Brain Topography (2019) 32:882–896
https://doi.org/10.1007/s10548-019-00718-8

ORIGINAL PAPER

Statistical Significance Assessment of Phase Synchrony in the Presence of Background Couplings: An ECoG Study

Parham Mostame[1] · Ali Moharramipour[1] · Gholam-Ali Hossein-Zadeh[1] · Abbas Babajani-Feremi[2,3,4]

Received: 4 October 2018 / Accepted: 13 May 2019 / Published online: 25 May 2019

Abstract

Statistical significance testing is a necessary step in connectivity analysis. Several statistical test methods have been employed to assess the significance of functional connectivity, but the performance of these methods has not been thoroughly evaluated. In addition, the effects of the intrinsic brain connectivity and background couplings on performance of statistical test methods in task-based studies have not been investigated yet. The background couplings may exist independent of cognitive state and can be observed on both pre- and post-stimulus time intervals. The background couplings may be falsely detected by a statistical test as task-related connections, which can mislead interpretations of the task-related functional networks. The aim of this study was to investigate the relative performance of four commonly used non-parametric statistical test methods—surrogate, demeaned surrogate, bootstrap resampling, and Monte Carlo permutation methods—in the presence of background couplings and noise, with different signal-to-noise ratios (SNRs). Using simulated electrocorticographic (ECoG) datasets and phase locking value (PLV) as a measure of functional connectivity, we evaluated the performances of the statistical test methods utilizing sensitivity, specificity, accuracy, and receiver operating curve (ROC) analysis. Furthermore, we calculated optimal p values for each statistical test method using the ROC analysis, and found that the optimal p values were increased by decreasing the SNR. We also found that the optimal p value of the bootstrap resampling was greater than that of other methods. Our results from the simulation datasets and a real ECoG dataset, as an illustrative case report, revealed that the bootstrap resampling is the most efficient non-parametric statistical test for identifying the significant PLV of ECoG data, especially in the presence of background couplings.

Keywords Functional connectivity · Phase locking value (PLV) · Background couplings · Statistical test · Electrocorticography (ECoG)

Introduction

Cognitive functions of the brain are derived by brain networks (Fries 2005; Koutsoukos et al. 2015; Varela et al. 2001). Brain networks have been investigated using functional connectivity based on Blood Oxygen-Level Dependent (BOLD) signal or electrophysiological signals, such as electrocorticographic (ECoG) recordings, during rest or cognitive tasks (Micheli et al. 2015; Srinivasan et al. 2007; Takahashi et al. 2018; Chen et al. 2018). Electrophysiological signals have a high temporal resolution and are an appropriate choice for investigating dynamics of functional connectivity. Various measures have been introduced to estimate functional connectivity of electrophysiological signals: coherency, imaginary part of coherency (ImC), phase locking value (PLV), phase slope index (PSI), phase lag index (PLI), and weighted PLI (wPLI) (Lachaux et al. 1999; Nolte

Handling Editor: Kevin Whittingstall.

Parham Mostame and Ali Moharramipour contributed equally to this work.

✉ Abbas Babajani-Feremi
ababajan@uthsc.edu

[1] School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran

[2] Department of Pediatrics, University of Tennessee Health Science Center, Memphis, TN, USA

[3] Department of Anatomy and Neurobiology, University of Tennessee Health Science Center, Memphis, TN, USA

[4] Neuroscience Institute, Le Bonheur Children's Hospital, Memphis, TN, USA

et al. 2004; Nolte et al. 2008; Stam et al. 2007; Vinck et al. 2011).

The performance of the functional connectivity measures derived from electrophysiological signals can be affected by several factors, such as signal-to-noise ratio (SNR), strength of connections, volume conduction effect, and artifacts (Bastos and Schoffelen 2016; Gordon et al. 2013; Greenblatt et al. 2012). In addition, the significant connections of functional brain networks may not be similarly identified if different statistical significance test methods are utilized. Particularly in cases with background couplings, some of these couplings may be falsely identified by a statistical test that yields misleading interpretations of the brain networks in task-based studies. Moreover, the calculated null distribution by a statistical test can be severely affected by the presence of background couplings, and thus the final results obtained from different statistical test methods may differ significantly. Although several studies have compared performances of different functional connectivity measures (Bastos and Schoffelen 2016; Greenblatt et al. 2012), there have been limited studies that evaluated the performance of statistical tests, and no study has yet evaluated the performance of various statistical test methods in the presence of background couplings.

There have been many studies reporting presence of intrinsic brain networks in the literature of resting-state functional magnetic resonance imaging (fMRI) (Cohen 2017; Sadaghiani et al. 2015; Barbey 2018). Similarly, few studies have recently reported presence of intrinsic brain networks in resting-state ECoG data (Fox et al. 2018; Kucyi et al. 2018). The resting-state intrinsic brain networks can also be present across a wide variety of task states (Krienen et al. 2014; Cole et al. 2014). Background coupling can be defined as any connection within intrinsic brain networks, independent of cognitive state. Such connections, that may be persistently present on both pre- and post-stimulus intervals, are not directly associated with a cognitive task and should be ignored by the statistical test in a task-based study. In this regard, distinguishing background couplings from task-related connections is a great challenge for statistical tests. Addressing this challenge by using a statistical test provides a precise and reliable functional brain network in task-based studies.

The standard practice is to use a pre-defined threshold corresponding to a significance level (usually $p < 0.05$) in functional connectivity studies. It is noteworthy that the detected brain networks change by the selection of different significance levels (i.e. p values), especially in high gamma band or with low SNR signals, where the values of connectivity are weak. Finding an optimal value for the significance level of the various statistical tests is very important, although this is a challenging task and no study has yet addressed this issue.

Two types of statistical tests have been utilized in the literature: parametric tests and non-parametric tests. Non-parametric tests are data-driven and parametric tests use a known probability density function (PDF) of the measure under test. Although parametric tests are more time-efficient, it has been shown that non-parametric tests are more accurate and have been widely applied to various kinds of studies (Maris and Oostenveld 2007). Several non-parametric statistical tests have been introduced and employed in previous studies (Micheli et al. 2015; Lachaux et al. 1999; Stam et al. 2007; Gordon et al. 2013; Maris and Oostenveld 2007; Guthrie and Buchwald 1991; Phillips et al. 2014; Theiler et al. 1992), although performances of these tests in extracting significant functional connectivity have not been evaluated, especially in the presence of background couplings. We evaluated the performances of the following four non-parametric statistical tests which have been frequently used in the literature of functional connectivity (Micheli et al. 2015; Lachaux et al. 1999; Stam et al. 2007; Gordon et al. 2013; Maris and Oostenveld 2007; Guthrie and Buchwald 1991; Phillips et al. 2014; Theiler et al. 1992): phase permuted surrogate data method, demeaned phase permuted surrogate data method, bootstrap resampling, and Monte Carlo permutation. For the ease of notations, we call the phase permuted surrogate data method the "surrogate" method. In addition, the demeaned phase permuted surrogate data method, here denoted as "demeaned surrogate" method, has been used in some studies instead of the regular surrogate method (Hagiwara et al. 2014). In the following sections, we first introduce these well-known non-parametric statistical tests. Then we compare the performances of these statistical tests based on sensitivity, specificity, and receiver operating curve (ROC) analysis using simulated ECoG signals. Then in our illustrative sample of real data, we utilized PLV in high gamma band (60–110 Hz) as a measure of functional connectivity. After identifying significant PLV by each statistical test method, the number of false connections detected in the pre-stimulus interval (for a wide range of p values) was used to evaluate the performance of each method in distinguishing background couplings from task-related connectivity.

Materials and Methods

Statistical Tests

Surrogate Method

In this method, a specific number (e.g. $n = 1000$) of surrogate datasets is created by random permutation of phase of electrophysiological signals. Then connectivity measures for each of these datasets are calculated (Maris and Oostenveld 2007; Astolfi et al. 2009) and used to build an empirical null

probability density function (PDF) of connectivity between each pair of connections. This empirical PDF is then used to detect the significant connections based on a pre-specified p value. It is noteworthy that the null PDF of the functional connectivity measures related to phase synchrony of electrophysiological signals is expected to have a mean close to zero in the surrogate data.

Demeaned Surrogate Method

This method is similar to the surrogate method, with the exception that the calculated connectivity in every step is demeaned with respect to its baseline value (Hagiwara et al. 2014):

$$Conn_{demeaned}(elec,f,t) = Conn(elec,f,t) - < Conn(elec,f,t)_{t\in Baseline} >_t \quad (1)$$

where $Conn(elec,f,t)$ represents the value of a connectivity measure between electrode pair $elec$ at frequency f and time point t, $Conn_{demeaned}(elec,f,t)$ represents demeaned connectivity measure, and $Baseline$ is a time interval before stimulus onset in task-based study. After calculating a null PDF for $Conn_{demeaned}(elec,f,t)$, significant connections are identified using a specified p value. As shown in Eq. (1), this method can remove baseline connectivity that might exist in post-stimulus. Thus, it is expected that this method works better than the surrogate method in distinguishing the background couplings from the task-related connections.

