EJN fens EUROPEAN JOURNAL OF NEUROSCIENCE

European Journal of Neuroscience, pp. 1–21, 2012 doi:10.1111/j.1460-9568.2011.07987.x

Scaled correlation analysis: a better way to compute a cross-correlogram

Danko Nikolić,[1,2] Raul C. Mureşan,[1,3] Weijia Feng[1,2] and Wolf Singer[1,2]
[1]Max Planck Institute for Brain Research, Frankfurt am Main, Germany
[2]Frankfurt Institute for Advanced Studies, Johann Wolfgang Goethe-University, Frankfurt am Main, Germany
[3]Center for Cognitive and Neural Studies, Romanian Institute of Science and Technology, Cluj-Napoca, Romania

Keywords: auto-correlation, cat, cross-correlation, oscillation, rate co-variation, synchrony

Abstract

When computing a cross-correlation histogram, slower signal components can hinder the detection of faster components, which are often in the research focus. For example, precise neuronal synchronization often co-occurs with slow co-variation in neuronal rate responses. Here we present a method – dubbed scaled correlation analysis – that enables the isolation of the cross-correlation histogram of fast signal components. The method computes correlations only on small temporal scales (i.e. on short segments of signals such as 25 ms), resulting in the removal of correlation components slower than those defined by the scale. Scaled correlation analysis has several advantages over traditional filtering approaches based on computations in the frequency domain. Among its other applications, as we show on data from cat visual cortex, the method can assist the studies of precise neuronal synchronization.

Introduction

A common problem with the use of the cross-correlation histogram (CCH) is the presence of correlated signal components that occur with a slower time course than those of interest. For example, when precise neuronal synchrony is investigated, a component with a 25 ms time scale (e.g. 40 Hz gamma oscillation) may be considered interesting, whereas a component with a 100 ms time scale or larger (≤ 10 Hz) – here referred to as rate co-variation – would normally be considered uninteresting. Rate co-variation, also known as noise correlation (Averbeck *et al.*, 2006), can impair the study of precise neuronal synchrony by modulating the size of the center peak in the CCH (Brody, 1999). Center peaks produced by rate co-variation are broader than those produced by neuronal synchrony but, nevertheless, the two processes are often difficult to dissociate. In particular, rate co-variation becomes a problem when it co-exists with synchrony (Staude *et al.*, 2008). For example, if changes in the strength of synchronization (such as those induced by stimulus properties, e.g. Gray *et al.*, 1989; Biederlack *et al.*, 2006) are investigated, in a case of peak superposition, it is difficult to tell whether a change in the size of the peak is due to changes in the strength of synchronization or, alternatively, due to a change in the strength of rate co-variation. Thus, for the analyses based on CCHs, it is an advantage if slow rate co-variation can be separated efficiently from the faster neuronal synchronization.

The most widely used method for this purpose is the so-called shift predictor (SP) (or shuffle predictor) (Gerstein & Perkel, 1972; Gray & Singer, 1989; Munk *et al.*, 1995; Nowak *et al.*, 1999). A SP is a CCH computed between pairs of signals recorded at different times (i.e. from different experimental trials and in response to the same stimulus). The assumption underlying the SP is that the slow temporal dynamics of rate co-variation does not change across repeated trials (i.e. the dynamics is time-locked to the stimulus onset) and, thus, the SP reveals this stimulus-evoked correlation in a center peak. In contrast, precise neuronal synchrony vanishes from a SP, as these correlations are usually generated by internally-timed processes not locked to the stimulus onset. The contribution of neuronal synchrony can then be isolated by simply subtracting from the standard CCH the stimulus-locked component estimated by the SP (Gerstein & Perkel, 1972).

Unfortunately, SPs do not provide a complete account for rate co-variations. Changes in neuronal firing rates are not always stimulus-locked across repeated trials and every deviation from such regularity results in an underestimation of the magnitude of rate co-variation (e.g. Brody, 1999). Example sources of rate co-variations that lack time-locking to the stimulus are changes in firing rates, which can occur due to fluctuations of attention, eye movements, spontaneous changes in the cortical states or non-stimulus-locked oscillatory rhythms slower than those supporting precise neuronal synchronization. Another problem with SPs is that, in some experimental designs, repeated trials do not exist and, thus, it is not even possible to define a SP, as is the case for example in the analyses of spontaneous neuronal activity (Beggs & Plenz, 2003; Ikegaya *et al.*, 2004; Hahn *et al.*, 2010). Therefore, methods better than SPs are needed to separate rate co-variation from precise neuronal synchrony. Such methods already exist for estimating the significance of joint spike events detected across a larger number of neurons (Grün *et al.*, 2002; Pipa *et al.*,

Correspondence: Danko Nikolić, [1]Max Planck Institute for Brain Research, as above.
E-mail: danko.nikolic@gmail.com

Received 4 October 2008, revised 8 October 2011, accepted 5 December 2011

2008). Also, the methods based on coherence can estimate the strength of synchrony between pairs of spike trains for a given frequency range (Fries *et al.*, 1997; Pesaran *et al.*, 2002). However, these methods cannot be used to construct CCHs as they either do not provide a means for removing the contributions of the slow components or, if they do, as we will show, they alter the spike trains such that the original relationships between the signals cannot be reconstructed accurately. The present work proposes a method that is designed for use with CCHs and auto-correlation histograms, dubbed scaled correlation analysis (SCA). The method attenuates the contributions of the slow component of a CCH (e.g. slow rate co-variation) and reveals the cross-correlation for the fast components of the signals (e.g. strength of precise neuronal synchronization).

Materials and methods

The experimental methods and setup were similar to those reported in several other studies (Biederlack *et al.*, 2006; Nikolić, 2007; Schneider & Nikolić, 2006; Schneider *et al.*, 2006; Nikolić *et al.*, 2009). Here, we report the methods in a shorter form.

Preparation and recordings

In two cats, anesthesia was induced with ketamine (0.2 ml/kg, 5% diluted) and, following the tracheotomy, was maintained with a mixture of 70% N_2O and 30% O_2 and with halothane (0.6%). To prevent eye movements, two cats were paralysed with pancuronium bromide applied intravenously (Pancuronium, Organon, 0.15 mg/kg/h). After the completion of the experiment, the animal was killed by a 3 ml dose of Narcoren (pentobarbital) applied intravenously. All of the experiments were conducted according to the guidelines of the Society for Neuroscience and German law for the protection of animals, approved by the local government's ethical committee and overseen by a veterinarian.

On each of the 16 channels, neuronal activity was recorded extracellularly from multiple neurons (multi-unit activity) by using a silicon-based probe (channels organized in a 4×4 spatial matrix) supplied by the Center for Neural Communication Technology at the University of Michigan. The probe had minimal intercontact distances of 200 μm (0.3–0.5 mega Ohm impedance at 1000 Hz). Electroencephalogram signals were recorded with intracranial silver-ball electrodes placed above the visual cortex. Signals were amplified 1000×, filtered between 500 Hz and 3.5 kHz, and digitized with 32 kHz sampling frequency. The probe was inserted into the cortex approximately perpendicular to the surface, which allowed recording simultaneously from neurons at different depths and with different orientation preferences. In one cat we used nine multi-unit activity signals recorded simultaneously that responded well to visual stimuli and had orientation selectivity that was appropriate for eliciting large responses with the presently used stimuli. All of the receptive fields were overlapping and were thus all stimulated simultaneously by a single stimulus.

Visual stimulation

Stimuli were presented on a 21 inch computer monitor (Hitachi CM813ET) with 100 Hz refresh rate. The software for visual stimulation was the stimulation tool, ActiveSTIM (http://www.ActiveSTIM.com). The stimuli were presented binocularly and the eyes were fused by mapping the borders of the respective receptive fields and then aligning the optical axes of the eyes with an adjustable prism placed in front of one eye. The orientation preferences of the multi-units used in the present analysis were determined by sinusoidal gratings drifting in 12 different directions. Responses were analysed to stimuli that consisted of a single moving bar (Figs 7C and H, and 8A), grating stimuli with either sinusoidal (Figs 7A and F, and 8B–E) or rectangular luminance profiles (Fig. 7D and I), or plaid stimuli constructed by overlapping two rectangular gratings (Fig. 7E and J). With the exception of the single-trial analysis in Fig. 8F, the responses for 20 stimulus presentations were averaged, each in a duration of 2–5 s. To create CCHs smoother than those obtained by 20 stimulus presentations, in several plots we averaged responses to several different stimulation conditions (Fig. 7A, B, F and G). The orientation and speed of all presented stimuli were optimal for eliciting large rate responses.

Artificial continuous signals

To generate artificial continuous signals (i.e. in which samples can take any real value), we used combinations of sinusoid modulated components and high-frequency noise. We first generated two perfect sinusoidal signals, one having a low (ω_{SLOW}) and the other a high (ω_{FAST}) frequency, and constructed the reference signal as a sum of these two components, as follows

$$S_{SLOW}(t) = a_{SLOW} \cdot \sin(\omega_{SLOW} \cdot t) \quad (1)$$

$$S_{FAST}(t) = a_{FAST} \cdot \sin(\omega_{FAST} \cdot t) \quad (2)$$

$$A_t = S_{SLOW}(t) + S_{FAST}(t) \quad (3)$$

where t is time, a_{SLOW} and a_{FAST} are the amplitudes of the slow and fast component, respectively, and ω_{SLOW} and ω_{FAST} are the (angular) frequencies of the slow and fast component, respectively.

The target signal was generated by mixing S_{SLOW} and S_{FAST} with noise, such as to obtain correlated versions of slow and fast components, which were then subsequently summed

$$B(t) = r_{SLOW} \cdot S_{SLOW}(t) + \sqrt{1 - r_{SLOW}{}^2} \cdot y(t) + r_{FAST} \cdot S_{FAST}(t) + \sqrt{1 - r_{FAST}{}^2} \cdot z(t) \quad (4)$$

where r_{SLOW} and r_{FAST} are the correlations of the slow and fast components, respectively, and $y(t)$ and $z(t)$ are noise signals drawn from uniform distributions with zero mean and amplitudes equal to $1.25 \times$ the amplitude of their respective sinusoidal components, S_{SLOW} and S_{FAST}. When the amplitudes of the slow and fast components were scaled, the amplitude of the noise functions was also scaled accordingly.

Artificial binary signals

To generate artificial binary signals (i.e. values 1 or 0), we used non-homogeneous Poisson processes, in which the likelihood of firing a spike in a 1 ms bin, M, was given by a scaled sum of sinusoidal probability functions, as follows

$$P(t) = \max[(a_{SLOW} \cdot \sin(\omega_{SLOW} \cdot t) + a_{FAST} \cdot \sin(\omega_{FAST} \cdot t)), 0] \quad (5)$$

$$R = \frac{\text{desired\_rate}}{\int P(t)} \cdot \frac{1}{1000} \quad (6)$$

$$M(t) = R \cdot P(t) \quad (7)$$

where P is the unnormalized probability function, t is time, a_{SLOW} and a_{FAST} are the amplitudes of the slow and fast component probability

functions, respectively, ω_{SLOW} and ω_{FAST} are the (angular) frequencies of the slow and fast component, respectively, R is a scaling factor that sets the desired average firing rate, and M is the actual probability function used to generate spikes.

The probability function $P(t)$ was a rectified sum of sinusoids, i.e. when the sum was negative it was rectified to 0. Desired firing rates were obtained by computing the integral of P and deriving a scaling factor R of the firing probability function M (see Eqns 6 and 7). At each moment in time, t, a spike was produced if an independent draw from a uniform random distribution in the range [0…1] yielded a value $\leq M(t)$. In previous studies (Moca *et al.*, 2008; Schneider & Nikolić, 2008; Havenith *et al.*, 2011) it was demonstrated that non-homogeneous Poisson processes produce artificial CCHs with properties similar to those of real CCHs.

Results

We first introduce the algorithm for the computation of SCA, and then the results of its applications on simulated data with known correlation properties. Next, SCA is applied to real neurophysiological data recorded from the cat visual cortex. Finally, we compare SCA with other techniques designed for the estimations of synchrony and that are based on cross-spectrum and coherence.

Computation of scaled correlation and its principles of operation

Correlations in restricted sampling ranges

The mechanism underlying the removal of slow components in SCA relies on the reduction of the correlations that emerge from the slow components while retaining the correlations of the fast components. This is achieved through restrictions in the sampling range of the original data. One characteristic of correlation measures, such as Pearson's r, is that any reduction in the variability of a sample relative to the variability in the population affects the measured strength of correlation. Depending on the structure of the data set, the estimate may either decrease or increase relative to the one with full variability, including even the possibility that the correlation changes its sign (Alexander *et al.*, 1984; Millsap, 1989; Held & Foley, 1994; Aguinis & Whitehead, 1997; Raju & Brand, 2003). In Fig. 1A we show an example scatter plot in which two restricted samples (created by dividing the scatter into two unequal halves by the vertical dashed line) exhibit much smaller variability than the entire sample. Importantly, the correlations in the sub-samples are reduced relative to that in the entire sample. In Fig. 1B we show an example of two continuous signals (i.e. data sets with temporal structure), which are more relevant for the present purposes. The slow components of these two signals are negatively correlated (when one increases, the other decreases) but, as indicated by the zoom-in, the fast components can, at the same time, be positively correlated (concomitant increases and decreases). Scaled correlation can be used to focus the analysis on this fast component. Correlations between continuous signals can also be shown in a scatter-plot-like diagram, known as a phase diagram. An example for another pair of continuous signals is shown in Fig. 1C. Here the time flows along the line. In this plot, a restriction in the sampling range is made by cutting out segments of data along the time axis, which is illustrated by the alternate blue and red colors. Although the overall plot indicates a negative correlation (it is oriented from top-left to bottom-right), the segments are predominantly correlated positively (oriented bottom-left to top-right). In binary data (i.e. neuronal spiking activity), restriction of the sampling range can also alter correlations, as illustrated in Fig. 1D. Although the entire

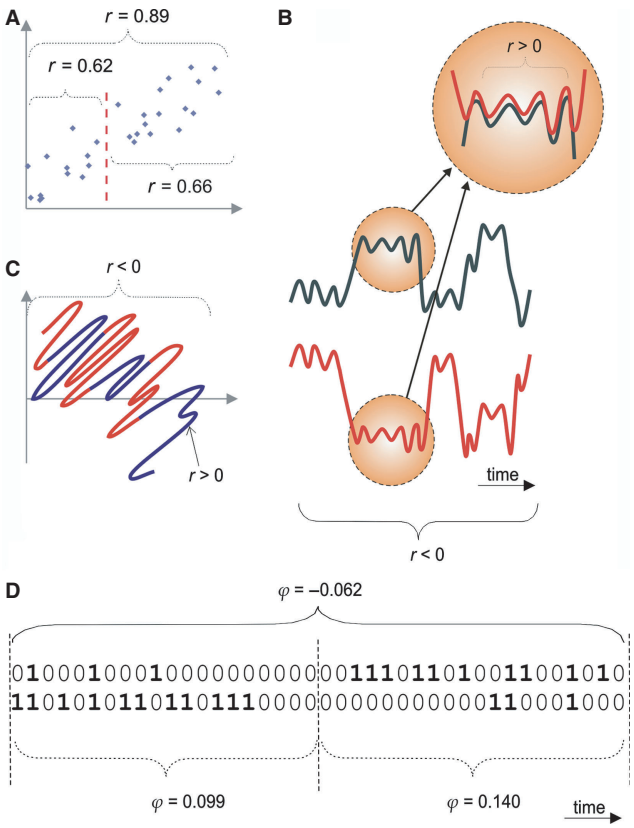


FIG. 1. Effects of restricted sampling on coefficients of correlation. (A) An illustration of how a reduction in sample variance can decrease the measured correlation in a scatter plot. Here, the sample variance is reduced by setting a threshold on the x-axis (red dashed line). (B) An example of restricted sampling by selecting only a sub-segment of continuous signals. The correlation is negative for the entire segment and the correlation is positive for the sub-segment. (C) An example similar to B but shown as a phase plot. In case of LFPs, the two axes would indicate voltage, whereas time flows along the line. The entire data set exhibits a negative correlation, whereas the sub-segments (alternating red and blue colors) have predominantly positive correlations. (D) Segmentation of binary processes (e.g. neural spiking events) works in the same way. It can also reduce variability and hence produce changes in correlation strength. For interpretation of color references in figure legend, please refer to the Web version of this article.

segment exhibits a negative correlation, both sub-segments exhibit positive correlations. Similar segmentation processes are the basis for the computation of SCA, as explained next.

