

## Measuring transient phase-amplitude coupling using local mutual information



Ramón Martínez-Cancino<sup>a,b,\*</sup>, Joseph Heng<sup>a,c</sup>, Arnaud Delorme<sup>a</sup>, Ken Kreutz-Delgado<sup>b</sup>, Roberto C. Sotero<sup>d</sup>, Scott Makeig<sup>a</sup>

<sup>a</sup> Swartz Center for Computational Neurosciences, UCSD, La Jolla, CA, USA

<sup>b</sup> Electric and Computer Engineering Department, Jacobs School of Engineering, UCSD, La Jolla, CA, USA

<sup>c</sup> Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

<sup>d</sup> Department of Radiology and Hotchkiss Brain Institute, University of Calgary, Canada

### ARTICLE INFO

#### Keywords:

Phase-amplitude coupling  
Cross-frequency coupling  
ECoG  
Mutual information

### ABSTRACT

Here we demonstrate the suitability of a local mutual information measure for estimating the temporal dynamics of cross-frequency coupling (CFC) in brain electrophysiological signals. In CFC, concurrent activity streams in different frequency ranges interact and transiently couple. A particular form of CFC, phase-amplitude coupling (PAC), has raised interest given the growing amount of evidence of its possible role in healthy and pathological brain information processing. Although several methods have been proposed for PAC estimation, only a few have addressed the estimation of the temporal evolution of PAC, and these typically require a large number of experimental trials to return a reliable estimate. Here we explore the use of mutual information to estimate a PAC measure (MIPAC) in both continuous and event-related multi-trial data. To validate these two applications of the proposed method, we first apply it to a set of simulated phase-amplitude modulated signals and show that MIPAC can successfully recover the temporal dynamics of the simulated coupling in either continuous or multi-trial data. Finally, to explore the use of MIPAC to analyze data from human event-related paradigms, we apply it to an actual event-related human electrocorticographic (ECoG) data set that exhibits strong PAC, demonstrating that the MIPAC estimator can be used to successfully characterize amplitude-modulation dynamics in electrophysiological data.

### 1. Introduction

Since the first observations of human brain (8–12 Hz) alpha waves by Hans Berger in 1929 (Berger, 1929), additional modes of rhythmic or semi-rhythmic cortical field activity have since been observed, including characteristic activities in the (1–4 Hz) delta, (4–8 Hz) theta, (12–30 Hz) beta and (above 30 Hz) gamma frequency bands. For many decades, the study of the spectral dynamics of cortical field oscillations was limited to analysis of phenomena occurring within single frequency bands. Over the last few years, the growing evidence that oscillations at different frequencies interact with each other has gathered attention, leading to heightened interest in cross-frequency coupling (CFC) in brain dynamics. Since high-frequency field activity is often more topologically localized than low frequency rhythms (Buzsaki et al., 2004), CFC could be a mechanism for integrating information across spatial as well as temporal scales.

Although CFC could refer to any possible interaction between frequencies, phases and amplitudes of oscillatory phenomena (Sotero, 2016), most experimental work has focused on three types of CFC: amplitude-amplitude coupling (AAC) or comodulation, phase-phase coupling (PPC) including bicoherence, and phase-amplitude coupling (PAC). Among them, PAC has attracted increasing interest given the growing amount of evidence of its potential role in brain information processing and its changes under pathological conditions including epilepsy (López-Azcárate et al., 2010; De Hemptinne et al., 2013). In PAC, the instantaneous amplitude of a higher frequency band within a signal is modulated by (or otherwise linked to) the instantaneous phase of a lower-frequency band of the same (or a different) signal.

PAC has been observed between various frequency bands, in multiple brain regions, in different task conditions, and in multiple species. Table 1 illustrates the diversity of recent reports. In its diversity, PAC has

\* Corresponding author. Swartz Center for Computational Neurosciences, UCSD, La Jolla, CA, USA.

E-mail address: [ram033@eng.ucsd.edu](mailto:ram033@eng.ucsd.edu) (R. Martínez-Cancino).

<https://doi.org/10.1016/j.neuroimage.2018.10.034>

Received 30 January 2018; Received in revised form 4 October 2018; Accepted 11 October 2018

Available online 18 October 2018

1053-8119/© 2018 Elsevier Inc. All rights reserved.

**Table 1**

Example of studies on Phase-Amplitude Coupling. DBS: Deep Brain Stimulation. LFP: Local field potential. MEG: Magnetoencephalography. MVLmi: Mean Vector Length Modulation Index. KLmi: Kullback-Leibler Modulation Index. PLV: Phase Locking Value. DPAC: Direct PAC estimator. SI: Synchronization Index. PMI-PAC: Permutation Mutual Information based Phase-Amplitude Coupling.

Study	CFC Frequencies	Brain Region	Modality	Study focus	Species	Estimation Method
(Canolty et al., 2006a)	$\theta$ - high- $\gamma$	Neocortex	ECoG	Perception, memory, motor	Human	MVLmi
(Cohen et al., 2009)	$\alpha$ - low- $\gamma$	Nucleus Accumbens	DBS electrodes	Reward	Human	MVLmi
(Tort et al., 2009)	$\theta$ - $\gamma$	Hippocampus	LFP	Learning	Rat	KLmi
(Voytek et al., 2010)	$\theta/\alpha$ - high- $\gamma$	Posterior cortex	ECoG	Perception	Human	PLV
(Yanagisawa et al., 2012)	$\alpha$ - high- $\gamma$	Sensorimotor cortex	ECoG	Motor	Human	SI
(Spaak et al., 2012)	$\alpha$ - low- $\gamma$ /high- $\gamma$	Visual cortex	Laminar electrodes	Spontaneous activity	Monkey	KLmi
(Szczepanski et al., 2014)	$\delta/\theta$ - high- $\gamma$	Fronto-parietal cortex	ECoG	Attention	Human	PLV
(Florin and Baillet, 2015)	$\delta/\theta/\alpha$ - low- $\gamma$ /high- $\gamma$	Resting state networks	MEG	Spontaneous activity	Human	DPAC
(Sotero et al., 2015)	$\delta/\theta/\alpha$ - $\beta$ /low- $\gamma$ /high- $\gamma$	S1	Laminar electrodes	Spontaneous activity	Rat	MVLmi
(Cheng et al., 2017)	$\theta$ - low- $\gamma$	Hippocampus CA3, CA1	LFP	Spontaneous activity	Rat	PMI-PAC

shown that it can reveal unsuspected relationships, e.g., the intriguing PAC found between the basal gastric rhythm and alpha band power in cortex (Richter et al., 2017).

If PAC is a functional aspect of normal brain dynamics, then abnormal PAC could be a cause or symptom of unhealthy brain function. Associations between brain disease and PAC have been found in epilepsy (Nonoda et al., 2016), Parkinson's disease (López-Azcárate et al., 2010; De Hemptinne et al., 2013), Alzheimer's disease (Goutagny et al., 2013), schizophrenia (Allen et al., 2011), obsessive-compulsive disorder (OCD) (Bahramisharif et al., 2016), and mild cognitive impairment (MCI) (Dimitriadis et al., 2015). This makes PAC estimation of interest for clinical studies.

Several methods have been proposed to measure PAC. However, none is currently a gold standard (see Table 1). The three measures most often cited in the PAC literature are: the Mean Vector Length Modulation Index (MVLmi) (Canolty et al., 2006a), the Kullback-Leibler Modulation Index (KLmi) (Tort et al., 2010), and the General Linear Model Modulation Index (GLMmi) (Penny et al., 2008a).

The data processing pipelines for these methods share a common beginning. A first step is to band-pass filter the signal or signals of interest to isolate the frequency ranges to be investigated. Then, an analytical signal is obtained for each frequency band using the Hilbert transform, from which instantaneous phase (of the low-frequency modulatory signal) and instantaneous amplitude (of the high-frequency modulated signal) are obtained. These time series are used then to compute a PAC measure as explained in detail below. To compute statistical significance, the collected sets of phases and/or amplitudes are shuffled many times to create surrogate data sets. Significance is then estimated by determining whether the PAC measure for the actual data belongs or not to the distribution of PAC measures computed on the surrogate data (Tort et al., 2010; Hurtado et al., 2004). The same process can be applied when PAC is assessed between low- and high-frequency dynamics of two different signals, e.g., between different electrophysiological recording channels or separated source activities in different brain areas.

A common drawback of most of the existing PAC measures is their need for a large number of data points to compute a robust PAC estimate. A typical approach is to perform the PAC calculation using a semi-arbitrary time window (Penny et al., 2008b; Tort et al., 2010; Voytek et al., 2013). The minimum length of this window is constrained by the frequency of the lower (phase) frequency of interest; at that frequency the analysis window should be at least one full cycle in length. However, a recent study (Tort et al., 2010) has proposed that because of the sensitivity to noise of existing PAC measures, the number of cycles needed to compute a reliable PAC estimate may be 200 cycles or more.

PAC was first measured across blocks of continuous data, providing relatively low sensitivity to changes in PAC across time. Subsequent work (Cohen, 2008) tried to fill this gap by applying both automated and supervised heuristics to determine the frequencies involved in the coupling and by a combination of existing PAC measures to a moving time window. This method has been proven useful for characterizing dynamics of

electrophysiological processes supporting cognition. However, this method fails to detect CFC in different scenarios; for example, when the amplitude time series does not have prominent components at the frequencies at which it is coupled (Cohen, 2008).

Voytek et al. (2013) addressed the lack of temporal resolution by computing PAC for latencies within multiple trials time locked to a set of similar events, thereby introducing Event-Related Phase Amplitude Coupling (ERPAC). In this approach, a variant of the GLMmi PAC estimator was applied in an event-related manner. However, this approach still cannot provide a trial-by-trial description of PAC variability and requires many experimental events that may be presumed to produce equivalent dynamics.

To disentangle the role(s) of PAC in neurophysiology, the question of *when* it occurs is as central as *where* or *how*. As suggested by Lizier (2014a), local or pointwise information theoretic measures can help answer this question. In this work we propose a new approach to PAC estimation, MIPAC, based on a local mutual information (MI) measure, that allows a dynamic description of PAC within a relatively short time window. The new method is first tested on simulated and then on actual human event-related electrocorticographic (ECoG) data. We also compare MIPAC estimates with the results of other PAC estimators.

The manuscript is organized as follows: In Section 2, we describe the data on which we test the new method, as well as the required data processing. Then, in Section 2.1.3 we comment on the estimation of the instantaneous phase and amplitude in the signal of interest. Next, we give an introduction to traditional PAC measures and briefly review three current PAC estimation methods: Mean Vector Length Modulation Index (Section 2.2.1), Kullback-Leibler Modulation Index (Section 2.2.2) and Modulation Index based on a General Linear Model (Section 2.2.3). Following this, we present an introduction to Information-Theoretical measures and then introduce the MIPAC approach. Section 3 presents validation results of the new methods on simulated data and its application to human electrophysiological data. In Section 4 we discuss the results, some pitfalls and caveats of the proposed method, and its possible uses and extensions.

## 2. Methods

### 2.1. Ethics statement

Ethics statement: The patient data we study here were recorded during monitoring of implanted ECoG electrode grid and strip signals conducted during planning for surgery for epilepsy. The patient participated in a purely voluntary manner, after providing informed written consent, under experimental protocols approved by the Institutional Review Board (IRB) of the University of Washington (# 12193). All patient data were anonymized according to IRB protocol, in accordance with HIPAA mandates. These data originally appeared in the manuscript, "Spontaneous Decoding of the Timing and Content of Human Object Perception from Cortical Surface Recordings Reveals Complementary

Information in the Event-Related Potential and Broadband Spectral Change”, published in *PLoS Computational Biology* in 2016 (Miller et al., 2016).

### 2.1.1. Experimental setting

A single data set (from subject *mv* in the data reported in (Miller et al., 2016)) was selected to demonstrate the methods proposed in this manuscript. The patient was clinically implanted with platinum subdural electrode grids and strips (Ad-Tech, Racine, WI) in frontal, parietal, temporal, and occipital cortical areas for extended clinical monitoring and localization of seizure foci. Lateral frontoparietal ECoG grid signals were not used in the analysis; we focused on signals from one strip located over inferior temporal cortex. The electrodes were 4 mm in diameter (2.3 mm exposed), with 1 cm inter-electrode spacing, and were embedded in silastic. Electrode locations relative to gyral surface anatomy were determined by projecting the post-implant CT to the pre-operative axial T1 magnetic resonance head image using the normalized mutual information approach in SPM, and the CTMR package of Hermes et al. (2010). Cortical surface mesh reconstruction was performed using Freesurfer (Dale et al., 1999).

During the experiment, the subject was presented with a sequence of images of either a face (‘face stimulus’) or house (‘house stimulus’) in random sequence. There were three runs of 100 presentations each (50 houses and 50 faces in random order). Each stimulus presentation lasted 400 ms and was followed by a blank inter-stimulus interval of 400 ms. The ECoG potentials were measured with respect to a scalp reference and ground, subjected to recording system-imposed band-pass filtering between 0.15 Hz and 200 Hz, and sampled at 1000 Hz per channel. The protocol is described in more detail in (Miller et al., 2016).

### 2.1.2. Data preprocessing

Data processing, PAC computation and analysis were performed using EEGLAB functions (Delorme and Makeig, 2004) and custom scripts written in MATLAB (The Mathworks, Inc.). First, segments in the data containing strong non-brain artifact (e.g., high-frequency muscle noise and other irregular artifacts) as identified by visual inspection were removed. A common-average reference computation involving all channels, was applied to the remaining data. The data were then resampled to 512 Hz and 60-Hz line noise was suppressed using a Hamming-windowed (sinc) FIR notch filter [order 847] (EEGLAB function *pop\_eegfiltnew.m*) to minimize activity between 58 Hz and 62 Hz and between 118 Hz and 122 Hz (i.e., at the fundamental and first harmonic of the 60 Hz line noise). MATLAB scripts and functions for computing and visualizing different PAC measures in the EEGLAB environment have been collected into a toolbox developed by the authors (the Event-Related Phase-Amplitude Coupling Toolbox, ERPACT) that is now available from the EEGLAB Extension manager<sup>1</sup> and from a GIT repository.<sup>2</sup>

### 2.1.3. Extracting phase and amplitude time series

As we mentioned earlier, most PAC estimation methods begin with computing phase and amplitude time series from the data. When analyzing PAC within a single signal, the original raw signal  $S_{raw_t}$  is first band-pass filtered to isolate the two frequency ranges in which the PAC coupling is to be assessed, respectively centered on a lower central frequency  $f_{phase}$  for the phase time series, and a higher central frequency  $f_{amp}$  for the amplitude time series. In the case of analyzing PAC between two signals the process will be similar to the one described, but extracting one frequency range from each of the signals under analysis.

Due to the uncertainty as to the ideal bandwidth for applying the Hilbert transform, the criteria for selecting the bandwidth for band-pass filtering diverges among authors. Usually, the bandwidth used in constructing the

phase time series is required to be sufficiently narrow to derive a meaningful phase measure (Aru et al., 2015). This requirement is often met. Meanwhile, when selecting the bandwidth used to obtain the amplitude time series, side-bands resulting from the amplitude-modulation (AM) process should be included within the measured bandwidth. Disregarding this criteria may result in inaccurate PAC estimates (Aru et al., 2015).

In the current work, the bandwidth of the band-pass filter used to estimate the phase time series was  $f_{PhaseBand} = [f_{phase} - 1\text{Hz}, f_{phase} + 1\text{Hz}]$ , while for the amplitude time series the bandwidth was  $f_{AmpBand} = [f_{amp} - (f_{phase} + 1\text{Hz}), f_{amp} + (f_{phase} + 1\text{Hz})]$  to include the AM side-bands. Below, we denote by  $S_\varphi$  and  $S_A$  the signals obtained by band-pass filtering the signal(s) of interest in the ranges  $f_{PhaseBand}$  and  $f_{AmpBand}$  respectively.

To filter the data within these frequency bands, a combination of high-pass and low-pass FIR filters were coded using EEGLAB function *pop\_eegfiltnew.m*. Analytic signals were then obtained by applying the Hilbert transform to the  $S_\varphi$  and  $S_A$  signals. Next, instantaneous phase  $\varphi_t$  and amplitude  $A_t$  time series were estimated from the analytic signals of  $S_\varphi$  and  $S_A$  respectively.

## 2.2. Traditional PAC measures

In this section we briefly review some of the methods most commonly used to estimate PAC. Later in the manuscript, we will compare results of these methods to results of the proposed MIPAC measure.

