

UNCOVERING SHORT-TIME CORRELATIONS BETWEEN MULTICHANNEL RECORDINGS OF BRAIN ACTIVITY: A PHASE-SPACE APPROACH

R. FDEZ GALÁN*

*Department of Biological Sciences, 4400 Fifth Avenue
Mellon Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

R. RITZ and A. V. M. HERZ

**Institute for Theoretical Biology, Humboldt Universität Zu Berlin
Invalidenstr. 43, D-10115 Berlin, Germany*

P. SZYSZKA

*Institute for Neurobiology, Freie Universität Berlin
Königin-Luise-Str. 28/30, D-14195 Berlin, Germany*

Short-time correlations in multivariate time series are notoriously difficult to detect. Extending the classical concept of correlation, we present a method to tackle this problem with little computational cost. In essence, the method uncovers multidimensional phase-locking and is especially useful for revealing changes of the dynamical coupling between different brain areas. The approach also permits us to estimate the shortest time-window in which the coupling occurs. Furthermore, the coupling can be quantified by a measure that defines a metric space. Therefore it may be used to identify task-dependent coupling hierarchies between two brain areas. We illustrate the analysis technique by studying the dynamical coupling between two brain regions involved in olfactory processing in the honeybee. We show that the neural activity in both areas is coupled with each other and that the coupling increases significantly during odor processing.¹

Keywords: Short-time correlation; multivariate time-series; phase-space; multidimensional phase-locking; embedding dimension; odor processing.

1. Introduction

Several experimental approaches in brain research, such as electroencephalography, magnetoencephalography, functional magnetic resonance, calcium-imaging and multielectrode recording provide multivariate time series from which the extraction of correlations between recording channels is of prime importance to understand how different brain regions interact. It is also of great interest to study whether these interactions change

depending on the performed task. Many of such interactions emerge presumably over a short temporal window in which neither the probability density function (p.d.f.) $p(\mathbf{X})$ of the original high-dimensional data $\mathbf{X}(t)$ nor its moments can be properly estimated from the small set of available multivariate observations $\mathbf{X}(t) = (x_1(t), x_2(t), \dots, x_n(t))$, where n is the number of channels. The limited size of a single data set must then be compensated by pooling data from several

¹This paper is based on results presented at the *1st European Interdisciplinary School on Nonlinear Dynamics for System and Signal Analysis (Euroattractor2000)* in Warsaw, 6–15 June, 2000, [Galán *et al.*, 2002] and at the workshop on *Analysis and Modelling of Event-Related Brain Potentials: Cognitive and Neural Approaches* in Potsdam, 29 November–1 December, 2001.

trials under the same experimental conditions. The shorter the correlation to be uncovered, the larger the number of trials. However, because of experimental constraints, the number of trials is often less than required and therefore short-time correlations remain unobserved. To overcome this problem, we propose a method that reduces finite-sample effects and at the same time generalizes the concept of correlation.

Consider the case of two signals $x_1(t)$ and $x_2(t)$. Our method does not rely on estimating the joint probability $p(x_1, x_2)$ or its moments. Instead we estimate the p.d.f. of the scalar $\phi(t)$, which is the instantaneous angular distance between the two signals in phase space. Observe that for the same amount of observations N , the p.d.f. of a scalar can be better estimated than the p.d.f. of a two-element variable because the N realizations of the scalar are spread along one dimension whereas the N realizations of the two-element variable are spread over two dimensions.

We demonstrate that the angle $\phi(t)$ and the estimator of its p.d.f. carry useful information about the interdependence of the signals, even over short-time windows. We provide a metric to quantify such interdependence and call it "dynamical coupling". The coupling can be understood as a measure of stochastic synchronization or phase-locking in the phase space where the signals are embedded. Before we present our method, we briefly review standard ways to estimate statistical interdependence between signals.

1.1. Methods based on the probability density function or its moments

A first step to quantify the degree of similarity between two signals $x_i(t)$ and $x_j(t)$ from the observed values $\mathbf{x}_i = (x_i(1), x_i(2), \dots, x_i(N))$ and $\mathbf{x}_j = (x_j(1), x_j(2), \dots, x_j(N))$ is to calculate the correlation coefficient

$$\text{corr}(\mathbf{x}_i, \mathbf{x}_j) = \frac{(\mathbf{x}_i - \langle \mathbf{x}_i \rangle) \cdot (\mathbf{x}_j - \langle \mathbf{x}_j \rangle)}{|\mathbf{x}_i - \langle \mathbf{x}_i \rangle| \cdot |\mathbf{x}_j - \langle \mathbf{x}_j \rangle|}, \quad (1)$$

i.e. the scalar product between both channels after removing mean values and normalizing to unity. For n channels one can analogously calculate the correlation matrix, whose entry in the i th row and j th column corresponds to $\text{corr}(\mathbf{x}_i, \mathbf{x}_j)$. If the signals are not normalized to unity after removing the mean value, one obtains the covariance matrix instead of

the correlation matrix. The covariance matrix is an estimator of the second-order central moment of \mathbf{X} . The n th moment μ_n of \mathbf{X} is calculated as

$$\mu_n^{i,j,k,\dots,p} = E(x_i \cdot x_j \cdot x_k \cdot \dots \cdot x_p) \quad \underbrace{i, j, k, \dots, p}_{n \text{ indices}} \quad (2)$$

where $E(x)$ stands for the expectation value of x . After removing the mean values of each channel x_i , formula (2) gives the n th central moment. If the signals are generated by a linear system or if the signals can be expressed as a multivariate autoregressive moving average (ARMA) process with normally distributed residuals [Priestley, 1996; Honerkamp, 1993], then the second-order statistics of $p(\mathbf{X})$, i.e. the covariance matrix, suffices to describe the relative dependence of the signals. However, as soon as the signals are nonlinear or are nonlinearly coupled, higher-order moments of the p.d.f. have to be taken into account. In those cases, the higher-order moments also contain information about the detailed statistical dependence between the signals, which cannot be uncovered by the covariance matrix.

For a finite number N of observations, the estimation of higher moments deteriorates quickly with increasing order of the moment. For example, if the data are bounded between the values $-L/2$ and $L/2$ and the discretization size is 1, the first-order moment has to be estimated with N realizations out of L possibilities, the second-order moment with N realizations out of L^2 and the n th moment with N realizations out of L^n . A way to partially balance this effect is to pool data from several trials. But since the number of trials is usually limited because of technical constraints, the search for short-time correlations through higher-order moments is often impractical as well.

