ORIGINAL ARTICLE

A comparison of binless spike train measures

António R. C. Paiva · Il Park · José C. Príncipe

Abstract Several binless spike train measures which avoid the limitations of binning have been recently been proposed in the literature. This paper presents a systematic comparison of these measures in three simulated paradigms designed to address specific situations of interest in spike train analysis where the relevant feature may be in the form of firing rate, firing rate modulations, and/or synchrony. The measures are first disseminated and extended for ease of comparison. It also discusses how the measures can be used to measure dissimilarity in spike trains' firing rate despite their explicit formulation for synchrony.

Keywords Distance measures · Spike train analysis

1 Introduction

Spike train similarity measures or, conversely, dissimilarity measures are important tools to quantify the relationship among pairs of spike trains. Indeed, the definition of such a measure is essential for classification, clustering or other forms of spike train analysis. For example, just by using a distance (dissimilarity) measure it is possible to decode the applied stimulus from a spike train [1–4]. This is possible because the measure is used to quantify how much the spike train differs from a "template" or sets of reference spike trains for which the input stimulus is known and, hence, classified accordingly (see Fig. 1). However, naturally the success of this classification is dependent on the discriminative ability of the measure.

A traditional measure of similarity between two spike trains is to measure the (empirical) cross-correlation of the binned spike trains [5]. If the bin size is large compared to the average inter-spike interval (ISI), binning provides a crude estimate of the instantaneous firing rate [6]. In this perspective, and assuming ergodicity, cross-correlation is a similarity measure of the estimated intensity functions, which is plausible only under the hypothesis that neurons encode information through modulation of the firing rates [6, 7]. However, recent studies have found evidence that information may also be encoded in the precise timing of action potentials [8–10]. Again, binning has also been employed with small bin sizes [9, 11–14]. But the small bin size can lead to boundary effects due to the quantization of the spike times [15] and estimation problems, which require longer averaging windows where stationarity must be assumed. For these reasons, measures based on binning of the spike trains are discouraged.

To avoid the difficulties associated with binning and to prevent estimation errors of information when binning is done, several binless spike train dissimilarity measures have been proposed. These include the Victor-Purpura's (VP) distance [1, 2],[1] van Rossum's distance [16], the correlation-based measure proposed by Schreiber et al. [17], the inter-spike interval (ISI) distance [18], the

A. R. C. Paiva (✉) · I. Park · J. C. Príncipe
Department of Electrical and Computer Engineering,
University of Florida, Gainesville, FL 32611, USA
e-mail: arpaiva@cnel.ufl.edu

I. Park
e-mail: memming@cnel.ufl.edu

J. C. Príncipe
e-mail: principe@cnel.ufl.edu

---

[1] Actually, in their works, Victor Purpura [1, 2] proposed not one but several spike train distances. Namely, $D^{spike}[q]$, $D^{interval}[q]$, $D^{count}[q]$ and $D^{motif}[q]$. In this study, and as in most references to their work, VP distance refers to $D^{spike}[q]$.

Fig. 1 Typical experimental setup for classification using spike train dissimilarities. In this setup the measure is utilized to quantify the dissimilarity between the new spike train and the reference spike trains for each of the stimulus. Then, the unlabeled stimulus is inferred as the one corresponding to the class for which the new spike train has smaller average dissimilarity
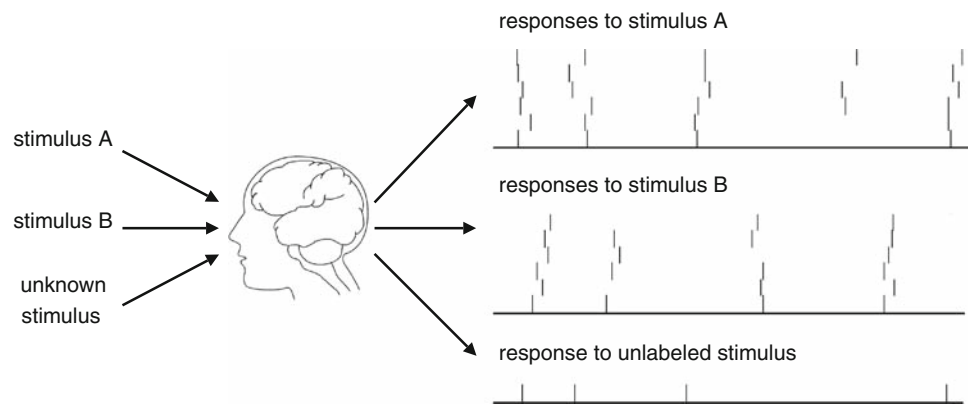
reliability (similarity) measure proposed by Hunter and Milton [19], and the metrics recently introduced by Houghton [4] which generalize the van Rossum's distance. Moreover, the VP and van Rossum's distance have been generalized to simultaneously measure the distance between sets of spike trains [20–22].