### 2.2.1. Mean Vector Length Modulation Index (MVLmi)

The method proposed by Canolty et al. (2006a) relies on generation of the composite values  $z_t$  in the complex plane, constructed by combining the instantaneous phase of the lower frequency signal ( $\varphi_t$ ) and instantaneous amplitude of the higher frequency signal ( $A_t$ ) as in Eq. (1).

$$z_t = A_t \cdot e^{i\varphi_t} \quad (1)$$

The distribution of the composite vectors  $z_t$  in the complex plane is then analyzed. If there is no PAC, since there is no affinity between a specific value of phase and amplitude, the distribution should be nearly radially symmetric. However, if the local mean magnitude  $A_t$  is higher at some specific phase (or phases), then PAC is present and the phasor representation of the phase-amplitude distribution will exhibit a ‘bump’ at these phase values. The length of the composite mean vector obtained in Eq. (2) is used as a measure of phase-amplitude coupling. Here,  $N_c$  denotes the number of composite values averaged in the interval  $t = [1, T]$ . Alternatively, some authors (Penny et al., 2008a) use in their PAC analysis the z-score of the MVLmi in Eq. (2) obtained from a surrogate data analysis (see Section 2.7)

$$MVLmi = \left| \frac{1}{N_c} \sum_{t=1}^T z_t \right| \quad (2)$$

### 2.2.2. Kullback-Leibler Modulation Index (KLmi)

Tort et al. (2010) pointed out that the MVLmi approach assumes that coupling only occurs at one preferred phase and thus may perform poorly for more complex coupling relationships. To remedy this, they proposed a different approach. First, the instantaneous phase time series  $\varphi_t$  is subdivided into an arbitrary fixed number of bins  $N_{bin}$  and the mean of the instantaneous amplitude  $A_t$  is computed for each  $j$ -bin  $j = \{1 \dots N_{bin}\}$ , here denoted by  $\langle A_t \rangle_\varphi(j)$ . These mean values are then divided by the sum of all the amplitudes over the bins, giving a fractional value  $P(j)$  for each bin  $j$  (Eq. (3)).

$$P(j) = \frac{\langle A_t \rangle_\varphi(j)}{\sum_{k=1}^{N_{bin}} \langle A_t \rangle_\varphi(k)} \quad (3)$$

<sup>1</sup> [https://scn.ucsd.edu/wiki/Plugin\\_list\\_process](https://scn.ucsd.edu/wiki/Plugin_list_process).

<sup>2</sup> [https://bitbucket.org/ramonmc/pop\\_pac](https://bitbucket.org/ramonmc/pop_pac).

As  $P(j) \geq 0 \forall j$  and  $\sum_{j=1}^{N_{bin}} P(j) = 1$ ,  $P$  meets the criteria for a probability distribution function. If there is no coupling, this distribution should be uniform. If there is coupling, the deviation of this distribution from uniformity will be proportional to the strength of the coupling. This deviation is quantified by computing the Kullback-Leibler divergence ( $D_{KL}$ ) between  $P$  and a uniform distribution  $U$ . The divergence measure is used to define the KLmi (Eq. (4)), where the  $\log(N_{bin})$  is used as a normalization term.

$$KLmi = \frac{D_{KL}(P||U)}{\log N_{bin}} \quad (4)$$

$$\text{with } D_{KL}(P||U) = \sum_{k=1}^{N_{bin}} P(k) \log \frac{P(k)}{U(k)} \quad (5)$$

### 2.2.3. Modulation index based on general linear model (GLMmi)

The use of correlation to compute PAC was first proposed by [Bruns and Eckhorn \(2004\)](#). In this method, also referred as envelope-to-signal correlation (ESC), correlation between the time series (band-pass filtered in the range of the frequency of the phase of interest) and amplitude in the range of the higher frequency of interest was used as a measure of PAC. Later, [Penny et al. \(2008a\)](#) pointed out Bruns and Eckhorn ESC inability to capture the PAC phenomena in all instances (e.g., ESC can be confounded with amplitude co-modulation). They took, nonetheless, this measure as a starting point of their own approach. They first proposed to estimate PAC by means of the correlation between the amplitudes and the *cosines* of the phase signals, but pointed to the inability of this approach to detect coupling at the so called ‘null phases’ (e.g.  $\cos(1/4)2\pi = 0$ ). To overcome this limitation, they proposed an approach using the general linear model (GLM) framework as in Eqs. (6) and (7). Here, the PAC estimate is inferred from the portion of variance explained by the linear model.

$$A_t = X\beta + e \quad (6)$$

$$\text{with } X = \begin{pmatrix} \cos \varphi_{t_1} & \sin \varphi_{t_1} & 1 \\ \vdots & \vdots & \vdots \\ \cos \varphi_{t_{max}} & \sin \varphi_{t_{max}} & 1 \end{pmatrix} \quad (7)$$

In Eq. (6),  $\beta$  is the regression coefficient matrix and  $e$  is additive Gaussian noise. The design matrix  $X$  (Eq. (7)) is composed of three columns; the first two columns are the *sine* and *cosine* of the time series  $\varphi_t$  respectively, and the third is a constant vector of ones. This design effectively circumvents the limitations of the previous correlation-based method.

### 2.2.4. Event-Related Phase Amplitude Coupling (ERPAC)

The GLMmi method proposed by Penny et al. (see Section 2.2.3) was modified by [Voytek et al. \(2013\)](#) to allow estimation of PAC dynamics across a set of brief event-locked data epochs in event-related experiments.

To apply this method, the time series  $\varphi_t$  and  $A_t$  are broken into time windows of equal length, time-locked to the onset of equivalent experimental events. Time series for both magnitudes form a matrix with dimensions equal to the number of trials by the number of latencies. The method then follows as to the GLMmi method 2.2.3 by computing the PAC modulation index for each successive latency across trials. This way one can obtain a latency-resolved description of the evolution of PAC following a set of similar events.

## 2.3. Information-theoretic measures

Next, we will present an overview of the information-theoretic measures relevant to the approach proposed in this work. Let us start by

considering two discrete random variables  $X$  and  $Y$  with sets of values  $x$  and  $y$  respectively, probability distributions  $p(x)$ ,  $p(y)$ , conditional probabilities  $p(x|y)$ , and joint distribution  $p(x,y)$ . A central quantity in Information Theory is the Shannon Entropy ( $H$ ) (Eq. (8)). The quantity  $H$  represents the average amount of uncertainty associated with a measurement  $x$  of the random variable  $X$ :

$$H(X) = -\sum_x p(x) \log_2 p(x) \quad (8)$$

Shannon Entropy can also be defined as the average Shannon information content  $h(x)$  (Eq. (9)) over each outcome  $x$  of the random variable  $X$  used to build the probability distribution function (pdf)  $p(x)$  (Eq. (10)):

$$h(x) = -\log_2 p(x) \quad (9)$$

$$H(X) = E_x[h(x)] \quad (10)$$

By convention, base 2 or  $e$  logarithms are used, giving units of bits or nats respectively.

By generalizing Shannon's definition of entropy to two random variables  $X$  and  $Y$ , the joint entropy can be defined as in Eq. (11), and represents the expected uncertainty of their joint distribution ([Lizier, 2014a](#)). Alternatively, Eq. (11) can be written as a function of the Shannon information content as in Eq. (12).

$$H(X, Y) = -\sum_{x,y} p(x, y) \log_2 p(x, y) \quad (11)$$

$$H(X, Y) = E_{x,y}[h(x, y)] \quad (12)$$

The conditional entropy of  $X$  given  $Y$  is defined as the average uncertainty about  $x$  that remains when the value of  $y$  is known:

$$H(X|Y) = -\sum_{x,y} p(x, y) \log_2 p(x|y) \quad (13)$$

Using the definition of entropy in Eq. (8), we can formalize the mutual information (MI) between the random variables  $X$  and  $Y$  as a non-negative and symmetric measure defined in Eqs. (14) and (15):

$$I(X, Y) = \sum_{x,y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \geq 0 \quad (14)$$

$$= H(X) - H(X|Y) \quad (15)$$

This quantity can be interpreted either as the average reduction in uncertainty about  $X$  given knowledge of the value of  $Y$ , or as its name suggests, as the amount of information shared by the two variables. Note that Eq. (14) is equal to zero if and only if  $X$  and  $Y$  are independent, which means  $p(x,y) = p(x)p(y)$ . Therefore, the mutual information is a measure of dependency (both linear and nonlinear) between the two random variables  $X$  and  $Y$  ([Kraskov et al., 2004](#)).

## 2.4. Defining local mutual information

In Eq. (12) we saw that the joint entropy of the random variables  $X$  and  $Y$  can be seen as the average of the joint Shannon information content  $h(x,y)$  over its realizations  $x$  and  $y$  respectively. Given this interpretation, several authors have used the terms pointwise (particularly in natural language processing) or *local* information theoretic measures to describe quantities such as those defined in Eq. (9) or (16) ([Lizier, 2014a](#); [MacKay, 2003](#); [Fano, 1963](#); [Church and Hanks, 1990](#); [Hindle, 1990](#)). These *local* information theoretic measures characterize quantities from a specific subset of measurements  $x$  and  $y$  of the random variables  $X$  and  $Y$ , rather than the associated average measure computed over all available data ([Lizier, 2014a](#)). An interesting discussion on how frequently used information-theoretic measures can be viewed as averages over their associated local measures can be found in ([Lizier, 2014a](#)). As an example

of this, local MI values  $i(x,y)$  (Eq. (16)) may be averaged to compute overall MI  $I(X,Y)$  (Eq. (17)):

$$i(x,y) = \log_2 \frac{p(x|y)}{p(x)} \tag{16}$$

$$I(X,Y) = E_{x,y}[i(x,y)] \tag{17}$$

Negative local MI values might result if  $p(x|y) < p(x)$  in Eq. (16). Thus, in contrast to the averaged quantity, which cannot take negative values, local MI values can be either negative or positive. These negative values may be interpreted as misinformative since the expectation of observing  $x$  is lowered by knowing  $y$  (Lizier, 2014a).

In Eq. (16), the local value of  $i(x,y)$  is obtained by evaluating the pdf  $p(x)$  and  $p(x|y)$  at a set of local observations  $(x,y)$ . Meanwhile, the pdf here are defined over all the observations, both local and non-local to the observation under analysis. The term *local*, though, can be misleading when dealing with time series since it does not imply locality in the time-sense. Notice that until now, while introducing the basic information-theoretic measures and local MI, we have been dealing only with discrete random variables. This is, without assuming any type of temporal structure within the samples in question. This makes the *local* sense to be solely based upon the sample properties, e.g., distance between sample magnitudes. If a time dependency is assumed, as in the case of a time series, we will not be dealing anymore with a statistical sample, but the definition of *local* will remain as in the case of discrete random variables. Thus, the term *local* is not in relationship to the time structure of the process. Because of this, in the case of real time series, points locally close to the sample in the pdf are not necessarily local to the observation in the time sense. Thus, the stationarity of the time series used to build the pdf should be considered carefully. In fact, the validity of the application to time series of the information theory measures introduced until now is under the assumption of stationarity or quasistationarity over time. We will return to this point in the next sections.

### 2.5. Estimating local mutual information

There is a long history of application of information theory methods in neurosciences, specially MI (for a review see (Lizier, 2014a; Ince et al., 2017)). Although several methods for estimating MI have been proposed, the so called binning method remains the most commonly used approach. In this method, the supports of  $X$  and  $Y$  are partitioned into an arbitrary, limited number of bins, for which the joint probabilities are estimated by counting the number of samples in each bin. However, this method inevitably leads to serious bias problems (Wibral et al., 2015; Treves and Panzeri, 1995; Victor, 2002; Panzeri et al., 2007). This drawback can be alleviated to some extent by recent advances in development of MI nearest-neighbour estimators such as the one proposed by Kraskov, Stogbauer, and Grassberger (KSG) (Kraskov et al., 2004), together with a rigorous statistical test using appropriate surrogate data. Although computationally expensive for large amounts of data, the KSG estimator constitutes an effective non-parametric estimator of MI that is data efficient (resolve structures down to the smallest possible scales), resolution adaptive (binning scale changes according to the underlying data point density), and has minimal bias (Kraskov et al., 2004) (in consequence, it may have large variance).

Before presenting the local mutual information estimator, let us first introduce the KSG estimator and then move on to the local estimator. The KSG estimator was obtained by modifying and extending the nearest-neighbors-based Kozacheko-Leonenko estimator of entropy (Kozachenko, 1987). Instead of estimating directly the pdf needed to compute the MI, the KSG estimator makes use of the distances of each sample to its  $k^{th}$  nearest neighbor to count the neighbors in the marginal spaces, then plugs in these values in the MI estimator. To elaborate on this, let's assume a metric given in the space spanned by the discrete random variables  $X, Y$  and the joint space  $Z = (X, Y)$ . Then, the neighbors of each

point  $z_i = (x_i, y_i)$  are ranked by the distance  $d_{ij} = \|z_i - z_j\| : d_{i,j_1} \leq d_{i,j_2} \leq \dots$ . Similar rankings are done in the subspaces  $X$  and  $Y$ . While any norm can be used for these subspaces, the maximum norm is used for the joint space (Eq. (18)) (Kraskov et al., 2004). At this point, marginal variables, or alternatively the norms, can be scaled to make them comparable. Additionally, although not strictly required, if the distributions are very skewed and/or rough, it might be a good idea to transform them such as to become more uniform in order to reduce errors in general (Kraskov et al., 2004).

$$\|z - z'\|_z = \max\{\|x - x'\|, \|y - y'\|\} \tag{18}$$

Next, the  $k^{th}$  nearest neighbor for each point  $z_i$ , a distance  $\varepsilon(i)/2$  away, is computed. Let's denote by  $\varepsilon_x(i)/2$  and  $\varepsilon_y(i)/2$  the distances between the same points projected into the subspaces  $X$  and  $Y$  respectively. As the maximum norm is used in Eq. (18), it becomes obvious that  $\varepsilon(i) = \max\{\varepsilon_x, \varepsilon_y\}$ . In the KSG estimator used in our work, the number of points  $n_x(i)$  and  $n_y(i)$  are estimated by counting the points within the strips defined by  $(x_i - \varepsilon(i)/2, x_i + \varepsilon(i)/2)$  and  $(y_i - \varepsilon(i)/2, y_i + \varepsilon(i)/2)$  respectively. The process of computing the number of neighbors  $n_x(i)$  and  $n_y(i)$  of point  $z_i$  in the neighborhood defined by its  $k^{th}$  neighbor is depicted in Fig. 1.

After computing the number of neighbors  $n_x$  and  $n_y$  for each  $z_i$ , the estimate for MI is given by the first KSG estimator (Eq. (19)).

$$I(X,Y) = \psi(k) - \langle \psi(n_x + 1) + \psi(n_y + 1) \rangle + \psi(N) \tag{19}$$

In Eq. (19),  $N$  is the number of samples and  $\psi$  denotes the digamma function ( $\psi(x) = \Gamma(x)^{-1} \frac{d\Gamma(x)}{dx}$ ). MI values  $I(X, Y)$  are returned in nats. By using this approach, pdf evaluations are bypassed and the average measure  $I(X, Y)$  can be estimated for each observation from the nearest neighbors (counts  $n_x$  and  $n_y$ ) in the marginal spaces. Taking the KSG estimator (Eq. (19)) as a starting point, Lizier (2014a) proposed to obtain the local quantities by unrolling the expectation values ( $\langle \dots \rangle$ ) and computing the nearest neighbor count only at the given  $(x,y)$  (Eq. (20)):

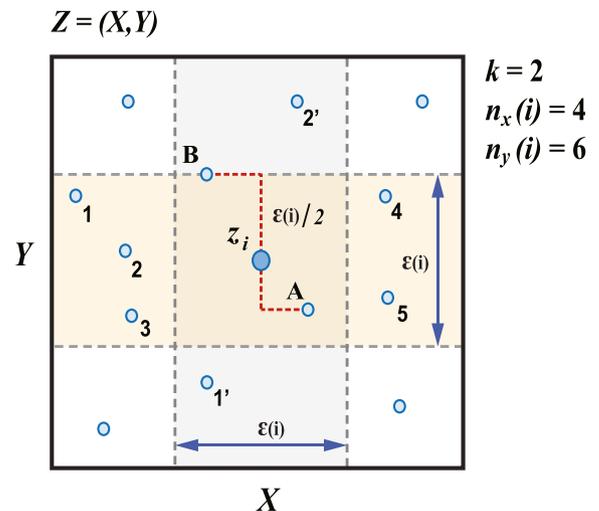


Fig. 1. Schematic of  $Z$ -space showing the  $k = 2$  nearest-neighbors search for data point  $z_i$ . Subspaces  $X$  and  $Y$  are represented in the horizontal and vertical axes respectively. The two nearest samples to  $z_i$  are **A** and **B**. Red dotted lines show the distance in the marginal spaces from the samples **A** and **B** to the sample  $z_i$ . Sample **B** is the  $k = 2$  neighbor and its maximum marginal distance to the sample  $z_i$  ( $\varepsilon(i)/2$ ) defines the width of the marginal neighborhoods for  $X$  and  $Y$ , represented by the vertical and horizontal blue and orange stripes. Four sample points (**1'**, **2'**, **A**, **B**) are within distance  $\varepsilon/2$  in the marginal space of  $X$  ( $n_x(i) = 4$ ); six sample points (**1**, **2**, **3**, **4**, **5**, **A**) are within distance  $\varepsilon/2$  in the marginal space of  $Y$  ( $n_y(i) = 6$ ). Several other samples (unlabeled) are outside the  $k = 2$  neighborhood.

$$i(x, y) = \psi(k) - \psi(n_x + 1) - \psi(n_y + 1) + \psi(N) \quad (20)$$

This estimator of local MI has been previously treated as a ‘time-resolved estimator’ in (Gómez-Herrero et al., 2015). If implemented in its naive version using nested loops through all  $N$  points, the KSG algorithm spends most of the CPU time searching for neighbors. This leads to a complexity  $O(N^2)$  (Kruskova et al., 2004). Although this may be acceptable for small data sets (say  $N \leq 3500$ ), strategies including fast nearest neighbor search algorithms are needed to deal with higher dimensionality. Here we used the naive version.

