Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 16(6): 6990–7008.
DOI: 10.3934/mbe.2019351
Received: 28 February 2019
Accepted: 18 July 2019
Published: 31 July 2019

Research article

Detection and evaluation of bursts in terms of novelty and surprise

Junji Ito[1,2,†,*], **Emanuele Lucrezia**[1,2,†], **Günther Palm**[1,3,†] **and Sonja Grün**[1,2,4]

[1] Institute for Neuroscience and Medicine (INM-6), Jülich Research Centre, 52425 Jülich, Germany

[2] Institute for Advanced Simulation (IAS-6) and Jara Brain Institute for Brain Structure and Function (INM-10), Jülich Research Centre, 52425 Jülich, Germany

[3] Institute of Neural Information Processing, University of Ulm, James-Franck-Ring, 89081 Ulm, Germany

[4] Theoretical Systems Neurobiology, RWTH Aachen University, Worringerweg 3, 52056 Aachen, Germany

† These authors contributed equally to this work.

* **Correspondence:** Email: j.ito@fz-juelich.de; Tel: +492461614748; Fax: +492461619460.

Abstract: The detection of bursts and also of response onsets is often of relevance in understanding neurophysiological data, but the detection of these events is not a trivial task. We build on a method that was originally designed for burst detection using the so-called burst surprise as a measure. We extend this method and provide a proper significance measure. Our method consists of two stages. In the first stage we model the neuron's interspike interval (ISI) distribution and make an i.i.d. assumption to formulate our null hypothesis. In addition we define a set of 'surprising' events that signify deviations from the null hypothesis in the direction of 'burstiness'. Here the so-called (strict) burst novelty is used to measure the size of this deviation. In the second stage we determine the significance of this deviation. The (strict) burst surprise is used to measure the significance, since it is the negative logarithm of the significance probability. After showing the consequences of a non-proper null hypothesis on burst detection performance, we apply the method to experimental data. For this application the data are divided into a period for parameter estimation to express a proper null hypothesis (model of the ISI distribution), and the rest of the data is analyzed by using that null hypothesis. We find that assuming a Poisson process for experimental spike data from motor cortex is rarely a proper null hypothesis, because these data tend to fire more regularly and thus a gamma process is more appropriate. We show that our burst detection method can be used for rate change onset detection, because a deviation from the null hypothesis detected by (strict) burst novelty also covers an increase of firing rate.

Keywords: burst detection; interspike intervals; significance; gamma process; response onset detection

1. Introduction

In extracellular recordings of single or multiple neurons in the mammalian cortex, clearly the most prominent feature is the occurrence of action potentials, or spikes. These recordings are usually called *spike trains* and modeled stochastically as processes of discrete events in continuous time. Another observation, which is also quite ubiquitous, is the tendency of neurons to produce 'bursts' of spikes, i.e., a number of spikes within a very short time period. In other words, some neurons produce more bursts than one would expect on the basis of independent 'random' repetitive firing. Thus, here the event of a 'burst' signifies the occurrence of comparatively many spikes in a comparatively short interval of time. One could also describe it as a short temporary increase of the neuron's spike rate.*

The occurrence of bursts in spike trains has been investigated by many authors. Both, potential biophysical mechanisms for the generation of bursts [1–3] and possible functional purposes of bursts in information transmission and learning have been discussed [1–5]. Depending on these assumptions or research hypotheses, several slightly different 'definitions' of bursts have been used and correspondingly different methods of burst detection have been proposed [1, 6–12]. They all share the basic intuition given above. Although the original methods for burst detection were meant for single spike trains, they are sometimes also applied to multi-unit recordings [13], which may require additional considerations and parameters. There are a few reviews [9, 14] that compare the various burst-detection methods, for example in terms the number of parameters they need or the type of information from the spike train they require for their definition.

1.1. Burst detection

From a statistical point of view the characterization of a burst points towards a significance test, i.e., a quantification of the deviation of the event from a naive null hypothesis of 'independent firing' in the direction of 'burstiness'. The measurement of 'burstiness', however, is a bit problematic because it involves two parameters: one could fix a time interval τ as a parameter and consider the number n of spikes in this interval as a test statistic (the greater the number n, the more bursty is the event), or one could fix a number n of spikes as a parameter and consider the time interval τ between the first and the last spike in sequence of n spikes (the shorter the interval τ, the more bursty is the event). Both approaches have been taken [6, 10], but the problem is that one has to examine the data first in order to make a reasonable choice of the parameter τ or n. Furthermore, because of non-stationarity of neural responses and of individual differences between neurons, it may in practice be allowed to examine only one particular neuron under one particular experimental condition, i.e., the same data that are subsequently evaluated statistically for burst detection. Another issue is that the comparison of the 'degree of burstiness' would be difficult if the degrees are obtained using different combinations of observed values of τ and n.

If one wants to analyze bursts for their properties, such as their duration or their spike frequency, more complex definitions are needed which do not fix one of these parameters in advance. One of the first methods of this type was introduced by Legendy [15]. He identified the significant events by minimizing the Poisson probability $\mathbb{P}[T_{k+n} - T_k \leq \tau]$ (where T_k is the time of the kth spike) over both n and τ, thus using this probability for trading off between τ and n. The logarithm of this probability he

*A problem with this description is that the determination of a spike rate usually requires temporal averaging across a reasonably long time interval that may be longer than the duration of the burst.

called 'burst surprise'. Legendy's surprise can be used to compare the degree of burstiness of different events of the type $[T_{k+n} - T_k \leq \tau]$ and it provides a useful and simple method of burst detection. Clearly one can also use any other theoretical or empirical distribution $\mathbb{P}$ for the computation of Legendy's surprise, but the Poisson probability is easy to calculate and needs only one parameter to be estimated from the data: the neuron's firing rate.

In this paper we provide a larger variety of burst detection methods, by introducing different probability distributions for the surprise computation, derived from various interspike interval (ISI) distributions other than the exponential distribution, which is assumed in Legendy's surprise. We also introduce different ways of minimizing the probability $\mathbb{P}[T_{k+n} - T_k \leq \tau]$, resulting in a different definition of novelty and surprise, by imposing stricter criteria on bursts than Legendy did.

1.2. Burst significance

It is tempting to interpret the burst surprise as a measure of the significance of burst occurrence, but this is a wrong interpretation (which has occasionally been made in previous studies) because the burst surprise as defined by Legendy does not represent a proper statistical significance of a burst event. Since the burst surprise is obtained through minimization of the Poisson probability $\mathbb{P}[T_{k+n} - T_k \leq \tau]$ over n and τ, the obtained probability is obviously biased toward smaller values. This problem can be solved in a principled way by using Legendy's surprise as the test statistics and determining its survival probabilities [16]. It turned out that this type of problem and the corresponding solution are also seen in many other scenarios (e.g., detection of 'pauses' in single spike trains, detection of spike coincidences in multi-unit spike trains [12, 17–19], and detection of spatio-temporal spike patterns [20, 21]), which asks for a broader statistical theory that could then also be related to information theory [22]. In the context of this broader theory Legendy's measure is renamed as 'novelty', saving the term 'surprise' for the proper significance measure to be used for statistical significance tests.

Several authors have suggested modifications of Legendy's method [9, 12, 23, 24], but none of them properly dealt with this transition from 'novelty' to 'surprise'. A comparison of the different new methods attempted in these papers remains somewhat debatable, because in each method the number of bursts detected depends strongly on a threshold value that cannot easily be compared among different methods. In fact, for a proper comparison one would need exactly the transition from 'novelty' to 'surprise' (and statistical significance) mentioned before.

