

## Unitary Events in Multiple Single-Neuron Spiking Activity: II. Nonstationary Data

**Sonja Grün**

*gruen@mpih-frankfurt.mpg.de*

*Department of Neurophysiology, Max-Planck Institute for Brain Research, D-60528 Frankfurt/Main, Germany*

**Markus Diesmann**

*diesmann@chaos.gwdg.de*

*Department of Nonlinear Dynamics, Max-Planck Institut für Strömungsforschung, D-37073 Göttingen, Germany*

**Ad Aertsen**

*aertsen@biologie.uni-freiburg.de*

*Department of Neurobiology and Biophysics, Institute of Biology III, Albert-Ludwigs-University, D-79104 Freiburg, Germany*

In order to detect members of a functional group (cell assembly) in simultaneously recorded neuronal spiking activity, we adopted the widely used operational definition that membership in a common assembly is expressed in near-simultaneous spike activity. Unitary event analysis, a statistical method to detect the significant occurrence of coincident spiking activity in stationary data, was recently developed (see the companion article in this issue). The technique for the detection of unitary events is based on the assumption that the underlying processes are stationary in time. This requirement, however, is usually not fulfilled in neuronal data. Here we describe a method that properly normalizes for changes of rate: the unitary events by moving window analysis (UEMWA). Analysis for unitary events is performed separately in overlapping time segments by sliding a window of constant width along the data. In each window, stationarity is assumed. Performance and sensitivity are demonstrated by use of simulated spike trains of independently firing neurons, into which coincident events are inserted. If cortical neurons organize dynamically into functional groups, the occurrence of near-simultaneous spike activity should be time varying and related to behavior and stimuli. UEMWA also accounts for these potentially interesting nonstationarities and allows locating them in time. The potential of the new method is illustrated by results from multiple single-unit recordings from frontal and motor cortical areas in awake, behaving monkey.

## 1 Introduction

---

In the companion article in this issue, unitary event analysis was introduced to detect a certain type of statistical dependency in the spiking activities of simultaneously recorded neurons: near-coincident spike constellations that occur more often than expected on the basis of independent firing rates. In the literature, such events are discussed as signatures of coherent cell assemblies, considered to be the building blocks of cortical processing (see the companion article for references). Unitary event analysis as described in the companion article was based on the assumption that the underlying processes are stationary. Typically, however, experimental data show modulations in firing rates. In fact, often stimuli are manipulated to enhance these “responses.” In this article, we describe an extension of the stationary method, unitary events by moving window analysis (UEMWA), specifically designed to enable an application to nonstationary data.

After introducing and describing the method in detail (section 2), we illustrate its performance and discuss its sensitivity using simulated spike sequences under different scenarios of nonstationarities in firing rate and nonstationarities in coincidence rate, on the same and on different timescales (section 3). Two experimental data sets from frontal and motor cortical recordings are used to illustrate the occurrence of unitary events in neuronal data and their relation to behavioral context (section 4). In section 5, we concentrate on practical aspects of the application of our new method. An assessment of the problems of false positives and false negatives is followed by guidelines for proper choice of analysis parameters.

## 2 Detecting Unitary Events by Moving Windows Analysis

---

The task is to develop a method that allows the detection of unitary events in nonstationary spike data. The basic idea is to segment the data into sections over which stationarity can be assumed and analyze the data in these sections separately by the method developed in the companion article for the stationary situation. Here, we construct a procedure in which a time window of width  $T_w$  is slid along the data, and unitary event analysis is performed separately at each window position, defining slightly different rate environments (an alternative approach is described in appendix D and discussed in section 5). The moving window has to be narrow enough such that the firing rates can be assumed to be stationary and at the same time long enough to obtain sufficient statistics. It turns out that a third criterion, the time-dependent rate of the spike coincidences to be detected, also influences the optimal choice of the analysis window size. In this section, we describe the procedure sketched above in detail and work out its statistical interpretation, using the results of the companion article.

Peristimulus time histograms (PSTHs) of spike data from motor and frontal cortical areas show that firing rates are usually sufficiently station-

ary (cf. section 4, Figures 6 and 7) over time windows on the order of 50 to 100 ms. However, considering a rate level of 1 to 50 s<sup>-1</sup>, we expect only 0.1 to 5.0 spikes in a single window. Clearly, these numbers are too low for a statistical comparison of the numbers of expected and observed coincidences. An important assumption of the PSTH is that firing rates are stationary across trials, and, thus, the average over trials allows computing a reliable estimate of the firing rate at any point in time. Using this assumption, the set of trials performed for a particular experimental condition can be combined to overcome the problem of the low numbers of counts stated above. Figure 1 illustrates how the data of all available trials are used to construct a new process. For a time window centered at  $t_i$ , the data from the  $M$  trials are concatenated to form a new set of parallel spike trains of length  $M \cdot T_w$ . Let  $\mathbf{v}_j(t)$  be the parallel (0, 1)-process (see the companion article) describing the neuronal spike data of trial  $j$ . In this article, all variables representing time are in units of the temporal resolution  $h$  of  $\mathbf{v}_j$ . In these units, the window width is an odd integer,

$$T_w = (2n + 1), \quad n \in \{0, 1, 2, \dots\}. \quad (2.1)$$

The new parallel process  $\mathbf{v}$  on the new time axes  $t'$  is given by

$$\mathbf{v}(t') = \mathbf{v}_j(t) \quad (2.2)$$

with

$$t' = (j - 1) \cdot T_w + t - \left( t_i - \frac{1}{2} (T_w - 1) \right) \quad (2.3)$$

where

$$t \in \left\{ t_i - \frac{1}{2} (T_w - 1), \dots, t_i + \frac{1}{2} (T_w - 1) \right\} \quad (2.4)$$

$$j \in \{1, \dots, M\}. \quad (2.5)$$

The stationary unitary event analysis is then performed on these new parallel spike trains. It can be summarized as follows. The average firing probabilities of the processes within the time window determine the expected number of coincidences—the empirical number of coincidences results from a counting process within the window over all trials. The significance for excess (or lacking) coincidences is evaluated by comparing the expected and empirical numbers using the joint-surprise measure, a logarithmic transform of the joint-p-value. The latter expresses the probability of observing the empirical number of coincidences by chance, under the null-hypothesis of independent Poisson processes. The full data set is analyzed by successively moving the time window from one position  $t_i$  to the next (usually in steps of one bin) and repeating the procedure sketched above.

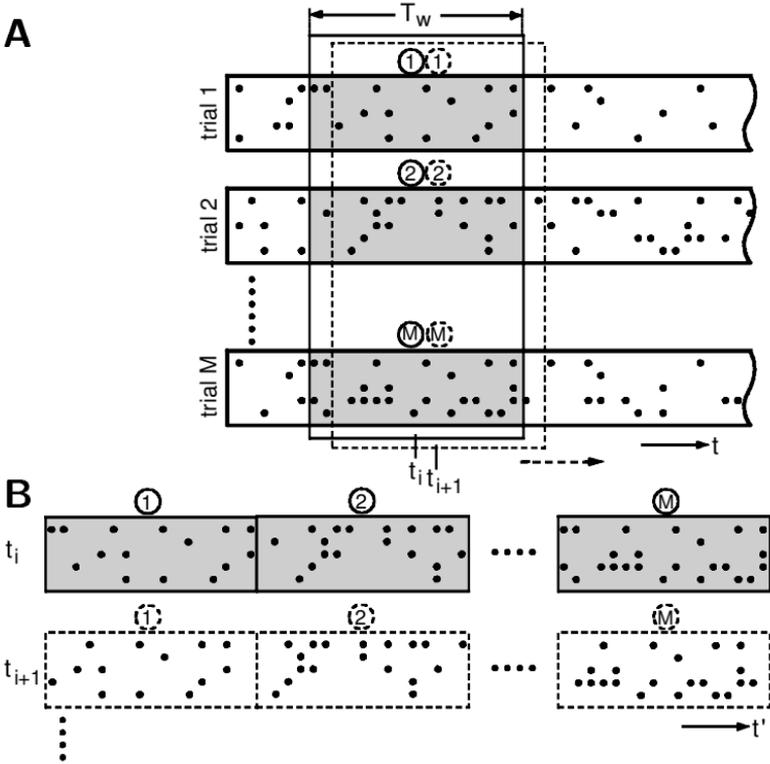


Figure 1: Sketch of the moving window analysis. (A) Parallel spike trains of five neurons (spikes marked as dots) for several repetitions 1, 2,  $\dots$ ,  $M$  (trials) of the same experiment. A window of width  $T_w$  centered at a given point in time  $t_i$  defines the segment of the data (shaded in gray), which enters the analysis at  $t_i$ . (B) From the data in each such time segment, a new time axis  $t'$  is constructed by concatenating the windows from all trials. Unitary events analysis is then performed on this new process. The full data set is analyzed by successively moving the window to the next point in time and repeating the above procedure (indicated by dashed). Typically, the window is shifted in steps of the time resolution of the spike data.

### 3 Dependence of Significance on Spike Rates

In this section we describe the performance of the UEMWA method under different conditions, including various time courses of the firing rates and the coincident rate. Special emphasis will be put on the width of the analysis time window, which sets detectability limits. After introducing a common theoretical framework, we discuss stationary and nonstationary rates. We first formulate a thought experiment in which we analytically describe the variables of the processes as expectation values. This will serve as a de-

scription of the average realization. Theoretical results will be illustrated by actual realizations in the form of simulated data.

**3.1 Description of the Performance Test.** For convenience we will restrict our considerations to two processes, both having the same rates. The results, however, are easily extended for  $N > 2$  processes and for differing rates. Dependencies between the two processes and their consequences on our measures are studied by injecting coincident Poisson events at a given rate level  $\lambda_c(t)$  into two independent Poisson processes with (background) rates  $\lambda(t)$  (for a detailed discussion, see Grün, Diesmann, Grammont, Riehle, & Aertsen, 1999). Temporal resolution  $h$  is assumed to be 1 ms. Let us now derive expectation values for the empirical number of coincidences  $n^{\text{emp}}$  in time interval  $T_w$  and for the number of coincidences  $n^{\text{pred}}$  we expect to find on the basis of the rates. The probability of finding a coincidence at time  $t$  is  $\lambda_c(t)h + (\lambda(t)h)^2$ . Therefore, the expectation value for the empirical coincidence count in  $T_w$  is

$$n^{\text{emp}}(t) = M \cdot \sum_{\tau=t-\frac{1}{2}(T_w-1)}^{t+\frac{1}{2}(T_w-1)} [\lambda_c(\tau)h + (\lambda(\tau)h)^2]. \quad (3.1)$$

Knowing only the rate of the processes ( $\lambda_c(t) + \lambda(t)$ ), the expected number of coincidences in a window  $T_w$  centered at  $t$  is

$$n_*^{\text{pred}}(t) = M \cdot \sum_{\tau=t-\frac{1}{2}(T_w-1)}^{t+\frac{1}{2}(T_w-1)} [(\lambda_c(\tau) + \lambda(\tau))h]^2. \quad (3.2)$$

However, given experimental data, the firing rates are not known and have to be estimated from the data. Assuming that the rates are stationary over the duration of a time window  $T_w$ , the rate can be estimated by the ratio of the spike count and the number of time steps. (The effect of rate estimation on stationary unitary events analysis is discussed in the companion article.) Having available only an estimate of the average firing rate, equation 3.2 reads

$$n^{\text{pred}}(t) = MT_w \cdot \left[ \frac{1}{T_w} \sum_{\tau=t-\frac{1}{2}(T_w-1)}^{t+\frac{1}{2}(T_w-1)} (\lambda_c(\tau) + \lambda(\tau))h \right]^2. \quad (3.3)$$

Comparing equation 3.3 with equation 3.2, we have, using the assumption of stationarity, effectively exchanged the sum and the squaring. The difference between  $n^{\text{pred}}(t)$  and  $n_*^{\text{pred}}(t)$  determines the error made when nonstationary processes are analyzed with an averaging window of width  $T_w$ . (See appendix B for a parametric study of this deviation in terms of false positives.) However, conceptually, rate estimation and coincidence statistics can

be separated and performed using different methods (see section 5). The following measures will be analyzed and illustrated for different time courses of the background rates  $\lambda(t)$  and the injected rate  $\lambda_c(t)$ :

- The number of coincidences expected to occur in the course of time. Both the number of occurrences assuming independence ( $n^{\text{pred}}$ ; equation 3.3) and the empirical number of coincidences ( $n^{\text{emp}}$ ; equation 3.1) will be evaluated.
- The joint-surprise  $S$  as a function of time, resulting from the comparison of  $n^{\text{pred}}$  and  $n^{\text{emp}}$ . The first characterizes the underlying distribution under the null-hypothesis of independence, and the second describes the deviation from independence.