Bootstrap Resampling Method

Suppose that a dataset (collected during a task) has N original trials. In each realization of the bootstrap resampling method, N trials are randomly selected with replacement from the original trials. The null PDF of a functional connectivity measure in the pre-stimulus interval is calculated by considering a large number of bootstrap realizations of pre-stimulus interval. Once the null PDF is obtained, a specific threshold value is extracted by means of a specified p value, and then the significant connections are identified (Efron 1982; Porcaro et al. 2013; Sekihara et al. 2004).

Monte Carlo Permutation Method

This method extracts significant connections based on calculating the difference between two cognitive conditions (Micheli et al. 2015; Maris and Oostenveld 2007). Suppose that a dataset has N trials and two conditions A and B which represent two cognitive states (e.g. pre- and post-stimulus). Assume that $\mathrm{Conn}^{A,B}_{\mathrm{diff}}(elec,f,t)$ represents the difference of connectivity between two conditions (A and B) in electrode pair $elec$ at frequency f and time point t. To form the null distribution, permuted datasets are generated by shuffling the order of conditions over trials randomly. Consider that $X(t)$ consists of N trials of data and $\left\{X(t_{A_1}), X(t_{A_2}), \ldots, X(t_{A_N})\right\}$ and $\left\{X(t_{B_1}), X(t_{B_2}), \ldots, X(t_{B_N})\right\}$ represent segments of data corresponding to conditions A and B, respectively. A permutation dataset is generated by randomly switching segments of conditions A and B among trials. For example, a realization of permutation dataset can be $\left\{X(t_{A_1}), X(t_{B_2}), X(t_{A_3}), \ldots, X(t_{B_N})\right\}$ in new condition A′ and $\left\{X(t_{B_1}), X(t_{A_2}), X(t_{B_3}), \ldots, X(t_{A_N})\right\}$ in new condition B′, and the connectivity difference between two conditions is calculated as $\mathrm{Conn}^{A',B'}_{\mathrm{diff}}(elec,f,t)$. When connectivity differences are calculated for a specific number of permuted datasets (e.g. $n = 1000$), the number of surrogate datasets with values of $\mathrm{Conn}^{A',B'}_{\mathrm{diff}}(elec,f,t)$ larger than $\mathrm{Conn}^{A,B}_{\mathrm{diff}}(elec,f,t)$ will be calculated and given as $\mathrm{M}(elec,f,t)$. The significant $\mathrm{Conn}^{A,B}_{\mathrm{diff}}(elec,f,t)$ is identified based on the following equation:

$$\frac{M(elec,t,f)}{n} < p\ value \quad (2)$$

where n is the total number of permuted datasets.

Simulations

In order to evaluate performance of the statistical tests introduced in the previous section, we simulated ECoG signals by manipulating the strength of connections over time and frequency. Different aspects of the simulated ECoG signals are specified below.

Simulated ECoG Signals

We simulated ECoG signals in three electrodes by mixing various components that exist in real ECoG data (Stephen et al. 2014) (Fig. 1). For each electrode, an ECoG signal with sampling rate of 1 kHz was simulated over 80 trials by mixing four components. The time range of trials was set from −1 to 2 s including three time intervals: baseline (−1 to 0 s), and two different post-stimulus intervals (0 to 1 s and 1 to 2 s). Descriptions for the four ECoG components are presented here.

1/f Component ("1/f")

It has been shown that the power spectral density of ECoG data has a shape similar to 1/f curve (He et al. 2010; Miller et al. 2009). We used a pink noise with mean value equal to 0 and standard deviation equal to 4 to model the 1/f characteristic of the ECoG data and refer to it as 1/f component hereafter (Fig. 1a).

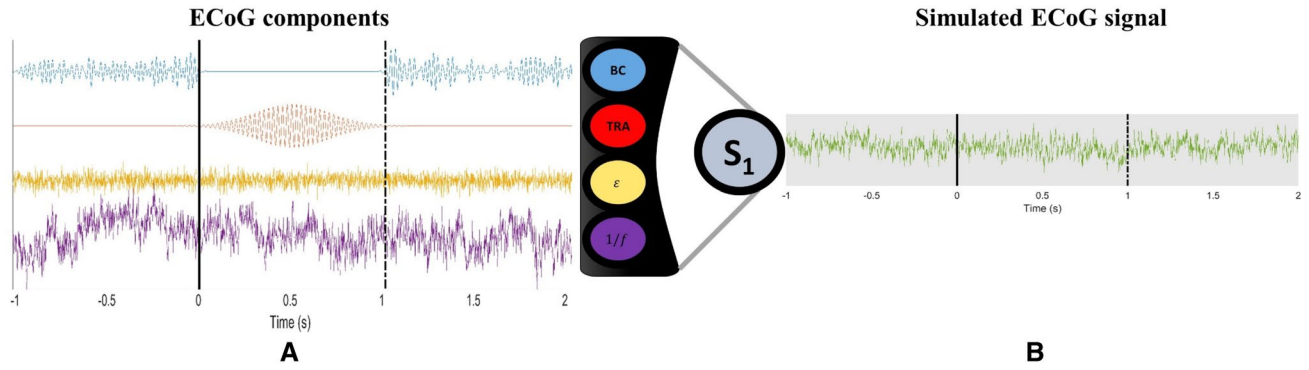

Fig. 1 Illustration of **a** four ECoG components including background couplings ("BC"), task-related activity ("TRA"), measurement noise ("ε"), and "$1/f$" components (top to bottom); and **b** a single-trial simulated ECoG signal (S_1) generated by adding all four components. Horizontal axis represents time with respect to stimulus onset and vertical axis shows amplitude of signals. The vertical solid black line at 0 s and the dashed vertical dashed black line at 1 s represent the stimulus onset and middle of the post-stimulus interval, respectively

Task-Related Activity Component ("TRA")

We used a Gaussian-tapered sinusoidal wave with a random phase shift over trials to simulate task-related connectivity in the simulated ECoG signals (Fig. 1a). Aforementioned random phase shift over trials was responsible for controlling the strength of such task-related connectivity. This random phase shift, denoted as φ, had a uniform random distribution with an adjustable variance over trials. The strength of connectivity was associated with the phase variance in that a larger phase variance generated a weaker connection. In each simulated connection between two electrodes, φ was set to zero for all trials of one electrode and to a random value with a known uniform distribution over trials in another electrode. The mean of this uniform distribution was set to zero and its variance (σ^2) was modified to generate different strengths for connectivity. The following equation represents the relationship between the desired connectivity in two electrodes and the phase difference ($\Delta\varphi$) between signals of the electrodes, considering the PLV as a measure of functional connectivity:

Background Coupling Component ("BC")

This component represents the background couplings. Since the background couplings should occur randomly over the time–frequency domain, we used a band-limited white noise to simulate these connections (Fig. 1a). In the same manner as with the "TRA" component, the background couplings were created by using a random phase difference between signals of two electrodes over trials. Strengths of the background couplings were considered to be equal to 0.3 ($\sigma = 1.36$ in Eq. (3)). In our simulations, we assumed that background couplings are suppressed when task-related connections are established. Thus, the "BC" component was suppressed using a rectangular time window, smoothed by a Gaussian function, during activation of task. Moreover, the power ratio between the "BC" component and the "1/f" component was set from 1 to 4.

Measurement Noise Component ("ε")

We modeled the measurement noise in ECoG data by using a white Gaussian noise with a desired variance (σ^2_{noise})

$$PLV = E\{e^{j\Delta\varphi}\} = \int_{-\infty}^{+\infty} e^{j\Delta\varphi} f_{\Delta\varphi}(\Delta\varphi)\cdot d(\Delta\varphi) \underline{\underline{\text{def}}} \int_{-\infty}^{+\infty} e^{j\varphi} f_{\varphi}(\varphi)\cdot d\varphi \underline{\underline{\text{def}}} \int_{-\sqrt{3}\sigma}^{+\sqrt{3}\sigma} e^{j\varphi}\cdot\frac{1}{2\sqrt{3}\sigma}\cdot d\varphi = \frac{Sin\left(\sqrt{3}\sigma\right)}{\sqrt{3}\sigma} \quad (3)$$

Based on the characteristic of our real ECoG datasets, we considered 1 to 4 as the power ratio between the "TRA" component and the "1/f" component in the simulations. In addition, the frequency range of task-related connection was specified by the frequency of sinusoidal wave in conjunction with the width of Gaussian tapper.

(Fig. 1a). We generated different SNRs from -10 dB to 10 dB by varying σ^2_{noise}. As an example, a single-trial simulated ECoG signal is illustrated in Fig. 1b.