Recently we realized that this transition, which was first described for the exponential ISI distribution (corresponding to Poisson spike train), can be done in the same way for any renewal process with an arbitrary ISI distribution. This resulted in a generalization of Legendy's method and its statistical improvement to arbitrary ISI distributions. In this paper we describe how this transition is carried out in detail.

1.3. Organization of the paper

In the following Section (Section 2), we first define 'burst novelty' and 'burst surprise' formally, and then describe how the surprise is derived from the novelty in practice. Based on these measures, we propose a method for detecting bursts in a given spike train. We extend the earlier method assuming the exponential ISI distribution to the one assuming arbitrary renewal processes allowing to incorporate more knowledge about the neuron's ISI-distribution into the null hypothesis.

Then in Section 3 we illustrate how the choice of the null hypothesis affects the detection of a burst by application of the method to artificial spike trains with gamma ISI distributions. We also show an application of these ideas to a set of 156 spike trains recorded from macaque motor cortex [25] and discuss the possibility to use it to determine the latencies of evoked activities of the single units in single trials.

Finally, in Section 4 we conclude the paper with a brief discussion, mentioning in particular the importance of the null hypothesis for this approach that is 'non-Bayesian' in spirit, because it does not involve any statistical model of burst generation, as e.g., in [26], or of the total network that generates the one (or several) observed spike train(s).

2. Materials and Methods

2.1. Definitions of burst novelty and burst surprise

We consider a spike train, i.e., a sequence of spike-event times: $T_1 < T_2 < T_3 < ...$, and the corresponding ISIs $X_i = T_{i+1} - T_i$. We define the burst novelty N_k and the burst surprise S_k for the spike train up to one particular spike k at time T_k. For mathematical convenience, and in order to avoid discussing the boundary effects in the beginning and the end of the spike train, we assume in our definitions a sequence $(X_i)_{i\in\mathbb{Z}}$.

As a null hypothesis we assume that the random variables X_i ($i \in \mathbb{Z}$) are independent and all have the same probability distribution (i.i.d.). For the experimentally most common situations where we do not have very long spike trains, it has been suggested to use the most naive assumption for the ISI distribution, namely the exponential distribution. This corresponds to the Poisson distribution of spike events and requires only one parameter to estimate [15, 16, 22]. If more data are available for the same neuron, one can also consider a more refined representation of the ISI distribution, for example, a gamma distribution (with two parameters to estimate), a gamma distribution with a fixed refractory period (i.e., a minimal value $d > 0$ of the random variables X_i), or the direct estimate of the neuron's empirical ISI distribution.

Our definition of burst novelty is based on accumulated ISIs defined as:

$$X_{l,k} = \sum_{j=1}^{l} X_{k-j}, \tag{2.1}$$

which is the sum of l consecutive ISIs preceding spike k; in other words, it is the time difference between the $(k - l)$th and the kth spikes: $X_{l,k} = T_k - T_{k-l}$. Its (cumulative) distribution function is $F_l(t) = \mathbb{P}[X_{l,k} \le t]$, which is independent of k. As in [22], we introduce the following definitions:

Definition 1. The novelty of an ISI sequence of length l ending at spike k is:

$$N_{l,k} = -\log_2[F_l(X_{l,k})] \tag{2.2}$$

and the **burst novelty** of the spike train up to spike k is defined as:

$$N_k = \max_l N_{l,k}. \tag{2.3}$$

Its distribution function $F(t) = \mathbb{P}[N_k \le t]$ is independent of k.

Definition 2.

a) The **surprise function** of the novelty (the log of the survival function) is defined as:

$$S : \begin{cases} \mathbb{R}_+ \to \mathbb{R}_+ \\ x \to -\log_2(1 - F(x)). \end{cases} \tag{2.4}$$

b) The **burst surprise** for the spike train up to spike k is:

$$S_k = S(N_k). \tag{2.5}$$

c) The **burst size** for a burst ending at spike k is defined as:

$$B_k = \arg\max_l N_{l,k}$$

and the corresponding **burst onset** and **offset** are defined at spikes $b_k^{\text{on}} = k - B_k$ and $b_k^{\text{off}} = k$, respectively.

It can be argued that this definition does not yet fully capture the intuition that a burst should be a brief outburst of activity and hence it should not consist of a too large number of spikes. Also for practical reasons it may be useful to restrict l up to a reasonable number. On the other hand, only one ISI should perhaps also not be considered a burst. These considerations lead to an idea of restricting the values of l over which $N_{l,k}$ is maximized in Eq 2.3, for example one could use $N_k^M := \max_{2\le l\le M} N_{l,k}$, which introduces two more parameters: a minimal and a maximal value of l.

We also introduce a stricter definition of a burst by requiring that each subsequent ISI in a sequence contributes to increasing the novelty when increasing l in $X_{l,k}$. In addition we require that the length of a sequence is at least 2 ISIs, i.e., 3 spikes.

Definition 3. We define the **strict burst novelty** as

$$\tilde{N}_k = N_k^L, \tag{2.6}$$

where L is the minimum $l \ge 2$ that satisfies $N_{l+1,k} < N_k^l - \delta$.

The parameter δ specifies how strictly the novelties $N_{l,k}$ for sequences should increase along with l. The strict burst surprise, burst size, onset and offset for strict burst novelty can be calculated in the same manner as for the original burst novelty.

2.2. Burst detection

The idea behind these definitions is to use the (strict) novelty to detect bursts in a spike train and to use the corresponding surprise to determine their statistical significance. Here we describe how the (strict) burst novelty is practically calculated from a given spike train and how it is used to detect bursts.

The first step is the accumulation of consecutive ISIs: $X_{l,k} = \sum_{j=1}^{l} X_{k-j}$ (Eq 2.1). For an incoming spike train we begin with computing $X_{1,2}$, then $X_{2,3}$, $X_{1,3}$, then $X_{3,4}$, $X_{2,4}$, $X_{1,4}$ and so on. In practice a limited number M of these accumulations will suffice, e.g., up to $l = 50$. Normally such a restriction of l should not influence the detected bursts. This can be checked by looking at the distribution of the detected burst sizes (see supplementary material).

For these values of l the cumulative distribution functions $F_l(X_{l,k})$ of the sums $X_{l,k}$ can be precomputed prior to the computation of $X_{l,k}$ from the spike train. The functions F_l for different values of l depend on the distribution of the random variable X_k. For the family of gamma distributions these functions are available in closed form; otherwise they have to be computed numerically or generated from simulated spike trains following the null hypothesis. Once F_l are obtained, $N_{l,k}$ can be calculated for each $X_{l,k}$ using Eq 2.2.

Next we compute the maximum of $N_{l,k}$ w.r.t. l for a fixed k, where we can introduce some reasonable stopping criterion, either the provisionally introduced maximal value M of l, or the (usually) stricter criterion introduced in the definition of the strict novelty (Def. 3). The obtained maximum is the burst novelty N_k up to spike k. We repeat this for every k and obtain a series of burst novelties N_k for all spike times T_k.

Finally we can use a predefined threshold N_{th} on the (strict) novelty for the detection of bursts. One can set the threshold according to a desired significance level as explained in Section 2.3, or one could use some conventional value that was used in previous studies (e.g., $N_{\text{th}} = 10$ as in [15]). Typically a burst event gives rise to novelty values greater than the threshold for several consecutive spikes (see Figure 1 for an example). If one is interested in identifying the spikes belonging to a burst event, firstly the end of a burst can be obtained as the spike k^* with the maximum novelty among the supra-threshold novelties, and then the beginning of this burst can be obtained as the burst onset $b^{\text{on}}_{k^*}$ of spike k^*, as shown in Figure 1. Otherwise, if one is only interested in detecting the onset of a burst as a latency from a certain reference event (e.g., stimulus onset time in a recording of stimulus evoked activity), one can take $b^{\text{on}}_{k'}$ of spike k' that first crossed the threshold after the reference event as the onset of a burst. This approach is taken in our application of the method to real spike train data shown in Section 3.2.

