Multiple Tests based on a Gaussian Approximation of the Unitary Events method

Christine Tuleau-Malot, Amel Rouis, Patricia Reynaud-Bouret, Franck Grammont

Multiple Tests based on a Gaussian Approximation of the Unitary Events method

Christine Tuleau-Malot
malot@unice.fr
Laboratoire J.-A. Dieudonné, UMR7351, UNS/CNRS, Université de Nice Sophia Antipolis, Parc Valrose, 06108 Nice Cedex 2, France

Amel Rouis
rouis.a@chu-nice.fr
Centre Hospitalier Universitaire de Nice, France

Patricia Reynaud-Bouret
reynaudb@unice.fr
Laboratoire J.-A. Dieudonné, UMR7351, UNS/CNRS, Université de Nice Sophia Antipolis, Parc Valrose, 06108 Nice Cedex 2, France

Franck Grammont [1]
grammont@unice.fr
Laboratoire J.-A. Dieudonné, UMR7351, UNS/CNRS, Université de Nice Sophia Antipolis, Parc Valrose, 06108 Nice Cedex 2, France

Keywords: Unitary Events - Synchronization - Multiple testing - Independence tests - Poisson processes - Neuronal assemblies

The Unitary Events (UE) method is one of the most popular and efficient method used this last decade to detect conspicuous patterns of coincident joint spike activity among simultaneously recorded neurons. The detection of coincidences was first based on a binning procedure (Grün, 1996), which lead to some defects. These defects were then corrected with the Multiple Shifts (MS) procedure (Grün et al, 1999). Starting from this last step, we propose here some new improvements. We mainly show that the delayed coincidences count cannot be Poisson distributed if it is assumed that the spike trains are both Poisson and stationary processes. The gap between the real distribution and the Poisson one increases with the firing rate and the allowed delay for the coincidences. Moreover we precisely compute the asymptotic Gaussian distribution of the difference between the observed coinci-

[1] Corresponding author

dences count and an estimate of the expected count, showing that the replacement of the expected count by an estimate changes the law itself. This leads to statistical tests that are proved to be asymptotically of prescribed level α. In practice, the UE method is applied simultaneously over different sliding windows and all the tests with p-values less than 0.05 are declared detected. However it is well known in multiple testing theory that this method does not guarantee any control in terms of False Discovery Rate. We combine our new tests on sliding windows with a Benjamini and Hochberg approach (1995) leading to the Multiple Tests based on a Gaussian Approximation of the Unitary Events (MTGAUE) method. MTGAUE is tested on simulated spike trains and applied on real neuronal data. Finally, MTGAUE is not only mathematically more rigorous but it also proves to be reliable over a larger range of parameters than the original UE method.

1 Introduction

1.1 Theoretical aspects

The study of how natural neural networks transmit activity and somehow code information implies to consider various aspects of cerebral activity. Several techniques exist to record different aspects of this activity or some by-products of it like hemodynamical fluctuations for instance. However there is broad agreement on the fact that action potentials (spikes) constitute the main neural substrate of real-time information processing in the brain. What dimensions of spike trains are relevant for the brain or at least sufficiently informative for neurophysiologists and how can we study them?

Since the origins until today, firing rate has been considered as the main way that neurons or populations of neurons use to transmit activity, from the peripheral level of sensory receptors or medullary neurons to the cortical level. In most studies in this domain, the aim is to establish a correlation between a motor, a sensory or a cognitive event in a behaving animal and the occurrence of a variation of the firing rate of specific neurons, before, during or after this event. This is how it can be shown that firing rate modulates as a linear function of the pressure on the skin in sensory receptors at a very basic level, or increases in individual cortical mirror neurons in relation to high level cognitive processes (Rizzolatti & Craighero, 2004). Such kinds of correlations can be shown by the use of Peri-Stimulus-Time-Histograms (PSTH) (Shinomoto, 2010).

One of the first and most famous approach to decipher the role of firing rate at the cognitive and motor level is due to Georgopoulos and collaborators who showed how a vectorial transformation of the firing rate of a sample of primary motor cortex neurons could lead to an understanding of how the direction of movements is coded in the primate primary motor cortex (Georgopoulos et al., 1986; Pouget et al., 2000, for a review). Beyond the estimation of firing rate, different features of spike trains have also been studied more precisely. The fact that the temporal structure of the spike trains of individual cortical neurons may also be important to transmit information has been explored in different ways (Victor & Purpura, 1996, 1997; Schreiber et al., 2003; Kreuz et al., 2007; Chicharro et al., 2011).

However, these various approaches and techniques focus on the spike trains of in-

dividual neurons. Actually, the fact that the activity of ensembles of neurons may be coordinated in the spatio-temporal domain (i.e. coordination of the production of spikes between different neurons) to form neuronal assemblies in a way or another (Hebb, 1949; Palm, 1990; Sakurai, 1999) has been a topic of numerous and hot debates until today (Masuda & Aihara, 2007). Theoretically, various historical conceptions used to consider that neurons are mostly independent computational entities (Barlow, 1972). In addition and more technically, the single unit activity of neurons has for a long time been recorded only one by one in living animals. There was thus no possibility anyway to study neuronal interactions because of technological constraints. As a matter of facts, even after the democratization of multi- electrode single unit recording in the eighties, these techniques were most often used to record more individual neurons at a time and to study their variations of firing rate without paying attention, theoretically or technically, to the cooperativity between neurons.

Since then however, numerous theoretical, experimental and simulation studies have accumulated evidences showing that neurons do interact by coordinating in time the production of their spikes. Indeed, it has been shown that despite the variability of the cortical neurons discharge (Softky & Koch, 1993), that generally justifies the averaging over a large number of trials, the mechanisms of spike generation can be very precise (Mainen & Sejnowski, 1995) under physiological conditions (Konishi et al., 1988; Lestienne, 2001; Prescott et al., 2008). Like this, different kinds of precise spike patterns have been detected in spike trains beyond what could be expected by chance (Lestienne, 1996; Abeles & Gat, 2001; Gerstein, 2004; Eyherabide et al., 2009).

Historically, spike synchronization, with or without oscillations, has first been shown to be involved in the so-called binding problem (Von der Malsburg, 1981; Singer & Gray, 1995; Engel & Singer, 2001). From there and progressively, various characteristics of spike synchronization and its relation to firing rate have been discovered (Singer, 1999). The major criticism that has been formulated against the possibility that neurons do interact by coordinating the production of their spikes to constitute neuronal assemblies is that spike synchronization might just be an epiphenomenon of the variations of the firing rate. That is why most studies in this domain tried to show the relative independence of synchronization over the firing rate (Riehle et al., 1997; Maldonado et al., 2000; Heinzle et al., 2007). Moreover, it has been shown that neurons, as expected from the theory (Hebb, 1949), are able to synchronize their spikes during short laps of time with certain neurons and to change rapidly their partnership to synchronize their activity with other neurons, supposed to belong to other assemblies (Grammont & Riehle, 1999). Beside experimental studies, how synchronization emerges and propagates in neural networks, with or without oscillations, has been studied extensively in simulation studies (Diesmann et al., 1999; Golomb & Hansel, 2000; Tiesinga & Sejnowski, 2004; Goedeke & Diesmann, 2008).

In the end, given the accumulation of evidences obtained these last 30 years, denying the existence and the role of neuronal cooperativity and its implication in information processing is now the matter of an obsolete scholastic debate. There is no reason anymore not to consider that natural neural networks are able to use different modes to transmit and to code information, notably by the variations of firing rate and/or the precise coordination of the production of spikes between neurons (Grammont & Riehle, 2003; Heinzle et al., 2007; Masuda & Aihara, 2007; Kumar et al., 2010). The question

may rather be now: how to characterize it in the best way ?

Indeed, it remains true that despite the accumulation of evidences about neuronal synchronization, the genuine neural mechanisms responsible for it are still broadly unknown (Heinzle et al., 2007). Indeed, we never know at the same time what are the activity of the neurons and the precise graph of their anatomical connections. Neither do we know the synaptic weights of these connections, their evolution after each spike and many other relevant parameters necessary to have a full understanding of the causal role of neuronal cooperativity on the underlying network (Vlachos et al., 2012). Actually, we might have to wait for decades before we acquire the technical capacity to characterize the nervous system in such a comprehensive way. That is why in the meantime, we must get the best out of the kind of data we are able to capture at the moment; that is in our case, spike trains from multiple neurons recorded at the same time during the behavior of an animal.

The approach we want to promote in this article is to combine the best statistical methods available to detect the expression of neuronal interactions with the constraints provided by the structure of Cognitive Neurosciences protocols. Both rigorous statistical methods and precise Cognitive Neurosciences protocols are necessary to make sense of neuronal data. One without the other cannot work. Such an approach has to respect 3 steps: 1. the method of analysis must be as mathematically rigorous as possible, 2. the method of analysis must be applied on compatible data, 3. the behavioral protocol must isolate the cognitive process studied and force its occurrence at precise moments. Indeed, a method can be perfectly rigorous mathematically and nonetheless be applied on non-compatible data or detect events that are statistically significant but biologically irrelevant, either because it detects epiphenomena or artefacts, or else. In Cognitive Neurosciences, behavioral protocols must be structured in order to rigorously isolate and identify the variables involved in the studied process so as to force their occurrence at certain moments of the protocol. If the method of analysis detects some positive events in a random manner along the protocol rather than at the specific expected moments, it may mean that it is not adapted to the neuronal and behavioral data studied, how rigorous it might be mathematically. Like this, the protocol constitutes a last constraint. At the end and in any case, the kind of event detected has to make sense biologically. Of course, we may remain unaware of some internal cognitive processes that could hardly be controlled in time and that would provide real positive events detected by the method of analysis. This gives the occasion to discover unexpected cognitive processes (Grammont & Riehle, 2003; Riehle et al., 2006).

1.2 Technical aspects

Historically, several methods have been developed to analyze single unit activity in terms of spike synchronization. There is a parentage between these different methods as well as between the researchers who developed them. Most often, these are the failures or limits of one method which have driven the development of a new method. In this article, we follow the same logic and try to bring our own contribution to this scientific inquiry. In the following we are thus going to quickly review several steps of the development of these analysis methods in order to identify the crucial questions at

stake, the failures and the improvements which have been brought up to now.

The cross-correlogram (Perkel et al., 1967) was one of the first methods used. It allows detecting systematically the delay between each spike of two distinct neurons (delay from 0 milliseconds up to a given limit). All these delays are then accumulated along each trial and averaged over the whole length of the trials in order to check if there is a specific temporal relation between the activities of the two neurons or if the various delays follow a random probability distribution. Depending on the result obtained, two kinds of interpretations can be formulated: one in terms of coding strategy (e.g. spike synchronization as a sign of neuronal cooperativity), the other in terms of neuronal connectivity (Glaser & Ruchkin, 1976). Indeed, if the cross-correlogram shows that a neuron fires regularly after another one has fired, it may be interpreted as the sign that these two neurons are linked by a direct excitatory connection. However, such a pattern could also be explained by the fact that these two neurons are both stimulated by a common source, directly for the first one and through one or several synapses for the second one. If the cross-correlogram is helpful to study neuronal synchronization and functional connectivity, we see that it cannot give any definitive answer about the real anatomical structure of the underlying network, as we already mentioned (Vlachos et al., 2012). Interpreting the precise anatomical connectivity of the underlying network remains thus out of the scope of this kind of methods, as well as for the method we propose in this article.

Another point concerning the cross-correlogram, which became crucial for the following methods of analysis, is the fact that the experimental protocols generally tend to provoke some increases in the firing rate of neurons at certain moments of the trial. This is the aim of such protocols in Cognitive Neurosciences. This covariation of the firing rate of the neurons generates mechanically more coincidences between the spikes of the two studied neurons, which are not necessarily meaningful in terms of coding strategy (Riehle et al., 1997). In order to counterbalance such artefacts, people have used the so-called shift-predictor (Gerstein & Perkel, 1969, 1972). Schematically, it compares the moments of occurrence of the spikes from a given trial of a first neuron with the moments of occurrence of the spikes of another trial of a second neuron. Today, we would call this a shuffling algorithm (Pipa & Grün, 2003; Pipa et al., 2003). Doing so, the supposedly real and functional coincidences are destroyed and the ones produced by the firing covariation are kept. It remains to subtract this last result to the raw cross-correlogram first obtained. This method limits the influence of the firing rate on the detection of meaningful coincidences and relevant delays by roughly discarding the coincidences produced just by chance. We will see in the following that this point became the cornerstone of this family of methods.

The other main weaknesses of the basic cross-correlogram and other similar methods are that it provides an averaged measure over the whole length of the trial and supposes the activity to be stationary all along. First, the activity of real neurons is not stationary over long periods of time, especially not during an experimental protocol designed to provoke some fluctuations of the activity at some specific moments (e.g. presentation of a sensory stimulus). More complex methods like the gravitational clustering (Gerstein et al, 1985; Gerstein & Aertsen, 1985) have proposed some solutions to deal with this problem. Second, we know from the theory (Hebb, 1949) and experimental results (Grammont & Riehle, 1999) that interactions between neurons may occur

during very short laps of time. Such short periods of coincident firing would be diluted in the averaging of the cross-correlation over the length of the trial. This problem of averaging has been tackled by Aertsen and collaborators with the Joint- PSTH by dealing with short time scale, stimulus-locked variations of near-coincident firing (Aertsen et al., 1989; Palm et al., 1988). Schematically, the Joint-PSTH detects the coincidences (with or without delay) between the spikes of two neurons recorded simultaneously as the cross-correlogram does. However, it provides a matrix representation in which all the coincidences are accumulated along the length of the trial. Like this, it becomes possible to detect some particular moments of correlation between two neurons during short periods of time, as a function of the stimuli occurring over the time of the trial.

However, as for the cross-correlogram there are several problems. More spikes are required for the Joint-PSTH than for the cross-correlogram as it detects coincidences along the duration of the trial but without averaging. Moreover, the Joint-PSTH analysis is maybe not structured at best to apply statistical methods in order to test the statistical significance of the number of coincidences detected. Its structure is too complex to allow the application of a statistical test in a simple and reliable manner. This is precisely one of the key points that is going to be addressed by the following analysis methods. Following these previous approaches and their inherent problems, Grün and collaborators developed the Unitary Events analysis method (Grün, 1996). This method has been continuously improved until today (Grün et al, 1999; Grün et al., 2001a,b; Grün, 2009; Grün et al, 2010). It is a popular method which has been used successfully in several experimental studies (Riehle et al., 1997, 2000; Grammont & Riehle, 1999, 2003; Maldonado et al., 2008). Our own work introduced in this article, is directly inspired from this last method. The Unitary Events (UE) analysis method came with several objectives. It is designed to detect precise spike coincidences between two or even more neurons. It does so along the time of the experimental protocol (i.e. the trial) and thus, as a function of the stimuli presented. Above all, it allows applying some tests to check the statistical significance of the difference between the number of coincidences detected and the number of coincidences expected on the basis of the firing rate of both neurons. This is a different approach than the ones discussed above which simply subtract an estimate of the number of coincidences occurring by chance from the total number of coincidences detected (Gerstein & Perkel, 1969, 1972). Additionally, the analysis is performed into small sliding windows of time in order to deal, at least partially, with the problem of non-stationarity of neuronal data. To describe the basis of this analysis method in a few more words, neuronal activity is represented under the form of spike trains with a time step usually equal to 1 ms (millisecond), given the duration of an action potential. In a first version (Grün, 1996), a binning procedure was applied on spike trains in order to detect the coincidences between the spikes of the studied neurons. The bin is defined as a function of the maximal delay accepted between two spikes of the two neurons to be considered as a coincidence. If one is interested by a delay of 5 ms, the whole spike train is discretized with a binning grid of 5 ms bins. It becomes easy in algorithmic terms to detect the coincidences as the presence of at least one spike in the same bin in both neurons. However, this technique suffers from several defects as the size of the bins increases. The larger the bins are, the more data are lost, as any additional spike present in a bin is clipped in order that its value never exceeds one. In the same vein, many coincidences between spikes may not be detected even if the

delay between the spikes is inferior to the size of the bin. Indeed, it is sufficient that the spike of the first neuron falls in a bin and that the spike of the second neuron falls one millisecond after, but in the next bin. These problems and others have been addressed by using the multiple shifts method (Grün et al, 1999). Here, there is no binning anymore and data are kept in their original time resolution (typically 1 ms per time step), (see Definition 2 in Section 2.1 for a precise formula in this case). Beyond the question of the accurate detection of coincidences with more or less delay, one has to determine whether these coincidences are statistically significant or not. If so, they will be considered as good candidates for being the sign of a real functional interaction between the two neurons concerned. For the first time, UE analysis proposed a framework in which it became much easier to apply statistical tests. Basically, the method compares the number of occurrences of the observed coincidences to the expected number of patterns, calculated on the basis of the joint probability of occurrence of the particular 0-1 configuration (presence or absence of a spike) assuming statistical independence. The difference is considered to be statistically significant if the number of observed coincidences deviates significantly from the center of the coincidence distribution resulting from independent processes (i.e. exceeds the joint-p-value; Grün, 2009). Classically in this domain, people use a Poisson distribution to describe analytically the distribution of spikes of real neurons, but also assume that the number of coincidences can be described by a Poisson distribution as well. At this point however several questions are raised. The statistical significance obviously depends on the nature of the chosen distribution. A given empirical value will be significant with a given reference distribution but not necessarily with another one. Choosing a Poisson distribution or another one is a matter of debates among researchers (Oram et al., 1999; Amarasingham et al., 2006; Maimon & Assad, 2009). We address this question in this article and propose a new approach. We still use a Poisson representation of the spike trains however we prove that this is incompatible with a Poisson distribution of the number of coincidences (see Section 2.1). We approximate very precisely the distribution of the coincidences count by a Gaussian law and we explain how the estimation of the underlying parameters modify the law itself in order to provide meaningful tests and p-values (see Section 2.2). This new precise approximation improve the range of parameters (firing rate, delay, windows length, number of trials) for which the previous analysis were valid. UE analysis is classically applied independently into short windows of time in order to deal with the non-stationarity of neuronal activity along the trial. Because some important changes in the firing rate can occur even into a given window, these windows are overlapped in order to smooth such effects, hence the term of sliding window. Once all the tests in all the sliding windows have been performed, their results are considered as a set. If all the tests are recklessly performed at level α, the set of answers is doomed to contain a lot of false detections. But as a set, there are different notions of errors (developed in Section 3.1) and depending on them, there are several ways to control the amount of false detections of dependence.

