Tutorial : phase-locking and phase-amplitude coupling analysis for a Local Field Potential recording

by lilian goepp

September 2019

1 Phase-amplitude Coupling

Phase-amplitude coupling is a special type of relationship between two frequency bands of the same signal. This coupling is a configuration where the amplitude of a high-frequency band is modulated by the frequency of a low-frequency band (fig 1, B and D).

If the analysis that we wish to execute only needs to establish the existence or nonexistence of phase-amplitude-coupling, we can use the script to answer this question. The method is based on the Modulation Index computation, a method further studied by Tort and al. [1] and Hülsemann and al. [2], and on a significance criteria built by a permutation test.

The Modulation Index is built as follows : first is isolated the phase and amplitude of the studied frequency bands. We therefore filter the raw signal in the two bands of frequencies, then we extract the instantaneous phase of the low-frequency band and the amplitude envelope of the high-frequency band by applying a hilbert transformation over the two filtered signals (fig 1, C and D).

After that, we cut the [0 ;2π] interval into n bins. Then, we average the amplitude envelope over each bin of the instantaneous phase. The instantaneous phase is approximately periodic, therefore there are several replications of a same bin. We therefore average the mean amplitude envelope over each replication. Finally, we build the normalised histogram of the averaged amplitude envelope over the different angular bins. This gives us a kind of empirical probability distribution of the amplitude of the high frequency band compared to the phase of the low frequency band (fig1, E).

Our idea to build statistics over the question of phase-amplitude coupling is to consider that a situation where there is no coupling is theoretically a uniform distribution of the amplitude over the phase (constant mean amplitude for every phase). The Modulation Index