2.3. Statistical significance measured by burst surprise

In order to determine the significance of an observed novelty value N_{obs} we have to calculate the so-called 'survival probability' $\mathbb{P}[N \geq N_{\text{obs}}]$. The burst surprise $S(N_{\text{obs}})$ (Eq 2.5) is the logarithm of this probability: it translates the (strict) novelty value into the logarithm of a burst significance probability. A closed-form representation of the surprise as a function of novelty can be obtained, for example for gamma distributions, but it contains Mth-order multiple integrations which allow neither the derivation of an analytic form nor a numerical integration within a reasonable computation time. Thus, in order to derive a surprise value for an arbitrary novelty value, the surprise S as a function of novelty N, or in other words the novelty-to-surprise (N-to-S) curve, needs to be estimated. In this study we take a Monte-Carlo approach for this estimation, where we sample a large number of ISIs from a very long spike train (specifically, for the results shown below, we used spike trains of $1,000,000$ spikes), generated according to the ISI distribution that we assumed in the null hypothesis for novelty computation. Given these samples, we numerically estimate the distribution function $F(x)$ of the burst novelty (computed with the maximum sequence length $M = 50$) in Eq 2.4.

A plot of the N-to-S curve (e.g., Figure 2 for the null hypothesis of an exponential ISI distribution) can be used to translate a given significance level to a novelty threshold for burst detection, or conversely, an arbitrary novelty value to a significance probability. Some examples for this are described in Section 3.1.

2.4. Statistical evaluation of burst properties

In order to analyze the statistics of various properties of individual bursts, like burst size (number of spikes in a burst), burst duration, in-burst spike rate, burst novelty or significance, one first has to count the burst events occurring in a spike train. The first step in such an analysis is to plot the (strict) burst novelty as a function $N(t)$ of time as in Figure 1. The natural candidates for burst events are the local maxima of this function. Depending on the purpose of the analysis one may decide to count only local maxima with statistically significant novelty values. In addition one may decide not to count local maxima close to other maxima, e.g., not to count burst events that overlap with a more significant neighboring event. In this paper we do not yet explore these possibilities, we just show the distribution of burst sizes of significant bursts in the supplementary material.

2.5. Experimental data

We apply our burst detection method for detection of single trial latency of rate increase in spiking activity of macaque motor and pre-motor cortex during a reach-to-grasp experiment [27]. Example datasets from this experiment are available from a data publication by Brochier et al. [25]. We use for our method application one of these datasets (session i140703), which was recorded with a 100-electrode Utah array and resulted after spike sorting in 156 simultaneously recorded spike trains. The behavioral task in the experiment was, briefly, that the monkey had to reach to an object, grasp it with a predefined grip type, and pull it at a predefined force level (see [25] for more details). The information about the grip type (either a precision grip or a side grip) was delivered by a visual cue at the beginning of a preparatory period of 1000 ms, followed by the GO signal, which also provided information about the force level (high or low). In this recording session the monkey performed the task in 141 trials. In our analysis we consider the preparatory period as the baseline, and aim to detect rate changes in the period starting at the GO signal. This rate change is known to be related to the arm movement [28]. Here we ignore the specific grip type and force level for the detection of the rate increase latencies, and leave a behavior related analysis for a further project.

3. Results

In this section we first calibrate the proposed burst detection method, with special emphasis on the effects of the choice of different null hypotheses on burst detection performance. For doing that we apply the method to artificial spike train datasets of which we know the ground truth, i.e., what the parameters of the process are and thus the correct null hypothesis. Then we apply the method to spike train data recorded from macaque motor cortex to demonstrate the effects of the selection of the null hypothesis in a practical application to real spike train data, where one needs to estimate the parameters for the null hypothesis from the data themselves.

Figure 1 shows an example application of our burst detection method to an artificial Poisson spike train with an injected burst, which is generated as a Poisson process with a higher rate ($\lambda_b = 3.5$ Hz) than the baseline rate ($\lambda_0 = 1$ Hz) for a short time interval. Here, the novelty values are computed with the null hypothesis of a Poisson ISI distribution of rate $\lambda = 1$ (the baseline rate), and the novelty threshold for burst detection is set to $N = 10$, as in [15].

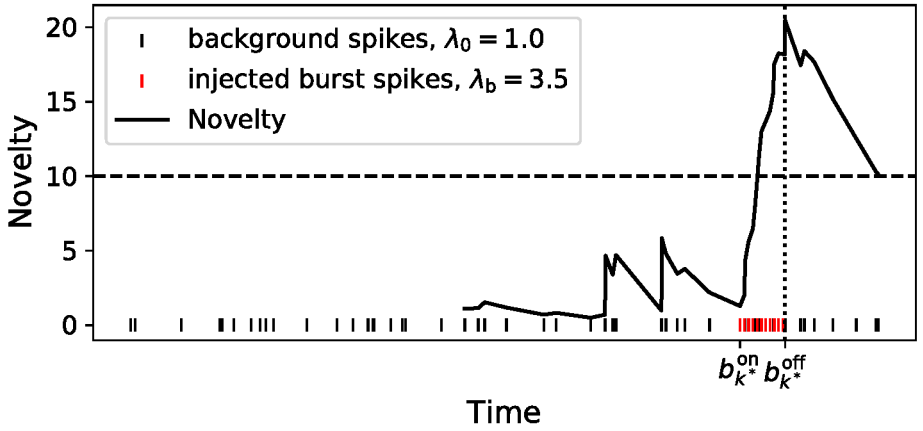

Figure 1. Example application of the burst detection method. Novelty values are computed at spike times in a Poisson spike train of rate $\lambda_0 = 1$ Hz (black tick marks) with injected burst spikes of rate $\lambda_b = 3.5$ Hz (red tick marks). The maximum ISI accumulation length $M = 20$ is used here for novelty computation. Novelty values above a fixed threshold of 10 are considered as signifying burst events. The last spike of a detected burst is identified as spike k^* with the maximum novelty value (vertical dotted line). The onset $b^{on}_{k^*}$ and the offset $b^{off}_{k^*}$ of the burst are obtained according to Def. 2c.

3.1. Relevance of the ISI distribution

In practical applications one would want to set the novelty threshold according to a conventional significance level, such as $\alpha = 0.05$. To derive a threshold novelty value corresponding to a given significance level, a novelty-to-surprise (N-to-S) curve needs to be estimated (Figure 2), as described in section 2.2. A N-to-S curve translates a given novelty value N to a corresponding surprise value S, which is related to the significance probability p of N as $S = -\log_2 p$. For example, the novelty threshold of 10 used in Figure 1 corresponds to a surprise of ~ 6.32 (i.e., $p \sim 0.0124$) for the original novelty, or a surprise of ~ 7.77 (i.e., $p \sim 0.00467$) for the strict novelty, under the assumption of a Poisson spike train. This is consistent with an intuition that, for a fixed novelty value, finding this value as a strict novelty is more surprising than finding it as an original novelty, thus it corresponds to a greater surprise in the case of the strict novelty than in the case of the original novelty. Conversely, one can derive a threshold novelty value for a given significance level α by applying the N-to-S curve inversely. For example, the novelty threshold corresponding to a significance level of $\alpha = 0.05$ is ~ 7.67 for the original novelty or ~ 6.12 for the strict novelty. This is again consistent with an intuition that, for a fixed significant level (coresponding to a fixed surprise value), it translates to a smaller strict novelty value than the respective original novelty value, because finding this smaller strict novelty value is as surprising as finding the larger original novelty value.