This is why in this article we propose to apply the multiple testing procedure due to (Benjamini & Hochberg, 1995) which controls the false discovery rate (FDR; see Section 3). To better understand the role of this procedure we need to introduce a definition of false positives based on testing theory. In the field of Neurosciences, false positive coincidences are often defined as the coincidences detected as significant because of

an artefact. For instance, if the expected number of coincidences is underestimated because of a problem of non-stationarity of the firing rate in time or across trials, it will be easier for the number of observed coincidences to reach the threshold of significance (Grün, 2009). However, in testing theory there is no need of a cause to explain false positive. Indeed, when a statistical test is performed with level $\alpha = 0.05$, it means that in 5 % of the cases and even in an ideal world, the test will wrongly reject and detect a false positive. Therefore if several tests are applied together (say 1000) in average 50 of them will reject the independence anyway. This is precisely what happens in the UE method. Thus, to avoid these false detections, or at least control their rate (and not systematically increase them when the number of tests increases), we couple the multiple testing procedure of Benjamini & Hochberg to our Gaussian Approximation of the UE method (see Section 3).

At the end, with this article we want to propose some new improvements, both at the level of mathematical rigor and in terms of the sensitivity of the method. Section 2 is devoted to the introduction of the Gaussian Approximation of the Unitary Events (UE) method on one single window. Then we detail the multiple testing procedure in Section 3. The new resulting method is applied first on simulated data (see Section 4) and then on single unit activity recorded in a behavioral context (see Section 5). These experimental data were already partially published (Grammont & Riehle, 2003; Riehle et al., 2006). We thus have elements of comparison to assess the relevance of our new method.

2 Single test of independence

As said previously, the aim of this work is to detect the local dependence between neurons. Before describing the multiple testing procedure, let us first describe what happens for a single test, on one window $[a, b]$. Those single tests will next be performed on several small windows of time (classically around 100 ms) and Section 3 will explain how we combine them with Benjamini and Hochberg's procedure.

Let N_1 and N_2 be the spike trains of two neurons, the single test aims at differentiating between:

$$H_0 \text{ (the null hypothesis): "}N_1 \text{ is independent of } N_2 \text{ on } [a, b]\text{"}$$
$$\text{versus}$$
$$H_1 \text{ (the alternative): "}N_1 \text{ and } N_2 \text{ are dependent on } [a, b]\text{"}$$

To perform this test, some models are required for N_1 and N_2. Here spike trains are considered as a random collection of points, that is a point process (Daley & Vere-Jones, 2003). From this we assume that:

Assumption 1. *The processes N_1 and N_2 are both Poisson processes.*

Assumption 1 can be actually reduced to an independence property between disjoint intervals: what happens in $[a, b]$ is independent of what happens in $[c, d]$ if $[a, b] \cap [c, d] = \emptyset$. Indeed if this property is true for any set $[a, b]$ and $[c, d]$ and if there is no accumulation of points[2], it can be proved that under H_0 both processes are Poisson processes

[2] This assumption is a direct consequence of the classical behavior of neurons.

(Daley & Vere-Jones, 2003). Note consequently that Assumption 1 is quite classical in Neurosciences even if it is not always precisely identified as a Poisson assumption (Roy et al., 2000; Grün et al., 2001a,b; Grün et al, 2010).

Assumption 2. *The processes N_1 and N_2 are stationary on the time period $[a, b]$.*

In practice, we just need the intensities (i.e. the firing rate) of both processes to be well approximated by constant functions.

Of course, only one realization of this couple of processes is not sufficient to decide whether the spike trains are dependent or not. During the experiment, M trials are recorded, which leads to the observation of $(N_1^{(1)}, N_2^{(1)}), \ldots, (N_1^{(M)}, N_2^{(M)})$, M realizations of the couple (N_1, N_2). As in the rest of the literature, we assume here that these realizations are independent and identically distributed.

In the sequel we denote by $\mathbb{E}(T)$, for some quantity T depending on (N_1, N_2), the expectation of T with respect to the law of (N_1, N_2). Heuristically this refers to the empirical mean over trials, when the number M of trials is infinite. In particular this is a deterministic quantity. This is only approximated by the empirical mean when the number of trials is finite. The empirical mean [3] for finite M is a random quantity which fluctuates around the corresponding expectation $\mathbb{E}(T)$, and can only be considered as an estimate of $\mathbb{E}(T)$, the fluctuation becoming smaller when M grows. To avoid confusion, all the estimates are denoted by $\hat{\square}$ or $\bar{\square}$, where $\square$ represents any letter, and those estimates have to be considered as random whereas the corresponding deterministic quantities have nothing on the top.

The probability associated to $\mathbb{E}$ is denoted $\mathbb{P}$: if we denote for any set A, by $1_A(T)$ the indicator function with value 1 when T is in A and value 0 otherwise, $\mathbb{E}(1_A(T)) = \mathbb{P}(T \in A)$. As an example of the previous notation, $p = \mathbb{P}(T \in A)$ is deterministic and only approximated by the empirical frequency $\hat{p}$ (i.e. the number of trials for which $T^{(i)}$, the realization of T for the trial i, lies in A divided by M), which is a random quantity.

2.1 Two definitions of coincidences

The statistic on which the single test is based is the coincidences count. However this notion actually covers two definitions: one within the binning framework (Grün, 1996), the other one within the Multiple Shifts method framework (Grün et al, 1999).

The binning framework In the binning approach (Grün, 1996; Grün et al, 2010, see section 10.2.1), the data are represented not as accurately as in the original spike trains. A part of the information is discarded (see Section 1.2). More precisely, the window $[a, b]$ is divided in small disjoint bins of length δ, denoted $I_1, \ldots, I_k$, such that they form a partition $[a, b]$ (and therefore $\delta = (b - a)/k$).

[3] The empirical mean of T is $\frac{1}{M}\sum_{i=1}^{M} T^{(i)}$ where $T^{(1)}, ..., T^{(M)}$ are M realizations of T.

Definition 1. *The binned coincidences count Y on $[a, b]$, is defined by:*

$$Y = \sum_{i=1}^{k} \mathbf{1}_{\{N_1(I_i)\geq 1,\ N_2(I_i)\geq 1\}}$$

where $N_j(I_i)$ denotes the number of spikes of N_j on I_i for $j = 1, 2$.

If we consider this definition, a coincidence on a bin is just the fact that the two neurons have both at least one spike in the same bin. As mentioned in the introduction, the loss of information comes from the fact that no difference is made between two spikes as long as they belong to the same bin. Their precise position is lost. Data are also clipped by deleting any additional spike in the bin. The next result shows how these transformations influence the shape of the expected coincidences count.

Proposition 1. *Under the null hypothesis H_0 and Assumptions 1 and 2,*

$$m_0 := \mathbb{E}(Y) = k\left(1 - \exp(-\lambda_1\delta)\right)\left(1 - \exp(-\lambda_2\delta)\right)$$

where λ_j is the constant intensity (firing rate) of N_j for $j = 1, 2$.

The original UE method is based on this binned coincidences count (Grün, 1996). However the approximation of the behavior of the observed coincidences count under H_0 done by the UE method may potentially fail at several levels.

- The first approximation is done on the expected count and consists in considering that m_0 is close to $m_{UE} = k\lambda_1\lambda_2\delta^2$. This is valid as long as the products $\delta\lambda_i$ are not too large (typically less than 1), which happens for instance when $\lambda_i \leq 50$ Hz and $\delta \leq 0.02$ s. So this first source of error may be quite negligible in our framework.

- The second approximation consists in assessing a law to the observed coincidences count $M\bar{m}$, where $\bar{m}$ the average coincidences count on $[a, b]$ is defined by

$$\bar{m} = \frac{1}{M}\sum_{i=1}^{M} Y^{(i)}, \tag{1}$$

  where $Y^{(i)}$ is the binned coincidences count for the ith trial. The UE method postulates that $M\bar{m}$ obeys a Poisson law with parameter $\tau := Mm_{UE} = Mk\lambda_1\lambda_2\delta^2$ under H_0 (Grün, 1996). However purely mathematically speaking, the random variable Y is by definition a binomial variable and therefore $M\bar{m}$ is also a binomial variable with parameters $n = Mk$ and $p = \left(1 - \exp(-\lambda_1\delta)\right)\left(1 - \exp(-\lambda_2\delta)\right)$ (see the Appendix Section A.2). Once again it is true that binomial variables can be approximated by Poisson variables when $n \geq 100$ and $p \leq 0.1$, (Hogg & Tanis, 2009, see p 159) which is usually the case here.

- The last source of potential error comes from m_{UE} itself which is unknown and can only be estimated. Since both λ_j's are unknown and can only be estimated by

$$\hat{\lambda}_j := \frac{1}{M(b-a)}\sum_{i=1}^{M} N_j^{(i)}([a, b]) \tag{2}$$

where $N_j^{(i)}([a,b])$ is the number of spikes in $[a,b]$ for neuron j during the ith trial, this leads to an estimate of the expected coincidences count

$$\hat{m}_{UE} = k\hat{\lambda}_1\hat{\lambda}_2\delta^2.$$

In (Grün, 1996), m_{UE} and $\hat{m}_{UE}$ are considered to be the same quantity and in particular $\hat{m}_{UE}$ is not considered as a random quantity. Therefore the UE rejection is obtained by comparing $M\bar{m}$ to the quantile of a Poisson distribution where the unknown parameter τ is replaced by $M\hat{m}_{UE}$.

This step is known in statistics as a plug-in step and is known to modify sometimes dramatically the law. One of the most famous example is the Gaussian law which has to be replaced by a Student law when the variance is unknown and estimated by an empirical mean over less than say 60 realizations [4].

One of the aims of this article is to sufficiently formalize the framework to avoid this kind of misleading plug-in steps that can potentially make all the practical studies worthless. The reader will find a proper Gaussian approximation of the binned coincidences count which fixes the plug-in problem in the appendix. However since the binning framework suffers from several drawbacks - loss of information, sensibility to the localization of the bins - as mentioned in the introduction, we focus the rest of the article on another notion of coincidences introduced in (Grün et al, 1999), where the plug-in step is not the only source of potential errors.

The delayed coincidences count This notion of coincidences refers to the kind of coincidences which can be detected by the Multiple Shifts method (Grün et al, 1999; see formula (10.6); Grammont & Riehle, 2003).

Definition 2. *The coincidences count with delay δ on the window $[a,b]$, denoted X, is defined by:*

$$X = \int_{[a,b]^2} \mathbf{1}_{|x-y|\leq\delta} dN_1(x)dN_2(y), \tag{3}$$

where dN_1 (resp. dN_2) is the point measure associated with N_1 (resp. N_2)[5]

If we consider this definition, the coincidences count with delay δ is the total number of couples (x,y) satisfying all the following properties:

- x corresponds to a spike in the spike train N_1,
- y corresponds to a spike in the spike train N_2,

[4] More precisely, if $X_1,...,X_n$ are iid Gaussian variables with mean m and variance σ^2 then $\sqrt{n/\sigma^2}\sum_{i=1}^n(X_i-m) \sim \mathcal{N}(0,1)$ whereas $\sqrt{n/\hat{\sigma}^2}\sum_{i=1}^n(X_i-m) \sim T(n)$ where $\hat{\sigma}^2$ is the unbiased estimate of σ^2

[5] This means for instance that for any function f,

$$\int f(x)dN_1(x) = \sum_{x\in N_1} f(x),$$

where x represents any spike registered on the first neuron.

- x and y belong to $[a, b]$
- the delay $|x - y|$ is less than δ.

Note that X is integrative and increases with δ. The next result gives the expectation and variance of X under the following assumption.

Assumption 3. *The delay δ is less than half the window size, i.e. $\delta \leq (b - a)/2$.*

Theorem 1. *Under Assumptions 1, 2, 3 and H_0, the expectation of X and the variance*[6] *of X are given by*

$$m_0 := \mathbb{E}(X) = \lambda_1 \lambda_2 \left[2\delta(b - a) - \delta^2\right]$$

and

$$\mathrm{Var}(X) = \lambda_1 \lambda_2 \left[2\delta(b - a) - \delta^2\right] + \left[\lambda_1^2 \lambda_2 + \lambda_1 \lambda_2^2\right] \left[4\delta^2(b - a) - \frac{10}{3}\delta^3\right]$$

where λ_j is the constant intensity of N_j for $j = 1, 2$.

The proof is given in the appendix.

In the original article (Grün et al, 1999), following the approximation done in the binning framework, X is thought to be a Poisson variable with mean

$$m_g := 2\lambda_1 \lambda_2 \delta(b - a)$$

under H_0 whereas Assumptions 1, 2 are also made. Theorem 1 shows that those assumptions are not compatible even if some approximations may make them valid in a certain range of parameters as we detail below.

- First of all, the quantity m_0 differs from m_g by a quadratic term in δ. This additional term results from the consideration of the edge effects. Indeed, in the integral which defines X, when x or y are near a or b, the length of the interval $|x - y| \leq \delta$ on which we consider the coincidences is strictly less than 2δ since x and y are both in the window $[a, b]$. Note that if $\delta > (b - a)/2$ then all the couples (x, y) in $[a, b]$ are affected by this edge effect and in this case, the above formula for the expectation and variance are not valid anymore. Other definitions of coincidences [7] which evacuate this edge effect in the computation of m_0 are non symmetric in (N_1, N_2) and are not considered in this article. Therefore m_0 is only correctly approximated by m_g when δ is small with respect to $(b - a)$.

[6] The variance of a quantity T is given by

$$\mathrm{Var}(T) = \mathbb{E}\left([T - \mathbb{E}(T)]^2\right).$$

The variance measures the fluctuation of a quantity around its expectation.

[7] such as $\int_{[a,b]} \int_{\mathbb{R}} \mathbf{1}_{|x-y| \leq \delta} dN_1(x) dN_2(y)$

- More importantly, X cannot be a Poisson variable. If we consider the ratio between the theoretical variance and the theoretical mean defined in Theorem 1[8]

$$\frac{\mathrm{Var}(X)}{\mathbb{E}(X)} = 1 + 2(\lambda_1 + \lambda_2)\delta(1 + o(1)). \tag{4}$$

  This ratio depends on $\lambda_1 + \lambda_2$. Note that for a Poisson variable, this ratio is 1 whereas here this ratio becomes much larger when $\lambda_1 + \lambda_2$ increases. Thus, the gap between the true distribution of X and the Poisson model increases with $\lambda_1 + \lambda_2$. Large values of the firing rates as well as large values of δ tends to make the approximation worse.