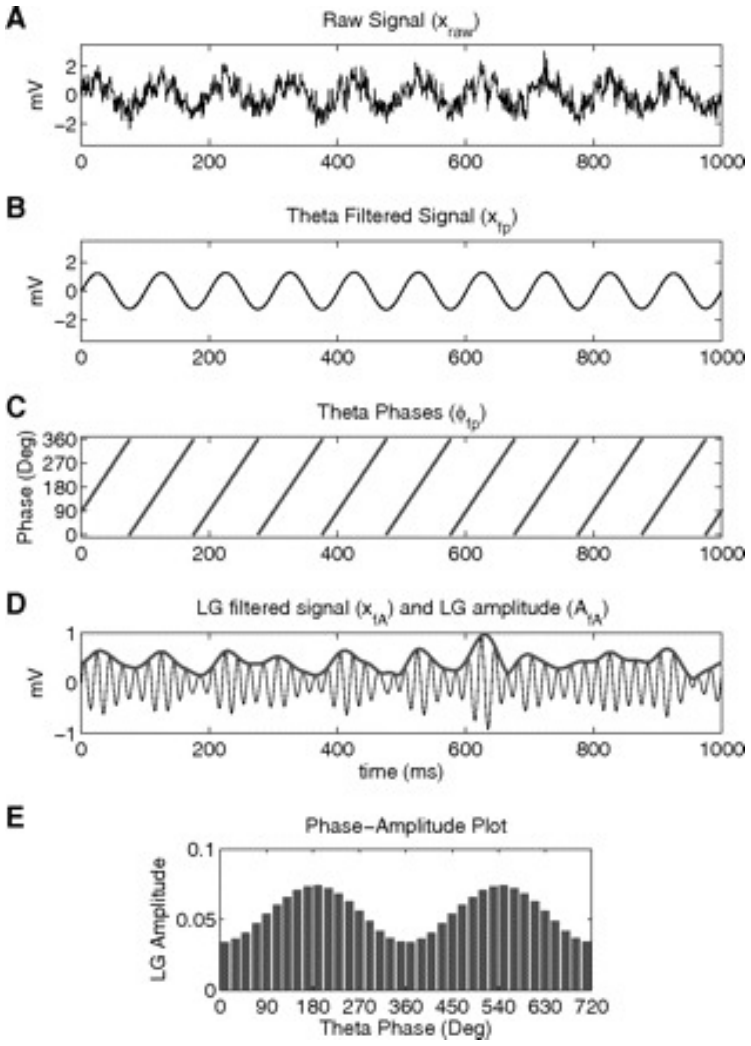

FIGURE 1 – A : example of recording of a Local Field Potential (simulated data). B : filtering for the theta frequency band. C : instantaneous phase obtained by Hilbert transform over B. D : filtering for the gamma frequency band, and the amplitude envelope obtained by Hilbert transform in bold. E : normalised histogram of the distribution of amplitude comparing to the phase (2 cycles showned for clarity). Figure adapted from the article of Tort et al. [1]

is therefore constructed as a Kullback-Leibler distance (1) normalised between the empirical probability distribution and the uniform

distribution over the n angular bins (2)

$$D_{KL}(P,Q) = \sum_{j=0}^{n} P(j) \log[\frac{P(j)}{Q(j)}] \tag{1}$$

Where : - D_{KL} is the Kullback-Leibler distance
- P and Q are probability distributions over the n bins of [0,2π]

$$\textbf{Modulation Index} = \frac{D_{KL}(P,U)}{\log(n)} \tag{2}$$

Where U is the probability distribution of a uniform law over the n bins

Finally, to assess the significance of the value obtained, we build a permutation test. The latter is a simulation of the value of the Modulation Index, but for artificially created data built by the following protocol : the amplitude of the high frequency band is randomly cut into two parts that we permute. We obtain this way a distribution of values of Modulation Index, and our test is simply a comparison between a critical value of this distribution and the value obtained at first.

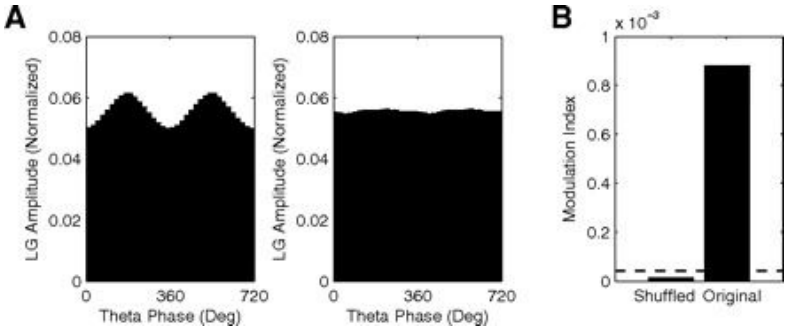

FIGURE 2 – A : left, normalised histogram of the distribution of amplitude comparing to the phase (2 cycles showned for clarity), right, example of the same histogram for a permuted amplitude. B : comparison between the MI value obtained and the mean of the MI computed for permuted amplitude. In dots, the quantile of the distribution of the simulated MI for an alpha risk of 0.01. Figure adapted from the article of Tort et al. [1]

2 Phase-locking

Phase-locking is a situation where the spikes of a neuron are synchronised with the phase of a frequency band of a Local Field Poten-

tial. In our case, spike sorting offers us for each Local Field Potential recording a list of neurons with the time of the different spikes.

Our analysis will allow us to assess if we observe a phase-locking situation. There are two different methods in the script, both of them using a sample of the instantaneous angular phase of the Local Field Potential at the moment when a spike occur.

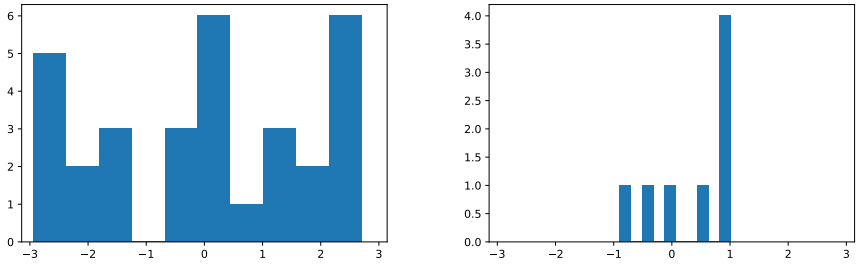

FIGURE 3 – Histogram of the disposition of the instantaneous phase of the filtered signal for each spikes of a neuron. Left, an example with chaotic disposition, right, a phase-locking situation.