These measures have been utilized in different neurophysiological paradigms (see Victor [23] and references within) and for different tasks, such as classification [1, 2] and clustering of spike trains [24–26]. However, in our opinion, in neither of these works was the choice of the measure used properly argued versus the candidates. Some of these metrics have been compared previously [18, 22, 27, 28]; however, the comparison has often focused on a single paradigm or recording dataset. Although these comparisons are clearly important, the comparison on a particular setting is not informative of the general asymptotic properties of the metrics for different cases. In this paper we analyze the discriminative performance of these spike train measures in multiple paradigms, from firing rate to synchrony. By analyzing the discrimination capability we are indirectly measuring which spike train metric is most informative about the differences between spike trains. In this paper we compare three from the aforementioned metrics: the VP and van Rossum's distance, and Schreiber's et al.'s correlation measure. These measures were chosen because, (1) they have been utilized for data analysis and/or in the development of machine learning algorithms, (2) they have been utilized in neurophysiological studies, and (3) they generalize directly across timescales.

An important issue for data analysis and for this comparison is that of the "spike encoding hypothesis." The three measures considered here were motivated by the perspective of a neuron as a coincidence detector [29], a fact which is explicitly stated by Victor and Purpura [1] and Victor [23] with regard to the VP distance, and in the presentation done by Schreiber et al. [17]. Yet, it is still unclear how neurons encode information [30]; if through

firing rate modulation, precise spike timing, or both (cf. de Ruyter van Steveninck et al. [31]). Nevertheless, as is shown here, despite the assumed hypothesis of temporal coding, these binless spike train distances do not distinguish these assumptions of the neural code. In fact, given that the "smoothing parameter" is appropriately set these measures can cope with any of the above-referred neural codes. Roughly speaking, the smoothing parameter controls the time-scale at which the distance analysis is done, much like the bin size, but without time quantization. If the time-scale associated with the smoothing parameter is small compared with the average ISI, then the measures quantify how closely the spikes from one spike train occur to spikes in the other. Conversely, if this time-scale is large, then the measures approximate a form of dissimilarity in the firing rate, and ultimately in the spike count [2, 16]. This multi-scale nature of the measures is explored in the analysis, where the measures are compared for their discriminative characteristics with regard to distinguishing features in spike trains such as firing rate, firing modulation phase, and synchrony.

As the presentation in Sect. 2 shows, each measure implies a given kernel function that measures similarity in terms of a single pair of spike times. Another issue addressed here was to what extent this kernel affects the performance of each measure. This factor was also explored, by first analyzing how the measures can be formulated in general, showing results for a set of four kernels commonly used. By evaluating the measures using all of these kernels we intended to make the comparison kernel independent, and show the connection and generality of the principles used in designing the measures.

2 Binless spike train dissimilarity measures

In this section, the VP distance, van Rossum's distance, and Schreiber's correlation measure are briefly reviewed. To aid the practitioner, we also discuss some recent

developments that allow for efficient implementation of the distances.

2.1 Victor-Purpura's distance

Historically, Victor-Purpura's (VP) distance [1, 2] was the first binless distance measure proposed in the literature. Two key design considerations in the definition of this distance were that it needed to be sensitive to the absolute spike times and would not correspond to Euclidean distances in a vector space. The first consideration was due to the fact that the distance was initially to be utilized to study temporal coding and its precision in the visual cortex. As stated by the authors, the basic hypothesis is that a neuron is not simply a rate detector but can also function as a coincidence detector. Within this respect the distance is well motivated by neurophysiological ideas. The second consideration is because, in this way it is "not based on assumptions about how responses should be scaled or combined" [1].