It has been known that experimental data tend to deviate from Poisson (see [29] and references therein). We want to avoid potential misses or false detections of bursts due to a wrong null hypothesis on the underlying process. In particular assuming a wrong ISI distribution could result in a wrong setting of the detection threshold. Therefore we explore here the effect of other processes and respective

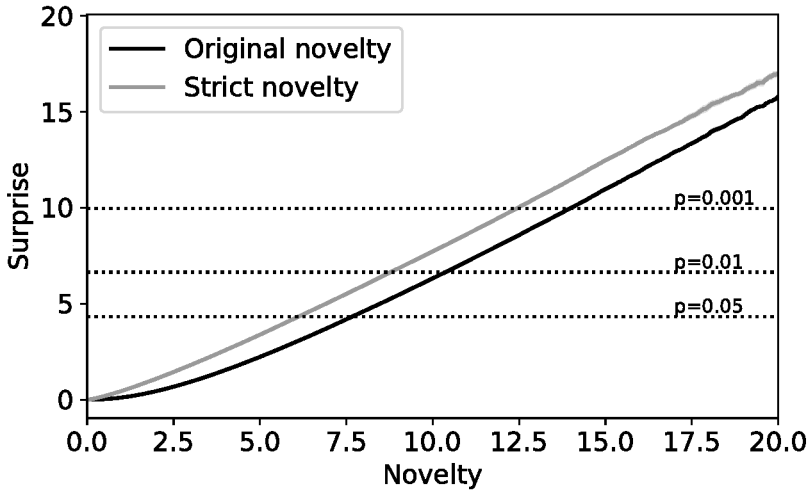

Figure 2. Novelty-to-surprise (N-to-S) curve estimated from a Poisson spike train, using the original novelty (black) or the strict novelty (with $\delta = 0$, gray). Horizontal dotted lines indicate surprise values corresponding to a few commonly used significance levels, i.e., 0.05, 0.01 and 0.001 in p-values.

null hypotheses on the N-to-S curve. For doing that we choose a gamma process, which enables to incorporate not only the mean, but also the variance of the ISI distribution, which quantifies the (ir-)regularity of ISIs into the null hypothesis. The (ir-)regularity of experimentally measured spike trains is commonly measured by the coefficient of variation (CV) of the ISIs, which ranges for experimental data between highly regular (CV < 1) and bursty (CV > 1) [25, 30]. This translates into the shape parameter κ of the gamma distribution by the relation $\kappa = \mathrm{CV}^{-2}$.

Here we examine the effect of the shape parameter on the N-to-S curve by deriving these curves with different shape parameters from within the physiological range: $\kappa = 3.33, 1.25, 1.0, 0.5$, which corresponds to CV values of $0.55, 0.894, 1.0, 1.41$. Figure 3 shows several N-to-S curves for null hypotheses of gamma distributed ISIs with the mentioned set of κ parameters. As a result the N-to-S curves show only slight differences. At first sight this seems to show that modeling the ISI distribution is not so important, but actually the N-to-S curve is only the second step in the evaluation; the first step, i.e., the novelty computation can give quite different results for different null hypotheses, as will be discussed in the following.

In real experimental situations, it may be the case that the amount of data available is not enough to correctly estimate the parameters of the ISI distribution (e.g., [31]). In such cases, one would need to bear with a poor estimate of the parameters, or with a naive assumption of a simple ISI distribution with less parameters to be specified, such as an exponential ISI distribution as expected from a Poisson spike train. Here we examine what type of effect such a mismatch between the empirical data and the assumed ISI distribution as the null hypothesis causes. In particular we explore the case where one assumes a shape parameter of $\kappa = 1$ (i.e., a Poisson spike train assumption) for the null hypothesis although the actual data is not Poissonian.

We take the following approach to examine the errors caused by this type of parameter mismatch. First, we generate sample spike trains of gamma distributed ISIs with several combinations of shape κ

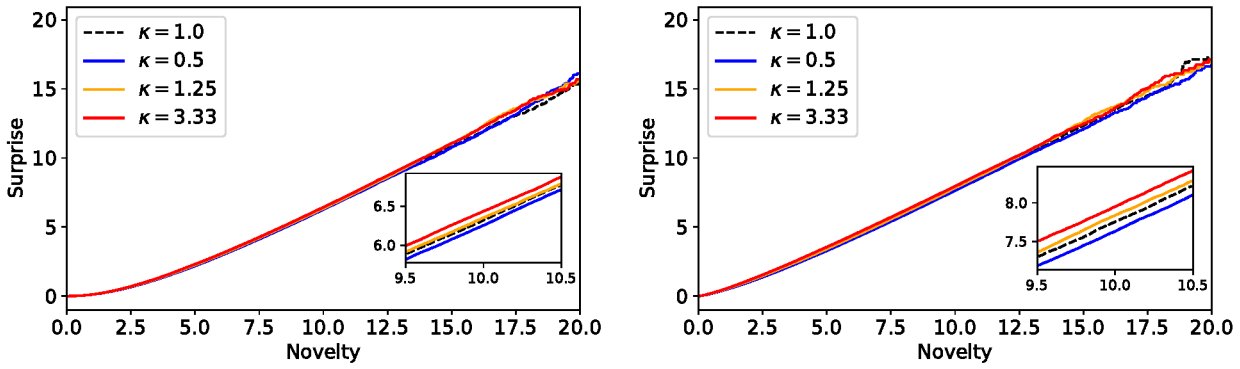

Figure 3. Dependence of the N-to-S curve on the ISI distribution. (Left) N-to-S curves, estimated using the original novelty, for gamma distributed ISIs with shape $\kappa = 1$ (i.e., a Poisson process), 0.5, 1.25 and 3.33. The latter two values are derived from the coefficients of variation of real spike trains reported in [30] and [25], respectively. The scale factor β is set to $1/\kappa$ for each shape κ. The inset shows the same graphs as in the main plot, but magnified around $N = 10$. (Right) same as the left panel, but the curves are estimated using the strict novelty.

and scale β. We vary the shape κ from 0.5 to 3.0 in steps of 0.1, and the corresponding scale factor β is set to $1/\kappa$ for each κ, such that all spike trains have the same firing rate. Then we estimate the N-to-S curve for each of the spike trains, *not* using their respective parameter values, but using $(\kappa, \beta) = (1, 1)$ for all of the spike trains (see Figure 4 left, colored curves). These curves represent the surprise of the novelty values found in the data, computed with the mismatched parameter, i.e., using the wrong null hypothesis. Differently from the case where we calculate N-to-S curves using the correct null hypothesis (Figure 3), the curves calculated using the wrong null hypothesis of Poisson assumption (Figure 4 left) strongly depend on the shape parameter κ of the data. This means that a given novelty value, which could be used as the burst detection threshold, corresponds to different surprise values (and therefore to different significance probabilities) depending on the shape parameter of the data, if one uses a Poisson null hypothesis to analyze gamma spike trains.

This point is further illustrated in Figure 4 (right) as follows. First we identify the novelty value that corresponds to a p-value of 0.05 (i.e., a surprise of $-\log_2 0.05$ (Figure 4 left, vertical dashed line), using the N-to-S curve obtained from the data with $(\kappa, \beta) = (1, 1)$ (Figure 4 left, black thick line, or equivalently Figure 2). This is the novelty threshold when the parameters are assumed as Poisson, i.e., $(\kappa, \beta) = (1, 1)$. Let's denote this novelty value as N_{th}, which is 7.67 in this case. Then we identify the surprise values that correspond to N_{th} for the other N-to-S curves, generated for non-Poisson processes but assuming the null hypothesis of Poisson in the calculation of novelty (Figure 4 left, colored curves and dots). The obtained values are denoted as $S_{\text{th}}(\kappa)$, where κ represents the shape parameter of the curve used to derive the surprise value. They represent the correct significance probabilities of N_{th}, for the data generated with shape parameter κ, but evaluated with the Poisson novelty. Naturally, $S_{\text{th}}(1) = -\log_2 0.05$ corresponding correctly to the significance level of $\alpha = 0.05$, but $S_{\text{th}}(\kappa)$ generally differs from this value for other values of κ.

