



A new similarity measure for spike trains: Sensitivity to bursts and periods of inhibition

David Lyttle^{a,*}, Jean-Marc Fellous^{a,b}

^a Program in Applied Mathematics, The University of Arizona, PO Box 210089, Tucson, AZ 85721-0089, United States

^b Department of Psychology, The University of Arizona, PO Box 210089, Tucson, AZ 85721-0089, United States

ARTICLE INFO

Article history:

Received 16 January 2011

Received in revised form 29 April 2011

Accepted 4 May 2011

Keywords:

Spike train metrics

Time series analysis

Synchrony

Neural coding

Neural data analysis

ABSTRACT

An important problem in neuroscience is that of constructing quantitative measures of the similarity between neural spike trains. These measures can be used, for example, to assess the reliability of the response of a single neuron to repeated stimulus presentations, or to uncover relationships in the firing patterns of multiple neurons in a population. While several similarity measures have been proposed, the extent to which they take into account various biologically important spike train features such as bursts of spikes, or periods of inactivity remains poorly understood. Here we compare these measures using tests specifically designed to assess the sensitivity to bursts and silent periods. In addition, we propose two new measures. The first is designed to detect periods of shared silence between spike trains, while the second is designed to emphasize the presence of common bursts. To assist researchers in determining which measure is best suited to their particular data analysis needs, we also show how these measures can be combined and how their parameters can be determined on the basis of physiologically relevant quantities.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

A fundamental issue in neuroscience research is that of quantifying the similarity or dissimilarity of temporal patterns of neuronal spiking activity, or “spike trains.” There are two main classes of experiments in which such an assessment is important. The first includes experiments *in vitro* or *in vivo* where single neurons respond to repeated presentations of the same stimulus. Understanding how a neuron encodes the stimulus relies on understanding how similar or dissimilar its responses are across these presentations. Typically, these similarities are based on a smoothed version of the binned firing rate of the responses. The second class of experimental paradigms involves the simultaneous recording of multiple neurons. In this case, understanding how populations of neurons work together relies upon an assessment of their partial correlations. The similarity or dissimilarity in the firing of a population of neurons at a particular time may carry information about the underlying population code. The need to quantify the similarity between spike trains also arises in computational work. For example, it can become important both in fitting neural models to data, and in comparing different models on the basis of how accurately they reproduce neural activity patterns. Both tasks require

that there be some method for comparing a spiking neural model's output to biological spike train data (Jolivet et al., 2008; Rossant et al., 2010). In classical approaches, statistical measures such as the cross-correlation coefficient at zero lag, or the entire cross-correlogram are used. Whether across multiple trials, or across multiple neurons, these measures are essentially blind to physiologically relevant features of the trains such as bursts or periods of shared inhibition. Because the measures are statistical, they also de-emphasize the role of single spikes, the timing of which may be important for computation.

As an alternative approach to this problem, a variety of spike train similarity measures have been proposed (Houghton, 2009; Kreuz et al., 2007; Quiroga et al., 2002; Schreiber et al., 2003; van Rossum, 2001; Victor and Purpura, 1997). Some of these quantitative measures of similarity are metrics in the strict mathematical sense, and all of them can be thought of as attempts to quantify the intuitive notion of a “distance” between two spike trains (Victor, 2005). In constructing or choosing a similarity measure, one faces the question of what exactly it means for two trains to be considered similar (close) or dissimilar (far apart), and how this definition of similarity is incorporated into the measure. One idea is that two patterns of neuronal activity should be “close” if they are responses to the same input, and “far apart” if they are responses to distinct inputs. This assumes that their response is deterministic, with one response per input. There is experimental evidence that this may not be the case in general (Fellous et al., 2004). Another possible way of defining similarity is to consider that two spike trains are

* Corresponding author. Tel.: +1 520 621 7447.

E-mail address: dlyttle@math.arizona.edu (D. Lyttle).

“close” if they elicit a similar response post-synaptically. To make these intuitive notions more explicit, throughout this work we say that a function of two spike trains qualifies as a similarity measure if certain minimal criteria hold. If x and y are two spike trains defined in the time interval from 0 to L , for some time L , then a function $d(x,y)$ is a similarity measure if the following hold:

1. $d(x,y) \geq 0$ for all x and y .
2. $d(x,y) = 0$ only when $x = y$ or x and y are “nearly identical.”
3. $d(x,y) > d(x,z)$ if x and y are “more dissimilar” than x and z .

Note that these criteria are subject to a great deal of interpretation. There is no universally agreed upon definition of what it means for two trains to be “nearly identical,” or for two trains to be “more dissimilar” than two other trains. It is entirely possible that a single definition may not be appropriate in all experimental contexts. In most cases, the analysis of spike train data requires that some assumption be made about which features of a spike train are physiologically relevant, what time scales are involved, and whether certain features are important for encoding information. For example, one may assume that all of the relevant information in a spike train is conveyed by the overall firing rate, in which case a measure based upon comparing total spike counts in a given window would be sufficient for assessing the similarity or dissimilarity between trains. However, this assumption neglects the possible importance of the precise timing of single spikes. Multiple studies addressing a diverse array of topics, including theta phase precession in the hippocampus (Maurer and McNaughton, 2007), replay of activity during sharp waves (Skaggs and McNaughton, 1996), and spike-timing dependent plasticity (STDP) (Bi and Poo, 1998), have suggested that the timing of single spikes may be important at time scales of less than 50 ms. Other studies have shown that many synapses away from the periphery, neurons could respond to sensory stimuli with a precision of less than 5 ms (Kara et al., 2000; Reinagel and Reid, 2002). Altogether, these studies suggest that at least in some cases, neurons can fire with exquisite precision, that this precision can have an impact on synaptic strength, and that it may be used for encoding information.

Various subtleties must be taken into account in the evaluation of similarity measures that are based on spike timing. For example, it may be reasonable to assume that differences between spike trains that result from small perturbations in the exact spike times (i.e. “jitter”), or those due to the insertion or deletion of single spikes, are likely to be the result of synaptic noise, and as such, should be de-emphasized by a good similarity measure.

In addition to firing rates and to the timing of individual spikes, there may be other, less obvious but nonetheless physiologically important features of spike trains that should be taken into account when quantitatively comparing two spike trains. For example, many cells fire tightly grouped “bursts” of spikes in addition to isolated action potentials. It has been argued elsewhere that these bursts may in some cases be more significant firing events than single spikes, in that they can overcome intrinsic synaptic unreliability and allow information to propagate more effectively across synapses (Lisman, 1997). Consequently, one may assume that the deletion or insertion of bursts of spikes, or changes in burst timing are more likely to be physiologically important post-synaptically than equivalent changes in the number or timing of isolated spikes. However the extent to which existing spike train similarity measures take this into account has not been explored. In the present study we propose a simple empirical test to determine how the different spike train similarity measures “weigh” bursts relative to single spikes. This test reveals striking and previously unexamined differences in how the various measures behave. We also propose a new similarity measure that allows for selective emphasis to be placed upon bursts relative to single spikes.

Another spike train feature of interest which we wish to emphasize is shared periods of inactivity or silence. Common periods of silence may in some cases be ‘information rich’, especially in cells with spontaneous activity. For example, one particularly interesting set of studies has found evidence that cerebellar Purkinje cells may be conveying information using the timing and duration of the periods in which the cell is silent (De Schutter and Steuber, 2009). More generally, silences may be the result of periods of active inhibitions, either stimulus driven across trials, or synchronized across neurons. We propose a set of tests for determining how existing measures respond to silent periods, and a new measure that has been designed specifically to detect shared silent periods of inactivity between spike trains.

It is doubtful that a single similarity measure is appropriate for all experimental data sets given the existence of a wide range of possible mechanisms through which spike trains could be encoding information. The best choice of similarity measure depends intimately on the specific features of the spike data under consideration. A recent study has compared several of these measures on the basis of their effectiveness in discriminating spike trains on the basis of firing rate and spike synchrony (Paiva et al., 2010). However, as we stated previously, the extent to which different measures respond to certain specific features of spike trains (bursts and silence) has for the most part not been explored. We examine the sensitivity of several existing similarity measures to periods of common silence and the presence of bursts through a set of simple computational tests using surrogate data, and propose a complementary pair of new measures that are designed specifically to address these features. As a demonstration of their potential, we subject our new measures to the same series of tests. Note that here we focus entirely on so-called “binless” measures. Previous works have pointed out that binned measures, in which spike trains are pre-processed and represented as binary vectors, may be undesirable in that an incorrect choice of bin width may result in a detrimental loss of information (Hunter et al., 1998; Kruskal et al., 2007; Masud and Borisyuk, 2011; Schreiber et al., 2003; van Rossum, 2001; Victor, 2005). Furthermore such measures typically assume that the underlying “space” of spike trains is inherently Euclidean, which may be unfounded (Aronov and Victor, 2004).

2. Overview of existing spike train similarity measures

2.1. Cost-based (Victor–Purpura) metrics

The Victor–Purpura family of metrics is a set of cost-based metrics which assigns distances to pairs of spike trains by determining a minimum ‘cost’ associated with transforming one spike train into another (Victor, 2005; Victor and Purpura, 1997). The basic principle is similar to that which underlies the various cost based or “edit-length” distances used in the analysis of genomic data (Gusfield, 1997). These cost based metrics are constructed using the following components:

- A list of elementary transformations (i.e. inserting or deleting a spike, or shifting a spike by some time difference dt).
- A set of costs associated with each transformation.

Typically, the cost of inserting or deleting a spike is set to a value of 1, while the cost of shifting a spike by an amount dt is $q \times dt$. Here q is a free parameter which allows for the adjustment of the sensitivity of the metric to the time scale at which similarity is considered. Given two arbitrary spike trains, there always exists some “path” or sequence of elementary transformations that transforms one spike train into another. Associated with each path is a cost, given by the sum of the costs of the elementary transformations that com-

pose individual steps of the path. The numerical distance that the metric assigns to each pair of spike trains is the minimum of the costs of all possible paths by which one train could be transformed into the other. It has been shown that this function satisfies the set of properties that define a metric in the strict mathematical sense (Victor and Purpura, 1997). The efficient implementation of this metric requires an algorithm for calculating the minimum cost of possible paths, which is non-trivial but has been discussed at length elsewhere (Victor and Purpura, 1997). One should note that this family of cost based metrics includes multi-neuronal extensions and variants that involve transforming sequences of ISIs rather than sequences of spike times (Victor, 2005; Victor and Purpura, 1997).

