The pairwise phase consistency: A bias-free measure of rhythmic neuronal synchronization

Martin Vinck a,⁎,1, Marijn van Wingerden a,2, Thilo Womelsdorf b,c, Pascal Fries b,d, Cyriel M.A. Pennartz a

⁎ Corresponding author.
E-mail address: M.A.Vinck@uva.nl (M. Vinck).

1 MV conceived the original idea of the pairwise phase consistency and was primarily responsible for theoretical and data analysis.

2 MvW performed the experiments and collected the data.

Article info
Article history:
Received 23 November 2009
Revised 7 January 2010
Accepted 21 January 2010
Available online 28 January 2010

Keywords:
Rhythmic
Synchronization
Oscillation
Phase-locking
Coherence
Pairwise phase consistency

Abstract
Oscillatory activity is a widespread phenomenon in nervous systems and has been implicated in numerous functions. Signals that are generated by two separate neuronal sources often demonstrate a consistent phase-relationship in a particular frequency-band, i.e., they demonstrate rhythmic neuronal synchronization. This consistency is conventionally measured by the PLV (phase-locking value) or the spectral coherence measure. Both statistical measures suffer from significant bias, in that their sample estimates overestimate the population statistics for finite sample sizes. This is a significant problem in the neurosciences where statistical comparisons are often made between conditions with a different number of trials or between neurons with a different number of spikes. We introduce a new circular statistic, the PPC (pairwise phase consistency). We demonstrate that the sample estimate of the PPC is a bias-free and consistent estimator of its corresponding population parameter. We show, both analytically and by means of numerical simulations, that the population statistic of the PPC is equivalent to the population statistic of the squared PLV. The variance and mean squared error of the PPC and PLV are compared. Finally, we demonstrate the practical relevance of the method in actual neuronal data recorded from the orbitofrontal cortex of rats that engage in a two-odor discrimination task. We find a strong increase in rhythmic synchronization of spikes relative to the local field potential (as measured by the PPC) for a wide range of low frequencies (including the theta-band) during the anticipation of sucrose delivery in comparison to the anticipation of quinine delivery.

© 2010 Elsevier Inc. All rights reserved.

Introduction

Oscillatory activity is abundant in nervous systems and has been implicated in numerous functions (Buzsáki and Draguhn, 2004; Fries, 2009; Gray et al., 1989; Pesaran et al., 2002). Two complementary approaches exist for studying oscillations. First, the rhythmic structure of the signals generated by a single source can be examined, for example, by computing the spectral power of EEG (electroencephalogram) or spike signals. Second, one can study the relationship between signals generated by two separate sources, for example, by computing the coherence between neuronal spike output and the LFP (local field potential) or between two EEG or MEG (magnetoencephalography) channels. When the signals generated by two separate sources demonstrate a consistent phase-relationship in a particular frequency-band, then we refer to this property as rhythmic synchronization. Studies have demonstrated task-modulated rhythmic synchronization between local and distant neuronal groups in many different areas and consistent task-modulated phase-relationships of spiking output to the surrounding LFP (Buschman and Miller, 2009; Fries et al., 2008; Fries, 2009; Gregoriou et al., 2009; Lansink et al., 2008; Pesaran et al., 2008; Siapas et al., 2005; Sirotta et al., 2008; Womelsdorf et al., 2007).

The two aspects of rhythmic synchronization that are linked to particular functions are the phase of rhythmic synchronization (Fries et al., 2007; Hopfield, 1995; O’Keefe and Recce, 1993; Vinck et al., 2009) and the strength or precision of rhythmic synchronization (Fries, 2005; Sejnowski and Paulsen, 2006; Singer and Gray, 1995; Varela et al., 2001). The consistency of the phase-relationship is typically quantified on the basis of a vector addition operation, i.e., the computation of the resultant vector length (Fisher, 1993), where the vector represents the relative phase between two signals. Two well-known instances of resultant vector length measures in the experimental neurosciences are the PLV (phase-locking value) (Lachaux et al., 1999) and the coherence measure (Mitra and Pesaran, 1999). Unfortunately, measures such as the PLV and the coherence measure have positively biased estimators for finite sample sizes. This
is often a particular problem to the neurosciences, wherein data are scarce and the number of observations (trials or spikes) is typically not under the control of the experimenter but varies across subjects or neurons.

Since the bias arises because of the vector addition operation, we might avoid the bias by defining a different kind of measure that avoids this operation completely. In this paper, we propose a new measure of the consistency of phase-relationships, namely, the pairwise phase consistency (PPC). The PPC computes the cosine of the absolute angular distance (the vector dot product) for all given pairs of relative phases, i.e., it computes how similar the relative phase observed in one trial is to the relative phase observed in another trial. We demonstrate that the PPC does not suffer from any bias, because it is based on sequential pairs of observations. Importantly, we prove that the population statistic of the PPC is equivalent to the population statistic of the squared PLV. We compare variance and mean squared error of the PPC and PLV and demonstrate the practical usefulness of the PPC for analyzing actual neuronal data.

**Material and methods**

**The problem of bias in measuring rhythmic synchronization**

We are interested in the problem of how to measure whether the signals that are generated by two separate sources demonstrate a consistent phase-relationship in a particular frequency-band. In other words, we want to measure the strength of rhythmic synchronization. In short, this problem will be referred to as how to measure phase consistency. Two of the most widely used measures of phase consistency are the PLV (phase-locking value) and the coherence measure. The PLV and coherence measure have traditionally been based on sequential pairs of observations. Importantly, we prove that the population statistic of the PPC is equivalent to the population statistic of the squared PLV. We compare variance and mean squared error of the PPC and PLV and demonstrate the practical usefulness of the PPC for analyzing actual neuronal data.