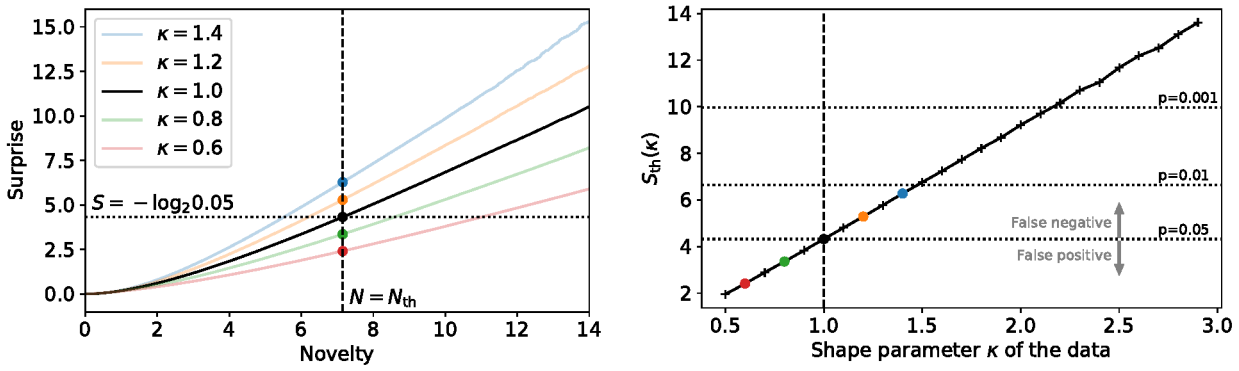

Figure 4. Bias in significance level due to parameter mismatch. (Left) Illustration of the derivation of $S_{\text{th}}(\kappa)$ from N_{th}. Left: N-to-S curves for various data parameters (colored curves; shape $\kappa = 0.6$, 0.8, 1, 1.2 and 1.4 are shown here as examples; scale $\beta = 1/\kappa$ for each shape κ) are estimated using a fixed parameter set: $(\kappa, \beta) = (1, 1)$, irrespective of the data parameters, for computation of the (original) novelty. Then, N_{th} (vertical dashed line) is obtained as $N_{\text{th}} = S^{-1}(-\log_2 0.05)$ using the N-to-S curve for $(\kappa, \beta) = (1, 1)$. $S_{\text{th}}(\kappa)$ (colored dots) is derived as $S_{\text{th}}(\kappa) = S(N_{\text{th}})$ using the N-to-S curve of each shape κ. Right: $S_{\text{th}}(\kappa)$ as a function of the shape parameter κ of the data when assuming a Poisson ISI distribution. Each colored dot corresponds to the value indicated in the left panel with the same color. Horizontal dotted lines indicate a few commonly used significance levels, i.e., 0.05, 0.01 and 0.001 in p-values.

In Figure 4 right we plot S_{th}, derived at a fixed N_{th} corresponding to the requested significance value of $p = 0.05$, as a function of the shape κ of the generated spike trains. This plot illustrates how the significance level p_κ corresponding to N_{th} (expressed as surprise values $S_{\text{th}}(\kappa)$ on the Y-axis) deviates from the desired significance value (i.e., $-\log_2 0.05$ in this case) if one wrongly assumes a Poisson ISI distribution in case of a non-Poissonian spike trains (shape parameters κ of which represented on the X-axis). The plot (Figure 4 right) shows a monotonically increasing function of κ. This means that, if one underestimates the shape parameter (to the right of the vertical line), the significance reaches a stricter level than the desired one, which implies false negative burst detections.

3.2. Application: Onset detection of single trial response activity

Here we demonstrate the effect of using a wrong null hypothesis (mismatched parameters, but correct rate) in an application of the method to real spike train data. In common neurophysiology experiments, neurons tend to show a temporary increase in their firing rates in response to external events, e.g., the onset of sensory stimuli, or to a cue to initiate motor action, and so on. Detection of such changes of neuronal spiking activity in single trials can be a demanding task given the stochastic nature of neuronal spiking.

Here we apply our burst detection method for the detection of single trial latency of rate increase in spike trains. We use a dataset recorded from macaque primary- and pre-motor cortex during a reach-to-grasp experiment (see Section 2.5 about the dataset, and see [25,27] for details of the experiment). This

dataset contains spike trains of 156 single units, and the majority of them exhibits a temporary increase in their firing rates after the GO signal to perform a reaching arm movement. The basic idea of latency detection is to calculate the burst novelty for a single trial spike train and identify the first significant burst after a certain trigger event, which is the GO signal in the present application. The identified burst is typically interpreted as the reaction of the system to the trigger event, to which neurons tend to respond by an increase of their firing rates, as mentioned above.

Thus, we are looking for a behaviorally induced rate increase, or a sequence of shorter ISIs. Essentially, we are looking for any deviation of the ISI statistics from those in the 'baseline period', here taken as the prepatory period before the GO signal. In that period we derive the shape κ and the scale β through the following relations: $\mathrm{E}(X_i) = \kappa\beta$ and $\mathrm{Var}(X_i) = \kappa\beta^2$, where $\mathrm{E}(X_i)$ and $\mathrm{Var}(X_i)$ are the estimates of the mean and the variance of the ISIs, respectively.

Given these parameters, the burst novelty for each spike in a single trial spike train is calculated, and the rate increase onset is detected as the onset of the first significant burst after the GO signal. Here the threshold novelty value for burst detection is determined based on a significance level (we use $\alpha = 0.05$), using a N-to-S curve estimated with the baseline ISI distribution parameters. To test how the performance depends on the choice of the ISI distribution parameters as the null hypothesis, we also apply the method with a mismatched parameter set: the shape parameter of 1 and scale parameter of $\kappa\beta$, which corresponds to the assumption of Poisson spike train. We hereafter refer to this mismatched assumption as Poisson assumption, while the original assumption is referred to as gamma assumption.

Figure 5 shows raster plots of representative units, which exhibit a clearly visible rate increase typically around 500 ms after the GO signal. The estimated rate change onset times in single trials are shown with orange plus marks (gamma assumption) and blue crosses (Poisson assumption). The trials are sorted along the Y-axis in ascending order of the detected latency (by gamma assumption, all panels). The unit shown in Figure 5 upper left appears to change clearly its firing rate in relation to the GO signal in every trial. In this case the rate changes were so clear that the application of method with either assumption detected significant rate changes in all trials (the number of detected rate changes for both cases are shown in the figure legend). The estimated rate change onset times were quite consistent between the two assumptions, but not completely identical.

The unit shown in Figure 5 upper right changes the rate less clearly in relation to the GO signal in the sense that not in every trial a rate change is detected, and the two null hypotheses yield large differences. With the gamma assumption, the method detects rate changes in 134 out of 141 trials, while the Poisson assumption results in detection of rate changes in 95 trials. Such a reduction in the number of detected rate changes is explained by the tendency for false negatives shown in Figure 4 right. Since the estimated κ of this unit is 2.3916, Poisson assumption results in a significance level of $p < 0.001$, which is much stricter than the desired level of $p = 0.05$. In many trials, the onset detected with the Poisson assumption is later than the one detected with the gamma assumption in the same trial. This happens when one burst is detected as significant only for the gamma assumption but not for the Poisson assumption, and then a second significant burst (for both assumptions) occurred after the first one. The opposite case, where the first burst is significant only for the Poisson assumption but not for the gamma assumption, is less likely to happen because of the above mentioned difference in the significance level of the two assumptions. When different bursts are detected by the different assumptions, the onset of the burst that first crossed the threshold is typically earlier than the onset of the other burst, but this is not always the case.

