ , and $\kappa_\Phi(\phi_i, \phi_j; \sigma)$ a Gaussian RBF kernel (see Equation A.6), where $\sigma \in \mathbb{R}^+$ is the kernel’s bandwidth. The second matrix has

elements $k_{ij}^l = \kappa_l(l_i, l_j)$ with $l_i, l_j \in \mathbf{l}$, and

$$\kappa_l(l_i, l_j) = \delta(l_i - l_j), \quad (\text{D.1})$$

a dirac kernel, where $\delta(\cdot)$ stands for the Dirac delta. Then, the CKA can be estimated as:

$$\hat{\rho}(\bar{\mathbf{K}}_\Phi, \bar{\mathbf{K}}_l) = \frac{\|\bar{\mathbf{K}}_\Phi, \bar{\mathbf{K}}_l\|_F}{(\|\bar{\mathbf{K}}_\Phi, \bar{\mathbf{K}}_\Phi\|_F \|\bar{\mathbf{K}}_l, \bar{\mathbf{K}}_l\|_F)^{1/2}}, \quad (\text{D.2})$$

where $\bar{\mathbf{K}} \in \mathbb{R}^{N \times N}$ is the centered version of \mathbf{K} , obtained as $\bar{\mathbf{K}} = \tilde{\mathbf{I}}\mathbf{K}\tilde{\mathbf{I}}$, where $\tilde{\mathbf{I}} = \mathbf{I} - \mathbf{1}\mathbf{1}^\top/N$ is the empirical centering matrix, $\mathbf{I} \in \mathbb{R}^{N \times N}$ is the identity matrix, $\mathbf{1} \in \mathbb{R}^N$ is an all-ones vector, and $\|\bar{\mathbf{K}}, \bar{\mathbf{K}}\|_F = \sqrt{\text{tr}(\bar{\mathbf{K}}\bar{\mathbf{K}}^\top)}$ denotes the matrix-based Frobenius norm. Now, for κ_Φ we select as distance operator the the Mahalanobis distance

$$d_A^2(\phi_i, \phi_j) = (\phi_i - \phi_j) \mathbf{\Gamma} \mathbf{\Gamma}^\top (\phi_i - \phi_j)^\top \quad (\text{D.3})$$

where $\mathbf{\Gamma} \in \mathbb{R}^{P \times Q}$, $Q \leq P$, is a linear projection matrix, and $\mathbf{\Gamma} \mathbf{\Gamma}^\top$ is the corresponding inverse covariance matrix. Afterward, the projection matrix $\mathbf{\Gamma}$ is obtained by solving the following optimization problem:

$$\hat{\mathbf{\Gamma}} = \arg \max_{\mathbf{\Gamma}} \log \left(\hat{\rho}(\bar{\mathbf{K}}_\Phi, \bar{\mathbf{K}}_l; \mathbf{\Gamma}) \right), \quad (\text{D.4})$$

where the logarithm function is used for mathematical convenience. $\hat{\mathbf{\Gamma}}$ can be estimated through standard stochastic gradient descent, as detailed in [Fernández-Ramírez et al. \(2020\)](#), through the update rule

$$\mathbf{\Gamma}^{r+1} = \mathbf{\Gamma}^r - \mu_{\mathbf{\Gamma}}^r \nabla_{\mathbf{\Gamma}^r} \left(\hat{\rho}(\mathbf{K}_\Phi, \mathbf{K}_l) \right), \quad (\text{D.5})$$

where $\mu \in \mathbb{R}^+$ is the step size of the learning rule, and r indicates a time step. Finally, we quantify the contribution of each feature to the projection matrix $\hat{\mathbf{\Gamma}}$, which maximizes the alignment between the feature and label spaces, by building a relevance vector index $\boldsymbol{\varrho} \in \mathbb{R}^P$, whose elements are defined as:

$$\varrho_p = \sum_{q=1}^Q |\gamma_{pq}|; \quad \forall p \in P, \quad \gamma \in \mathbf{\Gamma}. \quad (\text{D.6})$$

$\boldsymbol{\varrho}$ can then be used to rank the features in Φ according to their discrimination capability. A high ϱ_p value indicates that the p -th feature in Φ , in our case a connection between

a specific pair of channels, is relevant when it comes to distinguishing between the conditions contained in the label vector \mathbf{l} .

Appendix E

Cao's criterion

Cao's criterion is a practical method for determining the minimum embedding dimension of a scalar time series (Cao, 1997). The time-delay reconstruction of a time series $\mathbf{x} \in \mathbb{R}^N$ can be expressed as

$$\mathbf{x}_t(d) = (x_t, x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-(d-1)\tau}), \quad (\text{E.1})$$

where $t = \{1, 2, \dots, N - (d-1)\tau\}$, and $d, \tau \in \mathbb{N}$ are the embedding dimension and delay, respectively. Note that for the sake of clarity, and in order to make the dependence of the reconstructed vectors on d explicit, we have changed the notation used in Equation 1.4. The first step proposed by Cao to determine an adequate value for d , consists of computing the following quantity

$$a(i, d) = \frac{\|\mathbf{x}_i(d+1) - \mathbf{x}_{n(i,d)}(d+1)\|_\infty}{\|\mathbf{x}_i(d) - \mathbf{x}_{n(i,d)}(d)\|_\infty}, \quad (\text{E.2})$$

where $i = \{1, 2, \dots, N - d\tau\}$, $\mathbf{x}_i(d+1)$ is the i th reconstructed vector with dimension $d+1$, and $1 \leq n(i, d) \leq N - d\tau$. The value of $n(i, d)$ is chosen so that $\mathbf{x}_{n(i,d)}(d)$ is the nearest neighbor of $\mathbf{x}_i(d)$ in the d -dimensional embedded space, according to $\|\cdot\|_\infty$. If $\mathbf{x}_{n(i,d)}(d)$ and $\mathbf{x}_i(d)$ are equal, the second nearest neighbor is selected. Then, the average of the $a(i, d)$ values is obtained as

$$E(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} a(i, d) \quad (\text{E.3})$$

$E(d)$ depends only on d and τ (which can, for instance, be estimated as the time series autocorrelation time), and its variation from d to $d + 1$ is defined as

$$E'(d) = \frac{E(d+1)}{E(d)} \tag{E.4}$$

The target embedding dimension is the first value of d for which $E'(d) = E'(d - 1)$ (in practice one might need to search instead for $|E'(d) - E'(d - 1)| < \epsilon$, with ϵ an acceptable tolerance).