Instead of computing cross moments of successively higher order, one can quantify statistical dependences directly from the p.d.f. of each channel by calculating magnitudes like the mutual information [Shannon, 1948] or the Kullback–Leibler divergence [Kullback, 1959; Kullback & Leibler, 1951]. For two channels x_1 and x_2 the mutual information M takes the form

$$M = \int_{-\infty}^{\infty} p(x_1, x_2) \cdot \log \left[\frac{p(x_1, x_2)}{p(x_1)p(x_2)} \right] dx_1 dx_2.$$

The Kullback–Leibler divergence K is a measure of the deviation of a distribution $p(\mathbf{X})$ from a null-hypothesis distribution $p_0(\mathbf{X})$. For two observation

channels x_1 and x_2 , K is given by

$$K = \int_{-\infty}^{\infty} p(x_1, x_2) \cdot \log \left[\frac{p(x_1, x_2)}{p_0(x_1, x_2)} \right] dx_1 dx_2.$$

Observe that K reduces to M under the null-hypothesis that both channels are independent and therefore $p_0(x_1, x_2) = p(x_1) \cdot p(x_2)$. The use of these information theoretical quantities convey several practical problems that make them useless to detect short-time correlations. The main difficulty is to estimate the p.d.f., for which a large number of data points is needed. Even if many observations are available, the estimation of the mutual information and the Kullback–Leibler divergence depends on the discretization size ε of the bins used to estimate the probability density function.

Alternatively to the estimation of $p(\mathbf{X})$ or its moments, it is also possible to search for correlations in the frequency domain. For example, the *coherency* of two signals is defined as the correlation of their spectra at each frequency. Recently the concept of *analytic signal* has been successfully applied to estimate the degree of synchrony between two channels [Rosenblum *et al.*, 1998; Tass *et al.*, 1998]. Given a real signal $x(t)$, the function

$$Z_x(t) = x(t) + ih_x(t) = A_x(t)e^{i\theta_x(t)} \quad (3)$$

represents its extension in the complex plane. Assuming that $Z_x(t)$ is an analytic function, $x(t)$ and $h_x(t)$ are linked through the Hilbert transform, defined as

$$h_x(t) = \frac{1}{\pi} \text{P.V.} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau, \quad (4)$$

where P.V. stands for the Cauchy principal value of the integral. By inserting (4) into (3) one obtains the instantaneous amplitude $A_x(t)$ and phase $\theta_x(t)$ of the analytic signal. Given two signals $x(t)$ and $y(t)$ one can then calculate $\theta_x(t) - \theta_y(t)$ to study the instantaneous synchronization of two signals. This method performs well also for coupled low-dimensional chaotic oscillators [Rosenblum *et al.*, 1996]. Related approaches on phase synchronization are presented in [Allefeld & Kurths, 2004a] and in [Allefeld & Kurths, 2004b].

The Hilbert transform as well as the coherency and other methods of quantifying correlations in the frequency domain require the time-window to be long enough to permit a reliable estimation of the spectrum.

1.2. Searching for correlations in the phase-space

An interesting approach to estimate the statistical dependence of two signals, which combines information theory and nonlinear-physics methods, has been proposed by Pawelzik [1991, 1994]. It consists of estimating the mutual information, however not between the p.d.fs. of the observations, but rather between the local density function of each signal in the phase space. The phase space itself is a geometrical construction to represent the dynamics of deterministic and stochastic systems. It can be approximated with delay coordinates [Takens, 1981; Kantz & Schreiber, 1997]. Let $\mathbf{x} = (x(1), x(2), \dots, x(N))$ be the sequence of observations in channel x . Then the sequence of vectors

$$\mathbf{x}(t) = (x(t), x(t+1), x(t+2), \dots, x(t+(m-1)))$$

for $t = 1 \dots N - (m - 1)$ corresponds to an m -dimensional representation of the signal x . The local density of the signal x in the phase space is calculated as

$$p_t^\varepsilon(x) = \frac{1}{N - (m - 1)} \sum_{\substack{s=1 \\ s \neq t}}^{N-(m-1)} \Theta(\varepsilon - |\mathbf{x}(t) - \mathbf{x}(s)|)$$

where ε is the size of the discretization bins and $\Theta(x)$ is the Heaviside function. The same procedure is applied to the second channel y and to z , the joint event of x and y , whose phase space is constructed as the concatenation of the phase spaces of x and y : $\mathbf{z} = (\mathbf{x}, \mathbf{y})$. Then the mutual information is computed as

$$M(\varepsilon) = \frac{1}{N - (m - 1)} \sum_{t=1}^{N-(m-1)} M_t(\varepsilon)$$

where

$$M_t(\varepsilon) = \log \frac{p_t^\varepsilon(z)}{p_t^\varepsilon(x)p_t^\varepsilon(y)}$$

is defined as the time-resolved mutual information. If one plots $M(\varepsilon)$ as a function of $\log(\varepsilon)$, the graph exhibits a negative slope. The slope increases in absolute value with m and eventually saturates if the representation of the signals in the phase space has some low-dimensional structure, i.e. if the dynamics has a low-dimensional global attractor. It has been shown [Pawelzik, 1991] that the fractal dimension of a global strange attractor can be better estimated

from two time series by means of the mutual information than from only one time series by means of the entropy.

In principle, the time-resolved mutual information can be used to detect instantaneous correlations. In practice, however, problems arise when interpreting $M_t(\varepsilon)$ because it depends on the coarse-grain parameter ε and on the embedding dimension m , which does not converge unless the data have a low-dimensional structure. Thus, for signals of high-dimensional structure, the fluctuations of $M_t(\varepsilon)$ are mainly due to a poor sampling of the phase space rather than significant changes of correlation. In addition, the (time-resolved) mutual information is computationally expensive. However, this approach suggests that the phase space is a natural framework to search for correlations between signals. In the next section we present in detail a method to uncover (short-term) correlations in the phase space with little computational cost. Preliminary results have been given in [Ritz *et al.*, 2001].

2. Methods and Computational Experiments

2.1. Angular distance in the phase space

We first proceed to reconstruct the phase space of the signals by using delay coordinates with the same sampling interval τ for each channel. Let us consider only two channels x and y . The vectors

$$\begin{aligned} \mathbf{x}(t) &= (x(t), x(t+\tau), x(t+2\tau), \dots, x(t+(m-1)\tau)) \\ \mathbf{y}(t) &= (y(t), y(t+\tau), y(t+2\tau), \dots, y(t+(m-1)\tau)) \end{aligned}$$

represent a sliding time-window of length $m \cdot \tau$. For increasing time t the vectors draw two m -dimensional trajectories in the phase space. The dimension of the phase space m and the sampling interval τ are free parameters but we will show below (see Sec. 3.3) how to choose both in an optimal manner. Within this framework it is now very intuitive to define the degree of similarity between both signals over the sliding short-time window as the angle between their respective vectors:

$$\phi(t) = \arccos \left(\frac{\mathbf{x}(t) \cdot \mathbf{y}(t)}{|\mathbf{x}(t)| \cdot |\mathbf{y}(t)|} \right).$$

Observe that the expression

$$\cos \phi(t) = \frac{\mathbf{x}(t) \cdot \mathbf{y}(t)}{|\mathbf{x}(t)| \cdot |\mathbf{y}(t)|} \quad (5)$$

resembles the correlation coefficient over the sliding short-time window of length $m \cdot \tau$. However, there is a subtle difference to point out: in contrast to (1), the mean values of each signal over the short-time window are not removed in (5). Removing the mean values would be equivalent to stretch or shrink the axis of the phase space at each point in time. Thus, when calculating the sliding correlation coefficient over short-time windows, one should not forget that removing the mean value is inconsistent with the phase space representation of the signals.