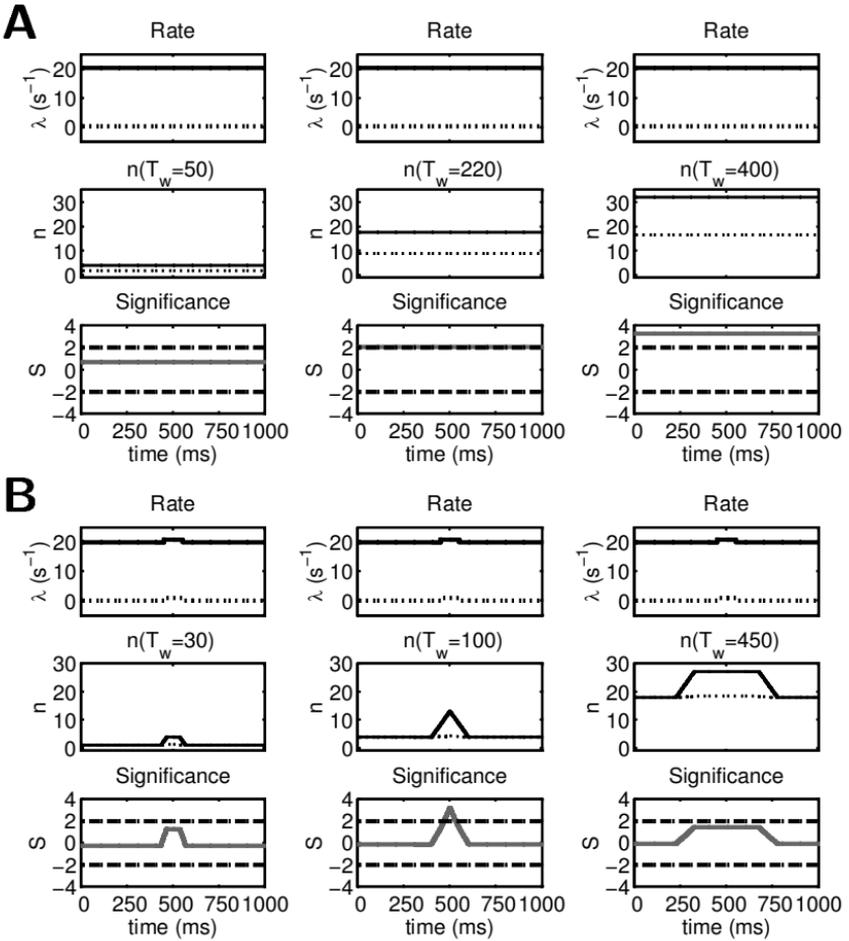
Data and results will be displayed as sequences of three subfigures (columns in Figures 2 and 4): at the top empirical (solid line) and expected coincidence rates (dotted line), in the middle the number of coincidences (empirical [solid] and expected [dotted]), at the bottom the joint-surprise (gray) including the upper and lower significance levels for  $\alpha = 0.01$  (dashed). All measures are shown as functions of time over the duration of the trial.

## 3.2 Stationary Background Rates

*3.2.1 Stationary Coincidence Rate.* The simplest “experimental” situation is given if both background rates and the injected coincidence rate are sta-

---

Figure 2: *Facing page.* Stationary background rates: relevance of window size. (A) Stationary coincidence rate. Three examples of sliding window analysis for different window widths (columns from left to right,  $T_w = 50, 220, 400$ ;  $M = 100$ ,  $h = 1$  ms). (Top) Coincident rate (dotted line) and compound rates (solid line). (Middle) Number of coincidences empirical (solid line) and expected (dotted line). (Bottom) Joint-surprise (gray curve) and upper and lower significance threshold (dash-dot lines) for significance level  $\alpha = 0.01$ . In the left column,  $T_w$  is smaller than the minimal window size  $T_\alpha$ ; the number of detected coincidences is not significant. In the middle column,  $T_w$  equals  $T_\alpha$  and thus is at the border of significance. Only when  $T_w > T_\alpha$  (right column) is the detected number of coincidences significant. All measures are independent of time because of stationarity of the underlying rates. (B) Nonstationary coincidence rate. Three examples of sliding window analysis for different window widths (columns from left to right  $T_w = 30, 100, 500$ ). Graphs as in A. In the left column,  $T_w$  is narrower than the minimal window size  $T_{\min}$ ; the number of detected coincidences is not significant. In the middle column,  $T_w$  equals  $T_c$  and, hence,  $S$  has a triangular shape. Since  $T_w$  is larger than  $T_{\min}$ , significance is reached for more than one window position. In the right column,  $T_w > T_{\max}$ ; the hot region is not detected as significant.



tionary, that is, independent of time (see Figure 2A). As shown in section C.1, for each combination of  $\lambda$  and  $\lambda_c$ , there is a minimal analysis window width  $T_\alpha$  needed to detect coincidences as significant events. If the window is chosen too narrow, the number of excess coincidences is too small to deviate significantly from the expected number. In the stationary case, detection of excess coincident events can always be ensured by increasing the analysis window, since the larger the analysis window is, the more excess coincidences are detected. The examples in Figure 2A demonstrate the dependence of the significance on the size of the analysis window (columns:  $T_w < T_\alpha$ ,  $T_w = T_\alpha$ ,  $T_w > T_\alpha$ ) for a fixed combination of  $\lambda$  and  $\lambda_c$ . Observe that in the left column,  $T_w$  is clearly too small to detect the injected coincidences, just on the border for detection in the middle column, and clearly large enough in the right column.

**3.2.2 Nonstationary Coincidence Rate.** We now discuss the case where coincidences occur clustered in time (i.e., form “hot regions”) on top of a constant background rate (see Figure 2B). In that case, the detection of excess coincidences is constrained to a range between a minimal and a maximal analysis window size.

Consider a situation where we place the analysis window in the middle of a hot region of width  $T_c$ , and gradually increase the width of the analysis window  $T_w$  (details are in section C.2; cf. Figure 10B). As long as  $T_w \leq T_c$ , we face the situation of stationary injected events, discussed in the preceding paragraph (cf. Figure 10A). Hence, we need a minimum width of the analysis window ( $T_\alpha$ ) to detect the cluster of injected events.  $T_\alpha$  depends only on the combination of  $\lambda$  and  $\lambda_c$ ; its value can be obtained from the calibration graph for the stationary situation (see Figure 10A, bottom). When the analysis window exceeds the hot region ( $T_w > T_c$ ), the total number of coincidences increases further; now, however, due only to coincidences occurring by chance on the basis of the background activity. Thus, by increasing the analysis window further, excess coincidences are averaged with independent coincidences, and at some window size,  $T_{\max}$  will not be detected as significant anymore.  $T_{\max}$  defines the maximal window for detecting excess coincidences as significant.

If  $T_c < T_\alpha$ , the cluster cannot be detected, even with arbitrarily large analysis windows (assuming the number of trials to be fixed). Thus, in contrast to the stationary condition, the existence of a  $T_{\min}$  depends on the width of the hot region  $T_c$ . A cluster of excess coincidences is not detectable if its duration  $T_c$  remains below the critical time span  $T_\alpha$ . If the cluster is detectable ( $T_c \geq T_\alpha$ ,  $T_{\min} = T_\alpha$ ), it may still go undetected if the analysis window is too small,  $T_w < T_{\min}$ , or too wide,  $T_w > T_{\max}$ . The range of appropriate window sizes can be obtained from calibration graphs as in Figure 10B in appendix C. Examples for analysis windows that are too narrow, appropriate, and too wide are presented in Figure 2B.

Let us now discuss the case where the analysis window is shifted gradually into a hot region. Once the window has overlap with the hot region, the injected coincidences contribute to the coincidence count. With increasing overlap, the number of contributed coincidences grows linearly, and the joint-surprise increases accordingly (see Figure 2B). A plateau is reached when the analysis window is completely inside the hot region ( $T_w < T_c$ ) or when the hot region is completely covered by the analysis window ( $T_w > T_c$ ). Further shifting eventually leads to a decrease in overlap and to a time course of the joint-surprise  $S$ , which is symmetrical around the center of the hot region. The trapezoidal shape of  $S$  degenerates to a triangle in the special case  $T_w = T_c$ .

The argument just given shows that  $S$  can pass the significance threshold (i.e., the plateau surpasses threshold) only if the analysis window  $T_w$  has the appropriate size  $T_{\min} \leq T_w \leq T_{\max}$ . The duration of the plateau is given by  $T_p = |T_c - T_w|$ . Table 1 summarizes the relationships between the observable

Table 1: Relationships of the Plateau Duration  $T_p$ , Width of the Analysis Window  $T_w$ , and Extent of the Hot Region  $T_c$ .

	$T_w < T_c$	$T_w = T_c$	$T_w > T_c$
$T_p = 0$	—	$T_c = T_w$	—
$T_p < T_w$	$T_c = T_w + T_p$	—	$T_c = T_w - T_p$
$T_p \geq T_w$	$T_c = T_w + T_p$	—	—

Notes:  $T_w$  is compared to  $T_p$  (rows) and to  $T_c$  (columns). Table entries marked by a dash represent nonexistent ( $T_p, T_w, T_c$ ) combinations. The trapezoidal shape of the joint-surprise  $S$  reaches significance if  $T_{\min}$  exists ( $T_c \geq T_\alpha$ ) and  $T_{\min} \leq T_w \leq T_{\max}$ .

variables  $T_w$ ,  $T_p$  and the variables generating the trapezoid  $T_w$ ,  $T_c$ . Using these relationships, one can determine the extent of the hot region  $T_c$  by measuring the size of the plateau and systematic variation of the analysis window (see section C.2 for a detailed derivation). Figure 3 summarizes the possible interactions of the width of an excess interval  $T_c$  and the width of the analysis window  $T_w$ .

If the significance threshold is reached at all, typically more than one window is significant. Only in the special case of  $T_w = T_c = T_\alpha$  is the maximum of  $S$  exactly at threshold level and for a single window only. For a given combination  $T_c$ ,  $T_w$ , we can compute the minimal overlap of the two windows needed to detect the injected coincidences as significant by the use of  $T_{\max}$  and can construct the extent of the region  $T_s$  in which injected coincidences are marked as significant (see section C.2)—that is, the time span between the two intersections of  $S$  with the significance threshold. According to the unitary event analysis, all coincidences in a significant window are marked as “special.” Thus, only in the case of minimal  $T_s$  exactly the coincidences in the region  $T_c$  are marked as special. The smaller  $T_w$ , the better the extent of the marked coincidences approximates  $T_c$ .

### 3.3 Nonstationary Background Rates

**3.3.1 Stationary Coincidence Rates.** We consider two different cases of nonstationary background rates: stepwise and gradual increase of  $\lambda(t)$ , both in combination with a stationary rate of injected coincidences  $\lambda_c$  (see Figures 4A and 4B). A stepwise increase in rate does not lead to a discontinuous change in the coincidence counts and the significance measure, due to the smoothing effect of the moving window (see Figure 4A). On a larger timescale,  $n^{\text{pred}}$  and  $n^{\text{emp}}$  increase parabolically because of their quadratic dependence on the background rate (see Figure 4A, middle graph). The difference between  $n^{\text{pred}}$  and  $n^{\text{emp}}$  is constant throughout the trial. However, due to the absolute increase in background rate, the significance decreases

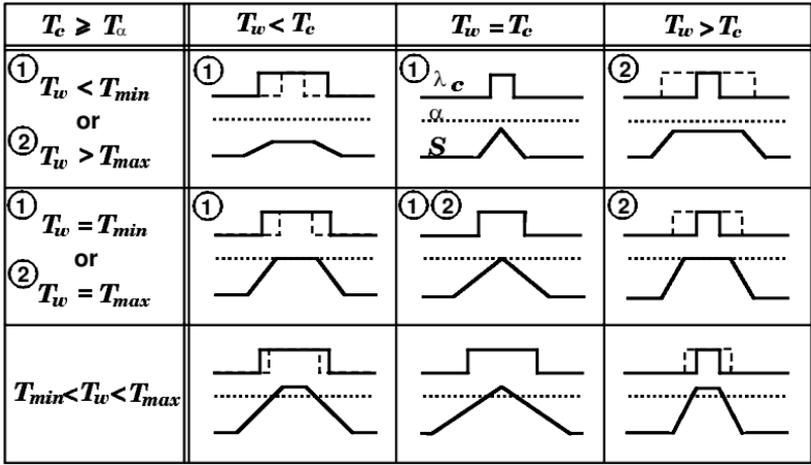


Figure 3: Detectability of short epochs of excess coincidences. The figure illustrates the influence of the choice of  $T_w$  for the detectability of a hot region  $T_c$ .  $\lambda$  and  $\lambda_c$  are constant, and  $T_c \geq T_a$  (detectability in principle) is assumed. Each individual graph shows the time course of  $\lambda_c$  (solid) in its top part (width of analysis window  $T_w$  indicated by the dashed line) and the time course of the significance measure (solid) relative to the significance threshold (the dotted line) in the bottom part. The ordinates of the significance measure are individually scaled for better visibility. The first column covers the constellations where  $T_w < T_c$ , the second  $T_w = T_c$ , and the third  $T_w > T_c$ . The rows are organized by the size of  $T_w$  relative to the interval  $[T_{min}, T_{max}]$  for which detection is possible. The first row depicts cases where injected coincidences are not detected as significant because  $T_w$  is outside the interval  $[T_{min}, T_{max}]$ . The second row shows cases where injected coincidences are just at the border of detectability, because  $T_w$  equals either  $T_{min}$  or  $T_{max}$ . For the case of  $T_w = T_c$  and  $T_w$  equal to one of the detection boundaries,  $T_{min}$  and  $T_{max}$  are identical. The third row illustrates cases where  $T_w$  is inside the interval  $[T_{min}, T_{max}]$ . Here, coincidences are detected as significant for a range of window positions.

