

Article

Estimating Directed Phase-Amplitude Interactions from EEG Data through Kernel-Based Phase Transfer Entropy

Iván De La Pava Panche ^{1,*} , Viviana Gómez-Orozco ¹ , Andrés Álvarez-Meza ² , David Cárdenas-Peña ¹ 
and Álvaro Orozco-Gutiérrez ¹ 

¹ Automatic Research Group, Universidad Tecnológica de Pereira, Pereira 660003, Colombia; vigomez@utp.edu.co (V.G.-O.); dcardenasp@utp.edu.co (D.C.-P.); aaog@utp.edu.co (Á.O.-G.)

² Signal Processing and Recognition Group, Universidad Nacional de Colombia, Manizales 170003, Colombia; amalvarezme@unal.edu.co

* Correspondence: ide@utp.edu.co



Citation: De La Pava Panche, I.; Gómez-Orozco, V.; Álvarez-Meza, A.; Cárdenas-Peña, D.; Orozco-Gutiérrez A. Estimating Directed Phase-Amplitude Interactions from EEG Data through Kernel-Based Phase Transfer Entropy. *Appl. Sci.* **2021**, *11*, 9803. <https://doi.org/10.3390/app11219803>

Academic Editors: Maurizio Varanini, Alessandro Tonacci and Lucia Billeci

Received: 7 September 2021

Accepted: 18 October 2021

Published: 20 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Cross-frequency interactions, a form of oscillatory neural activity, are thought to play an essential role in the integration of distributed information in the brain. Indeed, 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 estimating such interactions is the use of transfer entropy (TE), a non-linear, information-theory-based effective connectivity measure. The conventional method involves feeding instantaneous phase and amplitude time series, extracted at the target frequencies, to a TE estimator. In this work, we propose that the problem of directed phase-amplitude interaction detection is recast 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 the common confounding factors that affect connectivity measures, such as the presence of high noise levels. We implement our proposal using a kernel-based TE estimator, defined in terms of Renyi's α entropy, which has successfully been used to compute single-trial phase TE. We tested our approach on the synthetic data generated through a simulation model capable of producing a time series with directed phase-amplitude interactions at two given frequencies, and on EEG data from a cognitive task designed to activate working memory, a memory system whose underpinning mechanisms are thought to include phase-amplitude couplings. Our proposal detected statistically significant interactions between the simulated signals at the desired frequencies for the synthetic data, identifying the correct direction of the interaction. It also displayed higher robustness to noise than the alternative methods. The results attained for the working memory data showed that the proposed approach codes connectivity patterns based on directed phase-amplitude interactions, that allow for the different cognitive load levels of the working memory task to be differentiated.

Keywords: EEG data; phase-amplitude interactions; cross-frequency interactions; transfer entropy; kernel methods; Renyi's entropy; connectivity analysis

1. Introduction

Biological neural systems exhibit rhythmic activity across many scales, from the spiking activity of individual neurons to the electric potentials generated by large populations of neurons, measurable from the scalp in the form of electroencephalographic (EEG) recordings [1]. Interactions between oscillations of different frequencies, known as cross-frequency couplings (CFCs), have been hypothesized to be directly related to the integration of distributed information in the brain [2] by regulating and synchronizing multi-scale communication within and across neural ensembles [3]. 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) [2–4]. These phase-amplitude interactions seem to be linked to normal and pathological brain processes in different mammalian species, including humans [5], and they were locally and interregionally observed across a wide range of cognitive tasks [6]. Theoretically, PAC allows for information transfer from large-scale brain networks associated with low-frequency oscillations to local, fast cortical processing areas exhibiting high-frequency activity [7].

Phase-amplitude interactions are commonly assessed using electrophysiological data through metrics of statistical dependency, such as the modulation index, mean vector length, and variations in the concept of mutual information [8–10]. However, they are unable to capture the directionality and delay of phase-amplitude interactions, quantities that are intrinsic to the concept of information being sent from one neural assembly to another [1,5,11]. A natural solution to this limitation, within the framework of information theory, would be to assess PACs using transfer entropy [3,9,11–13]. Transfer entropy (TE) is a model-free connectivity measure that estimates the directed interaction, or information flow, between two dynamical systems [14,15]. It is specially well-suited to exploratory analysis in neuroscience because of its ability to detect unknown non-linear interactions [16,17]. However, as a model-free information-theoretic measure, it does not capture the details of how the information transfer is carried out [1]. Standard TE will not reveal anything about the spectral characteristics of the interactions it detects. The most common approach explored in the literature to assess directed phase-amplitude interactions through TE involves bandpass-filtering the signals in the target frequency bands before the extraction of the instantaneous phase and amplitude time series, which are then fed to a TE analysis [5,18]. Nonetheless, it has been argued that filtering before TE computation negatively affects the results of TE [19,20], leading to false positives, and that it may not have the desired frequency-specific effects [1].

In this work, we reframe the problem of estimating directed phase-amplitude interactions through TE as the computation of TE between two instantaneous phase time series, known as phase TE [21], by borrowing the underlying premise of the cross-frequency directionality (CFD), a linear connectivity measure capable of estimating the PAC direction [11]. We previously addressed the problem of phase TE estimation from single trial data [22] and hypothesized 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 [23].

To test our proposal, we use a simulation model that allows generating synthetic data with unidirectionally phase-amplitude couplings at two target frequencies [11], and real EEG data from a change detection task with several difficulty levels designed to study visuospatial working memory [24,25]. Working memory (WM) is a memory system of limited capacity with the ability to store and manipulate information for a short period of time [26,27]. Current hypotheses for the mechanisms underpinning WM highlight the role of PAC [28], pointing to bidirectional interactions between the θ (4 Hz–7 Hz), α (8 Hz–12 Hz), and β bands (13 Hz–30 Hz) (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 [6,29]. 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 the cognitive load associated with a trial of the change-detection task to be detected.

The remainder of the paper is organized as follows: in Section 2, we present our proposal to estimate directed phase-amplitude interactions through a phase TE-based approach. Section 3 describes the two experiments we carried out to test the performance of our method. In Section 4, we present and discuss our results and, finally, we provide conclusions in Section 5.

2. Methods

2.1. Transfer Entropy

Transfer entropy (TE) is an information-theoretic measure of directed interactions between two dynamical systems [14,15]. Given two time-series $\mathbf{x} = \{x_t\}_{t=1}^T$ and $\mathbf{y} = \{y_t\}_{t=1}^T$, with $t \in \mathbb{N}$ a discrete time index, $T \in \mathbb{N}$, TE is defined as:

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

where $\mathbf{x}_t^{dx}, \mathbf{y}_t^{dy} \in \mathbb{R}^{D \times d}$ are time-embedded versions of $\mathbf{x}$ and $\mathbf{y}$, $D = T - (\tau(d-1))$ with $d, \tau \in \mathbb{N}$ the embedding dimension and delay, respectively; $u \in \mathbb{N}$ stands for the interaction delay between the two systems, and $p(\cdot)$ represents a probability density function [30]. The time embeddings are defined as $\mathbf{x}_t^d = (x(t), x(t-\tau), x(t-2\tau), \dots, x(t-(d-1)\tau))$ [31,32]. TE, as in Equation (1), evaluates the statistical causality from $\mathbf{x}$ to $\mathbf{y}$ by measuring whether the information contained in the previous $\mathbf{x}$, alongside that of the previous $\mathbf{y}$, is better at predicting the future of $\mathbf{y}$ than the information from the past of $\mathbf{y}$ alone. If this is the case, then $\mathbf{x}$ causes $\mathbf{y}$ (in the sense of Wiener's definition of causality [14]). For estimation convenience, we can also express TE as a linear combination of Shannon entropies:

$$TE(\mathbf{x} \rightarrow \mathbf{y}) = H_S(\mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}) - H_S(y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}) + H_S(y_t, \mathbf{y}_{t-1}^{dy}) - H_S(\mathbf{y}_{t-1}^{dy}), \quad (2)$$

where $H_S(X) = -\sum_x p(x) \log(p(x))$, X is a discrete random variable ($x \in X$), and $H_S(\cdot, \cdot)$, $H_S(\cdot)$ stand for joint and marginal entropies.