Material and methods

The problem of bias in measuring rhythmic synchronization

We are interested in the problem of how to measure whether the signals that are generated by two separate sources demonstrate a consistent phase-relationship in a particular frequency-band. In other words, we want to measure the strength of rhythmic synchronization. In short, this problem will be referred to as how to measure phase consistency. Two of the most widely used measures of phase consistency are the PLV (phase-locking value) and the coherence measure. The PLV and coherence measure have traditionally been based on sequential pairs of observations. Importantly, we prove that the population statistic of the PPC is equivalent to the population statistic of the squared PLV. We compare variance and mean squared error of the PPC and PLV and demonstrate the practical usefulness of the PPC for analyzing actual neuronal data.

**Material and methods**

The problem of bias in measuring rhythmic synchronization

We are interested in the problem of how to measure whether the signals that are generated by two separate sources demonstrate a consistent phase-relationship in a particular frequency-band. In other words, we want to measure the strength of rhythmic synchronization. In short, this problem will be referred to as how to measure phase consistency. Two of the most widely used measures of phase consistency are the PLV (phase-locking value) and the coherence measure. The PLV and coherence measure have traditionally been based on sequential pairs of observations. Importantly, we prove that the population statistic of the PPC is equivalent to the population statistic of the squared PLV. We compare variance and mean squared error of the PPC and PLV and demonstrate the practical usefulness of the PPC for analyzing actual neuronal data.

**Material and methods**

The problem of bias in measuring rhythmic synchronization

We are interested in the problem of how to measure whether the signals that are generated by two separate sources demonstrate a consistent phase-relationship in a particular frequency-band. In other words, we want to measure the strength of rhythmic synchronization. In short, this problem will be referred to as how to measure phase consistency. Two of the most widely used measures of phase consistency are the PLV (phase-locking value) and the coherence measure. The PLV and coherence measure have traditionally been based on sequential pairs of observations. Importantly, we prove that the population statistic of the PPC is equivalent to the population statistic of the squared PLV. We compare variance and mean squared error of the PPC and PLV and demonstrate the practical usefulness of the PPC for analyzing actual neuronal data.
The finite sample size bias can be informally explained by the following example. Suppose that our relative phases are drawn from the uniform circular distribution, which has probability density \( p(\theta) = \frac{1}{2\pi} \). In other words, every phase is equally likely to be observed. If the sample size \( N \) goes to infinity, we would expect that all phases cancel each other out, such that the PLV approaches zero, i.e., \( \Psi = \int_{-\pi}^{\pi} \frac{1}{2\pi} \exp(i\theta) \, d\theta = 0 \). However, the expected circular distance between any pair of relative phases \( \theta_j \) and \( \theta_k \) (where \( j \) and \( k \) are indices for different trials or spikes) is equal to \( \frac{1}{2\pi} \cdot 180^\circ \) (90°). Thus, if we have only two observations, our resultant length is expected to be larger than zero. Precisely, for two observations \( x \) and \( y \), the expected value of our sample PLV will be equal to

\[
E\{\hat{\Psi}\} = \int_{-\pi}^{\pi} \frac{1}{2\pi} \sqrt{\left(1 + \cos^2\theta\right)} \, d\theta = \frac{2}{\pi}.
\]  

(6)

Here, we rotated both phases such that \( \varphi = 0 \) and only the variable \( \chi \) remains, and \( \cos(\chi) = 1 \) and \( \sin(\chi) = 0 \). If we now add a third observation, then the average distance of the new observation to the resultant vector of the first two observations will again be \( \frac{1}{2\pi} \), which will lead to a smaller resultant length. This process goes on until the resultant length reaches the value of zero (because the distribution is uniform).

The finite size sample problem is complicated by the fact that the way \( \hat{\Psi} \) decreases as a function of the sample size depends on the type of circular distribution. Suppose we have as a probability distribution \( P(\frac{1}{2}\pi) = \frac{1}{2} \) and \( P(-\frac{1}{2}\pi) = \frac{1}{2} \). As with the uniform circular distribution, the population PLV (Eq. 4) is zero for this distribution. Suppose that the sample consists of two observations. Either the observed phases are equal to each other, thus the circular distance will be zero, or they are orthogonal to each other, thus they cancel each other out completely. Thus, the expected resultant length for two observations will be \( E\{\hat{\Psi}\} = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0 = \frac{1}{2} \). Suppose that the sample consists of three observations. Then, we have \( (2^3)^3 = 8 \) possible outcomes, wherein two of the eight cases all observed phases are identical, such that \( \Psi = 1 \) for those observations. For all other possible outcomes, we have two identical phases and one orthogonal phase, such that \( \Psi = \frac{1}{2} \). We thus have \( E\{\hat{\Psi}\} = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} = \frac{5}{8} \). Thus, \( \hat{\Psi} \) converges in a non-continuous way to the population PLV (Eq. 4). Fig. 2B compares the dependence of \( \Psi \) on the sample size for this distribution with the uniform distribution. Both distributions have the same population PLV (Eq. 4), but the expected value of the sample estimate of the PLV (Eq. 2) converges in a very different way to the population PLV. Any correction to \( \hat{\Psi} \) thus depends on the assumption of a particular circular distribution.

Reducing the bias by fixing the number of spikes

Recently (Vinck et al., 2009; Womelsdorf et al., 2008), we introduced a technique to reduce the positive bias of the sample PLV and sample coherence estimate (Eqs. 2 and 3). We suggested that we can reduce this bias by using the same number of observations for every sample when comparing samples with a different number of observations. For example, if we have a thousand trials in the first sample and 50 trials in the second, we can reduce the bias by using only 50 observations from the first sample to compute the PLV. To use all available data, a bootstrap without replacement (i.e., every observation can enter one bootstrapped sample only) can be used to reduce the variance of the new estimates. Thus, we introduced the FSPLV (fixed-sample size PLV) estimate as

\[
\hat{\Psi}_F = \frac{1}{F \cdot b} \sum_{b=1}^{B} \left| \sum_{j=1}^{F} \exp(i\theta_{j,b}) \right|.
\]  

(7)

where \( b = (1, 2, ..., B) \), with \( B \) as the number of bootstraps without replacement and \( F \) as the fixed number of spikes, with \( B \leq N \).