2.2. Convolution metrics

The convolution metrics involve a different approach (Hunter et al., 1998; Schreiber et al., 2003; van Rossum, 2001). The basic procedure by which one computes the value of the metric for a given pair of spike trains is as follows: one first convolves each spike train with a kernel function (e.g. a decaying exponential, or a Gaussian), and then uses the L_2 function space norm of the difference to compute the distance between the resulting functions.

To illustrate this more explicitly, assume two spike trains, $x = \{x_1, x_2, x_3, \dots, x_n\}$ and $y = \{y_1, y_2, y_3, \dots, y_m\}$. The first step in computing $d(x, y)$ is to first perform the mappings:

$$x \rightarrow f(t) = \sum_j H(t - x_j)K(t - x_j, \tau) \quad \text{and}$$

$$y \rightarrow g(t) = \sum_i H(t - y_i)K(t - y_i, \tau)$$

Here H is the Heaviside theta function ($H(u) = 0$ for $u \leq 0$, $H(u) = 1$ when $u > 0$), and K is a one parameter kernel function. A typical choice of one-parameter kernel function for this measure is a decaying exponential with time constant τ , given by:

$$K(t - x_j) = \frac{1}{\tau} \exp(-(t - x_j))$$

The smaller the time constant, the more rapidly the exponential kernel decays, which implies that small time constants will result in an increased sensitivity to precise spike timing. Once the mappings are complete, the distance is given by:

$$d_R(x, y) = \sqrt{\int_0^L (f(t) - g(t))^2 dt}$$

Constructing the metric with a Gaussian kernel also involves the use of a free parameter, corresponding to the width of the Gaussian function. Like q in the Victor–Purpura method, these parameters set the time scale at which similarity is measured. It is largely up to the experimenter to decide what this value is, depending on the experiment, the type of cells and the brain area. Many properties of this metric, such as the sensitivity of the metric to the insertion of noisy spikes and jitter, have been explored elsewhere (Fellous et al., 2004; Paiva et al., 2010; van Rossum, 2001). A multi-neuronal extension to this metric has also been proposed (Gunay et al., 2008; Houghton, 2009).

2.2.1. Synapse-like convolution metric variant

An extension to the conventional convolution metric based upon synaptic depression has recently been proposed (Houghton, 2009). In the original convolution metric, spike trains are mapped to real-valued functions in the manner described above. An alternative, equivalent description of this mapping can be given which makes explicit the parallels between the construction of the functions used in the metric and a model of synaptic transmission. In

this description, the functions to which the spike trains are mapped are taken to be solutions of the following ODE:

$$\tau \frac{df}{dt} = -f$$

with the added condition that at each of the spike times t_i the value of the function f undergoes the following jump:

$$f(t_i) \rightarrow f(t_i) + \delta$$

where δ is typically equal to 1. The variation on this metric proposed by Houghton et al. modifies this mapping by making the magnitude of the jumps that occur at spike times dependent upon the value of the function at that time. Formally, the above expression is changed to the following:

$$f(t_i) \rightarrow (1 - \mu)f(t_i) + 1$$

Once both trains have been subjected to this mapping, the distance is computed as in the standard convolution metric, by taking the L_2 norm of the difference. Note that when $\mu = 0$, the mapping is the same as in the convolution metric. However for non-zero values of μ , spikes that occur shortly after a preceding spike produce a smaller jump in the function value. As a result, such spikes will have a reduced impact on the result of computing the distance between two trains using this measure. The biophysical motivation given by the authors is that it more realistically captures short-term synaptic dynamics by including synaptic depression (Houghton, 2009). This similarity measure is a true metric, because the mapping from spike train to a function is unique, and because the L_2 distance between functions is a true metric.

2.2.2. Spike correlation measure

A different, correlation based approach can be computed as follows: Given two spike trains $x = \{x_1, x_2, x_3, \dots, x_n\}$ and $y = \{y_1, y_2, y_3, \dots, y_m\}$, one first performs a mapping from each spike train to a real-valued function through convolution with a chosen kernel function (Schreiber et al., 2003; Wiskott et al., 1997). This step is identical to the corresponding step in the computation of the convolution metric, with the exception that the kernel function is typically Gaussian rather than a decaying exponential. This mapping yields two time-dependent functions, $f(t)$ and $g(t)$, which can be compared as follows:

$$d_{SC}(f, g) = 1 - \frac{\int_0^L f(t)g(t) dt}{\sqrt{\int_0^L f(t)^2 dt} \sqrt{\int_0^L g(t)^2 dt}}$$

Note that the output of this measure will always lie within $[0, 1]$. This measure proved efficient in a data clustering study (Fellous et al., 2004).

2.3. Event synchronization

The event synchronization similarity measure was first proposed to measure synchronization and time delays in datasets related to epileptic seizures (Quiroga et al., 2002). Unlike several of the other measures, it does not have any free time-scale parameters that need to be set a priori. For two spike times x_i and y_j , each from a different train, the value τ_{ij} is computed as half of the minimum of the four ISIs around these two spike times:

$$\tau_{ij} = \frac{\min(x_{i+1} - x_i, x_i - x_{i-1}, y_{j+1} - y_j, y_j - y_{j-1})}{2}$$

Next, one constructs the function $c(x|y)$ which intuitively is designed to count the number of times a spike from the train x

appears immediately after a spike in the train y . The function is given by:

$$c(x|y) = \sum_{i=1}^{M_x} \sum_{j=1}^{M_y} P_{ij}$$

where M_x is the number of spikes in train x , M_y is the number of spikes in train y , and:

- $P_{ij} = 1$ if $0 < x_i - y_j \leq \tau_{ij}$
- $P_{ij} = 1/2$ if $x_i = y_j$
- $P_{ij} = 0$ otherwise.

Then one can define the event synchronization as:

$$Q(x, y) = \frac{c(x|y) + c(y|x)}{\sqrt{M_x M_y}}$$

which becomes a similarity measure via the transformation:

$$d_Q(x, y) = 1 - Q(x, y)$$

Intuitively, this measure provides a normalized count of the number of “synchronous” spikes, where the criteria for determining whether two spikes are synchronous depends upon the smallest inter-spike interval in the two trains.

2.4. ISI-distance

The ISI-distance is computed as follows: Assume a spike train $x = \{x_1, x_2, \dots, x_k\}$ such that for every spike time x_i , $0 < x_i < L$, where L is the length of the recording time (Kreuz et al., 2007). The train is first modified to include artificial “spikes” corresponding to the beginning and end of the recording time, i.e. let $x = \{x_0 = 0, x_1, x_2, \dots, x_k, x_{k+1} = L\}$. Then each train is mapped to a function $f(t)$ as follows:

For $x_i < t < x_{i+1}$, where $i = 0, 1, \dots, k$, let $f(t) = x_{i+1} - x_i$.

Once this has been done for both trains, producing two functions $f(t)$ and $g(t)$, a third function is computed as follows:

$$I(t) = 1 - \frac{\min(f(t), g(t))}{\max(f(t), g(t))}$$

Finally, the value of the measure is determined by:

$$d_{ISI}(x, y) = \frac{1}{L} \int_0^L |I(t)| dt$$

This measure, like the event synchronization measure, does not have any free parameters which the experimenter is required to set. These two measures therefore do not allow for a ‘focus’ of the similarity on particular time scales. It is also unique in that inter-spike intervals rather than spikes form the basis for comparisons.

3. A new spike train similarity measure

3.1. Motivation

A current deficiency in the majority of existing spike train similarity measures is their inability to make a distinction between the following two cases illustrated in Fig. 1. The first two trains A_1 and A_2 are random Poisson trains in which the spikes are mostly uncorrelated, except at 3 times (gray boxes). In contrast, trains B_1 and B_2 share 3 times at which they both burst (gray boxes). Bursts are significant firing events, indicating a strong presynaptic input drive or the ability for the cell to overcome the unreliability of synaptic transmission to other cells at this particular time. For these reasons, a good similarity measure should deem B_1 and B_2 to be closer to each other than trains A_1 and A_2 . As will be shown below (Section

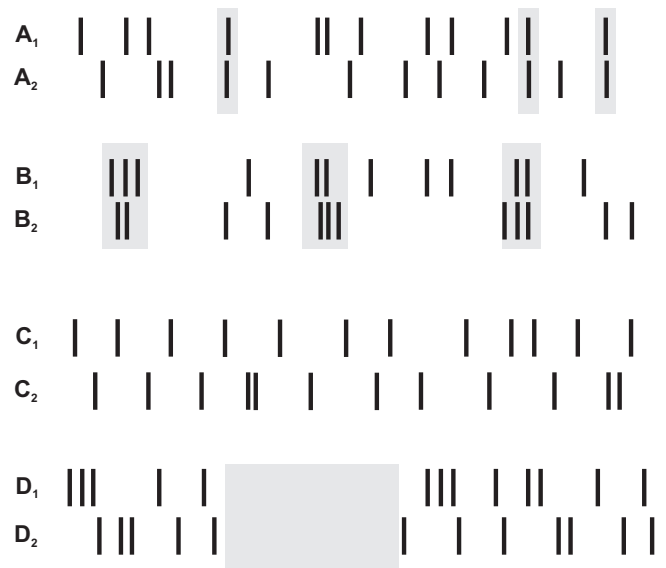


Fig. 1. Trains A–D all have the same firing rate. Trains A_1 and A_2 have 3 instances of single spike near-synchrony (gray boxes). Trains B_1 and B_2 share 3 common bursting times (gray boxes). Trains C_1 and C_2 are uncorrelated, whereas trains D_1 and D_2 are uncorrelated but share a common period of silence. We propose that shared periods of silence and common bursts are meaningful features. A similarity measure should recognize trains B_1 and B_2 as being “closer” to one another than A_1 and A_2 , and should identify D_1 and D_2 as being “closer” to one another than C_1 and C_2 .

4.2), some measures, such as the van Rossum and spike correlation measures, are intrinsically sensitive to bursting, however they do not provide a means for the experimenter to control the extent of this sensitivity.

Fig. 1 also illustrates a second criterion for comparing spike train similarity measures.

Trains C_1 and C_2 have the same average firing rate but little to no correlations in the spike times. These could be representing spike trains resulting from two responses of the same neuron to two different stimuli, or two spontaneously active and uncoupled simultaneously recorded neurons. The trains D_1 and D_2 also have mostly uncorrelated spikes, however a common period of silence has been inserted into the middle of both trains. These common periods of inactivity may be information-rich, and indicative of some active inhibition being imposed on both neurons. Because these trains share this common feature, a good similarity measure should deem them to be “closer” to one another than the first pair C_1 and C_2 , and assign them a lower dissimilarity value. However, most measures would treat these two pairs as equally dissimilar. For these measures, the difference in the dissimilarity values will on average be zero, as will be seen in Section 4.3.2.