By manipulating the contributions of the above four components in ECoG signals of subdural electrodes, we generated various strengths for connections between the three

electrodes of our simulated ECoG data. We considered the strength of task-related connections and background couplings to be equal to 0.5 and 0.3, respectively. These values are set fairly close to each other to allow comparing the performance of the statistical tests in a challenging case. We generated data with different SNRs by manipulating the variance of the white noise in the simulated signals. The SNR of the simulated signals varied between − 10 and 10 dB. We generally interpreted our results for low, moderate, and high SNRs with corresponding values equal to − 10, 0, and 10 dB, respectively. The composition of ECoG components in each electrode and their expected connections are illustrated in Fig. 2.

Functional Connectivity Analysis

The simulated data was filtered by a high pass filter with 2 Hz cutoff frequency in order to remove drifts and trends. We applied the time–frequency analysis using the short-time Fourier transform (STFT) over gamma and high gamma bands (30–110 Hz). We utilized the Discrete Prolate Spheroidal Sequences (DPSS) windowing method (Slepian 1978), window length of 20 cycles of the center frequency, and window shift (per step) of 25 ms. Moreover, the frequency resolution of time–frequency analysis was 4 Hz and the frequency smoothing in the DPSS method was set to 0.15 times the center frequency. We selected the PLV as the most basic measure of phase coupling for the functional connectivity measure in the current study:

$$PLV = E\{e^{j\Delta\varphi(f)}\} \tag{4}$$

where f is frequency and $\Delta\varphi(f)$ is the phase difference of two signals at frequency f, and E represents the expected value. PLV is a well-known phase synchrony measure that quantifies phase lag consistency over trials of the task. In other words, two electrodes are considered functionally connected by PLV if they have a consistent phase lag value over trials. In an extreme case, PLV is equal to 1 when the phase difference of corresponding frequency between two signals are equal over all trials, and is close to 0 when phase difference values of all trials are uniformly distributed between $-\pi$ to π. We estimated PLV for every single time–frequency bin of the STFT matrices, resulting in time–frequency matrices of PLV between pairs of electrodes (see Fig. 3a as an example). To estimate sensitivity, specificity, and accuracy of the statistical tests, we created an ideal mask in time–frequency domain for the PLV, shown in Fig. 3b. Each time–frequency bin corresponding to the task-related connection in the simulated ECoG signals was set to 1 in the ideal mask. The simulated PLV in the current study, shown in Fig. 3a, allowed us to test the performance of statistical tests in four cases:

– Case I: Task-related phase couplings in the presence of background couplings

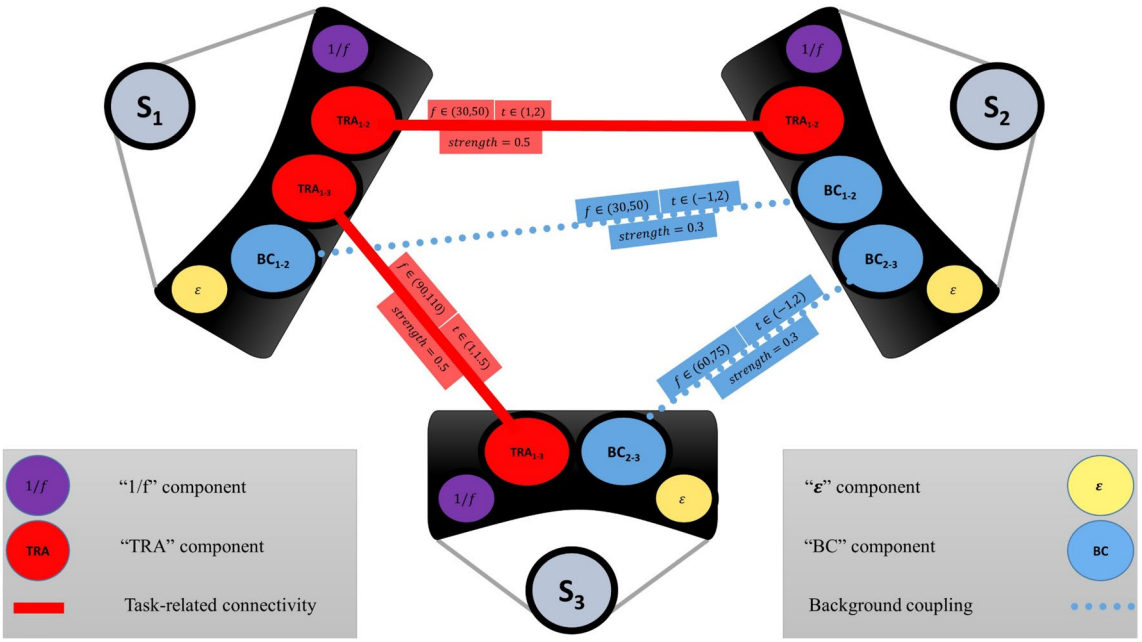

Fig. 2 Schematic view of designed connectivity pattern of three simulated ECoG signals is shown. S_i represents simulated ECoG signal in ith ($i = 1$, 2, and 3) intracranial electrode. For each electrode, items in the black dashboard represent the components that was used to generate the simulated signal of the electrode. Connections between electrode pairs are indicated by solid red (task-related connections) and dotted blue (background couplings) lines. Time–frequency range and strength of connections are noted on the connectivity links

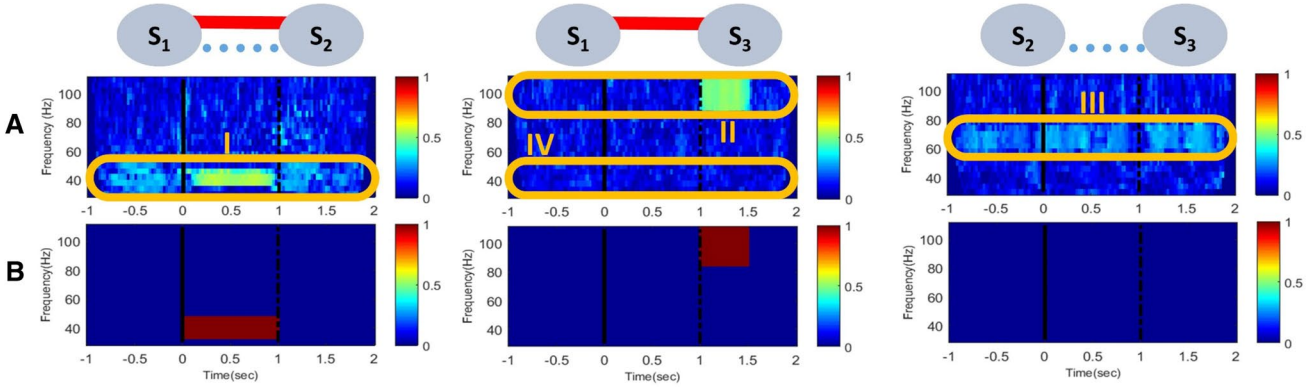

Fig. 3 **a** PLV measures of functional connectivity in simulated data and **b** the ideal time–frequency masks of connectivity. Electrode pairs and connections between them are indicated above every column (See also Fig. 2). Note that the strength of connections are color coded. Orange ovals identify four different cases of phase couplings: (I) presence of both task-related and background couplings; (II) presence of task-related connections and absence of background couplings; (III) absence of task-related connections and presence of background couplings; and (IV) absence of both task-related and background couplings

- Case II: Task-related phase couplings in the absence of background couplings
- Case III: Absence of task-related phase coupling and presence of background couplings
- Case IV: Absence of both task-related phase coupling and background couplings

Evaluating Performance of the Statistical Tests

By comparing the time–frequency connectivity matrices of the simulated data extracted by each statistical test with the ideal connectivity masks shown in Fig. 3b, we calculated true negative (TN), true positive (TP), false negative (FN), false positive (FP), true positive rate (TPR), and false positive rate (FPR). For all statistical tests, we calculated FPR for p values from 0.002 to 1 [Number of random permutations (n) = 500] to compare the ability of each method in controlling the false positive error. We also performed ROC analysis in different SNRs (low, moderate, and high) by using the calculated values of the FPR and TPR.

