Stimulus evoked causality estimation in stereo-EEG

Journal of Neural Engineering

PAPER

Stimulus evoked causality estimation in stereo-EEG

Andrea Cometa[1], Piergiorgio D'Orio[2,3], Martina Revay[2,4], Silvestro Micera[1,5,6] and Fiorenzo Artoni[1,5,6,*]

[1] BioRobotics Institute and Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna, Viale Rinaldo Piaggio, 34, Pontedera, 56025, Italy
[2] 'Claudio Munari' Center for Epilepsy Surgery, ASST GOM Niguarda Hospital, Piazza dell'Ospedale Maggiore, 3, 20162 Milano, Italy
[3] Institute of Neuroscience, CNR, via Volturno 39E, Parma 43125, Italy
[4] Department of Biomedical and Clinical Sciences "L. Sacco", Università degli Studi di Milano, Via Giovanni Battista Grassi 74, Milan 20157, Italy
[5] Ecole Polytechnique Federale de Lausanne, Bertarelli Foundation Chair in Translational NeuroEngineering, Center for Neuroprosthetics and School of Engineering, Chemin des Mines, 9, Geneva, GE CH 1202, Switzerland
[6] These authors contributed equally to this work.
[*] Author to whom any correspondence should be addressed.

E-mail: fiorenzo.artoni@unige.ch

Keywords: SEEG, partial directed coherence, event related causality, connectivity, non-parametric test

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
4 May 2021

REVISED
8 September 2021

ACCEPTED FOR PUBLICATION
17 September 2021

PUBLISHED
4 October 2021

Abstract

Objective. Stereo-electroencephalography (SEEG) has recently gained importance in analyzing brain functions. Its high temporal resolution and spatial specificity make it a powerful tool to investigate the strength, direction, and spectral content of brain networks interactions, especially when these connections are stimulus-evoked. However, choosing the best approach to evaluate the flow of information is not trivial, due to the lack of validated methods explicitly designed for SEEG. *Approach.* We propose a novel non-parametric statistical test for event-related causality (ERC) assessment on SEEG recordings. Here, we refer to the ERC as the causality evoked by a particular part of the stimulus (a response window (RW)). We also present a data surrogation method to evaluate the performance of a causality estimation algorithm. We finally validated our pipeline using surrogate SEEG data derived from an experimentally collected dataset, and compared the most used and successful measures to estimate effective connectivity, belonging to the Geweke–Granger causality framework. *Main results.* Here we show that our workflow correctly identified all the directed connections in the RW of the surrogate data and prove the robustness of the procedure against synthetic noise with amplitude exceeding physiological-plausible values. Among the causality measures tested, partial directed coherence performed best. *Significance.* This is the first non-parametric statistical test for ERC estimation explicitly designed for SEEG datasets. The pipeline, in principle, can also be applied to the analysis of any type of time-varying estimator, if there exists a clearly defined RW.

1. Introduction

Stereoelectroencephalography (SEEG) is traditionally used by clinicians to locate epileptic foci in patients with drug-resistant epilepsy [1, 2] and has recently gained importance to investigate the neuronal structures involved in important brain functions [3–5], or as driving signal for brain-computer interfaces [6]. Indeed, SEEG recordings allow us to gather data by combining electroencephalography (EEG) or magnetoencephalography (MEG) temporal resolutions of milliseconds with unmatched spatial specificity [7] and without the common limitations associated with non-invasive recordings (e.g. artifacts, inverse problem, source leakage, etc) [4].

Recent works using SEEG recordings have shown that cognitive tasks cause event-related spectral perturbations of brain activity at frequencies up to 300 Hz [3, 8–12]. Furthermore, high gamma activity arising from a portion of the network engaged by a task may also causally induce high gamma activity in another brain region [13]. Fast oscillatory neuronal activity can in fact play an important role in organizing neurons in large-scale networks [14–16].

The combination of high spatial and temporal resolution, together with the possibility to record

Table 1. Main advantages and disadvantages of the causality metrics considered in this paper.

Causality measures	Main advantages	Main disadvantages
DTF	Extends the Granger causality concept to multichannel data.	Cannot discriminate between direct and indirect connections.
dDTF	Selectively enhances direct influences over indirect ones.	Not well-suited to study interactions occurring during brief time scales.
SdDTF	Also adapts for examining short-time epochs.	Lack of use -and therefore validation- in literature.
PDC	Able to separate direct from indirect interactions and works well in relatively noisy data.	Compromised sensitivity to outflows.
wDTF and wPDC	Increases the physiological interpretability of the DTF or the PDC.	This type of weighting can bias the results by enhancing interactions coming from channels with a high-power spectral density.

wideband data, offered by SEEG, makes it a remarkable tool to identify brain structures engaged in task-related cognitive processes. Furthermore, SEEG allows to investigate the direction, the amplitude, and the characteristic frequencies of the interactions occurring among those structures. This set of causal connections under different functional conditions is traditionally defined as effective connectivity [17].

The work presented here stems from this background and aims to define a pipeline to evaluate effective connectivity in SEEG cognitive processing tasks.

One of the most interesting applications of effective connectivity is the study of the event-related causality (ERC), i.e. the changes in the connectivity network of brain structures time-locked to a stimulus [13]. Hence, to investigate the dynamical evolution of connectivity patterns in response to a specific component of a time-varying stimulus (i.e. the response window, RW) there is the need to validate a time-varying connectivity estimation framework.

Such approach may be useful to investigate reaching and grasping tasks [18, 19], or the RW of single words of a complete sentence [3, 5, 20].

Reliable time-varying methods are thus needed to identify the interactions among brain structures from their electrophysiological signals, and how these interactions vary with stimuli and time. In particular, connectivity metrics based on the Geweke-Granger causality [21, 22] such as the directed transfer function (DTF) [23], the partial directed coherence (PDC) [24], the direct DTF (dDTF) [25], the short-time direct DTF (SdDTF) [13] and the scalogram-weighted DTF and PDC (wDTF, wPDC) [26], are particularly interesting because they proved to be effective in correctly identifying interactions on simulated or benchmark data [13, 26–29]. These indexes are based on multivariate autoregressive (MVAR) models and allow the analysis of non-stationary data [30, 31] by also determining the spectral content of the interaction.

The DTF extends the Granger causality concept to multichannel data, which in principle should allow distinguishing between direct and indirect interactions. However, non-zero values of DTF can be found between two recording sites for which the causal influence is not direct [13]. The dDTF is the product between the DTF and the partial coherence [32], and it was designed to successfully discriminate between direct and indirect causal influences. dDTF is however unsuited for connectivity estimation in brief data epochs [13]. To overcome this limitation, the SdDTF was introduced. SdDTF assesses magnitudes and spectral characteristics of directed causal interactions between time series, by also examining short-time epochs [13]. The PDC can separate direct from indirect connections and can correctly identify interactions even in relatively noisy data [33–36]. The PDC, however, normalizes the outgoing connection strengths and this can, in turn, compromise the sensitivity to outflows. The wDTF and the wPDC were proposed to increase the physiological interpretability of the DTF and the PDC [26]. However, this type of weighting can potentially bias the results by enhancing interactions coming from channels with a high-power spectral density. The main advantages and disadvantages of these causality measures are summarized in table 1.

Choosing the best method to evaluate the ERC on SEEG recordings is not trivial. While the Geweke–Granger causality measures have been widely used and validated for the assessment of brain networks in EEG/MEG and fMRI scenarios [30, 37–45], it is still missing a ground truth of the application of these techniques on SEEG data. This, combined with the lack of a solid statistical framework to identify significant directed connections, leads to the need for developing and validating ERC estimation tools explicitly designed for SEEG recordings.