2.2 Test of independence on a window $[a, b]$

From now on, we focus on the notion of delayed coincidences count and the single test is based on this quantity. Once again, we compare two estimates of m_0: the observed averaged coincidences count

$$\bar{m} = \frac{1}{M}\sum_{i=1}^{M} X^{(i)}, \tag{5}$$

where $X^{(i)}$ is the coincidences count with delay δ for the ith trial, and an estimate of the expected coincidences count under H_0:

$$\hat{m}_0 = \hat{\lambda}_1\hat{\lambda}_2[2\delta(b-a) - \delta^2] \tag{6}$$

where the $\hat{\lambda}_j$'s are defined by (2). We want to reject H_0 when the difference between $\bar{m}$ and $\hat{m}_0$ is too large. If the Central Limit Theorem together with Theorem 1 tells us that under H_0, $\sqrt{M}(\bar{m} - m_0)$ can be well approximated[9] by $\mathcal{N}(0, \mathrm{Var}(X))$, this is not true when we add the plug-in step and replace m_0 by $\hat{m}_0$, as explained in the next mathematical result.

Theorem 2. *Under Assumptions 1, 2, 3 and H_0, the following convergence holds*

$$\sqrt{M}\,(\bar{m} - \hat{m}_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

where

$$\sigma^2 = \lambda_1\lambda_2\left[2\delta(b-a) - \delta^2\right] + \lambda_1\lambda_2\left[\lambda_1 + \lambda_2\right]\left[\frac{2}{3}\delta^3 - (b-a)^{-1}\delta^4\right].$$

Moreover σ^2 can be estimated by

$$\hat{\sigma}^2 = \hat{\lambda}_1\hat{\lambda}_2\left[2\delta(b-a) - \delta^2\right] + \hat{\lambda}_1\hat{\lambda}_2\left[\hat{\lambda}_1 + \hat{\lambda}_2\right]\left[\frac{2}{3}\delta^3 - (b-a)^{-1}\delta^4\right]$$

and

$$\sqrt{M}\,\frac{\bar{m} - \hat{m}_0}{\sqrt{\hat{\sigma}^2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

[8] In the following equation, $o(1)$ denotes a quantity that tends to 0 when δ tends to 0.

[9] $\mathcal{N}(m, \sigma^2)$ denotes a Gaussian law with mean m and variance σ^2.

The symbol $\xrightarrow{\mathcal{L}}$ means "convergence in law when M tends to infinity". This means for instance that the quantiles of $\sqrt{M}\,(\bar{m} - \hat{m}_0)$ tend to those of $\mathcal{N}(0, \sigma^2)$, when M becomes larger. Therefore the plug-in step which consists in substituting m_0 by $\hat{m}_0$ modifies the shape of the variance: $\mathrm{Var}(X)$ is not σ^2 and this even for an infinite number of trials! On the other hand the substitution $\sigma^2 \to \hat{\sigma}^2$ makes no difference from an asymptotic point of view. Figure 1 illustrates the fact that even if this is an asymptotic result, the approximation by the Gaussian law holds for a reasonable number of trials.

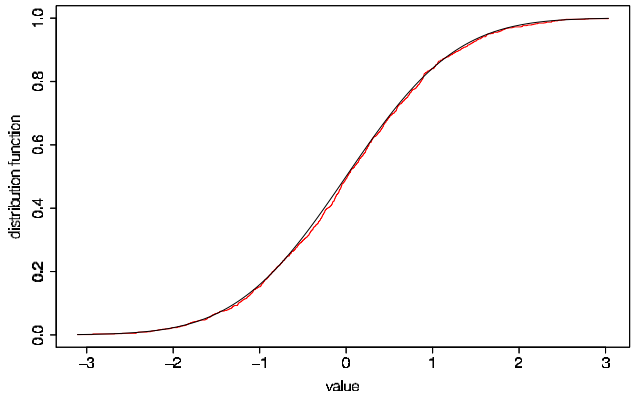

Figure 1: Cumulative distribution function (i.e. $t \to \mathbb{P}(T \leq t)$) under H_0 of $T = \sqrt{M}\,\frac{\bar{m}-\hat{m}_0}{\sqrt{\hat{\sigma}^2}}$ estimated over 1000 runs in red with $M = 41$, $\lambda_1 = \lambda_2 = 50$ Hz and $\delta = 0.02$ s. In black the cumulative distribution function of $\mathcal{N}(0, 1)$.

Theorem 2 shows another interesting behavior which may explain to some extent why the original Poisson approximation intuited by Grün et al. leads to reasonable tests: the renormalization σ^2 is much closer to m_0 than $\mathrm{Var}(X)$, the corrective term being now a cubic function of δ and not a quadratic term anymore.

This Gaussian Approximation of the UE method - denoted GAUE in the following - leads to three different single tests depending on what needs to be detected:

- the *symmetric test* $\Delta^{sym}_{GAUE}(\alpha)$ of H_0: "N_1 and N_2 are independent" versus H_1: "N_1 and N_2 are dependent", which rejects H_0 when $\bar{m}$ and $\hat{m}_0$ are too different:

$$|\bar{m} - \hat{m}_0| \geq z_{1-\alpha/2}\sqrt{\frac{\hat{\sigma}^2}{n}} \tag{7}$$

- the *unilateral test by upper value* $\Delta^{+}_{GAUE}(\alpha)$ which rejects H_0 when $\bar{m}$ is too large:

$$\bar{m} \geq \hat{m}_0 + z_{1-\alpha}\sqrt{\frac{\hat{\sigma}^2}{n}} \tag{8}$$

- the *unilateral test by lower value* $\Delta^{-}_{GAUE}(\alpha)$ which rejects H_0 when $\bar{m}$ is too small:

$$\bar{m} \leq \hat{m}_0 - z_{1-\alpha}\sqrt{\frac{\hat{\sigma}^2}{n}} \tag{9}$$

where z_t is the t-quantile of $\mathcal{N}(0,1)$, i.e. the real number z_t such that $\mathbb{P}(\mathcal{N}(0,1) \leq z_t) = t$.

Note that the original Unitary Event Multiple Shift method can be formalized in the same way by

- a *symmetric test* $\Delta_{UE}^{sym}(\alpha)$ which rejects H_0 when $M\bar{m} \leq q_{\alpha/2}$ or $M\bar{m} \geq q_{1-\alpha/2}$
- the *unilateral test by upper value* $\Delta_{UE}^{+}(\alpha)$ which rejects H_0 when $M\bar{m} \geq q_{1-\alpha}$
- the *unilateral test by lower value* $\Delta_{UE}^{-}(\alpha)$ which rejects H_0 when $\bar{m} \leq q_{\alpha}$

where q_t is the t-quantile of a Poisson variable whose parameter is $M\hat{m}_g := 2M\delta(b-a)\hat{\lambda}_1\hat{\lambda}_2$.

Theorem 2 tells that all the three Δ_{GAUE} tests are of level α asymptotically, which means that the type I error of those tests - that is the probability of wrong rejection of H_0 by one of those tests - is asymptotically less than α. If Figure 1 makes us confident in the fact that the level is correctly estimated even for finite M, it is less clear that the original UE method is valid. The plot in Figure 2 is purely heuristic and should help the reader in understanding the difference between the two approaches: the interesting property of Gaussian variables is that for every deterministic m and σ^2, $m + \sigma * \mathcal{N}(0,1)$ is a $\mathcal{N}(m, \sigma^2)$. Extrapolating this property - since the parameters are random - Theorem 2 tells us that $M\bar{m}$ heuristically obeys $\mathcal{N}(M\hat{m}_0, M\hat{\sigma}^2)$. Therefore its density can be plotted on the same space as the distribution postulated in the UE method, which is a Poisson with mean $M\hat{m}_g$ (Grün et al, 1999). This can obviously be done for just one run of both procedures because at each run the quantities $M\hat{m}_0, M\hat{\sigma}^2$ and $M\hat{m}_g$ will be different, even if the fact that $\hat{m}_g \geq \hat{m}_0$ will lead to the same systematic shift on the right for the UE approximation with respect to the GAUE one. Conversely and even heuristically, it is not possible to transform the UE approximation into something that can be plotted on Figure 1, because the parameter inside the Poisson law is random, and cannot be put outside of the law.

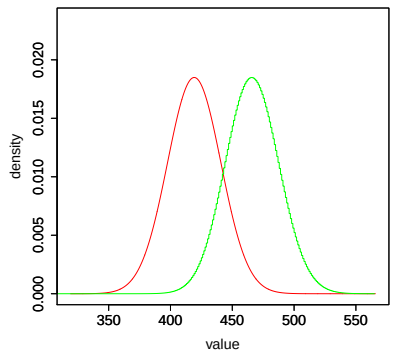

Figure 2: Density curve of a Gaussian distribution with mean $M\hat{m}_0$ and variance $M\hat{\sigma}^2$ in red and histogram curve of a Poisson distribution with parameter $M\hat{m}_g$ in green, for one run of simulation. $M = 41$, $\lambda_1 = \lambda_2 = 50$ Hz and $\delta = 0.02$ s.

Therefore, in order to produce more rigorous results, we estimate the type I error of both classes of tests by simulating 1000 times M independent and identically distributed couples $(N_1^{(i)}, N_2^{(i)})$ of independent Poisson processes and by counting how many times the tests have wrongly rejected H_0. The results are displayed on Figure 3.

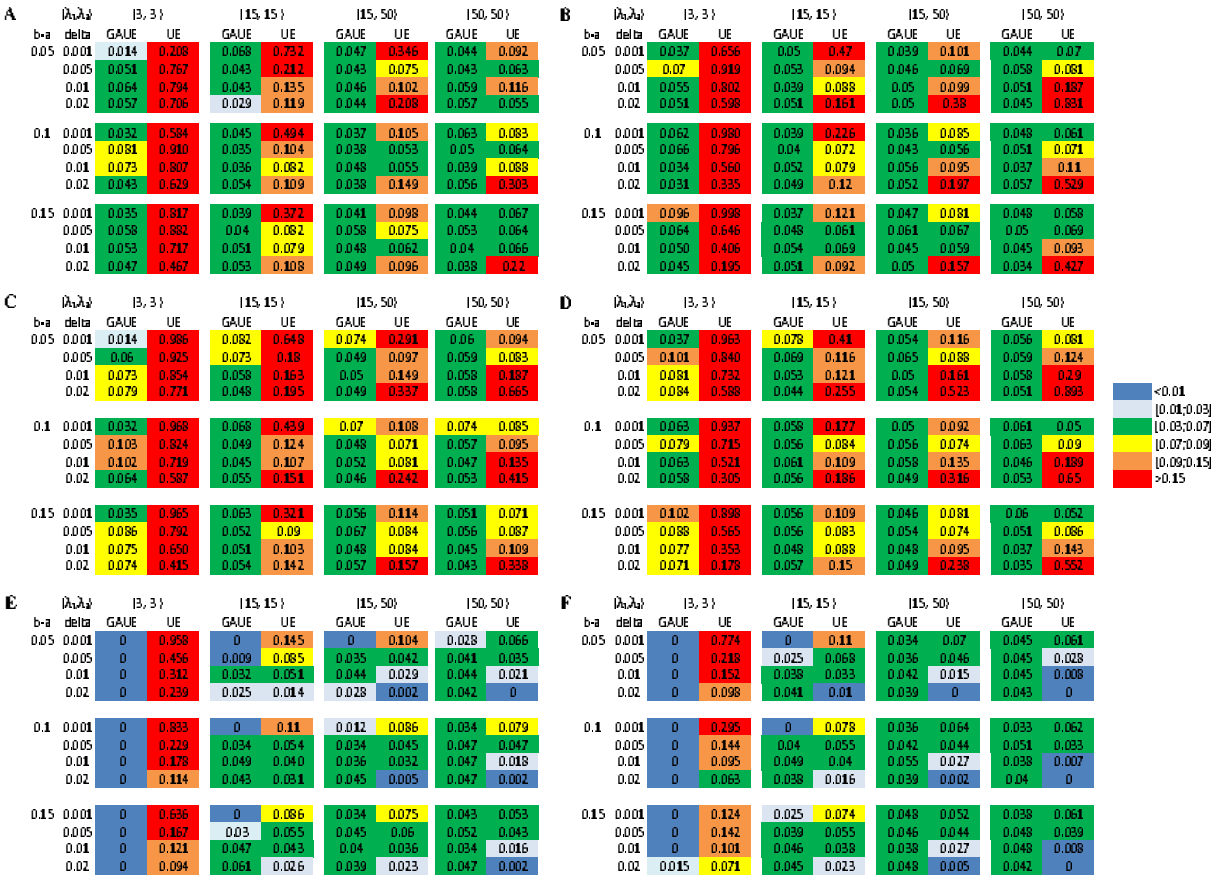

Figure 3: Estimates of the Type I errors of the single tests over 1000 runs. A, C and E are for $M = 19$ trials while B, D and F are for $M = 41$ trials. A and B are related to the Δ^{sym}'s, C and D are related to the Δ^{-}'s whereas E and F deal with the Δ^{+}'s. The window length $b - a$ varies in $\{0.05\ ,\ 0.1\ ,\ 0.15\}$ seconds, the delay δ in $\{0.001\ ,\ 0.005\ ,\ 0.01\ ,\ 0.02\}$ seconds and the λ_j are either $3, 15$ or 50 Hz. The parameter α is fixed to 0.05. A color code is given additionally to the exact values on the right.

Paradoxically having the smallest type I error is not the best case. Indeed, this indicates in testing theory a conservative test that has much more difficulty to detect a dependent case than a test of exact error α. Let us develop this notion. Since thousand runs are performed, it is possible "just by chance" that some of those runs "badly behaved" and lead to a detection whereas there should be none. The main point is consequently to control the error, that is the probability of wrong rejection. If this probability is zero, then the test would never reject anything even if there is a dependence: a similar case "just by chance" would have happened in our thousand runs. So we need to control the type I error not with 0 but with something small - here 5%. This is what should happen

for the Δ_{GAUE} tests because of Theorem 2, at least when M is large enough. It means that we allow our test to wrongly detect in 5% of the cases, in order to allow it to detect more when there will be a dependence. This is our aim: having a type I error as close to 5% as possible. The rejection of a more conservative test (type I error strictly less than 5%) can be trusted but it can potentially miss some adequate rejection when one works under the alternative H_1 that is when the spike trains are dependent. On the other hand, when the type I error is large, such a test can obviously not be trusted because it potentially detects something that does not exist (here dependence) with a probability as large as this type I error.

This leads to the following interpretation of Figure 3. The GAUE tests behave most of the time as expected, with an error close to 5%. If it is not the case, they have a tendency to be too conservative, mainly because there are not enough points for the Gaussian approximation to work and this can be due to various causes: small number of trials (M), small firing rates (λ_j) or too short period of time ($b-a$). Note in particular that for very small firing rates (here 3 Hz) which are much lower than the boundary 7 Hz usually used in the UE analysis (Roy et al., 2000), the first kind error of the GAUE tests is close to 5% for Δ^{sym}_{GAUE} and Δ^{-}_{GAUE} and too conservative (the error is null) for Δ^{+}_{GAUE}. In any case since the default is a conservative default, this means that one can trust the detections of the GAUE tests.

On the other side, the original UE tests behave very badly particularly for the symmetric and lower cases where the error can reach levels higher than 50%. This is particularly true for high firing rate λ and large δ and this fact deteriorates when the number of trials increases! This is because as said before (see (4)) the Poisson approximation is worse when λ and δ increase and increasing the number of trials does not make this approximation better.

Note however that the Δ^{+}_{UE} is on the contrary too conservative for not too small firing rates (larger than 15 Hz): this fact partly explains why the UE method works in practice when the detections are due to a statistically too large observed coincidences count. Note also that this confirms the intuition given by the heuristic Figure 2 where the UE approximation is shifted on the right. On the contrary, the Δ^{+}_{UE} reaches first type errors potentially as large as 80% for very small firing rates (3 Hz) and consequently the UE analysis is not suitable in practice in this range of firing rates as already noted by (Roy et al., 2000).

By definition and for both classes of tests (UE or GAUE), the symmetric test $\Delta^{sym}(\alpha)$ rejects H_0 if and only if either $\Delta^{+}(\alpha/2)$ or $\Delta^{-}(\alpha/2)$ rejects H_0. That is why for the remaining part of the paper, we focus the GAUE method on the symmetric test, which contains both Δ^{+} and Δ^{-} tests at a smaller level.

3 Detection of local dependence on $[0, T]$

All the previous computations are done under Assumptions 1, 2 and 3. But the trials recorded on $[0, T]$ are not stationary in time. That is why those tests must be performed on smaller time period than $[0, T]$, when the stationary assumption is (almost) true. Let $\mathcal{W}$ be any collection of windows, W, of $[0, T]$ (in particular they can potentially overlap) and let K denotes the cardinality of $\mathcal{W}$. Then we can perform on each window W

any of the previous Δ_{GAUE} tests, now denoted Δ_W to underline the dependence on the window on which the test has been applied.