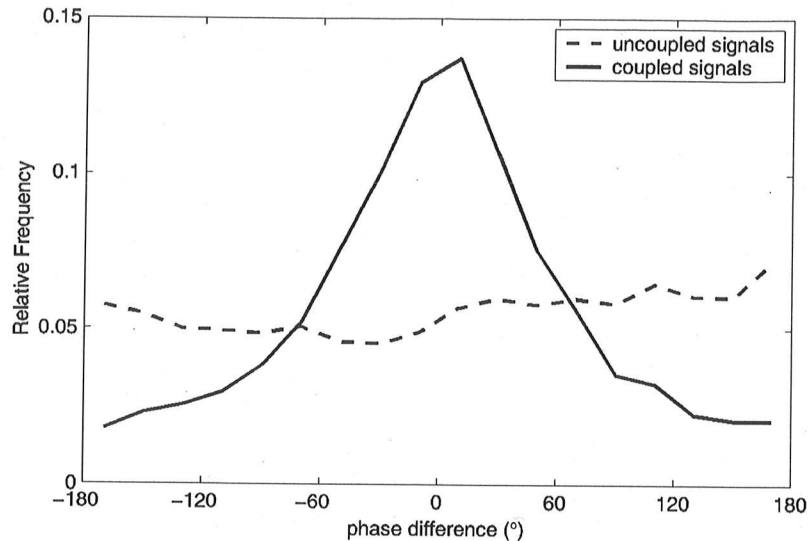
In the following, we will not consider $\cos \phi$, but the angle ϕ itself, because it has a direct geometrical meaning. The values of $\phi(t)$ can be regarded as realizations of a stochastic process. The basic idea in our approach is that the estimated p.d.f. of ϕ characterizes the coupling between signals with low computational cost. The following examples will support this statement. In the first example we consider two identical systems X, Y that are driven by two different stochastic forces and study the effects of correlation between the external forces. In the second example we consider two identical systems driven by correlated stochastic forces, however, the systems' parameters are changed at a certain time.

Example 1. Let the dynamics of systems X and Y be described by

$$\begin{aligned} x(t) &= \sum_{\tau=1}^n a(\tau) \cdot x(t-\tau) \\ &\quad + \sum_{\tau=0}^n b(\tau) \cdot \eta(t-\tau) \quad \text{and} \\ y(t) &= \sum_{\tau=1}^n a(\tau) \cdot y(t-\tau) \\ &\quad + \sum_{\tau=0}^n b(\tau) \cdot \xi(t-\tau), \end{aligned} \quad (6)$$

respectively. Each equation represents a linear ARMA model of order n . We set $n = 10$. The coefficients $a(\tau)$ and $b(\tau)$ are chosen as coefficients of a Butterworth pass-band filter of order $n/2 = 5$ [Rabiner & Gold, 1975], where the pass band is between 10 and 30 Hz and the sampling frequency is $1/\tau = 100$ Hz.

Case 1.1. The stochastic forces η and ξ are not correlated: The values of η and ξ are normally distributed with zero mean and unit variance and $\langle \eta(t) \cdot \xi(t') \rangle = 0 \forall t, t'$. For the numerical example



(a) Phase-difference-distribution calculated with the Hilbert transform

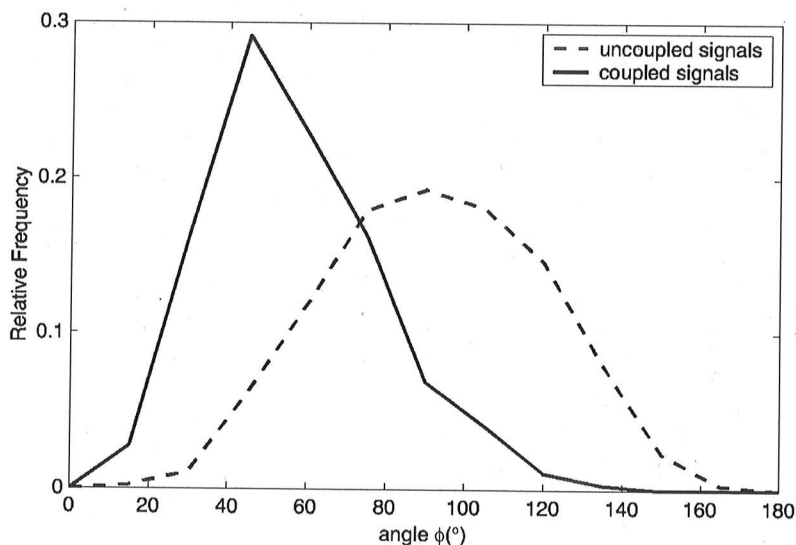
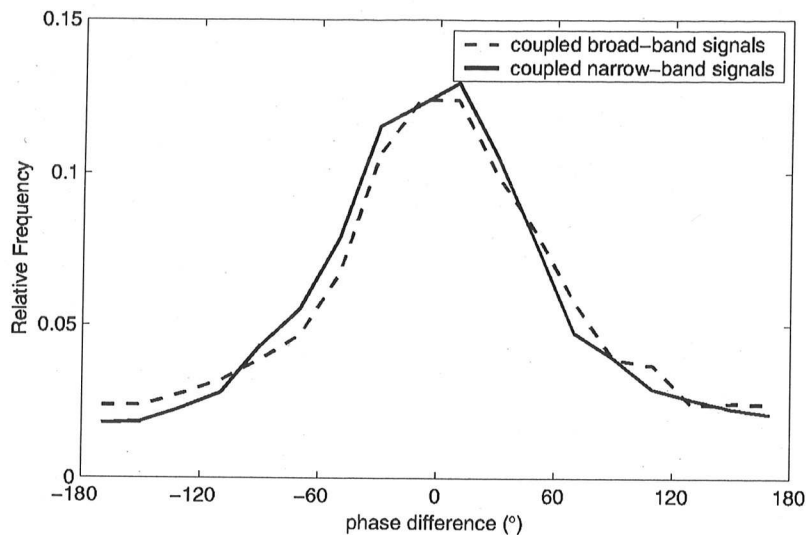
(b) Distribution of the instantaneous angular distance ϕ in phase space

Fig. 1. Comparison of the performance of two methods to study Example 1: (a) The instantaneous phase of x and y is calculated by applying the concept of analytic signal (see Introduction). Then we plot the distribution of the phase difference. This method detects the change in correlation of the inputs η and ξ . When there is no correlation at all, the phase-difference is uniformly distributed. As soon as the inputs are correlated, there appears a stochastic phase-locking of the outputs x and y , as revealed by the peak at $\phi = 0^\circ$. (b) Distribution of the instantaneous angle $\phi(t)$ between x and y in the ($m = 11$)-dimensional phase space. There is also a clear difference between both cases with this method. When the inputs η and ξ are uncorrelated, the distribution of ϕ is symmetric and centered at 90° . That means \mathbf{x}_t and \mathbf{y}_t lie in average orthogonal to each other in the phase space. For a given correlation of the inputs, the distribution is no longer centered at $\phi = 90^\circ$ and becomes more asymmetric.

shown in Fig. 1, the correlation coefficient between x and y turns out to be -0.02 . The distribution of the phase difference after the Hilbert transform is constant and the p.d.f. of ϕ is symmetric and centered at 90° . See also the dashed lines in Figs. 1(a) and 1(b).