The VP distance defines the distance between spike trains as the cost in transforming one spike train into the other. Three elementary operations in terms of single spikes are established: moving one spike to perfectly synchronize with the other, deleting a spike, and inserting a spike. Once a sequence of operations is set, the distance is given as the sum of the cost of each operation. The cost in moving a spike at t_m to t_n is $q|t_m−t_n|$, where q is a parameter expressing how costly the operation is. Because a higher q means that the distance increases more when a spike needs to be moved, the distance as a function of q expresses the precision of the spike times. The cost of deleting or inserting a spike is set to one.

Since the transformation cost for the spike trains is not unique, the distance is not yet well defined. Moreover, this criterion needs to guarantee the fundamental axioms of a distance measure for any spike trains S_i, S_j and S_k:

(a) Symmetry: $d(S_i, S_j) = d(S_j, S_i)$
(b) Positiveness: $d(S_i, S_j) \geq 0$, with equality holding if and only if $S_i = S_j$
(c) Triangle inequality: $d(S_i, S_j) \leq d(S_i, S_k) + d(S_k, S_j)$.

To ensure the triangle inequality and uniqueness of the distance between any two spike trains, the sequence which yields the *minimum cost* in terms of the operations is used. Therefore, the VP distance between spike trains S_i and S_j is defined as

$$d_{\text{VP}}(S_i, S_j) \triangleq \min_{C(S_i \leftrightarrow S_j)} \sum_l K_q\left(t^i_{c_i[l]}, t^j_{c_j[l]}\right), \tag{1}$$

where $C(S_i \leftrightarrow S_j)$ is the set of all possible sequences of elementary operations that transform S_i to S_j, or vice-versa, and $c_{(\cdot)}[\cdot] \in C(S_i \leftrightarrow S_j)$. That is, $c_i[l]$ denotes the index of

the spike time of S_i manipulated in the lth step of a sequence. $K_q(t^i_{c_i[l]}, t^j_{c_j[l]})$ is the cost associated with the step of mapping the $c_i[l]$th spike of S_i at $t^i_{c_i[l]}$ to $t^j_{c_j[l]}$, corresponding to the $c_j[l]$th spike of S_j, or vice-versa. In other words, K_q is a distance metric between two spikes.

Suppose two spike trains with only one spike each, the mapping between the two spike trains is achieved through the three aforementioned operations and the distance is given by

$$K_q(t^i_m, t^j_n) = \min\{q|t^i_m − t^j_n|, 2\}$$
$$= \begin{cases} q|t^i_m − t^j_n|, & |t^i_m − t^j_n| < 2/q \\ 2, & \text{otherwise.} \end{cases} \tag{2}$$

This means that if the difference between the two spike times is smaller than $2/q$, then the cost is linearly proportional to their time difference. However, if the spikes are farther apart, it is less costly to simply delete one of the spikes and insert it at the other location. Shown in this way, K_q is nothing but a scaled and inverted triangular kernel applied to the spike times. This perspective of the elementary cost function is key to extend this cost to other kernels, as we will present later.

At first glance it would seem that the computational complexity would be unbearable because the formulation of the algorithm describes the distance in terms of a full search through all allowed sequences of elementary operations. Luckily, efficient dynamic programming algorithms were developed which reduce it to a more manageable level of $\mathcal{O}(N_i N_j)$ [1], i.e., the scaled product of the number of spikes in the spike trains whose distance is being computed.