2.2. Transfer Entropy for Directed Phase-Amplitude Interactions

The conventional approach to estimate directed phase-amplitude interactions through TE consists of two stages. The first stage is a component extraction stage, which involves complex-filtering, or performing a phase/amplitude decomposition, to extract instantaneous phase and amplitude time series (see Figure 1A). Then, there is a TE computation stage that simply consists of estimating the information flow between the previously extracted data [5,18]. Formally, given two time series $\mathbf{x}$ and $\mathbf{y}$, 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 $\mathbf{s}_x(f_l) = \zeta^x e^{i\theta^x} \in \mathbb{C}^T$ and $\mathbf{s}_y(f_h) = \zeta^y e^{i\theta^y} \in \mathbb{C}^T$, which contain the filtered values of $\mathbf{x}$ and $\mathbf{y}$ at f_l and f_h , respectively; where $\theta^x, \theta^y \in [-\pi, \pi]_{t=1}^T$ are instantaneous phase time series, and $\zeta^x, \zeta^y \in \mathbb{R}^T$ are amplitude envelopes [21]. Then, we compute the desired TE as:

$$TE^{\theta\zeta}(\mathbf{x} \rightarrow \mathbf{y}, f_l, f_h) = H_S(\zeta_{t-1}^{y,dy}, \theta_{t-u}^{x,dx}) - H_S(\zeta_t^y, \zeta_{t-1}^{y,dy}, \theta_{t-u}^{x,dx}) + H_S(\zeta_t^y, \zeta_{t-1}^{y,dy}) - H_S(\zeta_{t-1}^{y,dy}), \quad (3)$$

where $\theta_t^{x,dx}$ and $\zeta_t^{y,dy}$ are time-embedded versions of θ^x and ζ^y .

2.3. 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 [11]. This is based on the phase slope index (PSI), which measures the coupling directionality between two oscillatory signals of similar frequencies [33]. 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 (ζ^y). Thus:

$$CFD(\mathbf{x} \rightarrow \mathbf{y}, f_h) = PSI(\mathbf{x} \rightarrow \zeta^y) = \Im \left(\sum_{f \in F} C_{\mathbf{x}\zeta}^*(f) C_{\mathbf{x}\zeta}(f + df) \right), \quad (4)$$

where $C_{x\zeta} = S_{x\zeta} / \sqrt{S_{xx} S_{\zeta\zeta}}$ is the complex coherence, $S_{x\zeta} \in \mathbb{C}$ is the cross-spectrum between x and ζ^y , $S_{xx}, S_{\zeta\zeta} \in \mathbb{C}$ are the auto-spectrums of x and ζ^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 [33]. Therefore, the CDF estimates the slope of the phase difference, between the phases of x and ζ^y , as a function of frequency. That is to say, it translates the problem of estimating the direction of phase-amplitude interactions to estimating interaction between phases.

2.4. Phase Transfer Entropy and Directed Phase-Amplitude Interactions

We begin by noting that the conventional approach to estimate directed phase-amplitude interactions through TE, as expressed by Equation (3), 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 ζ^y . In this work, we reformulate the problem of directed phase-amplitude interaction detection using TE as a phase TE-estimation task [21]. We do this by applying the idea behind the CFD: obtaining the directionality of interactions between phase and amplitude time series can be redefined as estimating the direction of interaction between two-phase time series. We hypothesize that such a change can improve the robustness of phase-amplitude TE estimates to signal degradation by noise and volume conduction effects.

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

$$TE^{\theta\theta^c}(x \rightarrow y, f_l, f_h) = H_S(\theta_{t-1}^{c,d\zeta}, \theta_{t-u}^{x,dx}) - H_S(\theta_t^c, \theta_{t-1}^{c,d\zeta}, \theta_{t-u}^{x,dx}) + H_S(\theta_t^c, \theta_{t-1}^{c,d\zeta}) - H_S(\theta_{t-1}^{c,d\zeta}), \quad (5)$$

where $\theta_t^{x,dx}$ and $\theta_t^{c,d\zeta}$ are time-embedded versions of θ^x and θ^c .

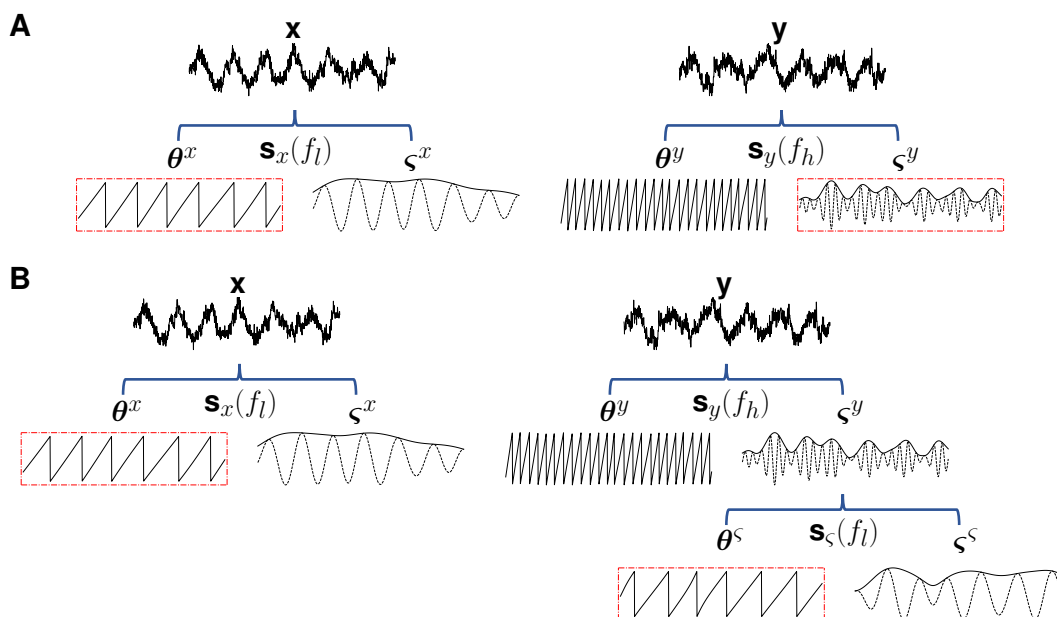


Figure 1. (A) The conventional approach to capture directed phase-amplitude interactions through TE consists of estimating TE from the instantaneous phase and amplitude time series that were extracted at frequencies f_l and f_h , respectively. (B) Our approach implies a further step, comprising the extraction of the phase of the amplitude time series at frequency f_l , to reformulate the problem as a phase TE estimation task.

The quantity in Equation (5) indicates the estimation of TE from two-phase time series extracted at the same frequency (f_l). Thus, it corresponds to the definition of phase TE, a phase-specific, nonlinear directed connectivity measure introduced in [21]. In that sense, a robust, kernel-based approach to estimating phase TE from single-trial data was recently proposed by our group [22]. Following this approach, we can redefine Equation (5) as:

$$TE_{\kappa\alpha}^{\theta\theta^c}(\mathbf{x} \rightarrow \mathbf{y}, f_l, f_h) = H_\alpha\left(\mathbf{K}_{\theta_{t-1}^c, d\zeta}, \mathbf{K}_{\theta_{t-u}^{x, dx}}\right) - H_\alpha\left(\mathbf{K}_{\theta_t^c}, \mathbf{K}_{\theta_{t-1}^c, d\zeta}, \mathbf{K}_{\theta_{t-u}^{x, dx}}\right) + H_\alpha\left(\mathbf{K}_{\theta_t^c}, \mathbf{K}_{\theta_{t-1}^c, d\zeta}\right) - H_\alpha\left(\mathbf{K}_{\theta_{t-1}^c, d\zeta}\right), \quad (6)$$

where $H_\alpha(\cdot)$ stands for the kernel-based formulation of Renyi's α entropy introduced in [34], $\mathbf{K}_{\theta_t^c}, \mathbf{K}_{\theta_{t-1}^c, d\zeta}, \mathbf{K}_{\theta_{t-u}^{x, dx}} \in \mathbb{R}^{(D-u) \times (D-u)}$ 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 $\mathbf{K}_{\theta_t^c}, \mathbf{a}_i, \mathbf{a}_j \in \mathbb{R}$ contain the values of the time series θ^c at times i and j . While for $\mathbf{K}_{\theta_{t-1}^c, d\zeta}$ 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^c and θ^x , $\theta_{t-1}^{c, d\zeta}$ and $\theta_t^{x, dx}$, respectively, at times i and j , in accordance with the time indexing of TE. Renyi's α entropy is a parametric family of entropies [35]:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log\left(\sum_x p(x)^\alpha dx\right), \quad (7)$$