Case 1.2. The stochastic forces η and ξ are correlated: $\langle \eta(t) \cdot \xi(t') \rangle = 0.6 \cdot \delta_{tt'}$. The correlation coefficient between x and y is 0.52 , in the example shown. The distribution of the phase difference after the Hilbert transform is symmetric and centered at 0° . The p.d.f. of ϕ is slightly asymmetric and no



(a) Phase-difference distribution calculated with the Hilbert transform

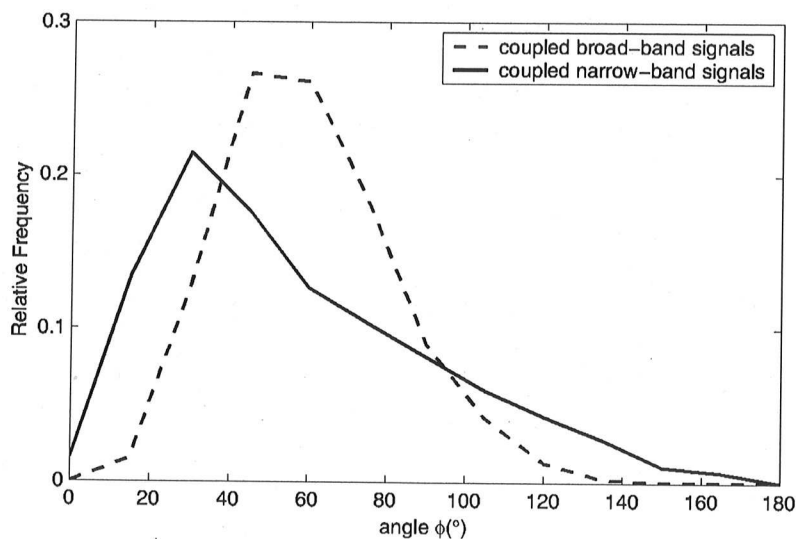
(b) Distribution of the instantaneous angular distance ϕ in phase space

Fig. 2. Comparison of the performance of two methods to study Example 2: (a) The instantaneous phase of x and y is calculated by applying the concept of analytic signal (see Introduction). Then we plot the distribution of the phase difference. We observe the same degree of stochastic synchronization of the analytic signals before and after the transition. In essence, this method detects the correlation of the inputs η , ξ in the outputs x , y . (b) The angular distances in the ($m = 11$)-dimensional phase space are not only sensitive to the correlation of the inputs but also to how the system processes such correlation. The difference between both methods resides in the embedding. The analytic signals “live” in the complex plane, which is two-dimensional, but the m number of degrees of freedom of x and y is $m = 11$. Thus the interdependence of the signals is to be found in a 11-dimensional phase space.

longer centered at 90° but at 59° , see also the continuous lines in Figs. 1(a) and 1(b).

We observe that in this first example all the three methods, (i) correlation coefficient, (ii) phase-difference of analytic signals and (iii) angular distance in the phase space uncover successfully

the underlying correlation between the signals. In essence, the correlation coefficient measures the correlation between the driving forces. Through the Hilbert transform, it is possible to detect a stochastic phase-locking between the analytic signals x and y , which is due to the correlation of the driving forces. The p.d.f. of ϕ is sensitive to the corre-

lated direction-sampling of the ($m = n + 1 = 11$)-dimensional phase space by x and y , as we will see later on.

Example 2. Let the dynamics of systems X and Y be again given by

$$\begin{aligned} x(t) &= \sum_{\tau=1}^n a(\tau) \cdot x(t-\tau) + \sum_{\tau=0}^n b(\tau) \cdot \eta(t-\tau) \quad \text{and} \\ y(t) &= \sum_{\tau=1}^n a(\tau) \cdot y(t-\tau) + \sum_{\tau=0}^n b(\tau) \cdot \xi(t-\tau), \quad (7) \end{aligned}$$

respectively. At some point in time, however the parameters a and b of both systems are now switched into a' and b' ,

$$\begin{aligned} x(t) &= \sum_{\tau=1}^n a'(\tau) \cdot x(t-\tau) + \sum_{\tau=0}^n b'(\tau) \cdot \eta(t-\tau) \\ y(t) &= \sum_{\tau=1}^n a'(\tau) \cdot y(t-\tau) + \sum_{\tau=0}^n b'(\tau) \cdot \xi(t-\tau). \end{aligned} \quad (8)$$

As before, each equation represents a linear ARMA model of order n . We set $n = 10$. The coefficients $a(\tau)$ and $b(\tau)$ are again chosen as coefficients of a Butterworth pass-band filter of order $n/2 = 5$ [Rabiner & Gold, 1975], where the pass band is between 5 and 30 Hz. The coefficients $a'(\tau)$ and $b'(\tau)$ are also chosen as coefficients of a Butterworth pass-band filter of order $n/2 = 5$, but the pass band is now between 15 and 20 Hz. The sampling frequency is $1/\tau = 100$ Hz. The stochastic forces η and ξ are correlated with $\langle \eta(t) \cdot \xi(t') \rangle = 0.6 \cdot \delta_{tt'}$ before and after the transition.

In this example, we notice important differences between the performance of the three methods: the correlation coefficient in the simulation is 0.52 before and 0.51 after the transition. The distribution of the phase-difference of the analytic signals does not show any significant change either [Fig. 2(a)]. In contrast, the p.d.f. of ϕ detects the transition of the dynamics: the distributions before and after the transition have similar mean values (60° and 59° respectively) but they have different variance and skewness [Fig. 2(b)].