2.2 van Rossum's distance

Similar to the VP distance, the distance proposed by van Rossum [16] utilizes the full resolution of the spike times. However, the approach taken is conceptually simpler and more intuitive. Simply put, van Rossum's distance [16] is the Euclidean distance between the exponentially filtered spike trains.[2]

A spike train S_i defined on the time interval $[0, T]$ and spike times $\{t^i_m: m = 1, ..., N_i\}$ can be written as a continuous-time signal as a sum of time-shifted impulses,

$$S_i(t) = \sum_{m=1}^{N_i} \delta(t − t^i_m), \tag{3}$$

where N_i is the number of spikes in the recording interval. In this perspective, the filtered spike train is the sum of the

---

[2] Filtered spike trains correspond to what is often referred to as "shot noise" in the point processes literature [32, Sect. 16.3].

time-shifted impulse response of the smoothing filter, $h(t)$, and can be written as

$$f_i(t) = \sum_{m=1}^{N_i} h(t - t_m^i). \qquad (4)$$

For the smoothing filter, van Rossum [16] proposed to use a causal decaying exponential function, written mathematically as $h(t) = \exp(-t/\tau)u(t)$, with $u(t)$ being the Heaviside step function (illustrated in Fig. 2). The parameter τ in van Rossum's distance controls the decay rate of the exponential function and, hence, the amount of smoothing that is applied to the spike train. Thus, it determines how much variability in the spike times is allowed and how it is combined into the evaluation of the distance. In essence, τ plays the reciprocal role of the q parameter (Eq. 2) for the VP distance. The choice for the exponential function was due to biological considerations. The idea is that an input spike will evoke a post-synaptic potential at the stimulated neuron which, simplistically, can be approximated through the exponential function [6].

In terms of their filtered counterparts, it is easy to define a distance between the spike trains. An intuitive choice is the usual Euclidean distance, $L^2([0, T])$, between square integrable functions. The distance between spike trains S_i and S_j is therefore defined as

$$d_{vR}(S_i, S_j) \triangleq \frac{1}{\tau} \int_0^\infty \left[f_i(t) - f_j(t) \right]^2 dt. \qquad (5)$$

The van Rossum distance also seems motivated by the perspective of a neuron as a coincidence detector. This perspective may be induced by the definition. When two spike trains are "close" more of their spikes will be synchronized, which translates into a smaller difference of the filtered spike trains and therefore yields a smaller distance. Despite this formulation, the multi-scale quantification capability of the distance was noticed before by van Rossum [16]. The behavior transitions smoothly from a count of non-coincidence spikes to a difference in spike count as the kernel size τ is increased. This perspective can be obtained from Eq. (4) if one

notices that it corresponds to kernel intensity estimation with function h [33]. In broader terms one can thus think of van Rossum's distance as the $L^2([0,\infty))$ distance between the estimated *intensity functions* at time scale τ. Thus, van Rossum's distance can be used to measure the dissimilarity between spike trains at any time scale simply by selecting τ appropriately.

Evaluation of the distance is numerically straightforward, as it directly implements the previous equations. But explicit computation of the filtered spike trains and integral in a discrete-time simulation is computationally more intensive than evaluating the VP distance which depends only on the number of spikes in the spike trains. Furthermore, the computation burden would increase proportional to the length of the spike trains and inversely proportional to the simulation step. However, as shown by Paiva et al. [34], and utilized in Paiva et al. [25], the van Rossum's distance can be evaluated in terms of a computationally effective estimator with order $\mathcal{O}(N_i N_j)$, given as

$$d_{vR}(S_i, S_j) = \frac{1}{2} \left[\sum_{m=1}^{N_i} \sum_{n=1}^{N_i} L_\tau(t_m^i - t_n^i) + \sum_{m=1}^{N_j} \sum_{n=1}^{N_j} L_\tau(t_m^j - t_n^j) \right] + \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} L_\tau(t_m^i - t_n^j), \qquad (6)$$

where $L_\tau(\cdot) = \exp(-|\cdot|/\tau)$ is the Laplacian kernel. Thus, this distance can be computed with the same computational complexity as the VP distance. It should be remarked that the Laplacian kernel plays a different role than that of the smoothing filter mentioned earlier. The smoothing filter describes the rate of change of the distance, whereas the Laplacian kernel contributes directly to the distance, much like the kernel K_q. This follows because the Laplacian kernel arises from the autocorrelation function (with integration over time) of the smoothing filter.