We propose two new similarity measures that are designed to specifically address these two issues. The first measure, which will be discussed in detail in Section 3.2.1, is based upon evaluating the normalized correlation between the inter-spike intervals of two trains, in the same manner as the spike correlation measure discussed in Section 2.2.2. The second measure, which we discuss in Section 3.2.2, is a modification of the original spike correlation measure in which convolved trains are subjected to a non-linear transformation as a way of emphasizing bursts. These two measures can be treated as independent measures in and of themselves, however they are both defined in a similar fashion and only take values between zero and one. As a result, they can be combined in a natural way by simply taking their weighted average to yield a single measure (we call the LF measure) that can be sensitive to both bursts and silence, given an appropriate choice of parameters.

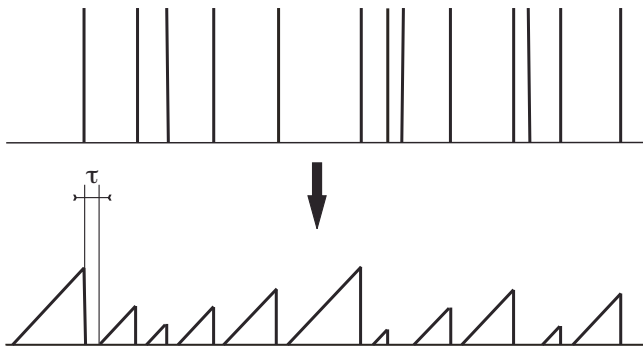


Fig. 2. Illustration of the “silence sensitive” component of the LF measure. Trains are mapped to functions that grow linearly in the time between spikes, but are reset to zero at each spike, and remain at zero for a time τ . The resulting functions are then compared using the standard normalized correlation measure.

3.2. Computing the LF measure

3.2.1. A measure that is sensitive to silences

We compute the silence-sensitive measure by first constructing a mapping from each spike train to a unique function in the following manner:

Assume a spike train $x = \{x_1, x_2, \dots, x_k\}$ such that for every spike time x_i , $0 < x_i < L$, where L is the length of the recording time. For convenience, we can amend the train to include artificial “spikes” corresponding to the beginning and end of the recording time, i.e. let $x = \{x_0 = 0, x_1, x_2, \dots, x_k, x_{k+1} = L\}$. There are other possible ways to treat the issue of what to do at the boundaries, another possibility is to begin the mapping at the first spike in each train, and end it at the last spike. Now we can map the train to a function, $f(t)$ as follows:

For every time t in every inter-spike interval $[x_i, x_{i+1}]$, let

$$f(t) = 0 \text{ for } t \in [x_i, x_i + \tau)$$

and

$$f(t) = t - (x_i + \tau) \text{ for } t \in [x_i + \tau, x_{i+1})$$

Here τ is an adjustable parameter indicating the maximal time interval at which common silences should not be considered significant. This value is set by the experimenter and depends on the specifics of the experiments, very much like the time scale parameters of the convolution and Victor–Purpura metrics. Section 3.4.1 discusses the choice of this parameter in details. Intuitively, beyond τ , the spike trains are mapped to functions that grow in a simple linear fashion between spiking events, but are reset to 0 at each new spike (see Fig. 2). Once both trains have been mapped to functions, the two resulting functions are then compared via the same method as in the spike correlation measure. Namely, given the functions $f(t)$ and $g(t)$, we compute:

$$d_S(f, g) = 1 - \frac{\int_0^L f(t)g(t) dt}{\sqrt{\int_0^L f(t)^2 dt} \sqrt{\int_0^L g(t)^2 dt}}$$

It is important to note that this particular construction of the measure is not symmetric in time, that is, if the two trains being compared were both reversed in time prior to applying the mapping, a different dissimilarity value would be obtained. This implies that the precise timing of a spike that occurs at the end of a long interval will be more important than the precise timing of a spike that occurs at the beginning of such an interval. However, if this time asymmetry is not desired, the measure can easily be made symmetric by applying the mapping both forwards and backwards in time, and taking the average of the resulting two values. This modification will not affect the sensitivity of the measure to long silent periods.

3.2.2. A measure with increased sensitivity to bursts

In order to emphasize the role of shared bursts, we use a modified version of the spike correlation measure. Because of the symmetric nature of the kernel (Gaussian), and the linearity of the convolution operation, the measure intrinsically emphasizes bursts over single spikes. However, the nature of the emphasis depends entirely on the width of the Gaussian kernel, so that bursts of spikes containing ISI smaller than this width will be weighted more than single spikes. Using the standard measures, the width of the kernel is therefore both a measure of the time scale of the similarity and a measure of what is considered a burst. There is no reason, a priori, to couple these quantities.

In order to provide a means of selectively filtering out the contribution of single spikes in favor of bursts, we modify the measure and introduce three new parameters. The first two parameters are physiologically relevant in that they are used to distinguish whether or not a particular fast sequence of spikes is considered a burst. The first parameter b sets the ISI time scale for what is considered a burst. Specifically, this parameter defines the maximum interval between any two consecutive spikes in the same burst. The second parameter n sets the minimum number of spikes that are required to be present in order for such a fast sequence to be considered a burst. The third parameter is simply a scaling factor taking values between 0 and 1, and is used to control the extent to which bursts are emphasized relative to single spikes.

The first step in computing this measure is to convolve the spike trains with a Gaussian kernel, in exactly the same manner as is done in computing the standard spike correlation measure to yield a continuous function $f(t)$. Note that isolated single spikes appear as Gaussians of height equal to 1.

We then apply a piecewise linear transformation $N(x)$ to each of the resulting convolved spike trains, where $N(x)$ is defined as follows:

$$N(f(t)) = H(f(t) - \eta T)(f(t) - \eta T)$$

where here H is the Heaviside theta function ($H(u) = 0$ for $u \leq 0$, $H(u) = 1$ when $u > 0$). Also, T is a function of the parameters n , b and σ , and dictates what is and is not considered a burst, while η in $[0, 1]$ is a scaling factor that determines the extent to which single spike information is discounted in favor of bursts. Setting this parameter equal to 1 is equivalent to completely ignoring all spikes that are not part of a bursts, whereas setting it to 0 recovers the original, unmodified spike correlation measure.

The threshold function $T(n, b, \sigma)$ is specified entirely in terms of the parameters that describe the minimum number of spikes (n) in a burst and the maximum ISI between consecutive spikes (b) in a burst, along with σ (the width of the Gaussian kernel used in the original convolution step). It is important to note here that T is a function of n , b and σ , and not a separate parameter in and of itself. We define $T(n, b, \sigma)$ as the maximum value of the function obtained by convolving a Gaussian kernel with a sequence of exactly n spikes, where each spike is exactly b ms away from the spikes that immediately precede or follow it. This function represents a sort of “worst-case scenario” in the sense that any sequence of spikes that has either more than n spikes or inter-spike intervals less than b will have a higher maximum value. Hence T can be computed from n , b and σ using the following formula:

$$T(n, b, \sigma) = \sum_{k=1}^{k=n} e^{-\frac{(p-kb)^2}{\sigma^2}}$$

where the variable p is obtained by noting that the maximum value of the function that results from convolving a Gaussian kernel with a train of equally spaced spikes is reached either at the midpoint (i.e. the average of the spike times), or at the exact time of one of

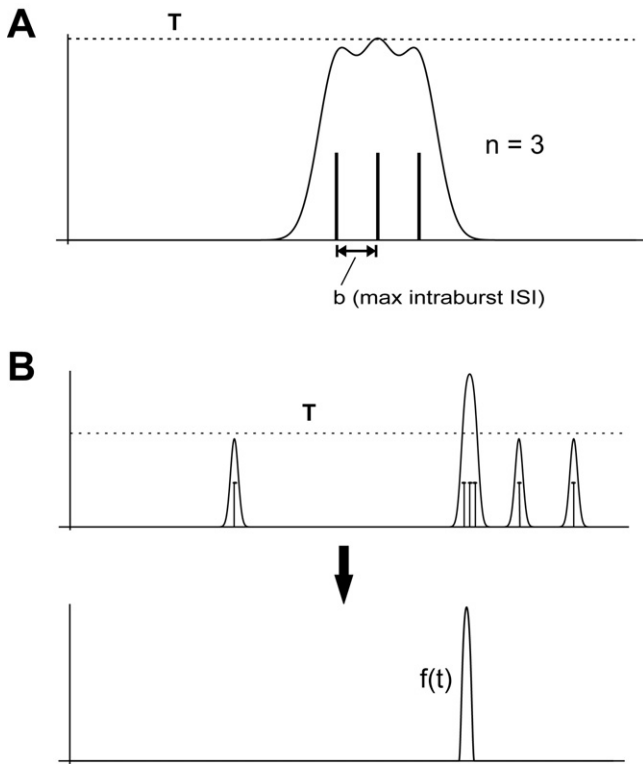


Fig. 3. Burst sensitive measure. (A) Cartoon illustrating how the threshold T is calculated on the basis of the minimum number of spikes η and maximum intra-burst ISI b . T is set to the maximum value of the function that results from convolving a Gaussian kernel with a single burst that meets these minimum requirements. (B) Illustration of the thresholding transformation of a convolved spike train. The top figure shows the result of convolving a given spike train (vertical marks) that contains a burst with a Gaussian kernel. The resulting function is then transformed via a thresholding operation, which leads to the function $f(t)$ shown at the bottom panel. The distance is computed using the resulting functions as inputs to the standard correlation measure. Note that after the convolved spike train is transformed, the single spikes are de-emphasized in favor of the burst.

the two spikes closest to the midpoint. This idea is illustrated in Fig. 3. Since we are looking only at a specific “worst case” example, it is easy to show that p is given either by:

$$p_1(n, b) = \frac{b(n+1)}{2} \text{ or } p_2(n, b) = \frac{b(n+2)}{2} \text{ and hence:}$$

$$T = \max(T(p_1), T(p_2))$$

The convolved spike trains will only take on values less than T when no bursts (as defined by n and b) are present, while the convolved spike trains will take on values greater than T over the course of individual bursts. Hence, the nonlinear transformation will de-emphasize single spikes and over-emphasize bursts. The severity of the transformation is controlled by η , which can be tuned to suit the needs of the experimenter.