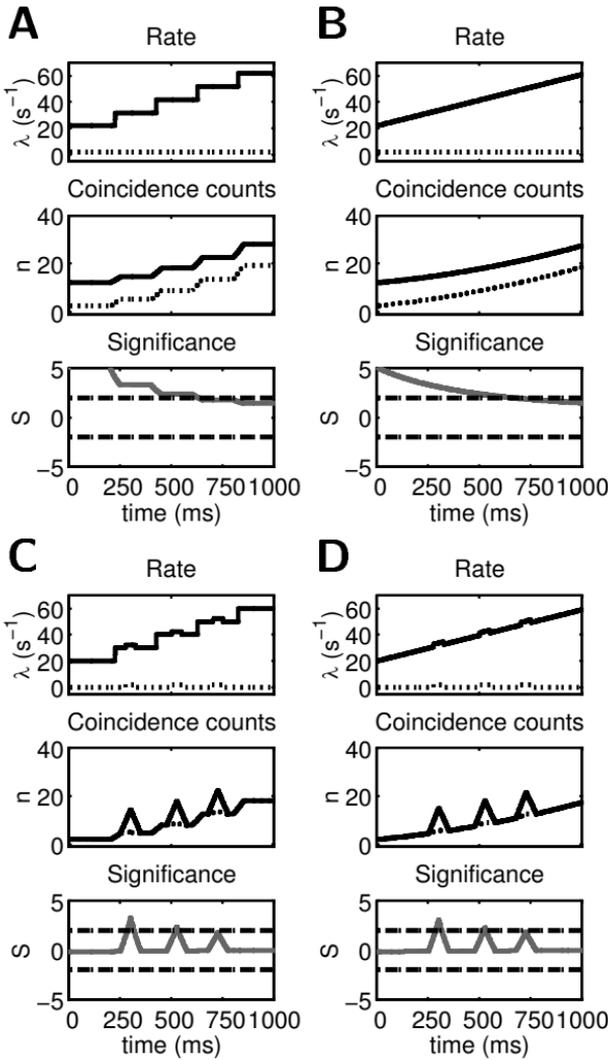
(see the companion article for an extended discussion of this issue). A linear increase of background rate (see Figure 4B) basically gives the same result. In both examples, the injected coincidences ( $\lambda_c = 2 \text{ s}^{-1}$ ) do not reach significance ( $\alpha = 0.01$ ) at a background rate of  $\lambda = 50 \text{ s}^{-1}$ . Above this rate, injected coincidences can be detected only with a larger analysis window  $T_w$ .

**3.3.2 Nonstationary Coincidence Rates.** Finally, we investigate the general case where the neuronal processes have time-dependent rates and the excess coincident activity occurs in a short interval, triggered by some external or internal event. When the neuronal processes are observed over repeated trials, the coincident activity appears to some degree locked to certain points

in time. For the purpose of this article, the situation described above is our model for the composition of neuronal spike trains. Firing rates and coincidence rates vary independently and consistently over the time course of a trial. Regions of increased coincidence rate may be accompanied with elevated firing rates, with suppressed activity, or without noticeable changes in firing rate. Such data are optimally suited for UEMWA. Reproducibility over trials allows for reliable estimates of firing rates and coincidence rate in relatively narrow time windows.

Consider a data set where several hot regions appear during the trial, while the firing rate is increasing with time. As in the preceding section, two types of increase (stepwise in Figure 4C and a constant slope in Figure 4D) are compared. The width of the analysis window is chosen as  $T_w = T_c$ . We again analyze the situation using the theory for the expectation values worked out in section 3.1. Figures 4C and 4D show the results of this analysis. Overall,  $n^{\text{pred}}$  and  $n^{\text{emp}}$  are increasing over time; in addition,  $n^{\text{emp}}$  exhibits strong peaks in the hot regions.  $S$  reflects the transients in the hot regions while staying at naught in between. This clearly demonstrates the rate-normalizing property of the joint-surprise measure. The triangular shape of the peaks is explained by the condition  $T_w = T_c$  (cf. Figure 2B, center column). Since  $\lambda_c$  is the same for each hot region, the peaks in  $n^{\text{emp}}$  relative to baseline are of equal height (see equation 3.1). With a constant number of excess coincidences and increasing background level, the significance decreases (cf. Figure 4 in the companion article and Figure 4A here). Thus, in our example, the height of the peaks in  $S$  (see Figures 4C and 4D) decreases over time. The last hot region is just on the border of detectability. Apart from differences in the fluctuations of  $S$ , caused by the discretization inherent in the joint-surprise measure (cf. Figure 10), the two types of rate variations are practically indistinguishable at the level of  $S$ . In case  $T_w \neq T_c$  (not shown here), the time course of  $S$  would exhibit plateau-like shapes around the hot regions. For the case of linearly increasing background rates, the plateaus would be oblique instead of flat. The slope of the plateau would then be comparable to the situation of stationary coincidences rates (cf. Figure 4B).

Before we apply UEMWA to experimental data, we will leave the theory for the expectation values and illustrate the procedure using simulated point processes with a realistic number of repetitions. Figure 5A shows simulations of two parallel processes in repeated trials. Both spike trains are simulated as independent Poisson processes (see the companion article for details). The first one has a nonstationarity in firing rate: at a certain point in time, the firing rate raises stepwise. The second process is stationary. Clusters of coincident events were injected around two points in time. In Figure 5A, all coincidences found (irrespective of their significance, termed “raw”) are marked by squares. The two clusters can clearly be seen. However, as expected, coincidences also occur outside the hot regions—more of them in the regime where the firing rate of one of the neurons is elevated. In Figure 5B, only those coincidences are marked that occur in windows



where  $S$  exceeds the significance threshold. The time course of the joint-surprise is shown in Figure 5C. As expected from our considerations above, the joint-surprise remains at baseline outside the hot regions. Thus, only coincidences in the hot regions and, because of the limited temporal resolution of UEMWA, in a small region around them are marked. The triangular shape of the peaks in  $S$  indicates that  $T_w$  was close to  $T_c$ . Only the top of the triangle is above threshold, meaning that  $T_w$  is not much larger than  $T_{min}$ . Consequently, the region of marked coincidences in Figure 5B gives a good estimate of the width of the cluster  $T_c$ . The fluctuations of  $S$  in repeti-

tions of the same experiment are illustrated in Figure 5D. The time course of the boundaries enclosing at least 70% of the realizations (dark gray curves) shows that the coincidences in the first hot region are detected with a probability exceeding 85% (already the lower boundary is above significance threshold). Sensitivity is lower for the second hot region because of a higher background rate. The behavior of the lower boundaries in the regime of low firing rates (the analysis window being outside the hot region) exemplifies the problem of detecting a lack in the number of coincidences. The size of the analysis window and the given number of trials does not allow for the detection of lacking coincidences at a reasonable significance level because the probability of finding no coincidences already is  $\approx 14\%$  (compare Aertsen & Gerstein, 1985).

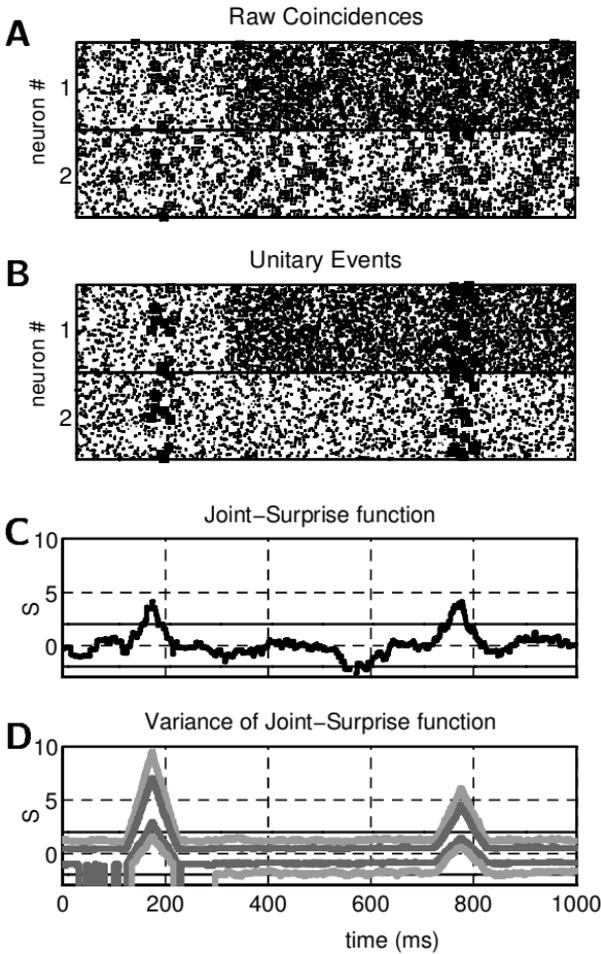
#### 4 Unitary Events in Cortical Activity

---

In the following, we present results from the analysis of simultaneously recorded multiple single-neuron spike trains from frontal and motor cortex in awake, behaving monkeys.

---

Figure 4: *Facing page*. Nonstationary background rates (graphs as in Figure 2; in all graphs the analysis window is  $T_w = 50$ ,  $M = 100$ ,  $h = 1$  ms). Stationary coincidence rates combined with (A) stepwise increasing background rates ( $\lambda = 20, 30, 40, 50, 60$  s $^{-1}$ ) and (B) continuously increasing background rates ( $\lambda = 20$  to  $60$  s $^{-1}$ ). Coincidences are injected at rate  $\lambda_c = 2$  s $^{-1}$  (dotted line). Compound rates are shown as solid curves. In A, coincidence counts (middle panel) reflect rate changes, with some smoothing due to the moving window (dotted: expected; solid: empirical). Above a certain background rate, injected coincidences are masked by coincidences expected from independent rates and are no longer significant (bottom panel, gray curve). For continuously increasing background rates (B, top panel, solid line), significance is lost for firing rates that are too high, as in A (bottom panel, gray curve). Nonstationary coincidence rates combined with (C) stepwise increasing background rates (as in A) and (D) continuously increasing background rates (as in B). Coincidences are injected in three “hot regions” (width 50 ms,  $\lambda_c = 2$  s $^{-1}$ ). Coincidence counts (middle panel) reflect increasing rates and injected coincidences in the hot regions (solid: empirical; dotted: expected). The triangular shape of coincidence counts results from smoothing by the moving analysis window, its width being equal to the widths of the hot regions (cf. Figure 3). In the bottom panels, joint-surprise remains at zero in regions where no coincidences are injected; significant excursions occur in the hot regions. Typically, several consecutive windows detect coincidences as significant. In the last hot region, only a single window is significant, because the size of the hot region matches the size of the analysis window, which is the minimal window for detectability at this combination of rates. Results in C and D are comparable. Dash-dotted lines indicate significance threshold ( $\alpha = 0.01$ ) for excess (upper) and lacking (lower) coincidences.



**4.1 Motor Cortical Activity.** In order to investigate the possible relation between the dynamics of neuronal interactions in the motor cortex and the behavioral reaction time (RT), a task was designed in which RT can be experimentally manipulated (Riehle, Seal, Requin, Grün, & Aertsen, 1995; Riehle, Grün, Diesmann, & Aertsen, 1997). Briefly, monkeys were trained to touch a target on a video display after a preparatory period (PP) of variable duration. To start a trial, the animal had to push down a lever. The preparatory signal (PS) was given by an open circle on the video display. After a delay of variable duration, during which the animal had to continue to press the lever, the response signal (RS) was indicated by a filling circle. Four durations of the PP, lasting 600, 900, 1200, and 1500 ms, occurred with equal probability and in random order. RT is defined as the period between the occurrence of the RS and the release of the lever, whereas movement

time (MT) is defined as the period between releasing the lever and touching the screen. After training, the monkeys were prepared for multiple single-unit recording. A multielectrode microdrive (Reitböck, 1983; Mountcastle, Reitböck, Poggio, & Steinmetz, 1991) was used to insert transdurally seven independently driven microelectrodes, spaced  $330\ \mu\text{m}$  apart, into the primary motor cortex (MI) (for details see Riehle et al., 1995; Riehle, Grün, Aertsen, & Requin, 1996; Riehle et al., 1997).

Figure 6 presents an example of modulation of coincident spiking activity during the preparation for movement. The first observation is that the number of coincidences marked as significant coincidences (see Figure 6C) is considerably reduced as compared to the raw coincidences (see Figure 6B). Second, unitary events show a distinct timing structure, with two phases of synchronized activity: about 100 ms after PS (lasting for about 200 ms) and after ES1 (lasting also about 200 ms). The composition of unitary events within these phases is the same: for a first short period, neurons 2 and 3 are synchronized; then neuron 3 switches its partner and is successively synchronized with neuron 1. Taking into account the condition under which

---

Figure 5: *Facing page*. Simulated nonstationarities. (A) Spike times (dots) of two parallel processes, simulated for 1000 ms ( $h = 1\ \text{ms}$ ) over 100 trials (upper panel process 1, lower panel process 2, trials displayed in consecutive rows). Process 1 has a nonstationarity in firing rate at 300 ms from trial start. Firing rate increases stepwise from  $20\ \text{s}^{-1}$  to  $60\ \text{s}^{-1}$ . Process 2 is stationary at  $20\ \text{s}^{-1}$ . Centered at 175 ms and 775 ms from trial onset, two hot regions ( $T_c = 50$ ) are generated by injecting additional coincidences at rate  $2\ \text{s}^{-1}$ . All coincidences occurring in the simulation are marked by squares (“raw coincidences”). (B) Same data as in A. Here, only coincidences are marked by squares that occur in analysis windows passing the significance threshold: unitary events (analysis parameters:  $T_w = 50$ ,  $\alpha = 0.01$ ). (C) Joint surprise corresponding to the data shown in A and B as a function of time (thick curve), representing at each instant in time the significance resulting from the analysis window centered around this point in time. Thin lines:  $\alpha = 0.01$  for excess (upper) and lacking (lower) coincidences. At around 175 ms and 775 ms, the joint-surprise function passes the significance level for excessive coincidences. (D) Variance of the joint-surprise functions estimated from 1000 repetitions of the simulation experiment shown in A–C. Significance level indicated as in C for orientation. Gray curves represent the width of the distribution as a function of time (dark gray: minimum 70%; light gray: 95% area). In the regime where both background rates are 20 Hz, the probability of finding no coincidences is  $\approx 14\%$ . Therefore, no lower boundary for the minimum 95% area region can be drawn. No coincidence count does exist such that the cumulative probability of lower counts is less than 2.5%. For the 70% area region, the lower boundary is at coincidence count 1. In some time steps, probability to obtain no coincidences exceeded 15% because of the finite number of repetitions.