Appendix F

Neural mass model equations

The neural mass model schematized in Figure 3.1A consists of the following set of stochastic differential equations (David et al., 2004):

$$\dot{z}_0 = z_3$$

$$\dot{z}_1 = z_4$$

$$\dot{z}_2 = z_5$$

$$\tau_e^1 \dot{z}_3 = H_e^1 S(wz_1 + (1-w)z_{10} - wz_2 - (1-w)z_6) - 2z_3 - z_0/\tau_e^1$$

$$\tau_e^1 \dot{z}_4 = H_e^1 (C_{21}^* (S(w(z_{13}(t-\nu) - z_{14}(t-\nu)) + (1-w)(z_{22}(t-\nu) - z_{18}(t-\nu))) - a) + (1 - C_{21})p_1 + P + c_2 S(c_1(wz_0 + (1-w)z_8))) - 2z_4 - z_1/\tau_e^1$$

$$\tau_i^1 \dot{z}_5 = H_i^1 c_4 S(c_3(wz_0 + (1-w)z_8)) - 2z_5 - z_2/\tau_i^1$$

$$\dot{z}_6 = z_7$$

$$\tau_i^2 \dot{z}_7 = H_i^2 c_4 S(c_3(wz_0 + (1-w)z_8)) - 2z_7 - z_6/\tau_i^2$$

$$\dot{z}_8 = z_9$$

$$\tau_e^2 \dot{z}_9 = H_e^2 S(wz_1 + (1-w)z_{10} - wz_2 - (1-w)z_6) - 2z_9 - z_8/\tau_e^2$$

$$\dot{z}_{10} = z_{11}$$

$$\tau_e^2 \dot{z}_{11} = H_e^2 (C_{21}^* (S(w(z_{13}(t-\nu) - z_{14}(t-\nu)) + (1-w)(z_{22}(t-\nu) - z_{18}(t-\nu))) - a) + (1 - C_{21})p_1 + P + c_2 S(c_1(wz_0 + (1-w)z_8))) - 2z_{11} - z_{10}/\tau_e^2$$

$$\dot{z}_{12} = z_{15}$$

$$\dot{z}_{13} = z_{16}$$

$$\dot{z}_{14} = z_{17}$$

$$\tau_e^1 \dot{z}_{15} = H_e^1 S(wz_{13} + (1-w)z_{22} - wz_{14} - (1-w)z_{18}) - 2z_{15} - z_{12}/\tau_e^1$$

$$\tau_e^1 \dot{z}_{16} = H_e^1 (C_{12}^* (S(w(z_1(t-\nu) - z_2(t-\nu)) + (1-w)(z_{10}(t-\nu) - z_6(t-\nu))) - a) + (1 - C_{12})p_2 + P + c_2 S(c_1(wz_{12} + (1-w)z_{20}))) - 2z_{16} - z_{13}/\tau_e^1$$

$$\tau_i^1 \dot{z}_{17} = H_i^1 c_4 S(c_3(wz_{12} + (1-w)z_{20})) - 2z_{17} - z_{14}/\tau_i^1$$

$$\dot{z}_{18} = z_{19}$$

$$\tau_i^2 \dot{z}_{19} = H_i^2 c_4 S(c_3(wz_{12} + (1-w)z_{20})) - 2z_{19} - z_{18}/\tau_i^2$$

$$\dot{z}_{20} = z_{21}$$

$$\tau_e^2 \dot{z}_{21} = H_e^2 S(wz_{13} + (1-w)z_{22} - wz_{14} - (1-w)z_{18}) - 2z_{21} - z_{20}/\tau_e^2$$

$$\dot{z}_{22} = z_{23}$$

$$\tau_e^2 \dot{z}_{23} = H_e^2 (C_{12}^* (S(w(z_1(t-\nu) - z_2(t-\nu)) + (1-w)(z_{10}(t-\nu) - z_6(t-\nu))) - a) + (1 - C_{12})p_2 + P + c_2 S(c_1(wz_{12} + (1-w)z_{20}))) - 2z_{23} - z_{22}/\tau_e^2$$

where

$$S(v) = \frac{e_0}{1 + \exp(r(v_0 - v))}$$

$$C_{12}^* = \frac{\sigma_p \sqrt{2C_{12} - C_{12}^2}}{\sqrt{\text{var}\{S(w(z_1 - z_2) + (1-w)(z_{10} - z_6))\}}}$$

$$C_{21}^* = \frac{\sigma_p \sqrt{2C_{21} - C_{21}^2}}{\sqrt{\text{var}\{S(w(z_{13} - z_{14}) + (1-w)(z_{22} - z_{18}))\}}}.$$

The signals originating from cortical areas 1 and 2, corresponding to the time series \mathbf{x} and \mathbf{y} in Figure 3.1A, are defined as

$$\text{signal}_{\text{Area 1}} = w(z_1 - z_2) + (1-w)(z_{10} - z_6), \quad (\text{F.1})$$

and

$$\text{signal}_{\text{Area 2}} = w(z_{13} - z_{14}) + (1-w)(z_{22} - z_{18}). \quad (\text{F.2})$$