We argue that the FSPLV (see also the discussion for a direct comparison with the pairwise phase consistency) still copes with several disadvantages: First, because the bias curves depend on the type of circular distribution, as shown in Fig. 2B, we can only compare the FSPLV between two samples in an unbiased way when they are drawn from the same probability distribution (that can have different parameters). Bimodal or even more complex phase distributions are not uncommon in neuronal data. For example, very strong phase-locking at a higher frequency can cause a bimodal phase distribution at the lower harmonic. Thus, when samples drawn from two different types of distributions are compared, a systematic difference between their samples’ FSPLV may arise, even though the population FSPLV can be the same for the two distributions. Second, one has to choose the arbitrary parameter \( F \), and different choices of this parameter might lead to systematically different conclusions. Moreover, if different parameter values of \( F \) are chosen across different experiments, it will be hard to compare the degree of phase consistency across experiments, because we are essentially using a different population statistic in every experiment. Third, it is not practical to use all available data in the same way, because computing all possible bootstraps is computationally not feasible. For example, if we have 10,000 observations, the number of unique bootstraps of size 50 equals \( \binom{10000}{50} \). That means that some observations might be used more often than other observations and that some combinations of observations are not taken into account. This will lead to a higher variance of our sample estimates and also a variance of the outcomes from run to run.

The pairwise phase consistency

The main problem with the sample estimate of the resultant length (or PLV) is that its finite sample estimator is biased by the number of observations. A measure that does not depend on the number of observations is suggested by looking at pairs of observations instead of all observations together. If phases are consistently clustering around some mean phase, they will on average have a small angular distance to each other.

We first introduce the average pairwise circular distance (APCD), which computes the absolute angular distance between relative phases. The APCD is defined by

\[
\hat{D} = \frac{2}{N(N-1)} \sum_{j=1}^{N-1} \sum_{k=j+1}^{N} d(\theta_j, \theta_k),
\]  

(8)

wherein \( d(\psi, \omega) \) is the absolute angular distance defined as the function

\[
d(\psi, \omega) = |\psi - \omega| \mod \pi
\]  

(9)

and \( \theta_j \) and \( \theta_k \) are the relative phases from two observations \( j \) and \( k \) (or spikes). There are in total \( \binom{N}{2} \) unique pairs that are composed of different observations. In essence, \( D \) computes the average absolute angular distance between all observed relative phases.

We define the population APCD by the Riemann–Stieltjes integral

\[
D = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} d(\psi, \omega) \, dP_{\phi}(\psi) \, dP_{\phi}(\omega).
\]  

(10)

Here, \( P_{\phi}(\psi) = P_{\phi}(\omega) \) is defined as the cumulative probability distribution of the relative phase \( \theta \). If the random variables \( \psi \) and \( \omega \) are absolutely continuously distributed, then \( dP_{\phi}(\psi) \, dP_{\phi}(\omega) = p(\psi) \, d\psi \, p(\omega) \, d\omega \), and Eq. 10 reduces to the standard (Riemann) integral

\[
D = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} p(\psi) \, d\psi \, p(\omega) \, d\omega.
\]  

To obtain a measure which has a similar dynamic range as the PLV and the coherence, we normalize the APCD, giving rise to a new
measure, which we call the pairwise circular distance index (PCDI) and which we will denote by the symbol $D^\ast$.

The PCDI is defined as

$$D^\ast = \frac{\pi - 2D}{\pi} \quad (11)$$

where the superscript $\ast$ in $D^\ast$ indicates normalization. We have $-1 \leq D^\ast \leq 1$, and $0 \leq D^\ast \leq 1$, with a value of 1 indicating complete phase consistency (corresponding to an average distance between pairs of 0) and a value of 0 indicating a complete absence of phase consistency, which is the case for example with the uniform circular distribution or a mixture of two von Mises distributions with an orthogonal mean phase and equal dispersion.

An important question pertains to what the relationship is between the population value of the PLV (or resultant length) and the APCD and PCDI. Obviously, the resultant vector length is a very important circular statistic, because it indicates whether the data have an angular mean direction. The angular distance (Eq. 9) can be alternatively written as $d(\psi, \phi) = \arccos (\cos(\pi) \cos(\alpha) + \sin(\pi) \sin(\phi))$. This leads to the following relationship between the PLV and the APCD:

$$\Psi = \sqrt{\left(\int_{-\pi}^{\pi} \cos(\theta)dP_{w}(\theta)\right)^2 + \left(\int_{-\pi}^{\pi} \sin(\theta)dP_{w}(\theta)\right)^2}.$$  

(12)

Hence we have the squared PLV as

$$\Psi^2 = \left(\int_{-\pi}^{\pi} \cos(\theta)dP_{w}(\theta)\right)^2 + \left(\int_{-\pi}^{\pi} \sin(\theta)dP_{w}(\theta)\right)^2 = \int_{-\pi}^{\pi} \cos(\theta)dP_{w}(\phi) + \sin(\theta)dP_{w}(\phi) \cos(\phi).$$  

(13)

The population statistic of the APCD (Eq. 16) is equivalent to the population statistic of the squared PLV, expect for an arc cosine transformation of the dot product inside the integral. The relationship between the APCD and the PLV implies that we can define a pairwise statistical measure, which gives an unbiased estimate of the population statistic of the squared PLV. We call this measure the pairwise phase consistency (PPC) and denote it by the symbol $T$. We have the sample estimate of the PPC as

$$\hat{T} = \frac{2}{N(N-1)} \sum_{j=1}^{N-1} \sum_{k=j+1}^{N} f(\theta_j, \theta_k).$$  

(14)

wherein the function $f$ computes the dot product between two unit vectors and is defined by

$$f(\varphi, \psi) = \cos(\varphi) \cos(\psi) + \sin(\varphi) \sin(\psi).$$  

(15)

The population statistic of the PPC is given by $T = \Psi^2$ (Eq. 13).