To study this sample, we decided to use the concentration index of a Von Mises distribution, by forcing the data to fit. It didn't seem relevant to do a test of goodness-of-fit, as it's not the nature of the distribution but a concentration criteria that interests us.

The estimation of the concentration index is obtained by the maximum likelihood estimator. It has been shown that this estimator is biased for small sample sizes and small values of the concentration index. To fix this problem, we use a corrected estimator for sample sizes smaller than 16 [3]

2.1 Bootstrap method

The first method is adapted to experimental replication or every other analysis where a signifiance threshold has already been established. It consists of the comparison between this threshold and the lower bound of a confidence interval built by bootstrap replications

The bootstrap replications are built by reestimating of the concentration index over an artificial sample. The latter is obtained by random sampling with replacement over the original sample. By several iterations, we obtain a distribution of several values. To construct a confidence interval, we simply cut this distribution by left and right in proportion to the alpha risk considered (percentile confidence interval method).

If the threshold value belongs to the confidence interval, we interpret this as a no-phase-locking situation. In fact, the comparison between the lower bound of the confidence interval and the threshold value can be understood as a one-sided test establishing, with an alpha risk defined in our confidence interval, if our estimation is significantly higher than the threshold value. The threshold value must therefore be constructed as the concentration index from which we consider that we are in phase-locking. The test is built as follows :

$$H0 : \text{we are in a phase-locking situation} \approx \kappa > \text{threshold} \tag{3}$$

$$H1 : \text{We are not in a phase-locking situation} \approx \kappa \leq \text{threshold} \tag{4}$$

Where κ is the Von Mises concentration index

If this threshold value is higher than the lower bound of the confidence interval, it means that the probability that the real concentration index is higher than the threshold value is itself higher than 1-α. It's therefore admitting an "alpha risk" higher than alpha TRADUIRE RISQUE DE PREMIERE ESPECE

2.2 Uniform method

On several points, the second method is very different. At first, the hypotheses are reversed. We first suppose that our sample is drawn from a uniform distribution over [-π,π], which means that there is no phase-locking. With this hypothesis, we can compute the probability for the concentration index to be over some value z, for a fixed sample size. We can therefore know over which value our estimation has a probability under alpha to be. If our estimation is above this one, we consider that our sample is phase-locked. Indeed, it means that the probability for a uniform distribution to exhibit his concentration index is lower than alpha. This method is therefore corresponding to this test :

$$H0 : \text{we are not phase-locked} \approx X_1, ..., X_n \rightsquigarrow uniform([-\pi, \pi]) \tag{5}$$

$$H1 : \text{We are phase-locked} \tag{6}$$

Where X_i is the random variable matching the i-th data in the sample, and n is the sample length.

The probability function is thereby computed like this :

$$\mathbb{P}(\hat{\kappa}_{uniform} > z) = \frac{e^{-n(zA(z)-\log(I_0(z)))}}{\sqrt{\frac{A(z)}{z}(1 - A(z)^2 + \frac{A(z)}{z})}} \tag{7}$$

where I_n is the modified bessel function of order n, and $A = \frac{I_1}{I_0}$

if n is larger or equal to 16. If n is under 16, as we use the corrected estimator, we need to adapt this probability like this :

$$\begin{aligned}\mathbb{P}(\hat{\kappa}_{corr} > z) = \mathbb{P}(\hat{\kappa}_{uniform} > \frac{n^3+n}{(n-1)^3}z) \times \mathbb{P}(\hat{\kappa}_{uniform} \geqslant 2) \\ +\mathbb{P}(\hat{\kappa}_{uniform} > \frac{z+\sqrt{z^2+\frac{8}{n}}}{2}) \times \mathbb{P}(\hat{\kappa}_{uniform} < 2)\end{aligned} \tag{8}$$

As soon as we have computed the z corresponding to the alpha risk that we consider, then an estimated concentration index higher than z means that the probability for our sample to be drawn from a uniform is less than the alpha risk considered. We can therefore suppose that our sample is phase-locked with an RISQUE DE PREMIERE ESPECE under our alpha risk.