Finally, for the experiments described in Section 3.2.1, the initial model parameters were set as in David et al. (2004), using the following values: $a = 3.501$, $c = 135$,

$c_1 = c_2 = 0.8c$, $c_3 = c_4 = 0.25c$, $H_e^1 = 3$, $H_e^2 = 7$, $H_i^1 = 20$, $H_i^2 = 150$, $\tau_e^1 = 0.0108$, $\tau_e^2 = 0.0046$, $\tau_i^1 = 0.022$, $\tau_i^2 = 0.0029$, $e_0 = 5$, $v_0 = 6$, $r = 0.56$, $w = 0.8$, $P = 220$, and $\sigma_p = 22$ (which stands for the standard deviation of the Gaussian noise terms $p_{1,2}$). Furthermore, the model's differential equations were numerically integrated using a second order Runge–Kutta algorithm.

References

- Ali Kareem Abbas, Ghasem Azemi, Sajad Amiri, Samin Ravanshadi, and Amir Omidvarnia. Effective connectivity in brain networks estimated using eeg signals is altered in children with ADHD. *Computers in Biology and Medicine*, 134:1–9, 2021. (page 1)
- U Rajendra Acharya, Hamido Fujita, Vidya K Sudarshan, Shreya Bhat, and Joel EW Koh. Application of entropies for automated diagnosis of epilepsy using EEG signals: a review. *Knowledge-Based Systems*, 88:85–96, 2015. (page 34)
- Amirmasoud Ahmadi, Saeideh Davoudi, Mahsa Behroozi, and Mohammad Reza Daliri. Decoding covert visual attention based on phase transfer entropy. *Physiology & behavior*, 222:112932, 2020. (pages 7, 18, 19, and 117)
- Hirotsugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974. (page 68)
- Juhan Aru, Jaan Aru, Viola Priesemann, Michael Wibral, Luiz Lana, Gordon Pipa, Wolf Singer, and Raul Vicente. Untangling cross-frequency coupling in neuroscience. *Current opinion in neurobiology*, 31:51–61, 2015. (pages 5 and 91)
- Alan Baddeley. Working memory: theories, models, and controversies. *Annual review of psychology*, 63:1–29, 2012. (pages 28 and 29)
- Hanieh Bakhshayesh, Sean P Fitzgibbon, Azin S Janani, Tyler S Grummett, and Kenneth J Pope. Detecting connectivity in EEG: a comparative study of data-driven effective connectivity measures. *Computers in Biology and Medicine*, 111:103329, 2019. (pages 4 and 9)
- Lionel Barnett, Adam B Barrett, and Anil K Seth. Granger causality and transfer entropy are equivalent for gaussian variables. *Physical review letters*, 103(23):238701, 2009. (page 12)
- André M Bastos and Jan-Mathijs Schoffelen. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience*, 9:175, 2016. (pages 1, 4, and 84)
- Michel Besserve, Bernhard Schölkopf, Nikos K Logothetis, and Stefano Panzeri. Causal relationships between frequency bands of extracellular signals in visual cortex revealed by an information theoretic analysis. *Journal of computational neuroscience*, 29(3):547–566, 2010. (pages 5, 8, 20, 21, and 25)

- Liangyue Cao. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D: Nonlinear Phenomena*, 110(1-2):43–50, 1997.
(pages 42, 68, and 120)
- Sezen Cekic, Didier Grandjean, and Olivier Renaud. Time, frequency, and time-varying granger-causality measures in neuroscience. *Statistics in medicine*, 37(11):1910–1931, 2018.
(pages 2, 5, 6, 7, 9, 46, and 57)
- Xiaoling Chen, Yuanyuan Zhang, Shengcui Cheng, and Ping Xie. Transfer spectral entropy and application to functional corticomuscular coupling. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(5):1092–1102, 2019.
(pages 5, 7, 8, 19, and 63)
- Ning Cheng, Qun Li, Sitong Wang, Rubin Wang, and Tao Zhang. Permutation mutual information: a novel approach for measuring neuronal phase-amplitude coupling. *Brain topography*, 31(2):186–201, 2018.
(pages 8 and 20)
- Michael X Cohen. Comparison of different spatial transformations applied to EEG data: a case study of error processing. *International Journal of Psychophysiology*, 97(3):245–257, 2015.
(pages 2, 28, and 113)
- Mike X Cohen. *Analyzing neural time series data: theory and practice*. MIT press, 2014.
(pages 113 and 114)
- DF Collazos-Huertas, AM Álvarez-Meza, CD Acosta-Medina, GA Castaño-Duque, and G Castellanos-Dominguez. CNN-based framework using spatial dropping for enhanced interpretation of neural activity in motor imagery classification. *Brain Informatics*, 7,8(1):1–13, 2020.
(pages 27, 67, 91, and 94)
- Christos Constantinidis and Torkel Klingberg. The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience*, 17(7):438–449, 2016. (page 55)
- Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh. Algorithms for learning kernels based on centered alignment. *The Journal of Machine Learning Research*, 13:795–828, 2012.
(page 117)
- Jolien Cremers and Irene Klugkist. One direction? a tutorial for circular data analysis using R with examples in cognitive psychology. *Frontiers in psychology*, 9:2040, 2018.
(page 7)
- Fernando Lopes Da Silva. EEG: origin and measurement. In *EEG-fMRI*, pages 19–38. Springer, 2009.
(pages 1 and 5)
- Zhongxiang Dai, Joshua De Souza, Julian Lim, Paul M Ho, Yu Chen, Junhua Li, Nitish Thakor, Anastasios Bezerianos, and Yu Sun. EEG cortical connectivity analysis of working memory reveals topological reorganization in theta and alpha bands. *Frontiers in human neuroscience*, 11:237, 2017.
(pages 28, 30, and 54)
- Jonathan Daume, Thomas Gruber, Andreas K Engel, and Uwe Frieze. Phase-amplitude coupling and long-range phase synchronization reveal frontotemporal interactions during visual working memory. *Journal of Neuroscience*, 37(2):313–322, 2017.
(pages 8 and 30)