The algorithm for computing scaled correlation analysis

To remove slow components, scaled correlation does not apply filtering of signals prior to the computation of correlation. Rather, scaled correlation can be understood as an algorithm that filters correlations directly by the means of restricted sampling, allowing only the correlations that occur at a given time scale or smaller to enter the calculation. Usually, restricted sampling is considered undesirable and has been studied in the context of estimating a correction that recovers the original correlation of the non-restricted sample (e.g. Alexander *et al.*, 1984; Millsap, 1989; Held & Foley, 1994; Aguinis & Whitehead, 1997; Raju & Brand, 2003). Scaled correlation, in contrast, takes advantage of restricted sampling, creating it intentionally in a controlled manner in order to estimate the correlations only for a sub-set of time scales.

To compute scaled correlation for the central bin in the CCH (i.e. without phase offset between the signals), the signals are cut into a

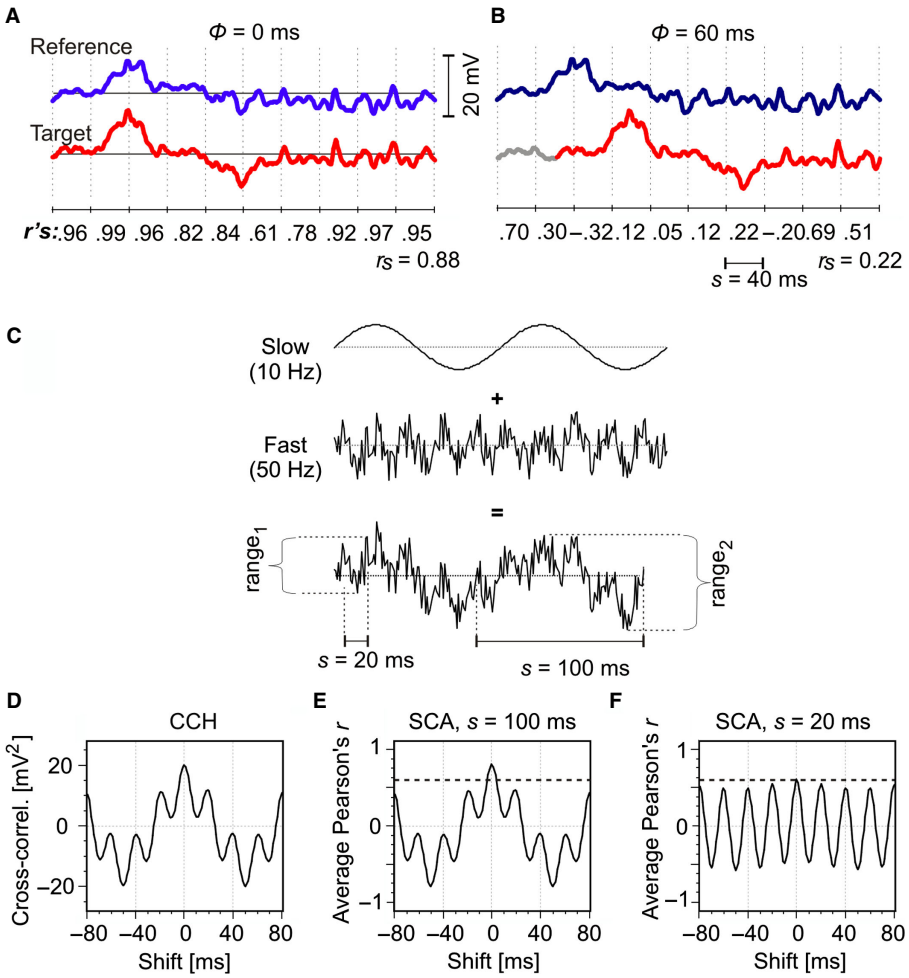


FIG. 2. Computation of SCA. (A) An illustration made on LFP signals recorded from two electrodes simultaneously. Signals are segmented into small pieces of length s (here, s = 40 ms). For each segment, a normalized measure of correlation is computed, r, and the values are averaged over longer stretches of the signal, resulting in $\bar{r}_s$. (B) To compute a complete CCH with all time shifts, φ, the segmentation must always be applied anew after each shift. (C) For simulations in D–F and in Figs 3 and 4, each signal is composed by the addition of two continuous components: a regular slow one and a noisy fast one. The added noise always had high frequency (1 kHz). Hence, the fast components were correlated with $r < 1.0$, the actual correlation depending on the amplitude of noise. In contrast, the slow components were always correlated perfectly (r = 1.0). (D) A classical CCH computed for a pair of signals generated by the methods in C. Both slow and fast components are visible. (E) Scaled correlation with s larger than the period of the slow component produces a cross-correlation plot of identical shape to the classical CCH. (F) When the scale is reduced to the period of the fast component, the contribution of the slow component to the CCH is removed and the correlation properties of the fast component are revealed much more accurately. The horizontal dashed line indicates the correlation inserted into the fast components, which should be recovered by the analysis.

series of adjacent segments of length s, and then coefficients of correlation are computed for each segment (Fig. 2A, left). The correlation coefficients are subsequently averaged across all segments in a trial, and these results are averaged next over repeated experimental trials (if any). In the present study, all of the averaging of correlations was made without Fisher's z-transformations of r-values (Fisher, 1915) (see a later Discussion and Appendices A and E on the pros and cons of using such transformations).

Thus, the scaled coefficient of correlation $\bar{r}_s$ at the scale s is given by

$$\bar{r}_s = \frac{1}{K}\sum_{k=1}^{K} r_k, \tag{8}$$

where r_k indicates the correlation computed for segment k, $k \in [1 \ldots K]$. K is the number of time segments that can be fit into the total length of the signal, T, without overlap between the segments, and is given by

$$K = \text{round}\left(\frac{T}{s}\right). \tag{9}$$

For example, for a total time period T = 2 s and segment size s = 25 ms (a value suitable for measuring synchrony supported by oscillations of 40 Hz or higher), the total time period will be divided into $K = 2000/25 = 80$ segments, and for each segment another value of r will be computed (later we also discuss cases in which some segments are skipped due to missing data). The duration of the segment is denoted as the scale of the analysis. Correlation coefficients other than r may also be used (e.g. Spearman's rho described in Appendix B).

As in the regular CCH, the above calculation also needs to be repeated for all necessary time shifts, φ, between the reference and the target. To obtain proper estimates of correlation for all possible shifts, it is necessary to segment the signals only after the target signal is shifted (as indicated in Fig. 2B). These repeated segmentation processes, i.e. following each shift, allow a scaled CCH to be computed legitimately for shifts that are larger than the scale of the

analysis, i.e. $\varphi > s$, which is the case for most of the CCHs shown in the present study (e.g. shifts of ± 80 ms with s = 30 ms). The final CCH is constructed simply by plotting $\bar{r}_s$ as a function of the time offset, φ (e.g. ± 80 ms in steps of 1 ms).

The chosen duration of the segments, s, is thus central to the present analyses as it determines the time scales of the signal components that will be preserved (the components with time scale ≤ s) or removed (those of the time scale > s). The segmentation process reduces the variance only for the slow components, whereas the fast components remain to be sampled with their full variance. This is illustrated in Fig. 2C, where the variability (vertical dispersion) of a composite signal is smaller within a 20 ms than within a 100 ms window ($\text{range}_1 < \text{range}_2$). See also the proof in Appendix C.

Mathematically identical cross-correlation functions are used to compute the correlogram for binary signals (e.g. action potentials) and continuous signals [e.g. local-field potentials (LFPs)] or for a combination of the two. However, traditionally, in statistics the correlations for each signal type (or their combination) have different names. For the pairs of binary signals, the measure is known as the phi coefficient of correlation and for the combination of binary and continuous signals, it is known as the point-biserial coefficient of correlation. Details on how Pearson's r can be calculated most speed-efficiently for various types of signals are given in Appendix B, which also provides information on how to access freely Matlab and C++ libraries for the calculation of SCA.

Example calculation of scaled correlation analysis

Example applications of SCA to simulated continuous signals are shown in Fig. 2C–F. Each of the signals in a correlated pair consisted of two oscillatory components, one component having a fast and the other a slow oscillatory rhythm (Fig. 2C) (usually, 5× ratio in oscillation frequencies). The examples in Fig. 2D–F were calculated with the maximal possible correlation between the corresponding slow components, r = 1.0, and a medium-range correlation between the corresponding fast components, r = 0.6.

The goal of the analysis was to accurately extract the correlation of the fast components and to remove, as much as possible, the contributions of the slow components. This was obviously not possible by computing a CCH without the use of scaled correlation (referred to here as classical CCH), as the resulting CCH was reflecting the superposition of the slow and fast components (Fig. 2D). A similar result was obtained when SCA was computed with a large scale that corresponded to the time scale (i.e. oscillation period) of the slow components (s = 100 ms) (Fig. 2E). In this case, the strength of correlation for the fast component at the zero phase offset was overestimated (r = ∼0.80 instead of 0.60) and the shape of the correlogram was the same as in the classical CCH (as in Fig. 2D), reflecting large contributions of the slow components. However, when the scale was shortened to the size that matched the time constant of the fast components (s = 20 ms), the contribution of the slow component was almost entirely eliminated despite its large correlation, and the shape of the resulting CCH was dominated by the fast components (Fig. 2F). The correlation estimated at the center peak (r = 0.5996) was very close to its true value. This illustrates that SCA with the scale larger than the time constants of signal components produces a correlogram of a shape identical to a classical CCH, whereas a decrease in the scale changes the shape to reflect the fast components.

How does scaled correlation analysis relate to a normalization of a cross-correlation histogram?

The absolute values on the y-axis depend on the voltages measured by field potentials or, in the case of spiking signals, on the firing rates of neurons. Hence, for comparisons, such CCHs often need to be normalized by various methods, which may involve fitting a Gabor function (König, 1994), or a computation of Pearson's r between spike trains (e.g. Lamme & Spekreijse, 1998). In the latter case, a single correlation coefficient is computed for the entire length of the analyzed signals, multiple r's being averaged only between repeated trials. Hence, this form of normalized CCH corresponds to scaled correlation with s equal to the length of the trial. As shown in Fig. 2D and E, given that s is equal to the total length of the signal, SCA returns a CCH of the shape that is identical to the classical CCH (proof provided as a part of Appendix D). Normalizations based on non-segmented data do not remove the slow components. They act only as scaling factors for the y-axis of the CCH. To remove a slow component from correlation it is necessary to use segments smaller than the period of that component. In that way, the correlation at each segment is normalized prior to the averaging.

Mathematical theory of scaled correlation

In Appendix E we provide a proof that, if all segments have identical means and variances, the average Pearson's correlation obtained with small s is identical to the value of the single correlation coefficient computed across the entire signal (i.e. with maximum segment size, s). This result justifies the averaging procedure in Eqn 8. Note also that this identity does not hold if Fisher's z-transformation is applied to the averaging process.

In Appendix D, we show analytically that scaled correlation removes the contribution of the slow components by estimating its contribution through a form of low-pass filtering and subtracting it from the total correlation. In effect, scaled correlation produces implicit low-pass filtering that is similar to a moving average filter.

Analysis of simulated signals

Pairs of continuous signals

By using simulations similar to those shown in Fig. 2C–F, we investigated systematically the importance of, and the dependencies between, the following three factors: (i) the segment size, s, (ii) the ratio between the amplitudes of the fast and the slow components, and (iii) the ratio of the oscillation frequencies of the two components.

To investigate the application on continuous signals we used both signals with a single oscillatory component (Fig. 3A and B) and signals composed of two oscillatory components (Fig. 3C–E). The composed signals were generated as illustrated in Fig. 2C (see also Materials and methods). Each data point in Fig. 3A–E is calculated on a 5 s simulated signal and, to focus the analysis on the magnitudes of correlations, only the values of the center peaks of scaled CCHs (i.e. φ = 0) are shown in these plots. In addition, representative complete CCHs are shown in Fig. 3F–J.

Single component. The relationship between segment size and measured strength of correlation was investigated for two different oscillation frequencies of a single oscillatory component (50 and 10 Hz) and four levels of correlation strength (r = 0.2, 0.4, 0.6 and 0.8; color coded). In Fig. 3A and B it can be seen that the correlations were estimated correctly only if the segment sizes were equal to or larger than the period of the oscillatory rhythm (dashed vertical lines). If the segment sizes were smaller than that, the correlations were always reduced, the magnitude of the reduction depending on the degree of discrepancy between the period of the