## 2.6. Estimating PAC using local mutual information

Here, we apply the local MI estimation approach presented in Section 2.5 to calculate PAC in brain electrophysiological signals. As mentioned before, to calculate local MI, as well as other information-theoretical measures, an assumption is made on the stationarity of the signal. This assumption is usually met in data collected in the field of behavioral neuroscience, in which many electrophysiological experiments are dominated by the repetition of equivalent (or considered equivalent) events. This type of data is non-stationary in its nature, but given the periodicity of the stimulus presentation (and the brain response), can be viewed as a cyclo-stationary process. However, in more recently emerging experimental paradigms, notably recordings acquired during ‘resting state’ and in continuous brain computer interface (BCI) applications, no explicit cycling through event presentations and participant responses is featured. This type of data, again, is non-stationary, but may be modeled as a stationary dynamic process within a time window of a small enough size so the stationarity criteria is met (locally stationary).

We thus here propose and report initial results for two variants of an approach to estimating mutual information-based phase-amplitude coupling (MIPAC) in locally stationary and cyclo-stationary time series respectively. We will use the terms *Instantaneous MIPAC* for continuous data estimation and *Event-related MIPAC* for latency-tagged PAC estimation in multiple-trial event-related data.

The first steps for both procedures are to obtain appropriate phase  $\varphi_t$  and amplitude  $A_t$  time series from the data as in 2.1.3. At this point in the manuscript, we make a switch from the random variables  $X$  and  $Y$  used in the presentation of the local MI to the time series of the phase  $\varphi_t$  and amplitude  $A_t$  previously introduced in Section 2.1.3. Next, the local MI is estimated then as in Eq. (20) in a manner appropriate to the data structure.

### 2.6.1. Instantaneous MIPAC

To estimate *Instantaneous MIPAC*, we first compute the local MI between  $\varphi_t$  and  $A_t$ . For this, let us assume the space  $Z = (\varphi_t, A_t)$ . Then, for each point  $z_i = (\varphi_i, A_i)$  in the  $Z$ -space, the number of neighbors  $n_\varphi(i)$  and  $n_A(i)$  are computed for a given value of  $k$  as done in Section 2.5. When estimating MIPAC, Eq. (18) can be rewritten as Eq. (21).

$$\|z - z'\|_z = \max\{\|\varphi - \varphi'\|_c, \|A - A'\|_2\} \quad (21)$$

Here, pairwise distances  $\|\varphi - \varphi'\|_c$  and  $\|A - A'\|_2$  are scaled by dividing each distance element by the maximum value of the pairwise distance matrix (alternatively, other type of scalings might be used with similar results, e.g., dividing by the spectral norm). Attending to the cyclic nature of  $\varphi_t$ , the circular ( $\|\cdot\|_c$ ) norm (Berens et al., 2009) is used to compute pairwise distance between points in the phase space; the Euclidean ( $\|\cdot\|_2$ ) norm is used for the amplitude space.

Then, to estimate the local MI between  $\varphi_t$  and  $A_t$ , the values of  $n_\varphi(i)$  and  $n_A(i)$  in the  $k$ -neighborhood are plugged into Eq. (20) as  $n_x(i)$  and  $n_y(i)$  respectively. The resultant local MI will be a vector of the same dimensions of  $\varphi_t$  and  $A_t$ .

To avoid the need of fixing a value for the parameter  $k$ , and as a convergence criteria for the estimation of the local MI, the local MI between  $\varphi_t$  and  $A_t$  is estimated in an iterative scheme for different values of

$k$ , starting at  $k_0 = 1$  increasing by  $\Delta k = 1$ , until the percent decrease in local MI variance from step  $k_i$  to  $k_{i+1}$  ( $\Delta Var_i$ ) reaches the threshold  $\Delta Var_{thres}$ , here defined as  $\Delta Var_{thres} = 0.05\%$ . The local MI for which the threshold  $\Delta Var_{thres}$  is reached is then selected for further processing.

When computing Eq. (21), a leakage from phase frequency  $f_{phase}$  to  $\|z - z'\|_z$  is observed. The characteristics of the leakage were determined empirically to lead to estimations of local MI with a high content of  $f_{phase}$  and its harmonics. As a heuristic to mitigate this leakage to obtain the *Instantaneous MIPAC* estimate, the estimated local MI with the optimal  $k$ -value is low-pass filtered to below  $f_{phase}$ . Given the empirical nature of the joint space  $Z$ , a theoretical explanation of the high frequency leakage into the local MI may prove hard to obtain. However, the proposed heuristics should work independently of the nature of the leakage observed and is supported by the fact that we do not intend to interpret modulating frequency components above  $f_{phase}$ . The estimated *Instantaneous MIPAC* is represented by a vector with the same dimensions of  $\varphi_t$  and  $A_t$ .

### 2.6.2. Event-related MIPAC

In the *Event-related MIPAC* approach, similar to Voytek et al. (2013) the low-frequency phase  $\varphi_t$  and high-frequency amplitude  $A_t$  time series are separated into non-overlapping epochs time-locked to a set of events of interest that are treated as being equivalent. Phase and amplitude epochs are then independently stacked to form two matrices (phase  $\varphi_{tr,t}$  and amplitude  $A_{tr,t}$ ) of dimensions number of trials by number of latencies. An across-trials local MI estimate between  $\varphi_t$  and  $A_t$  is then computed for each trial latency using the iterative scheme for estimating  $k$  presented in Section 2.6.1 on *Instantaneous MIPAC*. Notice here that estimated local MI at each latency will be a uni-dimensional vector of length the number of trials. To sample the full phase cycle on each latency neighborhood, the values of  $n_\varphi(i)$  and  $n_A(i)$  are estimated using a one-cycle  $f_{phase}$  frequency window centered on the latency of interest. This will also boost the nearest-neighbors count and account for possible jitter in the brain response to the stimulus. Estimation of local MI is then performed for each trial latency. Finally, to obtain the *Event-related MIPAC*, each epoch of the resulting local MI matrix (of length the number of latencies) is low-pass filtered to below  $f_{phase}$  to mitigate high frequency noise in the final MIPAC estimate.

Each resulting MIPAC epoch time series is interpreted as a time-varying measure of phase-amplitude coupling in the brain signal being analyzed. By taking the average of the *Event-related MIPAC* estimates across trials, we can obtain the *Trial-mean MIPAC*. Further collapsing the MIPAC time series by taking its grand mean gives a summary Mutual Information Modulation Index (MIMI) value. This MIPAC-derived measure can be regarded as equivalent to other PAC estimates computed over trials and latencies (see Section 2.2).

## 2.7. Surrogate analysis to compute MIPAC significance

For all the methods described in Section 2.2 including both MIPAC variants, a statistical significance value can be computed using a surrogate data approach (Penny et al., 2008b; Canolty et al., 2006b).

Here, for each PAC estimate,  $N_{surr} = 500$  surrogate measures were generated (unless otherwise specified). In the case of *Instantaneous MIPAC*, to obtain these surrogates, each  $\varphi_t$  and  $A_t$  time series were separated into  $N_{segm} = 20$  (unless otherwise specified) segments. Segments in  $\varphi_t$  and  $A_t$  were then uniformly randomly shuffled across latencies and a new PAC measure estimate was derived. This process was repeated  $N_{surr}$  times to obtain a distribution of the surrogate MIPAC estimates. This shuffling process aims to disrupt any meaningful relationship between  $\varphi_t$  and  $A_t$ , while preserving some of the temporal structure of the original signal, to allow testing the null hypothesis that no meaningful coupling exists in the data. This procedure has been assumed to be a more rigorous test than random latency-shuffling (He et al., 2010; Hurtado et al., 2004).

For *Event-related MIPAC*, surrogate signals for  $\varphi_{tr,t}$  and  $A_{tr,t}$  are constructed by randomly drawing as many time points as trials from a

baseline of the respective matrices  $\varphi_{tr,t}$  and  $A_{tr,t}$ . In this case, the baseline was assumed to be a latency window across all trials spanning from the beginning of the trials until the latency of the stimulus presentation.

For each surrogate data set we then compute the mean  $\mu_{surr}$  and standard deviation  $\sigma_{surr}$ , plus MIPAC z-scores defined as:

$$Z_{surr} = \frac{PAC_{measure} - \mu_{surr}}{\sigma_{surr}} \quad (22)$$

Here,  $Z_{surr}$  values satisfying  $|Z_{surr}| > 1.96$  were considered statistically significant for  $\alpha = 0.05$  (assuming the z-scores to be gaussian distributed).

### 3. Results

Next, we present the results obtained to validate the *Instantaneous MIPAC* and *Event-related MIPAC* approaches introduced in Section 2.6, first using simulated and then using actual brain data.

#### 3.1. Simulation results

As an initial validation of the MIPAC approach, we carried out five simulations. In the first two (Section 3.1.1 and 3.1.2) the performance of the *Instantaneous MIPAC* estimator is tested on simulated phase-amplitude coupled signals with and without added noise. In Section 3.1.3, the convergence of MIPAC in the previous simulations, as well as the interaction of parameters of simulated signals and MIPAC estimation are examined. In Section 3.1.4, we evaluate the performance of *Event-related MIPAC* and compare its estimates to those produced by the ERPAC measure (see Section 2.2.4). Finally, we compare the performance of Mimi (see Section 2.6.2) on a simulated PAC signal to the results of the PAC estimation methods reviewed in Section 2.2.

##### 3.1.1. Estimating MIPAC on simulated single-trial PAC data

Here we study the estimation of MIPAC in a simulated continuous noiseless signal using *Instantaneous MIPAC*. To simulate the PAC signal, a synthetic Amplitude-Modulated (AM) signal was generated using a time-varying modulation strength. The modulation strength is then estimated using *Instantaneous MIPAC* and later is compared to the ground truth.

To generate the synthetic PAC signal, a procedure similar to that used for modeling AM signals was used (see Appendix A). Here we adopt the notation used in the AM literature. The signal and frequency of the modulator ( $f_m$ ) in these simulations will correspond to the time series  $\varphi_t$  and the central frequency  $f_{phase}$  of the phase frequency band during MIPAC estimation respectively. In the same way, the signal and frequency of the carrier ( $f_c$ ) will correspond to the time series  $A_t$  and frequency  $f_{amp}$  of the amplitude frequency band used in the estimation.

In all simulations, two sinusoids featuring frequencies  $f_c = 40$  Hz and  $f_m = 5$  Hz were generated to act as carrier and information (modulator) signals respectively. The duration of the simulated signals was 5 s at the sampling rate  $S_{rate} = 500$  Hz. The signals were divided into five equal segments; coupling was introduced in Segments 2 and 4 only, leaving Segments 1, 3 and 5 with no coupling. In one simulation coupling was introduced in all segments. The threshold for variance reduction (see Section 2.6.1) was set to  $\Delta Var_{thres} = 0.05\%$ . Finally the signal was zero-padded by appending 1 s of 0's to the beginning and to the end of each time series. To compute MIPAC while accounting for filter edge effects, this padding was removed after band pass filtering the signal. Unless otherwise specified, these parameters were used in all the simulations.

To assess the ability of *Instantaneous MIPAC* to retrieve the time course of coupling, three simulated variations in coupling strength were constructed without adding noise. On/off boxcar and linearly increasing time courses were used in the first two cases to simulate coupling in signal Segments 2 and 4 only. In the third case, coupling was applied as the absolute value of a sinusoid across the five Segments.

Fig. 2 shows the estimated time course of MIPAC in a simulated PAC

signal in which modulation occurs only from 1 to 2 s and from 3 to 4 s; no coupling is induced outside of these data segments (Fig. 2A). To estimate MIPAC, the simulated signal was band-pass filtered using a combination of high-pass and then low-pass filtering to isolate the two narrow frequency bands used in the simulation ( $f_{amp} = f_c = 40$  Hz and  $f_{phase} = f_m = 5$  Hz) as described in Section 2.1.3. The frequency limits of the band-pass filters were  $f_{PhaseBand} = [4 \ 6]$  Hz and  $f_{AmpBand} = [34 \ 46]$  Hz. These values were kept constant across simulations (any changes noted below).

Fig. 2B depicts the relationship between the (red) time series  $\varphi_t$  and (blue)  $A_t$  in this simulation. The time series  $A_t$  is the envelope of the (grey) narrow-band AM or carrier signal within  $f_{AmpBand}$ . The frequency of the carrier signal remains constant both during and in the absence of coupling, but its amplitude varies as a function of the phase of the modulator signal  $\varphi_t$  in the segments where the modulation is in effect.

Fig. 2C shows the MIPAC estimate (red) using the *Instantaneous MIPAC* approach along with the local MI (pink) time series obtained before being low-pass filtered to obtain the MIPAC. The overall MI computed as the average of the local MI is shown as well here. Note that MIPAC successfully captures the simulated modulation dynamics shown in Fig. 2A.

When MIPAC methods were introduced in Section 2.6, we discussed the high-frequency leakage in the local MI while estimating the  $k$ -neighbors in Eq. (21). Fig. 2C shows the effect of this leakage through the presence of a 10-Hz oscillation in the local MI. This frequency is the first harmonic of the  $f_{phase}$  frequency. The figure also shows how the heuristic remedy we implemented – low pass filtering the local MI at  $f_{phase}$  – improves the MIPAC estimate considerably. To this end, a zero phase-lag, infinite impulse response (IIR) Butterworth low pass filter [order 6] with a normalized cutoff frequency of  $2f_{phase}/S_{rate} = 0.02$  was designed and implemented in MATLAB. The type of filter and filter order was chosen to avoid noticeable edge effects in view of the signal in question.

In performing the analyses to obtain the results shown in Figs. 3 and 4 we followed the same procedures as in Fig. 2 but used different time courses to vary the coupling: linearly increasing and absolute value of a sinusoidally varying coupling strength respectively. As can be seen in both cases, the MIPAC estimate successfully recovers the simulated variations in coupling strength.

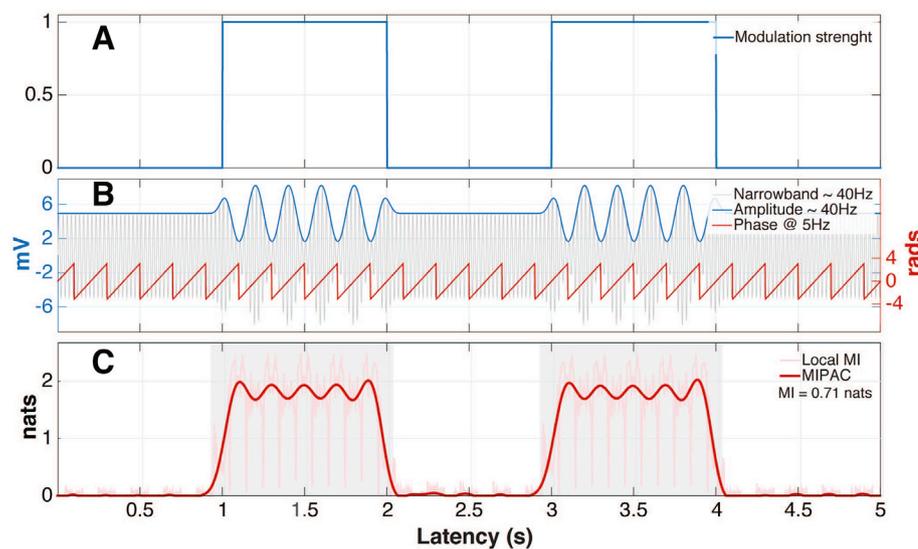
##### 3.1.2. Estimating MIPAC in simulated single-trial PAC data with added noise

In the previous section (Section 3.1.1), a set of three noiseless simulated PAC signals were used to test *Instantaneous MIPAC*. A similar MIPAC analysis was performed after adding white Gaussian noise to the same three simulated PAC signals using the MATLAB function *awgn.m*. For each signal, the signal-to-noise ratio (SNR) was set to 10. In Fig. 5, the local MI and MIPAC estimated from the noisy signal using the *Instantaneous MIPAC* approach are shown for simulated coupling time courses featuring a (on/off) boxcar, a linear ramp and absolute value of a sinusoid (Fig. 5A–C). The overall variations in PAC shown in Figs. 2A, 3A and 4A are again captured by MIPAC dynamics.