The correlation coefficient and the phase-difference of analytic signals basically measure the amount of correlation of the input to the system, i.e. the driving force, but they neglect the translation of such correlation into the output. This

happens because both measures are based on low-dimensional projections of the phase space and therefore do not capture its whole structure. The correlation coefficient measures correlations only in one dimension of the phase space and the Hilbert transform gives the phase of a signal on a two-dimensional embedding, the complex plane of the analytic signal. However, the signals x and y of both examples can be only embedded in a phase space of $m = n + 1 = 11$ dimensions. It is therefore interesting to ask to which quantity the p.d.f. of ϕ is particularly sensitive to. To answer this question, observe that each of Eqs. (6)–(8) can be rewritten as

$$\begin{aligned} \mathbf{a} \cdot \mathbf{x}_t &= \mathbf{b} \cdot \boldsymbol{\eta}_t \\ \mathbf{a} \cdot \mathbf{y}_t &= \mathbf{b} \cdot \boldsymbol{\xi}_t \end{aligned} \quad (9)$$

where

$$\begin{aligned} \mathbf{a} &= (1, a(1), \dots, a(n)) \\ \mathbf{b} &= (b(0), b(1), \dots, b(n)) \end{aligned}$$

are constant vectors in the ($m = n + 1 = 11$)-dimensional phase space;

$$\begin{aligned} \mathbf{x}_t &= (x(t), \dots, x(t-n)) \\ \mathbf{y}_t &= (y(t), \dots, y(t-n)) \end{aligned}$$

are the vectors that draw the signals in the phase space and

$$\begin{aligned} \boldsymbol{\eta}_t &= (\eta(t), \dots, \eta(t-n)) \\ \boldsymbol{\xi}_t &= (\xi(t), \dots, \xi(t-n)) \end{aligned}$$

are the vectors that represent random driving forces in the phase space. Now consider the projections of \mathbf{x}_t and \mathbf{y}_t onto the direction of \mathbf{a} :

$$\begin{aligned} x_a(t) &= \frac{1}{|\mathbf{a}|} \mathbf{b} \cdot \boldsymbol{\eta}_t \\ y_a(t) &= \frac{1}{|\mathbf{a}|} \mathbf{b} \cdot \boldsymbol{\xi}_t \end{aligned} \quad (10)$$

The product of the projections, whose geometrical meaning resembles that of the angle $\phi(t)$ between \mathbf{x}_t and \mathbf{y}_t , yields

$$x_a(t) \cdot y_a(t) = k^2 \cdot \eta_b(t) \cdot \xi_b(t)$$

where $k = |\mathbf{b}|/|\mathbf{a}|$ and η_b , ξ_b are the projections of $\boldsymbol{\eta}_t$, $\boldsymbol{\xi}_t$ onto \mathbf{b} respectively. The expectation value of this quantity is

$$\langle x_a(t) \cdot y_a(t) \rangle = k^2 \cdot \langle \eta_b(t) \cdot \xi_b(t) \rangle. \quad (11)$$

Recall that η and ξ are Gaussian variables. For random Gaussian variables a vanishing correlation

implies statistical independence [Hyvärinen *et al.*, 2001]. Thus, $\langle \eta(t) \cdot \xi(t') \rangle = 0 \forall t, t'$ implies that the vectors η_t and ξ_t scan all directions of the phase space independently and that their projections onto any given vector, say \mathbf{b} , are not correlated, i.e. $\langle \eta_b(t) \cdot \xi_b(t) \rangle = 0$. According to (11), neither are x_a and y_a , $\langle x_a(t) \cdot y_a(t) \rangle = 0$. From the independence of η_t and ξ_t , we analogously expect, that on average both vectors are orthogonal to each other, i.e. $\langle \phi(t) \rangle = 90^\circ$.

Higher order moments of $x_a(t) \cdot y_a(t)$ provide complementary information about the interdependence of the signals $x(t)$ and $y(t)$. The variance, or second central moment of $x_a(t) \cdot y_a(t)$ is given by

$$\begin{aligned} & \langle (x_a \cdot y_a - \langle x_a \cdot y_a \rangle)^2 \rangle \\ &= k^4 \cdot (\langle \eta_b^2 \cdot \xi_b^2 \rangle - \langle \eta_b \cdot \xi_b \rangle^2) \end{aligned}$$

and the skewness or third central moment by

$$\begin{aligned} & \langle (x_a \cdot y_a - \langle x_a \cdot y_a \rangle)^3 \rangle \\ &= k^6 \cdot (\langle \eta_b^3 \cdot \xi_b^3 \rangle + 2\langle \eta_b \cdot \xi_b \rangle^3 \\ & \quad - 3 \cdot \langle \eta_b^2 \cdot \xi_b^2 \rangle \langle \eta_b \cdot \xi_b \rangle). \end{aligned}$$

Observe that if η_t and ξ_t are independent the mean and the skewness vanish, but the variance does not. In general, for two identical correlated ARMA models, numerical simulations indicate that the mean value and the skewness of ϕ increase with increasing correlation of the input whereas the variance decreases with the bandwidth of the spectra.

Taking (9) into account, we may rewrite (10) as $x_a(t) \cdot y_a(t) = \eta_t \cdot Q \cdot \xi_t$ where

$$Q = \frac{\mathbf{b} \otimes \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}}.$$

From this expression we deduce that the product $x_a(t) \cdot y_a(t)$ measures the relative orientation of η_t and ξ_t with the *pseudometric* Q (the prefix "pseudo-" drops out if Q is positive definite), which is determined by the system. Thus, the angular distance between two signals in the phase space accounts for how a correlation in the input is transformed by the system into the output.

2.2. Optimal embedding for angular distances

In the previous examples we knew the dimensionality ($m = 11$) of the signals. But in applications it is usually difficult to find out a proper model for the measured signals and hence, the actual dimension of the phase space in which $\phi(t)$ has to be calculated

remains unknown. However, we show here that it is possible to estimate an optimal embedding for $\phi(t)$ by means of the symmetrized Kullback–Leibler divergence G [Schreiber, 1999],

$$G = \int_{-\infty}^{\infty} [P(s) - Q(s)] \cdot [\log P(s) - \log Q(s)] ds$$

where $P(s)$ and $Q(s)$ are two probability distributions of s to be compared. The symmetrized Kullback–Leibler divergence expresses the knowledge about one p.d.f. when we completely know the second. In other words: G measures how different two distributions are from an information-theoretical point of view. Take again the system of Example 2. Its dynamics can be completely embedded in an ($m = 11$)-dimensional phase space. Thus, increasing the dimension by taking more delay coordinates than $m = 11$ will not improve our knowledge of X , Y nor of the coupling between both, no matter what \mathbf{a} and \mathbf{b} are. This means that the symmetrized Kullback–Leibler divergence between the p.d.fs. of ϕ before and after the tran-

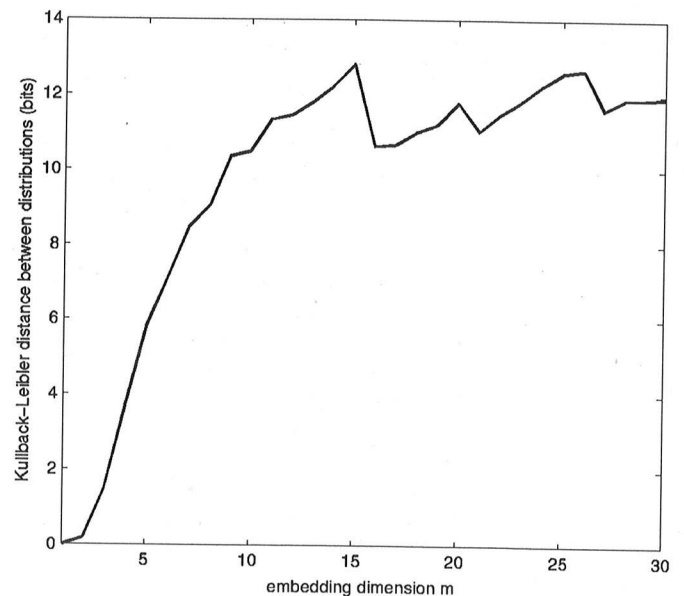


Fig. 3. Estimation of the embedding dimension to calculate angular distances: The symmetrized Kullback–Leibler divergence of two distributions measures the knowledge we gain about one of the two distributions, when we know the other. The dynamics of the systems from Example 2 are embedded in a ($m = 11$)-dimensional phase space. Increasing the m number of dimensions beyond 11 does not give us more information about the dynamics of X and Y , nor about the coupling of both before or after the transition. Therefore the symmetrized Kullback–Leibler divergence between the p.d.fs. of ϕ before and after the transition saturates around $m = 11$ dimensions.