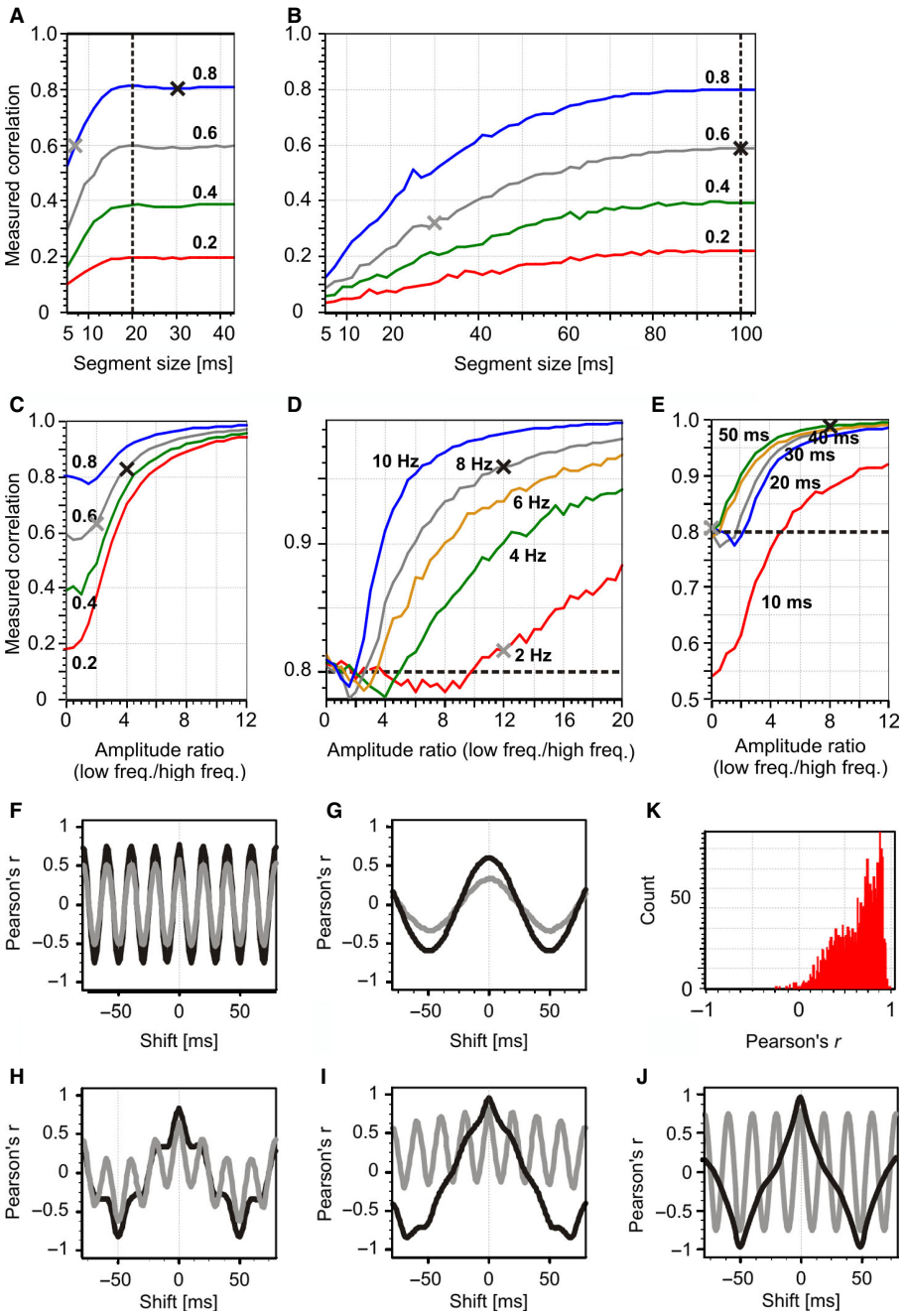


FIG. 3. Analysis of the degree to which SCA attenuates the contributions of slow components in simulated continuous signals. Signals have one oscillating component of either 50 Hz (A) or 10 Hz (B). The color code indicates the strength of correlation inserted into the signals (r = 0.2–0.8) and the vertical dashed line indicates the segment size that corresponds to the period of the oscillation cycle. (C–E) Signals with two oscillatory components composed as described in Fig. 2C. (C) The strength of correlation between high-frequency components (50 Hz) was manipulated (color coded), whereas the correlation between the slow components (10 Hz) was kept constant (r = 1.0) (s = 20 ms). (D) The frequency of the slow component was manipulated (color coded), whereas the correlation between fast components was kept constant (r = 0.8; dashed line) (s = 20 ms). (E) Scale size was manipulated (color coded), the correlations and frequencies of slow components being constant (r = 0.8 and f = 10 Hz, respectively). The crosses indicate data points for which cross-correlograms are shown in F–J. (F) CCHs for the data points in A. Gray, s = 7 ms; black, s = 30 ms. (G) CCHs for the data points in B. Gray, s = 30 ms; black, s = 100 ms. (H) CCHs for the data points in C. Gray, amplitude ratio = 2; black, amplitude ratio = 4. (I) CCHs for the data points in D. Gray, f(slow) = 2 Hz; black, f(slow) = 8 Hz. (J) CCHs for the data points in E. Gray, amplitude ratio = 0; black, amplitude ratio = 8. (K) Distribution of r-values across 250 segments of duration s = 20 ms used to calculate the center peak of the CCH in H (gray). For interpretation of color references in figure legend, please refer to the Web version of this article.

signal component and the segment size. Importantly, a small discrepancy between the scale and the signal period does not much affect the estimated correlation – a property of SCA that should be considered when choosing the segment size and when interpreting the results of an analysis.

Two components. Signals composed of two oscillatory components consisted of varying degrees of correlation (r_{FAST} = 0.2–0.8) for the fast component (50 Hz) and fixed correlation (r_{SLOW} = 1.0) for the varying frequencies of the slow component (2–10 Hz). A variable that strongly affected the measured correlation was the ratio between the

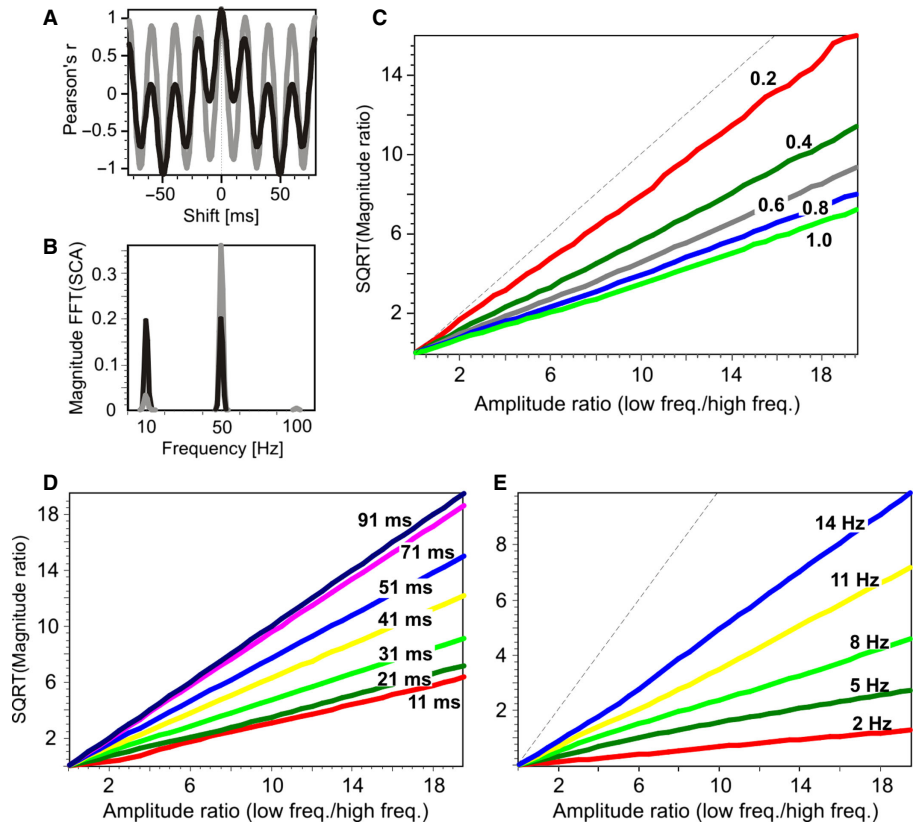


FIG. 4. Analysis of the magnitude spectrum of scaled cross-correlograms. (A) Scaled correlograms computed with long (black, s = 100 ms) and short (gray, s = 20 ms) scale for combined signals at 11 and 50 Hz of equal amplitude. (B) Magnitude of the FFT for the correlograms in A. (C) Square root (SQRT) of the ratio between peak sizes from B at 11 and 50 Hz, shown as a function of the ratio between the amplitudes of the slow and fast component (x-axis) and the strength of correlation between the fast components (color coded). (D) Instead of the correlation strength as in E, the sizes of the scale were manipulated (color coded). (E) Similar to C and D but the frequency of the slow component was manipulated (color coded). For interpretation of color references in figure legend, please refer to the Web version of this article.

amplitudes of the slow and fast components, and this is indicated on the x-axis of all plots. If not specified otherwise, the scale of the analysis was s = 20 ms.

The finding that the relative amplitude of components strongly determines the measured correlation is shown in Fig. 3C, where the frequency of the slow component was set to 10 Hz. SCA attenuated the slow components well when their amplitude was similar to that of the fast components. However, when the slow component had extremely large amplitudes (i.e. > ~4× larger), the measured correlation remained dominated by the slow components.

This result did not depend much on the true correlations between the fast components (Fig. 3C), each value of r being affected in a qualitatively similar way. However, the relative difference in the oscillation frequencies was an important variable (Fig. 3D). As this difference increased, increasingly larger differences in the amplitude could be absorbed by SCA and r_{FAST} could be estimated accurately. For example, 2 Hz components could have as much as 12× larger amplitude than that of 50 Hz components and still not prevent an accurate estimation of the correlation between the 50 Hz components (red curve in Fig. 3D).

Finally, the choice of the scale, s, affected the measured correlations between multi-component signals similarly to the single-component signals. To recover the original correlation between the fast components (r = 0.8), it was necessary that the scale was close to the oscillation period of that component (ideally, 20 ms for 50 Hz; Fig. 3E). However, the scale should not be much smaller, as it then also attenuates the correlation between the fast components (red line in Fig. 3E).

Representative examples of auto-correlograms and cross-correlograms are shown in Fig. 3F–J, which correspond to the data points marked with crosses in Fig. 3A–E, respectively. The choice of scale, s, does not affect the overall shape of the CCH when signals have only one component, but only the magnitudes of measured correlations are affected (Fig. 3F and G). In contrast, for multi-component signals the shapes of CCHs can change drastically as a function of scale. As the contributions of slow components to the measured r reduce with a decrease in s, so does their contribution to the shape of the correlogram (Fig. 3H–J).

Even when the average $\bar{r}_s$ is computed accurately it is a result of many r-values that can vary considerably across individual segments. For continuous signals, these distributions are also unimodal. An example distribution is shown in Fig. 3K, which has an average of 0.65 and corresponds to the center peak of the CCH in Fig. 3H (gray curve).

The contributions of slow and fast components to a CCH can be quantified by computing the fast Fourier transform (FFT) of a correlogram (here, this was done by computing correlograms for lags between ± 512 ms and considering the leftmost 1024 out of 1025 points for the FFT). For example, scaled correlograms applied to two-component signals with s = 100 and s = 20 ms in Fig. 4A (only parts of the correlograms are shown, spanning ± 80 ms); the magnitude spectrums are shown in Fig. 4B. With a long segmentation window, the magnitude of the fast and slow component is about the same (black curve). However, with the segmentation window that favored the fast component, the magnitude of the fast

component became about an order of magnitude higher than for the slow component (gray line).

The sizes of the peaks in the magnitude spectrum of the correlogram scale with the square of the amplitude of the slow and fast components (assuming that the target and reference have components of identical amplitudes). This follows from the Wiener–Khinchin theorem, which states that the Fourier transform of the auto-correlation function is equivalent to the power spectrum. Therefore, the magnitude spectrum of the cross-correlation is equivalent to the cross-power spectrum [described by Blackman & Tukey (1958)] and, thus, the magnitude peaks scale with the square of the amplitude of signal components. In scaled cross-correlograms, changes in the magnitude spectrum reflect the changes in the shape of the CCH.

To investigate systematically the relation between the spectral magnitude peaks of the correlogram, the properties of the signals, and the scale of the analysis, we plotted the ratio between the amplitude of the signals vs. the square root of the ratio of their magnitudes at 11 and 50 Hz (the two frequencies were chosen not to be integer multiples of each other to avoid harmonics in the magnitude spectrum from interfering with the calculations of peak ratios) as a function of correlation between the fast components (with s = 21 ms; Fig. 4C), segment size (with $r_{FAST} = r_{SLOW} = 1.0$; Fig. 4D), and frequency of the slow component (with s = 21 ms and $r_{FAST} = r_{SLOW} = 1.0$; Fig. 4E).

The relationships between the two types of ratios were linear in all cases, as indicated by straight lines in Fig. 4C–E. Thus, the square root of the ratio between the FFT magnitudes in the scaled CCH was always a linear function of the ratio of the amplitudes of the signal components. The correlations, segment sizes and frequencies affect only the slopes of these linear relationships. The less SCA removes the slow component, the closer is this slope to 1.0 (indicated by dashed lines in Fig. 4C–E). The slope approaches 1.0 as the correlation of the fast component approaches 0.0 (Fig. 4C), as the segment size approaches the period of the slow component (Fig. 4D), and as the frequency of the slow component approaches that of the fast component (Fig. 4E). This analysis indicates that, in continuous signals, SCA operates well across different parameter values and can hence be used as a versatile tool for removing slow components from CCHs.

Pairs of binary signals

Unlike the above examples of continuous signals, in which signal components were based on oscillatory patterns of given frequencies, we generated the spike events using non-homogeneous Poisson processes, i.e. probabilities modulated by a certain envelope function (half-sinusoidal in our simulations, Fig. 5A) (see also Materials and methods). The resulting spike trains can be represented as sequences of 1's and 0's (Fig. 5B), indicating the presence and absence of spikes within each time bin (we used bins of 1 ms in all cases). Therefore, the oscillatory components are not inserted into the signal directly but, rather, ''modulate'' another high-frequency signal. See Materials and methods for more details on generating artificial spike trains for the present analyses.

When Pearson's r is applied to binary signals it is known as the φ (phi) coefficient of correlation. Calculation can be made quickly based on a contingency table containing the counts of all four possible combinations of spiking events across the bins (Table B1 in Appendix B): those that contribute to positive correlations, either both neurons fire or both neurons are silent (entries b and c, respectively); and those that contribute to negative correlations, only one neuron fires and the other is silent (entries a and d). The resulting correlation is a net effect of those four counts, as described in Eqn B3 (Appendix B).

Empty segments. When computing scaled correlation between pairs of binary signals it is necessary to ensure a proper treatment of empty segments of the spike trains. As the duration of the segments reduces, inevitably, within many segments, one or both spike trains will not have a single spike (Fig. 5B). In those cases, φ cannot be calculated due to a division by zero (i.e. in Eqn B3, Appendix B, either the term $c + d$ or $a + c$ is 0). In other words, a signal without spikes does not have variance and, without variance, no correlation can be defined.

A mathematically proper approach is to define the value of correlation for this segment as not-available or not-a-number (e.g. indicated as NaN in Matlab), which means that these segments should not be taken into the average. Consequently, the number of averaged segments is smaller than the total number of segments K. The smaller the s, the larger proportion of segments is expected to be removed from the analysis. This can be seen from the fact that, for a Poisson process, the probability of observing no events in a given interval drops exponentially as a function of the size of the interval and the rate of the process.

The alternative would be to assign zero correlation values to these segments. This approach would not change the shape of a CCH but would affect the magnitudes of estimated correlations, the included zeros reducing the average φ considerably (see Fig. 5C). If zero segments were kept in the analysis, the averages would favor higher firing rates over lower ones and a direct comparison between scaled correlograms would become similar to a comparison of the absolute counts of coincident events between classical CCHs. Such correlations would largely reflect the relative duration of the periods of activity and non-activity for the given neurons, which is not interesting for most analyses. Hence, with the exception of Fig. 5C, all of the analyses presented here were made by removing from the averages the segments that lacked spikes, and this is how we define SCA for binary signals.

By using simulations similar to those shown in Figs 3 and 4, we investigated systematically the following four factors: (i) the segment size, s, (ii) the ratio between the amplitudes of the probability functions in Fig. 5A of the slow and fast components (which is a counterpart of the amplitude in continuous signals), (iii) the ratio of the oscillation frequencies of the two components, and (iv) the combined firing rate of the two spike trains.

Single component. With a non-homogeneous Poisson spike process, the correlation values are expected to be much lower than those obtained with continuous signals (Macke *et al.*, 2009). This is due to the nature of spiking signals, where a very small bin size (e.g. 1 ms) makes the probability of observing coincidences modulated by the signal of interest (e.g. the 60 Hz probability function) small, inducing a correlation smaller than the normalized co-variance of continuous signals. The positive contributions by spikes (i.e. entry b in Table B1 in Appendix B) require the events in both spike trains to occur in the same bin, the likelihood of which event is << 1.0 even if the firing probability functions in Fig. 5A are correlated perfectly. Thus, the count of negative contributions (entries a and d in Table B1 in Appendix B) is likely to be non-zero.