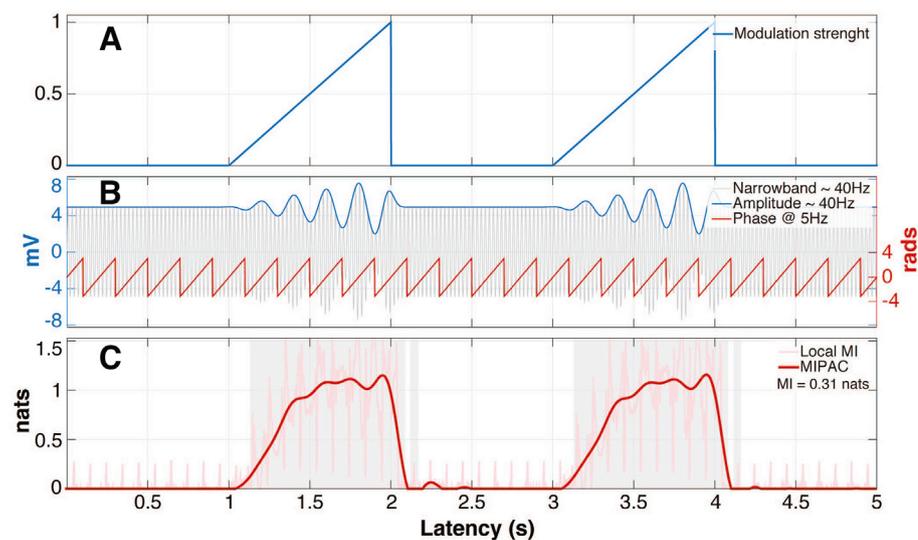
##### 3.1.3. Convergence in MIPAC estimation

The first row of Fig. 6 shows the percent decrease in the estimated local MI variance  $\Delta Var$ , (blue) and MIPAC variance (red) as a function of the number of  $k$ -neighbors in the simulations without noise added (Figs. 2–4). The maximum value of  $k$  on the abscissa represents the value at which the convergence criteria was reached for  $\Delta Var_{thres} = 0.05\%$  (see Section 2.6.1). In this simulation, MIPAC variance tended to decrease as the value of  $k$  increased. The same trend, although less evident, is shown by  $\Delta Var$ .

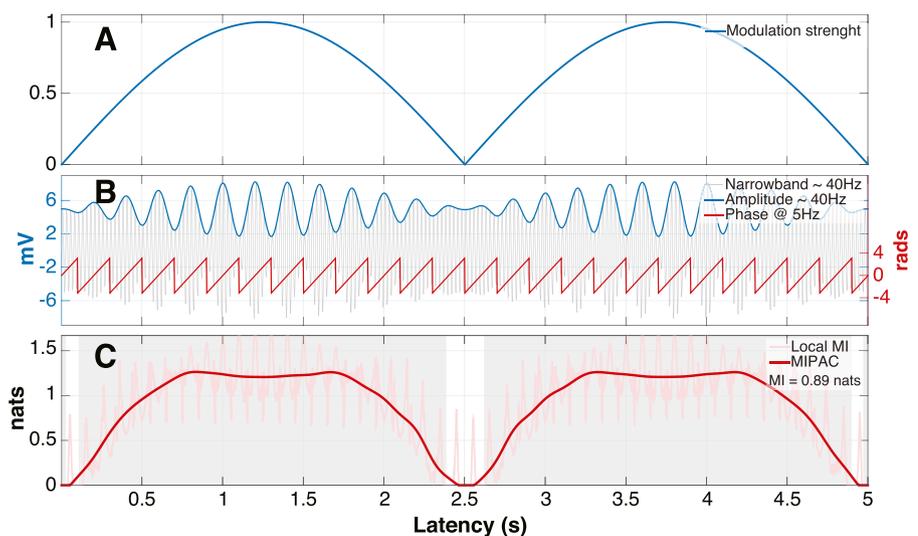
This result is intuitively correct. As  $k$  increases while keeping constant the number  $N$  of points in the signal, more neighboring time points are included in the estimation of the local MI for each specific point. This will lead to uniformity in the number of neighbors for all points, thus pushing the distribution of local MI values toward uniformity and therefore decreasing the variance of the estimated MIPAC signal. Notice



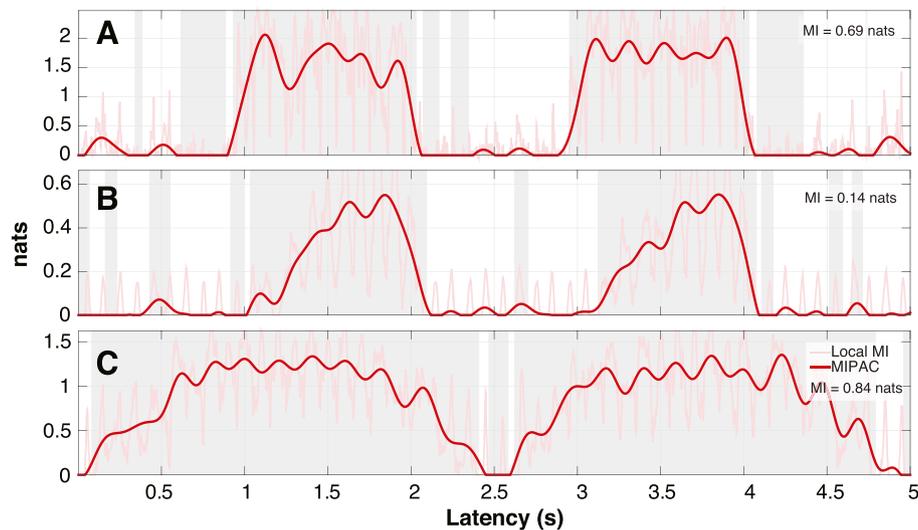
**Fig. 2.** MIPAC from a simulated PAC signal with on/off boxcar coupling Here the sampling rate of the simulated signal was  $S_{rate} = 500$  Hz, the frequency of the carrier was  $f_c = 40$  Hz and the frequency of the modulator was  $f_m = 5$  Hz. (A): Coupling strength alternated between ‘on’ and ‘off’ across signal segments. (B): In the simulated signal: instantaneous low-frequency phase at 5 Hz (red), narrow-band high-frequency signal at 40 Hz (grey), and instantaneous high-frequency amplitude at 40 Hz (blue). (C): MIPAC estimated from the simulated signal for  $f_{amp} = 40$  Hz,  $f_{phase} = 5$  Hz. Time courses of estimated MIPAC (red) and local MI (only positive values are shown) before low-pass filtering to obtain the MIPAC (pink) are shown; both have the same units (nats). Overall MI between instantaneous phase and amplitude computed as the average local MI is shown in the legend. Latencies with statistically significant coupling estimates ( $p < 0.05$ , uncorrected) appear in light grey.



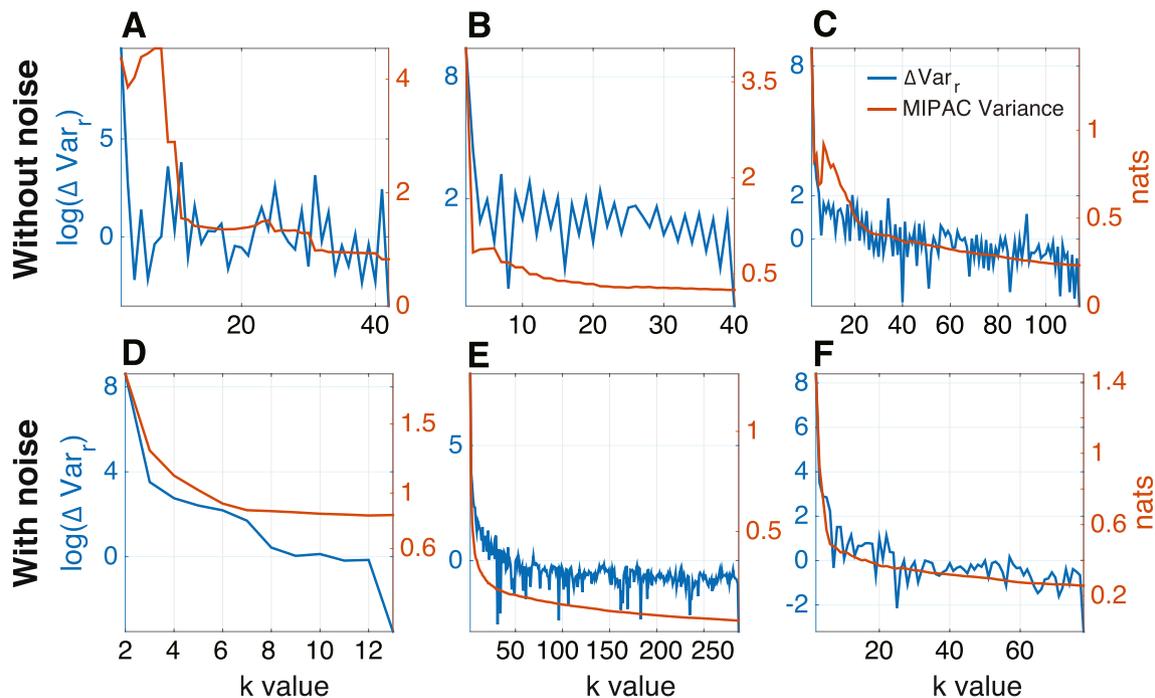
**Fig. 3.** MIPAC estimates of simulated PAC signals with linearly increasing coupling. (A): Linearly increasing ramps of coupling strength in a simulated PAC signal. (B): In the simulated signal, instantaneous phase (red) at 5 Hz, (grey) narrow-band signal at 40 Hz and (blue) instantaneous 40-Hz amplitude. (C): MIPAC estimate for the simulated signal  $f_c = 40$  Hz,  $f_m = 5$  Hz. The (red) MIPAC estimate and (pink) local MI (only positive values are shown) before being low-pass filtered. Overall MI between instantaneous phase and amplitude computed as the average local MI is shown in the legend. Latencies with statistically significant coupling estimates ( $p < 0.05$ , uncorrected) appear in light grey.



**Fig. 4.** MIPAC time course estimate for a simulated PAC signal with absolute value of a sinusoid used as coupling. (A): Absolute value of a sinusoid used as coupling strength in a simulated PAC signal. (B): In the simulated signal, instantaneous phase (red) at 5 Hz, (grey) narrow-band signal at 40 Hz and (blue) instantaneous amplitude at 40 Hz. (C): Time course of MIPAC estimated from the simulated signal with  $f_c = 40$  Hz,  $f_m = 5$  Hz. The (red) MIPAC estimate and (pink) local MI (only positive values are shown) before being low-pass filtered. Overall MI between instantaneous phase and amplitude computed as the average local MI is shown in the legend. Latencies with statistically significant coupling estimates ( $p < 0.05$ , uncorrected) appear in light grey.



**Fig. 5.** MIPAC estimation using Instantaneous MIPAC on simulated PAC signals with noise added. In this simulation, the time course of MIPAC was estimated for the same signals simulated in Figs. 2–4 but with noise added ( $SNR = 10$ ). Estimated (red) MIPAC, and (pink) local MI before low-pass filtering are shown. Latencies with statistically significant coupling estimates ( $p < 0.05$ , uncorrected) appear in light grey.



**Fig. 6.** Convergence in MIPAC estimation. Percentage variance reduction ( $\Delta Var_r$ ) (blue) and MIPAC variance (red) as a function of the number of  $k$ -neighbors. The decreasing of values of  $\Delta Var_r$  below the threshold defined by  $\Delta Var_r^{thresh}$  defines the convergence criteria for the MIPAC estimate. The three columns in the figure show the percentage variance reduction ( $\Delta Var_r$ ) (blue) and MIPAC variance for each of the three types of simulations shown earlier without noise (A–C) and with noise added (D–F).

that in the limiting case  $k = N - 1$  the number of neighbors for each point to be estimated become equal, which lead to  $\Delta Var_r = 0$  and MIPAC variance = 0.

The same convergence analysis performed on three noiseless signals was replicated for the noisy simulated signals ( $SNR = 10$ ) (lower row of Fig. 6). The pattern of behavior previously found in the noiseless simulations was preserved, though in this case as  $k$  increased the decreases in  $\Delta Var_r$  and MIPAC variance became smoother. This effect can be explained by the fact that by introducing noise, the distribution of points  $z_i$  in the joint space generated by the supports of  $A_i$  and  $\varphi_i$  become more uniformly distributed. This blurs the signatures of interdependence

between the signals in the  $Z$ - space and leads to a smoother MIPAC estimate. Notice that in the limiting case when  $SNR \rightarrow 0$ , the plots in second row of Fig. 6 will display an approximate similar  $\frac{1}{k}$  decay.

Another simulation was performed to study interactions between the sampling rate,  $S_{rate}$ , the  $SNR$  of the signal, and the parameter  $k$  used in MIPAC estimation. Simulated PAC signals of 5 s duration were generated using all combinations of the parameters  $SNR = [0.2, 0.5, 1.4, 3.8, 10]$  dB and  $S_{rate} = [128, 256, 512, 1024, 2048]$  Hz. Coupling was introduced as in Fig. 2. Instantaneous MIPAC in these signals was then computed using a fixed value of  $k = [2, 3, 4, 6, 8, 12, 17, 24, 35, 50]$ . For each combination of these parameters, the estimated MIPAC time series and the time course

of simulated coupling were correlated. The values of these correlations for all combinations of parameters  $S_{rate}$ ,  $SNR$  and  $k$  are shown in Fig. 7. Here, higher correlation values reflect a better MIPAC estimation of the simulated coupling.

Fig. 7 shows that MIPAC estimation is negatively impacted by higher noise levels but that higher  $S_{rate}$  improves MIPAC reliability. The dependence of the MIPAC estimate on  $k$  was not as strong as its dependence on  $S_{rate}$ . For values of  $k$  below 4 and above 24, MIPAC estimation performance was poor, particularly at lower SNR levels. Values between  $k = 6$  and  $k = 12$  appear to provide slightly better estimates for a given SNR. The effect of parameter  $k$  seems to affect the MIPAC estimate mainly at low values of  $SNR$  and  $S_{rate}$ .

### 3.1.4. Estimating MIPAC on simulated multi-trial data

Next, we tested the *Event-related MIPAC* approach, estimating MIPAC for a simulated multi-trial data set, and compared the results to the PAC time series computed for these data using ERPAC (Voytek et al., 2013). For this simulation, a synthetic PAC signal similar to that shown in Fig. 2 was first used. The signal was replicated  $N_{trials} = 200$  times and stacked to form a matrix of dimension  $N_{trials}$  by number of latencies. Then, each trial was circularly shifted a random number of points between 1 and 100 (generated from a uniform distribution), equivalent to one full cycle of the simulated phase frequency. This was aiming to simulate actual trial-to-trial variations of brain responses to a repeated and/or equivalent stimulus presentation. Gaussian white noise was then added to each trial, with  $SNR = 10$ . Phase and amplitude time series for each trial were estimated as described in Section 2.1.3 and 2.6.2. Then *Event-related MIPAC* was computed and its mean progression across the trials (*Trial-mean MIPAC*) calculated. The MIPAC time series estimate for each trial is shown in the central panel of Fig. 8; The *Trial-mean MIPAC* time series is shown in the lower panel. *Trial-mean MIPAC* (red trace) and the time-resolved ERPAC (blue trace) are similar (Fig. 8, bottom).