with $\alpha \neq 1$ and $\alpha \geq 0$, and X a discrete random variable. This is a generalization of Shannon's entropy, and tends towards it in the limiting case where $\alpha \rightarrow 1$. Its kernel-based formulation can bypass the need for direct probability estimation from the data, instead relying on the kernel matrices that capture similarity relationships [34]. It is defined as:

$$H_\alpha(\mathbf{A}) = \frac{1}{1-\alpha} \log(\text{tr}(\mathbf{A}^\alpha)), \quad (8)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a kernel matrix containing elements $a_{ij} = \kappa(x_i, x_j)$, n is the number of realizations of X , and $\text{tr}(\cdot)$ stands for the matrix trace. We used this formulation, along with its definition for joint probability distributions:

$$H_\alpha(\mathbf{A}, \mathbf{B}) = H_\alpha\left(\frac{\mathbf{A} \circ \mathbf{B}}{\text{tr}(\mathbf{A} \circ \mathbf{B})}\right) = \frac{1}{1-\alpha} \log\left(\text{tr}\left(\left(\frac{\mathbf{A} \circ \mathbf{B}}{\text{tr}(\mathbf{A} \circ \mathbf{B})}\right)^\alpha\right)\right), \quad (9)$$

where $\mathbf{B} \in \mathbb{R}^{n \times n}$ is another kernel matrix, and the operator $\circ$ stands for the Hadamard product, to successfully and robustly estimate TE for real-valued and instantaneous phase time series [16,22]. In this study, we apply it in the context of directed phase-amplitude interaction estimation, following the TE estimation approach presented in Equation (6).

3. Experiments

3.1. Simulated Phase-Amplitude Interactions

3.1.1. Simulation Model

To evaluate the performance of our proposal, we generate simulated time series using a modified version of the PAC modeling strategy introduced in [11]. 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 sinusoidal signal period:

$$\mathbf{x}'_i = A_i(\sin(2\pi f_i \mathbf{t}_i + 1.5\pi) + 1), \quad (10)$$

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 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 concatenated the d segments to obtain a continuous signal of varying amplitude, as depicted in Figure 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:

$$\mathbf{y}' = \zeta \left(1 - \frac{1}{1 + \exp(-a(\mathbf{x}' - c))} \right) (\sin(2\pi f_h \mathbf{t}) + 1), \quad (11)$$

where $\mathbf{t} = \{0, dt, 2dt, \dots, (N-1)dt\}$, $\zeta = \sqrt{f_l/f_h}$, $c = 0.6$, and $a = 10$ (see Figure 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 imposed 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 Figure 2D–F).

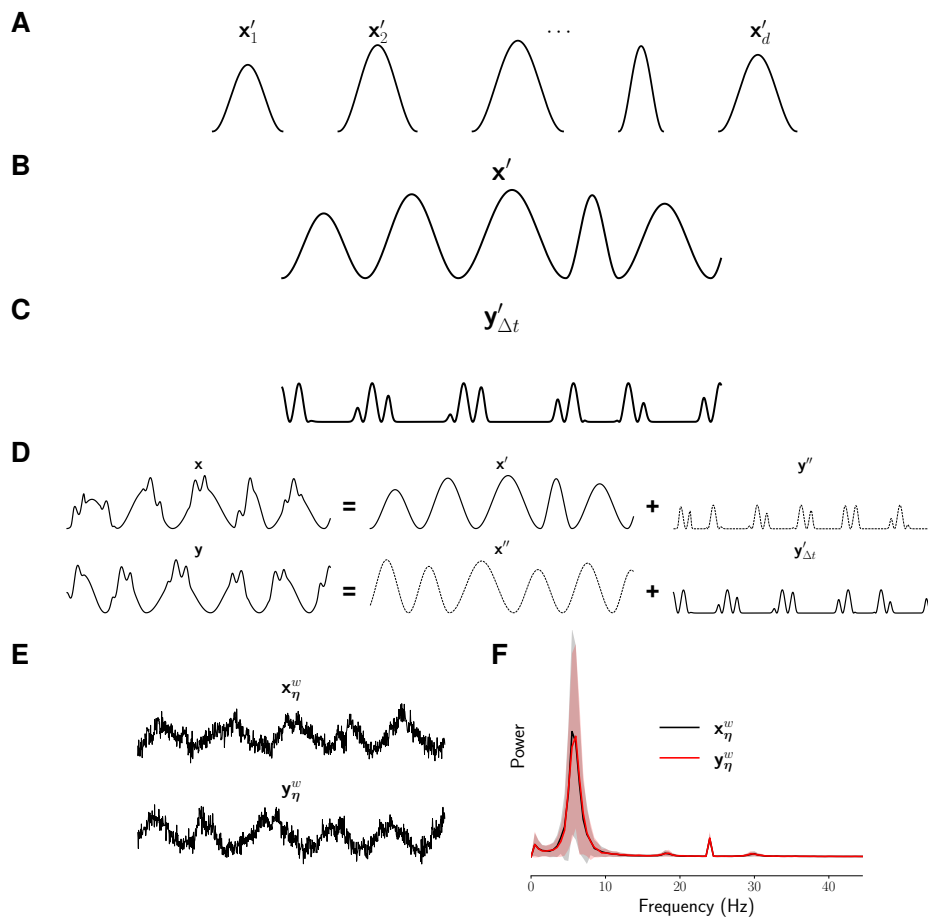


Figure 2. Schematic representation of the simulation model used to obtain data with directed phase-amplitude interactions. (A) Sinusoidal signal segments of varying amplitude and period ($\mathbf{x}'_i$). (B) A continuous, low-frequency signal ($\mathbf{x}'$) is obtained after concatenating the sinusoidal signal segments. (C) A high frequency signal ($\mathbf{y}'$) modulated by $\mathbf{x}'$ is computed through Equation (11). Then $\mathbf{y}'$ is time-shifted by Δt seconds ($\mathbf{y}'_{\Delta t}$). (D) $\mathbf{x}'$ and $\mathbf{y}'_{\Delta t}$ are combined with a pair of auxiliary, non-interacting signals $\mathbf{x}''$ and $\mathbf{y}''$ to generate signals $\mathbf{x}$ and $\mathbf{y}$. (E) Signals $\mathbf{x}$ and $\mathbf{y}$ are contaminated with noise and linearly mixed. (F) The resulting signals, $\mathbf{x}''_{\eta}$ and $\mathbf{y}''_{\eta}$, have very similar power spectra.

3.1.2. 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 timeshift of 20 ms. We detrend and normalized the simulated signals, before contaminating them with normalized white (η) and pink (η_p) noise, as follows:

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

where the parameter SNR controls the signal to noise ratio. We vary this to simulate low (SNR = 6), moderate (SNR = 3), and high (SNR = 1) noise conditions. For each scenario, we also mix the noisy signals, $\mathbf{x}_\eta$ and $\mathbf{y}_\eta$, aiming to reproduce the effects of volume conduction, by defining $\mathbf{x}_\eta^w = (1 - \frac{w}{2})\mathbf{x}_\eta + (\frac{w}{2})\mathbf{y}_\eta$, and $\mathbf{y}_\eta^w = (1 - \frac{w}{2})\mathbf{y}_\eta + (\frac{w}{2})\mathbf{x}_\eta$, with $w = 0.25$ the mixing strength. Then, we downsample $\mathbf{x}_\eta^w$ and $\mathbf{y}_\eta^w$ by a factor of 4 (we keep only every fourth sample) 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 2. Finally, to determine whether the estimated interactions are statistically significant, we perform permutation tests based on randomized surrogate trials [19,36] for each evaluated condition and frequency pair. The Bonferroni-corrected significance level for the tests is set at 4.4×10^{-5} .