The sample estimate is bound by $-1 \leq \hat{T} \leq 1$. The PPC has an essential advantage in comparison to the PCDI: The expected value of its estimator is equal to the population statistic of the squared value of the resultant length (PLV). The relationship between the APCD and the PPC is given by taking the cosine of the absolute angular distances,

$$T = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \cos(d(\phi, \theta))dP_{w}(\phi)dP_{w}(\theta).$$  

(16)

The PPC is graphically explained in Fig. 1.

The key difference between the APCD and the PPC is that for the computation of the PPC, we determine the absolute angular distances between relative phases, whereas for the computation of the PCDI, we determine the cosine of the absolute angular distances. For angular differences close to $90^\circ (\frac{\pi}{2})$, the angular distance and the cosine of the angular distances are approximately linearly related, since for these angles we have $\cos(x) \approx \frac{1}{2} \pi - x$. However, especially for small distances and for very large angular distances, the functions $d(\phi, \theta)$ (Eq. 9) and $f(\varphi, \psi)$ (Eq. 15) behave differently, since the derivative of the cosine of the angular distances is zero for these angular distances.

There exist distributions for which the PLV is zero, but the PCDI is not. In other words, for these distributions there is no circular mean direction, but the mean angular distance between relative phases is smaller than $90^\circ$. For example, suppose we have a circular distribution with probabilities defined by $P(\theta = \frac{1}{2} \pi) = \frac{1}{\sqrt{2}}$ and $P(\theta = 0) = \frac{\sqrt{2}}{2}$ for this distribution is zero, but the PCDI is not.

The PCDI is given by

$$D = \frac{1}{(2 + \sqrt{2})} \left( \frac{1}{2^\pi} + \frac{1}{2^\pi} \right).$$  

(17)

On the contrary, the PPC is given by

$$T = \frac{2 - 2 \cdot 1}{\left(2 + \sqrt{2}\right)^2} + \frac{2 - 2 \cdot 1}{\left(2 + \sqrt{2}\right)^2} + \frac{2 - 2 \cdot 1}{\left(2 + \sqrt{2}\right)^2} = 0.$$  

(18)

Bias and variance of the sample PPC estimator

In this paragraph, we will show that the sample estimator of the PPC is consistent and not biased by the sample size. The expected value of the sample PPC equals

$$E\{\hat{T}\} = \left( \frac{2}{N(N-1)} \sum_{j=1}^{N-1} \sum_{k=j+1}^{N} d(\theta_j, \theta_k) \right) = \frac{2}{N(N-1)} \sum_{j=1}^{N-1} \sum_{k=j+1}^{N} E\{d(\theta_j, \theta_k)\} = \hat{T}.$$  

(19)
Here, we used the property $E[f(X_1, X_2) + f(X_2, X_3)] = E[f(X_1, X_2)] + E[f(X_2, X_3)]$ and the assumption that $dP(\theta, \theta_k) = dP(\theta) dP(\theta_k)$, in other words that our observations are independent. Eq. 19 shows that the expected value of the sample PPC does not depend on the number of pairs in our sample. The pairs are of course not independent, but this does not affect the expected value, but the variance of our sample estimate. The variance of our sample PPC estimate is given by

$$\text{Var}(\hat{\Upsilon}) = E\{\hat{\Upsilon}^2\} - \Upsilon^2$$

(20)

We have

$$E\{\hat{\Upsilon}^2\} = E\left\{\frac{4}{N^2(N-1)^2} \left(\sum_{j=1}^{N-1} \sum_{k=(j+1)}^{N} d(\theta_j, \theta_k)\right)^2\right\}$$

(21)

We can decompose $E\{\hat{\Upsilon}^2\}$ in three terms: The product of the distance for pairs that are identical, the product of the distance for pairs that are independent and the product for pairs in which there is one shared observation, such that the total number $M$ of product of pairs (for $N>3$) equals

$$M = \binom{N}{2} \binom{N}{2} \frac{(N-2)(N-3)}{2} + \binom{N}{2}(2N-4)$$

$$= \binom{N}{2}^2$$

(22)

The variance of the PPC estimate is thus given by

$$\text{Var}(\hat{\Upsilon}) = \frac{2}{N(N-1)} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(\theta, \theta)^2 dP_\theta dP_\theta$$

$$+ \frac{6-4N}{N(N-1)} \Upsilon^2$$

$$+ \frac{4N-8}{N(N-1)} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(\theta, \theta)^2 f(\varphi, \varphi) dP(\theta) dP(\theta) dP(\varphi) dP(\varphi)$$

(23)

and the function $f$ is defined by Eq. 9. This means that the variance of the PPC estimate approaches zero asymptotically,

$$\lim_{N \to \infty} \text{Var}(\hat{\Upsilon}) = 0.$$

(24)

The asymptote of our estimate $\hat{\Upsilon}$ is thus equal to

$$\lim_{N \to \infty} \hat{\Upsilon} = \Upsilon.$$

(25)

This demonstrates that $\hat{\Upsilon}$ is a consistent and unbiased estimator of $\Upsilon$. Note that Eqs. 19–25 applies to the PCDI as well.