We investigated the relationship between the segment size and measured strength of correlation by simulations for two different oscillation frequencies of a single oscillatory component (50 and 10 Hz). For each component, we used two levels of correlation strength (color coded). As found for continuous signals, the correlations in binary signals were also estimated accurately only if the

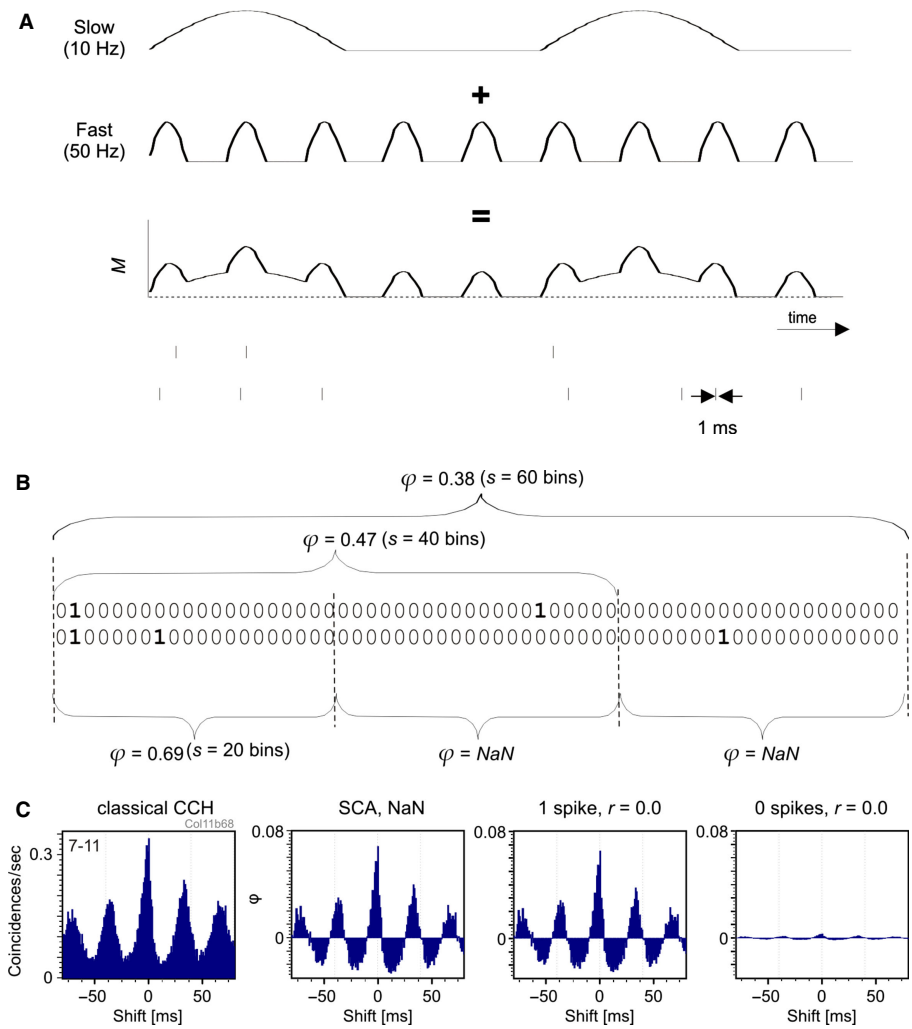


FIG. 5. Analysis of scaled correlation for binary signals. (A) For simulations in Fig. 6, each spike train is created from a non-homogeneous probability function that is created by adding two half-sinusoidal components: a slow one and a fast one. The probability functions of the pairs of spike trains always had a correlation of $r = 1.0$. (B) A segment in which one or both spike trains lack spikes (empty segment) needs to be assigned a not-a-number value (*NaN*) and taken out of analysis. (C) Assigning zero to empty segments does not change the shape of the CCH but reduces the correlation strength. Far left: classical CCH. Middle left: scaled correlation with $s = 40$ ms and with assigning *NaN* to empty segments. Middle right: the same scaled correlation but assigning zero only to segments that lack spikes in only one of the trains (when both spike trains are empty for the segment, the latter is discarded). Far right: zero is also assigned to segments that lack spikes in both spike trains. If the neurons are silent for a considerable amount of time (i.e. low firing rates), as is the case in this example, assignment of zeros reduces the magnitudes of correlations considerably.

segment sizes were equal to or larger than the period of the oscillatory rhythm (Fig. 6A and B). As expected, the correlations reduced when the size of the segment reduced to a sufficient degree.

There was one difference to the results with continuous signals. When the firing rates were relatively low (≤ ∼20 spikes/s), i.e. in the range of cortical single units, and the segment sizes were small (< ∼40 ms), i.e. corresponding to the periods of gamma and beta oscillations, the measured correlation values did not necessarily decrease with the reduction in segment size but sometimes exhibited an increase (Fig. 6A, green and red traces). The reason for that increase in correlation strength lies in the nature of modulated high-frequency signals, i.e. as the small segment sizes increase the number of segments not included in the analysis, they also increase the number of spikes disposed from the analysis. Notably, this disposal is selective for spikes that provide negative contributions to the computed correlation (entries *a* and *d* in Table B1 in Appendix B), as illustrated in Fig. 5B and, consequently, the number of positive contributions (entries *b* and *c*) relatively increases – leading to higher correlations. In other words, in terms of a reduction of sample variability, SCA achieves variance reduction by removing spikes that do not have matches in the counterpart spike trains – the variance of any binary signal being a direct function of the number of the events, i.e. var = $P(1 - P)$, where P indicates a probability to encounter an event and is given by c/n, where c is the number of events in a binary series of the length n.

Two components. In spiking signals it is not possible to create two binary signals, each with its own coefficient of correlation and then sum them up to produce a composite signal that would maintain both correlations, each at a different time scale. This is because, in spikes, all components (slow and fast) are represented by binary events, only the underlying probability functions being continuous. As a consequence, an analysis made for continuous signals in Fig. 3C–E cannot be defined for binary signals. Rather, we focused the analyses on the shapes of CCH (as in Fig. 4), as the shapes reflect the statistical properties of the underlying probability functions. In Fig. 6C–E we

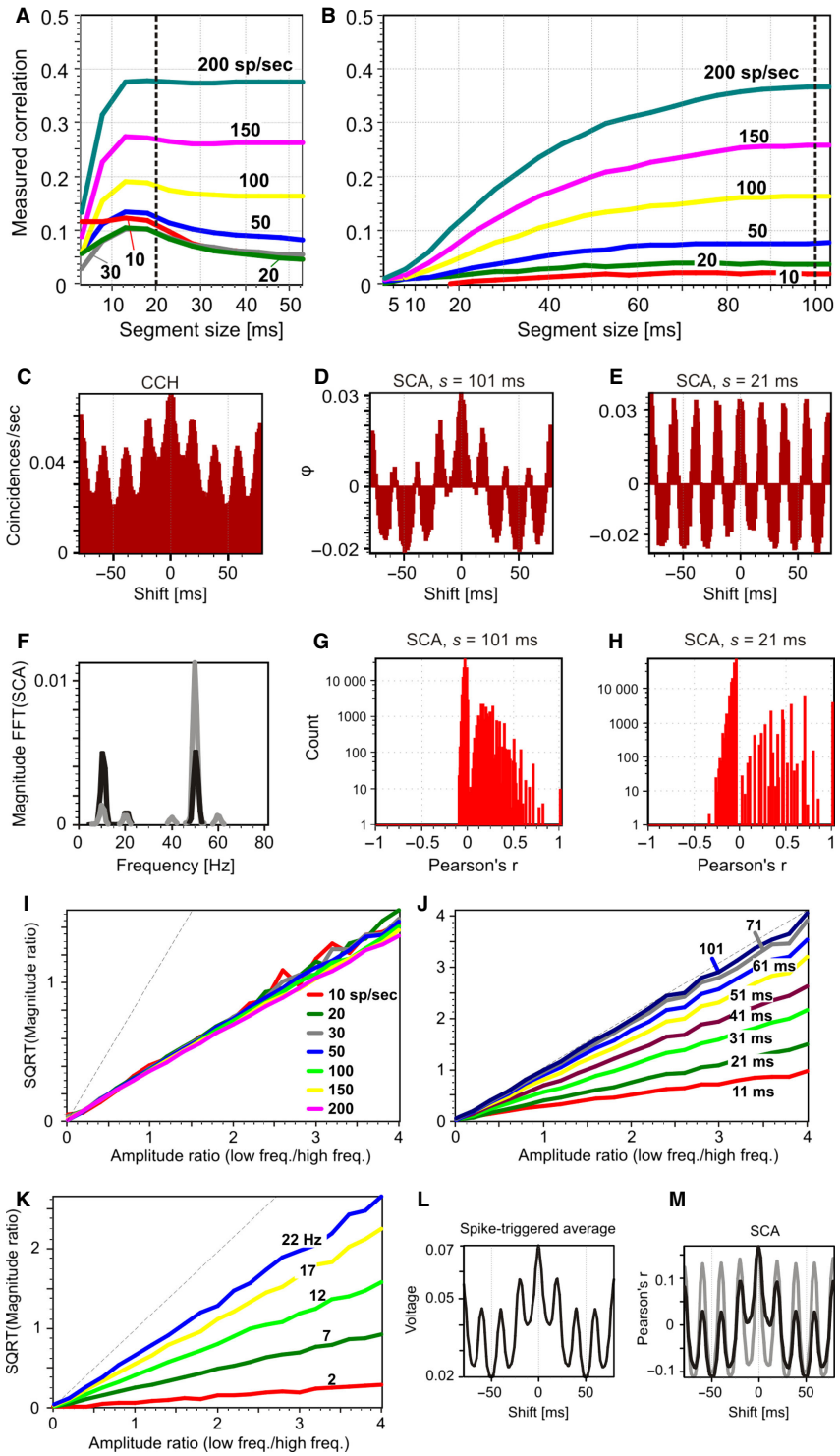


FIG. 6. Analysis of the degree to which SCA attenuates contributions of slow components in pairs of binary signals (A–K) and in combinations of binary and continuous signals (L and M). Binary signals have one oscillating component of either 50 Hz (A) or 10 Hz (B). The color code indicates the combined firing rate of the spike trains (10–200 spikes/s) and the vertical dashed line indicates the segment size that corresponds to the period of the oscillation cycle. (C–E) An example effect of scaled correlation on the shape of CCH that contains two components, slow and fast (frequencies of 11 and 50 Hz, respectively). (C) Original CCH. (D) Scaled correlogram with s = 101 ms. (E) Scaled correlogram with s = 21 ms. (F) Magnitude of the FFT for the correlograms in D (black) and E (gray). (G and H) Distribution of r-values in D and E, respectively. (I–K) Analysis of the magnitude spectrum of scaled cross-correlograms by computing the square root (SQRT) of the ratio between peak sizes as illustrated in F at 11 and 50 Hz, shown as a function of the ratio between the amplitudes of the slow and fast component probability functions (x-axis). Color codes were used to indicate different combined firing rates (I), segment sizes (J), and the frequency of the slow component (K). (L–M) The same analysis as in C–E but for combinations of continuous and binary signals with 10 and 50 Hz components. (L) Spike-triggered average of the continuous signal (equivalent to a CCH between the two signals). (M) Scaled correlation computed with s = 101 ms (black) and s = 21 ms (gray). For interpretation of color references in figure legend, please refer to the Web version of this article.

show an example of a change in the shape of CCH depending on the choice of the scale of analysis. The original CCH for a combination of 10 and 50 Hz signals of equal amplitude (combined firing rate = 40 spikes/s) indicates the presence of both components (Fig. 6C) and so does SCA with large scale, which reveals the identical shape (Fig. 6D). Importantly, SCA computed with short segments (s = 21 ms) removes the slow, revealing the fast component (Fig. 6E), much like the analysis of continuous signals (e.g. Fig. 4A). Also, the analysis of the magnitude of the FFT of the correlogram on these two scaled correlograms indicates an increase in the relative magnitude of the fast component and concomitant reduction in the magnitude of the slow component (black vs. gray trace in Fig. 6F), much as was the case for continuous signals (Fig. 4B).

For binary signals, the distributions of individual r-values are bimodal (Fig. 6G and H), with the cutting point at $r = 0$. With low firing rates, a segment is likely not to contain a single coincident event, in which case the correlation of the segment will have a small negative value. In the case that a segment contains a coincidence, the correlation will be most likely to be positive, and it may cover any value between 0 and 1.0, including the perfect correlation of $r = 1.0$ (when this coincidence is the only event within the segment).

We made a detailed analysis of the effects of scale size on the shape of the correlograms by using the same method based on FFT as reported for continuous signals. We plotted the amplitude ratio between probability functions of slow and fast components (Fig. 5A) against the square root of the magnitude ratio in FFT and systematically manipulated the firing rates of units (11 and 50 Hz; s = 21 ms; Fig. 6I), the segment size (firing rate = 40 spikes/s; Fig. 6J), and the frequency of the slow component (firing rate = 40 spikes/s; Fig. 6K). Again, the relationships between the two ratios stayed linear for all parameter values. Moreover, the results of the analysis were fully independent when the firing rates were changed (rates were changed by an equal factor for the slow and fast components; Fig. 6I). Much as for the continuous signals, the slope of the linear relationship for binary signals approached 1.0 as the scale approached the period of the slow component (Fig. 6J), and as the frequency of the slow component approached that of the fast component (Fig. 6K). The results indicate that, when the shape of the CCH is of concern, SCA removes slow components from binary signals in much the same way as from continuous signals.

Combinations of continuous and binary signals

The SCA can also be applied to analyses that correlate a binary signal with a continuous signal (e.g. LFP correlated to spiking activity) – also known as a spike-triggered average (Gray & Singer, 1989). When Pearson's r is computed with such combined signals, it is called the point-biserial coefficient of correlation and is described in Appendix B.

The results obtained in the analyses of those combined signals were very similar to and fully consistent with those obtained for continuous and binary signals separately. In Fig. 6L and M we show examples of the classical CCH and scaled CCH made with long (s = 101 ms) and short (s = 21 ms) scales. The size of the scale has the same effect as shown previously in Figs 4A and 6C–E for continuous and binary signals, respectively. For the sake of brevity, other analysis results are not shown.

Applying scaled correlation analysis to neurophysiological data

In Fig. 7A–J we show examples of CCHs computed from neuronal spiking activity recorded from the cat visual cortex under anesthesia and in response to visual stimulation (see Materials and methods for details). These examples compare the use of SPs with the use of SCA. In Fig. 7A, an example of a CCH is shown with a flat SP (i.e. no rate co-variation), and consequently the shape of the correlogram obtained with SCA (Fig. 7F) was very similar in many details to the classical one, also when the SP is subtracted (CCH − SP). Moreover, the detailed structures of the CCHs, i.e. the relative sizes of the neighboring bins, remained largely preserved across all of the plots. This illustrates that, without rate co-variation, SCA produces the same shape of a CCH as the classical approaches – including all of the details on the bin-by-bin basis [see Appendix F for the estimation of significance of averaged correlation coefficients in SCA based on the method of fixed effects by Hedges & Olkin (1985)].

In Fig. 7B we show an example of a CCH that consists only of slow components. The graph indicates slow rate co-variation without a narrow center peak. The SP was not fully successful in removing these slow components, suggesting that the slow rate co-variation was not time-locked perfectly to the stimulus onset – a problem that the SP cannot resolve. In contrast, SCA with s = 40 ms removed this rate co-variation much more efficiently, as indicated by a near-flat CCH (Fig. 7G). Also, as shown in Fig. 7C, SCA can remove rate co-variation even when the SP is completely flat and, thus, when rate co-variation is fully independent of the timing of the stimulus. In this case, the classical CCH exhibits a wide peak and the SP is nevertheless flat. With s = 40 ms, SCA removed this broad peak almost entirely (Fig. 7H), which should be contrasted to the failure to do so by a subtraction of the SP from the classical CCH.