3.2. Working Memory Data

3.2.1. Database

The database of brain activity during visual working memory [24] consists of the EEG data recorded from twenty-three subjects while they performed a change detection task [25] (<https://data.mendeley.com/datasets/j2v7btchdy/2>, accessed on 4 September 2021). The subjects had normal or corrected-to-normal vision, and did not have color-vision deficiency. The EEG data were acquired from 64 electrodes (Biosemi ActiveTwo), arranged according to the international 10/20 extended system, at a sampling rate of 2048 Hz. In addition to the EEG data, the database contains recordings from two external electrodes placed on the left and right mastoids, and four EOG channels. The goal of the change-detection task is to remember the colors of a set of squares (the memory array), and then compare them with the colors of a second set of squares (the test array), which appear in the same locations as the first set. The task has three levels: low-, medium-, and high-memory load. Each level has a different number of elements in the memory array: one square (low-memory load), two squares (medium-memory load), and four squares (high-memory load). At the beginning of each task trial, an arrow pointing to the left or right hemifield appears on the screen, signaling to the subject that they must remember only the stimuli that will be displayed on that side of the screen. Next, a memory array is presented for 0.1 s, followed by a retention period of 0.9 s. Afterward, a test array appears, and the subject reports whether the colors of the squares in the memory and test arrays are the same. Figure 3A depicts the task's experimental paradigm. Each subject performed 96 trials (32 trials for each memory load level). The colors of the squares in the test and memory arrays were different in 50% of the trials.

3.2.2. Preprocessing

First, the data were re-referenced to the average of the recordings from the electrodes located on the mastoids, bandpass-filtered between 0.01 Hz and 20 Hz using an order 2 Butterworth filter, and segmented using a 1.4 s squared window in order to extract the data corresponding to each trial. A trial segment started 0.2 s before the presentation of the memory array. Then, independent component analysis (ICA) was performed on the trial data to eliminate ocular artifacts using the fastICA algorithm, as implemented in the MNE python package, and the EOG information [24]. Next, all trials for which

the subjects incorrectly matched the memory and test arrays were dropped from further analysis. Afterward, the 32 channels ($C = 32$) depicted in Figure 3B were chosen from the 64 channels contained in the EEG data. Then, the trial data were downsampled by a factor of 2 (only every second sample is kept) to 1024 Hz, and underwent a final segmentation stage to select the part of the retention interval when the subject's working memories held the stimulus information. To that end, a 0.7 s long time window ($M = 717$) was used, starting 0.3 s after the onset of the memory array and ending before the test array appeared on the screen, as schematized in Figure 3A. Finally, to reduce the effects of volume conduction, the surface Laplacian of each trial was computed through the spherical spline method for source current density estimation [37–39].

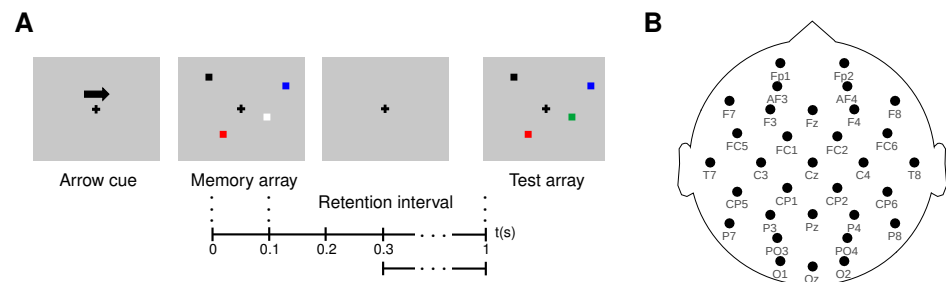


Figure 3. (A) Graphical representation of the change detection task for WM. (B) EEG channels selected from the montage used for the acquisition of the WM database.

Subjects number 11 and 17 were discarded during preprocessing because their data contained strong artifacts in a very large number of trials. Subjects number 22 and 23 were reassigned as subjects 17 and 11, respectively.

3.2.3. Classification Setup

Our aim is to set up a subject-dependent classification system (one classifier per subject) that allows the different levels of the change detection task, to be differentiated using relevant directed phase-amplitude interactions captured through the proposed $TE_{\kappa\alpha}^{\theta\theta^c}$ approach as inputs.

Feature Extraction

Let $\Psi = \{\mathbf{X}_n \in \mathbb{R}^{C \times M}\}_{n=1}^N$ be the EEG set holding N trials from the WM dataset, recorded from a single subject, with C as the number of channels and M as the number of samples of each trial. Let $\{l_n\}_{n=1}^N$ be a set whose n -th element corresponds to a label assigned to the n -th trial $\mathbf{X}_n$. The elements of l_n can take the values of 1, 2, and 3 to indicate low-, medium-, and high-memory loads, respectively. Our aim is to predict the label l_n from the transfer entropy features $(TE_{\kappa\alpha}^{\theta\theta^c})$ 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 TE 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 (6). 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 $\Lambda(f_c, f_{c'}) \in \mathbb{R}^{C \times C}$. 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, since activity in that frequency range has been associated with interactions between different brain regions during WM [6]. 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 $\Lambda(f_c, f_{c'})$ are averaged within each pairwise combination of bandwidths ($\theta - \alpha, \theta - \beta_l$, etc.). 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 $\Lambda' \in \mathbb{R}^{C \times C \times 6}$. For the N trials, a set of connectivity matrices $\{\Lambda'_n \in \mathbb{R}^{C \times C \times 6}\}_{n=1}^N$ is used. Then, by applying vector concatenation to Λ'_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 phase-amplitude TE measures. In practice, for each subject, the training data Φ have the following dimensions: N rows, which correspond to the number of trials of the task correctly performed by the subject ($N \leq 96$), and $P = C \times (C - 1) \times 6 = 5952$ features, where $C - 1$ accounts for the fact that the TE of a time series is not defined, and, therefore, the main diagonals of the Λ'_n matrices are empty (or all zeros, as in our case).

Feature Selection and Classification

In order to identify the TE features relevant to the discrimination between the cognitive load levels in the WM database, we rely on a relevance analysis methodology based on centered kernel alignment (CKA). CKA quantifies the similarity between two sample spaces by comparing two kernel functions that characterize each space [40]. This allows for a relevance vector index $q \in [0, 1]^P$ to be built, which can be used to rank the features in the characterizing matrix Φ according to their discrimination ability. Each element q_p in q signals how relevant the p -th feature is regarding its usefulness in distinguishing between the class labels in $\{l_n\}_{n=1}^N$, with higher values of q_p indicating higher relevance [22]. However, estimating CKA from data with a very low number of trials per class leads to unstable results. Given the fact that, in the WM database, each subject's dataset contains an average of less than 30 correct trials per class, we began by setting up an auxiliary cross-validation scheme to obtain an average q per subject, termed $\bar{q}$. This auxiliary cross-validation scheme has 10 iterations. For each iteration, 80% of the trials were randomly assigned to a training set and the remaining 20% were assigned to a validation set (80/20). After the 10 iterations were performed, $\bar{q}$ was obtained as the average of the q vectors computed in each iteration. Then, we set up a 10 iteration, 80/20, cross-validation scheme for use in subject-dependent classification. In this study, we used a support vector classifier (SVC) with an RBF kernel [41]. All classification parameters were tuned through a grid search and selected according to the classification accuracy, which was used as the system performance evaluation metric. For each iteration, we used $\bar{q}$ to rank the features in Φ from the highest to lowest relevance. Then, we selected a percentage of the ranked features, ranging from 5% to 100% in 5% increments, and input the chosen features to the classification algorithm, with the most relevant features being input first. The percentage of discriminant features was tuned along with the other classification parameters.