The effect becomes even more serious when the shape parameter is even greater in the data, e.g., Figure 5, lower left. Here the unit exhibits a highly regular spike train during the baseline period with a shape parameter of 4.6175 (highly regular), and also during the rate change period after the GO signal (visual inspection). The application of the burst detection method with the gamma assumption actually detects a rate change in almost all the trials, but the Poisson assumption detects them only in 109 out of the 141 trials. The estimated onset times of the rate changes agree to a large extent (10% mismatch of the detected rate changes).

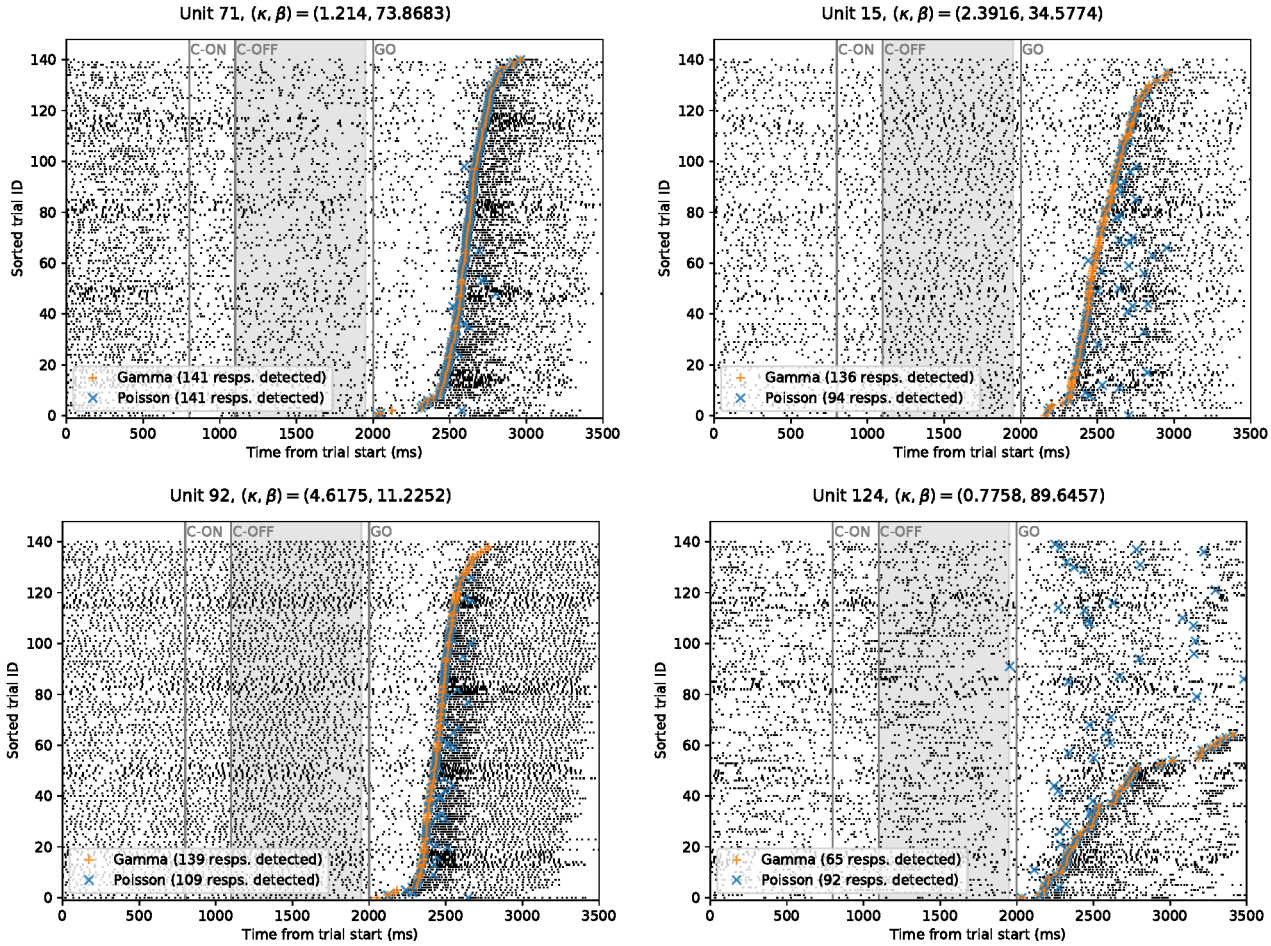

Figure 5. Raster plots of four different single units (from session i140703 of the dataset published in [25]) with rate change onset estimation results. Trials are sorted in the ascending order of the rate change onset times estimated using a gamma assumption based on parameter estimates in the baseline period (gray shaded). C-ON: onset of the cue signal. C-OFF: offset time of the cue signal. GO: occurrence of the GO signal. Gray-shaded area: baseline period, within which the spike data were used to estimate the parameters for the null hypothesis.

We also show example of the effect in the opposite direction (Figure 5, lower right) where visually there seems to be no rate change in every trial and the estimated shape parameter is less than unity (i.e., more bursty than Poisson). Here the application of our method using the Poisson assumption

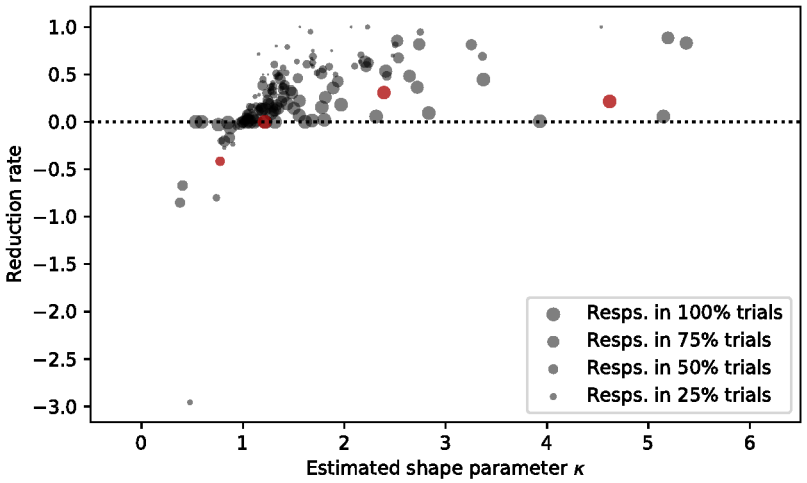

Figure 6. Reduction rate of rate change detection as a function the estimated shape parameter from the data. Each disk corresponds to a single unit, and the size of a disk represents the number of the rate changes detected (with the gamma assumption) for the corresponding unit. The disks corresponding to the units shown in Figure 5 are indicated in red.

results in rate changes detected in more trials than for the gamma assumption, i.e., shows a tendency for false positive detections. This is expected from the dependence of the significance level on the shape parameter as shown in Figure 4, right.

We quantify the dependence of the false negatives and false positives as a function of the shape parameter of the data by calculating the reduction rate of rate change detection defined as: $(N_{\text{gamma}} - N_{\text{Poisson}})/N_{\text{gamma}}$, where N_{gamma} refers to the number of rate change detections with a shape parameter deviating from 1 and N_{Poisson} with the Poisson assumption ($\kappa = 1$). Note that the reduction rate takes a negative value when the Poisson assumption detected more rate changes than the gamma assumption (as in Figure 5, lower right). N_{gamma} and N_{Poisson} are now derived for the same data as in Figure 5 and on the rest of the simultaneously recorded neurons in the dataset.