Finally, the value of the measure is obtained by applying the standard correlation measure to the non-linearly transformed, convolved spike trains.

$$d_S(f, g) = 1 - \frac{\int_0^L N(f(t))N(g(t)) dt}{\sqrt{\int_0^L N(f(t))^2 dt} \sqrt{\int_0^L N(g(t))^2 dt}}$$

3.3. Combining the two measures

Both intuition and the results of our computational tests indicate that focusing exclusively on the correlation between inter-spike intervals may result in missing important spike-based features such

as bursts. However, by construction, the ISI correlation measure can be combined quite naturally with the spike correlation measure by taking the weighted average of the two. This results in a new measure which can be computed as:

$$d_{LF}(f, g) = W_S d_S(f, g) + W_B d_B(f, g)$$

Here d_S and d_B are, respectively, the silence and burst sensitive measures, and W_S and W_B are weighting parameters such that $W_S + W_B = 1$. This weighting provides flexibility in terms of choosing which feature of the data one wishes to emphasize. By default, one may simply set $W_S = W_B = 0.5$.

3.4. Choosing parameters for a particular data set

Both of the measures that we propose contain free parameters. In this section we suggest a set of guidelines on how to choose these parameters for a given dataset. In addition to these parameters, if the measures are combined, then the respective weights of the two components must be set. The appropriate choice of weights depends on some a priori knowledge of the physiology of the system being studied, and thus there are no general rules regarding how they can be selected. However, the fact that the possible values of the weights are constrained (they must sum to 1) allows them to be systematically varied. Furthermore, a systematic search across the range of weights may reveal the extent to which single spikes, bursts, or silent periods individually contribute to the similarity (or lack thereof) between two spike trains, which could provide useful information regarding how the spike trains in question may be encoding information.

3.4.1. Parameters of the silence sensitive component

The silence sensitive component of our measure maps spike trains to functions that are set to zero at each spike, and after a short delay of length τ begin to grow linearly until the next spike occurs, at which point they are again reset to zero. The duration of this delay, τ , is a free parameter that allows for short or average inter-spike intervals to be ignored, such that only atypically long inter-spike intervals will affect the value of the similarity measure. This is motivated by the idea that such atypically long “silences” are features of interest, perhaps reflecting periods of inhibition upstream. Consequently, the choice of this parameter should reflect some timescale present in the data. For example, it could be set to the average length of the inter-spike intervals in the spike trains under analysis. Setting this parameter to be proportional to the mean inter-spike interval could allow selective emphasis to be placed upon inter-spike intervals that are longer than average. The slope of the linearly increasing function is by default set to 1, and the actual value of this parameter should be of no consequence, since the dot product in d_S is normalized.

3.4.2. Parameters of the burst sensitive component

The bursts sensitive component of our measure has four parameters. As introduced above, the first two are n , the minimum number of spikes for a fast sequence to be considered a burst, and b , the maximum inter-spike interval allowable for two subsequent spikes to be treated as part of the same burst. While both of these parameters can be chosen a priori or explored systematically, it is also possible to estimate them on the basis of the data under analysis. Spike trains in which bursts are a prominent feature typically display a bimodal distribution of ISIs, and the maximum intra-burst ISI can be estimated by examining the first peak in the ISI distribution. Once this parameter has been chosen, a simple burst detection algorithm can be used to determine the typical or minimum number of spikes in a burst. The third parameter η sets the severity of the thresholding that is applied to the convolved spike trains (Fig. 3).

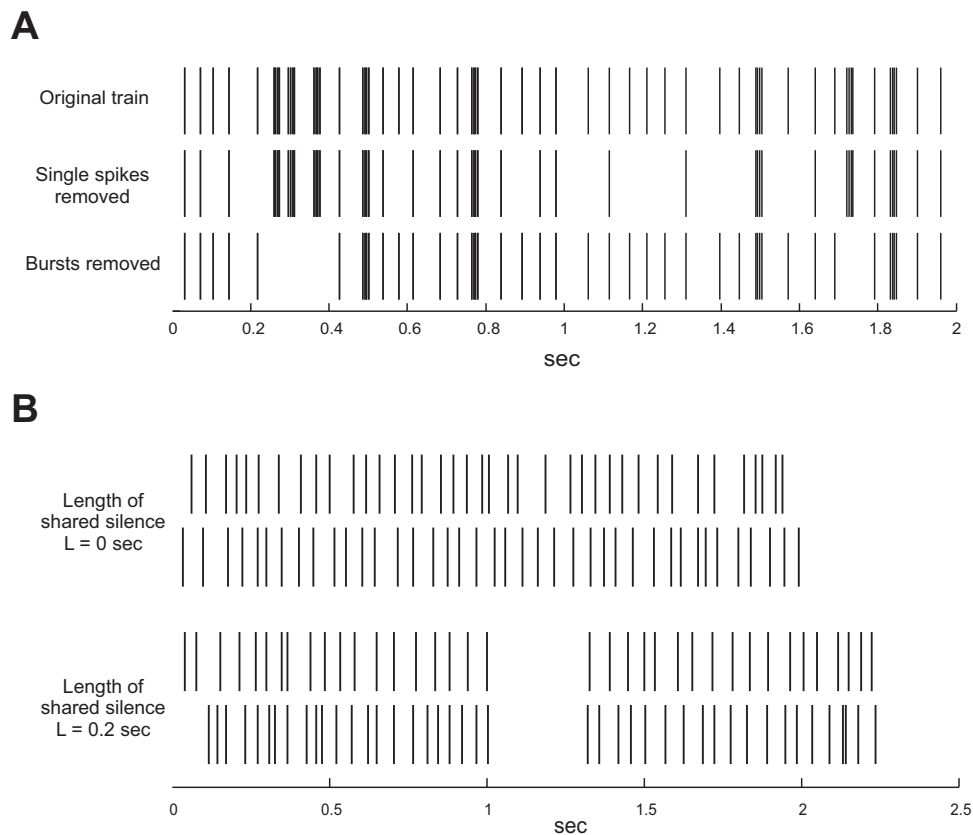


Fig. 4. Example spike trains used in the burst and silence sensitivity tests. (A) Spike trains used for testing burst sensitivity. The top spike train is the original, and contains both bursts and isolated spikes. The middle train has had 12 isolated spikes removed, while in the bottom train 4 bursts (3 spikes each) have been deleted, while isolated spikes are preserved. The burst sensitivity test involved subtracting the distance between the original train and the train with the single spike removals from the distance between the original train and the train with the burst deletions. (B) Illustration of trains used in the silence sensitivity test. The first pair consists of uncorrelated Poisson spike trains. The second pair also contains independently generated Poisson spike trains, but a silent period of length L has been inserted into both. For display purposes the trains here are shorter and have a lower firing rate, the trains used in the actual test were 5 s long and 40 Hz. The silence sensitivity test involves measuring how the pairwise distance between trains changes as a function of the length of their shared silent period.

This parameter takes values from 0 to 1, which makes it easy to vary systematically across the entire range. Intuitively, this parameter controls the extent to which simple spikes are ignored in favor of bursts. When it is set to zero, the resulting measure is simply the original spike correlation measure. Finally, the fourth parameter is inherited from the original spike correlation measure, and is the width of the Gaussian kernel used in the initial convolution step.

4. A comparison of different spike train similarity measures

4.1. Dataset and parameter tuning

For our two empirical tests, we use two different surrogate datasets. The first dataset, used for the burst sensitivity tests, consists of 50, 8-s long spike trains in which the timing of the “events” (either a burst or a single spike) in each train is given by a Poisson process with rate 40 Hz. Each spike train contains 25 bursts, where each burst consists of 4 spikes with an inter-spike interval of 5 ms. The second dataset contained no bursts, but each train had some period of silence of variable length inserted at random onset times.

Several of the published measures discussed above contain one or more free parameters, which typically are used to set the time scale at which two spikes are treated as being synchronous. In general, it was impractical to choose a single, ideal time constant parameter for each measure. As a result, for all measures containing a free parameter which could not be chosen in an obvious manner

from the data set, we simply ran our tests using both a “small” parameter value (typically 2–5 ms as a lower bound) and a large parameter (typically 20–25 ms as an upper bound). Each step in the analysis was repeated with both the short time constant and with the long time constant. We also attempted to fix parameters for the various previously published measures in such a way that our test results would be comparable across measures.

Encouragingly, we found that the qualitative results of all our tests were not affected by the choice of parameters in the different measures, as long as the parameter values used were within the physiological range (i.e. not infinitesimally small or unreasonably large). We do not re-state the results for multiple parameter values, since the differences in the results are minimal. Consequently we report only the results obtained using the “small” set of parameters.

In the case of our new measure, there are several parameters in addition to those used to tune the sensitivity to precise spike timing. For the burst sensitive component, the parameters corresponding to the minimal definition of a burst were chosen to correspond to the pre-chosen values used in the artificial data set. The measure also involves a parameter η that controls the extent to which bursts are emphasized over single spikes. For simplicity, this parameter was set to 0.5, although again, the results did not change appreciably unless the parameter was set to zero, in which case the results corresponding to the original spike correlation measure were reproduced. For the silence-sensitive component, the time scale parameter was set to match the mean ISI of the trains in the data sets being examined, which in this case was 25 ms.