Simulations based on artificial data

To compare the variance and mean squared error between the squared PLV and the PPC and to validate our analytical findings by means of numerical simulations, we created artificial samples drawn from a von Mises distribution, according to (Best and Fisher, 1979). For every sample drawn, we computed the PLV, the squared PLV, the PCDI and the PPC. We repeated this procedure 10^6 times to obtain estimates of the variance and the mean of the PLV, the squared PLV, the PCDI and the PPC. We repeated this procedure for a large range of the dispersion parameter $K$ from the von Mises distribution. The population PLV and squared PLV values are respectively given by

$$\Psi = \frac{l_1(K)}{l_0(K)}$$

(26)

and $\Psi^2$, where $l_1(K)$ and $l_0(K)$ are modified Bessel functions of order zero and one, and $K$ is the dispersion parameter from the von Mises distribution (Fisher, 1993). The MSE of the sample estimates is defined as

$$\text{MSE} = \frac{1}{M} \sum_{m=1}^{M} (\hat{S}_m - S)^2.$$

(27)

wherein $m$ is the index for sample number, $M$ is the total number of samples that we draw from the population, and $S$ is the sample estimate of the population statistic $S$ (either PLV or PPC). All simulations were performed in C using the gcc compiler.

Actual spike–LFP data

Three adult male Wistar rats (Harlan CPB, Horst, the Netherlands), weighing 370–450 g at the time of surgery, served as subjects in these experiments. Prior to training and surgery, the rats were housed two to a cage on a reversed light/dark cycle (lights off: 7AM, lights on: 7PM) with ad libitum food and water. After surgery, animals were housed individually in a transparent cage (40×40×40 cm), with other rats present in the climate controlled colony room. During training and recording, rats were maintained on food restriction, with 5–15 g of food available from >1 h after training, depending on the amount of reward collected in the session, amounting to 90% free-feeding intake. All experiments were conducted according to the National Guidelines on Animal Experiments and with approval of the Animal Experimentation Committee of the University of Amsterdam.

Odor discrimination training was conducted in an operant chamber (56×30×40 cm L W H). The chamber was equipped with an odor sampling port and trial light on a front panel and a tray for delivery of fluids placed at the opposing wall. The front panel was slanted at 45° (with respect to ground level) above the odor sampling port to allow unhindered nose poking into the odor port by implanted animals. Entries by the animals into the odor port, fluid well and licking into the well were recorded by photobeam interruptions. During recording sessions, behavioral events were synchronized with electrophysiological data acquisition running on a separate computer. Odors were delivered via separate glass vials and tubing to avoid mixing. Upon entering the application system, they were mixed 1:1 with clean air and released into the compartment directly behind the odor port by way of computer-controlled valves. Likewise, quinine and sucrose solutions were delivered to the fluid well via separate fluid lines and electronically controlled by valves. Training and task performance were devoid of human interference. The animals were trained on a two-odor go/no-go discrimination task. After habituation to the chamber and pre-training, rats were confronted with odor discrimination problems. Each session, one novel odor was associated with reward (100 μl of 15% sucrose solution), and a second novel odor with an aversive outcome (100 μl of 0.01 M quinine solution). Training sessions consisted of blocks of 5 + 5 pseudo-randomly ordered positive and negative odor trials. When a trial light was illuminated, rats could initiate a trial by poking their snout in the odor sampling port. After a 500-ms delay, air flow through the odor sampling port was switched from clean air to the selected clean air/odorant mixture. A correct nose poke in the odor port (wait for odor for 500 ms, constituting a pre-stimulus delay, and sample odor for at least 750 ms) was indicated by the trial light turning off. After sampling, the rats could move over to the fluid well, into which they were required to make a nose poke for >100 ms before the outcome (sucrose or quinine solution) was presented. We refer to the invariant 1000-ms delay period as the anticipatory period. This anticipation period allows sampling of neural activity during the waiting period that is not confounded by whole-body movement. When the rats left the fluid well, an inter-trial interval (ITI) of 10–15 s was observed before the
next trial started. A correct rejection was scored if the rat refrained
from entering the reinforcement tray for $>5\,$ s following sampling of
the negative odor. Responses during the ITI had no programmed
consequences, while prematurely ended responses (i.e., short pokes)
during the odor sampling or waiting period resulted in immediate
termination of the current trial and the start of a new trial.

Rats were implanted after they reached behavioral criterion, viz.
scoring $>85\%$ hits and correct rejections over a moving block of 20
trials. Animals were anesthetized by i.m. injection of 0.8 ml/100 g
Hypnorm ($0.2\,$ mg/ml fentanyl, $10\,$ mg/ml fluanisone; VetaPharma Ltd.,
Leeds, UK), followed by $0.04\,$ ml/100 g Dormicum ($5\,$ mg/ml
midazolam; Roche Nederland B.V., Woerden, the Netherlands) s.c.
and mounted in a stereotact. Body temperature was maintained
between 35 and $36\,$ °C. A microdrive, holding 14 individually
moveable electrode drivers, was chronically implanted onto a
crianotome (diameter: $2\,$ mm) in the left hemisphere dorsal to the
OFC at 3.4–3.6 mm anterior and 3.0–3.2 mm lateral to bregma. The
drivers were loaded with 12 tetrodes and 2 reference electrodes.

Using dental cement, the drive was anchored to six stainless steel
screws, one of which was positioned in the left parietal bone and
served as ground. Immediately after surgery, all tetrodes and
reference electrodes were advanced 0.8 mm into the brain. Next,
the animal was allowed to recover for 7 days with ad libitum food
and water, during which the 12 recording tetrodes were advanced in
daily steps to the upper border of the OFC according to a standardization
rat brain atlas ( Paxinos and Watson, 1982). The reference electrodes
were lowered to a depth of $1.2–2.0\,$ mm and adjusted to minimize
spiking activity on the reference channel. After surgery, saline was
injected s.c. ($2\,$ ml per flank), and pain relief was provided by 0.1 ml/
$100\,$ g pre-surgical weight of a 10% Finadyne (flunixin meglumine
50 mg/ml; Schering-Plough, Brussels, Belgium) solution adminis-
terated in saline s.c.