In Fig. 7D and E, we show two example CCHs for the same pair of units but in response to different stimuli (grating and plaid stimuli, respectively, see Materials and methods). Here, an estimate of the change in the strength of synchrony is hindered by the slow rate co-variation in the original CCHs. SPs also do not help much. Only SCA is helpful because – as judged by the shape of CCHs – it removes the slow components much more efficiently than does the SP (Fig. 7I and J). In addition, Fig. 7D illustrates a real-world example in which SCA can lead to a switch in the sign of correlation; the troughs of the oscillation that are positioned at half the period of the oscillation cycle indicate negative correlations detectable only when SCA is applied (e.g. Fig. 7I). This is not always achieved by subtraction of the SP as troughs remain positioned above the baseline of the CCH, which is another indicator that SPs do not account fully for the rate co-variation.

To illustrate further the versatility of scaled correlation in practice with spiking data, four examples of CCHs with unusual shapes and the corresponding SCA (s = 30 ms in all cases) are shown in Fig. 7K–N. Irrespective of the shape, scaled correlation removes the slow components and retains the fast components.

These results indicate that the conclusions obtained from simulations also apply to real data; if a classical CCH indicates large rate co-variation, SCA will remove these slow components but will retain detailed information about the fast components.

Multi-scale analysis

Further analysis of the non-stimulus-locked rate co-variation in Fig. 7C revealed that the underlying source is an oscillatory rhythm in the theta range, which is shown in detail in Fig. 8A. The theta component is visible when the cross-correlation window is expanded to ± 1000 ms. The resulting plots show an oscillatory rhythm of ~5 Hz. Interestingly, this is not the only source of slow activity in this CCH. Theta activity is superimposed on top of yet another slow component that had a time period of > 1 s. The SP captured this

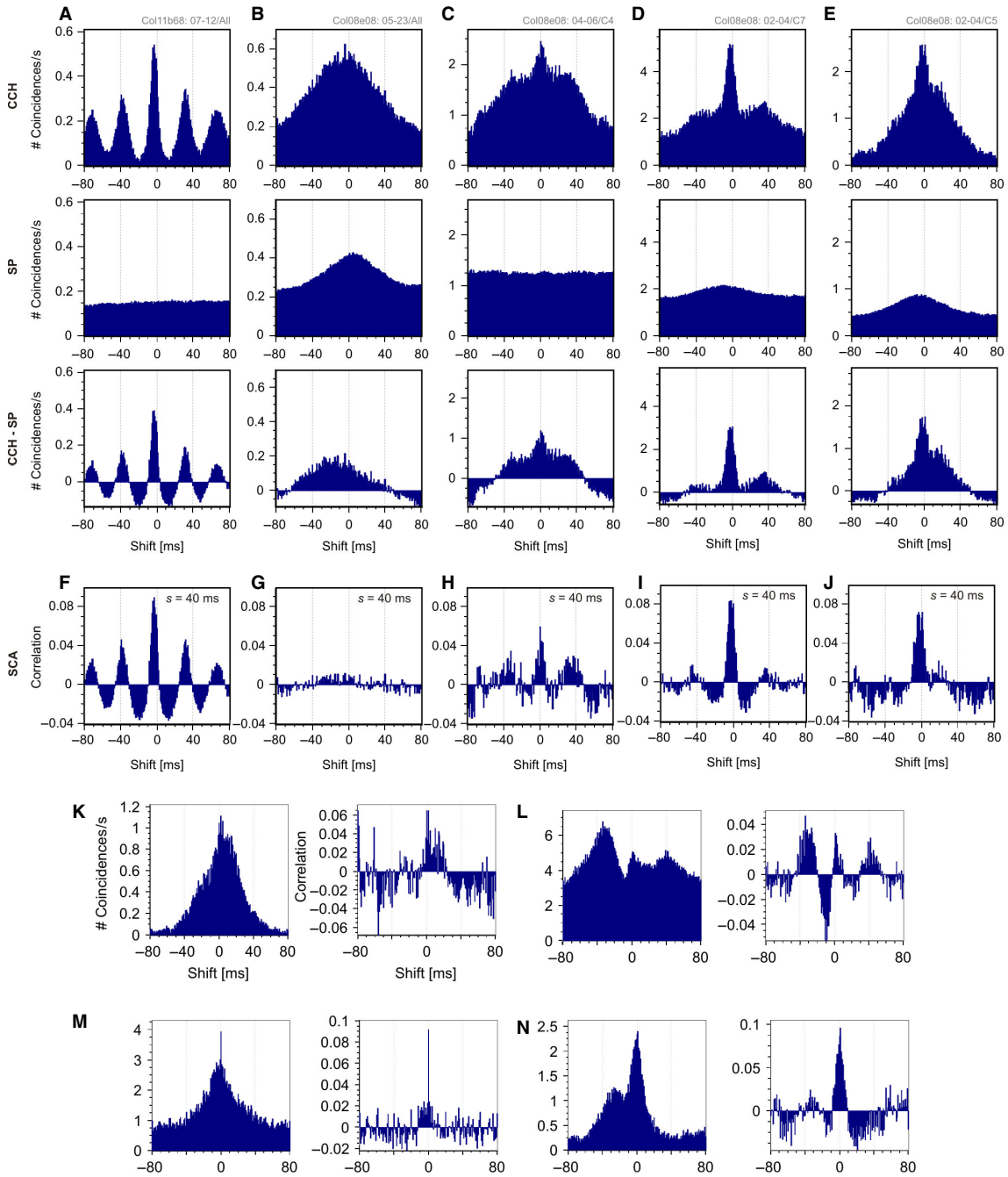


FIG. 7. Example application of scaled correlation to spiking signals recorded from the cat visual cortex. (A–E) CCH, SP and the difference between the two (CCH − SP) shown for five example pairs of multi-unit activity recorded from the cat visual cortex. (A) An example of precise neuronal synchrony (large center peak) supported by strong oscillatory activity in the beta range (large satellite peaks at ∼35 ms phase shift). (B) Example CCH that does not indicate neuronal synchronization but only slow rate co-variation as indicated by the wide center peak and the elevated SP. This result is typically obtained when CCH is computed between spike trains that include strong transient changes in rate responses due to, e.g. an abrupt onset of the stimulus (the so-called ON-responses). (C) Example CCH in which the center peak contains both a narrow component indicating neuronal synchrony and a much slower wide component. In this example, the slow component is not time-locked to the stimulus, as indicated by a flat SP. (D and E) The CCHs are computed for the same pair of units but in response to two different stimuli, illustrating the difficulty in determining how synchronization strength changes as a function of stimulus properties. (F–J) Application of SCA to the CCHs shown in A–E by using the scale s = 40 ms, which is suitable for extraction of signals supported by an oscillation frequency of 25 Hz or higher. (K–N) Further examples of CCHs of unusual shapes illustrating that SCA maintains the detailed structure of the correlogram at fast time scales. In all cases s = 30 ms.

slowest component well, the oscillatory pattern in the theta range being revealed accurately after its subtraction from the original CCH (CCH − SP, Fig. 8A). This indicates that this slow component is locked to the stimulus. As would be expected, SCA revealed the same oscillatory pattern when computed with the same cross-correlation window (± 1000 ms) and with an appropriate scale (s = 200 ms) (Fig. 8A, SCA). Again, SCA and CCH − SP match in much of the detailed relative relations between the neighboring bins, indicating that SCA is an accurate method for removing the slow components from CCHs.

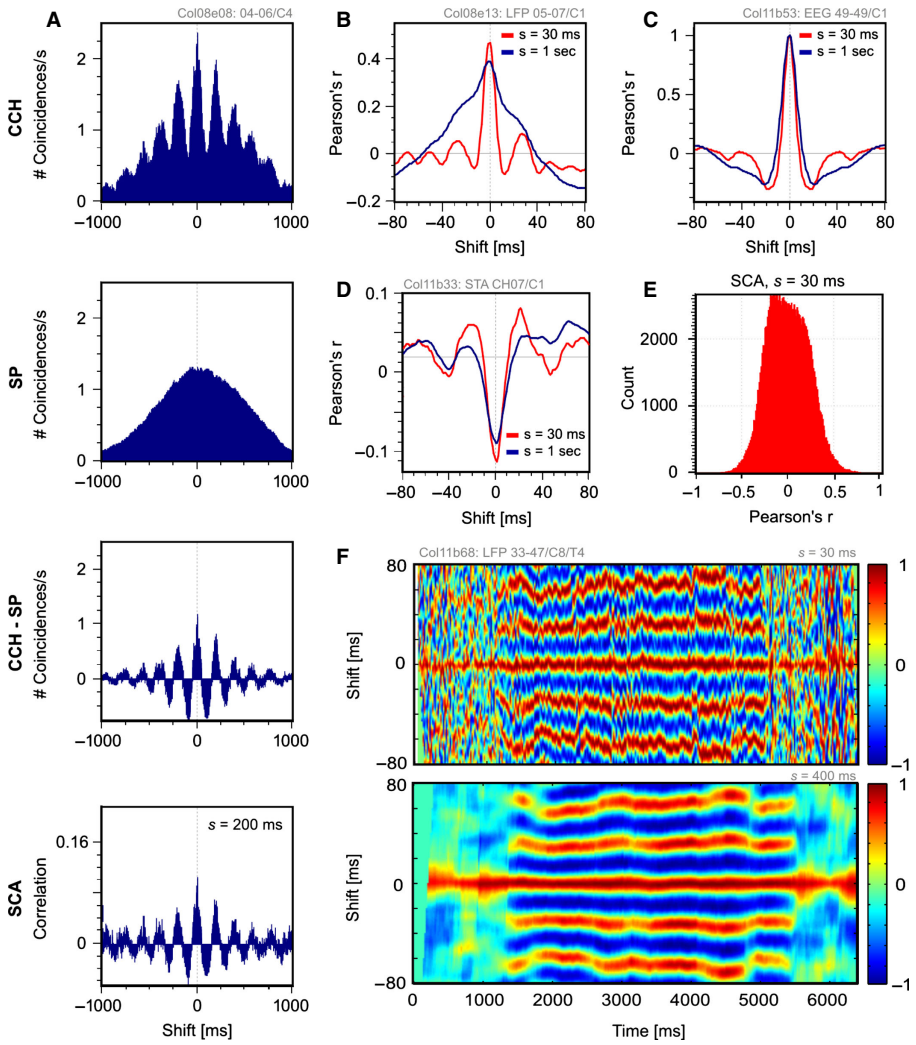


FIG. 8. Example of a multi-scale analysis and application of SCA to real continuous signals. (A) A raw CCH, SP, their difference and SCA computed for the same data as in Fig. 7C except that the cross-correlation window is ±1000 ms and the scale s = 200 ms. (B–D) SCA computed on signals other than pairs of spike trains. (B) SCA for continuous data obtained by recording LFPs computed with large (s = 1000 ms; blue) and small (s = 30 ms; red) scale. (C) The same as in B but calculated on an electroencephalogram. In this case an auto-correlation histogram is computed. (D) Scaled correlation computed for one continuous (LFP) and one binary signal. (E) The distribution of r-values of the scaled correlogram in D (s = 30 ms). (F) Intensity plot of sliding window analysis with different scales. SCA for a pair of LFP signals over a ∼6.5 s period computed either with s = 30 ms (upper graph) or s = 400 ms (lower graph) with sliding windows of a size equal to the scale (i.e. 30 or 400 ms) that were slid in steps of 20 ms. The stimulus was presented at 1000 ms and was removed at 5000 ms. For interpretation of color references in figure legend, please refer to the Web version of this article.

Local-field potentials and electroencephalogram signals

Example applications of SCA to continuous signals are shown for LFPs (Fig. 8B, CCH) and electroencephalogram signals (Fig. 8C, auto-correlation histogram). Similar to the application on spiking activity, the correlations between slow components of continuous signals are removed efficiently and the dynamics of fast components is revealed when small scales are chosen.

In Fig. 8D we show example spike-field correlograms with SCA computed at two different scales. Here, for small values of s, stronger correlations are detected than for large values of s, indicating that the slow components attenuated the correlations produced by the fast components. The distribution of r-values is shown in Fig. 8E.

Sliding window analysis is also a popular type of application of CCH to data analysis where the calculations are made over small consecutive time windows and it reveals in this way how synchronization and/or oscillations evolve over time (e.g. along an experimental trial) (e.g. Roelfsema *et al.*, 1997). Sliding window analysis can also be performed with SCA and an example for a pair of LFP signals is shown in Fig. 8F. Here a window of size equal to the scale was slid in steps of 20 ms. For each window, scaled correlation CCH was computed and plotted as an intensity plot. When sliding window analysis is performed with a small scale (s = 30 ms, top), many details of the underlying dynamics can be recovered in the plot. These details are hidden in the plots computed with a large scale (s = 400 ms, bottom) or with a classical CCH (not shown).

Comparison to spectral methods

Spectral techniques are widely used to study different frequency components of neurophysiologic signals (Jarvis & Mitra, 2001; Pesaran *et al.*, 2002) and have also been used to address issues similar to those for which SCA has been designed. Thus, it is important to know how SCA compares with analyses based on various spectral techniques. We first investigated the degree to which spectral techniques can be used as an alternative to SCA for removing the slow components from a CCH. Next, we compared the performance of

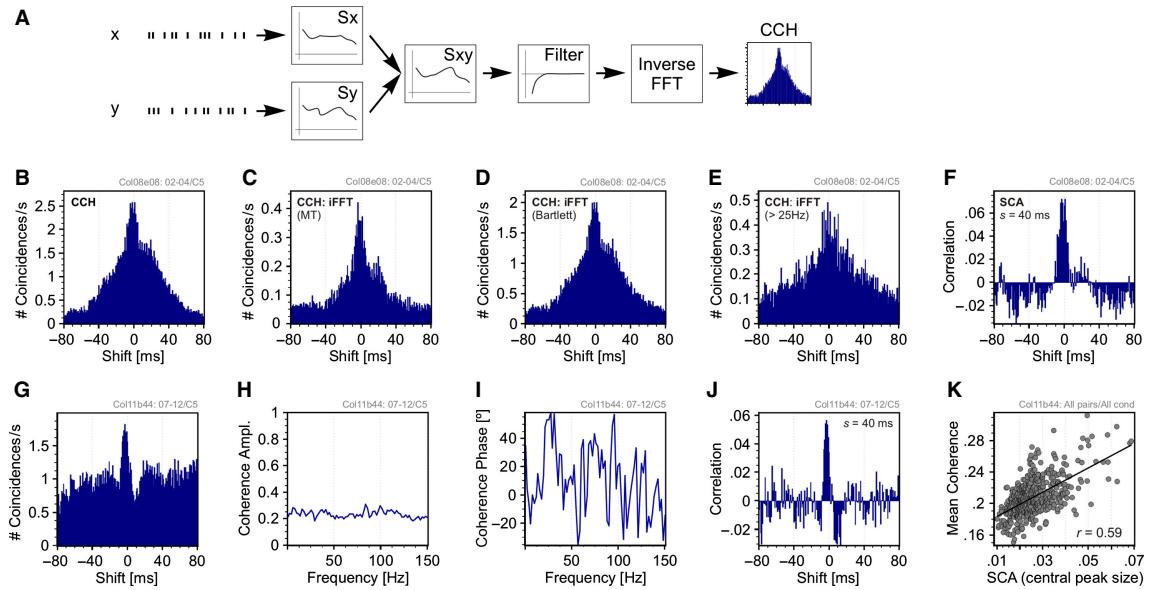


FIG. 9. Comparison between correlation analysis and spectral techniques. (A) Schematic representation of computing a filtered CCH on binary signals. Sx, Sy and Sxy are the spectra of the first, x, and second, y, spike train, and the cross-spectrum, respectively. (B) A CCH with a fast and slow component. The CCH in B is reconstructed by reversing the cross-spectrum (iFFT) estimated with multi-taper (MT) (C), or rectangular Bartlett window (D). (E) The CCH in B filtered by reversing a high-passed cross-spectrum estimated with the Bartlett window. (F) CCH in B computed with scale correlation, s = 40 ms. (G) An example CCH with very weak oscillatory component. (H) Coherence magnitude and (I) coherence phase for the data used to compute the CCH in G. (J) Scaled correlation corresponding to data in G. (K) Correlation between the size of the central peak in SCA (computed with s = 6000 ms) and average coherence, computed for all pairs of cells and all stimuli in a data set with weak oscillations.

coherence techniques as measures of synchronization with those based on cross-correlation computed with scaled correlation.