2.3 Schreiber et al. induced divergence

The third dissimilarity measure considered in this paper is derived from the correlation-based measure proposed by

Fig. 2 a Spike train and **b** corresponding filtered spike train utilizing a causal exponential function (Eq. 4)
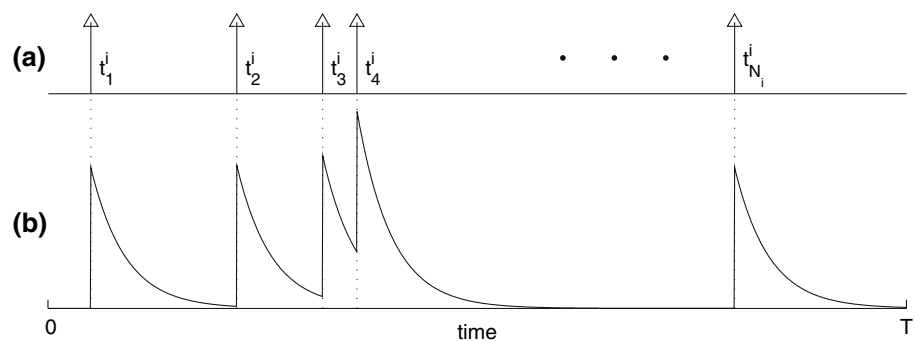
Schreiber et al. [17]. Like van Rossum's distance, the correlation measure was also defined in terms of the filtered spike trains. Instead of using the causal exponential function, however, Schreiber and coworkers proposed to utilize the Gaussian kernel. The core idea of this correlation measure is the concept of dot product between the filtered spike trains. Actually, in any space with an inner product two types of quadratic measures are naturally induced: the Euclidean distance, and a correlation coefficient-like measure, due to the Cauchy-Schwarz inequality. The former corresponds to the concept utilized by van Rossum, whereas the latter is conceptually equivalent to the definition proposed by Schreiber and associates. So, in this sense, the two measures are directly related. Nevertheless, this measure is non-Euclidean like the VP distance, since it is an angular metric [34].

In defining the measure, write the filtered spike trains as

$$g_i(t) = \sum_{m=1}^{N_i} G_{\sigma/\sqrt{2}}(t - t_m^i), \tag{7}$$

where $G_{\sigma/\sqrt{2}}(t) = \exp[-(t)^2/\sigma^2]$ is the Gaussian kernel. Notice the dependence of the filtering on σ which plays in this case the same role as τ in the exponential function in van Rossum's distance, and is inversely related to q in VP distance. Assuming a discrete-time implementation of the measure, then the filtered spike trains can be seen as vectors, for which the usual dot product can be used. Based on this, the Cauchy-Schwarz (CS) inequality guaranties that

$$|\vec{g}_i \cdot \vec{g}_j| \leq \|\vec{g}_i\| \|\vec{g}_j\|, \tag{8}$$

where g_i, g_j are the filtered spike trains in vector notation, and $\vec{g}_i \cdot \vec{g}_j$ and $\|\vec{g}_j\|$, $\|\vec{g}_i\|$ denotes the filtered spike trains dot product and norm, respectively. The norm is given as usual by $\|\vec{g}_i\| = \sqrt{\vec{g}_i \cdot \vec{g}_i}$. Because by construction the filtered spike trains are non-negative functions, the dot product is also non-negative. Consequently, rearranging the Cauchy-Schwarz inequality yields the correlation coefficient-like quantify,

$$r(S_i, S_j) = \frac{\vec{g}_i \cdot \vec{g}_j}{\|\vec{g}_i\| \|\vec{g}_j\|}, \tag{9}$$

proposed by Schreiber et al. [17]. Notice that like the absolute value of the correlation coefficient, $0 \leq r(S_i, S_j) \leq 1$. Equation (9), however, takes the form of a *similarity measure*. Utilizing the upper bound, a dissimilarity can be easily derived,

$$d_{CS}(S_i, S_j) = 1 - r(S_i, S_j) = 1 - \frac{\vec{g}_i \cdot \vec{g}_j}{\|\vec{g}_i\| \|\vec{g}_j\|}. \tag{10}$$

In light of the perspective presented here we shall hereafter refer to d_{CS} as the CS dissimilarity measure.