- Olivier David and Karl J Friston. A neural mass model for MEG/EEG: coupling and neuronal dynamics. *NeuroImage*, 20(3):1743–1755, 2003. (pages 60, 63, and 65)
- Olivier David, Diego Cosmelli, and Karl J Friston. Evaluation of different measures of functional connectivity using a neural mass model. *Neuroimage*, 21(2):659–673, 2004. (pages 63, 122, and 123)
- Ivan De La Pava Panche, Andres M Alvarez-Meza, and Alvaro Orozco-Gutierrez. A data-driven measure of effective connectivity based on renyi’s α -entropy. *Frontiers in neuroscience*, 13:1277, 2019. (pages 24, 51, 61, 68, 74, 75, 78, 96, and 99)
- Iván De La Pava Panche, Andrés Álvarez-Meza, Paula Marcela Herrera Gómez, David Cárdenas-Peña, Jorge Iván Ríos Patiño, and Álvaro Orozco-Gutiérrez. Kernel-based phase transfer entropy with enhanced feature relevance analysis for brain computer interfaces. *Applied Sciences*, 11(15):6689, 2021a. (pages 25 and 117)
- Iván De La Pava Panche, Viviana Gómez-Orozco, Andrés Álvarez-Meza, David Cárdenas-Peña, and Álvaro Orozco-Gutiérrez. Estimating directed phase-amplitude interactions from EEG data through kernel-based phase transfer entropy. *Applied Sciences*, 11(21):9803, 2021b. (pages 25 and 117)
- Stefan Debener, Falk Minow, Reiner Emkes, Katharina Gandras, and Maarten De Vos. How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology*, 49(11):1617–1621, 2012. (pages 2 and 71)
- Stavros Dimitriadis, Yu Sun, Nikolaos Laskaris, Nitish Thakor, and Anastasios Bezerianos. Revealing cross-frequency causal interactions during a mental arithmetic task through symbolic transfer entropy: a novel vector-quantization approach. *IEEE Trans Neural Syst Rehabil Eng*, 24(10):1017–1028, 2016a. (pages 1, 3, 5, 6, 15, 24, 30, 37, 42, 46, 57, 62, 78, 82, and 97)
- Stavros I Dimitriadis, Nikolaos A Laskaris, Vasso Tsirka, Sofia Erimaki, Michael Vourkas, Sifis Micheloyannis, and Spiros Fotopoulos. A novel symbolization scheme for multichannel recordings with emphasis on phase information and its application to differentiate EEG activity from different mental tasks. *Cognitive neurodynamics*, 6(1):107–113, 2012. (page 15)
- Stavros I Dimitriadis, Yu Sun, Nitish V Thakor, and Anastasios Bezerianos. Causal interactions between frontal θ –parieto-occipital α 2 predict performance on a mental arithmetic task. *Frontiers in human neuroscience*, 10:454, 2016b. (pages 3, 8, 20, and 30)
- Frank H Duffy, Aditi Shankardass, Gloria B McAnulty, and Heidelise Als. A unique pattern of cortical connectivity characterizes patients with attention deficit disorders: a large electroencephalographic coherence study. *BMC medicine*, 15(1):51, 2017. (page 1)
- Ali Ekhlasi, Ali Motie Nasrabadi, and Mohammad Reza Mohammadi. Direction of information flow between brain regions in ADHD and healthy children based on EEG by using directed phase transfer entropy. *Cognitive Neurodynamics*, 15(6):975–986, 2021. (page 19)

- Basem Elasuty and Seif Eldawlatly. Dynamic bayesian networks for EEG motor imagery feature extraction. In *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 170–173. IEEE, 2015. (pages 27, 52, and 75)
- Andreas K Engel, Christian Gerloff, Claus C Hilgetag, and Guido Nolte. Intrinsic coupling modes: multiscale interactions in ongoing brain activity. *Neuron*, 80(4):867–886, 2013. (page 1)
- J Fernández-Ramírez, A Álvarez-Meza, EM Pereira, A Orozco-Gutiérrez, and G Castellanos-Dominguez. Video-based social behavior recognition based on kernel relevance analysis. *The Visual Computer*, 36(8):1535–1547, 2020. (pages 25 and 118)
- Karl J Friston. Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36, 2011. (pages 1 and 4)
- Yu Fukuda, Teresa Katthagen, Lorenz Deserno, Leila Shayegan, Jakob Kaminski, Andreas Heinz, and Florian Schlagenhauf. Reduced parietofrontal effective connectivity during a working-memory task in people with high delusional ideation. *Journal of Psychiatry and Neuroscience*, 44(3):195–204, 2019. (page 1)
- Steven Galindo-Noreña, David Cárdenas-Peña, and Álvaro Orozco-Gutierrez. Multiple kernel stein spatial patterns for the multiclass discrimination of motor imagery tasks. *Applied Sciences*, 10(23):8628, 2020. (page 27)
- Jianbo Gao, Jing Hu, Thomas Buckley, Keith White, and Chris Hass. Shannon and renyi entropies to classify effects of mild traumatic brain injury on postural sway. *PLoS One*, 6(9):e24446, 2011. (page 34)
- Pedro García and R Mujica. A local approach for information transfer. *Communications in Nonlinear Science and Numerical Simulation*, 70:326–333, 2019. (pages 6 and 14)
- Daniel Guillermo García-Murillo, Andres Alvarez-Meza, and German Castellanos-Dominguez. Single-trial kernel-based functional connectivity for enhanced feature extraction in motor-related tasks. *Sensors*, 21(8):2750, 2021. (pages 26, 27, 74, and 76)
- Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O’Reilly Media, 2019. (pages 96, 106, 108, 109, and 117)
- Matthieu Gilson, Gorka Zamora-López, Vicente Pallarés, Mohit H Adhikari, Mario Senden, Adrià Tauste Campo, Dante Mantini, Maurizio Corbetta, Gustavo Deco, and Andrea Insabato. Model-based whole-brain effective connectivity to study distributed cognition in health and disease. *Network Neuroscience*, 4(2):338–373, 2020. (page 1)
- Luis Gonzalo Sanchez Giraldo, Murali Rao, and Jose C Principe. Measures of entropy from data using infinitely divisible kernels. *IEEE Transactions on Information Theory*, 61(1):535–548, 2015. (pages 5, 6, 13, 17, 24, 34, 43, 58, 68, 87, 106, 110, and 111)
- V Gómez, A Álvarez, P Herrera, G Castellanos, and A Orozco. Short time EEG connectivity features to support interpretability of MI discrimination. In *Iberoamerican Congress on Pattern Recognition*, pages 699–706. Springer, 2018. (pages 27, 28, 52, and 75)