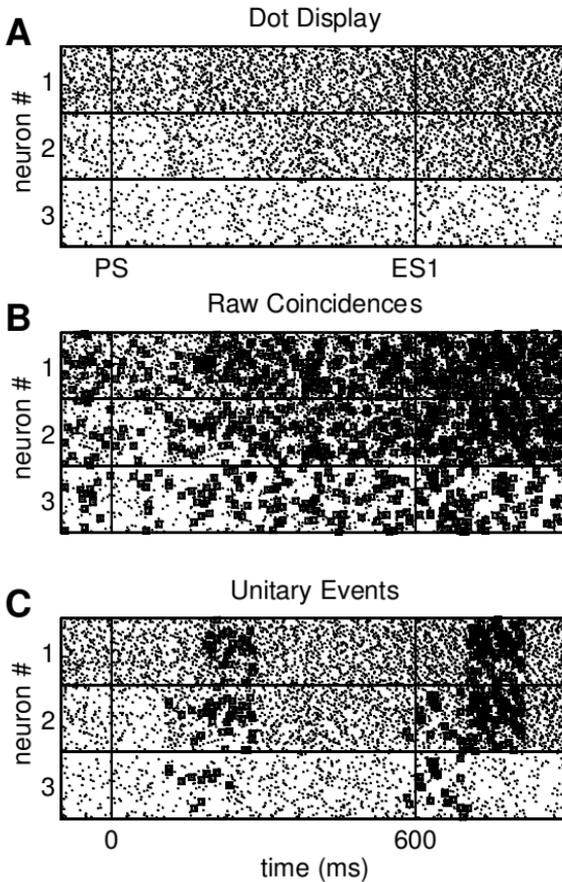


Figure 6: Time structure of coincident spiking activity. The three panels of dot displays show the same spike data from three simultaneously recorded neurons in the primary motor cortex of a monkey involved in a delayed-response task. (A) Spiking activity of three neurons (1, 2, 3) organized in separate displays showing 96 trials. Data are pooled from three types of trials (PP 900, 1200, and 1500 ms and aligned on the preparatory signal (PS, vertical line at time 0). Only the first 800 ms after PS are shown; this includes the end of the first potential end of the waiting period ES1 (“expected signal”; vertical line at time 600 ms). In the data analyzed here, no movement instruction occurred at ES1. (B) Same spike data as in A. All “raw” coincidences are marked by squares (bin width 5 ms). (C) Same spike data as in A and B. Unitary events are marked by squares (UEMWA, window width 100 ms,  $\alpha = 0.05$ ). (Modified from Riehle et al., 1997).

these unitary events occur, one may speculate that the occurrence of unitary events can be interpreted as activation of a cell assembly that is involved with the initiation (or reinitiation) of a waiting period.

**4.2 Frontal Cortical Activity.** In the second experimental study we discuss, rhesus monkeys were trained to perform a “delayed localization task” with two basic paradigms (localizing and nonlocalizing; an example of the latter is shown Figure 7). In both paradigms, the monkey receives a sequence of two stimuli (visual and auditory) out of five possible locations. After a waiting period, a “GO” signal instructs the monkey to move its arm in the direction of the stimulus relevant in the current trial. In the localizing paradigm, the relevant spatial cue was selected by the color of the GO signal. In the nonlocalizing paradigm, an indicator light between blocks of trials informed the monkey about the reinforced direction for arm movement. Thus, in the latter case, the animal had to ignore the spatial cues given before the GO signal. In the behavioral paradigm analyzed here (nonlocalizing), neither the spatial cues before the GO signal nor the GO signal itself could be used to determine the correct behavioral response (see Vaadia, Bergman, & Abeles, 1989; Vaadia, Ahissar, Bergman, & Lavner, 1991; Aertsen et al., 1991, for further details). The activity of several (up to 16) neurons from the frontal cortex was recorded simultaneously by six microelectrodes during performance of the task. In each recording session, the microelectrodes were inserted into the cortex with interelectrode distances of 300 to 600  $\mu\text{m}$ . Isolation of single units was aided by six spike sorters that could isolate activity of two or three single units, based on their spike shape (Abeles & Goldstein, 1977). The spike sorting procedure introduced a dead time of 600  $\mu\text{s}$  for the spike detection.

Using data from this study, we found that coincident activity in the frontal cortex can be specific to movement direction. We parsed the data of five neurons according to the movement direction and analyzed each of these subsets separately. Figure 7 shows the analysis results for two movement directions (*A*: to the left; *B*: to the front); for the three other movement directions, there was no significant activity. For each of the two movement directions, there is mainly one cluster of unitary events (besides some sparsely spread individual ones), occurring at the onset of the movement. The clusters of unitary events differ, however, in both their neuronal composition and their timing. During movement to the left, significant coincidences occur between neurons 6 and 9; for movement to the front, they occur between neurons 6 and 10. The timing of the unitary events differs when measured in absolute time after the GO signal (to the left: 355 ms; to the front: 400 ms); however, both occur shortly after LEAVE. Thus, unitary events appear to be locked better to the behavioral event (LEAVE) than to the external event (GO). The analysis of the same five neurons during the localizing task, where the color of the GO signal contained the information about the reinforced type of stimulus (data not shown), did not reveal any indications for unitary

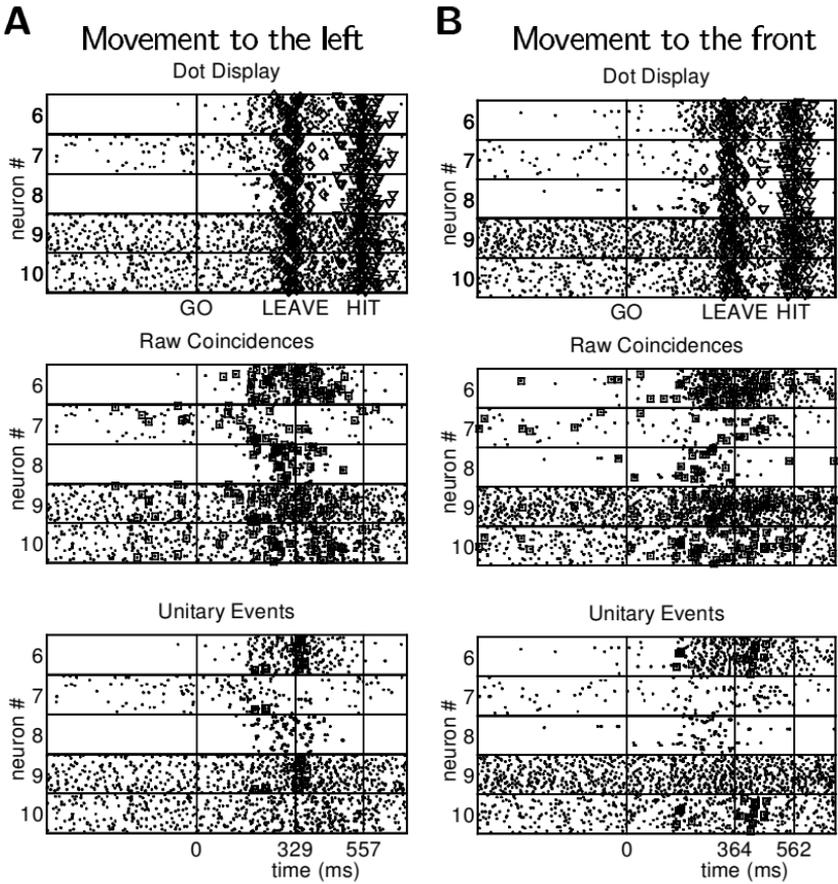


Figure 7: Task dependence of coincident spiking activity. The dot displays show the spiking activity of five simultaneously recorded neurons (labeled 6 to 10) from the frontal cortex of a monkey involved in a delayed localization task (28 trials). The two columns represent two different behavioral conditions: (A) movement to the left, (B) movement to the front. Organization of the columns (A, B) is the same as in Figure 5, with bin width of 3 ms for coincidence detection and an analysis window width of 60 ms, and 0.01 significance level for UEMWA. Data are taken from segments starting 500 ms before and ending 700 ms after the GO signal (vertical line at time 0 ms). The top row dot displays include two behavioral events, LEAVE (monkey leaves central key) and HIT (monkey hits target), marked by diamonds and triangles, respectively. Average times of behavioral events are indicated by vertical lines labeled LEAVE (A: 329 ms; B: 364 ms) and HIT (A: 557 ms; B: 562 ms).

events related to movement direction. Note that neuron 6 is participating in significant coincident activity in both movement directions, however, with another coincidence partner in each. This is indicative of a common membership of neuron 6 in two different cell assemblies, one of which is activated depending on the movement direction.

## 5 Discussion

---

We developed the unitary events analysis method to detect excess coincidences in multiple single-neuron spike trains. In the companion article, we evaluated the method for the case of stationary rates and calibrated the method for physiological relevant parameters: firing rates, coincidence rates, and number of neurons analyzed in parallel. The method was shown to be very sensitive to excess coincidences; their significance can be evaluated using the joint-surprise measure. In this article, we extended the method to incorporate nonstationary firing rates by introducing the UEMWA. This method performs the analysis for unitary events within analysis windows of fixed length, which are slid in small steps along the data, in order to follow the dynamic changes of firing rates. Within each window, we assume stationarity and apply unitary event analysis as in the stationary case. The resulting time-dependent joint-surprise function provides a convenient measure for the probability that the number of coincident spiking events in a certain observation interval represents a chance event. By imposing a threshold level on the joint-surprise function, certain time segments of the data are highlighted as potentially interesting regarding the presence of excess coincident spiking events, referred to as unitary events. Their neuronal composition, as well as the moments they occur in time, may give us information about the underlying dynamics of assembly activation.

**5.1 Appropriate Size of Analysis Window.** The width of the moving window is clearly an important parameter and may be adjusted according to the data. In the calibration study described in this article, (see section 3), we analyzed a model in which excess coincidences were injected into independent background activity. In a first step (see section 3.2.1), we studied the sensitivity under stationary conditions and in dependence of background and coincidence rate levels, using analytical descriptions for coincidence counts. For a given rate constellation, an increase of the analysis window leads to a linear increase of the coincidence count. In order to reach significance, a certain level of excess coincidences needs to be present to pop out from the chance coincidences due to background activity. This requires a minimum size of the analysis window, specific for the given background and coincidence rates. The larger the coincidence rate is, the smaller is the minimal window size. By contrast, the larger the background rate is, the larger is the minimal window size.

In a second step (in section 3.2.2), we explored the detectability of non-stationary coincidence rates. We studied the case that excess coincidences occurred only within a restricted time interval (“hot regions”), motivated by experimental observations (Riehle et al., 1997). Such hot regions may be the result of loose time-locking of synchronous spiking to an (external) trigger event. By studying the detectability in the symmetrical case (the analysis window is centered in the hot region), we found that there is not only a minimal window size as discussed in the stationary condition but also a maximal window size. By increasing the analysis window (starting from a window size smaller than the width of the hot region), more and more excess coincidences are “seen,” which, upon reaching the minimal window size, are detected as unitary events. If the analysis window covers exactly the hot region, all injected coincidences are detected, and maximal detectability is reached, that is, the joint-surprise reaches its maximum. When further increasing the window, the number of excess coincidences no longer grows; however, the contribution of chance coincidences will increase, leading to a decrease of detectability until the joint-surprise finally drops below significance. Thus, if the analysis window is too large or too small, the hot region is not detected, although the hot region would be detectable with the appropriate choice of analysis time window.

Using the results from the symmetrical case, we analyzed the situation when the analysis window is gradually shifted into the hot region. Once the two start to overlap, the injected coincidences contribute to the coincidence count. With increasing overlap, the joint-surprise increases accordingly, until it reaches its maximum at maximal overlap. When the analysis window leaves the hot region, the overlap decreases again, and so does the joint-surprise. For symmetrical shape of the hot region (including spike densities and coincidence densities, as is assumed here), the joint-surprise is symmetrical around the center of the hot region. Injected coincidences are detected as unitary events if the size of the analysis window is between  $T_{\min}$  and  $T_{\max}$ . Only in the special case that  $T_w$  is equal to or narrower than the width of the hot region and the joint surprise just reaches threshold at its peak, unitary events are restricted to the extent of the hot region. Generally, however, the epoch over which unitary events are detected does not coincide with the extent of the hot region: it may be narrower but also wider, depending on the size of the various windows. Figure 3 summarizes the possible combinations. The time course of the joint-surprise function indicates how the width of the analysis window can be optimized. From the extent of the plateau, we can derive the width of the hot region. Table 1 summarizes the necessary relationships.

By shifting the time window  $T_w$  in smaller time steps than the width of the analysis window (usually we shift by one bin), we introduce dependencies between the time windows, since the analysis is applied to partly overlapping time segments. When shifting the window in single bin steps, each single joint-event will be considered in  $T_w$  analyses, although in slightly

differing contexts. As a result, a single event may be evaluated with differing significance values in the various analyses. Assuming continuity of the processes, the significance does not change drastically from one window to the next. One possibility for dealing with these dependencies is to give “bonus points” each time a joint-event surpasses a certain significance level. Thus, an individual spike constellation would collect bonus points as the window is shifted along the data, indicating that its accumulated significance “counts.” This procedure would lead to a gradual evaluation of the “unitarity” of those events. The decision for the final selection of unitary events could be based on a threshold on bonus points. For reasons of simplicity, however, we chose a simpler version: we define an event as unitary once it fulfills a certain significance criterion (usually  $\alpha = 0.05$  or  $0.01$ ) in at least one window. In terms of bonus points, this implies a selection on the basis of a threshold set to 1. A practical evaluation of the performance of a more elaborate bonus point rule is currently under study.