3.3. 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. The value of the embedding delay τ was set to 1 autocorrelation time (ACT) [31]. The embedding dimension d was automatically selected using Cao's criterion [36,42] from the range $d = \{1, 2, \dots, 10\}$ [22]. The parameter u was selected as the delay producing the largest TE from the following ranges: $u = \{4, 8, \dots, 40\}$ for the simulated data, and $u = \{50, 60, \dots, 250\}$ for the WM data. The parameter α in Renyi's entropy was set to 2, which is neutral to weighting, an a reasonable, a priori choice [16,34]. A radial basis function (RBF) kernel with Euclidean distance [43], defined as:

$$\kappa(\mathbf{a}_i, \mathbf{a}_j) = \exp\left(-\frac{d^2(\mathbf{a}_i, \mathbf{a}_j)}{2\sigma^2}\right), \quad (12)$$

was employed as kernel function, where $d^2(\cdot, \cdot)$ is a distance operator, and $\sigma \in \mathbb{R}^+$ is the kernel's bandwidth. The bandwidth σ was set in each case as the median distance of the data [44]. For CFD estimation, we used a sliding window 5 frequency bins long. Furthermore, for all measures, the required phase and amplitude decompositions we carried out by convolving the real-valued data with a Morlet wavelet, defined as:

$$h(t, f) = \exp(-t^2/2\zeta_t^2)\exp(i2\pi ft), \quad (13)$$

where f is for the filter frequency, $\xi_t = m/2\pi f$ is the standard deviation of the wavelet in the time domain, and m controls the time/frequency resolution [21]. The parameter m was varied from 3 to 10 in a logarithmic scale, according to the selected frequency of the filter. Finally, all connectivity values were obtained through in-house implementations of the algorithms for the different measures studied. The Python implementation of the proposed $TE_{\kappa\alpha}^{\theta\theta^c}$ approach is available at https://github.com/ide2704/Directed_PAC_through_kernel-based_Phase_Transfer_Entropy (accessed on 7 September 2021).

4. Results and Discussion

4.1. Simulated Data

Figure 4 presents the results obtained through the proposed $TE_{\kappa\alpha}^{\theta\theta^c}$ approach for the simulated data in the case where $SNR = 3$. Figure 4A shows the average values obtained for $TE_{\kappa\alpha}^{\theta\theta^c}(\mathbf{x}_\eta^w \rightarrow \mathbf{y}_\eta^w, f_l, f_h)$, 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 the TE values that were computed following our proposal, assuming that the oscillation phases in $\mathbf{x}_\eta^w$ drove the amplitude envelopes of the oscillations in $\mathbf{y}_\eta^w$. Similarly, Figure 4B shows the average obtained $TE_{\kappa\alpha}^{\theta\theta^c}(\mathbf{y}_\eta^w \rightarrow \mathbf{x}_\eta^w, f_l, f_h)$ values, where the underlying assumption is that oscillation phases in $\mathbf{y}_\eta^w$ are causal to the amplitude envelopes of the oscillations in $\mathbf{x}_\eta^w$. Figure 4C,D display the results returned by the permutation tests carried out over the TE data that were estimated for all the simulated trials, and whose average values are displayed in Figure 4A,B, 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. Statistically significant results were only obtained when TE was estimated from the phase of $\mathbf{x}_\eta^w$ in the θ band, at around 6 Hz, to the amplitude of $\mathbf{y}_\eta^w$ in the β band, at around 24 Hz, and not when estimated assuming causality in the opposite direction.

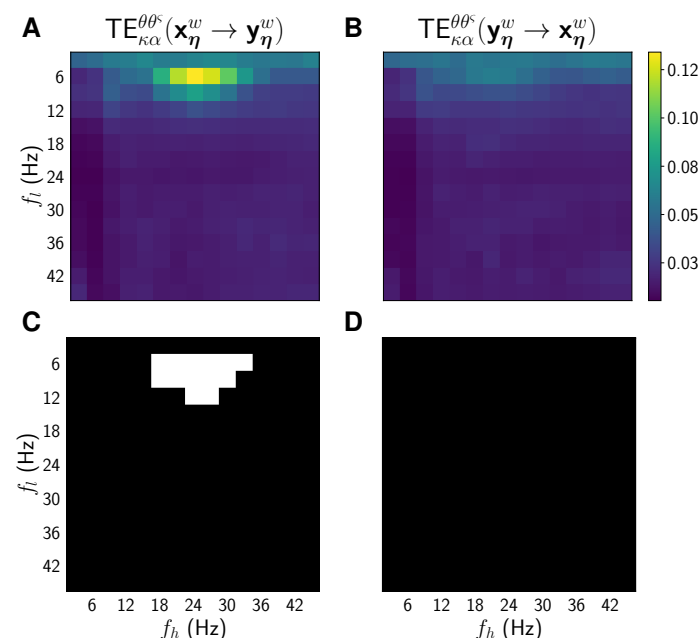


Figure 4. $TE_{\kappa\alpha}^{\theta\theta^c}$ results for the simulated data in the case when $SNR = 3$. (A) Average $TE_{\kappa\alpha}^{\theta\theta^c}(\mathbf{x}_\eta^w \rightarrow \mathbf{y}_\eta^w)$ values. (B) average $TE_{\kappa\alpha}^{\theta\theta^c}(\mathbf{y}_\eta^w \rightarrow \mathbf{x}_\eta^w)$ values. (C) Results of the permutation test performed on the $TE_{\kappa\alpha}^{\theta\theta^c}(\mathbf{x}_\eta^w \rightarrow \mathbf{y}_\eta^w)$ values. Statistically significant connectivities are indicated in white. (D) Results of the permutation test performed on the $TE_{\kappa\alpha}^{\theta\theta^c}(\mathbf{y}_\eta^w \rightarrow \mathbf{x}_\eta^w)$ values.

Figure 5 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, $TE_{\kappa\alpha}^{\theta^c}$ (conventional approach to estimate directed phase-amplitude through TE, using

the kernel-based Renyi's α TE estimator proposed in [16]), and $TE_{\kappa\alpha}^{\theta\theta\zeta}$, under the three modeled noise conditions, from left to right, low (SNR = 6), moderate (SNR = 3) and high (SNR = 1) noise levels. In this case, all connectivity measures were obtained assuming the correct direction of causality, from $\mathbf{x}_\eta^w$ to $\mathbf{y}_\eta^w$. 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 $TE_{\kappa\alpha}^{\theta\zeta}$ display significant results on narrower frequency ranges around the interacting frequencies that were actually present in the data than $TE_{\kappa\alpha}^{\theta\theta\zeta}$. This is likely to be associated with the additional filtering stages involved in the computation of $TE_{\kappa\alpha}^{\theta\theta\zeta}$ and, in the case of CFD, its high-frequency specificity is probably linked to its robustness against false positives [11]. However, unlike $TE_{\kappa\alpha}^{\theta\theta\zeta}$, both CFD and $TE_{\kappa\alpha}^{\theta\zeta}$ 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 and 5, we can argue that the proposed $TE_{\kappa\alpha}^{\theta\theta\zeta}$ approach allows for directed phase-amplitude interactions to be uncovered, 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.

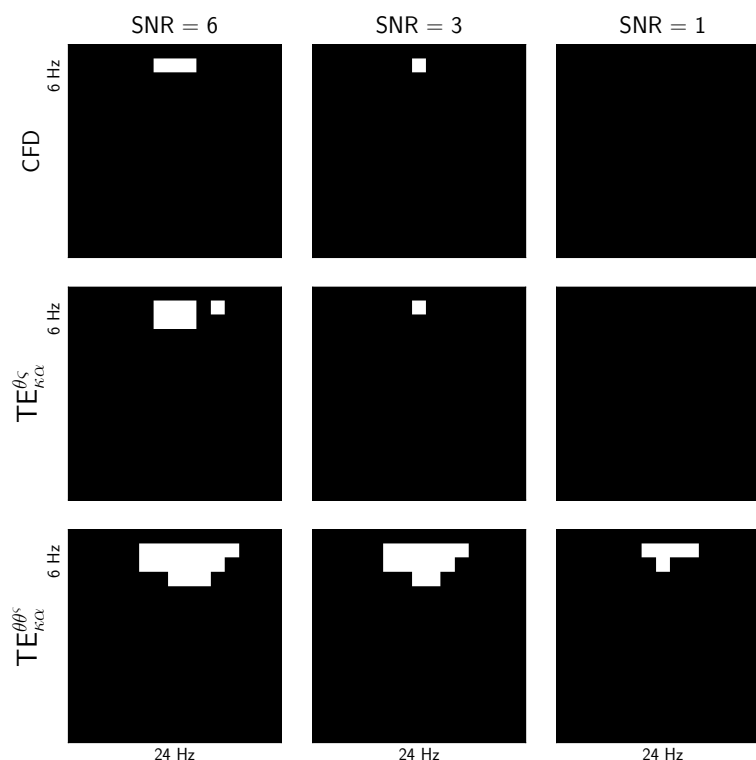


Figure 5. Results of the permutation tests carried out on the connectivity values estimated using CFD, $TE_{\kappa\alpha}^{\theta\zeta}$ and $TE_{\kappa\alpha}^{\theta\theta\zeta}$, for the three noise conditions modeled (SNR = [6, 3, 1]), assuming interactions in the simulated data from the phase of $\mathbf{x}_\eta^w$ to the amplitude of $\mathbf{y}_\eta^w$. Statistically significant results are shown in white.