To overcome this limitation, we propose here a new workflow for the ERC assessment on SEEG recordings, based on a statistical non-parametric mapping approach [46, 47]. The procedure relies on the computation of a significance metric tested against a null permutation distribution obtained by shuffling several times the time samples of the connectivity measure.

We tested the validity of our pipeline on surrogate SEEG data with physiologically plausible noise levels and with a known connectivity matrix as well as strength and distribution of the causality samples comparable to those of a real SEEG dataset. In this way, it was also possible to investigate which combination of the previously described causality measures and significance metrics displays results closer to this ground truth.

2. Methods

2.1. ERC assessment on SEEG recordings

The computational steps necessary to identify significant directed connections during the RW on a multi-trial SEEG dataset are graphically described in figure 1 and detailed below.

2.1.1. Causality estimation

Within the Granger causality framework, a time series $x_j(t)$ is deemed to cause another time series $x_i(t)$ if knowledge of past samples of $x_j(t)$ reduces the prediction error for the current sample of $x_i(t)$. The relation between $x_j(t)$ and $x_i(t)$ can be estimated by fitting a time-varying MVAR model on $\boldsymbol{X}(t)$:

$$\boldsymbol{X}(t) = [x_1(t), x_2(t), \ldots, x_D(t)]^T \tag{1}$$

where D is the total number of channels.

The MVAR model assumes a linear relationship between the channels in $\boldsymbol{X}(t)$ of the form:

$$\boldsymbol{X}(t) = -\sum_{p}^{k=1} A_k(t)\boldsymbol{X}(t-k) + \boldsymbol{e}(t) \tag{2}$$

where $A_k(t)$ is the time-varying DxD MVAR coefficients matrix, $\boldsymbol{e}(t)$ is a white noise process with covariance matrix W and p is the model order. The $A_k(t)$ matrices can be derived by using a general linear Kalman filter (GLKF) [31]. To estimate the model order p, the Bayesian information criterion (BIC) can be used [48].

To improve ERC estimation accuracy, single trials can be combined in two ways: (a) single-trial connectivity estimation followed by averaging across trials, and (b) simultaneous connectivity estimation on all trials. Here the single-trial modeling is used. This allows for the possibility of time-warping of the connectivity matrices before trial averaging to align epoch events while preserving the frequency content of the signal. Single-trial modeling is also more reliable and less prone to overfitting than multi-trial modeling [49].

After estimating the $A_k(t)$ matrices trial by trial, these are fed to the preferred causality measure algorithm (details on how to calculate the DTF, PDC, dDTF, SdDTF, wDTF, and wPDC can be found in [13, 23–26]), resulting in $K \times D(D-1)$ $\mathrm{Conn}_{ij}(f,t)$ matrices, where K is the number of trials, and $\mathrm{Conn}_{ij}(f,t)$ is the information flow from channel j to channel i, function of both time and frequency.

2.1.2. Connectivity in RW

All the next steps of the algorithm are applied independently at each frequency (or band $[f_1, f_2]$ if values from $\mathrm{Conn}_{ij}(f_1,t)$ to $\mathrm{Conn}_{ij}(f_2,t)$ are averaged first).

If the RW is not temporally aligned across trials, time-warping of the $\mathrm{Conn}_{ij}(f,\cdot)$ time series can now be safely performed without altering the frequency content of the original signal.

Baseline correction is then carried out by dividing $\mathrm{Conn}_{ij}(f,\cdot)$, trial by trial and for each i,j $(i \neq j)$ couple independently, by its mean baseline value. The $\overline{\mathrm{Conn}_{ij}}(f,t)$ matrices are finally obtained by averaging across trials, and a significance metric (sig) that quantifies the increase of the connectivity during the RW is calculated for each pair i,j $(i \neq j)$ of channels. Many significance metrics can be used, here we considered the following ones:

Significant Mean:

$$\mathrm{sig} = \frac{a \times T}{\left(\sum_s a_s \times T_s\right) + 1}. \tag{3}$$

RW mean:

$$\mathrm{sig} = a. \tag{4}$$

Mean ratio:

$$\mathrm{sig} = \frac{a}{\left(\sum_s a_s\right) + 1}. \tag{5}$$

Significant Max:

$$\mathrm{sig} = \frac{A \times T}{\left(\sum_s A_s \times T_s\right) + 1} \tag{6}$$

where a and A are respectively the mean and the maximum connectivity amplitude during the RW, T is the amount of time the RW connectivity is above the 95th percentile of the distribution of connectivity values in the baseline and a_s, A_s and T_s are the same quantities but pertaining to the other segments s of the stimulus.

The rationale behind the Significant Max metric is to determine the pairs of channels that highlight the maximum time-specific significant causality. The Significant Mean metric has been introduced as a variant of Significant Max to limit the effect of possible outliers in the causality estimation. The rationale behind the RW Mean metric is to highlight the pairs of channels with the highest absolute mean connectivity during the RW, while the Mean Ratio is designed to highlight the highest connectivity values in the RW relative to the causality values of other parts of the stimulus.

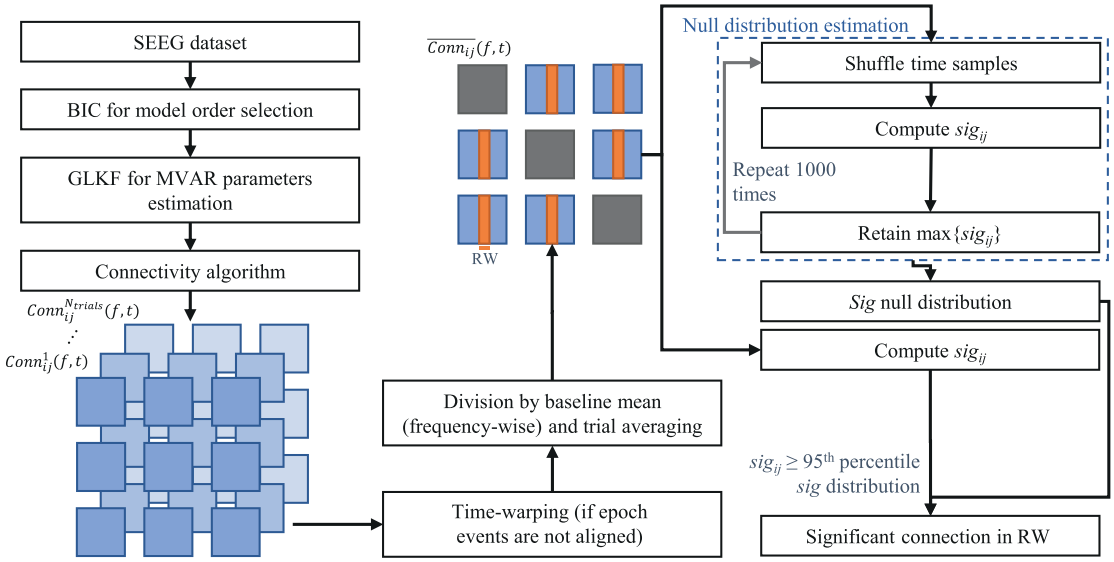

Figure 1. Processing steps needed to identify a significant connection in the response window (RW). First, the model order p of the time-varying multivariate autoregressive (MVAR) model is estimated using the Bayesian information criterion (BIC). Afterwards, a general linear Kalman filter (GLKF) is used to extract the MVAR model parameters that are then fed to the connectivity estimation algorithm. After the time-warping of the causality time series in order to temporally align the epochs (if needed), the causality values are baseline corrected and averaged across trials. A significance metric (sig) is subsequently computed in RW for each couple of channels and compared against a null permutation distribution. This distribution is obtained by shuffling 1000 times the causality time series for each pair i, j $(i \neq j)$ of channels, by computing the sig_{ij} values and retaining for each permutation only the maximum across all i, j pairs.