To quantify the performance of each statistical test in a practical region of FPR (or 1—specificity) and TPR (or sensitivity), we obtained the two-way pAUC values for each statistical test, where both sensitivity and specificity were set greater than 0.5 (Yang et al. 2017):

$$Two-way\ \ pAUC(0.5,0.5) = \int_{FPR=0}^{FPR=0.5} \left[ROC_{FPR,TPR}(FPR, TPR > 0.5) - 0.5\right] d(FPR) \quad (5)$$

We estimated the optimum p value for each statistical test by maximizing the product of sensitivity and specificity:

$$p\,\text{value}_{\text{optimal}} = \text{Argmax}_{p\,\text{value}} \{\text{Sensitivity}\ (p\,\text{value}) \times \text{Specificity}\ (p\,\text{value})\} \quad (6)$$

We also calculated sensitivity, specificity, and balanced accuracy at the optimal p values for each statistical test method in each SNR to compare optimal performances of these methods:

$$Sensitivity = \left.\frac{True\ positive}{True\ positive + False\ negative}\right|_{p\ value_{optimal}} \quad (7)$$

$$Specificity = \left.\frac{True\ negative}{True\ negative + False\ positive}\right|_{p\ value_{optimal}} \quad (8)$$

$$Balanced\ Accuracy = \frac{Sensitivity|_{p\ value_{optimal}} + Specificity|_{p\ value_{optimal}}}{2} \quad (9)$$

Additionally, we compared computational costs for the statistical tests (n = 500) running on a typical PC system with core i7 3.06 GHz CPU and 20 GB RAM to provide a comparison of the computational costs of the different statistical test methods.

To test possibility of replication of the simulation results in real applications, the same simulation procedures were applied on a real ECoG dataset collected form a patient with epilepsy (Fig. 9). The ECoG data were collected while the patient performed a word recognition task (WRT), described in detail elsewhere (Babajani-Feremi et al. 2018). The functional connectivity and statistical testing (all four methods; $p < 0.05$; n = 500) were implemented on the real data similar to that employed for the simulated data.

Results

Effects of Background Couplings

In the first simulation, we investigated performance of the statistical test methods in the absence of background couplings and presence of the 1/f component, the task-related activity component, and the measurement noise component providing a SNR of 0 dB (Fig. 4). The ideal connectivity in Fig. 4 shows that we simulated a connection between electrodes 1 and 2 in time–frequency window of 0.0–1.0 s and 34–46 Hz, and another connection between electrodes 1 and 3 in time–frequency window of 1.0–1.5 s and 86–110 Hz. The simulated connectivity and the results of four statistical tests using a p value of 0.05 are shown in Fig. 4. Our results show that all methods were able to identify the task-related connectivity in the absence of background couplings, although the Monte Carlo permutation and the bootstrap resampling methods had the most and least number of false positives within the post-stimulus interval, respectively. Moreover, the surrogate and demeaned surrogate methods had similar performances.

In the second simulation, we investigated performance of the statistical test methods in the presence of the background couplings (Fig. 5). In this simulation, we used the same setting specified in the first simulation, except we added the background couplings as shown in Fig. 2. In the same manner as the first simulation, the task-related connections were identified by all statistical test methods, which indicates that these methods all had a similar false negative error. The surrogate method falsely detected almost every background couplings as significant task-related connections, indicating that this method is very susceptible to the presence of background couplings. In contrast to the surrogate method, the demeaned surrogate method was able to partially reject the background couplings. The bootstrap resampling was successful in rejecting almost all of the background couplings. The Monte Carlo permutation test rejected some but not all of the background couplings. Results of this simulation reveal that the surrogate method and the bootstrap resampling had the worst and the best performances regarding the detection of false positives in the presence of background couplings.

False Positive Ratio (FPR)

p value, here denoted as selected FPR, determines the desired level of strictness of a statistical test. However, the detected empirical FPR in practice, here denoted as detected FPR, may differ from this value, resulting in an

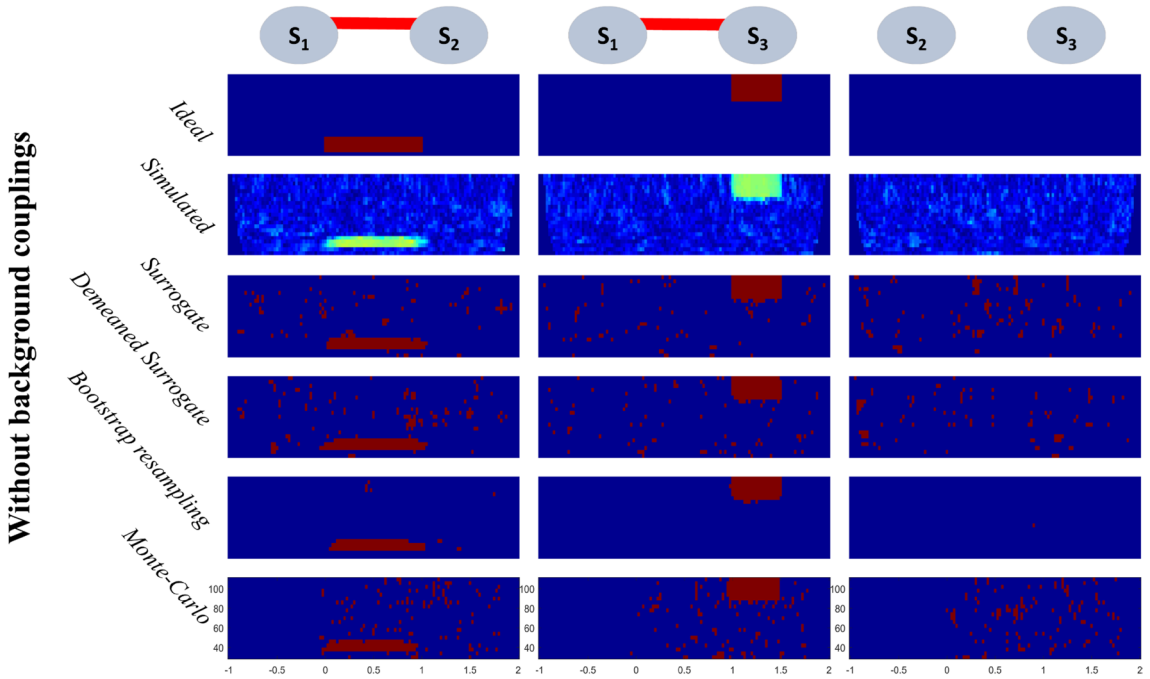

Fig. 4 Ideal connectivity masks, simulated time–frequency connectivity, and significant connectivity identified by surrogate, demeaned surrogate, bootstrap resampling, and Monte Carlo permutation methods in the absence of background couplings (SNR = 0 dB and p value < 0.05). Electrode pairs are indicated above every column with their schematic links. Horizontal and vertical axes show time after stimulus onset (seconds) and frequency (Hz), respectively

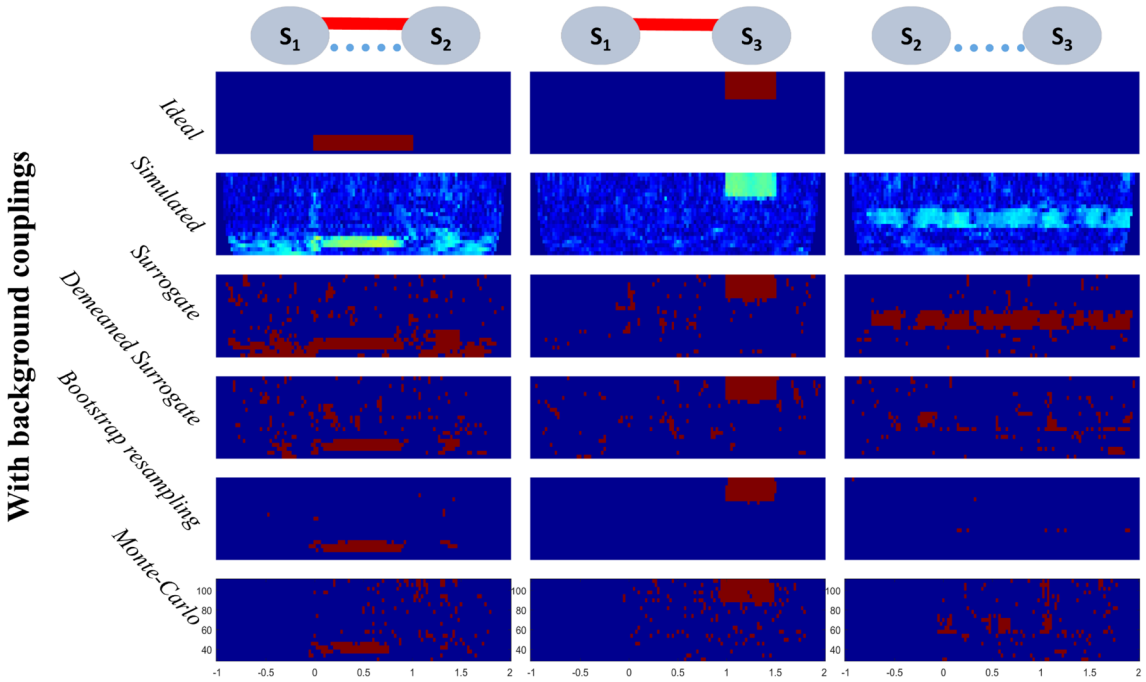

Fig. 5 Ideal connectivity masks, simulated time–frequency connectivity, and significant connectivity extracted by surrogate, demeaned surrogate, bootstrap resampling, and Monte Carlo permutation methods in the presence of background couplings (SNR = 0 dB and p value < 0.05). Electrode pairs are indicated above every column with their schematic links. Horizontal and vertical axes show time after stimulus onset (seconds) and frequency (Hz), respectively

actual level of strictness different from the desired theoretical one. Therefore, testing the relationship between selected and detected FPR values is an informative evaluation and shows how a statistical test under- or overestimates in terms of FPR. Accordingly, the relationships between the selected FPR and detected FPR of four statistical tests in low, moderate, and high SNR levels, and in the absence or presence of background couplings, are shown in Fig. 6. The statistical tests were not perfect in that the curve of detected FPR lied either above or below the identity line. In the absence of background couplings (Fig. 6a), FPR curves of the bootstrap resampling and Monte Carlo permutation tests lied below and above the identity line, respectively. In contrast, curves of surrogate and demeaned surrogate methods closely lied on the identity line. In the presence of background couplings (Fig. 6b), all curves except the one for the bootstrap resampling method lied above the identity line, indicating that the bootstrap resampling method is the most conservative method.