4.2. Working Memory Data

The ability to store and manipulate information for short periods of time, provided by WM, plays a key role in complex cognitive tasks such as comprehension, reasoning, planning and learning [29,45]. WM consists of three distinct stages of information-processing: encoding, maintenance or retention, and retrieval [6], with the retention interval being considered as a defining component of WM, since it differentiates it from other memory types [27]. The most widely recognized model of WM [26] describes it as a several separate but interacting subsystems: a central component (central executive), two stimuli-dependent

storage subsystems (the phonological loop and the visuo-spatial sketchpad), and a system of limited capacity that allows for interaction between the other components (episodic buffer) [46]. Moreover, multiple studies have found that neural oscillatory activity in a wide range of frequencies is modulated during WM [47]. High-frequency activity, in the β and γ (>30 Hz) bands, seems to play a role in the encoding, retrieval, and maintenance of the stimulus; activity at lower frequencies, in the θ and α ranges, especially in frontal areas, is associated with the coordination and integration of different cognitive processes during the execution of WM tasks [48]. This has led to hypotheses about the 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 [12,49]. PAC interactions are thought to play a crucial role in WM [28], with bidirectional interactions linking the prefrontal cortex to parieto-occipital and medial temporal regions.[6,29]

In this study, we test the proposed $TE_{\kappa\alpha}^{\theta\theta^c}$ approach for the estimation of directed phase-amplitude interactions in the context of an important topic in the study of WM: how the characteristics of brain activity are modulated by changes in memory loads, as induced by WM tasks with different difficulty levels [48]. This topic is closely related to the concept of WM capacity, the amount of information that can be maintained and manipulated in WM [45]. WM capacity is, in turn, linked to important abilities, including non-verbal reasoning and control of attention, among others, and can be altered in people with psychiatric disorders [50].

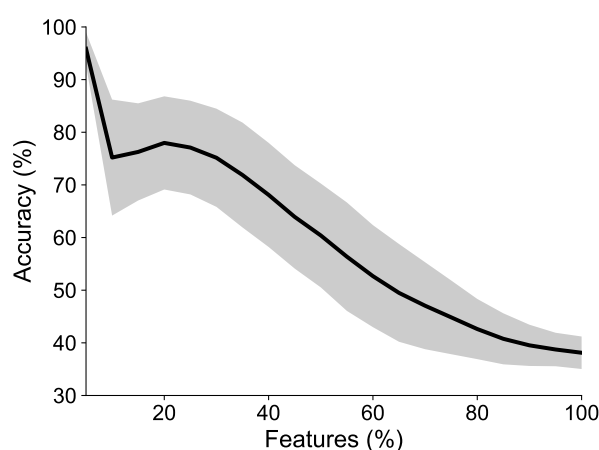


Figure 6. Average classification accuracy for all subjects as a function of the percentage of selected $TE_{\kappa\alpha}^{\theta\theta^c}$ features.

We built a subject-dependent classification system based on features extracted through the proposed $TE_{\kappa\alpha}^{\theta\theta^c}$ method from the EEG data recorded during the retention interval of a change-detection task: a task designed to test visuospatial working memory capacity. The goal of the classifier is to assign a particular cognitive load (number of elements in the memory array) to EEG data from a task trial characterized using our proposal. Figure 6 presents the classification accuracy for all subjects in the WM database as a function of the percentage of features used to train the classifier. The best results are obtained when 5% of the features are selected. These are those with the highest relevance values according to the CKA-based relevance vector $\bar{q}$, achieving 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. Figure 7 shows the highest accuracy obtained for each subject in the WM database, as well as the precision, recall, and F1 score values. Although the subjects differ in performance, the proposed classification system exhibits accuracies well above what would be expected in a three-class classification task for all of them. The precision, recall and F1 score values indicate that the accuracies obtained by the classifiers for each subject are not the result of class imbalances in the data or biased class selection

by the classifiers themselves. Additionally, all subjects, except subject 3, achieve peak performance when the classifier is trained with only 5% of the connectivity features (subject 3 does so when 10% of the features are selected). These results imply that our $TE_{\kappa\alpha}^{\theta\theta^c}$ characterization strategy successfully captures the discriminant directed phase-amplitude interactions 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 leads to a pronounced decrease in classification performance, as shown in Figure 6. This phenomenon can be explained because all channel connectivity analyses lead to datasets with a large number of features (see Section 3.2.3), 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 will be far away from each other, and that new instances will also be far away from those used to train the machine learning system. This makes it difficult to make good predictions [51]. In addition, many of the obtained connectivity features will not provide useful information to discriminate between the conditions of the cognitive paradigm of interest [16], and will only add 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 with classification performance.

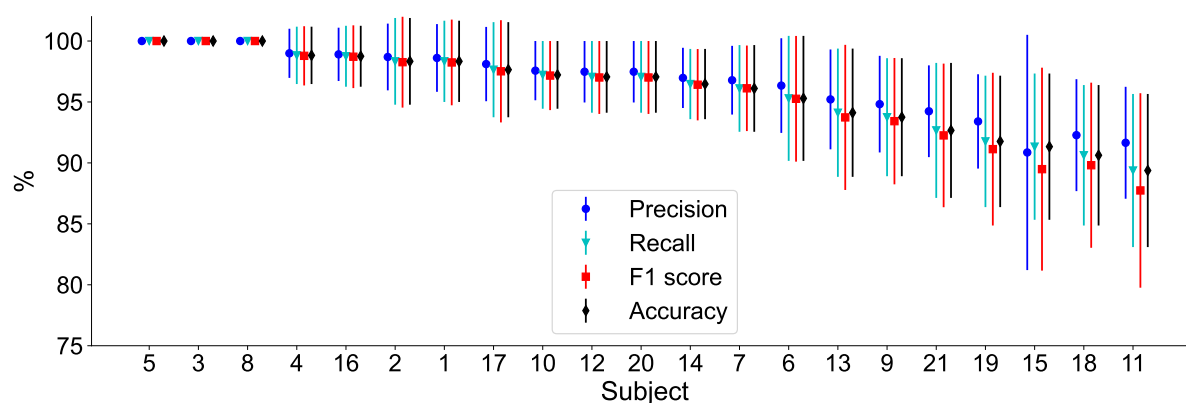


Figure 7. Highest average classification performance for each subject in the WM database. Mean $\pm$ standard deviation values are displayed for four classification performance measures: precision, recall, F1 score and accuracy. The subjects are ordered from the highest to lowest accuracy.

At this point, it is worth noting that the vector $\bar{q}$, 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 for classification. Therefore, our classification setup suffers from data leakage, and, as a consequence, the above results area is probably higher than that which 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 can successfully discriminate between the brain activity elicited at different levels of the change-detection task. Furthermore, the vector $\bar{q}$ provides valuable insights regarding the directed phase-amplitude interactions that arise while the subjects perform the task. It does this by assigning a relevance value q_p to each column, or feature, in the characterizing matrix Φ . In the case of our $TE_{\kappa\alpha}^{\theta\theta^c}$ analysis, q_p indicates whether a particular connectivity between two channels, for a specific frequency-band pair ($\theta - \alpha$, $\theta - \beta_1$, etc), allows for discrimination of the different cognitive loads of the task. A feature's discriminant ability is tantamount to its variation across classes. That is to say, the most relevant connectivities in $\bar{q}$ 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 activated during the task.