If all those tests are performed at level α, the potential detections (i.e. the windows W where Δ_W rejects) are actually not controlled at all. Assume for instance that all the Δ_W's are independent then if H_0 is satisfied on the whole interval $[0, T]$, the probability to have no detection at all (which we would like to be large, since all of the intervals are under H_0) is

$$\mathbb{P}(\forall W \in \mathcal{W}, \Delta_W \text{ accepts}) \rightarrow_{M\rightarrow\infty} (1-\alpha)^K \rightarrow_{K\rightarrow\infty} 0.$$

This means that when K grows, this procedure is doomed to reject at least 1 and in average $K\alpha$ tests. Note that the UE procedure follows this approach and is consequently not controlled in term of false detections (Grün et al, 1999).

3.1 Multiple tests: a Benjamini and Hochberg's approach

One way to control such a problem is to control the so called *Family Wise Error Rate* (FWER) (Hochberg & Tamhame, 1987), which consists in controlling $FWER = \mathbb{P}(\exists W \in \mathcal{W}, \Delta_W \text{ wrongly rejects})$. This can be easily done by Bonferroni bounds:

$$\mathbb{P}(\exists W \in \mathcal{W}, \Delta_W \text{ wrongly rejects}) \leq \sum_{W\in\mathcal{W}} \mathbb{P}(\Delta_W \text{ wrongly rejects}) \rightarrow_{M\rightarrow\infty} K\alpha.$$

So Bonferroni's method (Holm, 1979) consists in applying the Δ_W tests at level α/K instead of α to guarantee a FWER less than α, the detections (also named discoveries) being given by the rejected tests. However and as said before, the smaller the type I error is, the more difficult it is to make a rejection. So when K is large, it potentially leads to no discovery/detection at all even in cases where dependent structures exist.

Another notion, popularized by (Benjamini & Hochberg, 1995) has consequently been introduced in the multiple testing areas leading to a large amount of publications in statistics, genomics, medicine etc in the past ten years (Goeman & Solari, 2011, and the references therein). This is the *False Discovery Rate* (FDR). Actually a false discovery (also named false detection or false positive) is not that bad if the ratio of the number of false discoveries divided by the total number of discoveries is small.
More formally, let us use the notations given in Table 1.

Then the *False Discovery Rate* is defined by

$$FDR = \mathbb{E}\left(\frac{V}{R}1_{R>0}\right).$$

Note that when both spike trains are independent for all windows, $K_1 = 0$, which leads to $S = 0$ and $V = R$. Hence the FDR in the full independent case is also a control of $\mathbb{P}(\exists W \in \mathcal{W}, \Delta_W \text{ wrongly rejects})$, i.e. the FWER. In all other cases, $FDR \leq FWER$. This means that when there are some W for which the independence assumption does

Number of W such that	Δ_W accepts	Δ_W rejects	Total
Independence on W	U = " true negative"	V = "false positive"	K_0 = "number of H_0"
Dependence on W	T = "false negative"	S = "true positive"	K_1 = "number of H_1"
Total	$K - R$	R= "discoveries"	K

Table 1: Repartition of the answers in a set of K tests.

not hold, controlling the FDR is less stringent, whereas the relative confidence that we can have in the discoveries is still good: if we make 100 discoveries with a FDR of 5%, this means that in average only 5 of those discoveries are potentially wrong.
The question now is: how to guarantee a small FDR? To do so, Benjamini and Hochberg (Benjamini & Hochberg, 1995) proposed the following procedure. For each test Δ_W, the corresponding p-value[10] P_W is computed. They are next ordered such that:

$$P^{(1)}_{W(1)} \leq .... \leq P^{(m)}_{W(m)} \leq P^{(K)}_{W(K)}.$$

Let $q \in [0, 1]$ be a fixed upper bound that we desire on the FDR and define:

$$k = \max\{m \text{ such that } P^{(m)}_{W(m)} \leq mq/K\}.$$

Then the discoveries of this BH-method are given by the intervals $W(1), ..., W(k)$ corresponding to the k smallest p-values.
The theoretical result of (Benjamini & Hochberg, 1995) can be translated in our framework as follows: if the p-values are uniformly and independently distributed under the null hypothesis, then the procedure guarantees a FDR less than q.

3.2 MTGAUE: Multiple Testing based on a Gaussian Approximation of the Unitary Events

MTGAUE consists in combining Benjamini-Hochberg approach with the GAUE tests of independence. Note that in our case, the assumptions required in the approach of Benjamini-Hochberg are not satisfied. Indeed, the tests Δ_W are only asymptotically of

[10] A p-value is the random value of α for which the test $\Delta(\alpha)$ passes from "accept" to "reject". Note that usually when $\alpha = 0$, the test always accepts, whereas it always rejects when $\alpha = 1$: therefore there is a limit value which depends on the observations for which one passes from one decision to another one. If the test is of type I error exactly α for all α, then one can prove that the corresponding p-value is uniformly distributed on $[0, 1]$ under H_0.

type I error α, which is equivalent to the fact that asymptotically and not for fixed M, the p-values are uniformly distributed. Therefore, there is a gap between theory and what we have in practice. However, as we illustrate later on simulations, this difference does not seem to significantly impact the FDR.
Moreover, to have independent p-values we should consider disjoint windows $[a, b]$ (see the comments after Assumptions 1). However, in order to detect very local phenomena, it is preferable to consider sliding windows $[a, b]$ that overlap. In theory, few results exist in this context - see for instance (Benjamini & Yekutieli , 2001). In practice, we will see in the next section that this lack of independence does not impact the FDR as well.

4 Simulation study

Some simulations have been performed to understand the behavior of MTGAUE in practice with respect to the original UE method.

4.1 Description of the simulated data

Our data consist in M realizations of a succession of dependent and independent couple of processes (N_1, N_2) through time. To do so, dependence on particular time periods $J \subset [0, T]$ must be simulated. In the original UE method (Grün, 1996), the injection model is used. It can be described as follows: 3 independent Poisson processes A, B, C with respective intensity $\lambda_A, \lambda_B, \lambda_C$ are simulated and the point processes (N_1, N_2) are given by

$$N_1 = A \cup C \text{ and } N_2 = B \cup C.$$

The main drawback of this model is that it does not allow any random delay, since by construction, the additional coincidences are exact (with null delay). There is also no random fluctuation in the number of additional coincidences.
A simple version of Hawkes processes is consequently used here instead of the injection model to allow both randomness in the delay δ and in the number of additional coincidences (Hawkes, 1971; Daley & Vere-Jones, 2003). In this model, the first process N_1 is a Poisson process with intensity λ_1. Independently a second Poisson process B with intensity ν is simulated. To add coincidences, conditionally to N_1, each point x of N_1 gives birth to children on N_2 according to a Poisson process N^x of intensity θ on $[x + \mu, x + \kappa]$, for $\kappa - \mu$ small. The process N_2 is then given by

$$N_2 = B \cup \bigcup_{x \in N_1} N^x.$$

Note that a child y of a parent x forms with its parent a coincidence with random uniform delay $(y - x) \in [\mu, \kappa]$. In this article, such coincidence is called *imposed coincidence*. Moreover the number of children for a fixed parent x is random: it obeys a

Poisson law and can therefore be null eventually but also larger than 2. Consequently the number of imposed coincidences even for a fixed parent is random.
Here the main point is that the mean number of children per point is given by $\theta(\kappa - \mu)$, which is classically taken strictly smaller than 1 (Daley & Vere-Jones, 2003). Therefore a significant proportion of the points x have no children (with probability $1 - \exp(-\theta(\kappa - \mu))$) and some of the points have children as in the injection model. One can also rewrite this as follows: the intensity of N_2, given N_1, is given by

$$\lambda_2 = \nu + \int_{t-A}^{t} h(t-u) dN_1(u), \tag{10}$$

where $h = \theta 1_{[\mu,\kappa]}$.

More generally, multivariate Hawkes process can model cross and even self dependence. It has recently been used to model spike trains (Krumin et al., 2010; Pernice et al., 2011, 2012).

4.2 Validation of the MTGAUE method and comparison to the original UE method

Dependence in the different parameters The MTGAUE has been implemented with the symmetric tests as described in Section 3 with parameter q. The original UE method does not correct the detections for the multiplicity of the tests. So for comparison, all the tests Δ_{UE}^{sym} have been consequently performed with parameter $\alpha = q$ and the discoveries of the UE method are given by the rejections of the single tests at level α. The simulations have been performed in various situations all of them having two periods of dependence. The results are displayed on 6 simulations with different parameters in Figure 4. This leads to the following comments:

- Influence of the delay δ: Figure 4.A shows that MTGAUE is able to accurately recover the dependence periods. There are some extra detections but as expected the ratio false positive / detections is very small and this for all the values of δ. On the contrary the original UE method detects a lot of false positive, in particular when $\delta \geq 0.02$ s. This is compatible with the fact that nobody looked at a further range in the literature (Riehle et al., 2000; Grammont & Riehle, 2003).

- Influence of the level θ of interaction: On Figure 4.B, the interaction is very small. Indeed when in 4.A, 8 points over 10 in N_1 have a child on N_2 roughly speaking, only 1 point over 10 has a child in 4.B. Moreover on a window of length 0.1 s there is at most 2 points in average on N_1, which may potentially have a child. Therefore the number of imposed coincidences is much smaller in 4.B than in 4.A. So it is likely that nothing is detected at all in 4.B and this is what happens for MTGAUE whereas the UE method still detects a lot of non relevant windows.

- Influence of the number of trials M: On Figure 4.C, the number of trials is much smaller than on 4.A. MTGAUE is still very accurate in its detections whereas the UE method suffers from the same drawbacks as in Figure 4.A.

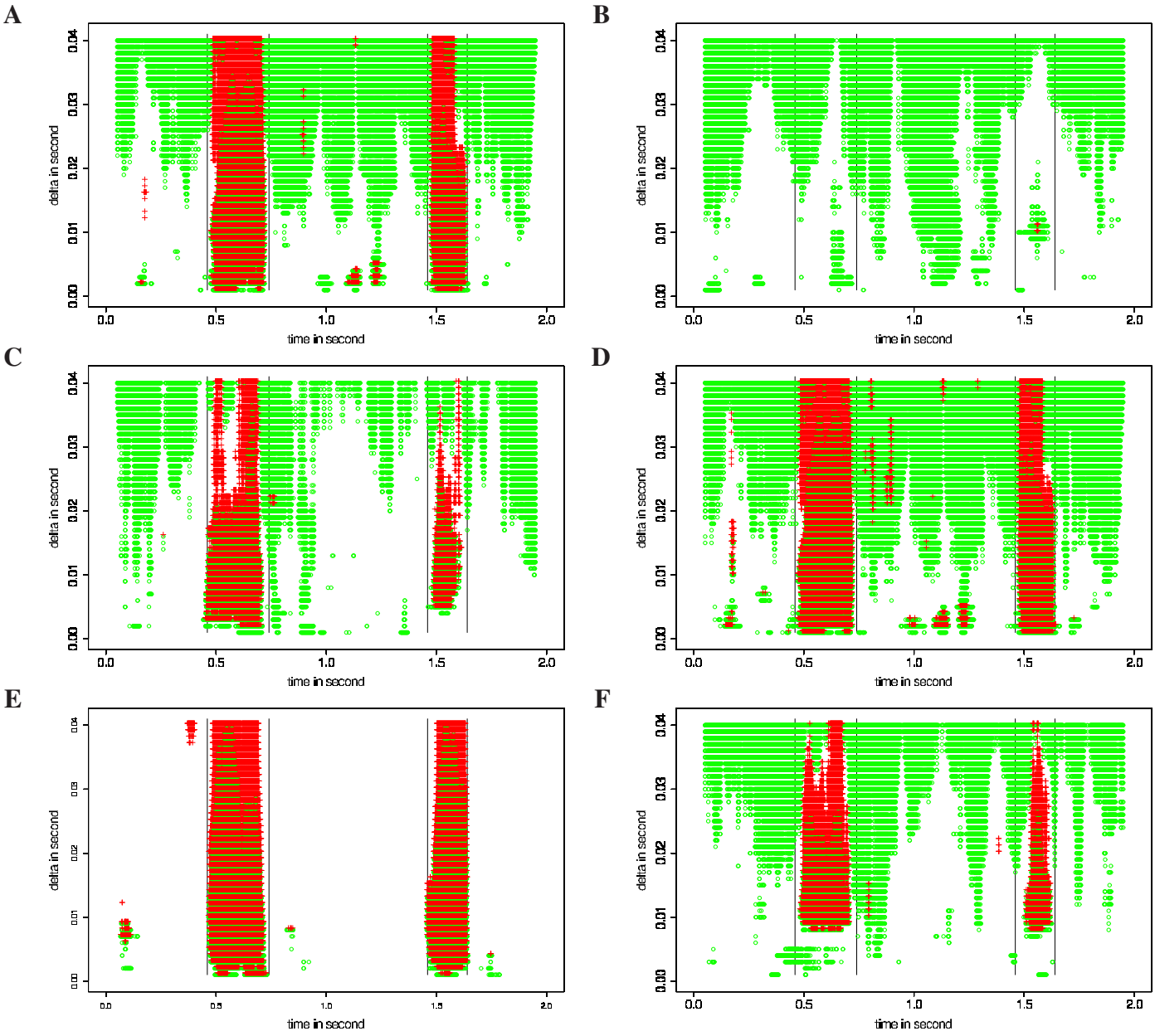

Figure 4: Detections made by the MTGAUE method (in red) and the original Multiple Shift UE method (in green). 1900 single tests were performed on 1900 overlapping (sliding) windows of length 0.1 s shifted by 0.001 s. The corresponding detection is marked by a color cross at the center of the window. Each line corresponds to a different delay: 40 different delays δ from 0.001 s to 0.04 s are considered. Two periods of dependence are considered $[0.5, 0.7]$ s and $[1.6, 1.7]$ s. The black vertical lines delimit the regions where the tests should detect a dependence, that is each time the window intersects a dependence period. The firing rates are given by : $\lambda_1 = 20$ Hz and $\lambda_2 = 50$ Hz, in the non-dependent periods. The dependent periods are modelled by (10) with $\lambda_1 = 20$ Hz and $\nu = 50$ Hz . A: $M = 41$ trials and $h = 801_{[0,0.01]}$, symmetric detections with $q = \alpha = 0.05$. B: same as A but with $h = 101_{[0,0.01]}$. C: same as A but with $M = 19$. D: same as A but with $q = \alpha = 0.01$. E: same as A but only the intervals where the observed coincidences count is significantly too large are displayed. F: same as A but with $h = 801_{[0.005,0.01]}$.

- Influence of the desired FDR level $q = \alpha$: On 4.D, the level q increases with respect to 4.A. If MTGAUE is very stable - it detects almost the same regions as in 4.A - the UE method detects on the contrary almost all the possible windows as soon as $\delta \geq 0.02$ s and detects a lot of non relevant periods even for $\delta \leq 0.02$ s. Note that as said previously, taking $\alpha = q = 0.1$ amounts to the detection with $\Delta^{+}_{UE}(\alpha/2 = 0.05)$ and $\Delta^{-}_{UE}(\alpha/2 = 0.05)$, i.e. the level classically used in those analysis.

- Influence of the upper/lower values: 4.E suggests an explanation of the previous phenomenon. Indeed we plot on this figure only the detections of the symmetric tests for which $\bar{m} \geq \hat{m}_0$, so they corresponds for the UE test to the discoveries done when the observed coincidences count is too large with respect to the estimate of the expected count. We see that MTGAUE is still very good and that almost all its detections were by upper values. More importantly we see that the UE method is very accurate too: in fact two errors come into play, on one side, the UE tests by upper values are too conservative (see Section 2.2) but on the other side, they are not adjusted for multiplicity and hence have a tendency to create too many false positive. At the end, except for high values of δ, the detections by upper values of the UE method are as good as MTGAUE. In particular for $\delta \leq 0.02$ s they are almost identical.

- Integrative aspects: Note that the coincidences count is integrative in δ and therefore it increases with δ. Quite logically, once a certain delay is detected it should remain detected. But another phenomenon may appear: if the coincidences have a maximal delay of say 0.01 s, then the proportion of imposed coincidences may become much smaller when $\delta >> 0.01$ s with respect to the coincidences appearing just by chance. So Figure 4.E also shows that there is enough coincidences for MTGAUE to detect dependence up to $\delta = 0.04$ s. This is not the case for the UE method because the Poisson approximation is too inadequate for large δ ($\delta \geq 0.03$ s).