For bootstrap resampling and Monte Carlo permutation tests, the detected FPR curves in each SNR level differed very slightly whether background couplings were absent (Fig. 6a) or present (Fig. 6b). This contends that these two methods are not susceptible to the presence of background couplings. The detected FPR curve of the demeaned surrogate method was moderately different in the presence or absence of background coupling, while the change of this curve was very extreme for the surrogate method. In agreement with previous simulations, results of simulations in Fig. 6 show that the surrogate method is extremely vulnerable to the presence of the background couplings.

Receiver Operating Curve (ROC) Analysis

ROC analyses are beneficial in that they can provide information about the trade-off between the sensitivity and specificity of a statistical test across different p values. The ROC curves of the four statistical test methods studied, in the presence of background couplings and in three SNR levels are shown in Fig. 7. The practical region of interest of ROC curves (specificity and sensitivity > 0.5), where two-way pAUC values are estimated, is highlighted in Fig. 7. The two-way pAUC values of the statistical test methods are listed in Table 1. The Monte Carlo method had the smallest two-way pAUC value compared to other methods in very low SNR (i.e. SNR = − 10 dB). In moderate to high SNRs (SNR = 0–10 dB), the bootstrap resampling method, followed by the demeaned surrogate method, provided the largest two-way pAUC value compared to other methods.

Optimal p values of the statistical tests were extracted from the ROC curves based on Eq. (5) and reported in

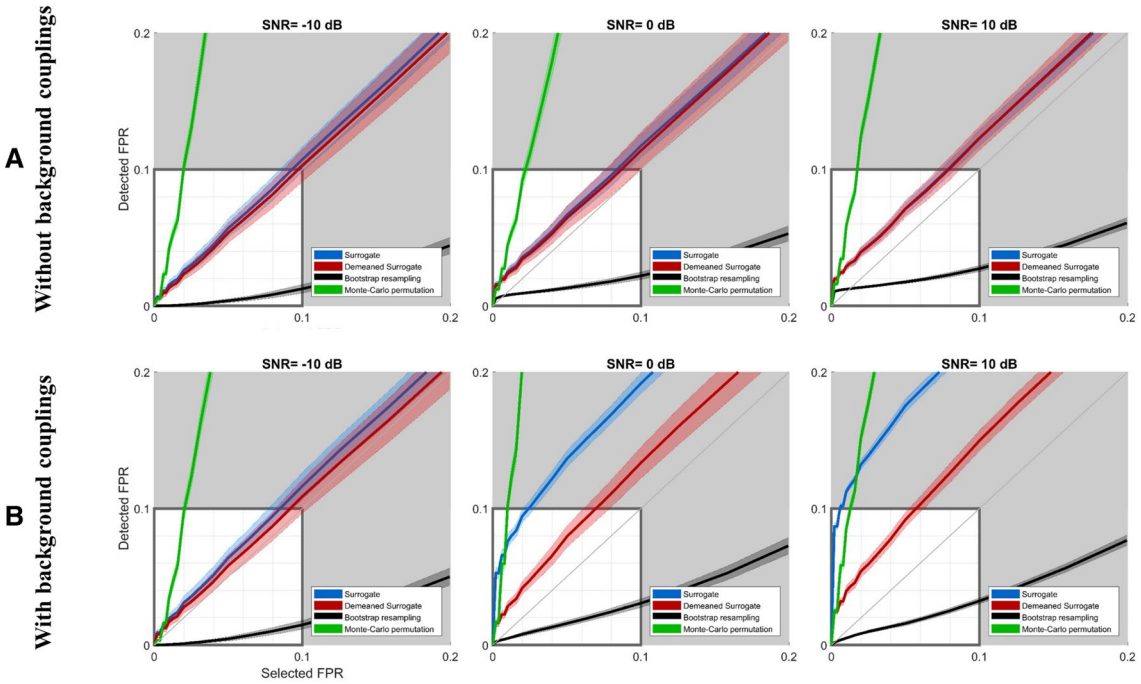

Fig. 6 Detected false positive rate (FPR) by the statistical tests *versus* the selected FPR (p value) in low, moderate, and high SNRs in **a** absence and **b** presence of the background couplings. In each subplot, the bold lines and the shaded areas represent mean values and standard deviations, respectively, of the detected FPR across all realizations of the simulated datasets. The gray area represent the impractical ranges of FPRs

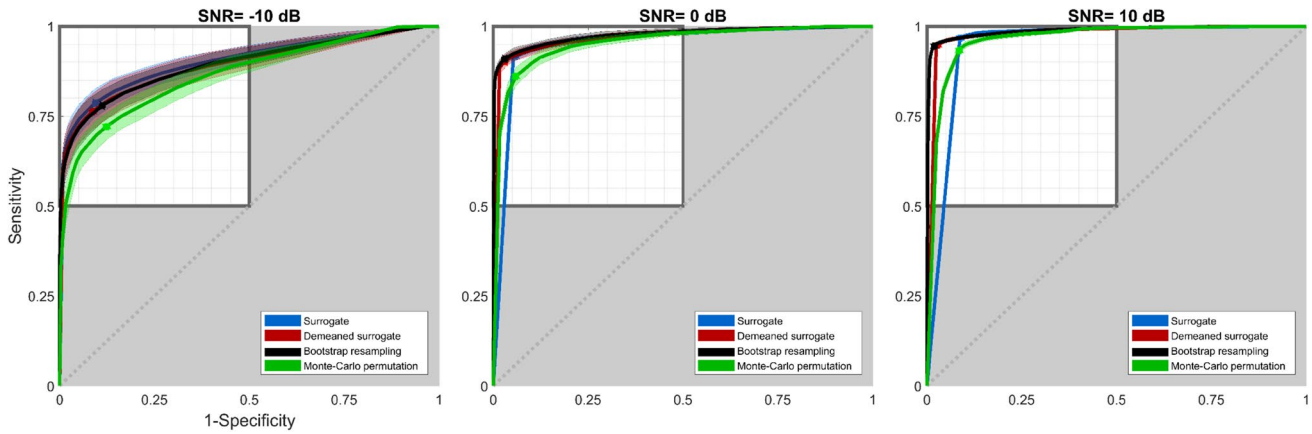

Fig. 7 ROC curves of the statistical test methods in low, moderate, and high SNRs. The ROC cut-off points corresponding to optimal p values are shown by filled circles. The gray area represents impractical ranges for the sensitivity and specificity. Area with a white background was selected to calculate the two-way pAUC

Table 1 Two-way partial area under ROC curve (pAUC; specificity and sensitivity > 0.5) of four statistical test methods for low, moderate, and high SNRs

	Surrogate	Demeaned surrogate	Bootstrap resampling	Monte Carlo permutation
SNR = − 10 dB	0.6912	0.6616	0.6603	0.5588
SNR = 0 dB	0.8620	0.9046	0.9219	0.8527
SNR = 10 dB	0.8594	0.9387	0.9624	0.8935

Table 2 (see also Fig. 7). Results in Table 2 show that the optimal p values increased with SNR decreasing. Selecting a larger p value in a noisy and low SNR condition will allow one to enhance detection of the true positives, but at the expense of having larger false positives. Results in Table 2 revealed that the optimal p value for bootstrap resampling was greater than that of other methods. The demeaned surrogate method had a slightly greater optimal p value compared

Table 2 Optimal p values of four statistical test methods at the cut-off point of ROC curve for low, moderate, and high SNRs

	Surrogate	Demeaned surrogate	Bootstrap resampling	Monte Carlo permutation
SNR = − 10 dB	0.100	0.100	0.398	0.031
SNR = 0 dB	0.006	0.016	0.125	0.008
SNR = 10 dB	0.004	0.006	0.079	0.012

to the surrogate method, indicating that the former method is slightly more conservative than the latter.