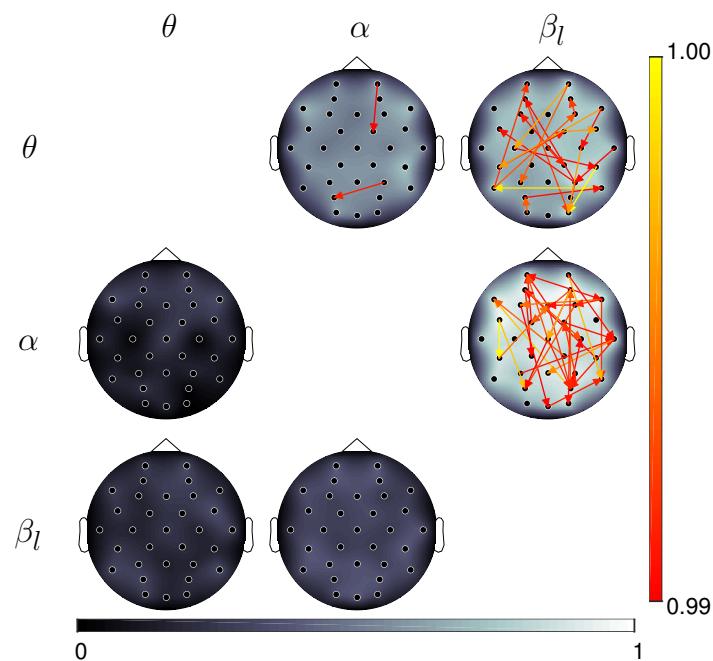


Figure 8. Topoplots of the average nodal (channel) relevance for each tested frequency band pair. The arrows represent the most relevant connections. For visualization purposes, only the connections with the highest average relevance values are depicted (1% of all connections). The rows indicate the frequency band driving the phase-amplitude interactions, while the columns correspond to the frequency band of the driven oscillations (e.g., the topoplot in the top right corner shows the nodal relevance and relevant connections obtained when $TE_{\kappa\alpha}^{\theta\theta^s}$ was computed, assuming directed interactions from the oscillations phases in the θ band to the amplitude envelope of oscillations in the β_l band).

Figure 8 displays the most relevant connectivities (average for all subjects), according to the relevance vector $\bar{q}$, discriminated by the frequency band pair. The background topoplots show the average nodal relevance, which corresponds to the relevance of the total information flow of every node. This is defined as the sum of all the q_p in $\bar{q}$ associated with all directed interactions originating from and targeting a specific node or EEG channel [22]. 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 way in terms of frequency. 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 the frontal and prefrontal areas and channels on the parietal and parieto-occipital areas. There are also several relevant long-range connections targeting the right temporal region. These results coincide with those previously obtained from the within-frequency, phase-based connectivity analysis carried out on the same database [22], as well as with studies that identified the presence of fronto-parietal and fronto-temporal interactions during cognitive tasks that activate visuo-spatial working memory [6,29,49].

4.3. Limitations

The results obtained for the simulated data show that our proposal is, unlike the other tested approaches, 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 to the analysis of noisy data, but care should be taken with results under more ideal conditions. The proposed approach

also relies on the kernel-based, phase TE-estimation method introduced in [22]; 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, as described in Section 3.3, but it is possible that more elaborate strategies [52] may lead to better results. Additionally, the TE estimator used assumes stationary or weakly non-stationary data, and cannot distinguish direct interactions from those originating from unobserved common causes [16]. 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 activities (phase-phase coupling, phase-frequency coupling, etc.) [3]. It was also not designed to be full information-theoretic analysis that simultaneously accounts for the multiple rhythms that can interact at both the sender and receiver side of coupled dynamical systems [1].

5. Conclusions

In this work, 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 this 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 an EEG database obtained under a WM paradigm. The obtained results for the synthetic data show that our approach successfully detects 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.

In future work, we will attempt to optimize the strategies for selecting the parameters involved in the filtering and time-embedding stages of the proposed approach and the kernel-based TE estimator.

Author Contributions: Conceptualization, I.D.L.P.P., V.G.-O. and A.Á.-M.; methodology, I.D.L.P.P., V.G.-O., and A.Á.-M.; software, I.D.L.P.P., A.Á.-M. and D.C.-P.; validation, I.D.L.P.P., A.Á.-M. and D.C.-P.; formal analysis, I.D.L.P.P., V.G.-O. and Á.O.-G.; investigation, I.D.L.P.P., V.G.-O. and A.Á.-M.; resources, I.D.L.P.P., A.Á.-M. and Á.O.-G.; data curation, I.D.L.P.P. and V.G.-O.; writing—original draft preparation, I.D.L.P.P., V.G.-O. and A.Á.-M.; writing—review and editing, A.Á.-M. and D.C.-P.; visualization, I.D.L.P.P.; supervision, Á.O.-G. and A.Á.-M.; project administration, Á.O.-G.; funding acquisition, I.D.L.P.P., V.G.-O. and A.Á.-M. All authors have read and agreed to the published version of the manuscript

Funding: Authors Iván De La Pava Panche and Viviana Gómez-Orozco were supported by the program “Doctorado Nacional en Empresa—Convocatoria 758 de 2016”, funded by Minciencias. A.M. Álvarez-Meza thanks the project *Prototipo de interfaz cerebro-computador multimodal para la detección de patrones relevantes relacionados con trastornos de impulsividad*-HERMES 50835, funded by Universidad Nacional de Colombia.

Institutional Review Board Statement: In this study, we use a public-access EEG database introduced in a previously published work, and made freely available by its authors [24]. We did not collect any data from human participants ourselves.

Informed Consent Statement: This study uses an anonymized public database introduced in a previously published work by another group [24].

Data Availability Statement: The database used in this study is public and can be found at the following link: database from brain activity during visual working memory <https://data.mendeley.com/datasets/j2v7btchdy/2> (accessed on 4 September 2021).