2.1.3. Identification of significant directed connections

Each sig metric is computed for each pair i, j $(i \neq j)$ of channels, and the $D(D-1)$ sig values are stored.

The time samples of the connectivity series are then shuffled N times and the sig value is re-computed for each permutation. To construct the null (permutation) distribution and to control for false discovery rate [46, 47], the maximum sig value across all the series for each permutation is retained. Significance is then assigned to connections between pairs of electrodes whose sig values are above the 95th percentile of the permutation distribution.

It is important to note that the null distribution can be also computed in many different ways, for example by circularly and independently shifting the time series [50] or by randomizing their phases [51]. Supplementary data figures S1 and S2 (available online at stacks.iop.org/JNE/18/056041/mmedia) show the results after testing the pipeline with the latter two methods of estimating the null distribution.

2.2. Case study

To validate our approach and evaluate which combination of causality measures and RW-significance metrics performs best in our SEEG scenario, six real data-derived channels with a known connectivity profile were simulated (figure 2(A)). These channels were then added to the real datasets from which they were derived to test the pipeline in a typical SEEG analysis scenario.

2.2.1. The dataset

The real SEEG data were recorded in ten patients (5 female, median age 32, range 17–44) who underwent surgical implantation of multiple leads with intracerebral electrodes for refractory epilepsy at the 'Claudio Munari' Epilepsy Surgery Center of Milan, Italy [52, 53]. The strategy of implantation was defined purely based on clinical needs. During the experiment, the patients were presented with a set of auditory stimuli, consisting of pairs of Italian sentences with a homophonous phrase (i.e. with the same acoustic content) which could be interpreted as a noun phrase or a verb phrase, depending on the words preceding and following it. The homophonous part of the stimuli was considered as the RW (figure 2(B)).

A total of 1439 recording contacts were implanted in the grey matter. Channels were referenced to two recording contacts located in the white matter, neutral to electrical stimulation (neutral reference).

Of all the implanted leads, 242 exhibited task-specific high gamma activity (150–300 Hz) and were thus retained for the subsequent analysis (median 14.5, range 2–71). The number of responsive recording contacts for each subject is in supplementary data table 1. The sampling rate was set to 1000 Hz, and trials were extracted from 1.5 s before stimulus onset, with a duration of 6 s.

2.2.2. Surrogate data modeling

For each subject, channels were modeled using a MVAR model with a model order p equal to that estimated by the BIC. The estimation was carried out

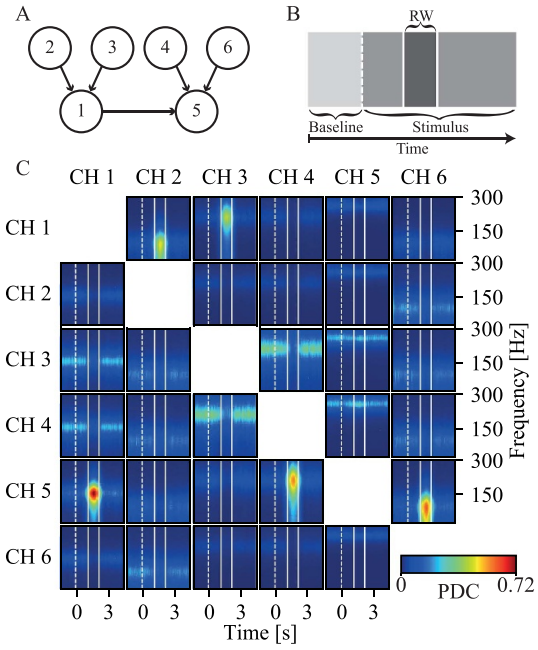

Figure 2. (A) Graph highlighting the directed connections (arrows) simulated in the six-nodes surrogate network during the RW. (B) Schematic of an experimental paradigm in which a time-varying stimulus is delivered, and the interest lies in the electrophysiological response specific for, exclusively, a part of the stimulus, the response window (RW). (C) Time-frequency plots of the partial directed coherence (PDC) of the six surrogate SEEG channels. The horizontal axis represents the time in seconds (relative to the onset of the stimulus), frequencies are on the vertical axis, and the colour scale (bottom right) encodes the value of the PDC, from 0 (blue) to its maximum (red). The plot in the *i*th row and *j*th column shows the PDC from channel *j* to channel *i*. The three white vertical lines in each plot indicate the onset of the stimulus (dashed line) and the start and end of the RW (solid line), respectively.

based on the concatenation of all trial baselines and independently for each subject (subject-by-subject model orders are in supplementary data table 1). By imposing the autoregressive coefficients' values, it was possible to obtain arbitrary connectivity relations.

To model the surrogate channels, the distribution of the MVAR coefficients of the real data (estimated using a GLKF) in the different segments of the stimulus was used. More specifically, to model the baseline and the response to the parts of the stimulus outside the RW, the time-varying MVAR coefficients were sampled from a normal distribution with mean and standard deviation equal to those estimated from real data. The estimation was done on the distribution of the coefficients for each of the three segment blocks (baseline, between baseline and the RW, and from the RW to the end of the stimulus), across time samples, trials, and channel pairs. To model the increase in connectivity during the RW, the simulation coefficients were set as the 99th percentile of the distribution of MVAR coefficients during the RW. Also, noise was added as various levels (diffusion) of high values of estimated connectivity outside the RW. It is worth noting that noise is also present during RW and baseline, even if it is not parametrized. This noise comes from the MVAR model that was used to surrogate data. It was extracted by sampling from a multivariate normal distribution with zero mean and covariance matrix equal to that estimated from the application of the GLKF to real data. This allowed increasing surrogate and real data similarity. Hence, seven equispaced noise levels were simulated as different percentages of the number of significant time points. A time point was deemed significant if it was above 95% of the amplitudes of the MVAR coefficients in the baseline. The 5th and the 95th percentiles of the distributions of the number of significant time points for each real dataset were used as boundaries of the physiological noise, leading to noise levels ranging from 5% to 47% of supra-threshold (the 95th percentile of baseline values) time samples across all subjects (the individual range is shown in supplementary data table 1). This type of noise modeling allowed to parametrize the strength of the connections outside the RW against the amplitude during the RW, i.e. the signal-to-noise ratio (SNR). The noise diffusion percentages are indeed inversely proportional to the SNR.

The MVAR model coefficient values of these time samples were extracted from a normal distribution with mean and standard deviation equal to those estimated from the significant time points in the real data. The noise levels were then increased beyond the physiological maximum up to 99% and with a step of 5%. This analysis aimed at locating the point at which performances started to heavily degrade, and to test which causality/significance combinations were more robust against substantial noise.

PDC values calculated on the resulting six surrogate channels for one example subject are shown in figure 2(C), with noise diffusion outside the RW corresponding to 7.63%.

2.2.3. Pipeline testing

To test the pipeline the surrogate channels were added to the corresponding real datasets. This is necessary to simulate a real-world analysis scenario, especially regarding the number of channels undergoing the analysis. In fact, most MVAR estimation algorithms suffer performance degradation with increasing dataset dimensionality [31, 49]. Subsequently, the connectivity metrics described before (DTF, wDTF, dDTF, SdDTF, PDC and wPDC,) are calculated.