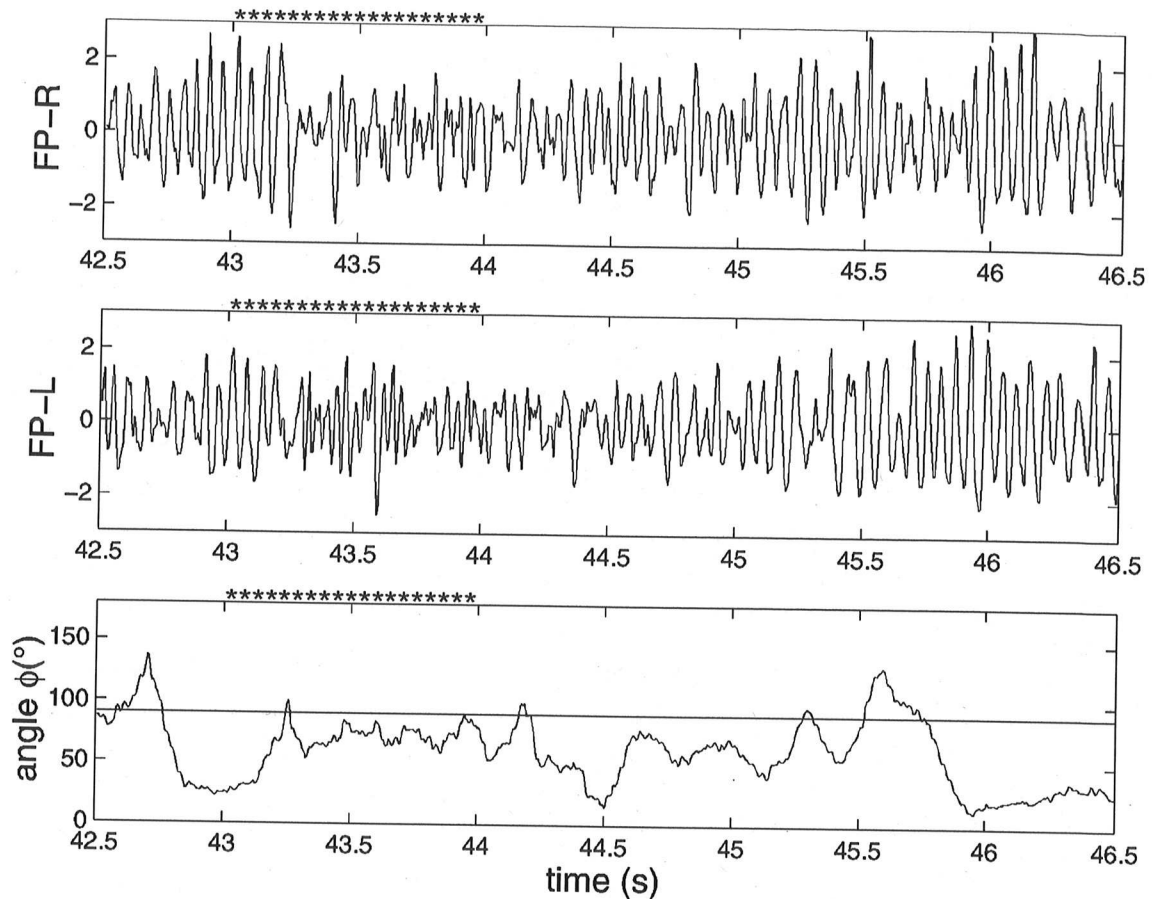


Fig. 4. Application to neurophysiological data. Two upper traces: Extracellular recordings of neural activity (field potentials) in the two hemispheres (left and right) of the honeybee brain. The stars indicate the presentation of an odor. Bottom: Instantaneous coupling between the signals measured as the angle between their trajectories in the phase space ($m = 30$, $\tau = 5$ ms).

sition should saturate around $m = 11$ dimensions, and that is actually what we observe (Fig. 3).

2.3. Remarks

Neurophysiological data often contain nonstationary slow-component artifacts, called *drifts*, that are due to movements of the subject, of the electrodes, etc. Before applying our method such drifts must be removed with an appropriate high-pass filter. Otherwise, the origin of the phase space would not be fixed. This would strongly reduce the diagnostic value of $\phi(t)$.

3. Analysis of Neural Data

3.1. Stimulus-induced change of the dynamical coupling

We now present an application of our method to neurophysiological data. In Fig. 4, the two

upper traces are simultaneous extracellular recordings of neural activity in the brain of the honeybee (sampling rate: 1 kHz; preprocessing: band-pass filter between 5 and 100 Hz). Each electrode was located at the top of the *mushroom body* of one brain hemisphere; the mushroom bodies are central processing stages in insects, necessary for spatial orientation and for the storage of long-term memory [Menzel, 1994]. We study whether the global neural activity of the mushroom bodies in both hemispheres is correlated and whether there is a change in the correlation during odor processing. The stars above the traces in Fig. 4 indicate an odor presentation (duration: one second). Observe the high degree of activity in both hemispheres in the absence of stimulation. The ongoing activity shows episodes of strong irregular oscillations with a frequency peak at around 20 Hz. During stimulation the peak smears out to lower and higher frequencies (data not shown).

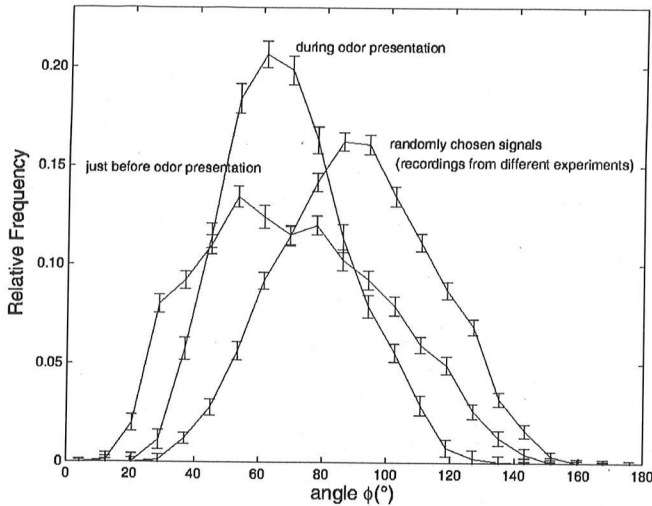


Fig. 5. Probability distributions of ϕ for recordings of the type shown in Fig. 4 ($m = 30$, $\tau = 5$ ms). Three distributions are particularly important: without stimulation, during stimulation and for randomly coupled signals. Observe that the three distributions of ϕ are significantly different from each other.