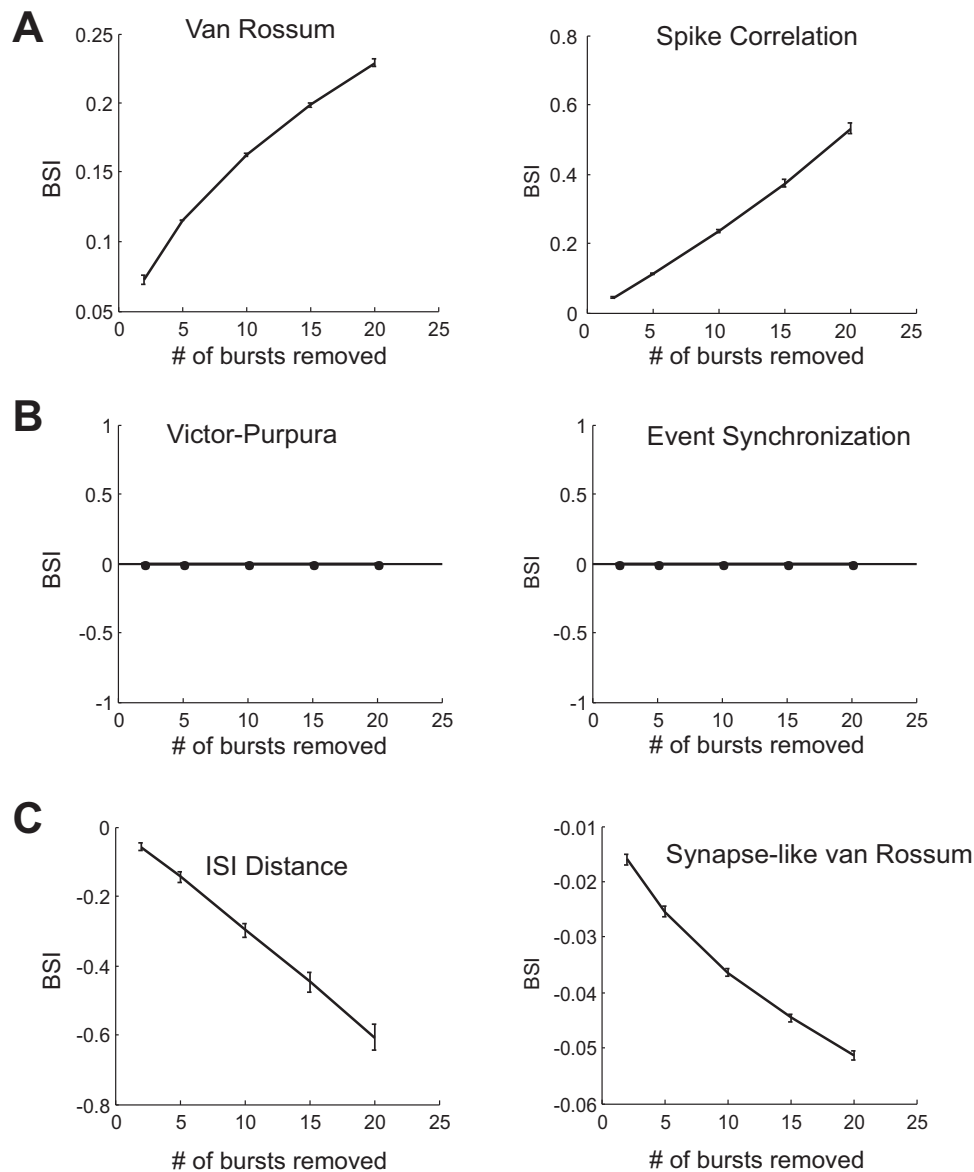


Fig. 5. Burst sensitivity test for previously published measures. Burst sensitivity is defined in terms of the Burst Sensitivity Index (BSI, see methods). A positive BSI implies that a measure emphasizes bursts, a negative BSI implies that a measure emphasizes single spikes. (A) For the van Rossum and spike correlation measures, the BSI is always positive and increases as more bursts are removed. (B) The BSI is always zero for both the Victor–Purpura and event synchronization measures. (C) The ISI-distance and “Synapse-like” variant on van Rossum metric show negative BSI values that decrease when more bursts are removed.

4.2. Testing for sensitivity to bursts

We developed a test to determine the extent to which the various similarity measures account for the presence of bursts. The basic rationale of the test is that the removal of a burst of spikes from a train should be detected by the measure as a more significant deviation from the original train than the removal of an equal number of isolated spikes. In other words, the distance between an arbitrary (burst-containing) train and a modified version of itself with one or more bursts deleted should be greater than the distance between that same train and a version of itself in which an equal number of isolated spikes have been deleted.

4.2.1. Test procedures

The basic procedure by which we tested the metrics for burst sensitivity began as follows: First, using our artificial dataset, we identified all of the bursts in each train. Once all of the bursts for every train had been identified, we constructed two new sets of

trains: one in which a specified number of bursts had been deleted at random from each train, and a second train where a corresponding number of isolated spikes were deleted at random. Thus, for example, if a spike train in the original set (spike train ‘A’) was modified in the second dataset to create a new train (spike train ‘B’), via the deletion of 5 bursts, each of which was made up of 4 spikes, then the corresponding train in the third data set (train ‘C’) will be the result of deleting 20 random isolated spikes from the original train ‘A’. Care was taken to ensure that these isolated spikes were not components of any bursts. These modified data sets were created independently for different numbers of burst removals, specifically 2, 5, 10, 15, and 20. An example set of trains constructed in this way can be seen in Fig. 4A.

When the modified data sets had all been constructed, the measures were compared in the following way: for each measure $d(x,y)$, we took a train ‘A’ from the original data set, its counterpart (train ‘B’) in the data set where some number of bursts had been deleted, and computed $d(A,B)$. We then compared the original train ‘A’ to its

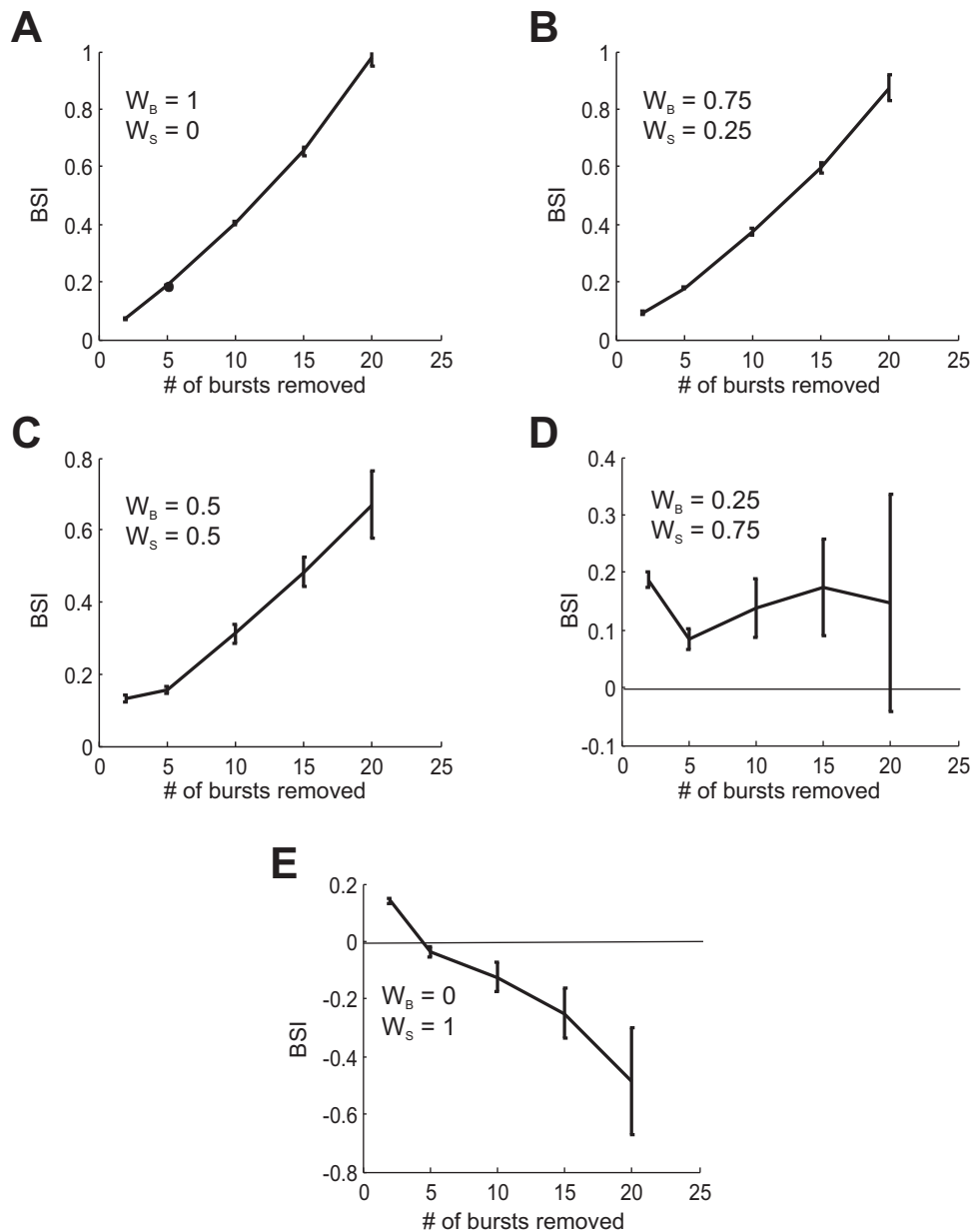


Fig. 6. Burst sensitivity test for the new LF measure (combined burst and silence sensitive components). Here W is the weight of the burst sensitive component, and W is the weight of the silence-sensitive component. Note that $W_B + W_S = 1$. (A) $W_B = 1, W_S = 0$. (B) $W_B = 0.75, W_S = 0.25$. (C) $W_B = 0.5, W_S = 0.5$. (D) $W_B = 0.25, W_S = 0.75$. (E) $W_B = 0, W_S = 1$.

counterpart in the third data set in which the equivalent number of isolated spikes had been deleted (train 'C') by computing $d(A,C)$. We wanted to see if, on average, the measure would detect the deletion of bursts as a more significant modification to the original train than the deletion of single spikes. If a measure has this property, then $d(A,B)$ will be greater than the quantity $d(A,C)$, and hence the quantity:

$$dDiff_A = d(A, B) - d(A, C)$$

will be positive across the entire data set. Since the various measures have different intrinsic scales, it is a priori difficult to compare them directly. To remove this intrinsic scaling, for each measure, we kept track of all the values of $d(A,B)$ and $d(A,C)$, for every spike train in the data set, and identified the maximum of all of these distance values. Then the entire array of $dDiff$ values was normalized by dividing each entry by this maximum distance. We refer to this normalized difference in distances as the Burst Sensitivity Index

(BSI), which became our criteria by which the burst sensitivity of the various measures was evaluated.