- Strictly positive lower bound for the delay: On 4.F, the delay of interaction lies in $[0.005, 0.01]$ s, therefore there should not be any detection when $\delta < 0.005$ s. Actually one has to wait until $\delta = 0.008$ s to be able to detect both dependence periods, that is when there is enough imposed coincidences to make a difference. So it is possible to detect a delay of interaction which is lower bounded by a strictly positive quantity and this phenomenon can be seen on the plots by areas of detections that are dense but not present for the smallest δ's. The detections of MTGAUE become scarce in the smallest dependence period when $\delta > 0.03$ s and this is due to the fact that as described before the imposed coincidences are diluted: with respect to 4.A, only roughly 4 points over 10 have a child in 4.F.

The graphs of Figure 4 are obtained on just one run of simulation. Figure 5 shows the estimated FDR over thousand runs for $q = \alpha = 0.05$. MTGAUE clearly respect the prescribed FDR of 0.05. Moreover, the UE method by upper value (at least when the interaction is high enough) respects it too. This confirms the results on one run of Figure 4.E: the conservative aspects of the UE single tests decrease the potential high

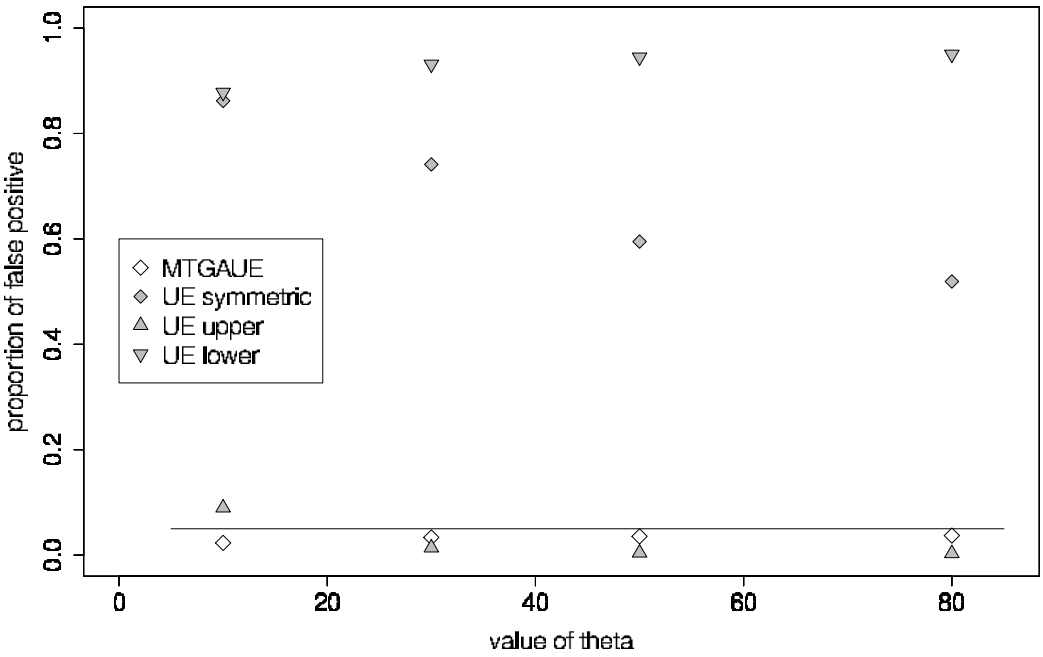

Figure 5: Estimated FDR over thousand runs in the conditions of Figure 4.A but with $h = \theta 1_{[0,0.01]}$ for $\theta = 10, 30, 50, 80$ represented on the abscissa. The white diamond corresponds to MTGAUE (symmetric) whereas the grey diamond corresponds to the symmetric UE. The up triangle corresponds to the UE detections by upper values whereas the down triangle corresponds to the UE detections by lower values.

number of false detections due to the fact that the UE procedure was not corrected for the multiplicity of the tests. On the other hand, the FDR of the UE method by lower value is so high (more than 80%) that one cannot trust any of its discoveries, the FDR of the symmetric UE being less high but still not trustworthy.

Robustness As said previously, the UE method by lower values is not trustworthy since it detects a lot of false positive. But it is not clear whether MTGAUE is able to detect anything. To produce simulated data with a negative interaction, the model given by (10) is transformed to allow h to take negative values. To do so, we use the positive part of the function as the new intensity of the process N_2, that is

$$\lambda_2 = \left(\nu + \int_{t-A}^{t} h(t-u) dN_1(u) \right)_+ \tag{11}$$

If this formula cannot have any interpretation in terms of parents/children and imposed coincidences with respect to Section 4.1, the following interpretation is still true: a negative interaction forces a low number of coincidences.

In Figure 6, $h = -501_{[0,0.01]}$. This is a very strong negative interaction which can be interpreted as follows: just after a spike on N_1, the firing rate of N_2 becomes null for a very short period of time of 0.01 s. Since the absolute value of h is smaller than the one used in Figure 4.A, the detections become more difficult. Figure 6 shows that even in this case, most of MTGAUE detections are at relevant positions of interaction. On the

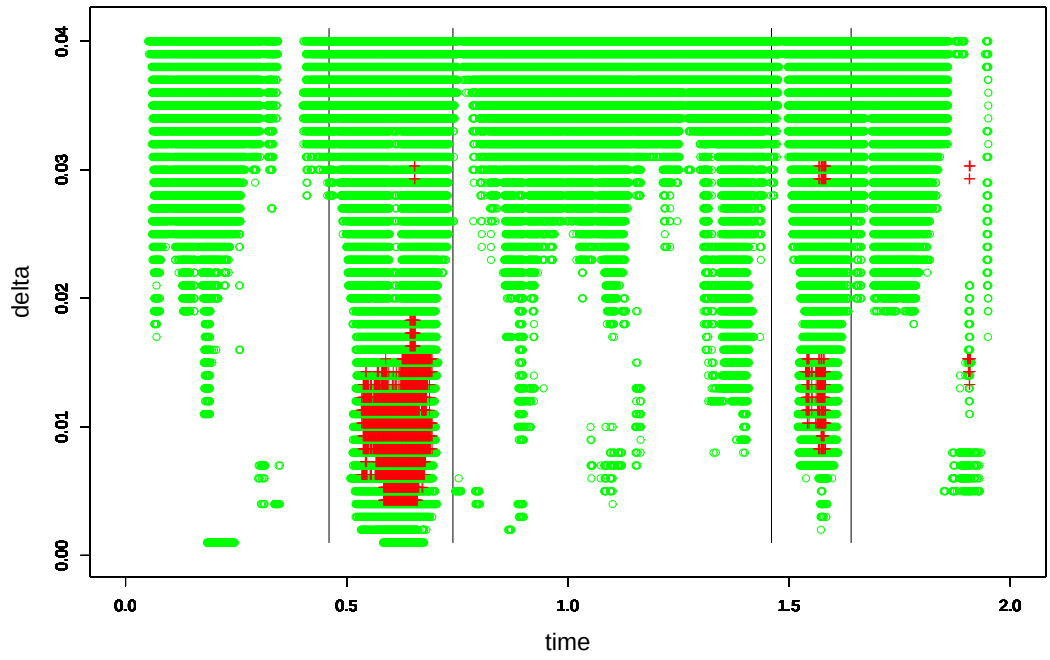

Figure 6: Same representation (and same parameters) as in Figure 4.A but with $h = -501_{[0,0.01]}$ in Equation (11).

contrary, the symmetric UE method detects almost the whole interval of observation for $\delta > 0.02$ s and even for smaller δ, the UE methods detects a lot of non relevant interactions at random positions.

Another potential problem that may be encountered in practice is whether the trials are truly iid or not. Indeed, it is possible (as we will see on the real data analysis) that different experimental conditions are regrouped in order to have a sufficiently high number of trials and therefore the identical distribution between trials would not be valid anymore. Other phenomenons that cannot be controlled by the experimental conditions may also appear and therefore it is not always possible to properly sort the truly iid trials.

Figure 7 shows that even if a small proportion of the trials (just one fourth) have indeed an interaction, MTGAUE is able to recover them. Moreover the number of discoveries increase with the ratio of trials having indeed an interaction.

As a conclusion of this simulation part, MTGAUE produces detections of dependence with a controlled False Discovery Rate on a large range of parameters. MTGAUE can correctly detect dependence even if the interaction is negative (too low number of coincidences) or if the trials do not all show the same pattern of interaction. On the contrary, following the present study, the UE analysis is valid only for the detection due to a significantly too large coincidences count for $\delta < 0.02$ s.

A

B

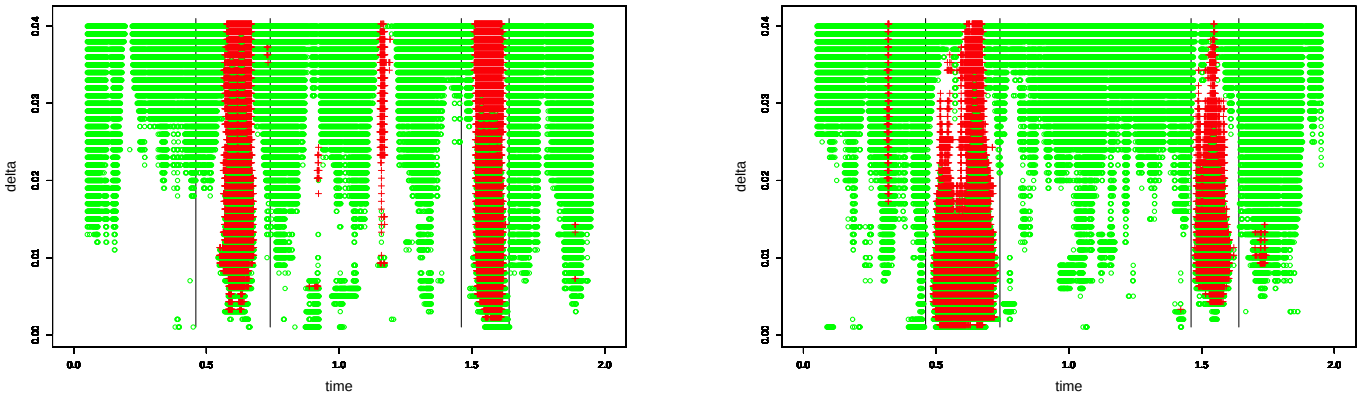

Figure 7: Same representation (and same parameters) as in Figure 4.A but in A, only 10 trials over 41 were simulated according to Figure 4.A, the other 31 trials being simulated as pure homogeneous independent Poisson processes during the whole observation time with $\lambda_1 = 20$ Hz and $\lambda_2 = 50$ Hz; in B, same situation with a ratio of 20 over 41 trials simulated according to Figure 4.A.

5 Real data study

MTGAUE being validated on simulated data, the method is applied on real data and some practical tools are provided to evaluate the quality of the detection.

5.1 Description of the data

Behavioral procedure The data used in this theoretical article to test the validity of the MTGAUE method were already partially published in previous experimental studies (Riehle et al., 2000; Grammont & Riehle, 2003; Riehle et al., 2006). These data were collected on a 5-year-old male Rhesus monkey who was trained to perform a delayed multidirectional pointing task. The animal sat in a primate chair in front of a vertical panel on which seven touch-sensitive light-emitting diodes were mounted, one in the center and six placed equidistantly (60 degrees apart) on a circle around it. The monkey had to initiate a trial by touching and then holding with the left hand the central target. After a delay of 500 ms, the preparatory signal (PS) was presented by illuminating one of the six peripheral targets in green. After a delay of either 600 or 1200 ms, selected at random with various probability, it turned red, serving as the response signal and pointing target. During the first part of the delay, the probability for the response signal to occur at 500+600 ms =1.1 s was either 0.3, 0.5 or 0.7, depending on the experimental condition. Once this moment passed without signal occurrence, the conditional probability for the signal to occur at 500+600+600 ms =1.7 s changed to 1. The monkey was rewarded by a drop of juice after each correct trial. Reaction time (RT) was defined as the release of the central target. Movement time (MT) was defined as the touching of the correct peripheral one.

Recording technique Signals recorded from up to seven microelectrodes (quartz insulated platinumtungsten electrodes, impedance: 25MO at 1000 Hz) were amplified and band-pass filtered from 300Hz to 10 kHz. Using a window discriminator, spikes from only one single neuron per electrode were then isolated. Neuronal data along with behavioral events (occurrences of signals and performance of the animal) were stored on a PC for off-line analysis with a time resolution of 1 kHz.

In the following study, only trials where the response signal occurs at 1.7 s are considered.

5.2 Quality of detection

Evaluation of the significance of detected windows is not as simple on real data as it is on simulated data. Indeed, the periods in which relevant dependence really lies are not a priori known.

For one single test, a criterion that assesses the quality of the detection is the well known p-value: the smaller the p-value is, the better the detection is. Performing the test at level $\alpha = 0.05$ consists in comparing the p-value to the threshold 0.05, rejecting H_0 when the p-value is smaller than the threshold.

However, when multiple testing procedures are applied, the single p-values are not sufficient anymore as explained in Section 3. More precisely their rank in the set of p-values is also actually a valuable information. Benjamini and Hochberg's procedure combines rank and values in order to provide a procedure with prescribed FDR less than q (Benjamini & Hochberg, 1995). Therefore, to provide the analogue of the single p-value for single tests, we consider *the multiplicity adjusted p-value* (Romano & Wolf, 2005; Goeman & Solari, 2011), defined for each tested interval by the smaller value of q that allows detection in the MTGAUE method. The underlying idea is that if the multiplicity adjusted p-value associated with an interval is weaker, it is more likely that the hypothesis H_0 is really at fault on this interval. Thus, the detections made by MTGAUE with $q = 0.05$ correspond to the detections for which the multiplicity adjusted p-value is less than 0.05.

Figure 8 displays the results when the MTGAUE and UE methods are applied on the activity of two pairs of neurons, recorded during the experiment described in Section 5.1.

The most trusted detections are the ones corresponding to a multiplicity adjusted p-value less than 0.01. Those intervals correspond to the detections of MTGAUE with $q = 0.01$ and consequently at the most 1% of those detections are false positive in average. The classical detections with $q = 0.05$ correspond to the intervals with a multiplicity adjusted p-value ≤ 0.05. For larger multiplicity adjusted p-values, the set of detected intervals contains too many false positives and it is therefore difficult to assert local dependence between neurons on those intervals. Thus, the multiplicity adjusted p-value is a practical tool to make a decision like the notion of p-value for single tests.

In this respect, on 8.A and 8.B several periods are detected which correlate with the occurrence of specific events of the behavorial protocol. Interestingly there is no real gain by looking at the detections for $q = 0.05$. Most of them are already detected

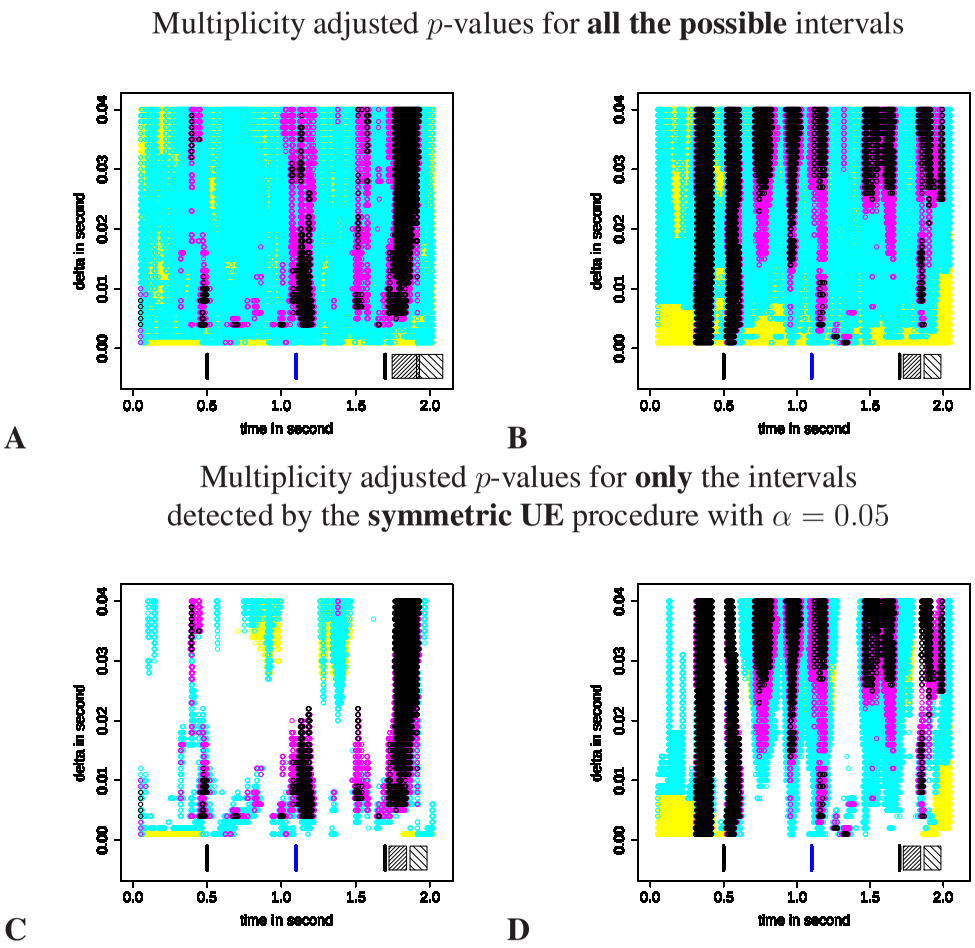

Figure 8: Representation of the detections according to MTGAUE and UE methods over sliding windows of length 0.1 s shifted by 0.001 s. Each detected window is associated to its multiplicity adjusted p-value thanks to the MTGAUE method: if the multiplicity adjusted p-value belongs to $[0, 0.01[$, a black circle is plot at the center of the corresponding window. For $[0.01, 0.05[$ the color is magenta. For $[0.05, 0.8[$ the color is cyan and yellow for $[0.8, 1[$. Each line corresponds to a different value of δ from 0 to 0.04 s. The first black vertical bar corresponds to the preparatory signal (PS), the blue vertical bar to the expected signal (ES), the second black vertical bar to the response signal (RS). The first hatched box correspond to the interval [mean reaction time (RT) minus its standard deviation, mean reaction time (RT) plus its standard deviation], the second hatched box corresponds to the same thing but for the movement time (MT). A and C correspond to the pair of neurons 13 in the experiment. B and D correspond to the pair of neurons 40 in our experiment. Graphs A and B show the MTGAUE multiplicity adjusted p-values of all the windows. Graphs C and D show the MTGAUE multiplicity adjusted p-values of the windows detected thanks to the symmetric UE procedure with $\alpha = 0.05$.