The CS dissimilarity, like the previous two measures, can also be utilized directly to measure dissimilarity in the firing rates of spike trains merely by choosing a large σ. Similar to van Rossum's distance, this is shown explicitly in the formulation of the measure in terms of the inner product of intensity functions, with the time scale specified by σ.

An important difference with regard to the VP and van Rossum's distances needs to be pointed out. d_{CS} is *not* a distance measure. Although it is trivial to prove that it verifies the symmetry and positiveness axioms, the measure does not fulfill the triangle inequality. Nevertheless, since it guaranties the first two axioms, it is what is called in the literature a semi-metric [35].

In the definition of the measure, and more importantly in the utilization of the concept of the dot product, the filtered spike trains were considered finite-dimensional vectors [17]. If this naïve approach is taken, then the computational complexity in evaluating the measure would suffer from the same limitations as the direct implementation of van Rossum's distance. But, like the latter, a data-effective method can be obtained in the same way to compute the distance [34],

$$d_{CS}(S_i, S_j)$$
$$= 1 - \frac{\sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \exp\left[-\frac{(t_m^i - t_n^j)^2}{2\sigma^2}\right]}{\sqrt{\left(\sum_{m,n=1}^{N_i} \exp\left[-\frac{(t_m^i - t_n^i)^2}{2\sigma^2}\right]\right)\left(\sum_{m,n=1}^{N_j} \exp\left[-\frac{(t_m^j - t_n^j)^2}{2\sigma^2}\right]\right)}}. \tag{11}$$

Evaluating the distance using this expression has a computational complexity of order $\mathcal{O}(N_i N_j)$, just like the two previously presented measures. Note that, as pointed out for the van Rossum distance, the kernel in Eq. (11) plays a different role than the smoothing filter in Eq. (7). The fact that one obtains the gaussian function in both cases is only because the autocorrelation of a gaussian function is another gaussian function (with different kernel size).

3 Extension of the measures to multiple kernels

From the previous presentation it should be observable that each measure was originally associated with a particular kernel function which measures the similarity between two spike times. Interestingly, the kernel function is found to be different in all three situations. In any case, it is remarkable that the measures are conceptually different, irrespective of the differences in the kernel function. To further complete our study we were also interested in verifying the impact of different kernel functions in each measure. In this section we further develop these ideas. In particular, we present the

details involved in replacing the default kernel for each dissimilarity measure and, whenever pertinent, intuitively explain how this approach reveals the connections between the measures. It should be remarked that similar considerations have been presented previously by Schrauwen and Campenhout [27], although under a different analysis paradigm.

In Sect. 2.1 the distance between two spikes for the VP distance is defined through the function K_q. This distance represents the minimum cost in transforming a spike into the other in terms of the elementary operations defined by Victor and Purpura. As briefly pointed out, this function is equivalent to having

$$K_q(t_m^i, t_n^j) = 2\left[1 - \kappa_{1/q}(t_m^i - t_n^j)\right], \tag{12}$$

where κ_α is the triangular kernel with parameter α,

$$\kappa_\alpha(x) = \begin{cases} 1 - |x|/(2\alpha), & |x| < 2\alpha \\ 0, & |x| \geq 2\alpha, \end{cases} \tag{13}$$

which is, in essence, a *similarity* measure of the spike times. Notice that this perspective does not change the non-Euclidean properties of the VP distance since those properties are a result of the condition in Eq. (1). Put in this way, it seems obvious that other kernel functions may be used in place of the triangular kernel, as pointed out by [2, Sect. 2.2.4].

The kernel in the VP distance is not explicit in the definition. Rather, is the cost associated with the three elementary operations. Similarly, in van Rossum's distance and the CS dissimilarity measure the perspective of a kernel operating on spike times is not explicit in the definition. The difference, however, is that the kernel arises naturally as an immediate byproduct of the filtering of the spike trains. This result is noticeable in the expressions for efficient evaluation given by Eqs. (6) and (11). Again, and just as proposed for the VP distance, alternative kernel functions can be utilized in the evaluation of the dissimilarity measures instead of the proposed kernel by the original construction.