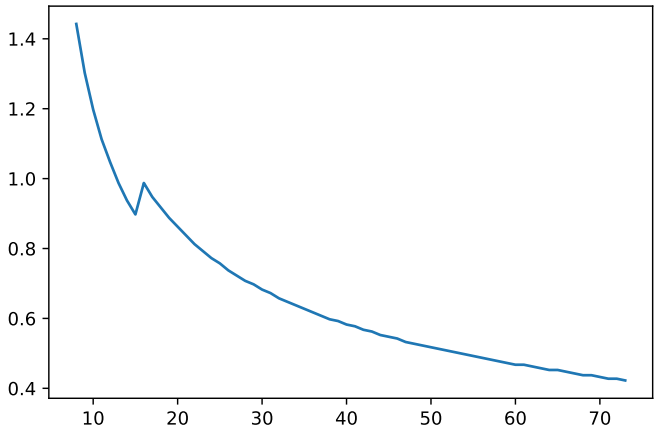

FIGURE 4 – threshold function computed for an alpha risk of 0.05 and a sample size from 8 to 73. We clearly see the cut at 16 between the classical and the corrected estimator.

This method is, I think, the most relevant when you don't know what to expect from your data, and takes particularly into account the sample size, and the principle that the smaller the sample, the more convincing it must be to assess anything. The hypotheses are also very simple : we are just wondering if it makes sense that the sample is drawn from a uniform. If not, it's necessarily phase-locked.

3 Script utilisation

In this section, we will present the in and out of the analysis, the parameters and the points were the user must be careful. We suppose that the signal is a set of recordings from several electrodes, grouped in channel groups that represent a localisation. Each record is a concatenation of the replication of an experience, without and with a difference in the protocol. We suppose that the recordings without the difference are placed at first. We also suppose that the difference doesn't affect the time length of the experience.
For the global analysis, the entries are :

- channel_groups : a list of each list corresponding to a channel group. A list corresponding to a channel group is the sequence of index identifying in the signal the records performed by the electrodes in the group in question
- nb_trial_no_stim : the number of experiences performed without a change of protocol
- t_stop : the time of the experience performed, which must be the same between the replications with and without the difference
- sampling_rate : the amount of data collected for one second of recording
- freq_band : a list of the bands of frequency you will want to be represented in a histogram at the end of the analysis.
- f_start : the frequency with which the analysis will start. It must be equal or under the lowest bound of freq_band
- f_stop : the frequency with which the analysis will stop.
- freq_bin : the order of the frequency analysis. Every frequency bands studied will be constructed as [f_start + n1 * freq_bin, f_start + n2 * freq_bin]
- scalo_threshold : the threshold in percentage of the maximum amplitude over which we consider that the signal filtered in a frequency band is not driven by noise, and is therefore used in our study.
- event_name : a list of the names of the different events of the experience performed, if you want to study events. If not, fill out one name anyway.
- time_interval : the list of time length associated with each event. If you're not studying different events, leave 0.
- file_loc_sig : the localisation of the signal file in your computer.
- alpha : the probability of false positive for the second hypothesis of each test performed.

For the phase-locking analysis, the entries are :

- file_loc_spike : the localisation in your computer of the spike sorting file. I suppose it's an excel array, and I suppose the last character of the localisation is the index of the channel group in channel_groups.
- kap_threshold_computation : boolean parameter. If true, we use for the phase-locking the second method based on the uniform hypothesis. If false, we use the bootstrap method.
- boot_kap_threshold : the critical value used in the bootstrap method.