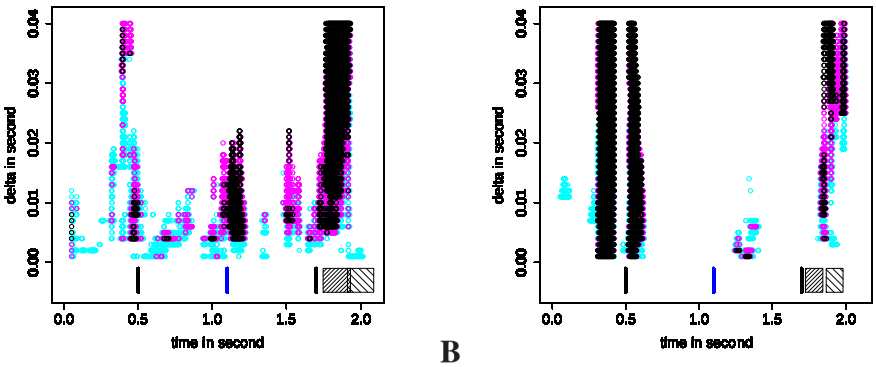

Figure 9: Representation of the detections according to UE and MTGAUE methods for which $\bar{m} - \hat{m}_0 > 0$, with the same convention as in Figure 8. A correspond to the couple of neurons 13 in the experiment. B correspond to the couple of neurons 40 in our experiment. Bars and boxes on the time axis as in Figure 8.

for $q = 0.01$ for larger δ. Since for these detections the FDR is less than 1% as we have seen in the simulation study, these detections are really significant. In particular, those appearing for $\delta \geq 0.02$ s on 8.B could not have been seen / trusted using the UE method. Indeed the detections using the UE method by upper values (see Figure 4.E) disappear for large values of δ and the detections using the UE method by lower values (see Figure 5) have a too large FDR. On the contrary, the simulation study makes us confident in the fact that such a large number of detections for $q = 0.01$ cannot be due to chance for MTGAUE. On 8.C and 8.D, we see that the symmetric UE method with $\alpha = 0.05$ detects a lot of intervals corresponding in reality to large multiplicity adjusted p-values (i.e. detected by MTGAUE only for $q \geq 0.8$). This corroborates the fact that the symmetric detections using the UE method may include a large number of false positive as we have seen in the simulation study.

5.3 Is the count significantly too low or too large ?

As a matter of fact (as already noted in Section 2.2), for single tests the detection of Δ^{sym} at level $\alpha = 0.05$ is just the detection of both tests Δ^+ and Δ^- at level $\alpha = 0.025$ and this for both UE and GAUE method. When dealing with a collection of tests, this result is still valid for the UE method which does not correct for multiplicity. However it is not valid anymore for MTGAUE since this method is based on the rank of all the p-values of all the symmetric tests, somehow intertwining detections by lower and upper values. Since the interest lies in both distinct detections (upper and lower values) on the experimental data, it is meaningless to independently perform tests by upper values and then tests by lower values. The correct way to use the method is consequently to perform MTGAUE with the symmetric tests. Once a multiplicity adjusted p-value has been assigned to the different windows, criterion measuring the quality of the detections of dependence, one can then ask the question whether this detection was due to a low count or large count by looking at the sign of $\bar{m} - \hat{m}_0$.

Figure 9 shows those detections when $\bar{m} > \hat{m}_0$ corresponding to a number of co-

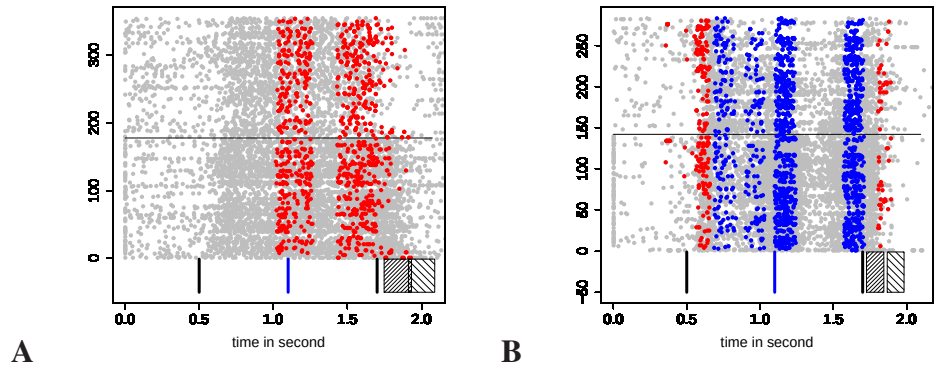

Figure 10: Localisation of the detected coincidences on the spike trains. Graphs A and B are respectively the raster plot of couple 13 and couple 40 with in red the significantly too large coincidences count and in blue the significantly too low coincidences count, the detection being made by the symmetric MTGAUE with $q = 0.05$ and $\delta = 0.02$ s. Bars and boxes on the time axis as in Figure 8.

incidences significantly too large. As already noticed in 4.E and 4.F, the UE detections due to large observed coincidences count are much more precise than the symmetric UE detections (the multiplicity adjusted p-values are low). The detections in 8.C and 8.D that are not found in 9.A and in 9.B corresponds to detection where the observed coincidences count is significantly too low. Those UE detections where the count is too low have sometimes a very large multiplicity adjusted p-value and are therefore not trustworthy. On the other hand, MTGAUE (with $q = 0.05$) only detects by definition periods with small multiplicity adjusted p-values. Those detections concern by definition both situations: significantly too large and too low coincidences count. In particular, on 8.D, those detections are due to a low coincidences count and correspond to $\delta \geq 0.02$ s, range that was not regarded by the previous UE method, indicating a negative interaction with lower bounded delay between both neurons. In order to view the results directly on the data, the rasters associated to the data sets used in Figure 8 for $\delta = 0.02$ s are represented in Figure 10.

5.4 Aggregation of the results obtained from several recording sessions

We run the symmetric MTGAUE method over three different data sets, corresponding to the various probabilities of occurrence of the response signal at 1.1 s. Once again, only trials where the response signal occurs at 1.7 s are treated by the method. The main point is to see the influence of these probabilities on the synchronized activity. In particular because of the very small number of usable trials (in particular in *Data70*) in a given movement direction, we decided to pool all directions together, as already in previous studies but in different ways (Grammont & Riehle, 2003). The trials are consequently not iid but Section 4.2 shows that the MTGAUE detections can be trusted anyway. In the same spirit, we did not discard pairs of neurons whose firing rate is smaller than 7 Hz (Roy et al., 2000; Riehle et al., 2000; Grammont & Riehle, 2003) because even in this

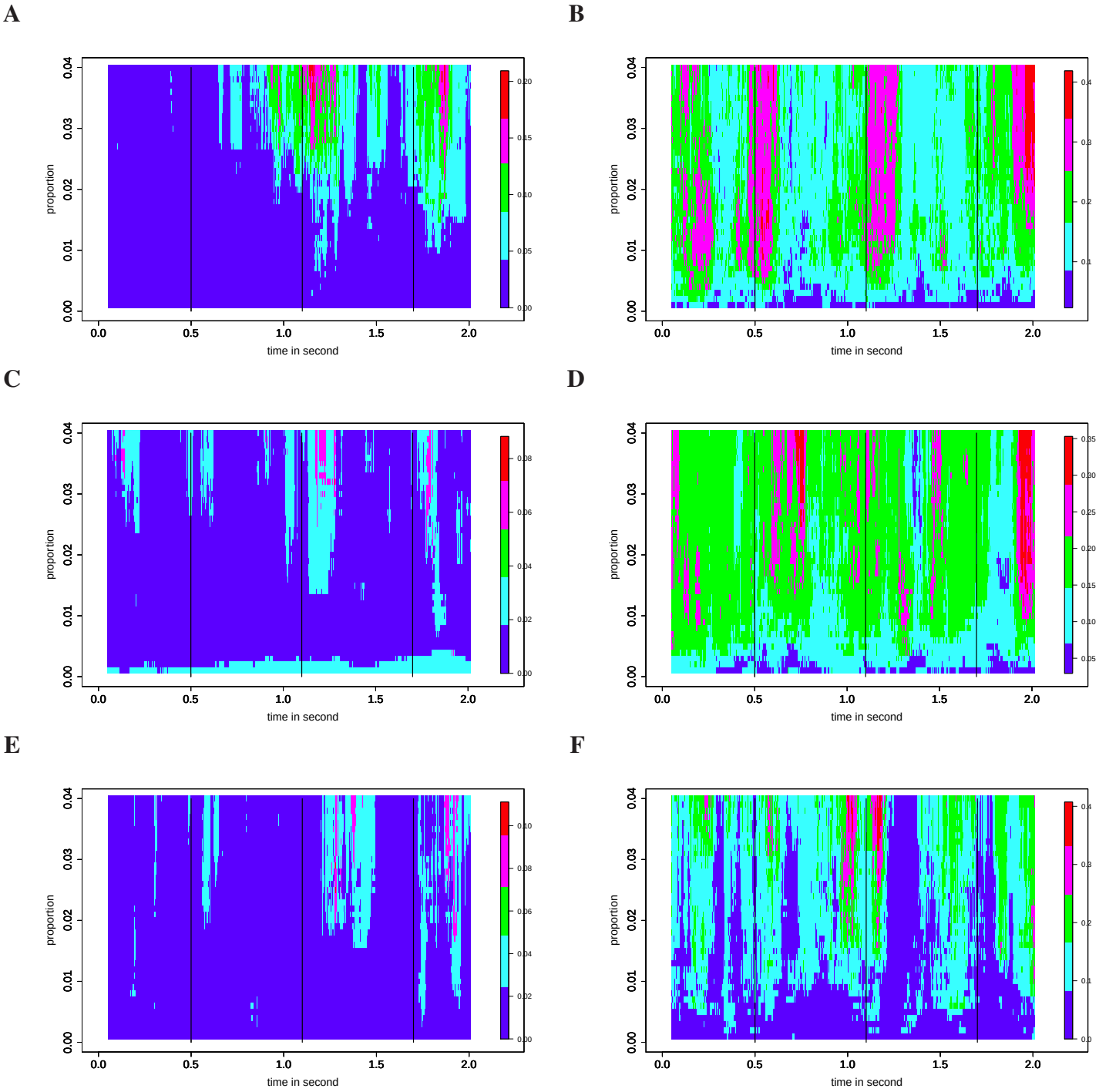

Figure 11: Proportion of pairs of neurons that have been detected by the MTGAUE method (sliding windows of length 0.1s shifted by 0.001 s, $q = 0.05$) through time. Each line represents a different parameter $\delta \in [0, 0.04]$ s. A, C and E corresponds to the proportion of pairs of neurons that have been detected by MTGAUE and for which the coincidences count is significantly too large whereas B, D and F corresponds to the proportion of pairs of neurons that have been detected by MTGAUE and for which the coincidences count is significantly too low. (PS) was presented at 0.5 s (first vertical black line), (RS) at 1.7 s (third vertical black line). Even if those trials are not used in the present average, a response signal at 1.1 s ((ES)=second vertical black line) was also presented in 30% of the cases for A and B (*Data30* - 43 pairs of neurons), in 50% of the cases for C and E (*Data50* - 34 pairs of neurons) and in 70% of the cases in E and F (*Data70* - 27 pairs of neurons).

case the MTGAUE detections can be trusted since MTGAUE is just too conservative (see Figure 3). Hence MTGAUE was performed on each pair of neurons, leading to the detection of windows where this pair is significantly synchronized. Then an aggregation of the detections over all pairs of each data set was performed. The proportion of these significantly synchronized pairs are displayed in Figure 11. As expected the maximal proportions of significantly synchronized pairs of neurons occur in correlation with the occurrence of the different behavioral events of the task (preparatory, expected and response signals). There is a huge proportion of detections (up to 40% of the pairs) that are due to a coincidences count significantly too low, whereas at the most 10% of the pairs have a coincidences count significantly too large.

- Significantly too large coincidences count.

  Comparing Figures 11 A, C and E, it appears that the proportion of pairs that have a significantly too large coincidences count is much more important on *Data30*, showing two maxima, just after the expected signal and just after the response signal. When the probability of the expected signal increases (*Data50* and *Data70*), those maxima still appear but in a lower proportion. In particular, when the response signal at 1.1 s is much more likely than the response signal at 1.7 s (in *Data70*), those maxima are shifted on the right as if the animal continue to expect the signal a longer time because it was more likely. Note finally that most of these maxima appear for large $\delta > 0.02$ s and could not have been seen using the classical UE method.

- Significantly too low coincidences count.

  Comparing Figures 11 B, D and F, it appears that all the three data sets have roughly the same maximal proportion of pairs that have a significantly too low coincidences count. The main difference comes from the localization and the fuzziness of the proportion of pairs. The maxima are large and strong in *Data30* in particular around the preparatory and expected signals. In *Data70*, the only localizations reaching such high scores are just before and after the expected signal. In *Data50*, the signal is fuzzier: in this case, the animal has one chance over two to receive a true signal at the expected signal and seems to prepare himself to both situations by having most of the time a certain fraction of the pairs of neurons with a significantly low coincidences count.

6 Discussion

The UE method (Grün, 1996; Grün et al, 1999, 2010) is a popular method that has shown its ability to detect relevant synchronizations from a neurophysiological point of view (Grammont & Riehle, 1999; Riehle et al., 2000; Grammont & Riehle, 2003; Maldonado et al., 2008). However, several points were not clearly explained and/or controlled at the mathematical level.

- The first main problem is to suppose at the same time that the spike trains and the coincidences counts are Poisson distributed when both spike trains are stationary and independent. If it is just a fairly good approximation in the binning

framework, this distribution is drastically different when the delayed coincidences count is involved (see Section 2.1). In particular, high firing rates or large delays δ tend to make the approximation by a Poisson law worse, fact which was ignored in the neuroscientific literature up to our knowledge.

- The second main source of error is a confusion between the expected coincidences count m_0 under H_0 and its estimate $\hat{m}_0$. Indeed the plug-in step which consists in using the estimate instead of the true parameter may also change drastically the law. Here this phenomenon appears on the variance of the test statistic (see Section 2.2).

- Some edge effects when the delay δ is large are also neglected in the original method, leading to a test statistic that systematically over-estimates the number of coincidences under H_0 for sufficiently large firing rates (larger than 7 Hz) (Roy et al., 2000; Grammont & Riehle, 2003). Summing up all those defects, the UE single test is a very conservative procedure when the rejection is due to a large coincidences count and a very wrongly calibrated procedure when the rejection is due to a low coincidences count (see Section 2.2).