- Anmin Gong, Jianping Liu, Si Chen, and Yunfa Fu. Time–frequency cross mutual information analysis of the brain functional networks underlying multiclass motor imagery. *Journal of motor behavior*, 50(3):254–267, 2018. (pages 54 and 68)
- Barbara Hammer and Kai Gersmann. A note on the universal approximation capability of support vector machines. *neural processing letters*, 17(1):43–53, 2003. (page 109)
- Mahmoud Hassan and Fabrice Wendling. Aiming for high resolution of brain networks in time and space electroencephalography source connectivity. *IEEE Signal Processing Magazine*, 35(3):81–96, 2018. (pages 1 and 2)
- Mahmoud Hassan, Olivier Dufor, Isabelle Merlet, Claude Berrou, and Fabrice Wendling. EEG source connectivity analysis: from dense array recordings to brain networks. *PloS one*, 9(8):e105041, 2014. (page 2)
- Mahmoud Hassan, Pascal Benquet, Arnaud Biraben, Claude Berrou, Olivier Dufor, and Fabrice Wendling. Dynamic reorganization of functional brain networks during picture naming. *Cortex*, 73:276–288, 2015. (page 2)
- Stefan Haufe, Vadim V Nikulin, Klaus-Robert Müller, and Guido Nolte. A critical assessment of connectivity measures for EEG data: a simulation study. *Neuroimage*, 64:120–133, 2013. (page 44)
- Sébastien Hétu, Mathieu Grégoire, Arnaud Saimpont, Michel-Pierre Coll, Fanny Eugène, Pierre-Emmanuel Michon, and Philip L Jackson. The neural network of motor imagery: an ale meta-analysis. *Neuroscience & Biobehavioral Reviews*, 37(5):930–949, 2013. (pages 54 and 78)
- Arjan Hillebrand, Prejaas Tewarie, Edwin Van Dellen, Meichen Yu, Ellen WS Carbo, Linda Douw, Alida A Gouw, Elisabeth CW Van Straaten, and Cornelis J Stam. Direction of information flow in large-scale resting-state networks is frequency-dependent. *Proceedings of the National Academy of Sciences*, 113(14):3867–3872, 2016. (pages 7 and 19)
- Alexandre Hyafil, Anne-Lise Giraud, Lorenzo Fontolan, and Boris Gutkin. Neural cross-frequency coupling: connecting architectures, mechanisms, and functions. *Trends in neurosciences*, 38(11):725–740, 2015. (page 5)
- Haiteng Jiang, Ali Bahramisharif, Marcel AJ van Gerven, and Ole Jensen. Measuring directionality between neuronal oscillations of different frequencies. *Neuroimage*, 118:359–367, 2015. (pages 7, 8, 18, 20, 62, 70, 84, 85, 88, 94, and 102)
- Viktor Jirsa and Viktor Müller. Cross-frequency coupling in real and virtual brain networks. *Frontiers in computational neuroscience*, 7:1–25, 2013. (pages 8, 18, and 19)
- Elizabeth L Johnson, Jenna N Adams, Anne-Kristin Solbakk, Tor Endestad, Pål G Larsson, Jugoslav Ivanovic, Torstein R Meling, Jack J Lin, and Robert T Knight. Dynamic frontotemporal systems process space and time in working memory. *PLoS biology*, 16(3):e2004274, 2018. (pages 3, 8, 19, 29, 30, 57, 67, 74, 78, 82, 91, 95, and 97)