### 3.1.5. Mutual Information Modulation Index (MImi)

Next, we compared the performance of the MImi (see Section 2.6.2) against the modulation indices obtained by using MVLmi (Section 2.2.1), KLmi (Section 2.2.2) and GLMmi (Section 2.2.3). We first simulated a signal similar to that shown in Fig. 2, with an on/off boxcar waveform modulating the coupling and using  $f_c = 50$  Hz,  $f_m = 7$  Hz. White Gaussian noise was then added to make  $SNR = 10$ . The simulated signal is shown in Fig. 9A. Modulation Index was computed on this signal for each of the

combinations of  $f_{phase} = [4, 5, 6, 7, 8, 9, 10, 11, 12]$  Hz and  $f_{amp} = [30, 35, 40, 45, 50, 55, 60, 65, 70]$  Hz by using MVLmi, KLmi and GLMmi. Additionally, *Instantaneous MIPAC* was computed for all combinations of  $f_{phase}$  and  $f_{amp}$  (Fig. 9B). MImi image (Fig. 9C) was then computed by taking the mean of MIPAC time series. MImi image was then compared with GLMmi (Fig. 9D), MVLmi (Fig. 9E) and KLmi (Fig. 9F) modulation index estimates. To facilitate comparison across methods, modulation indices were scaled by dividing the values on each figure by its maximum modulation value. The MImi estimates qualitatively replicate the results obtained using the other methods and even outperform the KLmi (Fig. 9F).

## 3.2. Applying MIPAC to actual ECoG data

### 3.2.1. Subject selection and data exploration

We applied *Event-related MIPAC* on actual electrocorticographic (ECoG) data from a human subject. Results obtained in this section are aiming to illustrate the use of the proposed method with actual data, not to support any specific interpretation of the results nor to generalize these results in advance to any other data sets.

The data described in Section 2.1.1, with electrode locations over the inferior temporal gyrus, fusiform gyrus, and lingual gyrus in the left brain hemisphere, were used after being preprocessed as described in Section 2.1.2. Combinations of  $f_{phase}$ , ranging from 2 Hz to 20 Hz in 2-Hz steps, and  $f_{amp}$  from 30 Hz to 120 Hz in 5-Hz steps, were investigated. The time series  $A_t$  and  $\varphi_t$  for all combinations of these frequency values were computed (as described in Section 2.1.3) and split into epochs time-locked to either ‘face’ image or ‘house’ image stimuli and extending from  $-400$  ms to  $800$  ms relative to stimulus onset. Next, *Event-related MIPAC* was estimated across all ECoG channels separately for face and house stimulus-locked datasets. Grand mean estimated MIPAC across trials and latencies was computed for each frequency combination to obtain comodulogram matrices of MImi estimate for each channel and stimulus type (Fig. 10).

A single channel exhibiting PAC was selected to illustrate the application of the proposed method to actual brain data. The selected channel (Channel 16) was located on the posterior side of the fusiform gyrus. An area of maximum MImi was obtained around 16 Hz (phase) and 95 Hz (amplitude) (Fig. 10 upper left dotted area) following face presentations only (Fig. 10). The channel Event-Related Potentials (ERPs) time-locked to face and house image presentations are plotted in Fig. 11 (left and right panels). For the selected channel, the responses to face presentations are more consistent across trials than responses to house presentations. Mean responses to face stimuli (Fig. 11, lower left) show a distinctive evoked negative peak (N170) at roughly 180 ms after presentation, while mean responses to house stimuli (Fig. 11, lower right) exhibit a weak positive peak near 420 ms after stimulus presentation. These ERP-image panels were plotted using EEGLAB function *erpimage.m* using a 3-trial vertical smoothing window.

To study the spectral dynamics of brain responses to presentations of face and house images, event-related spectral perturbation (ERSP) (Makeig, 1993) and inter-trial coherence (ITC) measures were computed for the same channel (16). Results are shown in Fig. 12. In the ERSP time locked to face-stimulus presentations (Fig. 12, upper left panel), beginning near 150 ms after stimulus onset a sharp power increase occurs across a broad band spanning from 4 Hz up to 120 Hz. At lower frequencies ( $\leq 50$  Hz) this power increase is brief, lasting  $\sim 100$  ms. At higher frequencies, power decreases slowly from its peak near 150 ms. ITC time locked to face stimulus presentations (Fig. 12, lower left panel) becomes significant ( $\geq 0.4$ ) in the (10–30 Hz) alpha-beta frequency range during the N170 negativity (between 150 ms and 250 ms). In ITC following house stimulus presentations (Fig. 12, lower right), a slight increase occurs near 400 ms at low frequencies ( $\leq 20$  Hz) during an ERP positivity, a feature that does not appear in the house stimulus ERSP (Fig. 12, upper right).

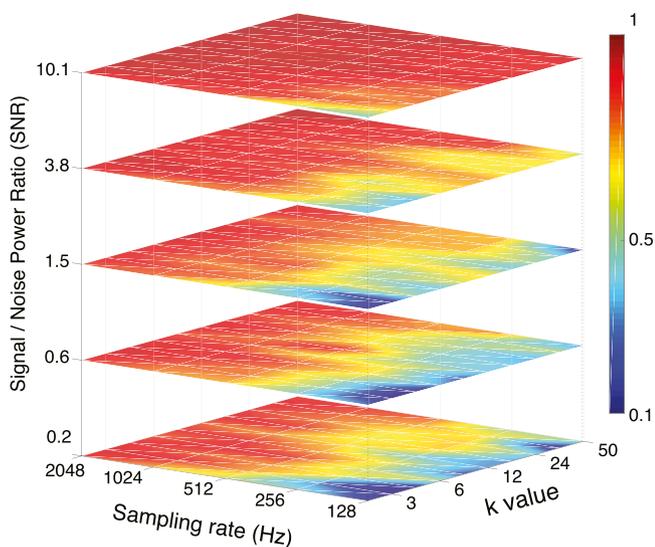
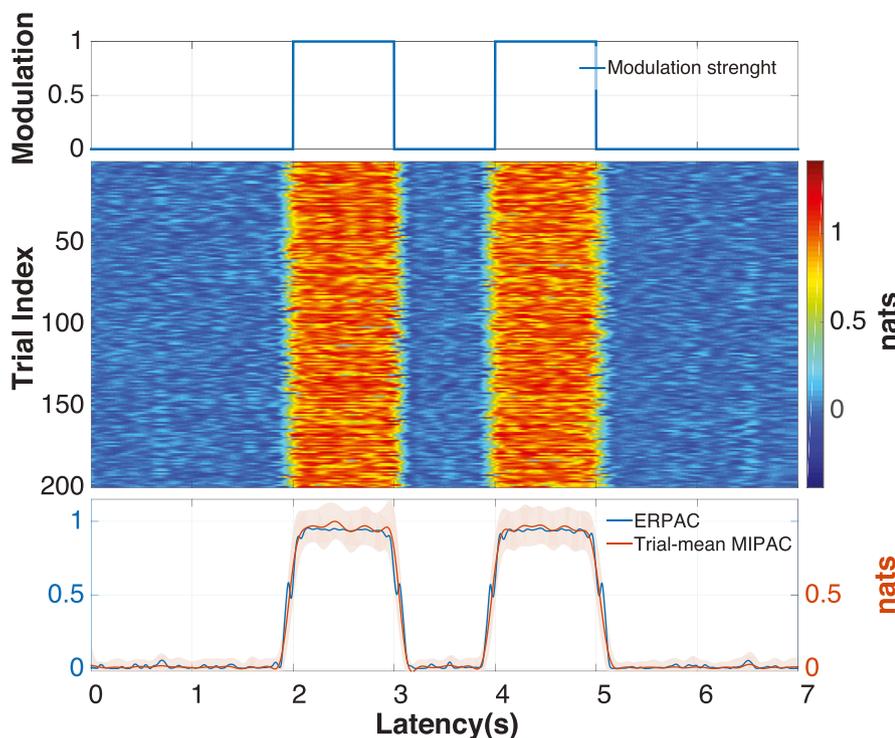


Fig. 7. Relationship between MIPAC estimates and signal parameters. Functional relationships between MIPAC estimates and parameters  $k$ ,  $S_{rate}$  and  $SNR$  were assessed by computing the correlation of the simulated coupling time courses and estimated MIPAC time series.



**Fig. 8.** Estimating MIPAC in a simulated noisy multi-trial PAC data set. A set of 200 simulated PAC signal trials like that depicted in Fig. 2 were generated to estimate Event-related MIPAC. Each trial was circularly shifted a random number of points (between 1 and 100), and Gaussian white noise was added to each trial, setting the SNR = 10. *Top:* An on/off boxcar waveform was used to vary the modulation strength of the simulated PAC signals. *Middle:* Estimated MIPAC for each trial and latency. *Bottom:* Estimated ERPAC (blue) and Trial-mean MIPAC (red). Standard deviation of MIPAC across trials is shown in light red.

Due to the non-stationarity in the signal indicated by the relatively sharp response to the face stimulus, it may be a concern that in the time-frequency analysis the power in high frequencies might be artificially created in response to sharp spike-like features in the (single-trial) signals. This issue is a well known and long discussed (Jensen et al., 2016) side effect of all Fourier-like techniques that may lead to spurious PAC, since artificial high frequencies returned by the analysis may then be coupled in time to slower oscillations. While this can occur in some cases, in our case, high-frequency components resembling wave complexes ( $\geq 50$  Hz) can be seen in single responses to face image presentations. In Fig. 13 we show two such trials (trials 113 and 91 in Fig. 11) time-locked to face stimulus that display this effect. Notice here that the increase in the power of the gamma band around 95 Hz (Fig. 13 red traces) is coherent with the high-frequency oscillation superimposed to the ERP starting around 200 ms (Fig. 13 blue traces). This argues for the validity of the high-frequency components found in the data (Fig. 12), but nevertheless is an issue that may need special attention in the general case.

### 3.2.2. Trial-mean MIPAC and ERPAC

In the face-image trials ERSP shown in Fig. 12, the simultaneous event-related increases in power in the beta and gamma bands (maximal in broad sub-bands containing 95 Hz (78–112 Hz) and 16 Hz (15–17 Hz) sub-bands respectively) coincide with a peak in ITC at  $\sim 16$  Hz surrounding the negative N170 trough. This indicates the presence of PAC at these frequencies. This finding is consistent with the relatively high degree of Mimi found for this channel (Fig. 10).

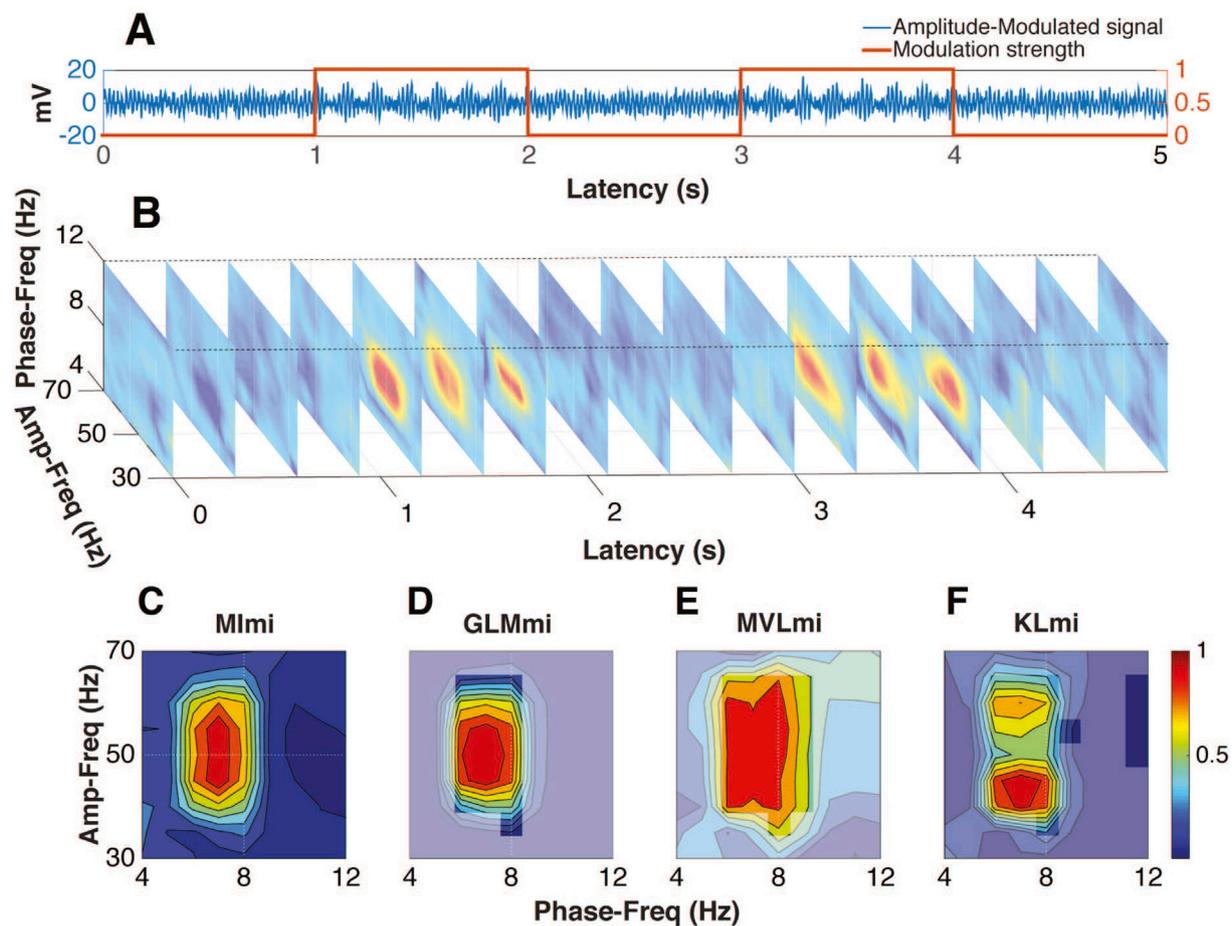
Next, Event-related MIPAC, Trial-mean MIPAC and ERPAC were computed for face and house image trials. Results of Trial-mean MIPAC and ERPAC were compared (Fig. 14). Trial-mean MIPAC for both stimulus trial sets (Fig. 14, upper panel) is consistent with the results from ERPAC (Fig. 14, lower panel). Statistical significance of increases in Trial-mean MIPAC and ERPAC following face-image presentations (compared to pre-stimulus baseline) were computed using a family-wise-error rate (FWER),  $t_{max}$  permutation test (Groppe et al., 2011), and are shown in Fig. 14 in pink shading. The empirical distribution of maximal  $t$ -values under the null hypothesis was obtained by generating 5000 100-time point

surrogate epochs from the pre-stimulus baseline. The null hypothesis was then rejected at a significance level  $\alpha = 0.05$  if the actual statistic  $t$  of the analyzed sample was beyond the  $1 - \alpha$  percentile of the empirical distribution. No significant differences from baseline were found in responses to house image presentations. Thus, further analyses focused on responses to face image presentations.

### 3.2.3. Event-related MIPAC

Results of MIPAC estimation using Event-related MIPAC ( $f_{phase} = 16$  and  $f_{amp} = 95$  Hz) for all face-presentation trials are shown in Fig. 15A, which we may refer to as a (peak MIPAC-sorted) MIPAC-image. In this figure, a statistical significance mask ( $p < 0.05$ , uncorrected) was computed by surrogate data analysis (see Section 2.7); non-significant values were set to zero and are shown in green color. In the MIPAC-image (Fig. 15A), trials were sorted by mean MIPAC across a window centered at 200 ms and spanning the equivalent of one full cycle of  $f_{phase}$  (the region between the dotted lines). In this time window, MIPAC consistently increases from bottom to top across the (peak MIPAC-sorted) trials. The same trial sorting order was then used to plot the time course of instantaneous phase at 16 Hz (Fig. 15C) and power at 95 Hz (Fig. 15B) in the same trials. The interval of increased MIPAC (180–250 ms) co-occurs with the N170 ERP peak (see Fig. 15D). However the MIPAC values are not homogeneous across trials. Trials with higher MIPAC in the sorting window also exhibit a distinct power increase at 95 Hz (Fig. 15B). This occurs concurrently with a partial phase reset at 16 Hz (Fig. 15C) beginning just after stimulus presentation and lasting until about 430 ms as also indicated by the ITC image. Here, latencies showing negative MIPAC may be interpreted as exhibiting no coupling.

The Fig. 15G shows the time course of 16-Hz phase in face image trials now explicitly sorted by phase at 16 Hz in a window centered at 200 ms spanning one full cycle of  $f_{phase}$  (e.g., between the thin dotted vertical lines). Using this trial-sorting order, the phase resetting after the stimulus presentation becomes even more obvious (Fig. 15G versus C). The same trial-sorting order was used then to plot the MIPAC-image (Fig. 15E) and the time course of power near 95 Hz in each trial (Fig. 15F). For the MIPAC-image (E), as in (A), a statistical significance mask ( $p < 0.05$ , uncorrected) was computed using surrogate data



**Fig. 9.** Comparing results of Mimi with those of other modulation indices (A): Simulated PAC signal with noise added ( $SNR = 10$ ) (blue). An on/off boxcar waveform is used to model the PAC modulation strength (red) in the simulated signal. In this simulation, the amplitude frequency is  $f_c = 50$  Hz, the phase frequency is  $f_m = 7$  Hz, and the sampling rate is  $S_{rate} = 500$  Hz. (B): MIPAC estimates for all combinations of phase frequencies from 4 to 12 Hz (1 Hz steps) and amplitude frequencies from 30 to 70 Hz (5-Hz steps). Comodulograms using (C) Mimi, (D) GLMmi (Penny et al., 2008b), (E) MVLmi (Canolty et al., 2006a) and (F) KLmi (Tort et al., 2010). Non significant values ( $p < 0.05$ , uncorrected) appear shaded in (D-F).

analysis (non-significant MIPAC estimates were set to zero and are shown in green color). We can see that the three measures depicted here, MIPAC (E), high-frequency amplitude (F), and low-frequency (16-Hz) phase (G), all appear to co-vary in latency across trials during the N170 negative peak (H).