Figure 6 shows the reduction rates of all 151 units in the dataset plotted against the shape parameters estimated for each of them. This figure shows a stronger reduction of rate change detection by use of a Poisson assumption for units with larger shape parameters, and vice versa. The reduction rate never takes a negative value for the units with estimated shape parameters greater than unity, while it does not become positive for the units with a shape parameter less than unity. This means that assuming a Poisson assumption for regular spike trains leads to an underestimation of rate onsets, which is consistent with our results in Figure 4.

4. Conclusions

In this paper we have extended the original definition of burst surprise by Legendy [15, 32] in two aspects. One is to incorporate an arbitrary ISI distribution into the null hypothesis for computing Legendy's burst surprise, which is renamed to burst novelty. The other is to assign a significance measure to the obtained novelty, which is now called burst surprise. We have shown that using a proper

null hypothesis for given data is important for detecting bursts in the data at a desired significance level; the use of a wrong null hypothesis results in false negative or false positive burst detections, depending on the property of the data. Our derivation of the significance measure, i.e., the burst surprise for spike trains with arbitrary ISI distributions enables us to examine quantitatively the false negative/positive effect of the use of a wrong null hypothesis. We have also demonstrated this effect in the application to real spike train data from macaque motor cortex, where neurons typically generate spike trains which deviate from a Poisson process. The result of burst detection actually shows the false negative/positive effects depending on the (ir-)regularity of the spike train, which are consistent with the quantification based on artificial spike trains.

The essential idea of measuring 'surprise' in neural activity was first suggested in [32] and later elaborated statistically leading to the definition and first evaluation of the burst surprise in [16]; it was first applied to cat visual cortex data in [15]. Our method consists of two stages: in the first stage bursts are detected by Legendy's method, now called burst novelty. This is computationally simple and could even be used online to detect interesting events, or to quickly scan a huge dataset of massively parallel recording. In the second stage the surprise is calculated from novelty in order to allow an evaluation in terms of statistical significance, which usually requires more computation time to generate and evaluate long spike trains from the null hypothesis. To our knowledge this second step has not been applied to real spike train data before.

We demonstrated the usefulness of our method in an application on latency detection, where we used gamma distributions with different shape parameters as null hypotheses. We have shown that the correct estimation of the shape parameter (in addition to the rate) is important[†] for a reasonable setting of the burst detection threshold and therefore for the detection of all 'responses' of individual neurons. Just assuming a Poisson distribution is not sufficient in this application. This would be due to a comparatively large regularity in the spiking activity and the ISI distribution of the neurons in motor cortex, which needs to be incorporated in the null hypothesis. Such a large regularity of motor cortex neurons has been shown to be a common feature throughout mammalian species, as well as a large irregularity of hippocampus neurons [33, 34]. It has also been reported that neurons in macaque parietal association areas show regular spike trains beyond Poisson assumption [35]. Our method would be particularly suitable for the detection of bursts in recordings from these areas, due to its ability to take into account the (ir-)regularity of spike trains in order to avoid false negative/positive detections of bursts.

The extension of the strict novelty method by improving the null hypothesis to include also other ISI distributions leads to an increased sensitivity of the method for spike trains that deviate from Poisson processes. However, bursts in spike trains may have several other reasons discussed in the literature, e.g., that other biophysical processes are activated and thus the spikes in the burst express a serial correlation in time [2, 36, 37]. Fortunately the surprise method does not require a statistical model of all these processes. Our null hypothesis considers ISIs that may deviate from Poisson, but are still following independently each other (renewal). Thus our burst detection is based on the deviation from the null hypothesis based on a) a temporary increase of the firing rate for the same shape parameter, b) a temporary deviation from the shape parameter assumed in the null hypothesis, or c) a deviation from the renewal hypothesis, or any combination of these.

[†]One needs to ensure the stationarity of the spike train used for this estimation, because firing rate modulation in a spike train leads to underestimation of its shape parameter. This calls for a careful selection of the baseline period from which the parameters are estimated.

In general, applying the burst surprise method to different sets of neurons and experimental settings shows the need of a careful choice of the null hypothesis, since the burst-surprise is a statistical measure of the deviation from this null hypothesis in the direction of burstiness. In the application to response onset detection we have presented here a pilot study on one data set. In a more extensive study we plan to apply this method to more spike train data and to test better models of the ISI distribution, including the use of the empirical ISI distribution. This would ensure that the measured deviation from the null hypothesis is (almost) only concerned with violations of the i.i.d. assumption, e.g., with increased correlation between subsequent ISIs or with temporary changes in the (baseline) ISI distribution, for example by increasing firing rate or increasing probability of small ISIs. In addition we plan to compare this method to other methods of response latency estimation [38–42], which are mostly based on statistical estimation of rate change.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under Specific Grant Agreement No. 785907 (Human Brain Project SGA2), Deutsche Forschungsgemeinschaft Grant GR 1753/4-2 Priority Program (SPP 1665) and Research Training Group 2416 MultiSenses-MultiScales.

Conflict of interest

The authors declare no conflict of interest.

References

1. K. D. Harris, H. Hirase, X. Leinekugel, et al., Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells, *Neuron*, **32** (2001), 141–149.
2. O. Avila-Akerberg and M. J. Chacron, Nonrenewal spike train statistics: causes and functional consequences on neural coding, *Exp. Brain Res.*, **210** (2011), 353–371.
3. F. Zeldenrust, W. J. Wadman and B. Englitz, Neural coding with bursts–current state and future perspectives, *Front. Comput. Neurosci.*, **12** (2018), 48.
4. A. Kepecs and J. Lisman, Information encoding and computation with spikes and bursts, *Network-Comp. Neural*, **14** (2003), 103–118.
5. D. A. Butts, P. O. Kanold and C. J. Shatz, A burst-based "hebbian" learning rule at retinogeniculate synapses links retinal waves to activity-dependent refinement, *PLoS Biol.*, **5** (2007), e61.
6. J. Cocatre-Zilgien and F. Delcomyn, Identification of bursts in spike trains, *J. Neurosci. Method.*, **41** (1992), 19–30.
7. D. C. Tam, An alternate burst analysis for detecting intra-burst firings based on inter-burst periods, *Neurocomputing*, **44** (2002), 1155–1159.
8. M. Chiappalone, A. Novellino, I. Vajda, et al., Burst detection algorithms for the analysis of spatio-temporal patterns in cortical networks of neurons, *Neurocomputing*, **65** (2005), 653–662.