Removing slow components from a cross-correlation histogram by filtering

One potential approach for removing slow components from pairs of signals comes from the Wiener–Khinchin theorem, from which it follows that the CCH is the inverse Fourier transform of the cross-spectrum (Blackman & Tuckey, 1958). The cross-spectrum can be computed by multiplying the Fourier spectrum of the first signal with the complex conjugate of the Fourier spectrum of the second signal (Jenkins & Watts, 1968). Thus, for a CCH between binary signals x and y, one may attempt to remove the slow components from the CCH by first computing the Fourier spectra of the two signals (S_x and S_y), next computing the cross-spectrum between the two signals (S_{xy}), and then by filtering the cross-spectrum (by setting the low-frequency components to zero). Finally, the 'filtered' CCH is computed as the inverse FFT of the 'filtered' cross-spectrum. See Fig. 9A for the computational steps required for such a calculation.

To obtain a smooth estimate of the frequency spectra, it is necessary to apply windowing techniques (Jenkins & Watts, 1968). We first considered the windowing method based on multi-tapers because such methods have been already used with spiking signals (Jarvis & Mitra, 2001; Pesaran *et al.*, 2002). With multi-tapers, the original signals are multiplied with a set of orthogonal tapering functions (e.g. Slepian sequences), the Fourier transforms of the resulting product signals are computed for each taper separately, and finally the spectral components are averaged across all of the tapers (Percival & Walden, 1993). To exemplify this, we first reconstructed the CCH in Fig. 9B (replotted from Fig. 7E) by using three Slepian tapers. The shape of the resulting CCH (Fig. 9C) only resembled the original but much of the structure was changed and the detailed structure was absent even in this non-filtered CCH. Thus, tapers produce spurious components and hence could not be expected to yield reliable results when filtering is applied.

Another, simpler, spectral smoothing method, the Barlett method (Djurić & Kay, 1999), produced better results. Here, the signals are windowed with a rectangular function and hence spurious components are not added to the reconstructed CCH. The reconstructed example CCH is shown in Fig. 9D. Therefore, we attempted to remove slow components from CCHs reconstructed by using Barlett's windowing method and by setting the low-frequency bins (≤ 25 Hz, corresponding to s = 40 ms) of the cross-spectrum to zero. This filtering did not work to a satisfactory degree. For example, when the CCH in Fig. 9B was 'filtered' by this method, the resulting CCH still contained large slow components, and fast components also seemed to be attenuated (Fig. 9E). In contrast, scaled correlation with s = 40 ms removed the slow component efficiently (Fig. 9F).

Scaled correlation and coherence

In the frequency domain, an equivalent of correlation is coherence (Jenkins & Watts, 1968). Coherence measures the periodic correlation between signals, and has been suggested as an alternative to correlation analysis (Pesaran *et al.*, 2002; Sandberg & Hansson, 2006). We investigated how this method relates to scaled correlation, in particular when it is applied to non-periodic signals. As an example, we show a CCH with a narrow center peak but with little or no oscillatory activity (Fig. 9G), suggesting that the signals, although synchronized, do not exhibit prominent periodicity. This is confirmed by the coherence analysis computed by using the multi-tapers (Zeitler *et al.*, 2006; again, by using three Slepian tapers). Spectra were estimated with non-overlapping sliding windows of 512 ms. Coherence showed no dominant frequency (Fig. 9H) and a noisy phase relationship between signals (Fig. 9I). Therefore, there was no peak frequency at which correlation between signals could be measured. In

contrast, SCA, which does not rely on periodicity, in this case showed a clear center peak (Fig. 9J).

We next investigated whether in these cases SCA correlated to the average coherence computed over the whole frequency range between 0 and 500 Hz (the sampling frequency was 1 kHz). To this end, we computed the correlation between the size of the peak of SCA and coherence averaged across all bands for 36 pairs (resulting from nine multi-units), the responses of which to 14 stimulation conditions (504 data points in total) were recorded simultaneously and did not exhibit periodic activity, as assessed from auto-correlation functions (average oscillation score in the 20–40 Hz band = 2.37, see Mureşan *et al.*, 2008). The scale, *s*, was set equal to the trial length. The correlation was positive ($r = 0.59$; see Fig. 9K). As would be expected, this correlation was much smaller with strong periodic activity in neuronal responses. In another data set, recorded from the same multi-units and stimuli but at a later time, neurons exhibited oscillatory responses in the beta-high band (oscillation frequency = ~27 Hz; average oscillation score in the 20–40 Hz band = 6.81). In this case, the correlation between the averaged coherence and SCA peak was low ($r = 0.28$) but the coherence at the oscillation frequency of 27 Hz correlated much better with SCA ($r = 0.57$) (results not shown). Therefore, coherence measures produce results that are similar to those obtained by SCA provided that there is good periodicity structure in the data. However, when periodicity is missing, another type of measure extracted from coherence results may be more optimal to describe the signals. In contrast, SCA does not rely on the assumptions of periodicity and, hence, its results are independent of the presence of this property in the data.

Discussion

In the present study we proposed a measure of neuronal synchrony based on the reduction of the sampled signal variances. The method enables the segregation of correlations at different time scales (e.g. frequency ranges). By basing the analysis on a fast (short) time scale, the contributions of correlations with slower time scales are strongly attenuated in the resulting CCH or auto-correlation histogram. This has the effect that unwanted (slow) rate co-variations are removed from the CCH. SCA can be applied to binary signals, continuous signals, and combinations of continuous and binary signals to remove the contributions of slow components and to extract precise neuronal synchronization between these signals. SCA is more efficient at removing slow components than the SP as it can be applied even when SPs are flat, i.e. when rate co-variation is not stimulus-locked.

The attenuations made by scaled correlation are not always perfect. The effects of the method are limited when the difference between the time scales is small and, in the same time, the signals that need to be attenuated have much larger amplitudes than the signals of interest. Nevertheless, as the simulations show, SCA attenuates slow signals well for a wide range of parameters and, as our experience indicated, it performs very well on neurophysiological data.

An advantage of using Pearson's coefficient of correlation is that its mathematical and statistical properties have been studied in much detail. The procedures for averaging multiple *r*-values have been investigated (Fisher, 1915; Corey *et al.*, 1998; Hunter & Schmidt, 1990; Field, 2001; Raju & Brand, 2003; Millsap, 1988, 1989) and the tables and algorithms for testing the statistical significance of *r* are available. We have also proposed a correction for multiple comparisons that is suitable for application to CCHs based on the requirement that three neighboring peaks are significant (Appendix F).

The calculation of SCA is fast and, to assist its easy implementation, we provide, as a free download, open-source, speed-optimized routines and compiled libraries (see Appendix B for details and links). Most of the demonstrations and tests provided in the present study were based on the extraction of correlations on fast time scales (e.g. < 50 ms), in the range of beta/gamma oscillations. However, the same principles apply for examining signals at any other time scales (e.g. preserving those in the range of theta oscillations), as the time scale of interest can be freely chosen. Also, the application of SCA is not limited to neuronal activity but the same tools can be applied to any type of signals if it is of interest to segregate the contributions of fast and slow components to cross-correlation or auto-correlation functions. Thus, the relevant scale may depend in each analysis on the research question at hand and may thus be decided *a priori* by the researcher. Alternatively, the scale of interest may be chosen after an initial inspection of the data by applying the classical CCHs, or a frequency analysis (e.g. Fourier spectrum). Scaled correlation may prove particularly useful for extracting small time delays in neuronal spiking activity. These delays are typically extracted from CCHs and require the analysis to focus on beta and gamma oscillations (König *et al.*, 1995; Schneider *et al.*, 2006; Nikolić, 2007; Havenith *et al.*, 2011).

Choosing between scaled correlation analysis and spectral methods

One straightforward approach to removing slow components from signals is to simply filter the input signals prior to the computation of the cross-correlation. The limitation of this approach is that it is applicable only to continuous signals because binary series (point processes) cannot be filtered such that another binary signal is obtained. In other words, for binary signals, filtering is not defined because it should be restricted only to the removal or insertion of binary events (spikes). Therefore, filtering cannot be used as a general approach for removing slow components from CCHs. Scaled correlation does not filter signals directly and, hence, SCA can be applied equally efficiently to signals of any nature, i.e. binary or continuous.

To test an alternative to direct filtering on binary data, we relied on the Wiener–Khinchin theorem (Blackman & Tuckey, 1958). We computed the cross-spectrum of the two spike trains, set its low-frequency bins to zero, and then inverse-Fourier transformed it to obtain a ''filtered'' CCH. We showed that, only without filtering, this approach can reconstruct the original CCH fairly accurately and only if certain windowing techniques (e.g. Bartlett) but not others are used. For example, the reconstruction of a CCH from the cross-spectrum estimated with multi-tapers has the problem that the reconstructed CCH also contains components corresponding to the tapers. Thus, the shape of the reconstructed CCH also does not correspond to the original. Consequently, although useful for a number of analyses (Pesaran *et al.*, 2002), multi-taper methods cannot be used to reconstruct a CCH and, hence, also cannot be used to remove slow components from a CCH. Problems also exist with other spectral methods because the spectral representation of binary and/or non-periodic signals is spread across multiple frequency components of the spectrum. Therefore, by filtering signals by setting the low-frequency bins to zero (Jenkins & Watts, 1968), the representation of the slow components is not eliminated completely.

We conclude that spectral filtering methods, although useful for a number of analyses, cannot be used for the accurate removal of slow components from CCHs. Particularly strong limitations are encountered when the correlated signals have representations spread across multiple components of the frequency spectrum, as is the case with trains of action potentials and/or with non-periodic signals. Therefore,

for any analysis based on CCHs, scaled correlation is a more efficient way to selectively extract the correlations between the fast components of the signals.

Our results also showed that coherence could not be a replacement for SCA (and vice versa) as the two measures are correlated only in some cases and to a limited degree. The strength of correlation depends on data properties and on the chosen features of the coherence function. Thus, as described previously by Jenkins & Watts (1968), correlation and coherence are not identical measures but provide complementary information.

In general, the main property that distinguishes the present methods from the spectral techniques is that scaled correlation does not assume a periodic nature of the signals, being equally applicable to the measurement of synchrony with and without oscillations. Consequently, depending on the researcher's preference and the scientific problem at hand, SCA may be chosen over other techniques.

Fisher's z-transformation

One commonly recommended procedure for averaging the values of r is to make a Fisher z-transformation of r-values prior to their averaging, and then transform the average back into an r-value (Fisher, 1915). This procedure is designed to reduce the skew in r-distributions that occurs with larger values of r and to thereby reduce the bias towards underestimating the average of such skewed distributions. However, there are different opinions among researchers about whether or not such a transform is beneficial to the analysis (Silver & Dunlap, 1987; Corey *et al.*, 1998; Hunter & Schmidt, 1990; Field, 2001). Fisher's z-transformation is an option when computing SCA with continuous signals (the equations are provided in Appendix A) but should not be used with spike signals for the following reasons. As we have shown, the distribution of r's has a large number of small (negative) values (see Appendix E). These small values do not suffer from the problem of skewed distributions. More importantly, the distribution is also very likely to contain values with $r = 1.0$, which occur when, within the segment, the two units fire one or more pairs of spikes in perfect coincidence. Fisher's z-transformation cannot be added to the sum with $r = 1.0$ as it produces $+\infty$.

In conclusion, in the measurements of precise neuronal synchrony, the issue of contaminating slow rate co-variation does not need to be a problem. With the application of SCA, slow rate co-variation can be efficiently separated from precise synchrony and the false detection of synchrony can be prevented. For these reasons, SCA may serve as a versatile tool for the investigation of correlations across a variety of neurophysiological signals.

Acknowledgements

The authors would like to thank Julia Biederlack for help with the acquisition of the data used to compute the example analyses, Saskia Nagel and Fabian Fußer for the help in acquiring extensive practical experience with the present methods, and Martha N. Havenith for fruitful discussions. Financial support was provided by a grant from the Deutsche Forschungsgemeinschaft (number NI 708/2-1), Hertie Foundation, Alexander von Humboldt Stiftung, LOEWE Neuronale Koordination Forschungsschwerpunkt Frankfurt (NeFF), Max-Planck Geselschaft, Frankfurt Institute for Advanced Studies, a grant from the Romanian Government (Human Resources Program, project number PNII-RU TE-11/2010, contract no. 23/28.07.2010 financed by CNCS/UEFISCDI), and a grant for the ''Max Planck – Coneural Partner Group''.

Abbreviations

CCH, cross-correlation histogram; FFT, fast Fourier transform; LFP, local-field potential; SCA, scaled correlation analysis; SP, shift predictor.

References

Aguinis, H. & Whitehead, R. (1997) Sampling variance in the correlation coefficient under indirect range restriction: implication for validity generalization. *J. Appl. Psychol.*, **82**, 528–538.

Alexander, R.A., Carson, K.P., Alliger, G.M. & Barrett, G.V. (1984) Correction for restriction of range when both X and Y are truncated. *App. Psychol. Meas.*, **8**, 231–241.

Averbeck, B.B., Latham, P.E. & Pouget, A. (2006) Neural correlations, population coding and computation. *Nat. Rev. Neurosci.*, **7**, 358–366.

Beggs, J.M. & Plenz, D. (2003) Neuronal avalanches in neocortical circuits. *J. Neurosci.*, **23**, 11167–11177.

Biederlack, J.M., Castelo-Branco, M., Neuenschwander, S., Wheeler, D., Singer, W. & Nikolić, D. (2006) Brightness induction: rate enhancement and neuronal synchronization as complementary codes. *Neuron*, **52**, 1073–1083.

Blackman, R.B. & Tukey, J.W. (1958) *The Measurement of Power Spectra From the Point of View of Communication Engineering*. Dover, New York, NY.

Brody, C.D. (1999) Correlations without synchrony. *Neural Comput.*, **11**, 1537–1551.

Corey, D.M., Dunlap, W.P. & Burke, M.J. (1998) Averaging correlations: expected values and bias in combined Pearson rs and Fisher's z transformations. *J. Gen. Psychol.*, **125**, 245–261.

Djurić, P.M. & Kay, S.M. (1999) Spectrum estimation and modeling. In Madisetti, V.K. & Williams, D.B. (Eds), *Digital Signal Processing Handbook*. CRC Press, Boca Raton, FL.