- On the other hand, the Unitary Events method is applied simultaneously over different sliding windows and all the tests with p-value smaller than 0.05 are declared detected. But it is well known in multiple testing theory that this method does not guarantee any control in terms of False Discovery Rate (see Section 3). The UE method by upper value - which combines a very conservative single procedure with this very non conservative way to treat several tests - fortunately leads to quite accurate detections in practice (at least when the window length is 0.1 s, when the number of trials is larger than 19, when the delay δ is smaller than 0.02 s and when the firing rates are larger than 7 Hz). On the contrary, the UE method by lower values can never be trusted because of its very high False Discovery Rate.

The main purpose of this article has been to propose some solutions to these problems in the context of the delayed coincidences count. The binning case is also fully treated in the appendix.

- When both spike trains are independent and stationary Poisson processes, we precisely compute the expected coincidences count and its variance showing that there are indeed edge effects in the formula (see Section 2.1). These effects increase with δ and we show that the distance with respect to a Poisson law increases as a function of both the firing rate and the delay δ.

- We propose a very precise Gaussian approximation of the test statistic when the expected coincidences count is replaced by its estimates leading to statistical tests that are proved to be asymptotically of prescribed level α (see Section 2.2).

- We show by a simulation study that this Gaussian Approximation of the Unitary Events (GAUE) is relevant even for δ in $[0.02, 0.4]$ s when the window length is 0.1 s and even for firing rates as low as 3 Hz and this both for upper and lower

values (see Section 2.2). The test by upper values becomes too conservative for firing rates as small as 3 Hz, but its detections can nevertheless be trusted.

- We combine those tests with a Benjamini and Hochberg procedure leading to the Multiple Testing based on a Gaussian Approximation of the Unitary Events (MTGAUE) on sliding windows. A simulation study shows that the False Discovery Rate is well controlled by the parameter q of the MTGAUE procedure (see Section 4). We show that MTGAUE is able to detect dependence as well as the UE method by upper values in the range of parameters where the UE method is valid, but that the MTGAUE method is also able to detect relevant periods, either when the firing rates are much lower or the delay much larger than the classical range or when the observed coincidences count is significantly too low. Moreover this method seems robust to non iid trials.

- All this factors make MTGAUE able to treat larger data sets where the firing rates, the number of trials and/or the non iid character prevented the use of the UE method (see Section 5). It is also able to look at larger delays δ that were ignored before and at detections due to a significantly too low coincidences count that can not be trusted in the case of the UE method. The patterns of synchronized activity detected by the MTGAUE method are not distributed in a random manner along the behavorial protocol. On the contrary, their localization is relevant with the different events occurring along the time of the behavorial protocol.

In this article, we show that the MTGAUE method is a reliable method for the analysis of the synchronized activity of pairs of neurons. However, one should be able to extend its use to the analysis of more than two neurons at a time. Actually, multielectrode recording techniques are more and more efficient and allow the recording of tens or even hundreds of neurons at the same time. Such techniques provide the possibility to study more directly neuronal assembly (i.e. without requiring the aggregation of data recorded at different moments). Mathematical tools must evolve in the same direction.

On one hand, it is possible to generalize the notion of delayed coincidences to more than 2 neurons as it has already been done for the binned coincidences (Grün et al, 2010). One possibility would be to provide a Gaussian approximation for this case, the multiple testing part remaining unchanged. On the other hand, we think that pure testing procedures do not give fully satisfying answers and that it would be legitimate to provide an estimation of the dependence notably, through the Hawkes model (Krumin et al., 2010). It is already known than this model can easily deal with more than two neurons (Daley & Vere-Jones, 2003; Pernice et al., 2011, 2012). However it is only in a recent work that we have proposed theoretical statistical methods to deal with large numbers of neurons and large potential delays of interaction through Lasso methods (Hansen et al, 2012). In a future work we will adapt this method to deal with non stationary data.

References

Abeles, M., & Gat, I. (2001). Detecting precise firing sequences in experimental data. *Journal of Neuroscience Methods*, *107*, 141–154.

Aertsen, A.M., Gerstein, G.L., Habib, M.K., & Palm, G. (1989). Dynamics of neuronal firing correlation: modulation of "effective connectivity". *Journal of Neurophysiology*, *61*(5), 900–917.

Amarasingham, A., Chen, T.L., Geman, S., Harrison, M.T., & Sheinberg, D.L. (2006). Spike count reliability and the poisson hypothesis. *Journal of Neuroscience*, *26*(3), 801–809.

Barlow, H.B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology?. *Perception*, *1*, 371–394.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, *57*(1), 289–300.

Benjamini, Y. & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, *29*(4), 1165–1188.

Bickel, P.J. & Doksum, K.A. (2000). *Mathematical statistics: Basic ideas and selected topics, vol. 1*. Pearson.

Chicharro, D., Kreuz, T., & Andrzejak, R.G. (2011). What can spike train ditances tell us about the neural code?. *Journal of Neuroscience Methods*, *199*, 146–165.

Daley, D.J., & Vere-Jones, D. (2003). *An introduction to the theory of point processes, vol. I.* Springer.

Diesmann, M., Gewaltig, M.O., & Aertsen, A.M. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature*, *402*, 529–533.

Engel, A.K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *TRENDS in Cognitive Sciences*, *5*(1), 16–25.

Eyherabide, H.G., Rokem, A., Herz, A V.M., & Samengo, I. (2009). Bursts generate a non-reducible spike-pattern code. *Frontiers in Neuroscience*, *3*(1), 8–14.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.

Georgopoulos, A.P., Schwartz, A.B., & Kettner, R.E. (1986). Neuronal population coding of movement direction. *Science*, *233*, 1416–1419.

Gerstein, G.L. (2004). Searching for significance in spatio-temporal firing patterns. *Acta Neurobiologia Experimentalis*, *64*, 203–207.

Gerstein, G.L., & Perkel, D.H. (1969). Simultaneous recorded trains of action potentials: analysis and functional interpretaion. *Science*, *164*, 828–830.

Gerstein, G.L., & Perkel, D.H. (1972). Mutual temporal relationships among neuronal spike trains. *Biophysical Journal*, *12*, 453–473.

Gerstein, G.L., & Aertsen, A.M. (1985). Representation of cooperative firing activity among simultaneously recorded neurons. *Journal of Neurophysiology*, *54*(6), 1513–1528.

Gerstein, G.L., Perkel, D.H., & Dayhoff, J.E. (1985). Cooperative firing activity in simulteneously recorded populations of neurons: detection and measurement. *Journal of Neuroscience*, *5*(4), 881–889.

Glaseer, E.M., & Ruchkin, D.S. (1976). *Principal of neurobiological signal analysis*. New York: Academic Press.

Goedeke, S., & Diesmann, M. (2008). The mechanism of synchronization in feed-forward neuronal networks. *New Journal of Physics*, *10*, 015007.

Goeman, J.J., & Solari, A. (2011). Multiple testing for exploratory research. *Statistical Science*, *26*(4), 584–597.

Golomb, D., & Hansel, D. (2000). The number of synaptic inputs and the synchrony of large sparse neuronal networks. *Neural Computation*, *12*(5), 1095–1139.

Grammont, F., & Riehle, A. (1999). Precise spike synchronization in monkey motor cortex involved in preparation for movement. *Experimental Brain Research*, *128*(1-2), 118–122.

Grammont, F., & Riehle, A. (2003). Spike synchronisation and firing rate in a population of motor cortical neurons in relation to movement direction ans reaction time. *Biological Cybernetics*, *88*, 360–373.

Grün, S. (1996). *Unitary joint-events in multiple-neuron spiking activity: Detection, significance and interpretation.* Thun: Verlag Harri Deutsch.

Grün, S. (2009). Data-driven significance estimation for precise spike correlation. *Journal of Neurophysiology*, *101*, 1126–1140.

Grün, S., Diesman, M., & Aertsen, A.M. (2001a). Unitary events in multiple single-neuron spiking activity: I. Detection ans Significance. *Neural Computation*, *14*, 43–80.

Grün, S., Diesman, M., & Aertsen, A.M. (2001b). Unitary events in multiple single-neuron spiking activity: II. Nonstationary Data. *Neural Computation*, *14*, 81–119.

Grün, S., Diesmann, M., Grammont, F., Riehle, A., & Aertsen, A.M. (1999). Detecting unitary events without discretization of time. *Journal of Neuroscience Methods*, *93*, 67–79.

Grün, S., Diesmann, M., & Aertsen, A.M.(2010). Unitary Events Analysis. *In Analysis of Parallel Spike Trains*, Grün, S., & Rotter, S., Springer Series in Computational Neuroscience.

Hansen, N.R., Reynaud-Bouret, P. & Rivoirard, V. (2012). Lasso and probabilistic inequalities for multivariate point processes. *http://arxiv.org/abs/1208.0570*

Hawkes, A. (1971). Point spectra of some mutually exciting point processes. *Journal of the Royal Society. Series B*, *33*(3), 438–443.

Hebb, D. O. (1949). *Organization of behavior; a neuropsychological theory*. New York: John Wiley & Sons.

Heinzle, J., König, P., & Salazar R.F. (2007). Modulation of synchrony without changes in firing rate. *Cognitive Neurodynamics*, *1*, 225–235.

Hochberg, Y., & Tamhame, A. (1987). *Multiple comparison procedures*. New York: Wiley.

Hogg, R.V., & Tanis, E.A. (2009). *Probability and statistical inference*. Pearson.

Holm, S. (1979). A simple sequentially rejection multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Konishi, M., Takahashi, T.T., Wagner, H., Sullivan, W.E., & Carr, C.E. (1988). Neurophysiological and anatomical substrates of sound localization in the owl. *In Auditory Function*, Edelman, G.M., Gall, W.E. and Cowan, W.M., Wiley, New-York.

Kreuz, T., Hass, J.S., Morelli, A., Abarbanel, H.D.I., & Politi, A. (2007). Measuring spike train synchrony. *Journal of Neuroscience Methods*, *165*, 151–161.

Krumin, M., Reutsky, I., & Shoham, S. (2010). Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input. *Frontiers in Computational Neuroscience*, *4*, article 147.

Kumar, A., Rotter, S., & Aertsen, A.M. (2010). Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nature Reviews Neuroscience*, *11*, 615–627.

Lestienne, R. (1996). Determination of the precision of spike timing in the visual cortex of anaesthetised cats. *Biological Cybernetics*, *74*, 55–61.

Lestienne, R. (2001). Spike timing, synchronization and information processing on the sensory side of the central nervous system. *Progress in Neurobiology*, *65*, 545–591.

Mainen, Z.F., & Sejnowski, T.J. (1995). Reliability of spike timing in neocortical neurons. *Science*, *268*, 1503–1506.

Maimon, G. & Assad, J.A. (2009). Beyond Poisson: increased spike-time regularity across primate parietal cortex. *Neuron*, *62*(3), 426–440.

Maldonado, P.E., Friedman-Hill, S., & Gray, C.M. (2000). Dynamics of striate cortical activity in the alert macaque: II. Fast time scale synchronization. *Cerebral Cortex*, *10*, 1117–1131.

Maldonado, P.E., Babul, C., Singer, W., Rodriguez, E., Berger, D. & Grün, S. (2008). Synchronization of neuronal response in primary visual cortex of monkeys viewing natural images. *Journal of Neurophysiology*, *100*, 1523–1532.

Masuda, N., & Aihara, K. (2007). Dual coding hypotheses for neural information representation. *Mathematical Biosciences*, *207*, 312–321.

Oram, M.W., Wiener, M.C., Lestienne, R., & Richmond, B.J. (1999). Stochastic nature of precisely timed spike patterns in visual system neuronal responses. *Journal of Neurophysiology*, *81*, 3021–3033.

Palm, G. (1990). Cell assemblies as a guideline for brain research. *Concepts in Neuroscience*, *1*, 133–148.

Palm, G., Aertsen A.M., & Gerstein, G.L. (1988). On the significance of correlations among neuronal spike trains. *Biological Cybernetics*, *59*, 1–11.

Perkel, D.H., Gernstein, G.L., & Moore, G.P. (1967). Neuronal spike trains and stochastic point processes. *Biophysical Journal*, *7*, 419–440.

Pernice, V., Staude, B., Cardanobile, S., & Rotter, S. (2011). How structure determines correlations in neuronal networks. *PLoS Computational Biology*, 7:e1002059.

Pernice, V., Staude, B., Cardanobile, S., & Rotter, S. Recurrent interactions in spiking networks with arbitrary topology. *Physical review E, Statistical, nonlinear, and soft matter physics*, 85:031916.

Pipa, G., Diesmann, M., & Grün, S. (2003). Significance of joint-spike events based on trial-shuffling by efficient combinatorial methods. *Complexity*, *8*(4), 1–8.

Pipa, G., & Grün, S. (2003). Non-parametric significance estimation of joint-spike events by shuffling and resampling. *Neurocomputing*, *52-54*, 31–37.

Pouget, A., Dayan, P., & Zemel, R.S. (2000). Information processing with population codes. *Nature Review Neuroscience*, *1*(2), 125–132.

Prescott, S.A., Ratté, S., De Koninck, Y., & Sejnowski, T.J. (2008). Pyramidal neurons switch from integrators in vitro to resonators under in vivo-like conditions. *Journal of Neurophysiology*, *100*, 3030–3042.

Riehle, A., Grammont, F., Diesmann, M. & Grün, S. (2000). Dynamical changes and temporal precision of synchronised spiking activity in monkey motor cortex during movement preparation. *Journal of Physiology*, *94*, 569–582.

Riehle, A., Grammont, F., & MacKay, A. (2006). Cancellation of a planned movement in monkey motor cortex. *Neuroreport*, *17*(3), 281–285.

Riehle, A., Grün, S., Diesmann, M., & Aertsen, A.M. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science*, *278*, 1950–1953.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.

Romano, J.P., & Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, *100*(469), 94–108.

Roy, A., Steinmetz, P.N., & Niebur, E. (2000). Rate Limitations of Unitary Event Analysis *Neural Computations*, *12*(9), 2063–2082.

Sakurai, Y. (1999). How do cell assemblies encode information in the brain?. *Neuroscience and Biobehavioral Reviews*, *23*, 785–796.

Schreiber, S., Fellous, J.M., Whitmer, D., Tiesinga, P., & Sejnowski, T.J. (2003). A new correlation-based measure of spike timing reliability. *Neurocomputing*, *52-54*, 925–931.

Shinomoto, S. (2010). Estimating the Firing Rate. *In Analysis of Parallel Spike Trains*, Grün, S., & Rotter, S., Springer Series in Computational Neuroscience.

Singer, W., & Gray, C.M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, *18*, 555–586.

Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations?. *Neuron*, *24*, 49–65.

Softky, W.R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, *13*(1), 334–350.

Tiesinga, P.H.E, & Sejnowski, T.J. (2004). Rapid temporal modulation of synchrony by competition in cortical interneuron networks. *Neural Computation*, *16*, 251–275.

Victor, J.D., & Purpura, K.P. (1996). Nature and precision of temporal coding in visual cortex: a metric-space analysis. *Journal of Neurophysiology*, *76*(2), 1310–1326.

Victor, J.D., & Purpura, K.P. (1997). Sensory coding in cortical neurons; recent results and speculations. *Annals of the New York Academy of Sciences*, *835*, 330–352.

Vlachos, I., Aertsen, A.M., & Kumar, A. (2012). Deyond statistical significance: implications of network structure on neuronal activity. *PLOS Computational Biology*, *8*(1), e1002311.

Von der Malsburg, C. (1981). *The correlation theory of brain function*. Internal Report. Göttingen: Max-Planck-Institute for Biophysical chemistry, Dept. Neurobiology.

A Binning framework

A.1 Notations:

- For $l \in \{1, 2\}$, $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, M\}$, let $N_l^{(j)}(I_i)$ be the number of points of N_l on I_i for the trial j.

- For $i \in \{1, \dots, k\}$, let $Z_i = \mathbf{1}_{\{N_1(I_i)\geq 1,\ N_2(I_i)\geq 1\}}$ be the binned coincidences count on I_i for the couple (N_1, N_2).
- For $i \in \{1, \dots, k\}$ and $j \in \{1, \dots, M\}$, let $Z_i^{(j)} = \mathbf{1}_{\{N_1^{(j)}(I_i)\geq 1,\ N_2^{(j)}(I_i)\geq 1\}}$ be the binned coincidences count on I_i for the couple $(N_1^{(j)}, N_2^{(j)})$, that is, $Z_i^{(1)}, \dots, Z_i^{(M)}$ are M realizations of Z_i.
- For $j \in \{1, \dots, M\}$, let $Y^{(j)} = \sum_{i=1}^{k} Z_i^{(j)}$ be the binned coincidences count on $[a, b]$ for the couple $(N_1^{(j)}, N_2^{(j)})$, that is, $Y^{(1)}, \dots, Y^{(M)}$ are M realizations of Y,where Y is defined in Definition 1.
- Let $\bar{Z} = \frac{1}{Mk} \sum_{j=1}^{M} Y^{(j)} = \frac{1}{Mk} \sum_{j=1}^{M} \sum_{i=1}^{k} Z_i^{(j)}$ be the average coincidences count on a bin. In particular, we have $\bar{Z} = k\bar{m}$, where $\bar{m}$ is defined by Equation (1).