4.2.2. Results

Figs. 5 and 6 show the results of this test applied to a set of existing measures and our new measure, respectively. The horizontal axis of each panel shows the number of bursts removed, while the vertical axis displays the average BSI of the measure (see above). The various previously proposed measures responded to bursts in three ways (Fig. 5). The convolution (van Rossum) and spike correlation measures, unambiguously "passed" the test, i.e. the values of $dDiff$ are always greater than zero, and increase as the number of burst removals increases. For the Victor–Purpura and event synchronization measures, the BSI values are identically zero in all cases, indicating that removing bursts and removing spikes are treated as equivalent modifications of the original trains. These measures are therefore insensitive to bursts. Finally, the ISI-distance and synapse-like variant of the convolution metric, display

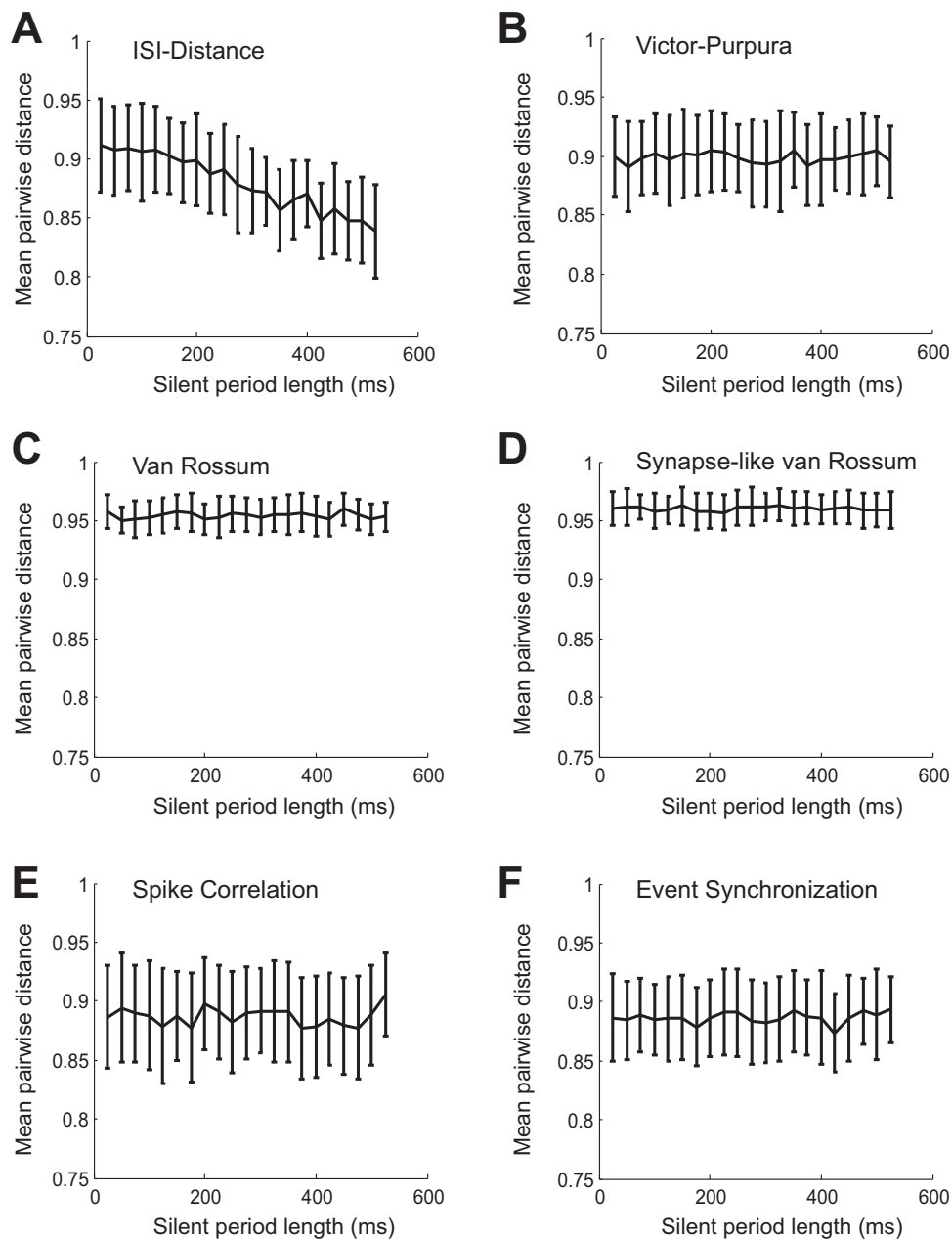


Fig. 7. Results of the silence sensitivity test for previously published measures. For all graphs, the horizontal axis is the length of the silent period (ms). Plotted on the vertical axis is the mean value of the similarity measure for 10 randomly generated pairs of trains sharing a common silent period of length L . (A) ISI-distance. (B) Victor–Purpura metric. (C) Convolution (van Rossum) metric. (D) Synapse-like convolution metric. (E) Spike correlation measure. (F) Event synchronization.

the opposite of burst sensitivity, in that the BSI values are always negative, and decrease as the number of bursts decreases. This indicates that the removal of isolated spikes in these measures represents a more significant change than the removal of bursts.

Fig. 6 shows the result of these tests for the LF measure for several pairs of weights (W_S , W_B) measuring the contribution of the silence and bursts sensitive components. One can see here that when the burst-sensitive component is taken alone (Fig. 5A), or is given a higher (or equal) weight than the silence-sensitive component (Fig. 5B and C), the measure displays burst-sensitive behavior. This sensitivity degrades smoothly when the silence-sensitive component is assigned a higher weight (Fig. 6E).

Qualitatively, it appears that the previously published measures fall into three categories: burst-sensitive (positive BSI values that increase with the number of bursts removed), burst-insensitive (BSI values identically or on average equal to zero), or spike-sensitive

(negative BSI values that decrease with the number of bursts removed). Thus, these tests are able to distinguish whether a particular measure places more, less, or equal emphasis on bursts vs. single spikes, which could be an important practical consideration in choosing which measure to use for a specific data analysis task. We also see that our new measure can display all three behaviors, for different combinations of weights, but that an unbiased choice of weights ($W_B = W_S = 0.5$) yields a clear burst-sensitivity behavior.

4.3. Testing for sensitivity to silences

Recalling the discussion in Section 3.1, we argue that long periods of inactivity may be a meaningful feature of spike trains, and therefore, pairs of trains that share a common period of silence should be considered “closer” than pairs which do not. The extent to which these trains are close should depend upon the length of the

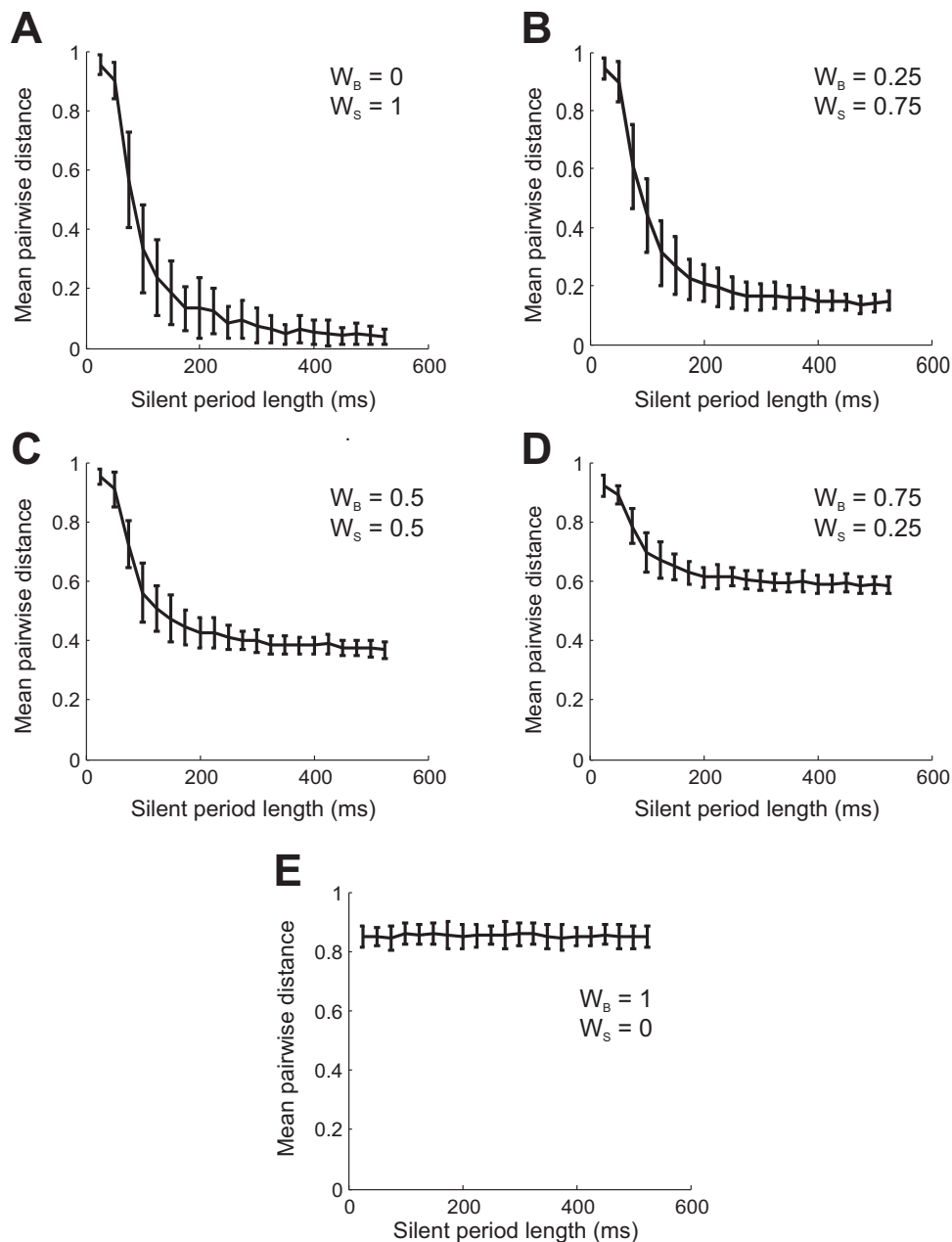


Fig. 8. Results of the silence sensitivity test for the new LF measure. Here W_B is the weight of the burst sensitive component, and W_S is the weight of the silence-sensitive component. Note that $W_B + W_S = 1$. (A) $W_B = 0$, $W_S = 1$. (B) $W_B = 0.25$, $W_S = 0.75$. (C) $W_B = 0.5$, $W_S = 0.5$. (D) $W_B = 0.75$, $W_S = 0.25$. (E) $W_B = 1$, $W_S = 0$.

shared silence, i.e. the longer the shared silent period, the “closer” the trains. Furthermore we claim that ideally, as the silent period grows longer, the value of the similarity measure should approach zero. In a more general sense, similarity measures should be able to distinguish between cases in which a shared inactive period existed and those in which it did not. Here we implement a simple test based upon this intuitive premise.