The lower panel of Fig. 4 depicts the angle ϕ between both signals in the phase space at each point in time ($\tau = 5$ ms, $m = 30$). During stimulation ϕ seems to stabilize around a fixed value whereas its variations are more pronounced during the ongoing activity. This points at an increase in the *dynamical coupling* of the signals: the angle between the signals, no matter whether it is large or small, it is less variable during stimulation. In other words, during stimulation, the movement of the vector $\mathbf{x}(t)$ is constrained more strongly by $\mathbf{y}(t)$ and vice versa. Thus, it is not only the angle $\phi(t)$ itself that represents the correlation, but also its temporal stability. In fact, as we have shown above (see Methods) the whole distribution of ϕ characterizes the coupling. Therefore we construct a histogram of ϕ values during one second of ongoing activity and another histogram of the values of ϕ during one second of stimulation. For the histograms we gather values of $\phi(t)$ from 50 trials (10 times 5 different odorants). As a reference, we also plot the histogram for the angle obtained between signals recorded from different bees on different days, which are supposed to be uncorrelated. For this case, we expect the histogram of ϕ to be centered at 90° . In fact, since $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are independent in this case, most of the time they will be orthogonal to each other in a high-dimensional space. The three histograms are shown in Fig. 5. They all are significantly different from each other.

The error bars were calculated by bootstrapping the set of ϕ values.

3.2. Defining a metric for the coupling

Interestingly, the mean values of the distributions just before and during stimulation are rather similar. However, the variance and the skewness are different. This reminds us of Example 2 (see Sec. 2.1) where the same amount of correlation led to the same mean value of ϕ before and after the transition. In any case, this result confirms again that the whole distribution of $\phi(t)$ carries information about the coupling dynamics and thus, about the (short-term) dependence of the signals. We therefore define the degree of correlation as the *distance* between a given distribution of $\phi(t)$ and the distribution of $\phi(t)$ for randomly chosen signals. In other words: the coupling is the deviation from randomness of $\phi(t)$. It only remains to find a proper measure of distance between distributions. We propose to use again the symmetrized Kullback–Leibler divergence G . This measure satisfies the triangular inequality and hence defines a metric on the space of distributions [Schreiber, 1999].

Our calculation for $m = 30$ yields the results shown in Table 1. The errors are calculated from the values of G between bootstrapped distributions. They delimit a 95% confidence interval. Observe that the distributions of ϕ during odor presentation but also during ongoing activity deviate significantly from randomness. Furthermore, during odor presentation the distance to randomness is significantly larger than during ongoing activity. This means that *there is a significant increase of correlation or coupling between the signals during odor processing and hence, between the mushroom-body activity of both brain hemispheres*. Such increase of coupling was completely overseen by the correlation coefficient and the phase-difference distribution of the analytic signals (data not shown). Concerning the angular distance in phase space, we notice that although the distributions for ongoing activity and odor presentation have similar mean values (approx 70°), the variance and skewness are different (see Fig. 5). This situation is analogous to that of Example 2 (see Sec. 2.1) and invites us to speculate about the biological meaning of the stimulus-induced change of coupling. The comparison with Example 2 suggests that both mushroom bodies are correlated to the same degree all the time

Table 1. Kullback–Leibler distance G between distributions ($m = 30$, $\tau = 5$ ms).

G (bits)	Ongoing Activity	Odor Presentation	Randomly Chose Signals
Ongoing activity	0	6 ± 1	11 ± 1
Odor presentation	6 ± 1	0	15 ± 2
Randomly chosen signal	11 ± 1	15 ± 2	0

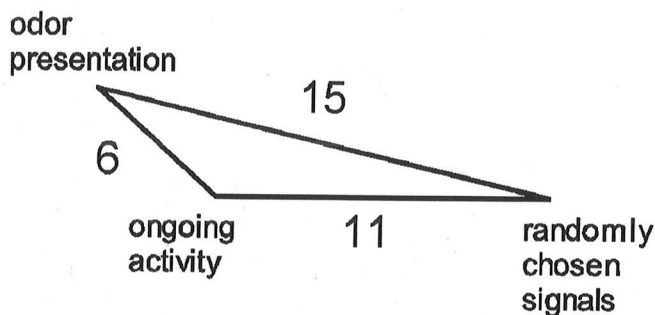


Fig. 6. Representation of the stimulus-induced changes of coupling: Due to the fact that the symmetrized Kullback–Leibler divergence G defines a metric space, we are able to represent our results geometrically. We consider the coupling as the deviation from the distribution of ϕ for randomly coupled signals. Observe that during odor-presentation the coupling between both hemispheres increases. G is expressed in bits ($m = 30$, $\tau = 5$ ms).

and that the stimulus triggers a transition in the dynamics that makes the *communication* between both mushroom bodies more efficient. The constant correlation of the mushroom bodies with or without stimulation may reflect the underlying anatomical connectivity. The stimulus-induced transition from a narrow-band dynamics to a broadband one may indicate an increase of the complexity in the global neural dynamics during odor processing.

One implication of G being a metric is that, when there are multiple dynamical states, clustering techniques can be applied to reveal task-dependent hierarchical structures among those states. Another interesting implication is that we can represent our results geometrically, as in Fig. 6.

3.3. Optimal choice of τ and m . Embedding the coupling

As mentioned above, τ and m can be chosen in a meaningful way. We now show how to make a proper choice of both.

The reconstruction of the phase space through delay coordinates is a transformation that preserves the topology but not the norm. That means that

we do not reproduce the phase space exactly, but a variant of it in which the reference axes are globally stretched or shrunk. This effect is especially pronounced when the time lag τ is very small. In this case, there is a strong linear correlation between the values of successive coordinates in phase space and hence, the angles between vectors are systematically distorted. To avoid this phenomenon, it is recommended to choose τ according to the standard procedures of attractor reconstruction [Takens, 1981; Fraser & Swinney, 1986]. For the measured electrophysiological signals the optimal value turns out to be $\tau = 5$ ms, given our time resolution of 1 ms.

The dimension of the phase space, or equivalently the width of the sliding (short-)time window, can be chosen according to the following criterion. As in Fig. 7, we plot the Kullback–Leibler distance G between the ϕ distributions for several m . We observe that the distance between ongoing activity and randomly chosen signals as well as the distance between odor presentation and randomly chosen signals increase with m .