Nonstationary coincidence rates (e.g., a hot region) may be the result of loose locking of assembly activation to an (external) trigger event. In this situation, the optimal window for UEMWA is determined by the degree of temporal locking. In physiological data, however, several internal triggers that we do not know of may lead to hot regions of different temporal widths, which cannot optimally be captured by one sliding window size. Thus, an interesting perspective would be to develop an algorithm that dynamically adapts the width of the analysis window to the varying width of the hot regions.

**5.2 An Alternative Method of Unitary Event Detection: Cluster Analysis.** UEMWA deals with nonstationary firing rates by sliding an analysis window that is narrow enough to obtain firing rates that are approximately stationary over the extent of the window for all positions, along the trial. There is a second approach to obtain segments of data with joint-stationary rates, based on cluster analysis. In the following, we discuss this option briefly, because some cortical data indeed exhibit joint rate states and the approach has interesting relationships to other methods of analyzing multiple single-neuron spiking activity (e.g., hidden Markov models (HMM); Abeles et al., 1995; Seidemann, Meilijson, Abeles, Bergman, & Vaadia, 1996; Gat, Tishby, & Abeles, 1997). The idea is to segment the data into (exclusive) joint-stationary subintervals, using a standard cluster algorithm (e.g., Hartigan, 1975). Subsequently the data are analyzed for unitary events in the time segments, defined by the joint stationary rate states. We call this method *unitary event by cluster analysis method* (UECA) (a detailed description is given in appendix D). In this method, the width of the analysis window is defined by the covariations of the firing rates of single neurons. However, from our theoretical results about the detectability of hot regions, we know that the optimal width of the analysis window is given by the width of the hot region or, in more general terms, the time course of the coincidence rate. Thus,

a clustering approach is useful if the occurrence of excess coincidences is connected to the rate state. This, however, is not in agreement with experimental data (Riehle et al., 1997). Moreover, in UECA, the data are analyzed in exclusive regions. This implies transitions in significance from one time step to the next. Moreover, if a segment is detected by UECA as significant, that entire segment is marked as containing unitary events, even if a hot region covers only part of the segment. In the case of UEMWA, the analysis window positions are independent from rate transitions. They cover the entire data set step by step, resulting in a measure that is at the same time more localized (only data from a connected time interval enter the analysis) and smoother (a single event is weighed in several consecutive windows) than UECA. Taken together, for the experimental data we have analyzed so far, UEMWA provides a more differentiated picture of the presence of unitary events. We cannot exclude, however, that experimental settings may arise in which UECA, with its variable window size and the property that the counting statistics are not limited by the local window size, might be a promising alternative.

**5.3 False Positives.** Various sources of false positives can be distinguished. Here, we discuss only those that arise in direct connection to the nonstationarity extension presented in this article. For a more general discussion of the topic of false positives in unitary event analysis, we refer to the companion article.

First, there are sources for false positives specific to the moving window analysis. The significance level  $\alpha$ , which we demand for events to be qualified as unitary, implies by definition a certain number of false positives. For example, if  $\alpha$  is set to 0.01, we expect in 1% of the experiments a detection of significant events by chance. In the case of UEMWA, we undertake, in fact, many of such experiments by analyzing step by step successive parts of a single data set. However, these experiments are not independent due to the overlap of the time windows. For a rough estimate of the number of windows that is expected to give rise to false positives, one has to calculate the number of nonoverlapping windows that would fit within the total length of the data segment and take a fraction  $\alpha$  of it. In our calibration experiments on simulated data, accidental crossings of significance threshold were extremely rare—typically, one window per data set at most, and mostly none. As a rule of thumb, we have developed the criterion that only those cases in which the number of windows that passes the significance threshold is clearly larger than this lower bound are considered as potentially interesting. A more systematic treatment of this issue is under development.

Important sources of false positives are nonstationarities of different flavors that give rise to significant results, although the processes observed actually do not violate the null-hypothesis of independence. The obvious source is a remaining nonstationarity of rate in the analysis window. In appendix B, we show that unitary event analysis is robust against moderate

violations of the assumption of rate stationarity and quantify the effect on the number of false positives. Specific types of nonstationarity across trials discussed in the literature are variations of “excitability” and of “latency variability” (e.g. Brody, 1999a, 1999b). Knowledge (or “educated guesses”) of the type of nonstationarity in the data is important, because it may allow for a compensation of its effects in the analysis. Variation of excitability describes a nonstationarity in which the time course of the firing rate is identical in each trial, although the amplitude is modulated (e.g. Arieli, Sterkin, Grinvald, & Aertsen, 1996). Latency variability describes a nonstationarity in which shape and amplitude of the firing rate are identical in each trial, but the position in time shifts from trial to trial. In both cases, the firing-rate estimation obtained by averaging across trials, as performed in the PSTH and the unitary event analysis, leads to a value that is not representative for a single trial. In the case of excitability variations, the rate is underestimated for some trials and overestimated for others. In the case of latency variability, the rate estimate will generally present a blurred picture of the rate dynamics in an individual trial due to the convolution with the latency distribution. Thus, “misalignment” of trials may lead to falsely detected unitary events (see the example in Figure 8A, bottom panel). A solution to this problem is to realign the data to an external or behavioral event, to which the single trial rate functions of the observed neurons have a more proper locking. In case no such events are available, one can try to find a consistent realignment directly based on the single trial rate functions themselves (as shown in Nawrot, Rotter, Riehle, & Aertsen, 1999; Nawrot, Aertsen, & Rotter, 1999; see also Baker & Gerstein, 2000). Figure 8 illustrates how proper realignment of trials is able to discard false positives (unitary events after RS in A).

Note that in order to maintain a common reference time frame, the same realignment should be performed for all neurons under consideration. This implies, however, that when the latency variabilities of (some of) the observed neurons do not co-vary, such joint realignment is not possible (Nawrot, Rotter, et al., 1999). In such a case, one needs an alternative method to estimate the firing probabilities and the associated coincidence expectancy on the basis of single-trial data. Several such methods have been proposed recently, including convolution-based methods (Nawrot, Rotter, et al., 1999; Nawrot, Aertsen, & Rotter, 1999) and inverse-interval-based methods (Nawrot, Rotter, & Aertsen, 1997; Pauluis & Baker, 2000). The incorporation of these methods into the unitary event analysis is in progress.

To ultimately demonstrate the consistent occurrence of unitary events, one may apply an additional test on a meta-level. This can be done by relating the unitary events to behavioral or external events or by an additional statistical test (Prut et al., 1998). An example where unitary events occurred in relation to behaviorally relevant events is shown in Figure 6. On the basis of a meta-analysis over many data sets, Riehle et al. (1997) demonstrated that unitary events occur in relation to stimulus and expected events.

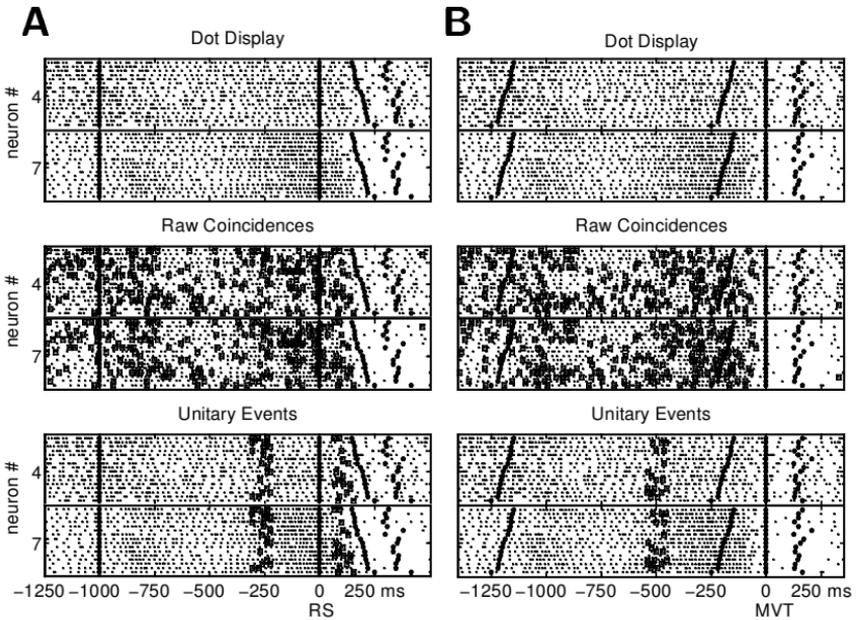


Figure 8: False positives due to nonstationarity across trials. (A, B) Identical spiking activities of two neurons (4,7) recorded from primary motor cortex while the monkey performed a delayed reaching task (1 direction out of 6; Grammont & Riehle, 1999). Columnar organization of the dot displays is the same as in Figure 6. In A, trials are aligned to the reaction signal (RS). In B, trials are aligned to the onset of movement (MVT). Trials are sorted according to increasing reaction time (time difference between RS and MVT) in both cases. All external and behavioral events are marked by filled circles. The events at the beginning of the data segments are the GO signals; the events late in the trials are the ends of movements. The unitary events in A about 250 ms before RS correspond to the ones in B at about 500 ms before MVT. Note, however, that the unitary events in A after the RS have vanished in B. These unitary events were due to nonstationarity across trials: an abrupt decrease in firing shortly before and locked to MVT dispersed to a different position in each trial by the incorrect alignment to RS. This misalignment of rate functions is removed in B, thereby discarding the falsely detected unitary events.

**5.4 Guidelines for Application to Experimental Data.** Based on the foregoing, we suggest the following procedure for unitary events analysis. First, inspect the data critically for across-trial nonstationarities (e.g., excitability and/or latency variability) by the use of the raster displays. For checking excitability variations, one may use as a rough estimate the number of spikes per trial. If necessary, eliminate the outlier trials. Latency variabil-

ity has to be eliminated by realignment of the trials, by either alignment of the data to another behavioral event or estimating the instantaneous rates per single trial and realigning the data by matching the shape of the rate functions (Nawrot, Aertsen, & Rotter, 1999).

Since generally we do not have prior knowledge about the time structure of the coincident events, we may get some qualitative information about their composition, their distribution in time, and whether there are obvious deviations from the firing rates by inspecting the raw coincident events.

The appropriate size of the analysis window cannot be known in advance. It has to be adjusted according to two aspects: rate changes and coincidence rate changes. In order to capture rate changes such that stationarity of rate within a window can be assumed, the window has to be adjusted to the timescale of the rate dynamics. Since we have no prior knowledge about the timescales of the coincidence rate dynamics, we have to scan the data with different window sizes. If we have an indication for a specific time structure from the raw data, we may start with a size in that range. In general, however, we found it best to start with a narrow analysis window and gradually increase its size. If the underlying time structure of coincident firing is clustered, the unitary events will appear at a certain minimum size of the analysis window and will be detected over a certain range of window widths beyond it. If the detected time structure is stable but only broadens due to the increased size of the analysis window, a hot region is detected. For an indication of the best analysis window size, the shape of the joint-surprise function may be used. Plateaus indicate a window size that is either too small or too large. A peaky shape indicates that the optimal window for a hot region size has been found.

False positives may be identified by evaluating whether the structure is stable for different alignments to external events. We are aware that this is time-consuming. In further work we intend to develop additional statistical tests on a meta-level, to filter out false positives automatically. However, to improve the analysis method further, we need to gain experience from applications of the unitary events analysis method to physiological data. In addition, we have found it useful to compare the outcome of our analyses to the result of other techniques (e.g. cross-correlation or joint peristimulus time histogram, JPSTH).

An additional parameter is the coincidence width. In order to determine the coincidence width of the experimental data, additional manipulations may be performed, for example, changing the bin width before analyzing the data for unitary events (see the companion article) or applying the multiple shift method (Grün et al., 1999; Riehle, Grammont, Diesmann, & Grün, 2000).

**5.5 Unitary Events in Cortical Activity.** We have analyzed simultaneously recorded multiple single-neuron spike trains from frontal and motor cortices in awake, behaving monkeys for the occurrence of significant co-

incident events. Our findings indicate that highly precise (1–5 ms) unitary events occur in these data. Their joint-surprise may be well above 2; that is, they are statistically highly significant. The composition and frequency of such patterns appear to be related to behavioral parameters. These results, together with results from other multi-neuron studies, are interpreted as expressions of cell assembly activity (for reviews, see Gerstein, Bedenbaugh, & Aertsen, 1989; Singer et al., 1997; Singer, 1999). The composition of unitary events is interpreted to reflect the common membership in a cell assembly. If the composition of the unitary events changes depending on the stimulus or the behavioral conditions, a different group and, hence, a different cell assembly may be activated in relation to the external event. In the example in Figure 7, we showed that the occurrence of unitary events was locked to the onset of the movement, but the composition of them was different for the different movement directions. Similar findings were made in visual and frontal areas in cross-correlation studies, when neurons were found to be correlated for one stimulus or behavioral condition but not for another (e.g., Vaadia et al., 1991; Aertsen et al., 1991; Vaadia et al., 1995; Freiwald, Kreiter, & Singer, 1995; Kreiter & Singer, 1996; Fries, Roelfsmema, Engel, König, & Singer, 1997; Castelo-Branco, Goebel, Neuenschwander, & Singer, 2000). Moreover, JPSTH results from frontal cortex show that the correlation between two neurons may dynamically change depending on the behavioral context, suggesting that the neurons rapidly change their associations into different functional groups (Aertsen et al., 1991; Vaadia et al., 1995).