A.2 Proof of Proposition 1

Proof. Since N_i is a Poisson process, we have:

- $\forall \{i, s\},\ N_i(I_s) \sim \mathcal{P}(\lambda_i \delta)$
- $\forall i$, the random variables $\{N_i(I_s)\}_s$ are independent.

Thus $\{Z_i\}_i$ are independent and identically distributed random variables whose distribution is a Bernoulli distribution with parameter $p = P(Z_1 = 1)$.
Thus the distribution of Y, the binned coincidences count on $[a, b]$, is a binomial distribution with parameters k and p.
Let us compute the expression of p.
Under H_0, we have:

$$\begin{aligned} p &= P(N_1(I_i) \geq 1,\ N_2(I_i) \geq 1) \\ &= P(N_1(I_i) \geq 1) P(N_2(I_i) \geq 1) \\ &= (1 - P(N_1(I_i) = 0))\,(1 - P(N_2(I_i) = 0)) \\ &= (1 - \exp(-\lambda_1 \delta))(1 - \exp(-\lambda_2 \delta)) \end{aligned}$$

Thus $m_0 = \mathbb{E}(Y) = kp$.
□

A.3 Detail on the correction

Theorem 3. *Under Assumptions 1, 2 and H_0, the following convergence holds:*

$$\sqrt{Mk}\left(\bar{Z} - \hat{p}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

where

$$\hat{p} = (1 - \exp(-\hat{\lambda}_1 \delta))(1 - \exp(-\hat{\lambda}_2 \delta)),$$

$$\sigma^2 = p(1-p) - \lambda_1 \delta \exp(-2\lambda_1 \delta)\left[1 - \exp(-\lambda_2 \delta)\right]^2 - \lambda_2 \delta \exp(-2\lambda_2 \delta)\left[1 - \exp(-\lambda_1 \delta)\right]^2,$$

and where the $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are defined by Equation (2). Moreover σ^2 can be estimated by

$$\hat{\sigma}^2 = \hat{p}(1-\hat{p}) - \hat{\lambda}_1\delta\exp(-\hat{\lambda}_1\delta)^2(1-\exp(-\hat{\lambda}_2\delta))^2 - \hat{\lambda}_2\delta\exp(-\hat{\lambda}_2\delta)^2(1-\exp(-\hat{\lambda}_1\delta))^2$$

and

$$\sqrt{Mk}\,\frac{\bar{Z}-\hat{p}}{\sqrt{\hat{\sigma}^2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Proof. First we apply the Vectorial Central Limit Theorem (see (Bickel & Doksum , 2000))

$$(Mk)^{-1/2}\sum_{i=1}^{k}\sum_{j=1}^{M}\left[\begin{pmatrix} Z_i^{(j)} \\ N_1^{(j)}(I_i) \\ N_2^{(j)}(I_i)\end{pmatrix} - \begin{pmatrix} p \\ \lambda_1\delta \\ \lambda_2\delta\end{pmatrix}\right] \xrightarrow{\mathcal{L}} \mathcal{N}_3(0,\Gamma),$$

where Γ is the corresponding covariance matrix, i.e.

$$\Gamma = \begin{pmatrix} p(1-p) & \lambda_1\delta\exp(-\lambda_1\delta)(1-\exp(-\lambda_2\delta)) & \lambda_2\delta\exp(-\lambda_2\delta)(1-\exp(-\lambda_1\delta)) \\ \lambda_1\delta\exp(-\lambda_1\delta)(1-\exp(-\lambda_2\delta)) & \lambda_1\delta & 0 \\ \lambda_2\delta\exp(-\lambda_2\delta)(1-\exp(-\lambda_1\delta)) & 0 & \lambda_2\delta \end{pmatrix}.$$

Indeed, we have:

$$\begin{aligned}
cov(Z_i^{(j)}, N_1^{(j)}(I_i)) &= \mathbb{E}(Z_i^{(j)}N_1^{(j)}(I_i)) - \mathbb{E}(Z_i^{(j)})\mathbb{E}(N_1^{(j)}(I_i)) \\
&= \mathbb{E}(\mathbf{1}_{N_1^{(j)}(I_i)\geq 1, N_2^{(j)}(I_i)\geq 1}N_1^{(j)}(I_i)) - \mathbb{E}(Z_i^{(j)})\mathbb{E}(N_1^{(j)}(I_i)) \\
&= \mathbb{E}(\mathbf{1}_{N_1^{(j)}(I_i)\geq 1}\mathbf{1}_{N_2^{(j)}(I_i)\geq 1}N_1^{(j)}(I_i)) - \mathbb{E}(Z_i^{(j)})\mathbb{E}(N_1^{(j)}(I_i)) \\
&= \mathbb{E}(\mathbb{E}(\mathbf{1}_{N_2^{(j)}(I_i)\geq 1}N_1^{(j)}(I_i)|N_2)) - \mathbb{E}(Z_i^{(j)})\mathbb{E}(N_1^{(j)}(I_i)) \\
&= P(N_2^{(j)}(I_i)\geq 1)\mathbb{E}(N_1^{(j)}(I_i)) - \mathbb{E}(Z_i^{(j)})\mathbb{E}(N_1^{(j)}(I_i)) \\
&= (1-P(N_2^{(j)}(I_i)=0))\lambda_1\delta - p\lambda_1\delta
\end{aligned}$$

By assumption on N_2, we have $N_2^{(j)}(I_i) \sim \mathcal{P}(\lambda_2\delta)$, thus:

$$\begin{aligned}
cov(Z_i^{(j)}, N_1^{(j)}(I_i)) &= (1-\exp(-\lambda_2\delta))\lambda_1\delta - p\lambda_1\delta \\
&= (1-\exp(-\lambda_2\delta))\lambda_1\delta - (1-\exp(-\lambda_1\delta))(1-\exp(-\lambda_2\delta))\lambda_1\delta \\
&= (1-\exp(-\lambda_2\delta))\lambda_1\delta(1-1+\exp(-\lambda_1\delta)) \\
&= \lambda_1\delta\exp(-\lambda_1\delta))(1-\exp(-\lambda_2\delta))
\end{aligned}$$

Next, we can rewrite

$$\begin{aligned}
\sqrt{Mk}(\bar{Z}-\hat{p}) &= \sqrt{Mk}\left[g\left(\frac{1}{Mk}\sum_{i=1}^{k}\sum_{j=1}^{M}Z_i^{(j)}, \frac{1}{Mk}\sum_{i=1}^{k}\sum_{j=1}^{M}N_1^{(j)}(I_i), \frac{1}{Mk}\sum_{i=1}^{k}\sum_{j=1}^{M}N_2^{(j)}(I_i)\right)\right. \\
&\qquad \left. -g(\mathbb{E}(Z_1), \lambda_1\delta, \lambda_2\delta)\right],
\end{aligned}$$

with $g(x,u,v) = x-(1-\exp(-u))(1-\exp(-v))$. Therefore the Delta method (see (Bickel & Doksum , 2000)) gives that

$$\sqrt{Mk}\left(\bar{Z}-\hat{p}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, D'\Gamma D),$$

where D is the gradient of g in $(\mathbb{E}(Z_1), \lambda_1(b-a), \lambda_2(b-a))$. That is

$$D = \begin{pmatrix} 1 \\ -\exp(-\lambda_1\delta)(1-\exp(-\lambda_2\delta)) \\ -\exp(-\lambda_2\delta)(1-\exp(-\lambda_1\delta)) \end{pmatrix}.$$

We recognise that $\sigma^2 = D'\Gamma D$ and the last result is just a classical application of Slustky's lemma (see (Bickel & Doksum , 2000)). □

B The delayed coincidences count

B.1 Proof of Theorem 1

Proof. Since N_1 et N_2 are now independent homogeneous Poisson processes with intensity λ_1 and λ_2, one has that

$$\mathbb{E}(X) = \mathbb{E}\left(\int_a^b \int_a^b 1_{|x-y|\leq\delta} dN_1(x) dN_2(y)\right)$$

One can prove (see (Daley & Vere-Jones, 2003)), that

$$\mathbb{E}(X) = \lambda_1\lambda_2 \int_a^b \int_a^b 1_{|x-y|\leq\delta} dxdy.$$

But

$$\begin{aligned}
\int_a^b \int_a^b 1_{|x-y|\leq\delta} dxdy &= \int_a^b [\min(b, x+\delta) - \max(a, x-\delta)]dx \\
&= \int_a^{a+\delta} [x+\delta-a]dx + \int_{a+\delta}^{b-\delta} 2\delta dx + \int_{b-\delta}^b [b+\delta-x]dx \\
&= \frac{(a+\delta)^2 - a^2}{2} + \delta(\delta-a) + 2\delta[(b-a)-2\delta] + \delta(b+\delta) - \frac{b^2-(b-\delta)^2}{2} \\
&= 2\delta(b-a) - \delta^2
\end{aligned}$$

since $a+\delta \leq b-\delta$. Similarly, we can compute $\mathbb{E}(X^2)$. If $Diag = \{(x,y) \mid x = y \in [a,b]\}$ and $[a,b]^{(2)} = [a,b]^2 \backslash Diag$, one can decompose

$$\begin{aligned}
\mathbb{E}(X^2) = \mathbb{E}&\left(\int_{[a,b]^4} 1_{|x-y|\leq\delta} 1_{|t-u|\leq\delta} dN_1(x)dN_1(t)dN_2(y)dN_2(u)\right) \\
&= \mathbb{E}\left(\int_{([a,b]^{(2)})^2} 1_{|x-y|\leq\delta} 1_{|t-u|\leq\delta} dN_1(x)dN_1(t)dN_2(y)dN_2(u)\right) \\
&+ \mathbb{E}\left(\int_{[a,b]^{(2)}\times Diag} 1_{|x-y|\leq\delta} 1_{|t-u|\leq\delta} dN_1(x)dN_1(t)dN_2(y)dN_2(u)\right) + \\
&+ \mathbb{E}\left(\int_{Diag\times[a,b]^{(2)}} 1_{|x-y|\leq\delta} 1_{|t-u|\leq\delta} dN_1(x)dN_1(t)dN_2(y)dN_2(u)\right) \\
&+ \mathbb{E}\left(\int_{Diag^2} 1_{|x-y|\leq\delta} 1_{|t-u|\leq\delta} dN_1(x)dN_1(t)dN_2(y)dN_2(u)\right)
\end{aligned}$$

This leads by classical properties of the moment measure of Poisson processes (see (Daley & Vere-Jones, 2003)) to

$$\begin{aligned}\mathbb{E}(X^2) &= \mathbb{E}(X)^2 + \lambda_1^2\lambda_2 \int_{[a,b]^3} 1_{|x-y|\leq\delta} 1_{|t-y|\leq\delta} dxdydt + \\ &\quad +\lambda_1\lambda_2^2 \int_{[a,b]^3} 1_{|x-y|\leq\delta} 1_{|x-u|\leq\delta} dxdydu + \lambda_1\lambda_2 \int_a^b \int_a^b 1_{|x-y|\leq\delta} dxdy\end{aligned}$$

Hence

$$\mathrm{Var}(X) = \lambda_1\lambda_2[2\delta(b-a) - \delta^2] + [\lambda_1^2\lambda_2 + \lambda_1\lambda_2^2] \int_a^b \left(\int_a^b 1_{|x-y|\leq\delta} dy \right)^2 dx.$$

It remains to compute as before

$$\begin{aligned}\int_a^b \left(\int_a^b 1_{|x-y|\leq\delta} dx \right)^2 dy &= \int_a^b [\min(b, x+\delta) - \max(a, x-\delta)]^2 dx \\ &= \int_a^{a+\delta} [x+\delta-a]^2 dx + \int_{a+\delta}^{b-\delta} [2\delta]^2 dx + \int_{b-\delta}^b [b+\delta-x]^2 dx \\ &= \frac{(a+\delta)^3 - a^3}{3} + [(a+\delta)^2 - a^2](\delta - a) + \delta(\delta-a)^2 \\ &\quad +[2\delta]^2[(b-a) - 2\delta] + \\ &\quad +\frac{b^3 - (b-\delta)^3}{3} - [b^2 - (b-\delta)^2](b+\delta) + \delta(b+\delta)^2 \\ &= 4\delta^2(b-a) - \frac{10}{3}\delta^3\end{aligned}$$

□

B.2 Proof of Theorem 2

Proof. Let $X^i = \int_{[a,b]^2} \mathbf{1}_{|x-y|\leq\delta} dN_1^{(i)}(x) dN_2^{(i)}(y)$ the coincidences count with delay δ on $[a,b]$ for the couple $(N_1^{(i)}, N_2^{(i)})$.
First we apply the Vectorial Central Limit Theorem

$$M^{-1/2} \sum_{i=1}^M \left[\begin{pmatrix} X^i \\ N_1^{(i)}([a,b]) \\ N_2^{(i)}([a,b]) \end{pmatrix} - \begin{pmatrix} \lambda_1\lambda_2[2\delta(b-a) - \delta^2] \\ \lambda_1(b-a) \\ \lambda_2(b-a) \end{pmatrix} \right] \xrightarrow{\mathcal{L}} \mathcal{N}_3(0, \Gamma),$$

where Γ is the corresponding covariance matrix, i.e.

$$\Gamma = \begin{pmatrix} \lambda_1\lambda_2[2\delta(b-a) - \delta^2] + [\lambda_1^2\lambda_2 + \lambda_1\lambda_2^2][4\delta^2(b-a) - \frac{10}{3}\delta^3] & \lambda_1\lambda_2[2\delta(b-a) - \delta^2] & \lambda_1\lambda_2[2\delta(b-a) - \delta^2] \\ \lambda_1\lambda_2[2\delta(b-a) - \delta^2] & \lambda_1(b-a) & 0 \\ \lambda_1\lambda_2[2\delta(b-a) - \delta^2] & 0 & \lambda_2(b-a) \end{pmatrix}.$$

This matrix is obtained by the previous computations, the fact that N_1 is independent of N_2 and the following fact,

$$\begin{aligned}
\mathbb{E}(XN_1([a,b])) &= \mathbb{E}(\int_{[a,b]^3} 1_{|x-y|\leq\delta} dN_1(x)dN_1(t)dN_2(y)) \\
&= \mathbb{E}(\int_{[a,b]^(2)\times[a,b]} 1_{|x-y|\leq\delta} dN_1(x)dN_1(t)dN_2(y)) \\
&\quad +\mathbb{E}(\int_{Diag\times[a,b]} 1_{|x-y|\leq\delta} dN_1(x)dN_1(t)dN_2(y)) \\
&= \lambda_1^2\lambda_2 \int_{[a,b]^3} 1_{|x-y|\leq\delta} dxdtdy + \lambda_1\lambda_2 \int_{[a,b]^2} 1_{|x-y|\leq\delta} dxdy \\
&= \mathbb{E}(X)\mathbb{E}(N_1([a,b])) + \lambda_1\lambda_2[2\delta(b-a) - \delta^2]
\end{aligned}$$

Next we can rewrite

$$\begin{aligned}
\sqrt{M}\,(\bar{m} - \hat{m}_0) &= \sqrt{M}\left[g\left(\frac{1}{M}\sum_i X^{(i)}, \frac{1}{M}\sum_i N_1^{(i)}([a,b]), \frac{1}{M}\sum_i N_2^{(i)}([a,b]) \right) \right. \\
&\quad \left. -g(\mathbb{E}(X^{(1)}), \lambda_1(b-a), \lambda_2(b-a)) \right],
\end{aligned}$$

where $\bar{m}$ and $\hat{m}_0$ are respectively defined by Equations (5) and (6), and with $g(x,u,v) = x - uv[2\delta(b-a) - \delta^2](b-a)^{-2}$. Therefore the Delta method gives that

$$\sqrt{M}\,(\bar{m} - \hat{m}_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, D'\Gamma D),$$

where D is the gradient of g in $(\mathbb{E}(X^{(1)}), \lambda_1(b-a), \lambda_2(b-a))$. That is

$$D = \begin{pmatrix} 1 \\ -\lambda_2[2\delta(b-a) - \delta^2](b-a)^{-1} \\ -\lambda_1[2\delta(b-a) - \delta^2](b-a)^{-1} \end{pmatrix}.$$

We recognise that $\sigma^2 = D'\Gamma D$ and the last result is just a classical application of Slustky's lemma. □