As said earlier, each of the spike train measures considered here was defined with a different kernel function. To provide a systematic comparison, each measure was evaluated with four kernels: the triangular kernel in Eq. (13), and the Laplacian, Gaussian, and rectangular kernels,

$$\text{Laplacian: } \kappa_\tau(x) = \exp\left(-\frac{|x|}{\tau}\right) \tag{14}$$

$$\text{Gaussian: } \kappa_\sigma(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{15}$$

$$\text{Rectangular: } \kappa_\alpha(x) = \begin{cases} 1, & |x| < \alpha \\ 0, & |x| \geq \alpha, \end{cases} \tag{16}$$

For reference, these four kernels and induced distance function K_q in terms of each of the kernels are depicted in Fig. 3. In this way each measure was evaluated for the kernel it was originally defined for and the other kernels for a fair comparison.

Note that if other kernels were to be chosen these would have to be symmetric, maximum at the origin, and always positive, to ensure the symmetry and positiveness of the measure. Additionally, for the VP distance to be well posed, the kernels need to be concave so that the optimization in Eq. 1 guarantees the triangle inequality. However, the Gaussian and rectangular kernels are not concave and thus for these kernels the VP measure is a semi-metric. This means that when these kernels are used the resulting dissimilarity is not a well-defined distance. Nevertheless, we utilize these kernels here regardless, since our aims are to study the effect of this kernel of the discrimination ability, and also to compare the measures apart from this factor.

It is interesting to consider the consequences in terms of the filtered spike trains associated with the choice of each of the four kernels presented. As motivated by van Rossum [16], the biological inspiration behind the idea in utilizing filtered spike trains is that they can be thought of as postsynaptic potentials evoked at the efferent neuron. In this sense, kernels are mathematical representations of the interactions involved with this idea. As shown before, the Laplacian function results from the autocorrelation of a one-sided exponential function. Likewise, the Gaussian function (with kernel size scaled by $\sqrt{2}$) results from its own autocorrelation. The triangular results from the autocorrelation of the rectangular function. The smoothing function associated with the rectangular function corresponds to the inverse of the square root of a sinc function. Based on these observations it seems to us that the Laplacian kernel is, from the four kernels considered, the most biologically plausible.

4 Results

In this section results are shown for the three dissimilarity measures introduced in terms of a number of parameters: kernel function, firing rate, kernel size, and, in the last paradigm presented, synchrony and jitter of the absolute spike times.

Three simulation paradigms are studied. In each paradigm we will be interested in verifying how well can the dissimilarity measurements discriminate differences in spike trains with regard to a specific feature. To quantify the discrimination ability of each measure in a *scale-free* manner, the results shall be presented and analyzed in terms of a discriminant index defined as