Linear interpolation time-warping was used to align the RW across all trials, before trial averaging of the causality time-frequency matrices [54–57]. The SNR is expected to be large in the context of averaged-across-trials ERC [58].

Note that all the causalities are functions of both time and frequency, but, for simplicity, the mean across all frequencies was retained and used for the subsequent computation of all the sig values. After

significance is assigned to a connection between two channels through the permutation test, a binary connectivity matrix with ones in case of detected connections and zeros otherwise is obtained for each combination of causality measure, sig metric, and level of outside RW noise diffusion. Accuracy, sensitivity, and specificity were thus computed by comparing the obtained connectivity matrices with the reference one, as follows:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}} \tag{7}$$

$$\text{sensitivity} = \frac{\text{TP}}{\text{P}} \tag{8}$$

$$\text{specificity} = \frac{\text{TN}}{\text{N}} \tag{9}$$

where TP is the number of correctly identified connections, TN is the number of correctly rejected connections, P is the number of RW connections within surrogate channels, and N is the difference between the number of all possible connections within surrogate channels and P.

Another important metric to assess the goodness of the estimation is the number of detected spurious connections, i.e. the number of significant connections between the surrogate and real data (ideally it should be zero).

All statistical testing was done using Friedman tests with Nemenyi post-hoc [59, 60]. Non-normality was assessed using the Shapiro–Wilk test [61]. Significance was assigned at $p < 0.01$.

3. Results

Figure 3 shows the accuracies, sensitivities, specificities, and the number of spurious connections of all the causality measures, noise diffusion levels, and significance metrics (RW mean, Mean Ratio, Significant Max, and Significant Mean), averaged across subjects. Only the combinations PDC-RW Mean and PDC-Mean Ratio achieved perfect accuracy (figure 3(A)). All the causality measures in combination with the RW Mean metric (apart from dDTF) showed the best specificity for all the noise diffusion levels, while only PDC, in combination with the *Mean Ratio* metric, achieved similar values. For the Significant Mean and Significant Max metrics, the PDC, wPDC, and SdDTF reached the highest specificity (figure 3(B)). All causality measures in combination with the Mean Ratio metric showed the greatest sensitivity for all physiological noise diffusion levels, with PDC performing equally well in combination with RW Mean (figure 3(C)). The number of spurious connections detected by the RW Mean was almost null, with just 8 for the wDTF in one subject and 2 for the SdDTF in another. All causality measures but PDC identified spurious connections in combination with all the other significance metrics for all the physiological noise diffusion levels, with numbers exceeding 100 using the Mean Ratio metric (figure 3(D)). Interestingly, the two Significant Mean and Significant Max metrics always performed equally.

To investigate the dependency of accuracy, sensitivity, and specificity on the level of outside RW noise diffusion, the trials of the surrogate data were randomly resampled with replacement. The ERC estimation pipeline was executed for each resampling. A series of linear regression analyses on the accuracy, sensitivity, and specificity values did not reveal any significant effect of the level of outside RW noise diffusion (Pearson's ρ not significantly different from zero, significance level adjusted with Bonferroni). Increasing the noise levels beyond the physiological maximum up to 99% revealed a decrease in the pipeline estimation accuracy starting well beyond physiological plausible levels (median 60%, range 30%–99%), with the PDC-RW Mean combination remaining the best performing one for almost every subject.

The general performance of each causality measure and significance metric was independently evaluated. For each causality measure, its accuracy, sensitivity, specificity, and the number of spurious connections were combined disregarding the significance metric and the level of noise diffusion outside the RW. This revealed that: (a) for accuracy, PDC performed better than all the other causality measures; (b) PDC, SdDTF and wPDC showed higher specificity than wDTF, dDTF, and DTF; (c) wPDC showed the lowest sensitivity; (d) PDC, SdDTF and wPDC manifested less spurious connections than wDTF, dDTF, and DTF (figure 4(A)).

The same procedure was applied to each significance metric, revealing that: (a) RW Mean had the highest accuracy and specificity and the lowest number of spurious connections; (b) Mean Ratio showed the best sensitivity and the largest number of spurious connections, while performing poorly in terms of specificity; (c) Significant Mean and Significant Max always performed equally, showing the lowest accuracies and sensitivities (figure 4(B)).

The subject-by-subject results are reported in table 2. The PDC-RW Mean and the PDC-Mean Ratio combinations reached 100% accuracy in classifying the presence or absence of connections. No spurious connections were detected using the RW Mean significance metric in 8 out of 10 subjects for all noise levels. Increasing the noise beyond physiological boundaries led to spurious connections in the PDC/Mean Ratio case for 7 out of 10 subjects. Interestingly, the only three subjects for which the PDC-Mean Ratio combination outperformed the PDC-RW Mean showed a higher physiological noise level (Mann–Whitney $U_{3,7} = 21$, $p < 0.05$).