Conflicts of Interest: The authors declare that this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Pinzuti, E.; Wollstadt, P.; Gutknecht, A.; Tüscher, O.; Wibral, M. Measuring spectrally-resolved information transfer. *PLoS Comput. Biol.* **2020**, *16*, e1008526.
- Jirsa, V.; Müller, V. Cross-frequency coupling in real and virtual brain networks. *Front. Comput. Neurosci.* **2013**, *7*, 78.
- La Tour, T.D.; Tallot, L.; Grabot, L.; Doyère, V.; Van Wassenhove, V.; Grenier, Y.; Gramfort, A. Non-linear auto-regressive models for cross-frequency coupling in neural time series. *PLoS Comput. Biol.* **2017**, *13*, e1005893.
- Seymour, R.A.; Rippon, G.; Kessler, K. The detection of phase amplitude coupling during sensory processing. *Front. Neurosci.* **2017**, *11*, 487.
- Martínez-Cancino, R.; Delorme, A.; Wagner, J.; Kreutz-Delgado, K.; Sotero, R.C.; Makeig, S. What can local transfer entropy tell us about phase-amplitude coupling in electrophysiological signals? *Entropy* **2020**, *22*, 1262.
- Johnson, E.L.; Adams, J.N.; Solbakk, A.K.; Endestad, T.; Larsson, P.G.; Ivanovic, J.; Meling, T.R.; Lin, J.J.; Knight, R.T. Dynamic frontotemporal systems process space and time in working memory. *PLoS Biol.* **2018**, *16*, e2004274.
- Shi, W.; Yeh, C.H.; Hong, Y. Cross-frequency transfer entropy characterize coupling of interacting nonlinear oscillators in complex systems. *IEEE Trans. Biomed. Eng.* **2018**, *66*, 521–529.
- Cheng, N.; Li, Q.; Wang, S.; Wang, R.; Zhang, T. Permutation mutual information: A novel approach for measuring neuronal phase-amplitude coupling. *Brain Topogr.* **2018**, *31*, 186–201.
- Martínez-Cancino, R.; Heng, J.; Delorme, A.; Kreutz-Delgado, K.; Sotero, R.C.; Makeig, S. Measuring transient phase-amplitude coupling using local mutual information. *NeuroImage* **2019**, *185*, 361–378.
- Malladi, R.; Johnson, D.H.; Kalamangalam, G.P.; Tandon, N.; Aazhang, B. Mutual information in frequency and its application to measure cross-frequency coupling in epilepsy. *IEEE Trans. Signal Process.* **2018**, *66*, 3008–3023.
- Jiang, H.; Bahramisharif, A.; van Gerven, M.A.; Jensen, O. Measuring directionality between neuronal oscillations of different frequencies. *Neuroimage* **2015**, *118*, 359–367.
- Dimitriadis, S.I.; Sun, Y.; Thakor, N.V.; Bezerianos, A. Causal interactions between frontal θ –parieto-occipital α 2 predict performance on a mental arithmetic task. *Front. Hum. Neurosci.* **2016**, *10*, 454.
- Shi, W.; Yeh, C.H.; An, J. Cross-Channel Phase-Amplitude Transfer Entropy Conceptualize Long-Range Transmission in sleep: A case study. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 4048–4051.
- Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461.
- Zhu, J.; Bellanger, J.J.; Shu, H.; Le Bouquin Jeannès, R. Contribution to transfer entropy estimation via the k-nearest-neighbors approach. *Entropy* **2015**, *17*, 4173–4201.
- De La Pava Panche, I.; Alvarez-Meza, A.M.; Orozco-Gutierrez, A. A data-driven measure of effective connectivity based on Renyi's α -entropy. *Front. Neurosci.* **2019**, *13*, 1277.
- Ursino, M.; Ricci, G.; Magosso, E. Transfer Entropy as a Measure of Brain Connectivity: A Critical Analysis With the Help of Neural Mass Models. *Front. Comput. Neurosci.* **2020**, *14*, 45.
- Besserve, M.; Schölkopf, B.; Logothetis, N.K.; Panzeri, S. Causal relationships between frequency bands of extracellular signals in visual cortex revealed by an information theoretic analysis. *J. Comput. Neurosci.* **2010**, *29*, 547–566.
- Weber, I.; Florin, E.; Von Papen, M.; Timmermann, L. The influence of filtering and downsampling on the estimation of transfer entropy. *PLoS ONE* **2017**, *12*, e0188210.
- Chen, X.; Zhang, Y.; Cheng, S.; Xie, P. Transfer spectral entropy and application to functional corticomuscular coupling. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1092–1102.
- Lobier, M.; Siebenhühner, F.; Palva, S.; Palva, J.M. Phase transfer entropy: A novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. *Neuroimage* **2014**, *85*, 853–872.
- De La Pava Panche, I.; Álvarez-Meza, A.; Herrera Gómez, P.M.; Cárdenas-Peña, D.; Ríos Patiño, J.I.; Orozco-Gutiérrez, Á. Kernel-Based Phase Transfer Entropy with Enhanced Feature Relevance Analysis for Brain Computer Interfaces. *Appl. Sci.* **2021**, *11*, 6689.
- Bastos, A.M.; Schoffelen, J.M. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front. Syst. Neurosci.* **2016**, *9*, 175.
- Villena-González, M.; Rubio-Venegas, I.; López, V. Data from brain activity during visual working memory replicates the correlation between contralateral delay activity and memory capacity. *Data Brief* **2020**, *28*, 105042.
- Vogel, E.K.; Machizawa, M.G. Neural activity predicts individual differences in visual working memory capacity. *Nature* **2004**, *428*, 748–751.
- Baddeley, A. Working memory: Theories, models, and controversies. *Annu. Rev. Psychol.* **2012**, *63*, 1–29.
- Pavlov, Y.G.; Kotchoubey, B. Oscillatory brain activity and maintenance of verbal and visual working memory: A systematic review. *Psychophysiology* **2020**, e13735. <https://doi.org/10.1111/psyp.13735>
- Liang, W.K.; Tseng, P.; Yeh, J.R.; Huang, N.E.; Juan, C.H. Frontoparietal beta amplitude modulation and its interareal cross-frequency coupling in visual working memory. *Neuroscience* **2021**, *460*, 69–87.
- Johnson, E.L.; King-Stephens, D.; Weber, P.B.; Laxer, K.D.; Lin, J.J.; Knight, R.T. Spectral imprints of working memory for everyday associations in the frontoparietal network. *Front. Syst. Neurosci.* **2019**, *12*, 65.

30. Wibral, M.; Pampu, N.; Priesemann, V.; Siebenhühner, F.; Seiwert, H.; Lindner, M.; Lizier, J.T.; Vicente, R. Measuring information-transfer delays. *PLoS ONE* **2013**, *8*, e55809.
31. Vicente, R.; Wibral, M.; Lindner, M.; Pipa, G. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* **2011**, *30*, 45–67.
32. Takens, F. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*; Springer: Berlin/Heidelberg, Germany, 1981; pp. 366–381.
33. Nolte, G.; Ziehe, A.; Nikulin, V.V.; Schlögl, A.; Krämer, N.; Brismar, T.; Müller, K.R. Robustly estimating the flow direction of information in complex physical systems. *Phys. Rev. Lett.* **2008**, *100*, 234101.
34. Giraldo, L.G.S.; Rao, M.; Principe, J.C. Measures of entropy from data using infinitely divisible kernels. *IEEE Trans. Inf. Theory* **2015**, *61*, 535–548.
35. Rényi, A. 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: Oakland, CA, USA, 1961.
36. Lindner, M.; Vicente, R.; Priesemann, V.; Wibral, M. TRENTOOL: A Matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC Neurosci.* **2011**, *12*, 119.
37. Perrin, F.; Pernier, J.; Bertrand, O.; Echallier, J. Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* **1989**, *72*, 184–187.
38. Cohen, M.X. Comparison of different spatial transformations applied to EEG data: A case study of error processing. *Int. J. Psychophysiol.* **2015**, *97*, 245–257.
39. Rathee, D.; Cecotti, H.; Prasad, G. Single-trial effective brain connectivity patterns enhance discriminability of mental imagery tasks. *J. Neural Eng.* **2017**, *14*, 056005.
40. Cortes, C.; Mohri, M.; Rostamizadeh, A. Algorithms for learning kernels based on centered alignment. *J. Mach. Learn. Res.* **2012**, *13*, 795–828.
41. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
42. Cao, L. Practical method for determining the minimum embedding dimension of a scalar time series. *Phys. D Nonlinear Phenom.* **1997**, *110*, 43–50.
43. Liu, W.; Principe, J.C.; Haykin, S. *Kernel Adaptive Filtering: A Comprehensive Introduction*; John Wiley & Sons: Hoboken, NJ, USA, 2011; Volume 57.
44. Schölkopf, B.; Smola, A.J.; Bach, F. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT press: Cambridge, MA, USA, 2002.
45. Zhang, D.; Zhao, H.; Bai, W.; Tian, X. Functional connectivity among multi-channel EEGs when working memory load reaches the capacity. *Brain Res.* **2016**, *1631*, 101–112.
46. Toppi, J.; Astolfi, L.; Riseti, M.; Anzolin, A.; Kober, S.E.; Wood, G.; Mattia, D. Different topological properties of EEG-derived networks describe working memory phases as revealed by graph theoretical analysis. *Front. Hum. Neurosci.* **2018**, *11*, 637.
47. Daume, J.; Gruber, T.; Engel, A.K.; Frieze, U. Phase-amplitude coupling and long-range phase synchronization reveal frontotemporal interactions during visual working memory. *J. Neurosci.* **2017**, *37*, 313–322.
48. Dai, Z.; De Souza, J.; Lim, J.; Ho, P.M.; Chen, Y.; Li, J.; Thakor, N.; Bezerianos, A.; Sun, Y. EEG cortical connectivity analysis of working memory reveals topological reorganization in theta and alpha bands. *Front. Hum. Neurosci.* **2017**, *11*, 237.
49. Dimitriadis, S.; Sun, Y.; Laskaris, N.; Thakor, N.; Bezerianos, A. 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.* **2016**, *24*, 1017–1028.
50. Constantinidis, C.; Klingberg, T. The neuroscience of working memory capacity and training. *Nat. Rev. Neurosci.* **2016**, *17*, 438–449.
51. Géron, A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media: Sebastopol, CA, USA, 2019.
52. Zhou, S.; Xie, P.; Chen, X.; Wang, Y.; Zhang, Y.; Du, Y. Optimization of relative parameters in transfer entropy estimation and application to corticomuscular coupling in humans. *J. Neurosci. Methods* **2018**, *308*, 276–285.