## 4. Discussion

### 4.1. MIPAC estimates of simulated data

We showed that both MIPAC variants, *Instantaneous MIPAC* and *Event-related MIPAC*, are able to successfully estimate the temporal dynamics of the coupling in simulated PAC signals both without and with added noise. In simulations, as the noise level in the simulated signal increased, MIPAC estimation accuracy decreased. Although analyzed here only in three comparable simulations, we expect this result to generalize. Noise introduced into the simulated signals produces an increased uniform spread of points in the joint space spanned by the supports of  $A_t$  and  $\varphi_t$ , leading to more uniformity in the numbers of neighbors counts used in the local MI computation and undermining the estimation of MIPAC. A quantitative evidence for this is the decrease of the overall MI values from the noiseless simulations in Figs. 2, 3, 4-C from 0.71, 0.31 and 0.89 nats to 0.69, 0.14 and 0.84 nats respectively in the simulations with noise added in Fig. 5A-C.

We also studied the reliability of estimating coupling dynamics using *Instantaneous MIPAC* in single trials. For this, we assessed the relationship

of the signal parameters  $SNR$  and  $S_{rate}$  to the number of neighbors  $k$  used in MIPAC estimation (Fig. 7). Correlation between the modeled coupling strength and estimated MIPAC time series was used as a measure of the quality of MIPAC estimation under the given parameters. Similar to other PAC estimation methods (Penny et al., 2008b), we showed that the reliability of the MIPAC estimation is inversely proportional to noise level and directly proportional to the signal sampling rate. In our method, the later correlation occurs since as the  $S_{rate}$  increases, the joint space spanned by the supports of  $A_t$  and  $\varphi_t$  becomes more locally populated, leading to increased estimates of marginal neighbor counts for points with relevant PAC. Here, with locally populated space, we want to highlight the difference between the increase of uniformity introduced by the increase of the noise level and of the local density increase in the joint space evoked by the increase of  $S_{rate}$  in signals displaying PAC.

In the same simulation, for a given  $S_{rate}$  we were able to identify a  $k$ -value interval that gave a high-reliability MIPAC estimate. This result supports our conclusion that, given a signal with PAC (real or simulated), an optimal value of  $k$  can be estimated. This supports our implementation of a heuristic in the form of the iterative scheme for the  $k$ -value used to estimate MIPAC here. However, the optimal values of  $k$  obtained cannot be generalized, since the result is dependent on the dynamics of the coupling in each signal. As a heuristic, the proposed iterative (or convergence) scheme is not a robust method to uniquely determine the MIPAC estimate, but rather gives a coarse approximation of the optimal  $k$ . Indeed, in our examples the combination of this heuristic and the low-pass filtering at the end of MIPAC estimation produced fair estimates. We

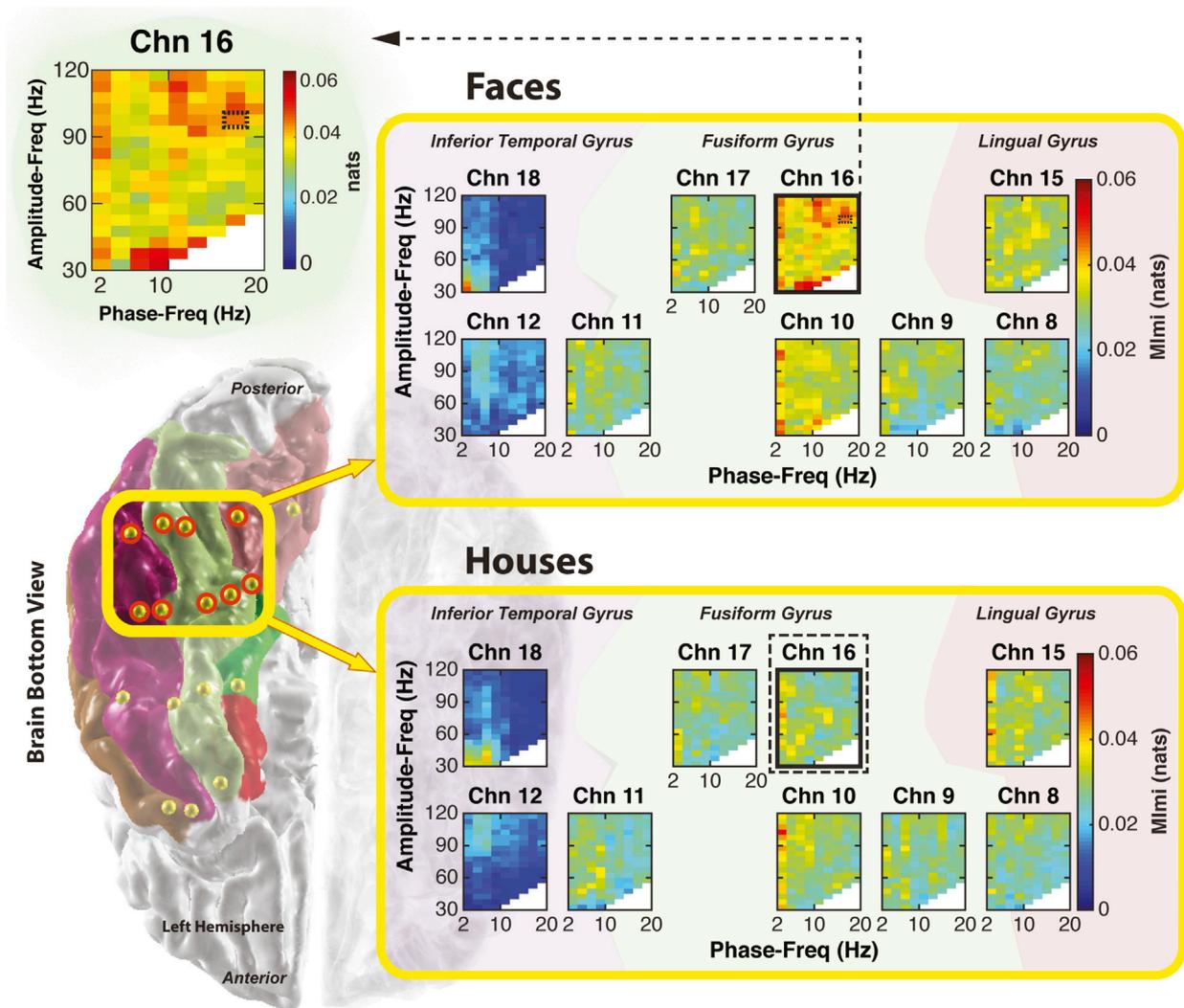


Fig. 10. Electrode locations (lower left) and Mimi comodulograms computed for face and house stimulus-locked trials at a subset of ECoG channels electrodes. The comodulogram for responses to face presentations in Channel 16 is magnified in upper left. Relevant Mimi at 95Hz (amplitude) and 16Hz (phase) is highlighted with a dotted box (upper left). Note that there is no Mimi following house image presentations for this channel (dashed box lower right). No significance testing was performed on the comodulograms.

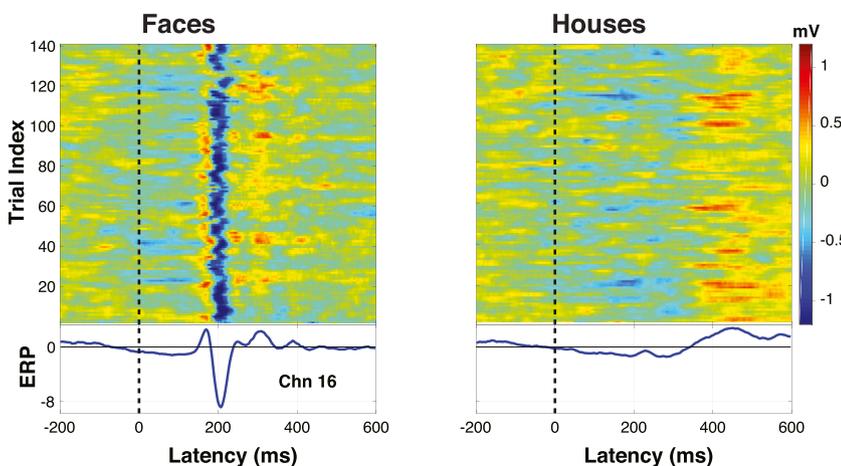
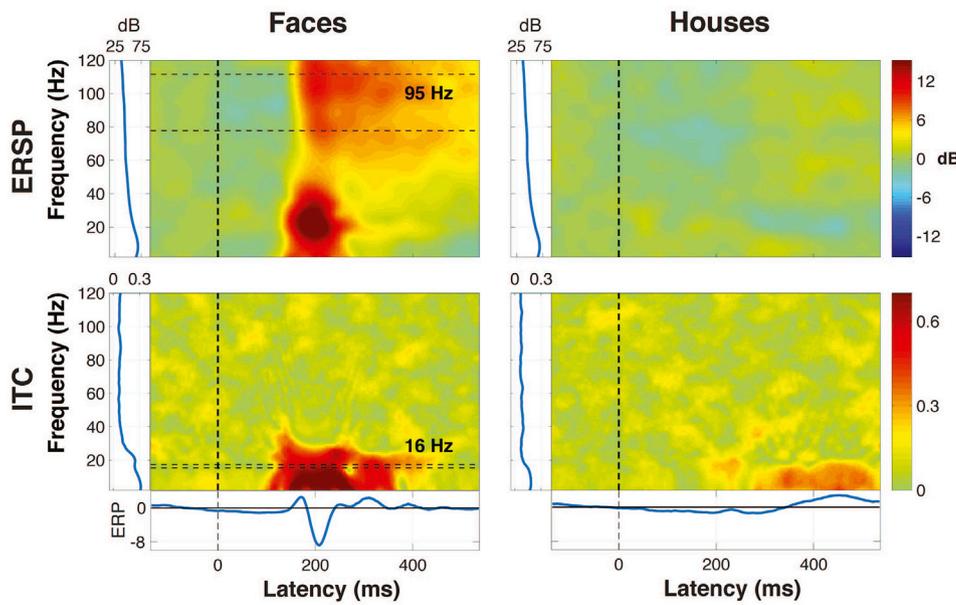
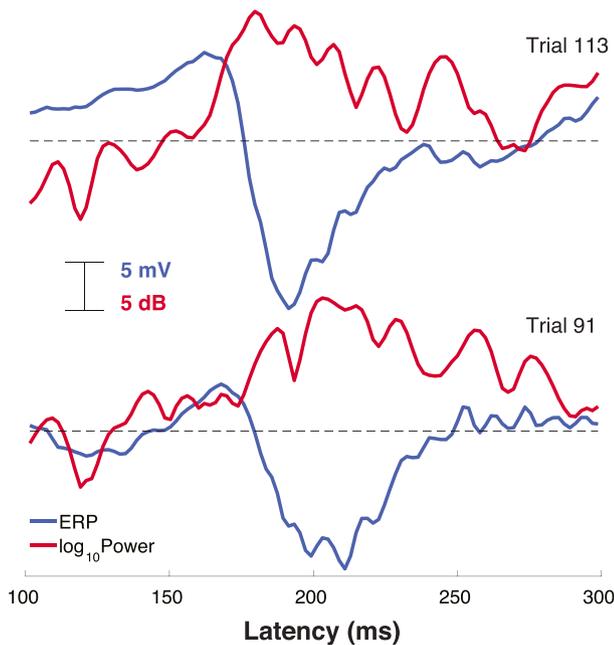


Fig. 11. ERP-images for Channel 16 trials time locked to presentations of face and house image stimuli. Stimulus onset is at 0 ms and trials have been (vertically) smoothed with a 3-trial moving window. Left: Single ECoG trials time-locked to the presentation of face stimuli. Right: Trials time-locked to presentations of house stimuli. Lower panels plot the trial-mean ERP for each trial subset. At Channel 16, located on the Fusiform Gyrus, the response to face stimuli includes a prominent and expected (N170) negativity that does not appear following presentations of house stimuli.



**Fig. 12.** Event-related spectral perturbations (ERSPs) and inter-trial coherences (ITCs) for the same channel (16) time locked to (left) face and (right) house image presentations. The trial-mean ERPs are shown in the traces below the ITC panels. The face-stimulus ERSF exhibits a broadband power increase near 200 ms after stimulus onset. Beta-band frequencies (16–30 Hz) exhibit strong ( $\geq 0.5$ ) ITC near 200 ms. The ERSF and ITC responses to house stimuli (right) are weaker and peak later (400 ms). Jointly, these phenomena may suggest the presence of event-related phase-amplitude coupling (PAC) near  $f_{\text{phase}} = 16$  Hz and  $f_{\text{amp}} = 95$  Hz. Attending to Section 2.1.3, these bands are represented with horizontal dotted lines with values between 78–112 Hz (upper left) and 15–17 Hz (lower left) respectively.



**Fig. 13.** Two single face image presentation trials displaying high-frequency activity: The figure shows (blue traces) trials 91 and 113 in Fig. 11 and time courses for the same two trials of power near 95 Hz (band limits, 78 Hz–112 Hz). The computed power increases are consistent with the appearance of high-frequency oscillations superimposed in the ERPs surrounding the negative (N170) ERP peak.

also note that small values of the parameter  $Var_{\text{thres}}$  may lead to the limiting case described in Section 3.1.3 where the estimated coupling can become constant, or analogously, have zero variance. As this is an undesirable feature, a careful assessment of parameter  $Var_{\text{thres}}$  is encouraged in real data applications. These results were obtained while estimating MIPAC in a single-trial signal using *Instantaneous MIPAC*. However, if we consider *Event-related MIPAC* as an iterative application of *Instantaneous MIPAC* at each latency across trials, then these results can be applied and generalized without loss of rigor to the case of multiple trials. For such data, *Event-related MIPAC* takes advantage of the presumed

cyclostationarity of the event-related trials to boost the neighbor count in the MIPAC estimation. The use of the one phase-cycle estimation window both allows the MIPAC estimates to sample from all parts of the full phase-frequency cycle, but also can better capture transient PAC across trials in which the event-related PAC time courses have some latency differences. On a practical level, these considerations support the reformulation of *Instantaneous MIPAC* into *Event-related MIPAC*.

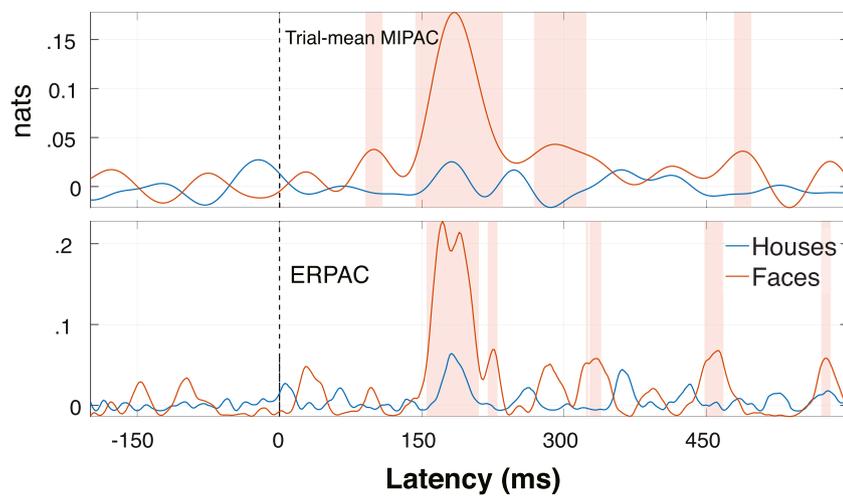
In the simulation 3.1.4, MIPAC computed on a set of simulated PAC signals was compared to the PAC result obtained using ERPAC. We demonstrated that both ERPAC and *Trial-mean MIPAC* could successfully estimate changes in PAC modulation strength. In addition, *Event-related MIPAC* estimated the modulation strength of PAC dynamics in each individual trial. From our view, this is one of the main benefits of the proposed MIPAC method vis a vis existing methods.