Sensitivity, Specificity, and Balanced Accuracy

We calculated the sensitivity, specificity, and balanced accuracy of the statistical test methods in the presence of background couplings, at their optimal p values, and at different levels of SNR (Fig. 8). As expected, results in Fig. 8 show that sensitivity and accuracy of all methods improved by increasing the SNR. As shown in Fig. 8c, the bootstrap resampling and the demeaned surrogate methods had similar accuracies and outperformed the other two methods. The Monte Carlo permutation method had the lowest accuracy compared to the other methods. Considering a practical range for SNR (i.e. SNR = − 5 to 5 dB), results in Fig. 8 show that the bootstrap resampling and demeaned surrogate methods resulted in overall superior performances compared to the Monte Carlo and surrogate methods.

Computational Cost

To have a sense about computational costs for the statistical test methods studied, the execution times of all methods were recorded for all simulations in 9 levels of SNRs (from − 10 to 10 dB in steps of 2.5 dB) considering $n = 500$ repetitions. The execution times for the bootstrap resampling, surrogate, demeaned surrogate, and Monte Carlo permutation methods were 11, 29, 30, and 74 min, respectively. It is noteworthy that the execution time of the Monte Carlo permutation will be much less than 74 min if pre- and post-stimulus time intervals have the same length (see discussion for details).

Real Data

Significant PLV connections in high gamma band during the WRT were identified using four statistical tests, and results are shown in Fig. 9. As shown in Fig. 9a, the number of significant connections in the pre-stimulus interval (i.e. false positives in the baseline) is comparatively large for the surrogate method, few for the demeaned surrogate, and zero for the bootstrap resampling and Monte Carlo permutation methods. By increasing the p value (from 0.005 to 0.1), the number of connections in the pre-stimulus interval increased for the surrogate and demeaned surrogate methods, yet stayed at zero for the bootstrap and Monte Carlo methods. It is worth mentioning that in a task-based study, connections in the baseline have to be disregarded by an ideal statistical test. In agreement with our simulations, results in Fig. 9 revealed that the surrogate, demeaned surrogate, Monte Carlo permutation, and bootstrap resampling methods were very susceptible, slightly susceptible, almost not susceptible, and not susceptible to the presence of background couplings, respectively. According to the total number of significant connections detected by the statistical methods in pre- and post-stimulus intervals (Fig. 9b), the bootstrap resampling was the most conservative method, with detection of a small number of significant connections, and the surrogate test was the most aggressive method, with identification of a large number of significant connections.

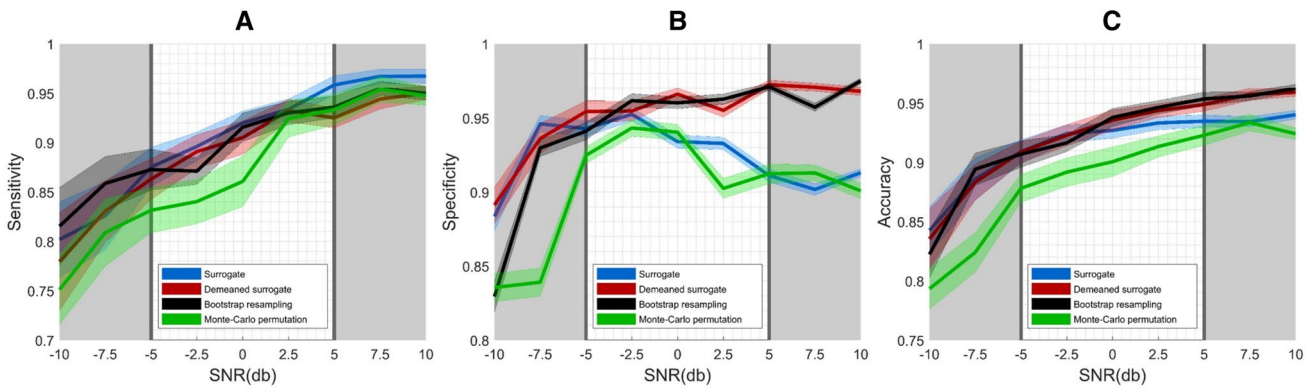

Fig. 8 **a** Sensitivity, **b** specificity, and **c** balanced accuracy of four statistical tests in a wide range of SNR (from − 10 to 10 dB) are shown. The bold lines and shaded areas represent mean and standard deviation of the measure, respectively, across simulated datasets. The gray area corresponds to the less realistic SNR values (− 10 to − 5 dB and 5 to 10 dB)

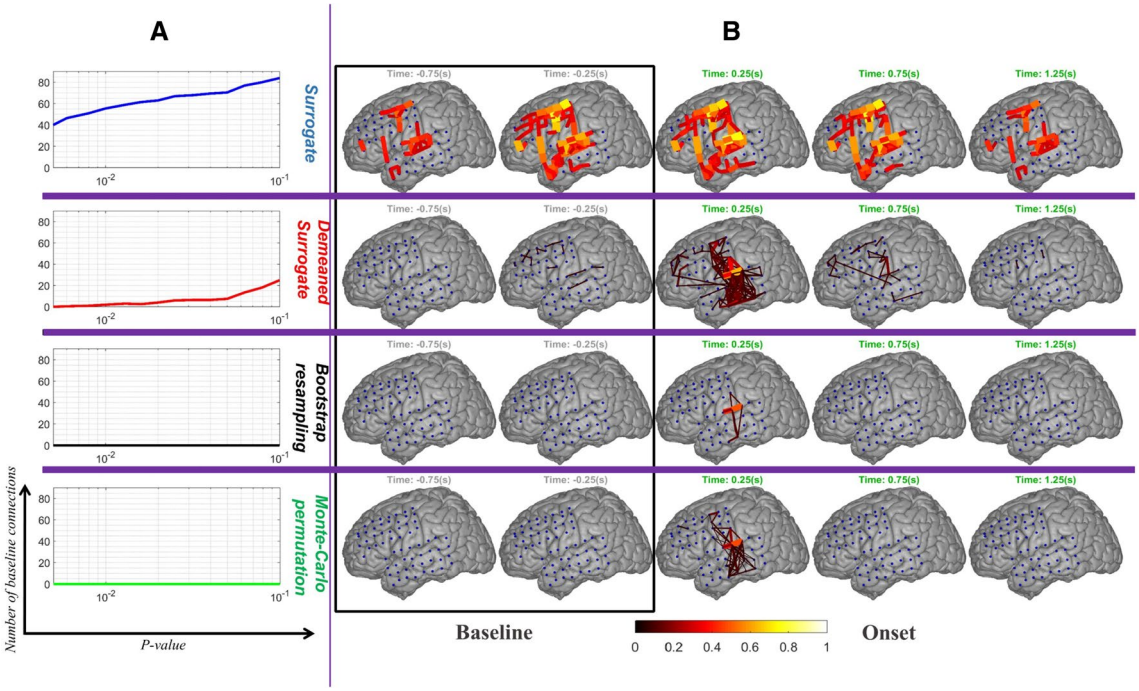

Fig. 9 An illustrative case report of a single subject real ECoG data. The data were collected from a patient with epilepsy (male, 40 years of age) who underwent a Phase II evaluation for treatment of drug-resistant epilepsy at the Le Bonheur Children's Hospital. Sixty-four subdural grid electrodes were implanted on the left temporo-frontal regions of this patient. As part of presurgical language mapping in this patient, ECoG data were collected while the patient performed a word recognition task (WRT). The WRT is a receptive language mapping task and is described in detail elsewhere (Babajani-Feremi et al. 2018). After preprocessing of ECoG data, as described in detail in (Babajani-Feremi et al. 2018), the functional connectivity and statistical testing (all four methods; $p < 0.05$; $n = 500$) were implemented on the real data similar to that employed for the simulated data. Then for visualization purposes, we calculated the average of significant PLV across high gamma frequency band (60–110 Hz) in overlapping time windows of 500 ms with 250 ms time shift. **a** Number of significant false connections detected within pre-stimulus interval by four statistical test methods where Y-axis indicates the number of false connections detected in the baseline and X-axis represents the utilized p values. **b** Dynamics of PLV functional connectivity in the high gamma band during the WRT task, detected by four statistical test methods. Each subplot refers to a certain time with respect to stimulus onset, which is shown on top of the subplot. Strength of the detected connections are coded by both width and color (see color bar)

Discussion

Different statistical tests have been utilized in the literature to assess significance in the study of the functional connectivity of electrophysiological signals. No study has yet evaluated their relative performance, specifically in the presence of background couplings. Background couplings known as connections within the intrinsic brain connectivity are present during both pre- and post-stimulus time intervals, and in a task-based study should be disregarded by an ideal statistical test. In this study, we investigated the efficiencies of four widely-used non-parametric statistical tests—surrogate, demeaned surrogate, bootstrap resampling, and Monte Carlo—in the presence of background couplings. We found that the surrogate method is extremely susceptible to the background couplings, while bootstrap resampling and Monte Carlo permutation methods are not. Moreover, we observed that the demeaned surrogate method is considerably more efficient than the surrogate method in the presence of background couplings. Our simulation results revealed that among the aforementioned statistical tests, bootstrap resampling is the most efficient method to assess significant functional connectivity in the presence of background couplings.