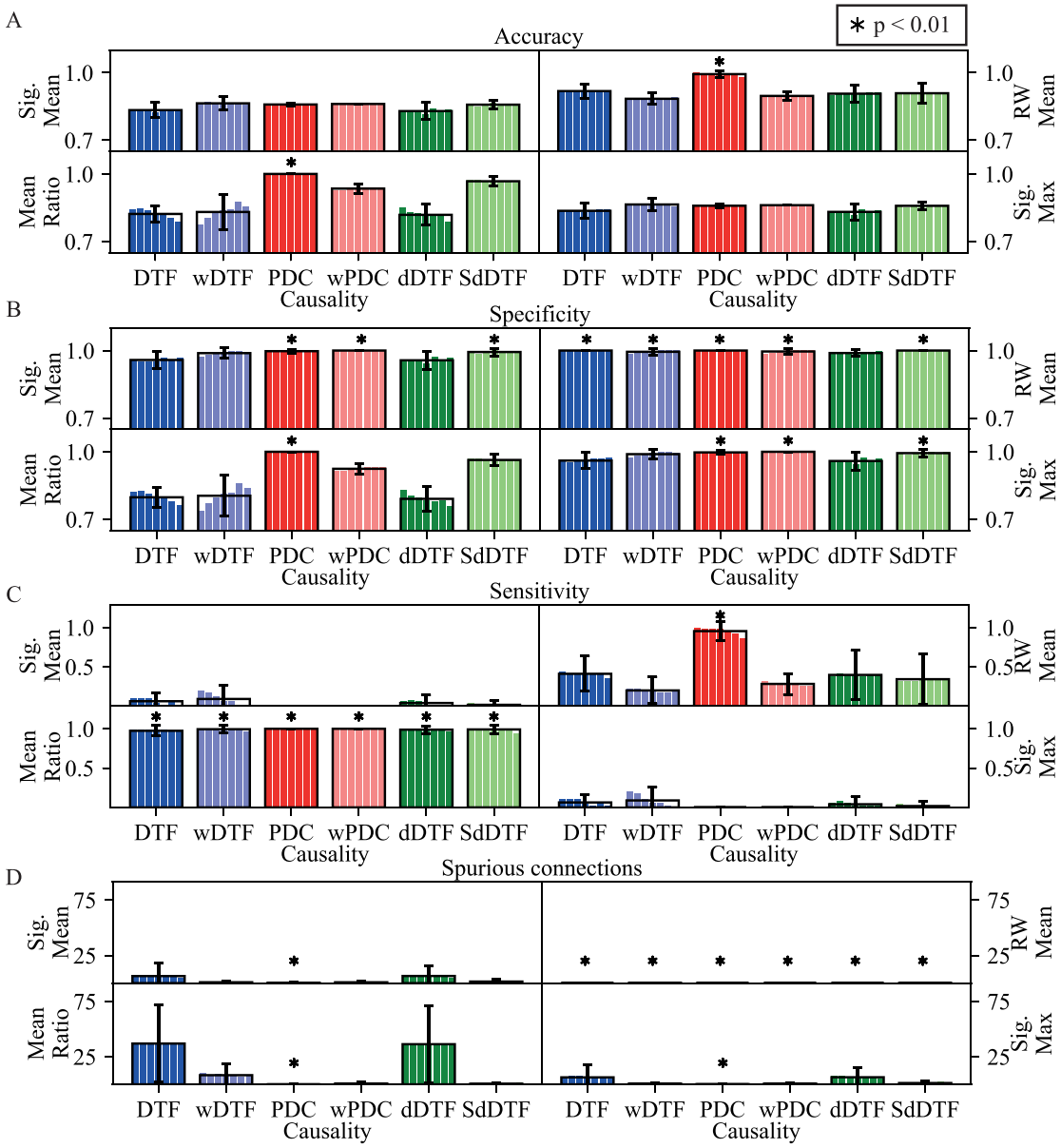

Figure 3. Bar plots of the accuracy (A), specificity (B), sensitivity (C) and the number of spurious connections (D) of the estimated connectivity matrices, averaged across subjects. Each colored group represents a causality metric, while each quadrant represents a significance metric, i.e. Significant Mean (top left), RW Mean (top right), Mean Ratio (bottom left), Significant Max (bottom right). The seven bars in each group represent the levels of diffusion of connectivity outside the RW. The black empty bars are the mean values calculated over subjects and noise levels. The error bars are their standard deviations. The asterisks represent all the combinations of causality and significance metrics which performed better (Nemenyi post-hoc after Friedman test, $p < 0.01$) for the evaluation metric in the corresponding panel title. Multiple asterisks in the same panel indicate that those combinations of causality and significance metric performed equally amongst them, but better with respect to the others.

The analysis was repeated using $p = 2$, yielding very similar results (supplementary data figures S3 and S4), indicating the efficacy of the pipeline even when lowering its computational complexity.

4. Discussion

Despite recent growing interest, the SEEG has not been often used before to investigate ERC patterns in brain circuitry. Here we propose a novel statistical approach to identify significant directed interactions in SEEG datasets. We showed the plausibility of our procedure by also testing the effectiveness of several causality measures on surrogate data.

The pipeline we defined to identify active directed connections during a RW can be so summarized: (a) estimation of the time-varying connectivity matrix; (b) computation of a significance metric that highlights the increase in connectivity during the RW; and (c) the comparison of this significance metric with a null distribution derived by randomly shuffling the connectivity time samples.

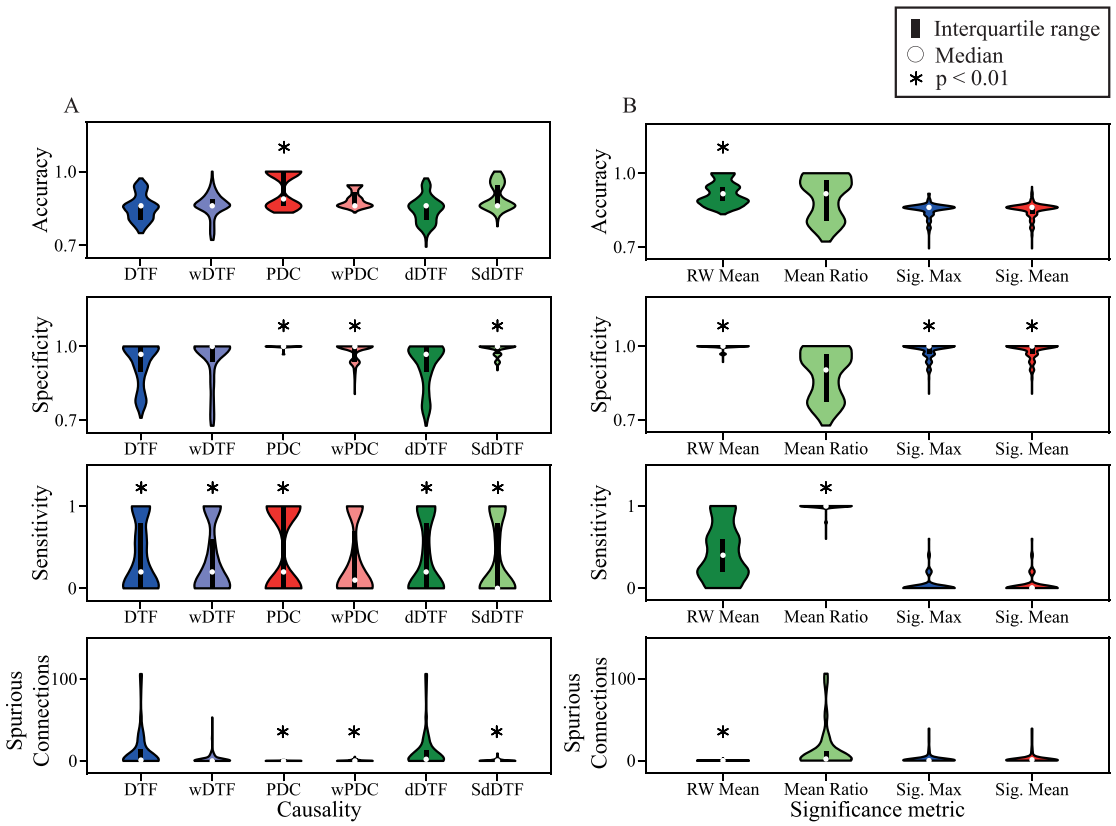

Figure 4. Violin plots of the accuracy, specificity, sensitivity, and number of spurious connections for each causality measure (A) and significance metric (B), merged across subjects and noise levels. The black bars and the white dots represent the interquartile ranges and the medians of the distributions, respectively. The asterisks represent all the best-performing causality measures or significance metrics (Nemenyi post-hoc after Friedman test, $p < 0.01$) according to the evaluation metric in the corresponding panel.

Table 2. The first three columns are causality measure(s), significance metric(s) and combination(s) yielding the highest accuracy for each subject (Nemenyi post-hoc after Friedman test, $p < 0.01$). The fourth column is the noise percentage threshold at which performances of the best combination start to degrade. In case of two or more combinations with no statistical difference in accuracy scores, the one with the higher noise threshold is selected as the best.

Subj	Best causality measure(s)	Best significance metric(s)	Best combination(s)	Noise threshold (%)
S01	PDC	RW Mean	PDC-RW Mean	60
S02	PDC	RW Mean	PDC-RW Mean	55
S03	PDC	RW Mean/Mean Ratio	PDC-Mean Ratio	90
S04	PDC/SdDTF	Mean Ratio	PDC-RW Mean	45
S05	PDC	RW Mean/Mean Ratio	PDC-Mean Ratio	50
S06	PDC	RW Mean	PDC-RW Mean	55
S07	PDC/SdDTF	RW Mean	PDC-RW Mean	75
S08	PDC/SdDTF	RW Mean/Mean Ratio	PDC-Mean Ratio	99
S09	PDC	RW Mean	PDC-RW Mean	60
S10	PDC/SdDTF	RW Mean	PDC-RW Mean	65

We also proposed an assessment procedure of an ERC pipeline based on surrogate data derived from a real SEEG dataset, to make the simulated data more physiologically plausible. A previous work on time-varying connectivity [13] proposed a method based on the assumptions of Gaussianity for the distribution of the connectivity values and of temporal independence, with Bonferroni correction for multiple comparisons. Here, instead, we propose a non-parametric approach, that is ideal to overcome the limitations caused by assuming the independence between causality time samples and by the overly-conservative Bonferroni correction of p-values [62].