Field, A.P. (2001) Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed- and random-effects methods. *Psychol. Methods*, **6**, 161–180.

Fisher, R.A. (1915) Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika*, **10**, 507–521.

Fries, P., Roelfsema, P.R., Engel, A.K., König, P. & Singer, W. (1997) Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proc. Natl. Acad. Sci. USA*, **94**, 12699–12704.

Gerstein, G.L. & Perkel, D.H. (1972) Mutual temporal relationships among neuronal spike trains. Statistical techniques for display and analysis. *Biophys. J.*, **12**, 453–473.

Gray, C.M. & Singer, W. (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl. Acad. Sci. USA*, **86**, 1698–1702.

Gray, C.M., König, P., Engel, A.K. & Singer, W. (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, **338**, 334–337.

Grün, S., Diesmann, M. & Aertsen, A. (2002) Unitary events in multiple single neuron activity. I. Detection and significance. *Neural Comput.*, **14**, 43–80.

Hahn, G., Petermann, T., Havenith, M. N., Yu, S., Singer, W., Plenz, D. & Nikolić, D. (2010) Neuronal avalanches in spontaneous activity in vivo. *J. Neurophysiol.*, **104**, 3312–3322.

Havenith, M.N., Yu, S., Biederlack, J., Chen, N.-H., Singer, W. & Nikolić, D. (2011) Synchrony makes neurons fire in sequence – and stimulus properties determine who is ahead. *J. Neurosci.*, **31**, 8570–8584.

Hedges, L.V. & Olkin, I. (1985) *Statistical Methods for Meta-Analysis*. Academic Press, Orlando, FL.

Held, J.D. & Foley, P.P. (1994) Explanations for accuracy of the general multivariate formulas in correcting for range restriction. *App. Psychol. Meas.*, **18**, 355–367.

Hunter, J.E. & Schmidt, F.L. (1990) *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Sage, Newbury Park, CA.

Ikegaya, Y., Aaron, G., Cossart, R., Aronov, D., Lampl, I., Ferster, D. & Yuste, R. (2004) Synfire chains and cortical songs: temporal modules of cortical activity. *Science*, **204**, 559–564.

Jarvis, M.R. & Mitra, P.P. (2001) Sampling properties of the spectrum and coherency of sequences of action potentials. *Neural Comput.*, **13**, 717–749.

Jenkins, G.M. & Watts, D.G. (1968) *Spectral Analysis and Its Applications*. Holden-Day, San Francisco, CA.

Kim, S.H. (1992) *Statistics and Decisions: An Introduction to Foundations*. CRC Press, Boca Raton, FL.

König, P. (1994) A method for the quantification of synchrony and oscillatory properties of neuronal activity. *J. Neurosci. Methods*, **54**, 31–37.
König, P., Engel, A.K., Roelfsema, P.R. & Singer, W. (1995) How precise is neuronal synchronization? *Neural Comput.*, **7**, 469–485.
Kuo, Z.-Y. (1930) The genesis of the cat's response to the rat. *J. Comp. Psychol.*, **11**, 1–30.
Lamme, V.A & Spekreijse, H. (1998) Neuronal synchrony does not represent texture segregation. *Nature*, **396**, 362–366.
Macke, J.H., Berens, P., Ecker, A.S., Tolias, A.S. & Bethge, M. (2009) Generating spike trains with specified correlation coefficients. *Neural Comput.*, **21**, 397–423.
Millsap, R.E. (1988) Sampling variance in attenuated correlation coefficients: a Monte Carlo Study. *J. Appl. Psychol.*, **73**, 316–319.
Millsap, R.E. (1989) Sampling variance in the correlation coefficient under range restriction: a Monte Carlo Study. *J. Appl. Psychol.*, **74**, 456–461.
Moca, V.V., Nikolić, D. & Mureşan, R.C. (2008) Real and modeled spike trains: where do they meet? In Vera Kurková, Roman Neruda & Jan Koutník. (eds.), *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, pp. 488–497.
Munk, M.H.J., Nowak, L.G., Nelson, J.I. & Bullier, J. (1995) Structural basis of cortical synchronization II. Effects of cortical lesions. *J. Neurophysiol.*, **74**, 2401–2414.
Mureşan, R.C., Jurjut, O.F., Moca, V.V., Singer, W. & Nikolić, D. (2008) The oscillation score: an efficient method for estimating oscillation strength in neuronal activity. *J. Neurophysiol.*, **99**, 1333–1353.
Nikolić, D. (2007) Non-parametric detection of temporal order across pairwise measurements of time delays. *J. Comput. Neurosci.*, **22**, 5–19.
Nikolić, D., Häusler, S., Singer, W. & Maass, W. (2009) Distributed fading memory for stimulus properties in the primary visual cortex. *PLoS Biol.*, **7**, e1000260.
Nowak, L.G., Munk, M.H.J., James, A.C., Girard, P. & Bullier, J. (1999) Cross-correlation study of the temporal interactions between areas V1 and V2 of the macaque monkey. *J. Neurophysiol.*, **81**, 1057–1074.
Percival, D.B. & Walden, A.T. (1993) *Spectral Analysis for Physical Applications*. Cambridge University Press, Cambridge.
Pesaran, B., Pezaris, J.S., Sahani, M., Mitra, P.P. & Andersen, R.A. (2002) Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.*, **5**, 805–811.
Pipa, G., Wheeler, D.W., Singer, W. & Nikolić, D. (2008) NeuroXidence: reliable and efficient analysis of an excess or deficiency of joint-spike events. *J. Comput. Neurosci.*, **25**, 64–88.
Raju, N.S. & Brand, P.A. (2003) Determining the significance of correlations corrected for unreliability and range restriction. *App. Psychol. Meas.*, **27**, 52–71.
Roelfsema, P.R., Engel, A.K., König, P. & Singer, W. (1997) Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature*, **385**, 157–161.
Runyon, R.P. & Huber, A. (1980) *Fundamentals of Behavioral Statistics*. Addison-Wesley Publishing Company, Reading, MA.
Sandberg, J. & Hansson, M. (2006) Coherence estimation between EEG signals using multiple Window Time-Frequency Analysis compared to Gaussian Kernels. *Proceedings of European Signal Processing Conference*, September 4–8. Florence, Italy.
Schneider, G. & Nikolić, D. (2006) Detection and assessment of near-zero delays in neuronal spiking activity. *J. Neurosci. Methods*, **152**, 97–106.
Schneider, G. & Nikolić, D. (2008) A stochastic framework for the quantification of synchronous oscillation in neuronal networks. *Proceedings of the Fifth International Workshop on Computational Systems Biology*. WCSB, Leipzig, Germany, pp. 169–172.
Schneider, G., Havenith, M.N. & Nikolić, D. (2006) Spatio-temporal structure in large neuronal networks detected from cross-correlation. *Neural Comput.*, **18**, 2387–2413.
Silver, N.C. & Dunlap, W.P. (1987) Averaging correlation coefficients: should Fisher's z transformation be used? *J. Appl. Psychol.*, **72**, 146–148.
Staude, B., Rotter, S. & Grün, S. (2008) Can spike coordination by differentiated from rate covariation? *Neural Comput.*, **20**, 1973–1999.
Toothaker, L.E. (1993) *Multiple Comparison Procedures*. Sage University Inc., Newbury Park, CA.
Zeitler, M., Fries, P. & Gielen, S. (2006) Assessing neuronal coherence with single-unit, multi-unit, and local field potentials. *Neural Comput.*, **18**, 2256–2281.

Appendix A: Fisher's *z*-transformation of *r*-values

To average the values of r by using Fisher's z-transformation, each r is first converted into a Fisher's z, denoted as z' (Fisher, 1915), by computing the inverse hyperbolic tangent

$$z' = \operatorname{arctanh}(r) = 1/2[\ln(1+r) - \ln(1-r)] \qquad (A1)$$

where ln is a logarithm with natural base e. Note that if $r = 1$ or $r = -1$, z' cannot be computed. After the transform, the values of z' can be averaged and the result, $\bar{z}'$, can be converted back to r by computing the hyperbolic tangent

$$r_z = \tanh(\bar{z}') = \left(e^{2\bar{z}'} - 1\right) / \left(e^{2\bar{z}'} + 1\right) \qquad (A2)$$

Note that Fisher's z-transformation is not the same as the standard z-score although both traditionally use the same symbol (z) to indicate the respective values.

Appendix B: calculating Pearson's coefficients of correlation for different types of signals

Pearson's r between variables x and y is defined theoretically as the product moment correlation, which for a sample is computed as follows

$$r = \frac{\Sigma z_x z_y}{N-1} \qquad (B1)$$

where N indicates the number of measurements and z is the standard score (z-score) for each measurement (not a Fisher's transform). Thus, Pearson's r is a measure of correlation normalized to a unit variance, and its square (r^2) can be interpreted as the proportion of variance explained in one variable by knowing the value in another variable. Other formulas are used for the computationally efficient calculation of r.

Continuous signals

For pairs of continuous signals, such as LFP or electroencephalogram, the following equation can be used to compute r quickly

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{N(\sum x^2 - (\sum x)^2)N(\sum y^2 - (\sum y)^2)}} \qquad (B2)$$

where x and y are the raw measurements of electric potentials for the two variables and N is the number of measurements used to compute r. For example, if the scale $s = 50$ ms and the sampling rate of the continuous signal is 1 kHz (i.e. 1 measurement/ms), there will be a total of $N = 50$ measurements for each r. Thus, prior to applying the equation, one needs to make one run through the data set (e.g. one 'for' loop) and compute five sums: Σx, Σy, Σx^2, Σy^2 and Σxy. If the measurements are given as integers, the calculation will be faster if: (i) these sums are also computed with integer arithmetic (type 'long' should be used) and (ii) floating point operations are then used only for the final calculation of Eqn B2. In MATLAB, function 'corr(X)' can be used to compute Pearson's coefficient of correlation.

Binary signals

For binary signals, such as the time stamps of action potentials (spikes), the version of Pearson's r known as the phi (φ) coefficient of

correlation (Kuo, 1930) should be used. The phi coefficient of correlation is mathematically equivalent to Pearson's r but the calculation is optimized for the use of dichotomous and nominal variables. The calculation assumes that data are organized into time bins, each bin having a short duration (e.g. 1 ms) and containing either a 0 or 1, indicating the absence or presence of an event, respectively (e.g. an action potential). For larger time bins (e.g. 5 ms), multiple action potential events can occur within a single bin, in which case only one of them can be counted (i.e. the values cannot exceed 1). The calculation of φ begins with a count of coincident spikes, b, coincident lack of spikes, c, bins with spikes only in one variable, a, and only in the other variable, d, which can be shown as a contingency table (Table B1).

TABLE B1. Contingency table for the computation of the φ coefficient of correlation between two binary variables X and Y

Y/X	0	1	Totals
1	a	b	$a + b$
0	c	d	$c + d$
Totals	$a + c$	$b + d$	N

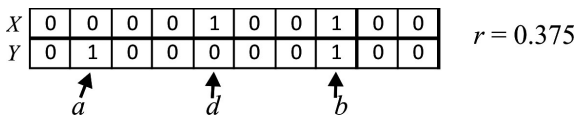
A value of 1 indicates the presence of at least one event (e.g. action potential) in a given bin (e.g. 1 ms size) and a 0 indicates the absence of the event. The count of coincident events is given by value b, whereas c represents the number of time bins in which no events occurred in either of the two signals, and a and d give the counts of events that did not have a coincident counterpart in the other respective signal. The coefficient φ is computed by entering the values $a - d$ into Eqn B3.

After the frequencies $a - d$ are determined, φ is computed by the following formula

$$\phi = \frac{(bc - ad)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}. \qquad (B3)$$

Here we provide an example of the calculation of the φ coefficient of correlation for a short segment of data with ten bins and two spikes in each signal. In this example, only one coincident event occurred (shown in Fig. B1).

Fig. B1. Two example spike trains with two action potentials each. r



indicates Pearson's coefficient of correlation computed with Eqn B2. The bins counted for values a, b and d are indicated by arrows. All other bins count for the value of c.

The counts in the contingency Table B1 are as follows

Y/X	0	1	Totals
1	1	1	2
0	7	1	8
Totals	8	2	10

By using Eqn B3, this leads to the following calculation of the phi coefficient of correlation

$$\phi = \frac{7 - 1}{\sqrt{2 \cdot 8 \cdot 2 \cdot 8}} = \frac{6}{\sqrt{256}} = 0.375.$$

Combination of continuous and binary signals

Finally, for correlation between one continuous variable, x, and one binary variable, a form of Pearson's r known as the point-biserial coefficient of correlation should be used. This coefficient is computed most efficiently by the following equation

$$r_{PB} = \frac{(x_1 - x_0)\sqrt{P(1-P)}}{S_x}, \qquad (B4)$$

where x_1 and x_0 indicate the averages for the continuous variable in cases where the binary variable had the values 1 and 0, respectively, P is the probability that a value of 1 will be observed in the binary variable, and S_x represents the population standard deviation for the continuous variable. Note that

$$P(1 - P) = (n_1 n_0)/n^2, \qquad (B5)$$

where n_1 and n_0 indicate sample sizes (counts) for x_1 and x_0, respectively, such that $n = n_1 + n_0$. This equation can be used to correlate continuous signals such as LFP or electroencephalogram to neuronal spiking activity. Note also that, when computed for all possible time offsets between target and reference, x_1 can be interpreted as the spike-triggered average of LFP (or electroencephalogram) and r_{PB} as the normalized spike-triggered average.

Spearman's rho

In cases in which the applicability of Pearson's r, which is a parametric method, is doubted, one possibility is to compute the non-parametric Spearman's rho (ρ) coefficient of correlation. This measure of correlation is a Pearson's r computed on ranked data instead of the original data. Thus, if data are first ranked with values $1 \ldots N$, where N is the number of data points, and if for each pair a difference in ranks, D, is computed, then Spearman's rho is given by

$$\rho = 1 - \frac{6 \sum_N D^2}{N^3 - N}. \qquad (B6)$$

The routines for the computation of SCA are freely available as a source code in C++, a dynamic-link library for Windows or a dynamic-link library specialized only for the Matlab environment (but also only for the Windows operating system). The package is downloadable from http://www.raulmuresan.ro/sources/corrlib/ and also contains the routines for the calculation of classical CCHs together with instructions on how to use the routines.

Appendix C: variance across the entire data set

Claim

If the segments are equal in length, the variance of the entire data set is larger than or equal to the arithmetic mean of the variances of the segments: $\mathrm{Var}_G \geq \mathrm{Var}_S$.

Proof

For any random variable x, the mean is the value around which the second moment of x is minimal (Kim, 1992; p. 224)

$$\frac{1}{N}\sum_i (x_i-\mu)^2 = \min . \quad (C1)$$

For K segments, each with L samples (bins), the variance of the entire data set is

$$\mathrm{Var}_G = \frac{1}{KL}\sum_{k=1}^{K}\sum_{l=1}^{L}(x_{kl}-\hat{\mu})^2, \quad (C2)$$

and that of a segment k is

$$\mathrm{Var}_S = \frac{1}{L}\sum_{l=1}^{L}(x_{kl}-\mu_k)^2, \quad (C3)$$

where $\hat{\mu}$ indicates the arithmetic mean of the entire signal and μ_k indicates the mean of the k'th segment. $\hat{\mu}$ will, by definition, be different to the means of the individual segments, k, i.e. $\hat{\mu} \neq \mu_k$ Therefore, it follows from Eqn C1 that the summands in $\sum_{k=1}^{L}$ in Eqn C2 will always be larger when computed with the mean of the entire signal, i.e.

$$\sum_{l=1}^{L}(x_{kl}-\hat{\mu})^2 > \sum_{l=1}^{L}(x_{kl}-\mu)^2,$$

leading to $\mathrm{Var}_G > \mathrm{Var}_S$.