We demonstrated that unitary events show a marked increase in temporal structure as compared to the spiking events of the participating neurons, including cases where the single neurons did not show any discernible response, as judged from the absence of systematic modulations of their firing rates. This may indicate that neuronal computation uses different kinds of timescales, usually referred to as rate coding and temporal coding (see also Abeles, 1982b; Neven & Aertsen, 1992; Koenig, Engel, & Singer, 1996; Riehle et al., 1997; Shadlen & Newsome, 1998). We have begun to investigate if and how these concepts are implemented in the cortical network. It would seem possible that both coding mechanisms—rate coding and precise time coding—are used in the brain, and that depending on the cortical area, either one might dominate (see Vaadia & Aertsen, 1992, for a detailed discussion on this issue). We hope that the unitary event method presented here may help to decipher the mechanisms of neuronal information processing in the brain.

## Appendix A: Notation

---

$h$	time resolution of data, $[h]$ = unit of time
$T_w$	size of the moving window, integer valued, odd
$M$	number of trials

$\lambda$	background firing rate, $[\lambda] = 1/\text{unit of time}$
$\lambda_c$	coincidence rate, $[\lambda_c] = 1/\text{unit of time}$
$n^{\text{pred}}$	expected coincidence count
$n^{\text{emp}}$	empirical coincidence count
$S$	joint surprise
$\alpha$	significance level
$n_\alpha$	number of coincidences needed for significance
$T_c$	duration of the “hot region”
$T_\alpha$	minimum analysis window width required by $n_\alpha$
$T_{\min}$	minimum analysis window width to reach significance
$T_{\max}$	maximum analysis window width to still reach significance
$T_p$	duration of the plateau of the joint-surprise function
$T_s$	duration of interval showing unitary events
$f_{\min}$	minimal overlap of analysis window $T_w$ and hot region for detection of unitary events in the sliding situation

All capital “T” variables specify time intervals in units of  $h$ ,  $[T] = 1$ .

## Appendix B: False Positives Induced by Nonstationarity

In order to estimate the error made by assuming stationarity of rates within the analysis window, we consider the worst-case scenario (with respect to mean count). It can be shown that the maximal error is given if the rate changes in stepwise fashion (in comparison to, say, a linear change of rate), if both neurons change their rates in parallel, and, if the analysis window is centered at the time of rate change. Thus, in the following we consider the situation where two neurons change their rate at the same time from the same rate level  $\lambda_1$  to  $\lambda_2$ . The analysis window is centered at the rate change; that is, the window  $T_w$  is divided into two regions of duration  $\frac{T_w}{2}$ , where the rates are stationary at level  $\lambda_1$  and  $\lambda_2$ , respectively. The equations used in the following correspond to equations 3.3 and equation 3.2, however here with  $\lambda_c = 0 \text{ s}^{-1}$ , and adjusted to the special case sketched here.

The mean rate within  $T_w$  is

$$\bar{\lambda} = \frac{(\lambda_1 \cdot \frac{T_w}{2} + \lambda_2 \cdot \frac{T_w}{2})}{T_w}. \quad (\text{B.1})$$

Using the averaged rate, the number of expected coincidences is (cf. equation 3.3)

$$\bar{n} = (\bar{\lambda} \cdot h)^2 \cdot T_w \cdot M. \quad (\text{B.2})$$

The exact number of expected coincidences is given by calculating the number of coincidences in each time segment separately and taking the sum of the two:

$$n_* = ((\lambda_1 \cdot h)^2 + (\lambda_2 \cdot h)^2) \cdot \frac{T_w}{2} \cdot M, \quad (\text{B.3})$$

expressing a special case of equation 3.2. The error in number of coincidences is

$$\tilde{n} - n_* = -T_w h^2 \frac{1}{4} (\lambda_1 - \lambda_2)^2. \quad (\text{B.4})$$

The larger the difference of the rate levels is, the larger is the error. The exact number of expected coincidences is larger than the approximated number  $\tilde{n}$ . Thus, if we estimate the number of coincidences by averaging the rates within the analysis window, we tend to overestimate the significance.

The relevant parameter for the significance test, the number of expected coincidences, defines the mean of the assumed Poisson distribution. This number, however, is determined by the rates, the size of the analysis window, and the number of trials. For simplicity, we assume a constant  $T_w$ . Different rate combinations of  $\lambda_1$  and  $\lambda_2$  can lead to the same number of expected coincidences. Thus, in order to study false positives, we derive the dependence of the exact number of coincidences  $n_*$  as a function of the approximated number of coincidences  $\tilde{n}$  by eliminating the rates. We define

$$r = \frac{\lambda_2}{\lambda_1} \quad (\text{B.5})$$

as the ratio of the rates.  $n_* = n_{*,1} + n_{*,2}$  is the total number of expected coincidences, with

$$n_{*,i} = \frac{1}{2} T_w M(\lambda_i)^2 h^2 \quad (\text{B.6})$$

the expected number of coincidences in each segment. Using equations B.5 and B.6, we yield

$$n_* = \frac{2(1+r^2)}{(1+r)^2} \cdot \tilde{n}. \quad (\text{B.7})$$

The slope of  $n_*$  ( $\tilde{n}$ ) is always  $\geq 1$ , that is, for all values of  $r$ ,  $n_*, \tilde{n} \geq 1$ .

In order to calculate the critical rate relation that leads to a significant outcome in expectation, we systematically vary  $n_*$  by varying  $r$  and compute the minimal number of coincidences needed for significance  $\tilde{n}_{\alpha r}$  given a significance level  $\alpha = 0.01$ . The intersection of  $n_*(r, \tilde{n})$  and  $\tilde{n}_{\alpha}(r, \tilde{n})$  gives the critical rate relation  $r$ . The larger  $r$  is, the steeper the slope of  $n_*(r, \tilde{n})$  and thus the smaller the minimal  $\tilde{n}$  for which false positives are obtained.

Interestingly, the mapping of  $(\lambda_1, \lambda_2)$  to  $(\tilde{n}, r)$  is invertible. The critical rate relation and the corresponding rates are shown as functions of  $\tilde{n}$  in Figure 9A.

Up to now, we have considered only the case where  $n_*$  is above  $\tilde{n}_{\alpha r}$ , corresponding to a significant outcome in expectation. Now we are interested in the effective significance level or, equivalently, the percentage of false

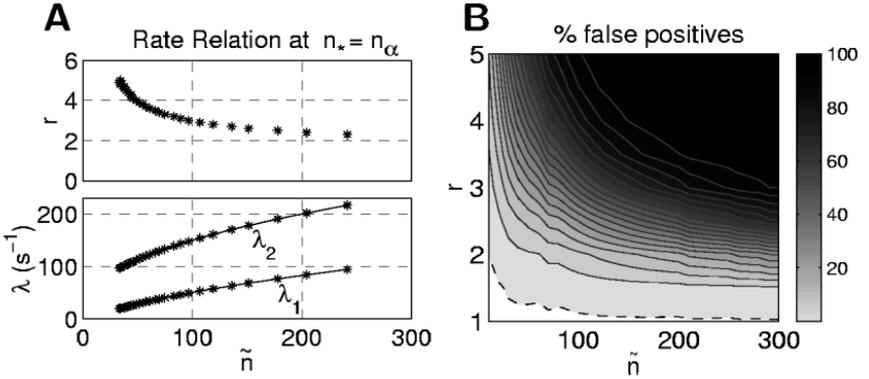
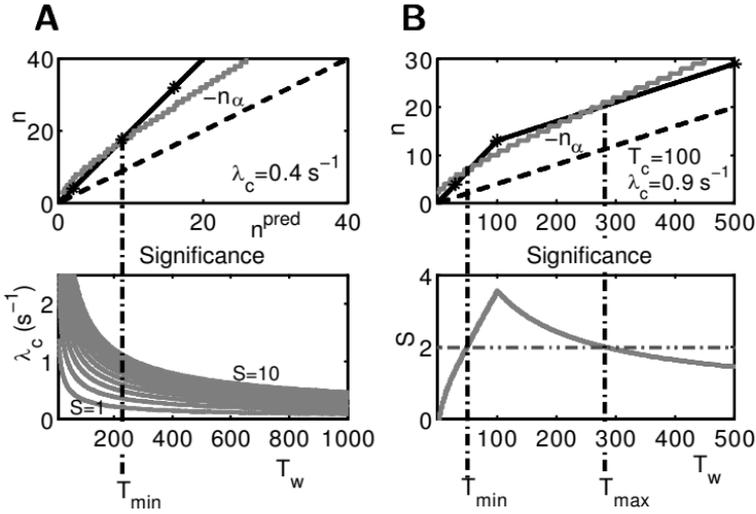


Figure 9: False positives induced by a stepwise rate change. As the worst-case scenario, two neurons are considered that change their rates in parallel stepwise from rate level  $\lambda_1$  to  $\lambda_2$ , and the analysis window is centered at the time of rate change. (A) Critical rate relation  $r = \lambda_2/\lambda_1$  (top) that leads to false positives at a given number of coincidences  $\tilde{n}$ . The corresponding rate levels are shown in the bottom panel. (B) Percentage of false positives. For each parameter constellation  $(\tilde{n}, r)$ , the contour plot shows the percentage of false positives for a significance level  $\alpha = 0.01$ : contour lines at 1% (dashed) and 5%, 10%, ..., 100% (solid). Gray scale indicates the percentage of false positives. In the stationary situation ( $r = 1$ ), the percentage of false positives equals  $\alpha$ .

positives at a given parameter constellation  $(\tilde{n}, r)$ . Therefore, we first calculate for each  $\tilde{n}$  the minimal number of coincidences  $\tilde{n}_\alpha$  at significance level  $\alpha = 0.01$  assuming a Poisson distribution with mean  $\tilde{n}$ . Then we determine the significance level for  $\tilde{n}_\alpha$ , assuming now the exact mean number of coincidences  $n_*$  and the corresponding distribution. Note that the distribution of coincidence counts is the convolution of the distributions for the two segments. Figure 9B illustrates that the larger  $\tilde{n}$  is, the lower the rate ratio  $r$  that can be tolerated.

## Appendix C: Size of the Analysis Window

**C.1 Stationary Coincidence Rate.** The need for a minimal window size in order to be able to detect excess coincidences will be derived for the case when both background rates and the injected coincidence rate are stationary. The minimal number of coincidences ( $n_\alpha$ ) that just fulfills the condition  $S(n_\alpha, n^{\text{pred}}) = S_\alpha$  depends nonlinearly on the number of occurrences expected at chance level (see Figure 4 in the companion article); for higher  $n^{\text{pred}}$ , disproportionately more coincidences  $n_c$  are needed to reach significance. Moreover, the minimum number of coincidences to reach threshold  $n_\alpha$  can take only discrete values. This induces discrete jumps in  $n_\alpha$  and in the



joint-surprise function (see Figure 10). In the stationary case, equations 3.1 and 3.3 reduce to

$$n^{\text{emp}} = [\lambda_c h + (\lambda h)^2] \cdot MT_w$$

$$n^{\text{pred}} = [(\lambda_c + \lambda)h]^2 \cdot MT_w.$$

$n^{\text{emp}}$  and  $n^{\text{pred}}$  are both linearly dependent on  $T_w$ . As a result, we can express  $n^{\text{emp}}$  as a linear function of  $n^{\text{pred}}$ :

$$n^{\text{emp}}(n^{\text{pred}}) = \frac{(\lambda h)^2 + \lambda_c h}{(\lambda h + \lambda_c h)^2} n^{\text{pred}} \quad (\text{C.1})$$

$$\approx \left( 1 + \frac{\lambda_c h}{(\lambda h)^2} \right) n^{\text{pred}}, \quad (\text{C.2})$$

where in the latter expression we have neglected second-order rate terms involving  $\lambda_c$ . This linear function has a single intersection with the significance threshold  $n_\alpha$  (assuming  $n_\alpha$  to be a smooth function), yielding the minimal number of coincidences considered to be significant.  $n_\alpha$  however, is a function of the width of the analysis window. Hence, for each combination of  $\lambda$  and  $\lambda_c$ , there is a minimal analysis window width  $T_\alpha$  needed to detect coincidences as significant events. In the situation described here, with coincidences injected over an arbitrarily long time interval, the minimal analysis window width can always be realized. The pair  $(\lambda, \lambda_c)$  determines the value of  $T_\alpha$  (here:  $T_\alpha \approx 220 \text{ ms}$ ), while the amount of data available determines whether an analysis window of this width can be realized. When  $T_\alpha$  can indeed be realized, we call  $T_{\text{min}} = T_\alpha$ ; in section C.2 we will see that this is not always the case in the nonstationary situation.

As it is clear from C.2, the function  $n^{\text{emp}}(n^{\text{pred}})$  is steeper if the rate of the injected coincident events is higher and hence crosses the curve of  $n_\alpha$  at a lower value of  $n^{\text{emp}}$  and, consequently, of  $T_\alpha$ . This behavior is summarized in the bottom graph of Figure 10A. Thus, in the stationary case, detection of excess coincident events can be ensured by increasing the analysis window. In addition, enlarging the window decreases possible effects of the stepwise behavior of the joint-surprise.