Utilizing Pre-stimulus Data in Statistical Testing

All statistical test methods presented in this study, except the surrogate method, use information from the pre-stimulus interval to identify significant connections. Subsequently, the surrogate method was the only method drastically susceptible to the presence of the background couplings. In addition, our results revealed that the demeaned surrogate method was significantly less sensitive to background couplings compared to the surrogate method. The advantage of the demeaned surrogate method over the surrogate method stems from the fact that the demeaning process removes the average of connectivity during the pre-stimulus interval from post-stimulus connectivity. By removing the average of connectivity in the pre-stimulus interval from the whole trial,

(a) values of connectivity in the pre-stimulus interval will approach zero and therefore will not be identified as false positives, and (b) connections that exist during both pre- and post-stimulus intervals (background couplings) are most likely to be identified as non-task-related connections. These observations imply that data from the pre-stimulus interval should be included in the statistical test in order to minimize the effects of background couplings.

Exploratory Versus Hypothesis-Driven Studies

There are two kinds of research studies in the literature of brain mapping: exploratory or hypothesis-driven studies (Cohen 2015). The aims of these two kinds of studies are different, compelling researchers to take different approaches in using functional connectivity measures and statistical tests. The goal of exploratory studies is to explore new findings without any possible theoretical hypothesis. In hypothesis-driven studies, one tries to test some known theoretical finding obtained by experiments (e.g. behavioral experiments or clinical findings). It is suggested that exploratory studies should be more specific to avoid false detections, and thus these studies should exert stricter control over FPR than over TPR (Cohen 2015). On the other hand, hypothesis-driven studies should be more sensitive to detect the hypothesized phenomenon under worst case conditions, and thus these studies should exert stricter control over TPR than FPR (Cohen 2015). Therefore, detected p values in exploratory and hypothesis-driven studies should not be greater and smaller than the selected p value, respectively. In other words, the FPR curves of statistical tests should not lie above the identity line for exploratory studies, nor below the identity line for hypothesis-driven studies. Considering Fig. 6b, bootstrap resampling and demeaned surrogate methods would be suitable choices for exploratory and hypothesis-driven studies, respectively, since their FPR curves are the closest curves below and above the identity line, respectively.

Real Data

Previous studies on auditory word recognition tasks reported that STG, MTG, superior temporal sulcus (STS), and anterior-inferior temporal gyrus (a-ITG) are involved in hearing words (Rimol et al. 2006; Binney et al. 2010; Hickok and Poeppel 2007), including phonological processing (Ashtari et al. 2004; Liebenthal et al. 2005) and semantic perception of either heard words or seen objects (meaning processing) (Binney et al. 2010; Nobre et al. 1994; Tomasello et al. 2017). During the WRT, it is expected that the STG is active at the beginning of the hearing, associated with primary auditory processing. Soon after primary hearing of the words, phonological processing and meaning processing (perception) are expected to take place within MTG, STS, and a-ITG (Rimol et al. 2006; Binney et al. 2010; Hickok and Poeppel 2007; Ashtari et al. 2004; Liebenthal et al. 2005; Nobre et al. 1994; Tomasello et al. 2017). Premotor and motor areas are involved in speech planning and articulation (Flinker et al. 2015; Hope and Price 2016), and since the WRT is a receptive language task, we do not expect involvement of these areas during this task. Results in Fig. 9 show that significant connections were identified by the statistical test methods over STG, MTG, and ITG during hearing and perception of a single word. However, all methods except the bootstrap resampling method identified many connections within the premotor and motor areas which are not expected to be active during a receptive language task. Therefore, the results of the bootstrap resampling method outperformed the results of other methods, in agreement with known anatomical and functional mapping of brain areas involved in the perception of single words (Rimol et al. 2006; Binney et al. 2010; Hickok and Poeppel 2007; Ashtari et al. 2004; Liebenthal et al. 2005; Nobre et al. 1994; Tomasello et al. 2017).

Qualitative Comparison

To encapsulate all of our findings derived from the simulated and real data, efficiencies of four statistical test methods are scored qualitatively from 1 (worst) to 4 (best) regarding several aspects, such as susceptibility to background couplings, specificity, and sensitivity (Fig. 10). It should be noted that all these qualitative measures are based on the observations and measurements in the current study. According to Fig. 10, the bootstrap resampling and surrogate methods generally seem to be the most and least efficient method for assessing the significance of functional connectivity, respectively. Moreover, the bootstrap resampling and demeaned surrogate methods are the most suitable choices for an exploratory versus a hypothesis-driven study, respectively.

Limitations

It has been shown that the results of functional connectivity analysis depend upon the choice of measure, e.g. PLV or PLI, for calculating connectivity (Bastos and Schoffelen 2016; Cohen 2015). We selected PLV as a measure of functional connectivity in this current study because: (a) PLV is a simple and efficient measure of functional connectivity that represents phase synchrony of electrophysiological signals; (b) PLV can detect a weak synchronization regime between signals having two areas, where the phases of signals are coupled but the amplitudes of signals may not be (Hramov et al. 2005); and (c) we and other investigators have demonstrated that PLV values are an efficient and reliable measure for studying functional connectivity in epilepsy and other

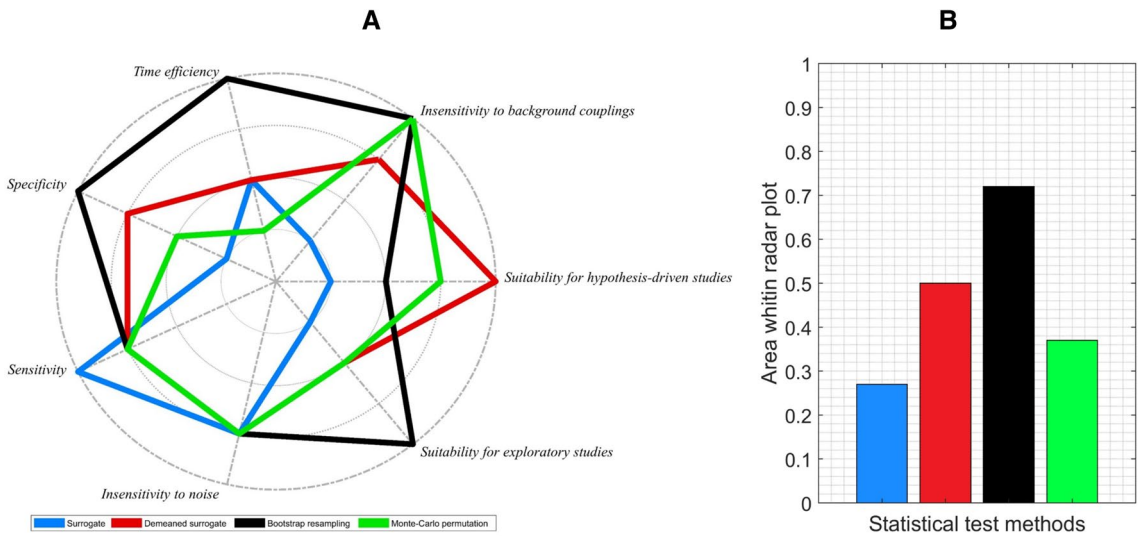

Fig. 10 **a** Qualitative scores of statistical tests in different aspects shown on a radar plot. The statistical test methods are color-coded (see legend). Four circles with different radii indicate the scores from 1 to 4. **b** Normalized areas within the curves in (**a**) showing overall performances of the four statistical test methods studied

diseases (Dimitriadis et al. 2015; Elahian et al. 2017; Babajani-Feremi et al. 2018). A limitation of this study is to compare performances of statistical methods using real ECoG data of one patient, and our results should be validated in future studies with large sample size. While the results of this study may not be able to be extended to all measures of functional connectivity, we think that our findings regarding performances of the statistical tests can be used as an initial assessment for other functional connectivity measures.

Conclusion

By generating realistic simulated ECoG signals, we compared the performance of four non-parametric statistical tests—phase permuted surrogate data (surrogate), demeaned phase permuted surrogate data (demeaned surrogate), bootstrap resampling, and Monte Carlo permutation methods—in the presence of different levels of noise and background couplings. We investigated the effects of background couplings on the performance of statistical test methods, which has not been explored in previous studies. We found that the surrogate method, but not the other methods, was dramatically susceptible to the presence of background couplings. In addition, the well-known Monte Carlo permutation test was comparatively less efficient than the bootstrap resampling and demeaned surrogate methods. We concluded that the bootstrap resampling method was the most efficient statistical test for assessing significance in PLV functional connectivity analysis, followed by the demeaned surrogate method.