4.3.1. Test description and explanation

We generated a set of 50 pairs of 5 s long spike trains with uncorrelated Poisson statistics, all with firing rates set to approximately 40 Hz. We then inserted a silent period of some length L at the same point (2.5 s) into every train by simply shifting all spikes occurring after 2.5 s forward in time by L ms. An illustration of pairs of trains both with and without a shared silent period can be seen in Fig. 4B. We computed the mean distance for each pair of trains using the

similarity measures discussed above. This was repeated 10 times for each value of L ranging from $L = 0$ ms to $L = 500$ ms in increments of 25 ms. As in the test for burst sensitivity, for each measure, we normalized all values by dividing by the maximum distance obtainable by any two trains in the artificial dataset. Our objective was to assess the manner in which the mean distance between random, uncorrelated Poisson trains would decrease as a function of the shared silent period.

4.3.2. Results

The results of this test can be observed in Figs. 7 and 8. For a measure to be deemed “silence-sensitive” in the context of this test, the pairwise distances between the random trains should decrease as the length of the shared silent period increases. First note that all the measures we tested failed to behave in this way, although the ISI-distance seemed to display a slight downward trend as a

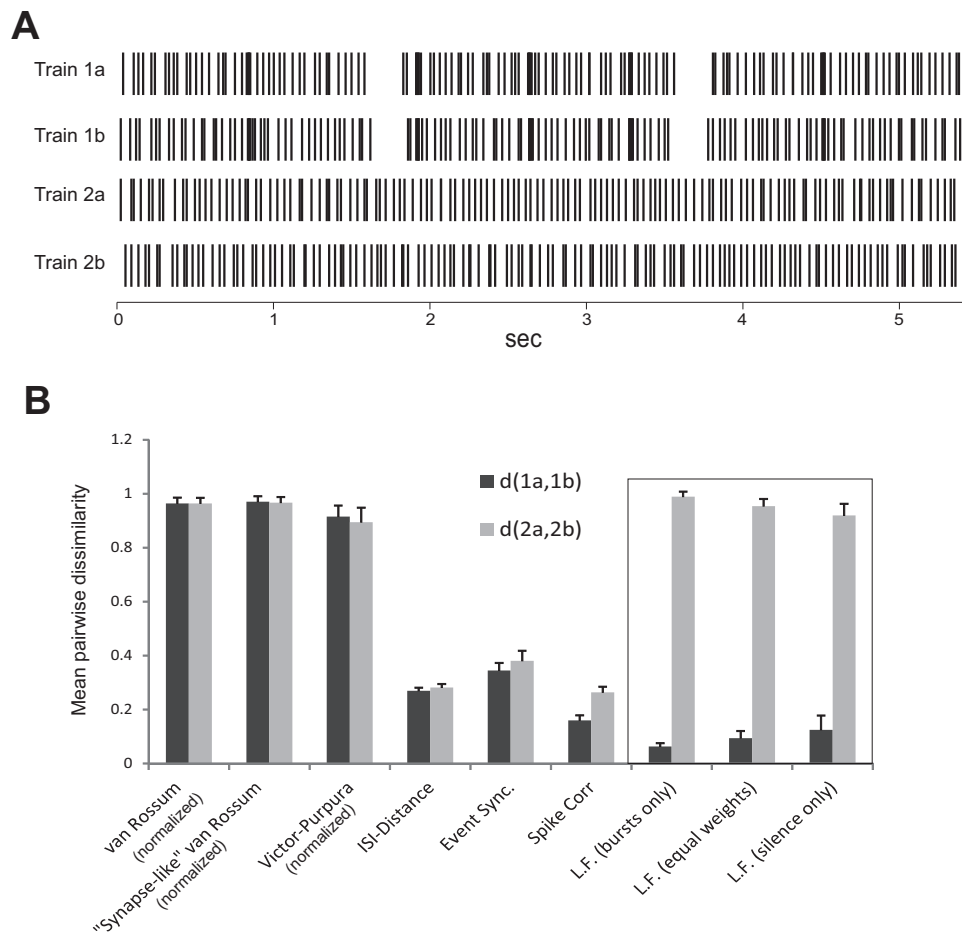


Fig. 9. Comparisons of the measures. (A) Sample test trains. Trains 1a and 1b contain random spikes, but share 3 bursts and two periods of silence. Single spikes are uncorrelated, and assumed to be “noise.” Trains 2a and 2b are simply uncorrelated, random trains. All four trains have approximately the same rate. (B) Comparison of the distance between the first pair vs. the distance between the second pair averaged across 25 sets of such trains using all measures. Error bars show standard deviations. For the van Rossum, synapse-like van Rossum, and Victor–Purpura measures, the distances are all normalized by dividing by the largest of the 25 different distance values in the set, prior to computing the mean. The LF distances (with 3 different sets of weights) outperform the others in their ability to distinguish between the first and second pairs of spike trains.

function of distance (Fig. 7A). It is arguable, however, whether this trend is pronounced enough to reliably detect shared periods of silence. This is evident in the negligible average difference in the values obtained via this measure between pairs of trains with no shared silent period and trains sharing a silent period that is 10 times the length of the average ISI (specifically, 250 ms and 25 ms, respectively). Also, note the small dynamic range of values (Y axes, at most between 0.8 and 1) in all these tests.

In contrast, with our new measure, even when the silence-sensitive component is weighted less heavily than the burst-sensitive component (Fig. 8A–D), the distinction is quite clear at these time scales. Some downward trend is observed in all cases except when the burst-sensitive component is considered alone (Fig. 8E). Again, the unbiased LF measure ($W_B = W_S = 0.5$, Fig. 8C) gives a quite acceptable sensitivity to silences.

4.4. Comparing existing measures to the LF measure for spike trains containing bursts and periods of inhibition

To explicitly illustrate the potential utility of the new measure, we constructed a new set of 50 pairs of spike trains. The spikes in the first 25 pairs of spike trains are uncorrelated, but there are 3 bursts and 2 silent periods that are aligned (Fig. 9A, trains 1a and 1b). The second set of 25 pairs consists of two totally uncorrelated Poisson random trains (Fig. 9A, trains 2a and 2b). We then

applied each of the similarity measures discussed here to both sets of pairs, to study the extent to which the measures could distinguish between the two cases. Fig. 9 shows the results (note the results for the van Rossum and Victor–Purpura measures have been scaled for comparison, see below). These results clearly show that most previously published measures do not distinguish between the two cases. In other words, the inclusion of common bursts and common periods of inhibition are ignored, and the two spike trains are considered as dissimilar as two random Poisson processes (Fig. 9B). As expected and unlike other measures, the new LF measure we propose here could effectively distinguish between the two cases (Fig. 9B, rightmost 3 pairs of bars). Note that the unbiased (equal weights) measure outperforms all the other measures tested.

4.5. Other considerations

In addition to considering the sensitivity of the various measures to specific, computationally relevant spike train features, it is also important to examine other aspects of these measures and how they may or may not be relevant for practical applications. In particular, we shall discuss briefly here the implications (or lack thereof) of whether a given measure is a metric in the strict mathematical sense, whether or not a measure has tunable free parameters, and whether it is bounded or unbounded.

4.5.1. Metrics vs. non-metrics

All of the measures considered here can be thought of as functions $d(x,y)$, where x and y are spike trains. In all cases the output of the function lies within the set of real numbers (or some subset of them, i.e. the interval $[0,1]$). Such a function is called a *metric* if it satisfies the following conditions:

1. $d(x,y) \geq 0$, for all x, y , and $d(x,y) = 0$ if and only if $x = y$ (non-negativity)
2. $d(x,y) = d(y,x)$ for all x and y (symmetry)
3. For all x, y , and z , $d(x,y) \leq d(x,z) + d(z,y)$ (triangle inequality)

The van Rossum and Victor–Purpura measures are metrics in the true sense of the word, in that they satisfy all three constraints (van Rossum, 2001; Victor, 2005). The other measures (with the exception of the ISI Distance, which satisfies the triangle inequality but may fail to satisfy the first constraint in a special case) all satisfy the first two but not the third constraint (see Appendix).

We note this distinction as an interesting means of classifying the measures along theoretical lines. However, for practical purposes, it remains an open question as to whether it actually matters if a similarity measure is a metric in the strict sense of the word or not. Since the non-metric measures typically only fail to satisfy the triangle inequality, examining the importance of this constraint in practical applications such as spike train clustering appears to be a promising route to understanding the significance of this distinction. However a formal investigation of this possibility has not yet been done, and is beyond the scope of this paper.

4.5.2. Free parameters

The convolution, Victor–Purpura, spike correlation, Silence Correlation, and the LF measures all contain one or more tunable free parameters. In contrast, the event synchronization measure and ISI-distance do not contain any free parameters. It was argued that these parameter-free measures are advantageous, making them self-adaptive to the intrinsic time scales of the data and hence making them more objective (Kreuz et al., 2007; Quiroga et al., 2002). Furthermore, the choice of free parameters may sometimes be difficult and non-intuitive and may require that the parameter space be explored, which adds an extra step to the analysis. However, it could be argued that having free parameters gives the experimenter more control over the interpretation of the data analysis. In many cases, it may be advantageous to have this additional flexibility. For example, when a free parameter controls the sensitivity of a measure to precise spike timing, the ability to vary such a parameter systematically may be useful in investigating hypotheses about the importance of spike timing in particular neural systems at particular time scales. Furthermore, in the measure we propose, one may adjust the weights of the different components of the measure in such a way as to be sensitive to either bursts or silent periods, on the basis of some prior knowledge or hypotheses of which features may be more important.

4.5.3. Bounded and unbounded measures

Some measures are bounded, meaning that the maximum dissimilarity value assigned to any pair of spike trains is less than some finite limiting value. Examples of such measures include the correlation-based measures, the LF measure as well as the ISI distance. Conversely, other measures are in principle unbounded, so that given any arbitrarily large number, it is possible to construct a pair of spike trains such that the dissimilarity value of that pair is greater than that number. The Victor–Purpura and van Rossum metrics are examples of such measures. This distinction may be important for practical applications. For example, in clustering applications, bounded measures may benefit from the application of a nonlinear filtering of the dissimilarity values, in order to accen-

tuate small difference in the values (Fellous et al., 2004). Fully understanding the practical consequences of using bounded vs. unbounded measures remains an interesting question to explore.

5. Conclusions

Experimental and theoretical work, from sensory (Lestienne, 2001) to central areas (Averbeck and Lee, 2004) have shown that neural responses in isolation, or seen within an ensemble can carry significant information about stimuli. The basis of the neural code is however in general unknown. The most common and parsimonious method to understand the manner in which a neuron's activity contains information about a stimulus is to present that stimulus multiple times, record the activity of the cell, and assess what (if anything) is similar from trial to trial. Inherently in this assessment is the notion of similarity between neural responses. Whether the information is in single spikes, bursts, instantaneous firing rates, or a combination thereof depends on the specifics of the experiment and of the neuron recorded.