Following surgery, animals were retrained on the familiar odor
pair with which initial training took place until performance was back
at criterion level. Recording of neural activity started in the
subsequent session. On each recording session, rats were confronted
with a new odor pair. Using tetrodes, neural activity was recorded by a
64-channel Cheetah setup (Neuralynx, Bozeman MT). Signals were
passed through a unity-gain pre-amplifier headstage, a 72-channel
commutator (Dragonfly, Ridgeley, West Virginia, USA), amplified
5000 and bandpass filtered between 600 and 6000 Hz for spike
recordings. If a signal on any of the leads of a tetrode crossed a pre-set
threshold, activity on all four leads was sampled at 32 kHz for $1\,$ ms
and stored for off-line analysis. Local field potentials recorded on all
tetrodes were amplified 10000, continuously sampled at $1874\,$ Hz,
and bandpass filtered between 1 and $475\,$ Hz. Events in the behavioral task
were co-registered and time-stamped by the Cheetah system. Spike
trains were sorted to isolate single units using a semi-automated
clustering algorithm ( BubbleClust) followed by manual refinement
using MCList. Automated and manual clustering of spikes was done
using the waveform peak amplitude, area, squared amplitude integral
and the first three principal components. Clusters were accepted as
single units when having no more than $0.1\%$ of inter-spike intervals
shorter than $2\,$ ms.

All data analysis was performed in MATLAB, with additional use of
the Fieldtrip Toolbox and custom-made Mex-files, using the gcc
compiler. For every neuron, we computed the spike–LFP phase for
every spike by Fast Fourier Transforming an LFP segment of length $T$
samples,

$$X_j(f) = \sum_{t=1}^{T} w(t) x_j(t) \exp(2 \pi i n t f ) . \tag{28}$$

Here, $w(t)$ is a Hanning window, $x_j(t)$ is the unfiltered LFP
segment that is centered around the spike, $f$ is the frequency, and $j$ is
an index for spike number.

Results

Simulations based on artificial data

We first tested whether the PPC (pairwise phase consistency)
indeed has a bias-free sample estimator by computing the PPC for
samples drawn from a von Mises distribution with different sample
sizes and levels of dispersion. Fig. 3A shows that the PPC does not
have any bias a function of the sample size, as we predicted
analytically by Eq. 19. As predicted from Eqs. 13 and 16, we observed
identical population values for the PPC and the squared PLV (Fig. 3B).
In addition, for samples drawn from a von Mises distribution,
we obtained a near linear relationship between the population squared
PLV (phase-locking value) and the population PCDI (pairwise circular
distance index) for a large range of the population PLV (PLV $< 0.9$
(Fig. 3B). We can approximate the population squared PLV with great
accuracy by $\Psi^2 = \beta D^2$ wherein $\beta = 1.21$ and $\Psi$ and $D^2$
are defined respectively by Eq. 2 and Eq. 11.

Typically, a reduction in bias for some estimator comes at the
expense of an increase in variance and MSE (mean squared error).
Thus, we were interested in comparing the MSE and variance of the
PPC with the squared PLV. For a given value of the von Mises
dispersion parameter $\kappa$, the population statistics of both statistics are
known and derived from Eq. 26. Fig. 3C compares the log ratio of
the MSE of the PPC to the MSE of the squared PLV. A value of zero
indicates that the PPC is the same for both statistics. Values lower than zero
indicate that MSE is smaller for the PPC than for squared PLV. For small
values of the population PLV ($0-0.3$), where the bias and variance
problem is most severe, the MSE of the PPC compares favorably to the
MSE of the squared PLV. For higher values of the PLV ($0.4–1$), the MSE
of the PPC is only slightly greater than the MSE of squared PLV. Fig. 3D
displays the difference in variance of the squared PLV with the
variance of the PPC. A value of zero indicates that the variance is equal,
whereas a higher value than zero indicates that the variance of the
PPC is relatively higher. For small samples ($N<10$), the PPC estimate
suffers from higher variance than the squared PLV estimate, showing
that there is some trade-off between bias and variance. It should be
noted however, that for small sample sizes (e.g., $N=2$), the PPC has a
dynamic range that still extends from 0 to 1, whereas the squared PLV
has a much smaller dynamic range (the lower bound is pushed
towards 1 because of the bias). Given a fixed variance, we are less
sensitive in detecting differences in the population statistic if the
dynamic range of the sample estimate is smaller.

Analysis of neuronal data

We first verified that our technique indeed provides unbiased
results for actual neuronal data. For that purpose, we selected cells
from which we recorded more than 200 spikes during the period in
which the rat anticipated sucrose delivery at the fluid well, in order to
obtain a relatively smooth circular distribution of spike theta ($6\,$ Hz)
phases. The sample of all recorded spikes now functioned as the
population probability distribution. Then, for every selected neuron,
we drew small-size samples of spike phases from the distribution of
all recorded spike phases. Samples were drawn without replacement.
Draws were repeated 50,000 times to obtain an estimate of the
effective value of the sample estimates. Fig. 4A shows that there was a
negative relationship between the expected value of the PLV
estimates and the sample size, similar to the results of the numerical
simulations shown in Fig. 2A. As predicted (Fig. 3A and Eq. 19),
no systematic bias was observed when using the PPC (Fig. 4B). This
means that for the expected value of our PPC sample estimate, it
would not have made a difference whether we observed 200 or 10
spikes from a given neuron (of course, for the variance it would).