- Elizabeth L Johnson, David King-Stephens, Peter B Weber, Kenneth D Laxer, Jack J Lin, and Robert T Knight. Spectral imprints of working memory for everyday associations in the frontoparietal network. *Frontiers in systems neuroscience*, 12:65, 2019. (pages 28, 30, 67, 74, 82, 95, and 97)
- Huan Kang, Xiaofeng Zhang, and Guangbin Zhang. Phase permutation entropy: a complexity measure for nonlinear time series incorporating phase information. *Physica A: Statistical Mechanics and its Applications*, 568:125686, 2021. (pages 7 and 15)
- Katherine H Karlsgodt, David C Glahn, Theo GM van Erp, Sebastian Therman, Matti Huttunen, Marko Manninen, Jaakko Kaprio, Mark S Cohen, Jouko Lönqvist, and Tyrone D Cannon. The relationship between performance and fMRI signal during working memory in patients with schizophrenia, unaffected co-twins, and control subjects. *Schizophrenia research*, 89(1-3):191–197, 2007. (page 104)
- Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004. (pages 6, 15, 24, and 62)
- Rafal Kus, Maciej Kaminski, and Katarzyna J Blinowska. Determination of EEG activity propagation: pair-wise versus multichannel estimate. *IEEE transactions on Biomedical Engineering*, 51(9):1501–1510, 2004. (page 38)
- Tom Dupre La Tour, Lucille Tallot, Laetitia Grabot, Valérie Doyère, Virginie Van Wassenhove, Yves Grenier, and Alexandre Gramfort. Non-linear auto-regressive models for cross-frequency coupling in neural time series. *PLoS computational biology*, 13(12):e1005893, 2017. (pages 5, 8, 18, 19, 20, and 99)
- Agatha Lenartowicz and Sandra K Loo. Use of EEG to diagnose ADHD. *Current psychiatry reports*, 16(11):498, 2014. (pages 1 and 2)
- Duan Li, Hongxin Zhang, Muhammad Saad Khan, and Fang Mi. A self-adaptive frequency selection common spatial pattern and least squares twin support vector machine for motor imagery electroencephalography recognition. *Biomedical Signal Processing and Control*, 41:222–232, 2018. (pages 27, 52, and 75)
- Kan Li and Jose C Principe. Fast estimation of information theoretic learning descriptors using explicit inner product spaces. *arXiv preprint arXiv:2001.00265*, 2020. (pages 12 and 109)
- Shuang Liang, Kup-Sze Choi, Jing Qin, Qiong Wang, Wai-Man Pang, and Pheng-Ann Heng. Discrimination of motor imagery tasks via information flow pattern of brain connectivity. *Technology and Health Care*, 24(s2):S795–S801, 2016. (pages 27, 52, and 75)
- Wei-Kuang Liang, Philip Tseng, Jia-Rong Yeh, Norden E Huang, and Chi-Hung Juan. Frontoparietal beta amplitude modulation and its interareal cross-frequency coupling in visual working memory. *Neuroscience*, 460:69–87, 2021. (pages 30 and 95)
- Zhenhu Liang, Yinghua Wang, Xue Sun, Duan Li, Logan J Voss, Jamie W Sleight, Satoshi Hagihira, and Xiaoli Li. EEG entropy measures in anesthesia. *Frontiers in computational neuroscience*, 9:1–17, 2015. (page 34)

- George C Linderman and Stefan Steinerberger. Clustering with t-SNE, provably. *SIAM Journal on Mathematics of Data Science*, 1(2):313–332, 2019. (page 77)
- Michael Lindner, Raul Vicente, Viola Priesemann, and Michael Wibral. TRENTOOL: a Matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC neuroscience*, 12(119):1–22, 2011. (pages 6, 10, 16, 24, 39, 42, 43, 48, 64, 67, 68, 82, 90, and 115)
- Weifeng Liu, Jose C Principe, and Simon Haykin. *Kernel adaptive filtering: a comprehensive introduction*, volume 57. John Wiley & Sons, 2011. (pages 42, 106, 107, and 109)
- Muriel Lobier, Felix Siebenhühner, Satu Palva, and J Matias Palva. Phase transfer entropy: a novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. *Neuroimage*, 85:853–872, 2014. (pages 5, 7, 18, 19, 21, 24, 60, 61, 62, 63, 65, 71, 74, 82, 85, and 86)
- Rakesh Malladi, Don H Johnson, Giridhar P Kalamangalam, Nitin Tandon, and Behnaam Aazhang. Mutual information in frequency and its application to measure cross-frequency coupling in epilepsy. *IEEE Transactions on signal processing*, 66(11):3008–3023, 2018. (pages 8 and 20)
- Nadia Mammone, Jonas Duun-Henriksen, Troels Kjaer, and Francesco Morabito. Differentiating interictal and ictal states in childhood absence epilepsy through permutation rényi entropy. *Entropy*, 17(7):4627–4643, 2015. (page 34)
- Eric Maris and Robert Oostenveld. Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, 164(1):177–190, 2007. (page 115)
- Ramón Martínez-Cancino, Joseph Heng, Arnaud Delorme, Ken Kreutz-Delgado, Roberto C Sotero, and Scott Makeig. Measuring transient phase-amplitude coupling using local mutual information. *NeuroImage*, 185:361–378, 2019. (pages 8 and 20)
- Ramón Martínez-Cancino, Arnaud Delorme, Johanna Wagner, Kenneth Kreutz-Delgado, Roberto C Sotero, and Scott Makeig. What can local transfer entropy tell us about phase-amplitude coupling in electrophysiological signals? *Entropy*, 22(11):1262, 2020. (pages 5, 8, 16, 19, 20, 21, and 25)
- Moemi Matsuo, Naoki Iso, Kengo Fujiwara, Takefumi Moriuchi, Daiki Matsuda, Wataru Mitsunaga, Akira Nakashima, and Toshio Higashi. Comparison of cerebral activation between motor execution and motor imagery of self-feeding activity. *Neural regeneration research*, 16(4):778, 2021. (page 26)
- Maarten Mennes, Heidi Wouters, Bart Vanrumste, Lieven Lagae, and Peter Stiers. Validation of ICA as a tool to remove eye movement artifacts from EEG/ERP. *Psychophysiology*, 47(6):1142–1150, 2010. (pages 2 and 71)
- Bratislav Mišić and Olaf Sporns. From regions to connections and networks: new bridges between brain and behavior. *Current opinion in neurobiology*, 40:1–7, 2016. (page 1)
- Alessandro Montalto, Luca Faes, and Daniele Marinazzo. MuTE: a Matlab toolbox to compare established and novel estimators of the multivariate transfer entropy. *PloS one*, 9(10):e109462, 2014. (pages 6, 14, 15, and 103)