It should also be noted that the permutation method can be applied to any type of time-varying estimator, whenever a clearly defined RW is available. Common scenarios are the event-related potential analysis [63] if it is applied directly on the signal time series, or event-related spectral perturbation [64] if it is computed on the signal scalogram.

Our pipeline relies on two important parameters: (a) the causality measure used to estimate the connectivity matrix; (b) the significance metric that highlights the increase in connectivity during the RW. We were able to identify the combination of these two parameters showing the theoretical maximum similarity to the ground truth. We achieved that by evaluating the accuracy, sensitivity, specificity, and the number of spurious connections derived from the application of our algorithm on benchmark surrogate data. We also investigated the effect of increasing levels of noise diffusion outside the RW. Finally, lowering the model order p of the causality estimation did not significantly alter the results, suggesting a way to make the pipeline less computationally intensive without affecting its effectiveness.

4.1. Performance of the causality measures

The causality measures we focused on in this work are based on the framework of the Geweke–Granger causality. More in detail, these measures can be divided into two groups: those derived from the partial directed coherence (the PDC itself and its weighted version, namely, the wPDC), and those built upon the directed transfer function (the DTF itself, the wDTF and the ones weighted by the partial coherence, i.e. the dDTF and the SdDTF).

Among these measures, the PDC was the only one with perfect accuracy in estimating the connectivity matrix active during the RW of our SEEG surrogate data. Its power spectral density-weighted version, the wPDC, showed decreased accuracy and sensitivity, with no significant difference in specificity or number of spurious connections. These results suggest that the weighting of the PDC by the power spectral density, while potentially enhancing the biological interpretability of the observed directed connections, may, in turn, degrade the accuracy in the estimation of the connectivity matrix by limiting the number of detected significant connections.

The DTF performed poorly in terms of accuracy, specificity, and number of spurious connections while reaching the same values of the PDC for the sensitivity. These findings show evidence of a higher number of estimated directed connections than the ground truth, which may be caused by its inability to distinguish between direct and indirect connections. Finally, the SdDTF offered a significant improvement over the DTF. Specifically, the higher specificity is an indicator of its ability to distinguish between direct and indirect connections.

In summary, the PDC showed to be the most effective among the causality measures tested here, in agreement with previous studies [27–29].

However, it should be noted that the same causality measures may perform differently in combination with other significance metrics or other ERC estimation techniques. Here we also leveraged a small set of all the possibly available causality measures, but a full benchmark of these algorithms is beyond the scope of the paper. This section aims to suggest to the reader which measures to use together with this pipeline.

4.2. Performance of the significance metrics

We defined a significance metric as a mathematical expression able to quantify the increase in the causality amplitude during a RW when compared against a null distribution. In this work, we suggested four different significance metrics. Nonetheless, other metrics (e.g. a metric based on the SNR [50]) may also be used with this pipeline.

Among all significance features (RW Mean, Mean Ratio, Significant Max, and Significant Mean), RW Mean reached the highest accuracy, perfect sensitivity, and an almost null number of spurious connections. Mean Ratio reached, most of the time, a perfect sensitivity, but at the expense of presenting the lowest specificity and the highest number of spurious connections. While the RW Mean metric in itself does not give any hint about the increase in connectivity during the RW, the comparison of its value with the estimated permutation distribution, does: shuffling the time samples in each permutation allows the comparison with the causality values corresponding to the other segments of the stimulus. By retaining, for each permutation, only the highest sig value among channel pairs, the RW Mean is able to highlight the causality values that are among the strongest and truly significant.

4.3. Performance of the causality measure and significance metric combinations

In our case study, the combinations PDC-RW Mean and PDC-Mean Ratio achieved the highest accuracies in estimating the active connections between channels during the RW. This, together with the RW Mean's lowest number of spurious connections, average higher accuracy and specificity, and the overall better performance of the PDC, suggest that the PDC-RW Mean combination might be the best choice for the estimation of the ERC in SEEG data. However, one may be interested in minimizing the type II error of the test (number of false negatives) at the expense of a higher type I error (i.e. a higher number of spurious connections and reduced specificity). In this scenario, we suggest the use of the PDC-Mean Ratio combination instead.

Also, linear regression analysis performed on the causality estimated on the resampled data highlighted the absence of any effect of the noise diffusion level outside the RW (for physiologically plausible values). Even increasing the noise levels beyond the threshold that degrades the accuracy of all parameters combinations, the PDC-RW Mean almost always showed the best potential.

Finally, we observed high consistency of the results among different subjects. Perfect accuracy was reached for all of them in identifying the simulated connections using the pipeline with PDC-Mean Ratio and/or PDC-RW Mean. Out of ten subjects, seven showed an overall better accuracy using the PDC and the RW Mean significance metric. Gender and number of recording contacts did not have a significant effect on the results, while physiological noise levels did. In fact, in the three subjects for which the Mean Ratio metric outperformed the RW Mean, we observed a lower SNR. This paves the way to future work, in which the dependency of the noise on the effectiveness of the significance metrics may be further investigated, for example, by changing the noise parametrization method and parametrizing even the noise in the RW.

However, the procedure outlined is quite robust against physiologically plausible noise levels, as modeled in our surrogate SEEG data.

4.4. Other factors affecting causality estimation

It is worth pointing out that the performance of an ERC estimation workflow depends on several elements. For example, the choice of the algorithm used to estimate the MVAR model parameters is crucial. Here, we used the GLKF, which is a parametric time-varying estimation method. This family of methods is to be preferred to the ones based on moving windows, which would require a stationarity assumption on the signal contained in the windows. This would bring further limitations to the results [29]. Among time-varying MVAR estimation methods, the GLKF outperformed other algorithms, such as the recursive least square, the multivariate adaptive autoregressive estimator, the classic Kalman filter, and the dual extended Kalman filter [29, 31, 49]. The GLKF is a parametric method, but there exist approaches that allow approximating the system transfer function through multitaper or wavelet transform spectral estimation. However, these approaches are dependent on parameters that may in turn reduce the spectral resolution. This limitation is absent in parametric methods [65]. However, the GLKF relies on an adaptation constant. This adaptation coefficient was set equal to 0.03, a value laying in the optimal range for the GLKF [31, 49].

Although it is possible to also use non-linear methods to estimate the connectivity matrices, linear ERC evaluation techniques are able to detect any type of information flow changes, regardless of whether they are due to linear or non-linear dynamics [5, 66–69].

Another factor possibly affecting the performance of our pipeline is the number of permutations performed to estimate the null distribution. Ideally, to get the true permutation distribution, all possible permutations of the time samples should be tested. Since this is not feasible due to computational constraints, there is the need to choose a number of permutations large enough to have a reliable estimation. Previous studies [47, 70] showed that 1000 permutations are enough to well approximate the null distribution.

The way the null distribution was estimated did not affect the main results. Supplementary data figures S1 and S2 show comparable results (i.e. PDC-RW Mean as the best combination, with no significant effect of physiological noise on the accuracy) when estimating the null distribution by randomly shifting the connectivity time series or by randomizing their phases.