Next, we computed the PLV and PPC for every neuron using all
available spikes and compared the PLV (Fig. 4C) and PPC (Fig. 4D)
across the neuronal population. For cells with small sample sizes (few spikes), we observed much higher PLVs and a higher variance. Figs. 4C and D demonstrate an important downside of the FS-PLV when applied to spike data, namely: The number of spikes we choose to compute the FS-PLV will lead to a selection of only a subset of neurons that have a higher number of spikes (and likely higher firing rates) than on average, for example, interneurons or neurons that increase their firing rate during the period under consideration. Conclusions drawn from analyses on this subset of neurons might not generalize to the population of all cells, especially since neurons with a large number of spikes might have different firing rate correlates.

As a demonstration, we then examined the relationship between task conditions (rat behavior) and the PPC spectra. We investigated the structure of PPC spectra for different inclusion criteria for our cells (i.e., the minimal number of spikes that was required to include a cell). Figs. 5A–D show the PPC spectra for the sucrose waiting (positive outcome; 1000 ms before fluid delivery) and quinine waiting (negative outcome; 1000 ms before fluid delivery) task periods. During this period, orbitofrontal cortex neurons are selective by means of their firing rate for the upcoming task outcome, based on a cue that was received a few seconds before (one of two odors). Every plot represents a different number of minimal spikes that had to be recorded from a neuron before we included it (2, 10, 30, 50). As Fig. 4D shows, when the minimal number of spikes is small, neurons are included whose individual sample estimates have a high variance (especially when they are weakly locked), which might obscure differences between task conditions. Figs. 5A–D show that theta-band spike–LFP phase-locking is enhanced during the sucrose waiting period relative to other task periods. Even when all cells are included, regardless of their sample size, we obtain similar results, showing that the difference in theta-band rhythmic synchronization between the sucrose and the quinine waiting period holds at the level of the complete sample of recorded cells, not only for the sample of cells with a relative high firing rate. Second, we show that across a large range of frequencies (2 to 10–15 Hz), rhythmic synchronization is selectively enhanced during the sucrose waiting period versus the quinine waiting period.

Discussion

We investigated the problem of how to measure phase consistency, i.e., whether signals that are generated by two separate sources display a consistent phase-relationship in a particular frequency-band. The sample estimators of the PLV (phase-locking value) and the coherence measure are strongly biased by the number of observations (trials or spikes) in the sample. This bias is difficult to control, because the way the bias behaves depends on the circular distribution (for example, unimodal versus bimodal). We have introduced the pairwise phase consistency (PPC) as a new measure of phase consistency. We demonstrated analytically, by means of simulations and by application on actual neuronal data, that the sample estimate of the PPC is an unbiased and consistent estimator of the population PPC. These properties, unbiasedness and consistency, hold independently of the circular distribution of the relative phases. We demonstrated, both analytically and by means of simulation, that the population statistic of the PPC is identical to the population statistic of the squared PLV. This means that we can interpret the PPC both in terms of (the cosine of) angular distances and in terms of the population value of the squared PLV. Furthermore, we showed that the MSE (mean square error) of the PPC estimate compared favorably to that of the squared PLV for low values of the PLV (0–0.3), where the problems of variance and bias are most severe. Finally, we demonstrated the use of the PPC in actual neuronal data. Thus, the PPC and the squared PLV have the same population statistic, but the key advantage of the PPC is that it has a bias-free sample estimator. In addition, we introduced the pairwise circular distance index (PCDI), which computes the absolute angular distances between relative phases. Although the PCDI has an unbiased and consistent estimator, it is more difficult to interpret the PCDI than it is to interpret the PPC. Although for a large range of dispersion of relative phases (R < 0.9), we demonstrated by numerical simulations that there is a linear relationship between the PCDI and the population value of the squared PLV for a von Mises distribution, the relationship between the PCDI and the population value of the (squared) PLV remains to be investigated for other probability distributions, for example, probability distributions that are not unimodal. For a uniform distribution, the PCDI and PLV are both zero, but we have shown that for example for a trimodal probability distribution, the PLV can be zero, while the PCDI is not.
The PPC has several advantages to the coherence and PLV in several respects. A main advantage is that the PPC’s sample estimate has no bias. Experimentally, this is especially an advantage when the sample consists of spike phases from sparsely firing neurons, or when the number of behavioral trials depends on the response of the behavioral subject and not on the experimenter. The unbiasedness gives several advantages. First, in neuroscience we often average results across subjects or neurons, because individual measurements are often relatively noisy. When there is a systematic bias across subjects or neurons, then this bias will appear in the group average as well. However, when the error of the sample estimates (relative to the population statistic) is only determined by the variance of the sample estimates, then we can squeeze this variance out of our group average, since it is expected to be largely independent across subjects or neurons. Unbiasedness ensures that our data will not have a systematic tendency for false positives. At the same time, unbiasedness increases the sensitivity of our statistical comparisons. A difference in sample size, which causes the PLV to go in the opposite direction of a hypothesized effect, will not affect the PPC. Second, unbiasedness facilitates the comparison of outcomes across different experiments. Such a comparison is currently complicated because of the sample-size bias. Depending on the behavioral task and recording area, large differences in the number of recorded spikes per neuron or behavioral trials can exist. The comparison of the strength of rhythmic synchronization across areas and stimulus and behavioral condition takes place mostly across experiments, and the sample size might be very different between experiments. Thus, we expect that the use of the PPC can give rise to interesting meta-analyses across experiments. Third, some research questions are explicitly concerned with the relationship between firing rates and rhythmic synchronization: a bias-free measure of rhythmic synchronization is very important in these cases. Another main advantage of the PPC is that the PPC has a smaller MSE (mean squared error) than the PLV. This means that the reduction in bias does not come at the expense of an even larger variance. Despite these advantages of the PPC, it should be emphasized that the PLV (and the coherence) remain important descriptive measures of the data. Although the PPC provides an unbiased estimate of the population value of the squared PLV, it does not for the PLV itself. The current paper did not investigate under what conditions the PLV can be reliably estimated from the square root of the PPC. Since the sample estimator of the PPC can attain negative values, this is at least impossible for combinations of lower values of the (population) PLV and PPC and small sample sizes.