Here we proposed and implemented a set of methods aimed at assessing the extent to which spike train similarity measures can capture similarities in bursts and period of inhibition between spike trains. While others have attempted to classify measures in terms of their response to other spike train features, such as rate, or spike "jitter" (Paiva et al., 2010), our analysis is unique in its focus on bursts and silences as computationally relevant features. It reveals clear differences between existing measures, and allows for clearer distinctions to be made between them. The importance of shared silence in particular has typically been neglected despite the potential physiological relevance that this feature may have in some systems (De Schutter and Steuber, 2009). Similarly, most measures are not sensitive to the occurrence of bursts, which many have thought of as significant features of neural activity (Kepecs and Lisman, 2003; Lisman, 1997).

We introduced a new spike train measure that is unique in its ability to group together spike trains on the basis of shared silences. In addition, we suggest a modification of the spike correlation measure of Schreiber et al. (2003) that emphasizes bursts, and show how these two measures can be combined. Given that these features may be of particular interest in certain experimental contexts, we hope to extend the range of tools available to experimenters. Further work will involve subjecting this new measure to the types of analysis conducted in (Paiva et al., 2010) to further elucidate its properties. Another promising avenue is to compare this new measure to others on the basis of clustering performance, as has been done with other measures (Kreuz et al., 2007).

Acknowledgements

Thanks to the Fellous laboratory and Kevin Lin for helpful discussions and proofreading. We also thank Romain Brasselet for assistance in proving the triangle inequality for the ISI distance. Work was supported in part by NIH training grant GM084905, and in part by a NSF/NIH Collaborative Research in Computational Neuroscience Program IIS-1010172.

Appendix A. The correlation-based and event synchronization measures do not satisfy the triangle inequality

A.1. Correlation based metrics

We show that this measure (Schreiber et al., 2003) does not satisfy the triangle inequality by means of a counterexample. Consider three spike trains, labeled A, B, and C, respectively. Train A consists

of a single spike at time $t=1$, train B consists of a single spike at time $t=3$, and train C consists of two spikes, one at time $t=1$ and a second at time $t=3$. We would like to show that, when $d(x,y)$ is a correlation-based distance, that

$$d(A, B) > d(A, C) + d(B, C)$$

For simplicity we assume a boxcar kernel function of unit height and unit width. This is to make the calculations easier, and this counterexample should generalize to a wider class of kernel functions.

Using this assumption:

$$d(A, B) = 1 - \frac{\langle f(A), f(B) \rangle}{|f(A)|_2 |f(B)|_2} = 1 - \frac{0}{1} = 1$$

And

$$d(A, C) = 1 - \frac{\langle f(A), f(C) \rangle}{|f(A)|_2 |f(C)|_2} = 1 - \frac{1}{\sqrt{2}} < \frac{1}{2}$$

$$\text{Similarly, } d(B, C) = 1 - \frac{1}{\sqrt{2}} < \frac{1}{2}$$

Therefore the triangle inequality does not hold, and hence the correlation-based measure is not a true metric.

A.2. Event synchronization

One can show that the event synchronization measure (Quiroga et al., 2002) does not satisfy the triangle inequality by using the same counterexample. Take A, B, and C as before. Then:

$$d_{ES}(A, B) = 1 - \frac{0}{\sqrt{2}} = 1$$

and

$$d_{ES}(A, C) = d_{ES}(B, C) = 1 - \frac{(1/2) + (1/2)}{\sqrt{2}} = 1 - \frac{1}{\sqrt{2}} < \frac{1}{2}$$

Appendix B. A note on the ISI-distance

It remains unclear whether the ISI-distance (Kreuz et al., 2007) satisfies all of the properties of a metric. It is worth noting that two periodic trains that are different by a phase lag will be assigned a distance very near to zero. In fact only the boundary conditions prevent this distance from being identically zero, and one could in principle construct choices for the boundary conditions of this measure where this may be the case. If this is done, then the first condition for a measure to be a true metric, that $d(x,y)=0$ iff $x=y$, would be violated. However, Kreuz et al. have recently introduced a new, spike-centered similarity measure that is not subject to these problems (Kreuz et al., 2011).

One can show, however, that this measure does satisfy the triangle inequality. We reproduce here an outline of the proof courtesy of Romain Brasselet (personal communication):

Let X, Y and Z be three spike trains defined on the time interval $[0, T]$, and let t be an arbitrary time within this interval. We first show that the triangle inequality holds locally at t . The first step in calculating the ISI distance is to construct functions for each spike train such that the value of each function at time t is given by the length of the inter-spike interval in which t lies. Let f, g , and h be such functions defined for X, Y , and Z , respectively. We assume, without loss of generality, that at time $t, f(t) \leq g(t) \leq h(t)$. We must then show that:

- (i) $1 - f(t)/h(t) \leq 1 - f(t)/g(t) + 1 - g(t)/h(t)$
- (ii) $1 - f(t)/g(t) \leq 1 - f(t)/h(t) + 1 - g(t)/h(t)$
- (iii) $1 - g(t)/h(t) \leq 1 - f(t)/g(t) + 1 - f(t)/h(t)$

To see that the first statement is true, note that it is equivalent to the expression: $0 \leq g(t)f(t) - h(t)f(t) - g(t)g(t) + h(t)g(t)$, which can be re-written as $0 \leq (f(t) - g(t))(g(t) - h(t))$, which is always true, since $f(t) \leq g(t) \leq h(t)$. Propositions (ii) and (iii) can also be easily checked using a similar argument. Since the triangle inequality holds for all t in $[0, T]$, it will continue to hold after integrating both sides over t from 0 to T , and hence, $d_{ISI}(X, Y) \leq d_{ISI}(X, Z) + d_{ISI}(Z, Y)$ for arbitrary spike trains X, Y and Z .

References

- Aronov D, Victor JD. Non-Euclidean properties of spike train metric spaces. *Physical Review E* 2004;69.
- Averbeck BB, Lee D. Coding and transmission of information by neural ensembles. *Trends in Neurosciences* 2004;27:225–30.
- Bi GQ, Poo MM. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience* 1998;18:10464–72.
- De Schutter E, Steuber V. Patterns and pauses in Purkinje cell simple spike trains: experiments, modeling and theory. *Neuroscience* 2009;162:816–26.
- Fellous JM, Tiesinga PH, Thomas PJ, Sejnowski TJ. Discovering spike patterns in neuronal responses. *Journal of Neuroscience* 2004;24:2989–3001.
- Gunay C, Smolinski TG, Lytton WW, Morse TM, Gleeson P, Crook S, et al. A method for discovering spatio-temporal spike patterns in multi-unit recordings. In: Smolinski TG, Milanova MG, Hassani AE, editors. *Applications of computational intelligence in biology: current trends and open problems*. Springer; 2008. p. 325–59.
- Gusfield D. *Algorithms on strings, trees, and sequences: computer science and computational biology*. UK: Cambridge University Press; 1997.
- Houghton C. Studying spike trains using a van Rossum metric with a synapse-like filter. *Journal of Computational Neuroscience* 2009;26:149–55.
- Hunter JD, Milton JG, Thomas PJ, Cowan JD. Resonance effect for neural spike time reliability. *Journal of Neurophysiology* 1998;80:1427–38.
- Jolivet R, Kobayashi R, Rauch A, Naud R, Shinomoto S, Gerstner W. A benchmark test for a quantitative assessment of simple neuron models. *Journal of Neuroscience Methods* 2008;169:417–24.
- Kara P, Reinagel P, Reid RC. Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron* 2000;27:635–46.
- Kepecs A, Lisman J. Information encoding and computation with spikes and bursts. *Network-Computation in Neural Systems* 2003;14:103–18.
- Kreuz T, Chicharro D, Greschner M, Andrzejak RG. Time-resolved and time-scale adaptive measures of spike train synchrony. *Journal of Neuroscience Methods* 2011;195:92–106.
- Kreuz T, Haas JS, Morelli A, Abarbanel HDI, Politi A. Measuring spike train synchrony. *Journal of Neuroscience Methods* 2007;165:151–61.
- Kruskal PB, Stanis JJ, McNaughton BL, Thomas PJ. A binless correlation measure reduces the variability of memory reactivation estimates. *Statistics in Medicine* 2007;26:3997–4008.
- Lestienne R. Spike timing, synchronization and information processing on the sensory side of the central nervous system. *Progress in Neurobiology* 2001;65:545–91.
- Lisman JE. Bursts as a unit of neural information: making unreliable synapses reliable. *Trends in Neurosciences* 1997;20:38–43.
- Masud MS, Borisyuk R. Statistical technique for analysing functional connectivity of multiple spike trains. *Journal of Neuroscience Methods* 2011;196:201–19.
- Maurer AP, McNaughton BL. Network and intrinsic cellular mechanisms underlying theta phase precession of hippocampal neurons. *Trends in Neuroscience* 2007;30:325–33.
- Paiva ARC, Park I, Principe JC. A comparison of binless spike train measures. *Neural Computing and Applications* 2010;19:405–19.
- Quiroga RQ, Kreuz T, Grassberger P. Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. *Physical Review E* 2002;66.
- Reinagel P, Reid RC. Precise firing events are conserved across neurons. *Journal of Neuroscience* 2002;22:6837–41.
- Rossant C, Goodman DF, Platkiewicz J, Brette R. Automatic fitting of spiking neuron models to electrophysiological recordings. *Frontiers in Neuroinformatics* 2010;4:2.
- Schreiber S, Fellous J-M, Tiesinga PH, Sejnowski TJ. A new correlation-based measure of spike timing reliability. *Neurocomputing* 2003;52–54:925–31.
- Skaggs WE, McNaughton BL. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* 1996;271:1870–3.
- van Rossum MC. A novel spike distance. *Neural Computation* 2001;13:751–63.
- Victor JD. Spike train metrics. *Current Opinion in Neurobiology* 2005;15:585–92.
- Victor JD, Purpura KP. Metric-space analysis of spike trains: theory, algorithms and application. *Network-Computation in Neural Systems* 1997;8:127–64.
- Wiskott L, Fellous J-M, Kruger N, von der Malsburg C. Face recognition by elastic bunch graph matching. *IEEE Spectrum* 1997;19:775–9.