**C.2 Nonstationary Coincidence Rate.** We now derive the conditions for the appropriate window size in the case when coincidences occur clustered in time on top of a constant background rate. Consider first a situation where

---

Figure 10: *Facing page.* Minimal and maximal window size. (A) Minimal window size. Existence and parameter dependence of the minimal analysis window size  $T_\alpha$  needed to detect injected coincidences as significant. (Upper graph) Independent from window size, the number of coincidences minimally needed ( $n_\alpha$ ) to reach the significance criterion is a function of the significance level  $\alpha$  and the expected number of coincidences  $n^{\text{pred}}$  (gray curve, here for  $\alpha = 0.01$ ). With increasing  $n^{\text{pred}}$ ,  $n_\alpha$  increasingly deviates from the diagonal. The solid line is the number of coincidences  $n^{\text{emp}}$  for  $n^{\text{pred}}$  coincidences due to the rate of the independent processes ( $\lambda = 20 \text{ s}^{-1}$ ) and additional injected coincidences ( $\lambda_c = 0.4 \text{ s}^{-1}$ ). At a certain  $n^{\text{pred}}$  (here  $n_{\text{min}} \approx 9$ ),  $n^{\text{emp}}$  intersects  $n_\alpha$  from below. The minimal size of the analysis window is defined by the requirement that the expected number of coincidences is at least  $n_{\text{min}}$ . (Lower graph) Assuming stationary rates,  $n^{\text{pred}}$  is proportional to the size of the analysis window  $T_w$ . Thus, by proper scaling, the abscissa can as well be expressed in  $T_w$ . The dashed-dotted line connecting the two graphs indicates the minimal window size  $T_{\text{min}} = T_\alpha$  for the example pair  $(\lambda, \lambda_c)$ . The contour plot (contour lines shown for  $S = 1, 2, \dots, 10$ ) shows the dependence of the joint-surprise  $S$  on  $T_w$  and  $\lambda_c$  for the fixed example value of  $\lambda$ . For  $\lambda_c = 0.4 \text{ s}^{-1}$ , a joint surprise value of 2 corresponding to  $\alpha = 0.01$  is reached at a window size of about 220 ms. The sensitivity of the method decreases rapidly when windows narrower than 100 ms are used. For windows wider than 220 ms, sensitivity increases only slowly. (B) Maximal window size. Same analysis as in A for a nonstationary coincidence rate. Coincidences are injected at rate  $\lambda_c = 0.9 \text{ s}^{-1}$  in a “hot region” of duration  $T_c = 100$ , centered on  $t = 500$ ,  $M = 100$ ,  $h = 1$  ms (background rate  $\lambda = 20 \text{ s}^{-1}$ ). Detectability as a function of the width of the analysis window, centered in the hot region. (Upper graph) Dashed line is the diagonal, where as many coincidences occur as expected for independent rates; the gray curve describes the minimal number of coincidences  $n_\alpha$  needed to fulfill the significance criterion. For analysis windows below  $T_c$ , the number of coincidences increases faster than the minimal number and surpasses the minimal number at  $T_{\text{min}}$ . For  $T_w$  wider than  $T_c$ , the number of coincidences increases with unit slope, slower than  $n_\alpha$ . There is a second intersection at  $T_{\text{max}}$ , above which the number of coincidences is no longer significant. The lower graph illustrates the dependence of the joint-surprise on  $T_w$ .

the analysis window is placed in the middle of a hot region of width  $T_c$ , and the width of the analysis window  $T_w$  is gradually increased (see Figure 10B). As long as  $T_w \leq T_c$ , we face the situation of stationary injected events, discussed in the preceding section (cf. Figure 10A). Hence, we need a minimum width of the analysis window ( $T_\alpha$ ) to detect the cluster of injected events.  $T_\alpha$  depends on only the combination of  $\lambda$  and  $\lambda_c$ ; its value can be obtained from the calibration graph for the stationary situation (see Figure 10A, bottom). When the analysis window exceeds the hot region ( $T_w > T_c$ ), the total number of coincidences increases further. However, the slope is now reduced to  $(\lambda h)^2$ , since the contribution of the injected coincidences remains constant ( $MT_c \cdot \lambda_c h$ ).

If  $n^{\text{emp}}(T_w)$  increases faster than  $n_\alpha(T_w)$ ,  $n^{\text{emp}}$  can intersect  $n_\alpha$  from below, yielding the minimal window size  $T_\alpha$  (compare to section C.1). Since the width of the hot region defines the size of the analysis window, from where on  $n^{\text{emp}}(T_w)$  shows a reduced slope, a cluster can be detected only if  $T_c \geq T_\alpha$  with  $T_{\min}$  equal to  $T_\alpha$ . For  $T_c < T_\alpha$ ,  $n^{\text{emp}}(T_w)$  bends before reaching  $n_\alpha$ . As a result,  $T_{\min}$  does not exist. If  $n^{\text{emp}}(T_w)$  increases faster than  $n_\alpha(T_w)$ , there is always a  $T_\alpha$ , but in contrast to the stationary situation, the existence of  $T_{\min}$  depends on the width of the hot region  $T_c$ .

When  $T_c < T_\alpha$ , the cluster cannot be detected, even with arbitrarily large analysis windows. The reason is that at  $T_w = T_c$ , we have  $n^{\text{emp}}(T_w) < n_\alpha(T_w)$ , and for  $T_w > T_c$ , the slope  $\dot{n}^{\text{emp}}(T_w) < \dot{n}_\alpha(T_w)$ . As a result,  $n^{\text{emp}}(T_w)$  remains below  $n_\alpha(T_w)$ . This argument also implies that when the cluster can be detected ( $n^{\text{emp}}(T_w) \geq n_\alpha(T_w)$  for  $T_\alpha \leq T_w \leq T_c$ ), a second intersection of  $n^{\text{emp}}(T_w)$  and  $n_\alpha(T_w)$  must exist for some  $T_w > T_c$ . Hence, in that case, there is a  $T_{\max} \geq T_c$  at which  $n^{\text{emp}}(T_w)$  intersects  $n_\alpha(T_w)$  from above. Two conclusions can be drawn from this. First, a cluster of excess coincidences is not detectable if its duration  $T_c$  remains below the critical time span  $T_\alpha$ . Second, even if the cluster is detectable ( $T_c \geq T_\alpha$ ), it may still go undetected if the analysis window is either too small  $T_w < T_\alpha$  or too wide  $T_w > T_{\max}$ . The range of appropriate window sizes can be obtained from calibration graphs as in Figure 10B.

When the analysis window is shifted gradually across the hot region, the time course of the joint-surprise  $S$  appears symmetrical around the center of the hot region.  $S$  has a trapezoidal shape in case of  $T_w \neq T_c$  that degenerates to a triangle in the special case that  $T_w = T_c$ . The duration of the plateau as a function of the parameter pair  $T_c, T_w$  is summarized in Table 1. Note that in two of the three possible outcomes regarding the observables  $T_p$  and  $T_w$  ( $T_p = 0$  and  $T_p \geq T_w$ ), we have a unique expression for  $T_c$  (in the remaining case  $T_p < T_w$ , two possibilities exist for  $T_c$ ). This uniqueness can be exploited to determine the extent of the hot region from the size of the plateau:

$$T_c = \begin{cases} T_w & \text{for } T_p = 0 \\ T_w + T_p & \text{for } T_p \geq T_w. \end{cases} \quad (\text{C.3})$$

Because we are free in choosing the size of the analysis window  $T_w$ , the requirement of equation C.3 can always be met. However, even for the case where the relationship is not unique ( $T_p < T_w$ ), a small variation of  $T_w$  immediately allows differentiating between the two possible values of  $T_c$ :

$$T_c = \begin{cases} T_w & \text{for } T_p = 0 \\ T_w + T_p & \text{for } T_p \geq T_w \\ T_w + T_p & \text{for } T_p < T_w \text{ and } \dot{T}_p(T_w) > 0 \\ T_w - T_p & \text{for } T_p < T_w \text{ and } \dot{T}_p(T_w) < 0. \end{cases} \quad (\text{C.4})$$

From the shape of the significance curve and the relationships worked out in equations C.3 and C.4, we can estimate  $T_c$ . If the significance threshold is reached at all, typically more than one window is significant. Only in the special case of  $T_w = T_c = T_\alpha$  is the maximum of  $S$  exactly at threshold level and for a single window only.

For a given combination  $T_c, T_w$ , we can compute the minimal overlap  $f$  of the two windows needed to detect the injected coincidences as significant. The overlap  $f$  describes the part of  $T_c$  “seen” by  $T_w$ , expressed as a fraction of  $T_w$ :

$$T'_c = \begin{cases} f \cdot T_w & \text{for } f \cdot T_w < T_c \\ T_c, & \text{otherwise.} \end{cases} \quad (\text{C.5})$$

We can now use  $T_{\max}$  to compute the minimal  $f$ . To this end, we reverse the question that originally led to the definition of  $T_{\max}$ : given a fixed  $T_w$ , what is the minimal  $T'_c$  (and hence,  $f_{\min}$ ) needed to detect the injected coincidences as significant? Formally, this  $f_{\min}$  can be computed as follows:

$$T_{\max}(T'_{c,\min}) = T_w \quad (\text{C.6})$$

$$T_{\max}(f_{\min} \cdot T_w) = T_w \quad (\text{C.7})$$

$$f_{\min}(T_w) = \frac{1}{T_w} \cdot T_{\max}^{-1}(T_w). \quad (\text{C.8})$$

Having found the minimal overlap  $f_{\min}$ , we can construct the extent of the region  $T_s$  in which injected coincidences are marked as significant (the time span between the two intersections of  $S$  with the significance threshold):

$$T_s = (T_w - f_{\min} \cdot T_w) + T_c + (T_w - f_{\min} \cdot T_w) \quad (\text{C.9})$$

$$= 2 \cdot T_w + T_c - 2 \cdot f_{\min} \cdot T_w \quad (\text{C.10})$$

$$= T_c + 2 \cdot (1 - f_{\min}) \cdot T_w. \quad (\text{C.11})$$

The minimal  $T_s$  is obtained for  $T_w = T_\alpha$ : in that case,  $f_{\min} = 1$ . Hence, it follows that  $T_s = T_c$ .

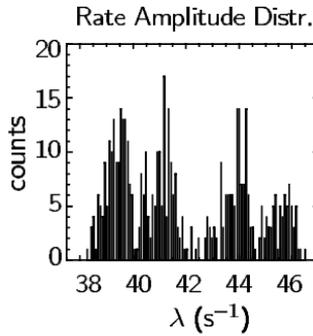


Figure 11: Multimodal rate amplitude histogram of a frontal cortex neuron, observed over 58 trials of 1500 ms duration. The rate was estimated by calculating a PSTH, smoothed with a window of 200 ms.

#### Appendix D: Detecting Unitary Events by Cluster Analysis

The approach in UECA is based on the assumption that the neurons' firing rates observed in parallel can only be in one of a finite number of joint rate "states" (Abeles et al., 1995; Seidemann et al., 1996). Using a clustering algorithm, segments of joint-stationary rates can be detected. In each of the resulting time segments, the unitary events analysis for the stationary case (see the companion article) is then performed separately. The amplitude distributions of the neurons' firing rates in many cases show indications of multimodality, suggesting that the firing rates can be in any of a finite number of states (e.g., Figure 11). Accordingly, we try to separate the combined rate activity of several simultaneously recorded neurons into joint stationary regions, defined as the time segments in which all the firing rates of the  $N$  observed neurons are stationary in parallel. This is achieved on the basis on estimates of the instantaneous firing rates (e.g., PSTHs) of each neuron. At each instant of discretized time, we derive a joint-rate vector, its components being the rate of neuron  $i$  at time  $t$ :

$$\vec{\lambda}(t) = \begin{bmatrix} \lambda_1(t) \\ \vdots \\ \lambda_i(t) \\ \vdots \\ \lambda_N(t) \end{bmatrix}, \quad i = 1, \dots, N. \quad (\text{D.1})$$

These vectors are grouped in  $N$ -dimensional  $\lambda$ -space into clusters of similar joint-rate vectors by the k-means clustering algorithm (Hartigan, 1975). It clusters according to  $R$  centers of gravity  $\vec{\mu}_k$ . Since we usually do not know the number of the underlying rate states of our data beforehand, we have

to vary the number of clusters and check the clustering result in each case. The stopping criteria for the “correct” number of clusters are given by the constraints that each potentially underlying joint state should be captured and that states should not be split into artificial ones (“overfitting”). For this purpose, we defined the following pairwise relative cluster distance:

$$d_{k,l} = \frac{|\vec{\mu}_l - \vec{\mu}_k|}{|\vec{\sigma}_l| + |\vec{\sigma}_k|}, \quad \forall k, l \in 1, \dots, R, k \neq l, \quad (\text{D.2})$$

$\vec{\sigma}_k$  representing the vector of standard deviation in the  $k$ th cluster. The stopping criteria are fulfilled if

$$d_{k,l} \geq 1, \quad \forall k, l \in 1, \dots, R, k \neq l. \quad (\text{D.3})$$

We call the final set of cluster joint-stationary rate states. Estimates of the firing rates are obtained by identifying the membership of each joint-rate vector  $\vec{\lambda}$  with the cluster mean it belongs to. Projecting the cluster mean back to the associated position on the time axis, we can observe the time course of cluster membership (PSTHs in Figure 12). The set of time instances identified as belonging to the same cluster defines the time segmentation of a state. Data from all segments belonging to one state are analyzed as one single stationary data set. Note that the set of time steps belonging to a single cluster is not necessarily compact and may in fact contain a number of separate time intervals, over which the neurons have approximately the same rates.