As already argued in the methods, the PPC also compares favorably to the FSPLV. First, the FSPLV is not an unbiased measure, in the sense that different types of circular distributions (e.g., unimodal vs.
bimodal) can have a different dependence of the sample FSPLV on sample size. Second, to use the FSPLV, an arbitrary value has to be chosen (the fixed number of observations we use per sample). Different choices of this value might result in different conclusions. Moreover, if we choose different parameter values of $F$ across different experiments, it will be hard to compare the degree of phase consistency across experiments, because we are essentially using a different population statistic in every experiment. Finally, some subjects or neurons might not have enough trials or spikes to compute the FSPLV, so that results might not generalize to the whole population. For the PPC, all subjects and neurons can contribute to the group average. Third, the PPC uses all available data, whereas for the FSPLV, it is typically not practical to use all available data in the same way, because computing all possible bootstraps is computationally often not feasible. That means that some observations might appear more often in the bootstraps than other samples and that some combinations of observations are not taken into account. This will lead to a higher variance of our sample estimates and also a variance of the outcomes from run to run. The PPC leads to exactly the same outcome every time we compute it.

The main disadvantage of the PPC is that it is computationally inefficient (order $N^2$ larger) in comparison to the PLV and the coherence. However, $N^2$ computations are still practically feasible, which is demonstrated by the fact that we could easily compute the PPC across 525 neurons. Another potential disadvantage is that negative values are possible when using the normalized version of the PPC (whose expected value runs from 0 to 1). This is a consequence of the unbiasedness of the PPC. In general, if a statistic is bounded from below by zero, then its expected value can only be zero then the variance is zero. On the contrary, the PLV estimate has the highest variance when its population value is zero. Percentages of change can only be computed when all values are larger than zero, so we might not be able to compute percentages in all cases. On the other hand, the percentage of change that can be obtained (when the values are positive) will not be influenced any more by the sample size (thus it will be larger). For the PLV and coherence, we will typically obtain smaller effect sizes, because small sample sizes effectively decrease the dynamic range of the statistic. Often, it is not even clear what the dynamic range of reported PLV or coherence values is. The PPC has the advantage that it has the same dynamic range, regardless of the

---

**Fig. 4.** Application of pairwise phase consistency (PPC) to neuronal data. (A) Bias of PLV (phase-locking value) as a function of spike count for individual cells. Theta-band spike–LFP synchronization was investigated while rats anticipated reward at a fluid well. For neurons with more than 200 spikes during reward anticipation period, we drew samples of different sizes from the distribution of all recorded spike theta (6 Hz) phases. 10 PLV showed a strong bias as a function of sample size. Every line corresponds to a single unit. Color scale indicates different levels of phase consistency. (B) Bias of PPC as a function of number of spikes for individual cells. Same conventions as in (A), except that here the PPC is plotted. PPC showed no bias as a function of sample size. (C) Distribution of squared PLVs across neuronal population. Each dot represents a neuron. All recorded spikes were used to calculate the squared PLV. Neurons with few recorded spikes demonstrated higher squared PLV. (D) Distribution of PPC across neuronal population. Same conventions as in (C), except that here the PPC is plotted. The strong bias of the PLV (C) disappeared for the PPC.

---

120 M. Vinck et al. / NeuroImage 51 (2010) 112–122
number of samples we take. As a rather philosophical remark, we would like to point out that a percentage increase in the degree of phase consistency (e.g., a 100% change from 0.05 to 0.1) can as well be interpreted as a much smaller decrease in the amount of randomness of our spikes if we had let our scale of randomness run from 0 to 1 instead of 1 to 0 (e.g., a 5% decrease from 0.95 to 0.9).

Finally, we demonstrated the use of the PPC in neuronal data. Spiking activity from orbitofrontal neurons was recorded from well isolated sparsely firing neurons, while the rat engaged in a two-odour discrimination task. We demonstrated that the PLV was influenced by the firing rate, while the PPC was not. Furthermore, we demonstrated the use of the PPC by analyzing group differences in spike–LFP rhythmic synchronization during different task conditions. We showed that for a large range of low frequencies, rhythmic synchronization was selectively enhanced during anticipation of sucrose delivery versus anticipation of quinine delivery.

To conclude, we provide a measure of rhythmic synchronization, which is easily computed, has an unbiased and consistent estimator, an intuitive interpretation and is straightforwardly related to existing measures of rhythmic synchronization. We expect that, for many experimental questions, our method will provide increased sensitivity for true positives and a decreased sensitivity for false negatives in comparison to existing methods. The pairwise phase consistency has great applicability when we want to measure rhythmic synchronization for both EEG–EEG, MEG–MEG, spike–LFP, and spike–spike pairs but might very well be applicable in the biomedical sciences in general and other sciences (e.g., physics or geosciences) in which circular statistics play an important role.

Acknowledgments

We thank Prof. A.J. Vinck, Dr. Jan Lankelma, Dr. Francesco Battaglia and Dr. Conrado Bosman for helpful comments. Furthermore, we would like to acknowledge the developers of Fieldtrip (fieldtrip.fcdonders.nl), in particular Dr. Robert Oostenveld; Peter Lipa (University of Arizona, Tucson AZ) for the use of BubbleClust; and A. David Redish (University of Minnesota, Minneapolis MN) for the use of MClast.

References