9. B. Gourévitch and J. J. Eggermont, A nonparametric approach for detection of bursts in spike trains, *J. Neurosci. Method.*, **160** (2007), 349–358.
10. L. Chen, Y. Deng, W. Luo, et al., Detection of bursts in neuronal spike trains by the mean inter-spike interval method, *Prog. Nat. Sci.*, **19** (2009), 229–235.
11. V. Pasquale, S. Martinoia and M. Chiappalone, A self-adapting approach for the detection of bursts and network bursts in neuronal cultures, *J. Comput. Neurosci.*, **29** (2010), 213–229.
12. D. Ko, C. Wilson, C. Lobb, et al., Detection of bursts and pauses in spike trains, *J. Neurosci. Method.*, **211** (2012), 145–158.
13. I. A. Välkki, K. Lenk, J. E. Mikkonen, et al., Network-wide adaptive burst detection depicts neuronal activity with improved accuracy, *Front. Comput. Neurosci.*, **11** (2017), 40.
14. D. J. Bakkum, M. Radivojevic, U. Frey, et al., Parameters for burst detection, *Front. Comput. Neurosci.*, **7** (2014), 193.
15. C. R. Legendy and M. Salcman, Bursts and recurrences of bursts in the spike trains of spontaneously active striate cortex neurons, *J. Neurophysiol.*, **53** (1985), 926–939.
16. G. Palm, Evidence, information and surprise, *Biol. Cybern.*, **42** (1981), 57–68.
17. S. Grün and S. Rotter (eds.), *Analysis of Parallel Spike Trains*, Springer, 2010.
18. S. Grün, M. Diesmann and A. Aertsen, 'Unitary Events' in multiple single-neuron spiking activity. I. Detection and significance, *Neural Comput.*, **14** (2002), 43–80.
19. S. Grün, M. Diesmann and A. Aertsen, 'Unitary Events' in multiple single-neuron spiking activity. II. Non-Stationary data, *Neural Comput.*, **14** (2002), 81–119.
20. E. Torre, D. Picado-Muiño, M. Denker, et al., Statistical evaluation of synchronous spike patterns extracted by frequent item set mining, *Front. Comput. Neurosci.*, **7** (2013), 132,
21. P. Quaglio, A. Yegenoglu, E. Torre, et al., Detection and evaluation of spatio-temporal spike patterns in massively parallel spike train data with spade, *Front. Comput. Neurosci.*, **11** (2017), 41.
22. G. Palm, Novelty, Information and Surprise, Springer Science & Business Media, 2012.
23. M. R. DeWeese, M. Wehr and A. M. Zador, Binary spiking in auditory cortex, *J. Neurosci.*, **23** (2003), 7940–7949.
24. H. Câteau and A. Reyes, Relation between single neuron and population spiking statistics and effects on network activity, *Phys. Rev. Lett.*, **96** (2006), 058101.
25. T. Brochier, L. Zehl, Y. Hao, et al., Massively parallel recordings in macaque motor cortex during an instructed delayed reach-to-grasp task, *Scientific Data*, **5** (2018), 180055,
26. S. Tokdar, P. Xi, R. C. Kelly, et al., Detection of bursts in extracellular spike trains using hidden semi-markov point process models, *J. Comput. Neurosci.*, **29** (2010), 203–212.
27. A. Riehle, S. Wirtssohn, S. Grün, et al., Mapping the spatio-temporal structure of motor cortical LFP and spiking activities during reach-to-grasp movements, *Front. Neural Circuit.*, **7** (2013), 48,
28. E. Torre, P. Quaglio, M. Denker, et al., Synchronous spike patterns in macaque motor cortex during an instructed-delay reach-to-grasp task, *J. Neurosci.*, **36** (2016), 8329–8340,

29. A. Riehle, T. Brochier, M. Nawrot, et al., Behavioral context determines network state and variability dynamics in monkey motor cortex, *Front. Neural Circuit.*, **12** (2018), 52.

30. M. P. Nawrot, C. Boucsein, V. Rodriguez Molina, et al., Measurement of variability dynamics in cortical spike trains, *J. Neurosci. Method.*, **169** (2008), 374–390.

31. M. P. Nawrot, Analysis and interpretation of interval and count variability in neural spike trains, in *Analysis of Parallel Spike Trains* (eds. S. Rotter and S. Grün), Springer, Berlin, 2010.

32. C. R. Legendy, Three principles of brain function and structure, *Int. J. Neurosci.*, **6** (1975), 237–254.

33. S. Shinomoto, H. Kim, T. Shimokawa, et al., Relating neuronal firing patterns to functional differentiation of cerebral cortex, *PLOS Comput. Biol.*, **5** (2009), e1000433.

34. Y. Mochizuki, T. Onaga, H. Shimazaki, et al., Similarity in neuronal firing regimes across mammalian species, *J. Neurosci.*, **36** (2016), 5736–5747.

35. G. Maimon and J. A. Assad, Beyond poisson: Increased spike-time regularity across primate parietal cortex, *Neuron*, **62** (2009), 426–440.

36. F. Farkhooi, M. F. Strube-Bloss and M. P. Nawrot, Serial correlation in neural spike trains: Experimental evidence, stochastic modeling, and single neuron variability, *Phys. Rev. E*, **79** (2009), 021905.

37. J. Csicsvari, H. Hirase, A. Czurko, et al., Reliability and state dependence of pyramidal cell–interneuron synapses in the hippocampus: an ensemble approach in the behaving rat, *Neuron*, **21** (1998), 179–189.

38. S. N. Baker and G. L. Gerstein, Determination of response latency and its application to normalization of cross-correlation measures, *Neural Comput.*, **13** (2001), 1351–1377.

39. M. Nawrot, A. Aertsen and S. Rotter, Elimination of response latency variability in neuronal spike trains, *Biol. Cybern.*, **5** (2003), 321–334.

40. A. Bollimunta, K. H. Knuth and M. Ding, Trial-by-trial estimation of amplitude and latency variability in neuronal spike trains, *J. Neurosci. Method.*, **160** (2007), 163–170.

41. M. Messer, M. Kirchner, J. Schiemann, et al., A multiple filter test for the detection of rate changes in renewal processes with varying variance, *Ann. Appl. Stat.*, **8** (2014), 2027–2067.

42. M. Levakova, M. Tamborrino, S. Ditlevsen, et al., A review of the methods for neuronal response latency estimation, *Biosystems*, **136** (2015), 23–34.

Supplementary

Restricting the novelty maximization in the computation of the burst novelty

The computation of the burst novelty N_k in Eq 2.3 in principle involves a maximization over all the spikes preceding spike k until the beginning of the spike train. In practice this may be too time consuming and therefore it would be reasonable to restrict this process to some fixed number M of preceding spikes. If one chooses a number M, e.g., $M = 50$ as we did in our experimental evaluation, one can plot the distribution of resulting burst sizes, and argue that the bursts contributing to the last bin (i.e., size 50) could have been longer without the restriction. The figure below shows some plots

of this kind. One can see that in the presented case a restriction to $M = 50$ is fully sufficient for the strict burst novelty, but potentially underestimates the size of ~ 5 percent of the largest bursts for the original burst novelty.

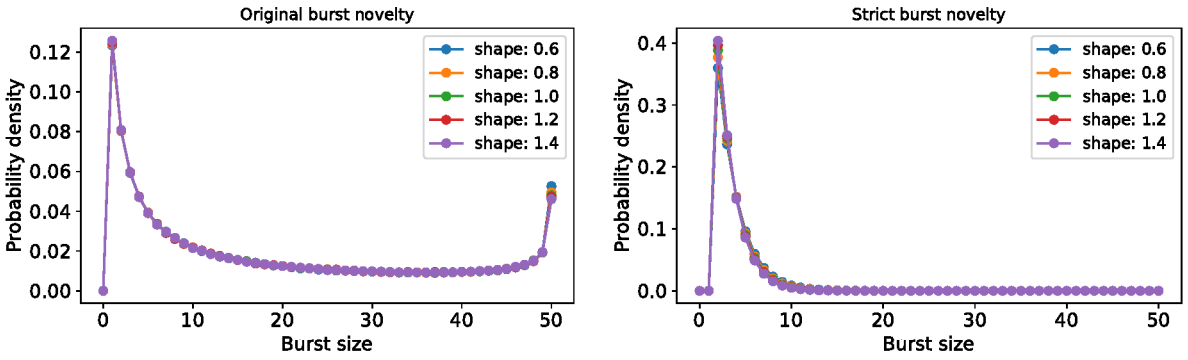

Figure S1. Distribution of burst sizes for original novelty (left) and strict novelty (right) in the case of the maximum ISI accumulation length $M = 50$. These are computed from spike trains (length: 1,000,000 spikes) of gamma distributed ISIs with different shape parameters, as indicated in the figure legend, using the gamma null hypothesis with the correct parameters. Note that the burst size is in terms of the number of ISIs and therefore a burst of size n contains $n + 1$ spikes.