- Guido Nolte, Andreas Ziehe, Vadim V Nikulin, Alois Schlögl, Nicole Krämer, Tom Brismar, and Klaus-Robert Müller. Robustly estimating the flow direction of information in complex physical systems. *Physical review letters*, 100(23):234101, 2008. (pages 7, 18, 62, 70, 85, and 86)
- Tianne Numan, Arjen JC Slooter, Arendina W van der Kooi, Annemieke ML Hoekman, Willem JL Suyker, Cornelis J Stam, and Edwin van Dellen. Functional connectivity and network analysis during hypoactive delirium and recovery from anesthesia. *Clinical Neurophysiology*, 128(6):914–924, 2017. (pages 7 and 19)
- Paul L Nunez, Ramesh Srinivasan, et al. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2006. (page 113)
- Yuri G Pavlov and Boris Kotchoubey. Oscillatory brain activity and maintenance of verbal and visual working memory: a systematic review. *Psychophysiology*, page e13735, 2020. (pages 28 and 29)
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. (page 40)
- François Perrin, J Pernier, O Bertrand, and JF Echallier. Spherical splines for scalp potential and current density mapping. *Electroencephalography and clinical neurophysiology*, 72(2):184–187, 1989. (pages 28, 40, and 113)
- Edoardo Pinzuti, Patricia Wollstadt, Aaron Gutknecht, Oliver Tüscher, and Michael Wibral. Measuring spectrally-resolved information transfer. *PLOS Computational Biology*, 16(12):e1008526, 2020. (pages 3, 5, 7, 8, 17, 19, 20, 21, 82, and 99)
- Jose C Principe. *Information theoretic learning: Renyi’s entropy and kernel perspectives*. Springer Science & Business Media, 2010. (pages 13, 34, 58, and 109)
- Dheeraj Rathee, Hubert Cecotti, and Girijesh Prasad. Single-trial effective brain connectivity patterns enhance discriminability of mental imagery tasks. *Journal of neural engineering*, 14(5):056005, 2017. (pages 1, 2, 27, 28, and 52)
- Alfréd Rényi et al. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California, 1961. (pages 12, 13, 34, and 110)
- Andres Roman-Urrestarazu, Päivi Lindholm, Irma Moilanen, Vesa Kiviniemi, Jouko Miettunen, Erika Jääskeläinen, Pirjo Mäki, Tuula Hurtig, Hanna Ebeling, Jennifer H Barnett, et al. Brain structural deficits and working memory fMRI dysfunction in young adults who were diagnosed with ADHD in adolescence. *European child & adolescent psychiatry*, 25(5):529–538, 2016. (page 104)
- Vangelis Sakkalis. Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Computers in biology and medicine*, 41(12):1110–1117, 2011. (pages 1, 2, 4, 5, and 71)

- Koichi Sameshima and Luiz Antonio Baccala. *Methods in brain connectivity inference through multivariate time series analysis*. CRC press, 2016. (page 12)
- Ralf Schlösser, Thomas Gesierich, Bettina Kaufmann, Goran Vucurevic, Stefan Hunsche, Joachim Gawehn, and Peter Stoeter. Altered effective connectivity during working memory performance in schizophrenia: a study with fMRI and structural equation modeling. *Neuroimage*, 19(3):751–763, 2003. (page 104)
- Bernhard Schölkopf, Alexander J Smola, Francis Bach, et al. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002. (pages 43, 57, and 68)
- Thomas Schreiber. Measuring information transfer. *Physical review letters*, 85(2):461–464, 2000. (pages 3, 4, 6, 9, and 14)
- Anil K Seth. A Matlab toolbox for granger causal connectivity analysis. *Journal of neuroscience methods*, 186(2):262–273, 2010. (page 12)
- Anil K Seth, Adam B Barrett, and Lionel Barnett. Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, 35(8):3293–3297, 2015. (pages 4 and 11)
- Robert A Seymour, Gina Rippon, and Klaus Kessler. The detection of phase amplitude coupling during sensory processing. *Frontiers in neuroscience*, 11:487, 2017. (pages 8, 18, and 19)
- Wenbin Shi, Chien-Hung Yeh, and Yang Hong. Cross-frequency transfer entropy characterize coupling of interacting nonlinear oscillators in complex systems. *IEEE Transactions on Biomedical Engineering*, 66(2):521–529, 2018. (page 20)
- Wenbin Shi, Chien-Hung Yeh, and Jianping An. Cross-channel phase-amplitude transfer entropy conceptualize long-range transmission in sleep: a case study. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4048–4051. IEEE, 2019. (pages 8 and 20)
- Chaitra Sridhar, Shreya Bhat, U Rajendra Acharya, Hojjat Adeli, and G Muralidhar Bairy. Diagnosis of attention deficit hyperactivity disorder using imaging and signal processing techniques. *Computers in biology and medicine*, 88:93–99, 2017. (page 1)
- Floris Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981. (page 10)
- Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J Miller, Gernot Mueller-Putz, et al. Review of the BCI competition IV. *Frontiers in neuroscience*, 6(55):1–31, 2012. (pages 27 and 28)
- Nicholas M Timme and Christopher Lapish. A tutorial for information theory in neuroscience. *eNeuro*, 5(3):1–40, 2018. (pages 4, 5, and 6)