Funding This study was supported by the Children's Foundation Research Institute at Le Bonheur Children's Hospital and the Le Bonheur Associate Board, Memphis, TN.

Compliance with Ethical Standards

Conflict of interest None of the authors have any conflicts of interest to disclose.

References

Ashtari M, Lencz T, Zuffante P, Bilder R, Clarke T, Diamond A et al (2004) Left middle temporal gyrus activation during a phonemic discrimination task. NeuroReport 15:389–393

Astolfi L, Fallani FDV, Cincotti F, Mattia D, Marciani M, Salinari S et al (2009) Estimation of effective and functional cortical connectivity from neuroelectric and hemodynamic recordings. IEEE Trans Neural Syst Rehabil Eng 17:224–233

Babajani-Feremi A, Holder CM, Narayana S, Fulton SP, Choudhri AF, Boop FA et al (2018a) Predicting postoperative language outcome using presurgical fMRI, MEG, TMS, and high gamma ECoG. Clin Neurophysiol 129:560–571

Babajani-Feremi A, Noorizadeh N, Mudigoudar B, Wheless JW (2018b) Predicting seizure outcome of vagus nerve stimulation using MEG-based network topology. Neuroimage Clin 19:990–999

Barbey AK (2018) Network neuroscience theory of human intelligence. Trends Cogn Sci 22:8–20

Bastos AM, Schoffelen J-M (2016) A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. Front Syst Neurosci 9:175

Binney RJ, Embleton KV, Jefferies E, Parker GJ, Ralph MA (2010) The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel

direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. Cereb Cortex 20:2728–2738
Chen Z, Caprihan A, Damaraju E, Rachakonda S, Calhoun V (2018) Functional brain connectivity in resting-state fMRI using phase and magnitude data. J Neurosci Methods 293:299–309
Cohen MX (2015) Effects of time lag and frequency matching on phase-based connectivity. J Neurosci Methods 250:137–146
Cohen JR (2017) The behavioral and cognitive relevance of time-varying, dynamic changes in functional connectivity. NeuroImage 180:515–525
Cole MW, Bassett DS, Power JD, Braver TS, Petersen SE (2014) Intrinsic and task-evoked network architectures of the human brain. Neuron 83:238–251
Dimitriadis SI, Zouridakis G, Rezaie R, Babajani-Feremi A, Papanicolaou AC (2015) Functional connectivity changes detected with magnetoencephalography after mild traumatic brain injury. Neuroimage Clin 9:519–531
Efron B (1982) The jackknife, the bootstrap, and other resampling plans, vol 38. SIAM, Philadelphia
Elahian B, Yeasin M, Mudigoudar B, Wheless JW, Babajani-Feremi A (2017) Identifying seizure onset zone from electrocorticographic recordings: a machine learning approach based on phase locking value. Seizure 51:35–42
Flinker A, Korzeniewska A, Shestyuk AY, Franaszczuk PJ, Dronkers NF, Knight RT et al (2015) Redefining the role of Broca's area in speech. Proc Natl Acad Sci 112:2871–2875
Fox KCR, Foster BL, Kucyi A, Daitch AL, Parvizi J (2018) Intracranial electrophysiology of the human default network. Trends Cogn Sci 22:307–324
Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn Sci 9:474–480
Gordon SM, Franaszczuk PJ, Hairston WD, Vindiola M, McDowell K (2013) Comparing parametric and nonparametric methods for detecting phase synchronization in EEG. J Neurosci Methods 212:247–258
Greenblatt RE, Pflieger ME, Ossadtchi AE (2012) Connectivity measures applied to human brain electrophysiological data. J Neurosci Methods 207:1–16
Guthrie D, Buchwald JS (1991) Significance testing of difference potentials. Psychophysiology 28:240–244
Hagiwara K, Ogata K, Okamoto T, Uehara T, Hironaga N, Shigeto H et al (2014) Age-related changes across the primary and secondary somatosensory areas: an analysis of neuromagnetic oscillatory activities. Clin Neurophysiol 125:1021–1029
He BJ, Zempel JM, Snyder AZ, Raichle ME (2010) The temporal structures and functional significance of scale-free brain activity. Neuron 66:353–369
Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393
Hope TMH, Price CJ (2016) Why the left posterior inferior temporal lobe is needed for word finding. Brain 139:2823–2826
Hramov AE, Koronovskii AA, Kurovskaya MK, Moskalenko OI (2005) Synchronization of spectral components and its regularities in chaotic dynamical systems. Phys Rev E Stat Nonlinear Soft Matter Phys 71:056204
Koutsoukos E, Maillis A, Papageorgiou C, Gatzonis S, Stefanis C, Angelopoulos E (2015) The persistent and broadly distributed EEG synchronization might inhibit the normal processing capability of the human brain. Neurosci Lett 609:137–141
Krienen FM, Yeo BT, Buckner RL (2014) Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. Philos Trans R Soc Lond B Biol Sci 369:20130526
Kucyi A, Schrouff J, Bickel S, Foster BL, Shine JM, Parvizi J (2018) Intracranial electrophysiology reveals reproducible intrinsic functional connectivity within human brain networks. J Neurosci 38:4230–4242
Lachaux J, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. *ma.utexas.edu*
Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005) Neural substrates of phonemic perception. Cereb Cortex 15:1621–1631
Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190
Micheli C, Kaping D, Westendorff S, Valiante TA, Womelsdorf T (2015) Inferior-frontal cortex phase synchronizes with the temporal–parietal junction prior to successful change detection. NeuroImage 119:417–431
Miller KJ, Sorensen LB, Ojemann JG, den Nijs M (2009) Power-law scaling in the brain surface electric potential. PLoS Comput Biol 5:e1000609
Nobre AC, Allison T, McCarthy G (1994) Word recognition in the human inferior temporal lobe. Nature 372:260–263
Nolte G, Bai O, Wheaton L, Mari Z, Vorbach S, Hallett M (2004) Identifying true brain interaction from EEG data using the imaginary part of coherency. Clin Neurophysiol 115:2292–2307
Nolte G, Ziehe A, Nikulin VV, Schlögl A, Krämer N, Brismar T et al (2008) Robustly estimating the flow direction of information in complex physical systems. Phys Rev Lett 100:234101
Phillips JM, Vinck M, Everling S, Womelsdorf T (2014) A long-range fronto-parietal 5- to 10-Hz network predicts "top-down" controlled guidance in a task-switch paradigm. Cereb Cortex 24:1996–2008
Porcaro C, Coppola G, Pierelli F, Seri S, Di Lorenzo G, Tomasevic L et al (2013) Multiple frequency functional connectivity in the hand somatosensory network: an EEG study. Clin Neurophysiol 124:1216–1224
Rimol LM, Specht K, Hugdahl K (2006) Controlling for individual differences in fMRI brain activation to tones, syllables, and words. Neuroimage 30:554–562
Sadaghiani S, Poline J-B, Kleinschmidt A, D'Esposito M (2015) Ongoing dynamics in large-scale functional connectivity predict perception. Proc Natl Acad Sci USA 112:8463–8468
Sekihara K, Sahani M, Nagarajan S (2004) Bootstrap-based statistical thresholding for MEG source reconstruction images. In: 26th annual international conference of the IEEE on engineering in medicine and biology society, 2004. IEMBS'04, pp 1018–1021
Slepian D (1978) Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: the discrete case. Bell Labs Tech J 57:1371–1430
Srinivasan R, Winter WR, Ding J, Nunez PL (2007) EEG and MEG coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics. J Neurosci Methods 166:41–52
Stam C, Nolte G, Daffertshofer A (2007) Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. Wiley Online Library, New York
Stephen EP, Lepage KQ, Eden UT, Brunner P, Schalk G, Brumberg JS et al (2014) Assessing dynamics, spatial scale, and uncertainty in task-related brain network analyses. Front Comput Neurosci 8:31
Takahashi T, Goto T, Nobukawa S, Tanaka Y, Kikuchi M, Higashima M et al (2018) Abnormal functional connectivity of high-frequency rhythms in drug-naïve schizophrenia. Clin Neurophysiol 129:222–231
Theiler J, Eubank S, Longtin A, Galdrikian B, Doyne Farmer J (1992) Testing for nonlinearity in time series: the method of surrogate data. Physica D 58:77–94
Tomasello R, Garagnani M, Wennekers T, Pulvermüller F (2017) Brain connections of words, perceptions and actions: a neurobiological model of spatio-temporal semantic activation in the human cortex. Neuropsychologia 98:111–129

Varela F, Lachaux J-P, Rodriguez E, Martinerie J (2001) The brainweb: phase synchronization and large-scale integration. Nat Rev Neurosci 2:229–239

Vinck M, Oostenveld R, Wingerden MV, Battaglia F (2011) An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. Elsevier, Amsterdam

Yang H, Lu K, Lyu X, Hu F (2017) Two-way partial AUC and its properties. Stat Methods Med Res 28:184–195