The performance of the clustering algorithm on a set of simulated, parallel Poisson processes is illustrated in Figure 12. Two extreme cases of firing-rate variations are chosen, covering the possible features of gradual (Figure 12A) and stepwise (Figure 12B) changes. To avoid overfitting, that is, fitting of states to variations that are due only to statistical fluctuations, we smoothed the PSTHs by using a moving average (Abeles, 1982a; Kendall, 1976). The choice of the width of the smoothing window has to be a compromise between large enough to reduce the noise level and small enough not to flatten out meaningful changes of the firing rate (see also Nawrot, Aertsen, et al., 1999). Observe that trajectories of the joint-rate vectors form clouds in rate space, well separated for rapid changes of the firing rates (see Figure 12A). Clusters can clearly be identified. By contrast, and not unexpectedly, gradual changes of firing rates (see Figure 12B) give rise to less well separated clouds or even smooth trajectories.

## Acknowledgments

---

We thank Moshe Abeles, George Gerstein, and Günther Palm for many stimulating discussions and help in the initial phase of the project. We especially thank Moshe Abeles, Hagai Bergmann, Eilon Vaadia, and Alexa

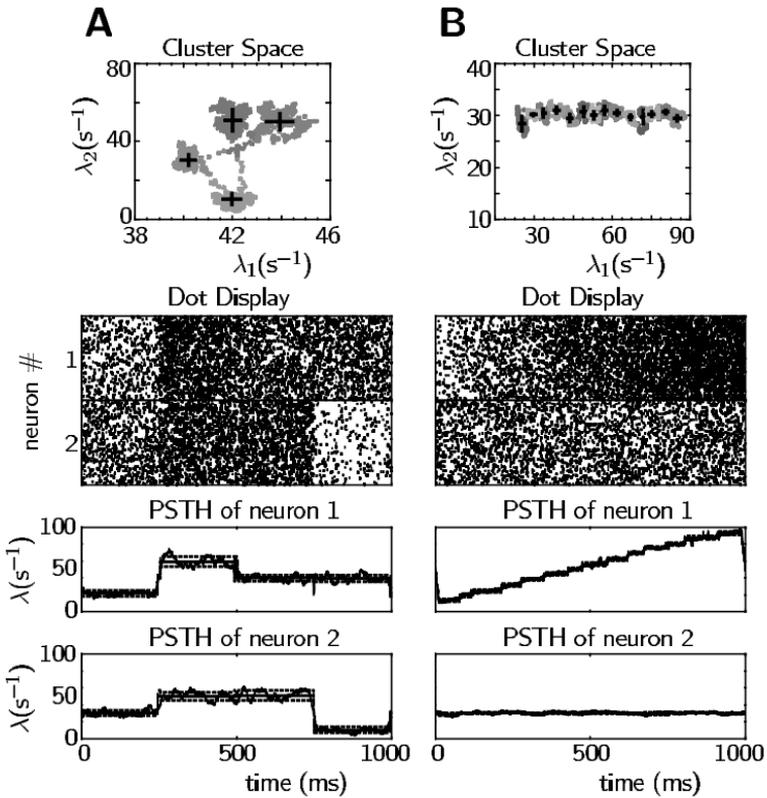


Figure 12: Unitary events by cluster analysis. Dot displays (middle) show the spiking activity of two parallel processes, the rates of which were varied in stepwise (A) or gradual (B) fashion. (A) 100 trials on the basis of time-varying rates in stepwise fashion:  $\lambda_1(i) = 30 \text{ s}^{-1}$ , for  $i \in [1, 250]$ ;  $\lambda_1(i) = 60 \text{ s}^{-1}$ , for  $i \in [250, 500]$ ;  $\lambda_1(i) = 40 \text{ s}^{-1}$ , for  $i \in [500, 1000]$ ;  $\lambda_2(i) = 30 \text{ s}^{-1}$ , for  $i \in [1, 250]$ ;  $\lambda_2(i) = 50 \text{ s}^{-1}$ , for  $i \in [250, 750]$ ;  $\lambda_2(i) = 10 \text{ s}^{-1}$ , for  $i \in [750, 1000]$ . (B) 100 (out of the 1000) trials in which neuron 2 was simulated with a constant rate of  $30 \text{ s}^{-1}$ , whereas the rate of neuron 1 is changing linearly throughout the trial from  $\lambda_1 = 20 \text{ s}^{-1}$  to  $90 \text{ s}^{-1}$ . The PSTHs (bottom panels) represent the time course of the instantaneous firing rates (smoothing window of 20 ms). (Top panels) Trajectories of the joint-rates in rate space. Each instance of a joint-rate vector is represented by a small gray star (abscissa: rate of neuron 1, ordinate: neuron 2). Members of different clusters are indicated by different gray levels. Cluster means are indicated by black crosses and standard deviations by the length of the cross-lines. Rates corresponding to the cluster means are superimposed on the PSTHs in the bottom panels. Cluster means approximate the original stationary rate levels; standard deviations (dashed lines) illustrate the variability of the rates within each cluster. Stepwise changes of cluster means indicate the timing segmentation used in further analysis. In B, clustering resulted in an average width of a time segment of 75 ms.

Riehle for kindly putting their experimental data (AR: motor cortex; MA, HB, EV: frontal cortex) at our disposal and for the many exciting discussions on our results. We also thank Robert Gütig, Stefan Rotter, and Wolf Singer for their constructive comments on an earlier version of the manuscript for this article. This work was partly supported by the DFG, BMBF, HFSP, GIF, and Minerva.

## References

---

- Abeles, M. (1982a). Quantification, smoothing, and confidence limits for single-units' histograms. *J. Neurosci. Meth.*, *5*, 317–325.
- Abeles, M. (1982b). Role of cortical neuron: Integrator or coincidence detector? *Israel J. Med. Sci.*, *18*, 83–92.
- Abeles, M., Bergman, H., Gat, I., Meilijson, I., Seidemann, E., Thishby, N., & Vaadia, E. (1995). Cortical activity flips among quasi stationary states. *Proc. Nat. Acad. Sci. USA*, *92*, 8616–8620.
- Abeles, M., & Goldstein, M. H. (1977). Multispikes train analysis. *Proc. IEEE*, *65*(5), 762–773.
- Aertsen, A., & Gerstein, G. L. (1985). Evaluation of neuronal connectivity: Sensitivity of cross-correlation. *Brain Research*, *340*, 341–354.
- Aertsen, A., Vaadia, E., Abeles, M., Ahissar, E., Bergman, H., Karmon, B., Lavner, Y., Margalit, E., Nelken, I., & Rotter, S. (1991). Neural interactions in the frontal cortex of a behaving monkey: Signs of dependence on stimulus context and behavioral state. *J. Hirnf.*, *32*(6), 735–743.
- Arieli, A., Sterkin, A., Grinvald, A., & Aertsen, A. (1996). Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science*, *273*(5283), 1868–1871.
- Baker, S. N., & Gerstein, G. L. (2000). Improvements to the sensitivity of gravitational clustering for multiple neuron recordings. *Neural Comp.*, *12*, 2597–2620.
- Brody, C. D. (1999a). Correlations without synchrony. *Neural Comp.*, *11*, 1537–1551.
- Brody, C. D. (1999b). Disambiguating different covariation types. *Neural Comp.*, *11*, 1527–1535.
- Castelo-Branco, M., Goebel, R., Neuenschwander, S., & Singer, W. (2000). Neural synchrony correlates with surface segregation rules. *Nature*, *8*(405), 685–689.
- Freiwald, W., Kreiter, A., & Singer, W. (1995). Stimulus dependent intercolumnar synchronization of single unit responses in cat area 17. *Neuroreport*, *6*, 2348–2352.
- Fries, P., Roelfsema, P., Engel, A., König, P., & Singer, W. (1997). Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proc. Nat. Acad. Sci. USA*, *94*, 12699–12704.
- Gat, I., Tishby, N., & Abeles, M. (1997). Hidden Markov modelling of simultaneously recorded cells in the associative cortex of behaving monkeys. *Network: Comp. Neural Sys.*, *8*, 297–322.
- Gerstein, G. L., Bedenbaugh, P., & Aertsen, A. (1989). Neuronal assemblies. *IEEE Trans. Biomed. Eng.*, *36*, 4–14.

- Grammont, F., & Riehle, A. (1999). Precise spike synchronization in monkey motor cortex involved in preparation for movement. *Exp. Brain Res.*, *128*, 118–122.
- Grün, S., Diesmann, M., Grammont, F., Riehle, A., & Aertsen, A. (1999). Detecting unitary events without discretization of time. *J. Neurosci. Meth.*, *94*, 67–79.
- Hartigan, J. A. (1975). *Clustering algorithms*. New York: Wiley.
- Kendall, M. (1976). *Time-series* (2nd ed.). London: Charles Griffin and Company.
- Koenig, P., Engel, A. K., & Singer, W. (1996). Integrator or coincidence detector? The role of the cortical neuron revisited. *TINS*, *19*(4), 130–137.
- Kreiter, A., & Singer, W. (1996). Stimulus-dependent synchronization of neuronal responses in the visual cortex of awake macaque monkey. *J. Neurosci.*, *16*(7), 2381–2396.
- Mountcastle, V. B., Reitböck, R. J., Poggio, G. F., & Steinmetz, M. A. (1991). Adaptation of the Reitboeck method of multiple electrode recording to the neocortex of the waking monkey. *J. Neurosci. Meth.*, *36*, 77–84.
- Nawrot, M., Aertsen, A., & Rotter, S. (1999). Single-trial estimation of neuronal firing rates. *J. Neurosci. Meth.*, *94*, 81–91.
- Nawrot, M., Rotter, S., & Aertsen, A. (1997). Firing rate estimation from single trial spike trains. In N. Elsner & H. Waessle (Eds.), *Göttingen Neurobiology Report 1997* (Vol. 2, p. 623). Stuttgart: Thieme Verlag.
- Nawrot, M., Rotter, S., Riehle, A., & Aertsen, A. (1999). Variability of neuronal activity in relation to behaviour. In N. Elsner & U. Eysel (Eds.), *Proceedings of the 1st Göttingen Neurobiology Conference of the German Neuroscience Society 1999* (Vol. 1, p. 101). Stuttgart: Thieme Verlag.
- Neven, H., & Aertsen, A. (1992). Rate coherence and event coherence in the visual cortex: A neuronal model of object recognition. *Biol. Cybern.*, *67*, 309–322.
- Pauluis, Q., & Baker, S. N. (2000). An accurate measure of the instantaneous discharge probability, with application to unitary joint-event analysis. *Neural Comp.*, *12*(3), 647–669.
- Prut, Y., Vaadia, E., Bergman, H., Haalman, I., Hamutal, S., & Abeles, M. (1998). Spatiotemporal structure of cortical activity: Properties and behavioral relevance. *J. Neurophysiol.*, *79*(6), 2857–2874.
- Reitböck, H. J. (1983). A multi-electrode matrix for studies of temporal signal correlations within neural assemblies. In E. Basar, H. Flohr, H. Haken, & A. J. Mandell (Eds.), *Synergetics of the brain* (pp. 174–182). Berlin: Springer-Verlag.
- Riehle, A., Grammont, F., Diesmann, M., & Grün, S. (2000). Dynamical changes and temporal precision of synchronized spiking activity in motor cortex during movement preparation. *J. Physiol. (Paris)*, *94*(5–6), 569–582.
- Riehle, A., Grün, S., Aertsen, A., & Requin, J. (1996). Signatures of dynamic cell assemblies in monkey motor cortex. In C. von der Malsburg, J. Vorbrüggen, & B. Sendhoff (Eds.), *Artificial Neural Networks—ICANN '96* (pp. 673–678). Berlin: Springer-Verlag.
- Riehle, A., Grün, S., Diesmann, M., & Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science*, *278*, 1950–1953.
- Riehle, A., Seal, J., Requin, J., Grün, S., & Aertsen, A. (1995). Multi-electrode recording of neuronal activity in the motor cortex: Evidence for changes in

- the functional coupling between neurons. In H. J. Hermann, D. E. Wolf, & E. Pöppel (Eds.), *Supercomputing in brain research: From tomography to neural networks* (pp. 281–288). Singapore: World Scientific.
- Seidemann, E., Meilijson, I., Abeles, M., Bergman, H., & Vaadia, E. (1996). Simultaneously recorded single units in the frontal cortex go through sequences of discrete and stable states in monkeys performing a delayed localization task. *J. Neurosci.*, *16*(2), 752–768.
- Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J. Neurosci.*, *18*(10), 3870–3896.
- Singer, W. (1999). Neural synchrony: A versatile code for the definition of relations. *Neuron*, *24*, 49–65.
- Singer, W., Engel, A. K., Kreiter, A. K., Munk, M. H. J., Neuenschwander, S., & Roelfsema, P. R. (1997). Neuronal assemblies: Necessity, signature and detectability. *Trends in Cognitive Sciences*, *1*(7), 252–261.
- Vaadia, E., & Aertsen, A. (1992). Coding and computation in the cortex: Single-neuron activity and cooperative phenomena. In A. Aertsen & V. Braitenberg (Eds.), *Information processing in the cortex* (pp. 81–121). Berlin: Springer-Verlag.
- Vaadia, E., Ahissar, E., Bergman, H., & Lavner, Y. (1991). Correlated activity of neurons: A neural code for higher brain functions? In J. Krüger (Ed.), *Neuronal cooperativity* (pp. 249–279). Berlin: Springer-Verlag.
- Vaadia, E., Bergman, H., & Abeles, M. (1989). Neuronal activities related to higher brain functions—theoretical and experimental implications. *IEEE Trans. Biomed. Eng.*, *36*(1), 25–35.
- Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., & Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature*, *373*, 515–518.