Appendix D: scaled correlation and filtering of correlations

Consider two time series X and Y, defined as follows

$$X = X_s + X_f + X_n \quad (D1)$$

and

$$Y = Y_s + Y_f + Y_n \quad (D2)$$

where X_s represents a time series that is slowly oscillating in amplitude (slow component), X_f is a time series that oscillates much faster than X_s (fast component), X_n is the noise, X_s, X_f and X_n are assumed to be mutually independent and Y_t is similarly defined as the summation of the three independent components Y_s, Y_f and Y_n. Also, in line with the conventions for computing cross-correlograms and given that we are interested primarily in time-averaged results, all of these time series are assumed to be homogeneous.

The cross-correlation between X and Y is by definition

$$\begin{aligned} E(XY) =& E((X_s+X_f+X_n)(Y_s+Y_f+Y_n)) \\ =& E(X_sY_s+X_sY_f+X_sY_n+X_fY_s+X_fY_n+X_nY_s+X_nY_f \\ &+X_nY_n+X_fY_f), \end{aligned} \quad (D3)$$

where E denotes the expected value.

The goal is to isolate the cross-correlation of the fast component, i.e. $E(X_fX_f)$.

Using Eqn D3, it can be written

$$\begin{aligned} E(X_fY_f) = E(XY) - E(&X_sY_s + X_sY_f + X_sY_n + X_fY_s \\ &+ X_fY_n + X_nY_s + X_nY_f + X_nY_n). \end{aligned} \quad (D4)$$

Assuming that the noisy processes X_n and Y_n have a mean of zero, Eqn D4 can be simplified into

$$\begin{aligned} E(X_fY_f) =& E(XY) - E(X_sY_s + X_sY_f + X_fY_s) \\ =& E(XY) - (E(X_sY_s) + E(X_s)E(Y_f) + E(X_f)E(Y_s)). \end{aligned} \quad (D5)$$

For the cases in which $E(X_s)$ and $E(Y_s)$ are of a moderate magnitude and $E(X_f) << 1$ and $E(Y_f) << 1$, the magnitude of the negative term in Eqn D5 is determined mainly by $E(X_fX_f)$, which is the cross-correlation between the slow processes X_s and Y_s.

It follows that the center bin of the classical cross-correlogram of the fast component is estimated as

$$E(X_fY_f) = E(XY) - X(X_sY_s) - \epsilon, \quad (D6)$$

where ε is zero if X_s is independent of Y_f, X_f is independent of Y_s, and the means of X_f and Y_f are 0.

Implicitly, scaled correlation estimates and subtracts a normalized form of the term $E(X_fY_f)$ from Eqn D6, which can be shown in the following way.

We first show that scaled correlation can be expressed in terms of a subtraction of the expected product between the means of the segments, $E(\mu_f\mu_f)$, from the classical cross-correlation.

Equation 8 from the main text can be linked to the average of correlation coefficients if the following is considered

$$\begin{aligned} \overline{r_s} =& \frac{1}{K}\sum r_k = \frac{1}{K}\sum_k \frac{E[(X_k-E(X_k))(Y_k-E(Y_k))]}{\sigma_{xk}\sigma_{yk}} \\ =& \frac{1}{K}\sum_k \frac{E(X_kY_k)-E(X_k)E(Y_k)}{\sigma_{xk}\sigma_{yk}}, \end{aligned}$$

where X_k represents the time series of the k-th segment (ranging between 1 and K).

Next, assuming that X and Y have the same variance across all segments, it follows that

$$\begin{aligned} \overline{r_s} =& \frac{1}{\sigma_x\sigma_y}\frac{1}{K}\sum_k [E(X_kY_k)-E(X_k)E(Y_k)] \\ \approx& \frac{1}{\sigma_x\sigma_y}(E(XY)-E(\mu_x\mu_y)), \end{aligned} \quad (D7)$$

where μ_X is the mean of segments of X.

Equation D7 states that, by calculating correlations on short segments for time series X and Y, scaled correlation in effect also computes the expected values for the product between μ_X and μ_Y, i.e. $E(\mu_x\mu_y)$. The averages μ_X and μ_Y are computed for each consecutive segment, and have $n = N/K$ sampling points (a positive integer; $s = n$*bin size), N being the total length of X and Y in sampling points ($t = N$*bin size), and K being the number of segments.

We next show that the subtraction of $E(\mu_x\mu_y)$ is an approximation of the $E(X_sY_s)$, and a part of this approximation is a procedure that acts as a moving average filter with low-pass properties.

Low-pass filtering is defined as a filter that passes low-frequency signals and attenuates signals with high frequencies, which can be achieved by calculating a moving average. For each data point, the moving average is computed by averaging the values of several neighboring data points, which is called the window of moving average, the size of which is given by the total number of averaged data points, d. Typically, the moving average is computed by sliding the window by a step of one data point, resulting in overlaps between subsequent window positions. In scaled correlation, due to the segmentation, the overlap is not necessary. It is sufficient to compute one average for each segment such that the length of the moving window, d, corresponds to the segment size, s.

To be precise, let us define X_c and Y_c as low-pass-filtered versions of X and Y by applying moving averages. Thus, μ_X and μ_Y are a sub-set of X_c and Y_c sampled at a constant interval, d.

It follows that the expectation of the product of the low-pass-filtered X and Y can be approximated, i.e. $E(X_cY_c)$ by the $E(\mu_x\,\mu_y)$.

The approximation error $|E(\mu_x\mu_Y) - E(X_cY_c)|$ is reduced 1when the segmentation length s reduces. Keep in mind that the length of this window also determines the properties of the low-pass filtering. Thus, the approximation error cannot be reduced without changing in the same time the cut-off frequency of the signal components that will be removed from the scaled correlation.

Finally, when s is equal to the total length of the signal, the term $E(\mu_x\,\mu_y)$ in Eqn D7 corresponds to the magnitude of the baseline of a normalized classical CCH, i.e. the product of the means of the X and Y. Therefore, a scaled cross-correlogram with s = data length will have a shape that is identical to that of the classical CCH, when the classical one is centered around zero after being corrected for the baseline.

Appendix E: conditions under which the average of segmented correlations equals non-segmented correlation

Claim

If the means and variances of signals are constant across all segments, the average scaled correlation obtained with small s is identical to the value of the single correlation coefficient computed across the entire signal (i.e. overall correlation determined with the maximum segment size, s).

Proof

Consider r computed across the entire pairs of signals x and y as the product moment correlation for population

$$r = \frac{\Sigma z_x z_y}{N}. \qquad (E1)$$

Note the division by N, rather than $N - 1$, which indicates the correlation estimate for the population. The standard score, z, is defined as

$$z = \frac{x - \mu}{\sigma},$$

where μ and σ indicate the mean and SD, respectively. If the signals are segmented and the mean and variance (i.e. σ^2) stay constant for each variable across all segments, correlation $\bar{r}$ averaged from r_1 to r_K can be written as

$$\bar{r} = \frac{\sum_{l=1}^{L} Z_{x_{1l}} Z_{y_{1l}} + \sum_{l=1}^{L} Z_{x_{2l}} Z_{y_{2l}} + \ldots + \sum_{l=1}^{L} Z_{x_{Kl}} Z_{y_{Kl}}}{KL} = \frac{\sum_{k=1}^{K}\sum_{l=1}^{L} Z_{x_{kl}} Z_{y_{kl}}}{KL}, \qquad (E2)$$

where indices $1\ldots K$ represent different segments and L is the number of measurements (bins) within a segment.

Following from the definition above, the value of the z-score does not depend on the actual number of segments, when μ and σ are constant across segments. For the correlations computed across the entire signal (Eqn E1), $N = KL$. Also, due to the consecutive segmentation of the signals, the nested sums over L and K can be expressed as a single sum over N. It follows that

$$\bar{r} = \frac{\sum_{n=1}^{N} Z_{xn} Z_{yn}}{N} \equiv r. \qquad (E3)$$

This concludes the proof.

To illustrate the applicability of this proof on binary signals, we give an example of two spike trains having in total $N = 21$ bins for which the correlation $\phi = -0.028$ (computed with Eqn B3). Each of the three segments has seven bins and the spike counts are the same across all segments, i.e. three and four spikes in all segments for the variables X and Y, respectively (Fig. E1). Equal spike counts make the means and variances equal across segments.

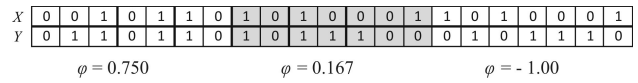


Fig. E1. Two hypothetical spike trains with different firing rates but with the same number of action potentials within each of the three segments. Segments are separated by the gray/white pattern. φ-values indicate correlation coefficients computed within each segment.

Correlations computed for each segment are different as they depend on the number of coincident events (3, 2 and 0 in the present example). Nevertheless, the average of these correlations, $\bar{\phi} = (0.75 + 0.167 - 1.0)/3 = -0.028$, is identical to that computed for the entire stretch of data.

Implications for Fisher's *z*-transformation

When Fisher's z-values are computed for the summands, Eqn E2 no longer applies and, thus, the identity Eqn E3 no longer holds. Foremost, Fisher's z-transformations cannot even be computed for the above example due to $\varphi = -1.0$ for which the transformation cannot be made. When the calculations were made for only the first two segments in which the correlations were suitable for Fisher's z-transformation, the 'corrected' average was $\bar{\phi} = 0.516$. This value was different from the single correlation coefficient computed across the two segments, $\varphi = 0.458$. Therefore, Fisher's z-transformation cannot be used with binary signals.

Appendix F: estimating the significance of *r*-values

Sometimes it is necessary to estimate whether a correlation coefficient obtained from SCA is significantly different from a zero correlation. The significance of a phi coefficient of correlation, which is based solely on binary signals, is usually determined by computing a chi-square test (e.g. Runyon & Huber, 1980). The significance of a coefficient of correlation that includes at least one continuous variable (hence including the point-biserial coefficient of correlation) is based on the Student's t-distribution. The t-value can be computed from r as follows

$$t = \frac{r}{\sqrt{\frac{1-r^2}{N-2}}}, \qquad (F1)$$

and is distributed with $N - 2$ degrees of freedom. N indicates the number of samples, which, in the case of spiking signals, corresponds to the number of bins. For example, with a 1 kHz sampling rate of the signal (1 ms bin), $r = 0.50$ is significant with $N = 12$ and 22 at $\alpha = 0.05$ and 0.01, respectively (t-values = 1.83 and 2.58, respectively; one-tailed test). With a 1 ms bin size, this amounts to a duration of segments of only 12 or 22 ms. Equation F1 should not be used for $N < 6$ as the estimates become inaccurate. High accuracy may be achieved with $N \geq 60$ (Millsap, 1988, 1989). For binary signals, the correlations are typically much smaller and will require longer stretches of signal to reach significance.

Significance of average correlations

Due to the nature of scaled correlation, it is necessary to compute the significance of an average of correlation coefficients rather than that of a single coefficient. For testing the significance of $\bar{r}$, we propose the method of fixed effects of Hedges & Olkin (1985) developed for testing the significance of average correlation coefficients in meta-studies. According to Hedges & Olkin (1985) (see also Field, 2001), the SE of the average correlation coefficient can be estimated from the number of estimates of r used to compute the average, which corresponds to the number of segments, K, and from the sample size for each r, which corresponds to the number of samples (bins) within each segment, L. Hence, the SE is given by

$$\mathrm{SE}(\bar{r}) = \sqrt{\frac{1}{K(-3L)}}. \qquad (F2)$$

The likelihood of obtaining the given $\bar{r}$ by chance is then obtained simply from the following z-score

$$z_{\bar{r}} = \frac{\bar{r}}{\mathrm{SE}(\bar{r})}. \qquad (F3)$$

For example, for $K = 400$ segments and for 25 samples in each segment (e.g. $s = 25$ ms with a 1 kHz sampling rate; a total of 10 s of signal), the SE is

$$\mathrm{SE}(\bar{r}) = \sqrt{\frac{1}{1000(22)}} = 0.01066,$$

which, even for a very small $\bar{r} = 0.05$ leads to a z-score = 4.69 and a highly significant value of $P = 1.36 \times 10^{-6}$.

However, with spiking signals these significances will be smaller because usually only a portion of segments contain spikes for both neurons and, thus, only a portion of segments will be taken into analysis. For example, with a firing rate of 10 spikes/s and with the assumption of independence between the neurons, it is expected that only 15 segments out of 40 (37.5%) will be valid. Equations F2 and F3 lead to $\mathrm{SE}(\bar{r}) = 0.0714$ and, for the same $\bar{r} = 0.05$, to $P = 0.002$ (z-score = 2.87).

Controlling for multiple comparisons

The next step is to control the type I error in statistical inference, which is the problem of false-positive detection of significant correlations. When computing CCHs, this can lead to a considerable problem due to the large number of estimates made across different offsets in each CCH. For example, if two signals are not correlated, and the significance is tested at $\alpha = 0.01$ for $m = 161$ correlations (±80 ms shift with 1 ms binning plus the center bin), it is expected that in $p(m) = 1 - (1 - \alpha)^m = 0.80$ or in 80% of cases at least one bin will nevertheless be significant. Thus, without a correction, a researcher is in danger of concluding incorrectly that, at some time delay, the two variables are correlated.

A number of correction methods have been proposed for similar problems with multiple comparisons in which one can, for example, reduce the critical α value (see e.g. Toothaker, 1993). We propose here a correction method that is suitable for the large number of estimates made typically in a CCH and for the studies in which correlation peaks of interest are wider than the binning resolution of the CCH, as is usually the case. This correction procedure requires that a peak is considered significant only if at least three neighboring peaks are found to be significant by a standard, not corrected, test.

This requirement that three neighboring, i.e. consecutive, bins are significant controls well the probability of making a type I error in statistical inference. For our example with nominal $\alpha = 0.01$ for $m = 161$ correlation bins, the probability that any combination of three bins will be significant by chance already reduces from 80% to $p(m)*p(m-1)*p(m-2) = 0.51$ or 51%. When it is required that the three bins are neighbors, the conditional probability of corrected alpha decreases to $\alpha_{\mathrm{CORR}} = p(m)*\alpha^2 = 0.00008$. With the nominal $\alpha = 0.05$ this requirement leads to $\alpha_{\mathrm{CORR}} = 0.0025$ and with the nominal α-value relaxed further to 0.10, $\alpha_{\mathrm{CORR}} = 0.01$. Therefore, if three neighboring bins are significant (and the correlations point to the same direction), we can accept the significance of the peak confidently because the likelihood of making a type I error is small.

Finally, it is important to note that it will not always be necessary to compute the significance for individual CCHs. In many studies, the significance tests are made by pooling a number of CCHs obtained from different pairs of units and/or different subjects. Comparisons are then made across, for example, different experimental conditions (e.g. Biederlack *et al.*, 2006). Here, the sample size is defined by the number of pairs of units investigated, and the properties of such correlation estimates (e.g. their distributions) usually allow for the application of standard parametric statistics such as ANOVA. These latter types of tests are also often more desirable as they allow the researcher to generalize to the entire population rather than being limited to a conclusion about a single pair of units, as is necessarily a limitation of a significance test of an individual CCH.