Finally, in Section 3.1.5 we compared Mimi performance on a simulated signal to the modulation indices obtained by GLMmi, MVLmi and KLmi. Here we demonstrated the degree of agreement among the results across all these methods. In our work, Mimi was introduced as an alternative to the modulation index used by other authors (Penny et al., 2008a; Canolty et al., 2006a; Tort et al., 2010). Our alternative came from the equivalence of MIPAC, when computed without low pass filtering of the local MI, to the overall MI. This measure has the potential to be used in analyses like the one performed at the beginning of Section 3.2, but note that this improvement may come at the cost of compromising critical temporal information about the PAC process.

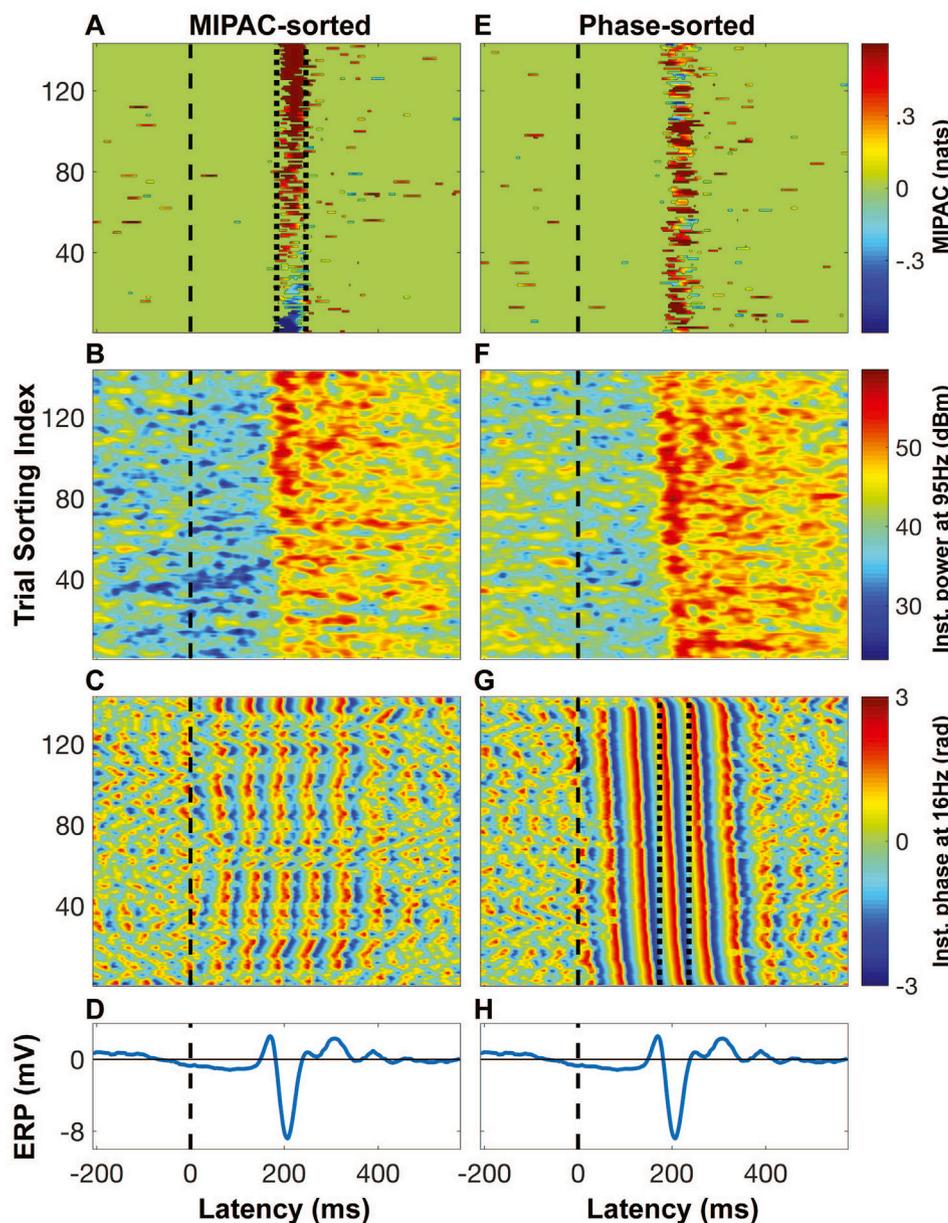
#### 4.2. MIPAC estimates of experimental data

To illustrate the potential of the proposed method, *Event-related MIPAC* was applied to real ECoG data in Section 3.2. Here a single, average-referenced ECoG channel with the active electrode located in the fusiform gyrus and exhibiting high Mimi in response to face image presentations was selected for further investigation. Selecting a channel because of a particular feature in its comodulogram, then analyzing the very same feature on the channel, might constitute a classic example of *double dipping* (Kriegeskorte et al., 2009). Also, most of the presented statistical analysis lacks multiple comparison correction (e.g., over sensors, conditions, frequencies). However, as mentioned previously, this analysis was presented here for illustrative purposes, without investigating whether the results might or not generalize to new subjects, etc.

The specificity of the fusiform gyrus response to visual face image presentations, and its expected (N170) ERP response to face image pre-



**Fig. 14.** Trial-mean MIPAC and ERPAC computed for Channel 16 at  $f_{phase}$  16 Hz and  $f_{amp}$  95 Hz. *Top:* Trial-mean MIPAC following face image (red) and house image (blue) presentations. *Bottom:* ERPAC following face image (red) and house image (blue) presentations. Pink shaded areas indicate intervals of significant PAC ( $p < 0.05$ , FWER corrected over both time and conditions) following face image presentations. Both event-related MIPAC and ERPAC measures give similar results that feature a PAC maximum near 185 ms following face image presentations only.



**Fig. 15.** MIPAC-images showing event-related MIPAC time series for individual face image trials. (A): MIPAC-image with trials sorted by the mean MIPAC value near 200 ms (between the vertical dotted lines). Points with statistically non-significant MIPAC estimates ( $p > 0.05$ , uncorrected) were set to zero and are shown in green color. (B): Power near 95 Hz in trials sorted as in A. (C): Phase at 16 Hz, trials again sorted as in A. (D): Trial-mean ERPs. (E-H) as in A-D, trials here sorted by 16-Hz phase in a latency interval centered at 200 ms (see dotted lines in panel G). Images in panels A and E are smoothed with a 3-trial moving window.

presentations are well known (McCarthy et al., 1997). However, the observed increase of beta-gamma band PAC during the surface-negative N170 ERP peak is a novel observation that may deserve further investigation. Here, the high-frequency broadband (HfB) activity increase in the ERSP (Fig. 12) following face presentations is consistent with the phase-amplitude coupling increase. The simultaneous  $\sim 16$ -Hz ITC and  $\sim 95$ -Hz HfB power increases occur together during the PAC period. The increase in power near 16 Hz is a necessary condition for emergence of meaningful PAC, since filtering may create an artificial appearance of activity at the phase frequency, which can lead to a spurious correlation. Interestingly, the local PAC values across trials (Fig. 15A, E) are not homogeneous.

Fig. 14 shows that results of *Trial-mean MIPAC* are similar to those obtained using ERPAC on the same data. This supports the validity and potential utility of the new MIPAC measure. Here the *Trial-mean MIPAC* was computed from the *Event-related MIPAC* estimates only for comparison to another PAC estimation method (ERPAC). MIPAC offers a richer description of PAC dynamics than other methods by providing a time-resolved PAC estimate for each trial.

Here, whether or not PAC is a brain mechanism supporting brain information transfer remains unanswered. However, MIPAC has the potential to add correlative evidence to this question by allowing a more complete description of PAC dynamics. Indeed, applying MIPAC to more complex data and relating it to behavior (reaction times, performance, etc.) could add compelling correlative evidence for the neurophysiological role of PAC. Appropriate modeling should also be carried out to disentangle the relationship of PAC to the measures involved in its computation (e.g., other spectral characteristics of the brain processes).

Since the MIPAC measure was developed within the framework of mutual information, we can interpret its moment-to-moment and trial-to-trial variability as indexing changes in the amount of information shared by the phase and amplitude signals. The fact that MIPAC can be used to estimate PAC with high temporal resolution is valuable. In an analysis of actual brain ECoG data, we were able to detect the appearance of PAC time locked to face stimulus presentations and to demonstrate the use of MIPAC for event-related PAC estimation, giving an insight into the dynamics of the coupling process without collapsing the dimensionality of the data. However, as we will discuss shortly, any conclusions based on these results should be drawn carefully.

#### 4.3. Limitations, caveats and future directions

One should be aware that MIPAC gives a measure of the degree of relationship of any sort between higher-frequency amplitude and lower-frequency phase time series. Thus, a MIPAC peak may not mark a high-frequency amplitude peak only, but might signal some other form of consistent relationship between the two time series.

An increasing concern in PAC reports is the inability of all PAC methods to deal with spurious CFC in non-stationary signals. Generally, non-stationary processes show spectral correlations between components of their Fourier expansions (Aru et al., 2015; Davis et al., 2011; Jensen et al., 2016) that can be misinterpreted as CFC. For example, the Fourier decomposition of a non-stationary signal with sharp spike-like features may lead to the spurious conclusion that some high-frequency activity is coupled to slower rhythms. This has been called spurious CFC. Some authors (Aru et al., 2015) have suggested that statistical analysis using surrogate data may help in dealing with the ambiguity introduced by the spurious CFC. Alternatively, a causal analysis between the CFC spectral components may be able to overcome these limitations. The new MIPAC method proposed here, like traditional PAC metrics, do not solve the problem of spurious CFC. Thus, physiological interpretation of MIPAC results should be subject to stringent statistical analysis supported by evidence for its physiological plausibility and function.

By defining the MIPAC-image (Fig. 15 A, E) we were able to visualize dynamics in the single-trial MIPAC estimates and to look for patterns – here, in particular, single-trial co-variations of event-related changes in

high frequency power and low frequency phase that were otherwise invisible. Thus, the proposed method might be a more powerful tool for exploratory analysis to gain insight into the evolution of the PAC process and its dynamic relationship to other electrophysiological measures.

The local information-theoretic measures used to create the MIPAC measures might also be applied to different types of cross-frequency coupling, since procedures we used made no assumptions as to the nature of the signals being analyzed. Thus, any type of CFC might potentially be estimated by a similar MI-based approach, perhaps simply by replacing the phase and amplitude signals with any pair of measures of interest applied to the same or to two different signals.

In addition to the potential problem of spurious coupling, we have identified three main limitations of the MIPAC approach. First, is that due to the scaling performed to compute the nearest neighbors in Section 2.6.1, MIPAC may be sensitive to outliers. For this, an alternative scaling may be used, or one may simply clean the data ahead of computing MIPAC. The fact that signals are band-pass filtered in a narrow band may also help to overcome the problem of the outliers. Second, like other PAC measurement methods, MIPAC does not provide information about the directionality of the phase-amplitude interaction (e.g., whether lower (phase) frequency oscillations appear to drive changes in the power of the high frequency signals and/or vice versa). A natural solution to this limitation within the framework of information theory might be to combine PAC measurement with transfer entropy estimation (Schreiber, 2000). So far as we know, this has not yet been applied using local information measures.

The third limitation may be inaccuracy in MIPAC estimation arising from sub-optimal sampling of the phase and amplitude values across trials. For *Event-related MIPAC*, estimation of the number of neighbors within a phase-frequency full cycle window alleviates this problem. However, if the trial-to-trial variations in PAC time course vary across a wider interval than one cycle, MIPAC may be underestimated. Another possible way to compute event-related MIPAC might be first to compute MIPAC on the continuous signal (or on the concatenated set of signal epochs) using *Instantaneous MIPAC* and then to extract MIPAC estimates in segments time-locked to events of interest. For this, some computational means would have to be implemented to handle the large dimensionality of the distance matrix needed to find the nearest neighbors. Fast nearest-neighbor search algorithms bypassing the need to compute the entire distance matrix exist and might be usable for this purpose (e.g., (Grassberger, 1990; Friedman et al., 1977)). However, this approach would not make use of the cyclostationarity assumption that the *Event-related MIPAC* approach capitalizes on, and so might work well only if local stationarity in the continuous (or concatenated) data can be assumed (see Section 2.6).

It has been widely demonstrated that in the hippocampus (Bragin et al., 1995; Colgin et al., 2009; Lisman and Jensen, 2013; Igarashi et al., 2014; Tort et al., 2009), human medial temporal lobe (Axmacher et al., 2010) and in the visual system in both human and non-human primates, (Osipova et al., 2008; Voytek et al., 2010; Spaak et al., 2012; Foster and Parvizi, 2012; Bahramisharif et al., 2013; Roux et al., 2013), low-frequency phase of local field activity can be coupled to high-frequency power. It is often assumed that PAC indexes transfer of information from low-frequency phase to high-frequency amplitude, i.e., low-frequency phase ‘drives’ high-frequency amplitude. However, most current methods for estimating PAC cannot distinguish the direction of this information flow (but see (La Tour et al., 2017)). The possibility of bi-directional interaction is supported by a recent report on ECoG data from epileptic patients that the envelope of gamma oscillations drives alpha phase (Jiang et al., 2015). No doubt, estimation of the directionality of the coupling is important for proper interpretation of PAC. Given this evidence and the temporal description provided by MIPAC, a natural question would be how to explore the directionality of the phase-amplitude interactions underlying MIPAC.

Quite recently, a new method (La Tour et al., 2017) has been developed to estimate cross-frequency coupling using non-linear

auto-regressive models. This method has the potential to model causal cross-frequency interactions by estimating time delays in interactions between the different frequency components. Further, a recent report (Barnett et al., 2009) has suggested a unified framework for data-driven causal inference that bridges both information-theoretic and auto-regressive methods. The report demonstrates the equivalence of Granger causality and transfer entropy under Gaussian assumptions on the variables. Information theoretic measures, indeed, may provide a non-linear model-free approach suitable for PAC estimation. Thus, based on previous work (La Tour et al., 2017) and on the findings provided by Barnett et al., 2009, one might propose a data-driven estimation method for CFC including causal analysis based on local transfer entropy (Lizier, 2014b) rooted in the same framework presented here.

## 5. Conclusions

Here we propose a novel method, MIPAC, for estimating the dynamics of phase-amplitude coupling (PAC) in (continuous) single-trial and (event-related) multi-trial data using a measure of local mutual information. Two mutual-information PAC (MIPAC) variants, *Instantaneous*

## Appendix A. Phase-amplitude coupled signal modeling

Let us start by considering a carrier signal  $S_c$  of frequency  $f_c$  and amplitude  $A_c$  as in Eq. (A.1).

$$S_c(t) = A_c \sin(2\pi f_c t) \quad (\text{A.1})$$

The amplitude of this carrier wave will be modulated by a signal  $S_m$  of frequency  $f_m$  and amplitude  $A_m$  (Eq. (A.2)).

$$S_m(t) = A_m \cos(2\pi f_m t) \quad (\text{A.2})$$

Then, the simulated PAC signal is expressed as the sum of the AM and modulator signal (Eq. (A.3)). Where the modulation is controlled by the parameter  $M(t)$ .

$$S_{sim}(t) = [1 + M(t)S_m(t)]S_c(t) + S_m(t) \quad (\text{A.3})$$

In the simulations presented in this manuscript,  $A_c = 5$ ,  $f_c = 40$  and  $f_m = 5$ , the time series of the modulation strength  $M(t)$  were constructed according to the type of simulation. For this, the signal was split in five identical segments each of 1 s duration. Three different functions were used to generate the  $M(t)$  time series simulating dynamic changes in the amplitude modulation. In the first two cases (on/off) boxcar and linearly ramped time courses were used to simulate coupling in alternate signal segments. In the third case, coupling was applied as the absolute value of a sinusoid across all the segments. The scripts and functions used to simulate the PAC signals in the manuscript can be obtained from the authors upon request.