5. Conclusions

Investigating the strength, direction and spectral content of brain network interactions with ever-increasing detail is the goal of most neuroscientists. The unparalleled spatial specificity and sub-ms temporal resolution of SEEG make it a very promising technique for analyzing effective connectivity in cognitive tasks. However, the lack of standardized and validated approaches explicitly designed for SEEG has, so far, limited its potential.

Here we proposed a pipeline to evaluate causal information flow in SEEG datasets, that notably adds to the tools currently available to neuroscientists. Among the various estimator we tested with this pipeline, we suggest the use of PDC together with the here defined RW Mean as they performed the best in time-varying causality evaluation. To our knowledge, this is the first validated ERC analysis approach designed explicitly for SEEG data.

Finally, we note that the proposed workflow can be applied to any type of time-varying estimator, whenever a clearly defined RW is available.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This work was financed by the Italian Ministry of Education, University and Research (MIUR) through PRIN-2017 ‘INSPECT’ (Project 2017JPMW4F) and by the Bertarelli Foundation.

We would like to acknowledge Professor Andrea Moro for defining the stimuli and Professor Stefano Cappa for the precious discussion about the work.

ORCID iDs

Andrea Cometa https://orcid.org/0000-0002-5771-0316
Piergiorgio D'Orio https://orcid.org/0000-0001-9757-2061
Martina Revay https://orcid.org/0000-0003-2270-5733
Silvestro Micera https://orcid.org/0000-0003-4396-8217
Fiorenzo Artoni https://orcid.org/0000-0002-0967-6643

References

[1] Iida K and Otsubo H 2017 Stereoelectroencephalography: indication and efficacy *Neurol. Med. Chir.* **57** 375–85
[2] Bartolomei F, Lagarde S, Wendling F, McGonigal A, Jirsa V, Guye M and Bénar C 2017 Defining epileptogenic networks: contribution of SEEG and signal analysis *Epilepsia* **58** 1131–47
[3] Artoni F *et al* 2020 High gamma response tracks different syntactic structures in homophonous phrases *Sci. Rep.* **10** 7537
[4] He B, Astolfi L, Valdes-Sosa P A, Marinazzo D, Palva S O, Benar C-G, Michel C M and Koenig T 2019 Electrophysiological brain connectivity: theory and implementation *IEEE Trans. Biomed. Eng.* **66** 2115–37
[5] Korzeniewska A, Franaszczuk P J, Crainiceanu C M, Kuś R and Crone N E 2011 Dynamics of large-scale cortical interactions at high gamma frequencies during word production: event related causality (ERC) analysis of human electrocorticography (ECoG) *NeuroImage* **56** 2218–37
[6] Herff C, Krusienski D J and Kubben P 2020 The potential of stereotactic-EEG for brain-computer interfaces: current progress and future directions *Front. Neurosci.* **14** 123
[7] Lachaux J P, Rudrauf D and Kahane P 2003 Intracranial EEG and human brain mapping *J. Physiol.* **97** 613–28
[8] Canolty R T, Edwards E, Dalal S S, Soltani M, Nagarajan S S, Kirsch H E, Berger M S, Barbaro N M and Knight R T 2006 High gamma power is phase-locked to theta oscillations in human neocortex *Science* **313** 1626–8
[9] Ray S, Niebur E, Hsiao S S, Sinai A and Crone N E 2008 High-frequency gamma activity (80–150 Hz) is increased in human cortex during selective attention *Clin. Neurophysiol.* **119** 116–33
[10] Edwards E, Soltani M, Deouell L Y, Berger M S and Knight R T 2005 High gamma activity in response to deviant auditory stimuli recorded directly from human cortex *J. Neurophysiol.* **94** 4269–80
[11] Kucewicz M T *et al* 2014 High frequency oscillations are associated with cognitive processing in human recognition memory *Brain* **137** 2231–44
[12] Gaona C M, Sharma M, Freudenburg Z V, Breshears J D, Bundy D T, Roland J, Barbour D L, Schalk G and Leuthardt E C 2011 Nonuniform high-gamma (60–500 Hz) power changes dissociate cognitive task and anatomy in human cortex *J. Neurosci.* **31** 2091–100
[13] Korzeniewska A, Crainiceanu C M, Kuś R, Franaszczuk P J and Crone N E 2008 Dynamics of event-related causality in brain electrical activity *Hum. Brain Mapp.* **29** 1170–92
[14] Buzsáki G 2004 Neuronal oscillations in cortical networks *Science* **304** 1926–9
[15] Engel A K and Singer W 2001 Temporal binding and the neural correlates of sensory awareness *Trends Cogn. Sci.* **5** 16–25
[16] Varela F, Lachaux J-P, Rodriguez E and Martinerie J 2001 The brainweb: phase synchronization and large-scale integration *Nat. Rev. Neurosci.* **2** 229–39
[17] Penny W D, Stephan K E, Mechelli A and Friston K J 2004 Modelling functional integration: a comparison of structural equation and dynamic causal models *NeuroImage* **23** S264–74
[18] Iturrate I, Chavarriaga R, Pereira M, Zhang H, Corbet T, Leeb R and Millán J D R 2018 Human EEG reveals distinct neural correlates of power and precision grasping types *NeuroImage* **181** 635–44
[19] Randazzo L, Iturrate I, Chavarriaga R, Leeb R and Millan J D R 2015 Detecting intention to grasp during reaching movements from EEG *37th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC) (Milan, Italy, 25–29 August 2015)* 1115–8
[20] Flinker A, Korzeniewska A, Shestyuk A Y, Franaszczuk P J, Dronkers N F, Knight R T and Crone N E 2015 Redefining the role of Broca's area in speech *Proc. Natl Acad. Sci.* **112** 2871–5
[21] Granger C W J 1969 Investigating causal relations by econometric models and cross-spectral methods *Econometrica* **37** 424
[22] Geweke J 1982 Measurement of linear dependence and feedback between multiple time series *J. Am. Stat. Assoc.* **77** 304–13
[23] Kaminski M J and Blinowska K J 1991 A new method of the description of the information flow in the brain structures *Biol. Cybern.* **65** 203–10
[24] Baccalá L A and Sameshima K 2001 Partial directed coherence: a new concept in neural structure determination *Biol. Cybern.* **84** 463–74
[25] Korzeniewska A, Mańczak M, Kamiński M, Blinowska K J and Kasicki S 2003 Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method *J. Neurosci. Methods* **125** 195–207
[26] Plomp G, Quairiaux C, Michel C M and Astolfi L 2014 The physiological plausibility of time-varying Granger-causal modeling: normalization and weighting by spectral power *NeuroImage* **97** 206–16
[27] Van Mierlo P, Lie O, Staljanssens W, Coito A and Vulliémoz S 2018 Influence of time-series normalization, number of nodes, connectivity and graph measure selection on seizure-onset zone localization from intracranial EEG *Brain Topogr.* **31** 753–66
[28] Omidvarnia A H, Mesbah M, Khlif M S, O'Toole J M, Colditz P B and Boashash B 2011 Kalman filter-based time-varying cortical connectivity analysis of newborn EEG *2011 Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (Boston, MA, USA, 30 August–3 September 2011)* 1423–6
[29] Ghumare E G, Schrooten M, Vandenberghe R and Dupont P 2018 A time-varying connectivity analysis from distributed EEG sources: a simulation study *Brain Topogr.* **31** 721–37
[30] Wilke C, Ding L and He B 2008 Estimation of time-varying connectivity patterns through the use of an adaptive directed transfer function *IEEE Trans. Biomed. Eng.* **55** 2557–64
[31] Milde T, Leistritz L, Astolfi L, Miltner W H R, Weiss T, Babiloni F and Witte H 2010 A new Kalman filter approach for the estimation of high-dimensional time-variant multivariate AR models and its application in analysis of laser-evoked brain potentials *NeuroImage* **50** 960–9
[32] Rosenberg J R, Halliday D M, Breeze P and Conway B A 1998 Identification of patterns of neuronal connectivity—partial spectra, partial coherence, and neuronal interactions *J. Neurosci. Methods* **83** 57–72
[33] Astolfi L, Cincotti F, Mattia D, Marciani M G, Baccala L A, Fallani F D V, Salinari S, Ursino M, Zavaglia M and Babiloni F 2006 Assessing cortical functional connectivity by partial directed coherence: simulations and application to real data *IEEE Trans. Biomed. Eng.* **53** 1802–12
[34] Astolfi L *et al* 2007 Estimate of causality between independent cortical spatial patterns during movement volition in spinal cord injured patients *Brain Topogr.* **19** 107–23