If we use the second method, this parameter will simply not be used.
- min_spike_per_event : the minimum number of spikes during an event required to perform an analysis. If below, the result array will write "not enough data".

For the phase-amplitude coupling, the entries are :
- nb_bin_MI_histogram : the number of bins used in the modulation index computation.
- nb_permutation_replication : the number of replications to construct the distribution used in the permutation test.

The result of the phase-locking analysis is the out of the overall_kap_analysis function, with 4 pieces of information :
- disp : an array of the disposition of the different values used in the estimation of the concentration index of the Von Mises distribution. It represents the phase of the filtered signal at the moment when a spike occurs. It's arranged by no difference or difference in the protocol, channel group, neuron, event and frequency band.
- kap : an array of the estimated concentration index. It's arranged by no difference or difference in the protocol, channel group, neuron, event and frequency band.
- freq : an array of the frequency band with which the neuron is phase-locked, and the concentration index computed. If there is no phase-locking with any frequency band considered, it contains the string "no phase-locking". If in this event the number of spikes is under min_spike_per_event, it contains "not enough data". It's arranged by no difference or difference in the protocol, channel group, neuron and event.
- cluster_freq : the list of the frequency bands considered in our analysis.

The result of the phase-amplitude coupling analysis is the out of the overall_MI_analysis function, with 5 pieces of information :
- phase : an array of the instantaneous phase computed. It's arranged by no difference or difference in the protocol, channel group, frequency band and event.
- amplitude : an array of the amplitude envelope computed. It's arranged by no difference or difference in the protocol, channel group, frequency band and event.
- MI : an array of the modulation index computed. It's arranged by no difference or difference in the protocol, channel group, low frequency band, high frequency band and event. If one MI is equal to NaN, it means that one phase bin is empty, and that your number of bin must be reduced for this analyse.
- sig_array : an array of the significiance of the modulation index computed. It's arranged exactly as MI
- cluster_freq_array : the list of the frequency bands considered in our analysis.

A few things must be kept in mind by the practitioner :

- The modulation index is biased for close frequency bands. When analysing the sig_array, you must ask yourself if the phase-amplitude coupling makes sense between the frequency bands considered. Indeed, phase-amplitude coupling between close frequency bands is absurd.
- The number of bins used in the MI computation must be smaller when analysing short epoch.
- As we compute our analysis with all the possible combinations, there is an intrinsic part (represented by the alpha) of the result array that are false. For example, when analysing the phase-locking with the second method with alpha = 0.05, 5% of the phase-locking assertions are false.
- The heaviest operation on this analysis is the permutation test. If you want to go faster, you will need to take a small number of replications.
- For the phase-locking analysis, be sure that before concatenating the different replications of an experience each one as the same time length.

4 Script construction

In this section, i will present the architecture of the script. The script is divided into two parts : one part is designed to give and manipulate the data, to select the parameters and to run the analysis, another is a library of functions used in the analysis.

Références

[1] Tort A. B., Komorowski R., Eichenbaum H., & Kopell N. (2010). Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *Journal of neurophysiology,* **Vol.104(2)**, *1195–1210*

[2] Hülsemann M.J, Naumann E. & Rasch B. (2019) Quantification of Phase-Amplitude Coupling in Neuronal Oscillations : Comparison of Phase-Locking Value, Mean Vector Length, Modulation Index, and Generalized-Linear-Modeling-Cross-Frequency-Coupling. *Front. Neurosci. https ://doi.org/10.3389/fnins.2019.00573*

[3] Best D.J. and Fisher N.I. (1981). The BIAS of the maximum likelihood estimators of the von mises-fisher concentration parameters. *Communications in Statistics - Simulation and Computation,* **Vol.10**, *493-502*

[4] Mardia K.V, Jupp P.E (1972). Directional Statistics. http ://palaeo.spb.ru/pmlibrary/pmbooks/mardia&jupp_2000.pdf

[5] Efron B. & Tibshirani R.J (1993). An introduction to the bootstrap.