However, the distance between the distributions of ϕ during ongoing activity and during odor processing saturates between $m = 20$ and $m = 30$. This implies that the coupling between signals can be embedded in a phase space of at most $m = 30$ dimensions, and equivalently, that the shortest time-window in which one has to search for significant changes of correlated activity between both hemispheres is $m \cdot \tau = 20 \cdot 5$ ms = 100 ms. The other curves do not saturate because the deviation from random coupling, G , is more pronounced, the higher-dimensional the space. This is due to the fact that with increasing m the randomly coupled vectors become orthogonal more quickly than vectors bound by some coupling.

Acknowledgments

We are very thankful to the anonymous referees for their useful suggestions. This work has been supported by a NaFöG-grant (R. F. Galán) and the DFG.

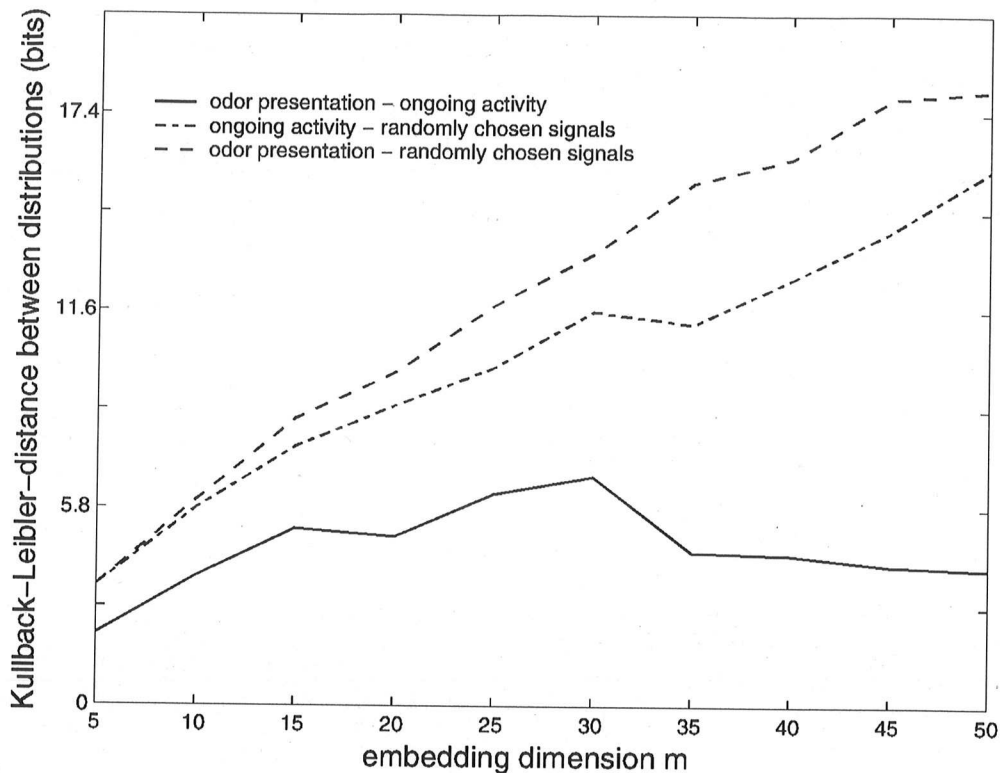


Fig. 7. Embedding the coupling dynamics: The Kullback–Leibler distance G between the distributions of ϕ during the ongoing activity and during stimulation saturates between $m = 20$ and $m = 30$. This indicates that the shortest time-window in which one can uncover significant changes of the correlation is approximately $m \cdot \tau = 20 \cdot 5 \text{ ms} = 100 \text{ ms}$.

References

- Allefeld, C. & Kurths, J. [2004a] "The synchronization cluster: An approach to multivariate phase synchronization analysis and its application to event-related potentials," *Int. J. Bifurcation and Chaos* **14**, 417–426.
- Allefeld, C. & Kurths, J. [2004b] "Testing for phase synchronization" *Int. J. Bifurcation and Chaos* **14**, 405–416.
- Fraser, A. M. & Swinney, H. L. [1986] "Independent coordinates for strange attractors from mutual information," *Phys. Rev.* **A33**, 1134–1140.
- Galán, R. F., Ritz, R., Szyszka, P. & Herz, A. V. M. [2002] "Detecting correlations in multi-site recordings of nonstationary brain activity," *Attractors, Signals and Synergetics*, ed. Klonowski, W. (PABST Science Publishers), pp. 377–386.
- Honerkamp, J. [1993] *Stochastic Dynamical Systems. Concepts, Numerical Methods, Data Analysis* (VCH Publishers).
- Hyvärinen, A., Karhunen, J. & Oja, E. [2001] *Independent Component Analysis* (John Wiley).
- Kantz, H. & Schreiber, T. [1997] *Nonlinear Time Series Analysis* (Cambridge University Press).
- Kullback, S. & Leibler, R. A. [1951] "On information and sufficiency," *Ann. Math. Stat.* **22**, 79–86.
- Kullback, S. [1959] *Information Theory and Statistics* (Dover Publications).
- Menzel, R. [1994] "The mushroom bodies in the honeybee: From molecules to behaviour," *Fortschritte der Zoologie* **39**, 81–102.
- Pawelzik, K. [1991] *Nichtlineare Dynamik und Hirnaktivität* (Verlag Harri Deutsch).
- Pawelzik, K. [1994] "Detecting coherence in neuronal data," in *Models of Neural Networks II*, eds. Domany, J. L. v. H. E. & Schulten, K. (Springer), pp. 253–285.
- Priestley, M. B. [1996] *Spectral Analysis and Time Series. Probability and Mathematical Statistics* (Academic Press).
- Rabiner, L. R. & Gold, B. [1975] *Theory and Application of Digital Signal Processing* (Englewood Cliffs, Prentice-Hall, NJ).
- Ritz, R., Galán, R. F. Szyszka, P. & Herz, A. V. M. [2001] "Analysis of odor processing in the mushroom bodies of the honeybee," *Neurocomputing* **38–40**, 313–318.
- Rosenblum, M. G., Pikovsky, A. & Kurths, J. [1996] "Phase synchronization of chaotic oscillators," *Phys. Rev. Lett.* **76**, 1804–1807.
- Rosenblum, M. G., Kurths, J., Pikovsky, A. Schäfer, C., Tass, P. & Abel, H.-H. [1998] "Synchronization in noisy systems and cardiorespiratory interaction," *IEEE Engin. Med. Biol.* **17**, 46–53.

- Schreiber, T. [1999] "Interdisciplinary application of nonlinear time series methods," *Phys. Rep.* **308**, chaodyn/9807001.
- Shannon, C. E. [1948] "The mathematical theory of communication," *Bell Syst. Tech. J.* **27**, p. 379.
- Takens, F. [1981] *Detecting Strange Attractors in Turbulence*, Lecture Notes in Mathematics, Vol. 898.
- Tass, P., Rosenblum, M. G., Weule, J., Kurths, J., Pikovsky, A., Volkman, J., Schnitzler, A. & Freund, H.-J. [1998] "Detection of n:m phase locking from noisy data: Application to magnetoencephalography," *Phys. Rev. Lett.* **81**, 3291-3294.