## References

- Allen, E.A., Liu, J., Kiehl, K.A., Gelernter, J., Pearson, G.D., Perrone-Bizzozero, N.I., Calhoun, V.D., 2011. Components of cross-frequency modulation in health and disease. *Front. Syst. Neurosci.* 5, 59. <https://doi.org/10.3389/fnsys.2011.00059>.
- Aru, J., Aru, J., Priesemann, V., Wibral, M., Lana, L., Pipa, G., Singer, W., Vicente, R., 2015. Untangling cross-frequency coupling in neuroscience. *Curr. Opin. Neurobiol.* 31, 51–61.
- Axmacher, N., Henseler, M.M., Jensen, O., Weinreich, I., Elger, C.E., Fell, J., 2010. Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proc. Natl. Acad. Sci. Unit. States Am.* 200911531.
- Bahramisharif, A., van Gerven, M.A., Aarnoutse, E.J., Mercier, M.R., Schwartz, T.H., Foxe, J.J., Ramsey, N.F., Jensen, O., 2013. Propagating neocortical gamma bursts are coordinated by traveling alpha waves. *J. Neurosci.* 33 (48), 18849–18854.
- Bahramisharif, A., Mazaheri, A., Levar, N., Richard Schuurman, P., Figeo, M., Denys, D., 2016. Deep brain stimulation diminishes cross-frequency coupling in obsessive-compulsive disorder. *Biol. Psychiatry* 80 (7), e57–e58. <https://doi.org/10.1016/j.biopsych.2015.05.021>. <http://www.sciencedirect.com/science/article/pii/S0006322315004709>.
- Barnett, L., Barrett, A.B., Seth, A.K., 2009. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* 103 (23), 238701.
- Berens, P., et al., 2009. Circstat: a matlab toolbox for circular statistics. *J. Stat. Software* 31 (10), 1–21.
- Berger, H., 1929. Über das elektroencephalogramm des menschen. *Eur. Arch. Psychiatr. Clin. Neurosci.* 87 (1), 527–570.
- Bragin, A., Jandó, G., Nádasdy, Z., Hetke, J., Wise, K., Buzsáki, G., 1995. Gamma (40–100 Hz) oscillation in the hippocampus of the behaving rat. *J. Neurosci.* 15 (1), 47–60.
- Bruns, A., Eckhorn, R., 2004. Task-related coupling from high- to low-frequency signals among visual cortical areas in human subdural recordings. *Int. J. Psychophysiol.* 51 (2), 97–116.

*MIPAC* and *Event-related MIPAC*, were here validated by applying them to simulated data and then to actual cortical surface (ECoG) brain data. In both cases, results were consistent with results of published PAC estimators. The main advantage of the proposed MIPAC measures is their more highly time-resolved description of PAC process dynamics. This makes MIPAC suitable, in particular, for the analysis of either spontaneous or event-related dynamics in electrophysiological data.

## Funding sources

This work was supported by National Institutes of Health grant 5R01-NS047293-13 and by a gift from The Swartz Foundation (Old Field, NY).

## Acknowledgements

The authors would like to thank Joseph Lizier and Andre Fonseca for discussions on the mutual information estimator and Yasser Aleman-Gomez for helping with Fig. 10. Many thanks as well to Johanna Wagner and Alejandro Ojeda for fruitful discussions and manuscript editing. Thanks also to the reviewers for their inputs and comments.

- Buzsáki, G., Draguhn, A., Buzsáki, G., 2004. Neuronal oscillations in cortical networks. *Science* 304 (5679), 1926–1929. <https://doi.org/10.1126/science.1099745>. <http://www.ncbi.nlm.nih.gov/pubmed/15218136>.
- Canolty, R.T., Edwards, E., Dalal, S.S., Soltani, M., Nagarajan, S.S., Kirsch, H.E., Berger, M.S., Barbaro, N.M., Knight, R.T., 2006. High gamma power is phase-locked to theta oscillations in human neocortex. *Science (New York, N.Y.)* 313 (5793), 1626–1628. <https://doi.org/10.1126/science.1128115>.
- Canolty, R.T., Edwards, E., Dalal, S.S., Soltani, M., Nagarajan, S.S., Kirsch, H.E., Berger, M.S., Barbaro, N.M., Knight, R.T., 2006. High gamma power is phase-locked to theta oscillations in human neocortex. *Science* 313 (5793), 1626–1628.
- Cheng, N., Li, Q., Wang, S., Wang, R., Zhang, T., 2017. Permutation mutual information: a novel approach for measuring neuronal phase-amplitude coupling. *Brain Topogr.* 1–16.
- Church, K.W., Hanks, P., 1990. Word association norms, mutual information, and lexicography. *Comput. Ling.* 16 (1), 22–29.
- Cohen, M.X., 2008. Assessing transient cross-frequency coupling in eeg data. *J. Neurosci. Methods* 168 (2), 494–499.
- Cohen, M.X., Axmacher, N., Lenartz, D., Elger, C.E., Sturm, V., Schlaepfer, T.E., 2009. Good vibrations: cross-frequency coupling in the human nucleus accumbens during reward processing. *J. Cognit. Neurosci.* 21 (5), 875–889. <https://doi.org/10.1162/jocn.2009.21062>. <https://doi.org/10.1162/jocn.2009.21062>.
- Colgin, L.L., Denninger, T., Fyhn, M., Hafting, T., Bonnevie, T., Jensen, O., Moser, M.-B., Moser, E.L., 2009. Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature* 462 (7271), 353.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage* 9 (2), 179–194.
- Davis, R.A., Lij, K.-S., Politis, D.N., 2011. Spectral analysis for harmonizable processes. In: *Selected Works of Murray Rosenblatt*. Springer, pp. 420–460.
- De Hemptinne, C., Ryapolova-Webb, E.S., Air, E.L., Garcia, P.A., Miller, K.J., Ojemann, J.G., Ostrem, J.L., Galifianakis, N.B., Starr, P.A., 2013. Exaggerated

- phase–amplitude coupling in the primary motor cortex in Parkinson disease. *Proc. Natl. Acad. Sci. Unit. States Am.* 110 (12), 4780–4785.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134 (1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>.
- S. I. Dimitriadis, N. A. Laskaris, M. P. Bitzidou, I. Tarnanas, M. N. Tsolaki, A novel biomarker of amnesic MCI based on dynamic cross-frequency coupling patterns during cognitive brain responses. *Front. Neurosci.* 9. <https://doi.org/10.3389/fnins.2015.00350>. URL <http://journal.frontiersin.org/article/10.3389/fnins.2015.00350/abstract>.
- Fano, R.M., 1963. *Transmission of Information. A Statistical Theory of Communication.* The MIT Press.
- Florin, E., Baillet, S., 2015. The brain's resting-state activity is shaped by synchronized cross-frequency coupling of neural oscillations. *Neuroimage* 111, 26–35.
- Foster, B.L., Parvizi, J., 2012. Resting oscillations and cross-frequency coupling in the human posteromedial cortex. *Neuroimage* 60 (1), 384–391.
- Friedman, J.H., Bentley, J.L., Finkel, R.A., 1977. An algorithm for finding best matches in logarithmic expected time. *ACM Trans. Math Software* 3 (3), 209–226.
- Gómez-Herrero, G., Wu, W., Rutanen, K., Soriano, M.C., Pipa, G., Vicente, R., 2015. Assessing coupling dynamics from an ensemble of time series. *Entropy* 17 (4), 1958–1970.
- Goutagny, R., Gu, N., Cavanagh, C., Jackson, J., Chabot, J.-G., Quirion, R., Krantic, S., Williams, S., 2013. Alterations in hippocampal network oscillations and theta-gamma coupling arise before A overproduction in a mouse model of Alzheimer's disease. *Eur. J. Neurosci.* 38 (10) <https://doi.org/10.1111/ejn.12446>, 3527–3527. <http://onlinelibrary.wiley.com/doi/10.1111/ejn.12446/abstract>.
- Grassberger, P., 1990. An optimized box-assisted algorithm for fractal dimensions. *Phys. Lett.* 148 (1–2), 63–68.
- Groppe, D.M., Urbach, T.P., Kutas, M., 2011. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48 (12), 1711–1725. <https://doi.org/10.1111/j.1469-8986.2011.01273.x>. <http://onlinelibrary.wiley.com/doi/10.1111/j.1469-8986.2011.01273.x/abstract>.
- He, B.J., Zempel, J.M., Snyder, A.Z., Raichle, M.E., 2010. The temporal structures and functional significance of scale-free brain activity. *Neuron* 66 (3), 353–369.
- Hermes, D., Miller, K.J., Noordmans, H.J., Vansteensel, M.J., Ramsey, N.F., 2010. Automated electrocorticographic electrode localization on individually rendered brain surfaces. *J. Neurosci. Methods* 185 (2), 293–298.
- Hindle, D., 1990. Noun classification from predicate-argument structures. In: *Proceedings of the 28th Annual Meeting on Association for Computational Linguistics.* Association for Computational Linguistics, pp. 268–275.
- Hurtado, J.M., Rubchinsky, L.L., Sigvardt, K.A., 2004. Statistical method for detection of phase-locking episodes in neural oscillations. *J. Neurophysiol.* 91 (4), 1883–1898.
- Igarashi, K.M., Lu, L., Colgin, L.L., Moser, M.-B., Moser, E.I., 2014. Coordination of entorhinal–hippocampal ensemble activity during associative learning. *Nature* 510 (7503), 143.
- Ince, R.A., Giordano, B.L., Kayser, C., Roussellet, G.A., Gross, J., Schyns, P.G., 2017. A statistical framework for neuroimaging data analysis based on mutual information estimated via a Gaussian copula. *Hum. Brain Mapp.* 38 (3), 1541–1573.
- Jensen, O., Spaak, E., Park, H., 2016. Discriminating valid from spurious indices of phase-amplitude coupling. *ENEURO* 0334.
- Jiang, H., Bahramisharif, A., van Gerven, M.A., Jensen, O., 2015. Measuring directionality between neuronal oscillations of different frequencies. *Neuroimage* 118, 359–367.
- Kozachenko, N.L.F., 1987. Sample estimate of the entropy of a random vector. *Peredachi Inf.* 23 (2), 9–16.
- Kraskov, A., Stögbauer, H., Grassberger, P., 2004. Estimating mutual information. *Phys. Rev.* 69 (6), 066138.
- Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., Baker, C.L., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* 12 (5), 535.
- La Tour, T.D., Tallot, L., Grabot, L., Doyère, V., Van Wassenhove, V., Grenier, Y., Gramfort, A., 2017. Non-linear auto-regressive models for cross-frequency coupling in neural time series. *PLoS Comput. Biol.* 13 (12), e1005893.
- Lisman, J.E., Jensen, O., 2013. The theta-gamma neural code. *Neuron* 77 (6), 1002–1016.
- Lizier, J.T., 2014. *Directed Information Measures in Neuroscience.* Springer.
- Lizier, J.T., 2014. Measuring the dynamics of information processing on a local scale in time and space. In: *Directed Information Measures in Neuroscience.* Springer, pp. 161–193.
- López-Azcárate, J., Tainta, M., Rodríguez-Oroz, M.C., Valencia, M., González, R., Guridi, J., Iriarte, J., Obeso, J.A., Artieda, J., Alegre, M., 2010. Coupling between beta and high-frequency activity in the human subthalamic nucleus may be a pathophysiological mechanism in Parkinson's disease. *J. Neurosci.: Off. J. Soc. Neurosci.* 30 (19), 6667–6677. <https://doi.org/10.1523/JNEUROSCI.5459-09.2010>.
- MacKay, D.J., 2003. *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press.
- Makeig, S., 1993. Auditory event-related dynamics of the eeg spectrum and effects of exposure to tones. *Electroencephalogr. Clin. Neurophysiol.* 86 (4), 283–293.
- McCarthy, G., Puce, A., Gore, J.C., Allison, T., 1997. Face-specific processing in the human fusiform gyrus. *J. Cognit. Neurosci.* 9 (5), 605–610. <https://doi.org/10.1162/jocn.1997.9.5.605>.
- Miller, K.J., Schalk, G., Hermes, D., Ojemann, J.G., Rao, R.P.N., 2016. Spontaneous decoding of the timing and content of human Object perception from cortical surface recordings reveals complementary information in the event-related potential and broadband spectral change. *PLoS Comput. Biol.* 12 (1), e1004660. <https://doi.org/10.1371/journal.pcbi.1004660>. <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004660>.
- Nonoda, Y., Miyakoshi, M., Ojeda, A., Makeig, S., Juhász, C., Sood, S., Asano, E., 2016. Interictal high-frequency oscillations generated by seizure onset and eloquent areas may be differentially coupled with different slow waves. *Clin. Neurophysiol.* 127 (6), 2489–2499. <https://doi.org/10.1016/j.clinph.2016.03.022>. <http://www.sciencedirect.com/science/article/pii/S1388245716300049>.
- Osipova, D., Hermes, D., Jensen, O., 2008. Gamma power is phase-locked to posterior alpha activity. *PLoS One* 3 (12), e3990.
- Panzeri, S., Senatore, R., Montemurro, M.A., Petersen, R.S., 2007. Correcting for the sampling bias problem in spike train information measures. *J. Neurophysiol.* 98 (3), 1064–1072.
- Penny, W.D., Duzel, E., Miller, K.J., Ojemann, J.G., 2008. Testing for nested oscillation. *J. Neurosci. Methods* 174 (1), 50–61. <https://doi.org/10.1016/j.jneumeth.2008.06.035>.
- Penny, W., Duzel, E., Miller, K., Ojemann, J., 2008. Testing for nested oscillation. *J. Neurosci. Methods* 174 (1), 50–61.
- Richter, C.G., Babo-Rebelo, M., Schwartz, D., Tallon-Baudry, C., 2017. Phase-amplitude coupling at the organism level: the amplitude of spontaneous alpha rhythm fluctuations varies with the phase of the infra-slow gastric basal rhythm. *Neuroimage* 146, 951–958. <https://doi.org/10.1016/j.neuroimage.2016.08.043>. <http://www.sciencedirect.com/science/article/pii/S1053811916304281>.
- Roux, F., Wibral, M., Singer, W., Aru, J., Uhlhaas, P.J., 2013. The phase of thalamic alpha activity modulates cortical gamma-band activity: evidence from resting-state meg recordings. *J. Neurosci.* 33 (45), 17827–17835.
- Schreiber, T., 2000. Measuring information transfer. *Phys. Rev. Lett.* 85 (2), 461.
- Sotero, R.C., 2016. Topology, cross-frequency, and same-frequency band interactions shape the generation of phase-amplitude coupling in a neural mass model of a cortical column. *PLoS Comput. Biol.* 12 (11), e1005180.
- Sotero, R.C., Bortel, A., Naaman, S., Mocanu, V.M., Kropf, P., Villeneuve, M.Y., Shmuel, A., 2015. Laminar distribution of phase-amplitude coupling of spontaneous current sources and sinks. *Front. Neurosci.* 9, 454–454.
- Spaak, E., Bonnefond, M., Maier, A., Leopold, D.A., Jensen, O., 2012. Layer-specific entrainment of gamma-band neural activity by the alpha rhythm in monkey visual cortex. *Curr. Biol.* 22 (24), 2313–2318.
- Szczepanski, S.M., Crone, N.E., Kuperman, R.A., Auguste, K.I., Parvizi, J., Knight, R.T., 2014. Dynamic changes in phase-amplitude coupling facilitate spatial attention control in fronto-parietal cortex. *PLoS Biol.* 12 (8), e1001936. <https://doi.org/10.1371/journal.pbio.1001936>. <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001936>.
- Tort, A.B.L., Komorowski, R.W., Manns, J.R., Kopell, N.J., Eichenbaum, H., 2009. Theta-gamma coupling increases during the learning of item–context associations. *Proc. Natl. Acad. Sci. Unit. States Am.* 106 (49), 20942–20947. <https://doi.org/10.1073/pnas.0911331106>. <http://www.pnas.org/content/106/49/20942>.
- Tort, A.B.L., Komorowski, R., Eichenbaum, H., Kopell, N., 2010. Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *J. Neurophysiol.* 104 (2), 1195–1210. <https://doi.org/10.1152/jn.00106.2010>. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2941206/>.
- Treves, A., Panzeri, S., 1995. The upward bias in measures of information derived from limited data samples. *Neural Comput.* 7 (2), 399–407.
- Victor, J.D., 2002. Binless strategies for estimation of information from neural data. *Phys. Rev.* 66 (5), 051903.
- B. Voytek, R. T. Canolty, A. Shestyk, N. E. Crone, J. Parvizi, R. T. Knight, Shifts in gamma phase–amplitude coupling frequency from theta to alpha over posterior cortex during visual tasks. *Front. Hum. Neurosci.* 4. <https://doi.org/10.3389/fnhum.2010.00191>. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2972699/>.
- Voytek, B., D'Esposito, M., Crone, N., Knight, R.T., 2013. A method for event-related phase/amplitude coupling. *Neuroimage* 64, 416–424. <https://doi.org/10.1016/j.neuroimage.2012.09.023>. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3508071/>.
- Wibral, M., Lizier, J.T., Priesemann, V., 2015. Bits from brains for biologically inspired computing. In: *Frontiers in Robotics and AI* 2, p. 5.
- Yanagisawa, T., Yamashita, O., Hirata, M., Kishima, H., Saitoh, Y., Goto, T., Yoshimine, T., Kamitani, Y., 2012. Regulation of motor representation by phase-amplitude coupling in the sensorimotor cortex. *J. Neurosci.: Off. J. Soc. Neurosci.* 32 (44), 15467–15475. <https://doi.org/10.1523/JNEUROSCI.2929-12.2012>.