[35] Fasoula A, Attal Y and Schwartz D 2013 Comparative performance evaluation of data-driven causality measures applied to brain networks *J. Neurosci. Methods* **215** 170–89
[36] Florin E, Gross J, Pfeifer J, Fink G R and Timmermann L 2011 Reliability of multivariate causality measures for neural data *J. Neurosci. Methods* **198** 344–58
[37] Astolfi L *et al* 2008 Tracking the time-varying cortical connectivity patterns by adaptive multivariate estimators *IEEE Trans. Biomed. Eng.* **55** 902–13
[38] Leistritz L *et al* 2013 Time-variant partial directed coherence for analysing connectivity: a methodological study *Phil. Trans. R. Soc.* A **371** 20110616
[39] Seth A K, Chorley P and Barnett L C 2013 Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling *NeuroImage* **65** 540–55
[40] Wen X, Rangarajan G and Ding M 2013 Is Granger causality a viable technique for analyzing fMRI data? *PLoS One* **8** 1–11
[41] Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R and Bressler S L 2004 Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality *Proc. Natl Acad. Sci. USA* **101** 9849–54
[42] Seth A K 2010 A MATLAB toolbox for Granger causal connectivity analysis *J. Neurosci. Methods* **186** 262–73
[43] Astolfi L *et al* 2007 Comparison of different cortical connectivity estimators for high-resolution EEG recordings *Hum. Brain Mapp.* **28** 143–57
[44] Wang H E, Bénar C G, Quilichini P P, Friston K J, Jirsa V K and Bernard C 2014 A systematic framework for functional connectivity measures *Front. Neurosci.* **8** 405
[45] Wu M-H, Frye R E and Zouridakis G 2011 A comparison of multivariate causality based measures of effective connectivity *Comput. Biol. Med.* **41** 1132–41
[46] Maris E and Oostenveld R 2007 Nonparametric statistical testing of EEG- and MEG-data *J. Neurosci. Methods* **164** 177–90
[47] Nichols T E and Holmes A P 2002 Nonparametric permutation tests for functional neuroimaging: a primer with examples *Hum. Brain Mapp.* **15** 1–25
[48] Schwarz G 1978 Estimating the dimension of a model *Ann. Stat.* **6** 461–4
[49] Pagnotta M F, Plomp G and Sendiña-Nadal I 2018 Time-varying mvar algorithms for directed connectivity analysis: critical comparison in simulations and benchmark EEG data *PLoS One* **13** e0198846
[50] Crowther L J, Brunner P, Kapeller C, Guger C, Kamada K, Bunch M E, Frawley B K, Lynch T M, Ritaccio A L and Schalk G 2019 A quantitative method for evaluating cortical responses to electrical stimulation *J. Neurosci. Methods* **311** 67–75
[51] Brunner M I, Bárdossy A and Furrer R 2019 Technical note: stochastic simulation of streamflow time series using phase randomization *Hydrol. Earth Syst. Sci.* **23** 3175–87
[52] Munari C, Hoffmann D, Fracione S, Kahane P, Tassi L, Russo G L and Benabid A L 1994 Stereo-electroencephalo graphy methodology: advantages and limits *Acta Neurol. Scand. Suppl.* **152** 56–67 discussion 68–69
[53] Cossu M, Cardinale F, Castana L, Citterio A, Francione S, Tassi L, Benabid A L and Lo Russo G 2005 Stereoelectroencephalography in the presurgical evaluation of focal epilepsy: a retrospective analysis of 215 procedures *Neurosurgery* **57** 706–18
[54] Artoni F, Fanciullacci C, Bertolucci F, Panarese A, Makeig S, Micera S and Chisari C 2017 Unidirectional brain to muscle connectivity reveals motor cortex control of leg muscles during stereotyped walking *NeuroImage* **159** 403–16
[55] Nordin A D, Hairston W D and Ferris D P 2019 Human electrocortical dynamics while stepping over obstacles *Sci. Rep.* **9** 4693
[56] Do T-T N, Lin C-T and Gramann K 2021 Human brain dynamics in active spatial navigation *Sci. Rep.* **11** 13036
[57] Gwin J T, Gramann K, Makeig S and Ferris D P 2011 Electrocortical activity is coupled to gait cycle phase during treadmill walking *NeuroImage* **54** 1289–96
[58] Handy T C 2005 *Event-related Potentials: A Methods Handbook* (Cambridge, MA: MIT Press)
[59] Friedman M 1937 The use of ranks to avoid the assumption of normality implicit in the analysis of variance *J. Am. Statist. Assoc.* **32** 675–701
[60] Nemenyi Peter Björn 1963 Distribution-free multiple comparisons *PhD Thesis* Princeton University
[61] Shapiro S S, Wilk M B 1965 An Analysis of Variance Test for Normality (Complete Samples) *Biometrika* **52** 591
[62] Nakagawa S 2004 A farewell to Bonferroni: the problems of low statistical power and publication bias *Behav. Ecol.* **15** 1044–5
[63] Luck S J 2012 Event-related potentials *APA Handbook of Research Methods in Psychology, Vol 1: Foundations, Planning, Measures, and Psychometrics* (Washington, DC: American Psychological Association) pp 523–46
[64] Grandchamp R and Delorme A 2011 Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials *Front. Psychol.* **2** 236
[65] Wang C, Rajagovindan R, Han S M and Ding M 2016 Top-down control of visual alpha oscillations: sources of control signals and their mechanisms of action *Front. Hum. Neurosci.* **10** 15
[66] Chávez M, Martinerie J and Le Van Quyen M 2003 Statistical assessment of nonlinear causality: application to epileptic EEG signals *J. Neurosci. Methods* **124** 113–28
[67] Freiwald W A, Valdes P, Bosch J, Biscay R, Jimenez J C, Rodriguez L M, Rodriguez V, Kreiter A K and Singer W 1999 Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex *J. Neurosci. Methods* **94** 105–19
[68] Gourévitch B, Le Bouquin-Jeannès R and Faucon G 2006 Linear and nonlinear causality between signals: methods, examples and neurophysiological applications *Biol. Cybern.* **95** 349–69
[69] Franaszczuk P J and Bergey G K 1999 An autoregressive method for the measurement of synchronization of interictal and ictal EEG signals *Biol. Cybern.* **81** 3–9
[70] Edgington E S 1969 Approximate randomization tests *J. Psychol. Interdiscip. Appl.* **72** 143–9