- Jlenia Toppi, Laura Astolfi, Monica Riseti, Alessandra Anzolin, Silvia E Kober, Guilherme Wood, and Donatella Mattia. Different topological properties of EEG-derived networks describe working memory phases as revealed by graph theoretical analysis. *Frontiers in Human Neuroscience*, 11:637, 2018. (page 29)
- Jennifer Townsend, Susan Y Bookheimer, Lara C Foland-Ross, Catherine A Sugar, and Lori L Altshuler. fMRI abnormalities in dorsolateral prefrontal cortex during a working memory task in manic, euthymic and depressed bipolar subjects. *Psychiatry Research: Neuroimaging*, 182(1):22–29, 2010. (page 104)
- Mauro Ursino, Giulia Ricci, and Elisa Magosso. 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. (pages 3, 4, 63, and 75)
- Raul Vicente, Michael Wibral, Michael Lindner, and Gordon Pipa. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of computational neuroscience*, 30:45–67, 2011. (pages 3, 4, 5, 6, 10, 35, 42, 45, 46, 57, and 68)
- Mario Villena-González, Iván Rubio-Venegas, and Vladimir López. Data from brain activity during visual working memory replicates the correlation between contralateral delay activity and memory capacity. *Data in brief*, 28:105042, 2020. (pages 30, 31, and 54)
- Edward K Vogel and Maro G Machizawa. Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428:748–751, 2004. (page 30)
- Shanshan Wang, Dajuan Zhang, Bei Fang, Xingping Liu, Guoli Yan, Guanghong Sui, Qingwei Huang, Ling Sun, and Suogang Wang. A study on resting EEG effective connectivity difference before and after neurofeedback for children with ADHD. *Neuroscience*, 457:103–113, 2021a. (pages 2, 7, and 19)
- Xiuli Wang, Bochao Cheng, Neil Roberts, Song Wang, Ya Luo, Fangfang Tian, and Suping Yue. Shared and distinct brain fMRI response during performance of working memory tasks in adult patients with schizophrenia and major depressive disorder. *Human brain mapping*, 42(16):5458–5476, 2021b. (page 104)
- Immo Weber, Esther Florin, Michael Von Papen, and Lars Timmermann. The influence of filtering and downsampling on the estimation of transfer entropy. *PloS one*, 12(11):e0188210, 2017. (pages 4, 5, 7, 8, 38, 39, 48, 64, 82, 90, and 115)
- Elvis Wianda and Bernhard Ross. The roles of alpha oscillation in working memory retention. *Brain and behavior*, 9(4):e01263, 2019. (pages 1 and 5)
- Michael Wibral, Nicolae Pampu, Viola Priesemann, Felix Siebenhühner, Hannes Seiwert, Michael Lindner, Joseph T Lizier, and Raul Vicente. Measuring information-transfer delays. *PloS one*, 8(2):e55809, 2013. (pages 2 and 10)
- Andreas Wilmer, Marc de Lussanet, and Markus Lappe. Time-delayed mutual information of the phase as a measure of functional connectivity. *PloS one*, 7(9):e44633, 2012. (page 18)

- Simon Wing, Kristin M Gunnarsdottir, Jorge Gonzalez-Martinez, and Sridevi V Sarma. Transfer entropy between intracranial EEG nodes highlights network dynamics that cause and stop epileptic seizures. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 6121–6125. IEEE, 2021. (page 2)
- Patricia Wollstadt, Mario Martínez-Zarzuela, Raul Vicente, Francisco J Díaz-Pernas, and Michael Wibral. Efficient transfer entropy analysis of non-stationary neural time series. *PloS one*, 9(7):e102833, 2014. (page 58)
- Ping Xie, Xiaohui Pang, Shengcui Cheng, Yuanyuan Zhang, Yinan Yang, Xiaoli Li, and Xiaoling Chen. Cross-frequency and iso-frequency estimation of functional corticomuscular coupling after stroke. *Cognitive Neurodynamics*, 15(3):439–451, 2021. (pages 7 and 18)
- Chunyao Xu, Chao Sun, Guoqian Jiang, Xiaoling Chen, Qun He, and Ping Xie. Two-level multi-domain feature extraction on sparse representation for motor imagery classification. *Biomedical Signal Processing and Control*, 62:102160, 2020. (pages 26 and 27)
- Pengbo Yang, Pengjian Shang, and Aijing Lin. Financial time series analysis based on effective phase transfer entropy. *Physica A: Statistical Mechanics and its Applications*, 468:398–408, 2017. (page 19)
- Pega Zarjam, Julien Epps, Fang Chen, and Nigel H Lovell. Estimating cognitive workload using wavelet entropy-based features during an arithmetic task. *Computers in biology and medicine*, 43(12):2186–2195, 2013. (page 34)
- Dan Zhang, Huipo Zhao, Wenwen Bai, and Xin Tian. Functional connectivity among multi-channel EEGs when working memory load reaches the capacity. *Brain research*, 1631:101–112, 2016. (pages 28 and 55)
- Yu Zhang, Chang S Nam, Guoxu Zhou, Jing Jin, Xingyu Wang, and Andrzej Cichocki. Temporally constrained sparse group spatial patterns for motor imagery BCI. *IEEE transactions on cybernetics*, 49(9):3322–3332, 2018. (page 52)
- Sa Zhou, Ping Xie, Xiaoling Chen, Yibo Wang, Yuanyuan Zhang, and Yihao Du. Optimization of relative parameters in transfer entropy estimation and application to corticomuscular coupling in humans. *Journal of neuroscience methods*, 308:276–285, 2018. (page 98)
- Jie Zhu, Jean-Jacques Bellanger, Huazhong Shu, and Régine Le Bouquin Jeannès. Contribution to transfer entropy estimation via the k-nearest-neighbors approach. *Entropy*, 17(6):4173–4201, 2015. (pages 3, 6, and 9)