**Fig. 3 a** Kernels utilized in this study and **b** the corresponding K_q function induced by each of the kernels
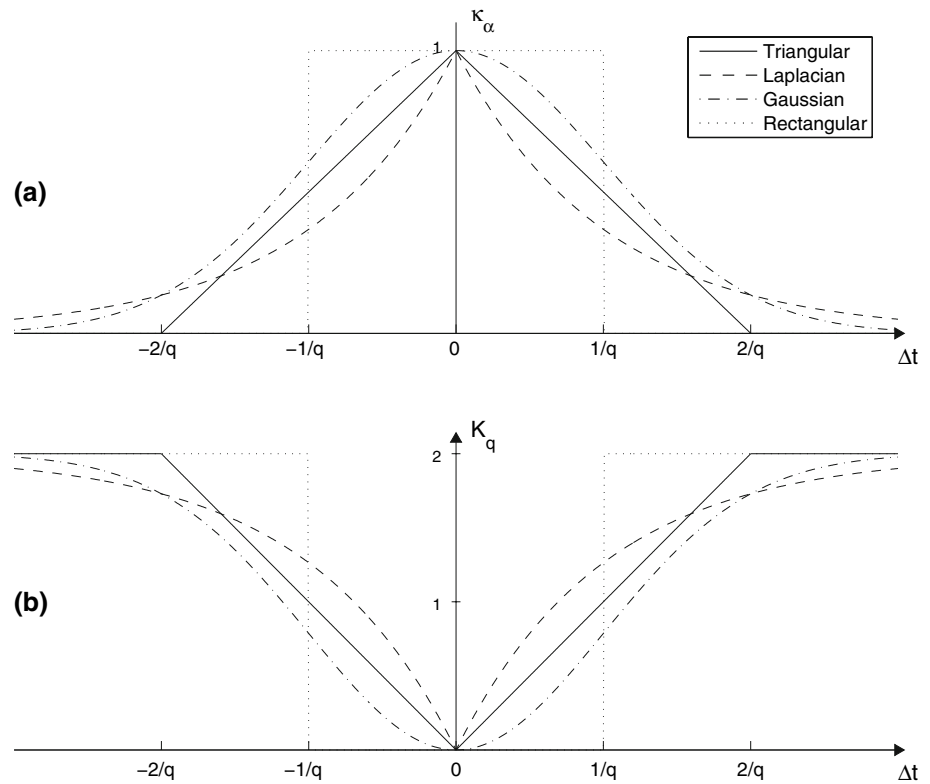
$$v(A, B) = \frac{\bar{d}(A, B) - \bar{d}(A, A)}{\sqrt{\sigma_{\bar{d}}^2(A, B) + \sigma_{\bar{d}}^2(A, A)}}, \qquad (17)$$

where $\bar{d}(A,A)$, $\bar{d}(A,B)$ denotes the mean of the dissimilarity measure evaluated between spike trains from the same and different situations, respectively, and $\sigma_{\bar{d}}^2(A, A)$, $\sigma_{\bar{d}}^2(A, B)$ denotes the corresponding variances. The use of a discriminant index was chosen instead of, for example, ROC plots for ease of display and analysis, and because in this way the conclusions drawn here are classifier-free. $v(A, B)$ quantifies how well the outcome of the measure can be used to differentiate the situation A from the situation B. In terms of Fig. 1, think that $\left[\bar{d}(A,A), \sigma_{\bar{d}}^2(A,A)\right]$ characterizes the distribution of the dissimilarity measure evaluation for spike trains in response to stimulus A, and $\left[\bar{d}(A,B), \sigma_{\bar{d}}^2(A,B)\right]$ characterizes a similar distribution but in which the dissimilarities are evaluated between a spike train evoked by stimulus A and a spike train evoked by stimulus B. This is supported by the fact that the distribution of the evaluation of the measures can be reasonably fitted to a Gaussion pdf (see Fig. 4). Therefore, the discriminant index is utilized in the simulated experimental paradigms to compare how well the dissimilarity distinguishes spike trains generated under the same versus different conditions, with regard to a parameter specifying how different spike trains from different stimulus are. The

discriminant index v is conceptually similar to that of the Fisher linear discriminant cost [36]. A key difference, however, is that the absolute value is not used. This is because negative values of the index correspond to unreasonable behavior of the measure; that is, the dissimilarity measure yields smaller values between spike trains generated under difference conditions than spike trains generated for the same condition. Obviously, intuitively the desired behavior is that the dissimilarity measure yields a minimum for spike trains generated similarly.

For contrast to the binless dissimilarity measures considered, results are also presented for a *binned* cross-correlation based dissimilarity measure, denoted d_{CC}. This measure is defined just like the CS dissimilarity through Eq. (10). The difference is that now $\vec{g}_i$ and $\vec{g}_j$ are finite dimensional vectors corresponding to the binned spike trains and, thus, $\vec{g}_i \cdot \vec{g}_j$ is the usual Euclidean dot product between two vectors. Notice that d_{CC} is in essence equivalent to quantize the spike times (with quantization step equal to the bin size) and evaluating d_{CS} using the rectangular kernel, with kernel size equal to half the bin size. Hence, d_{CC} can be alternatively computed utilizing Eq. (11). The former approach is more advantageous for large bin size, whereas the latter is computationally more effective for smaller bin size (larger number of bins).
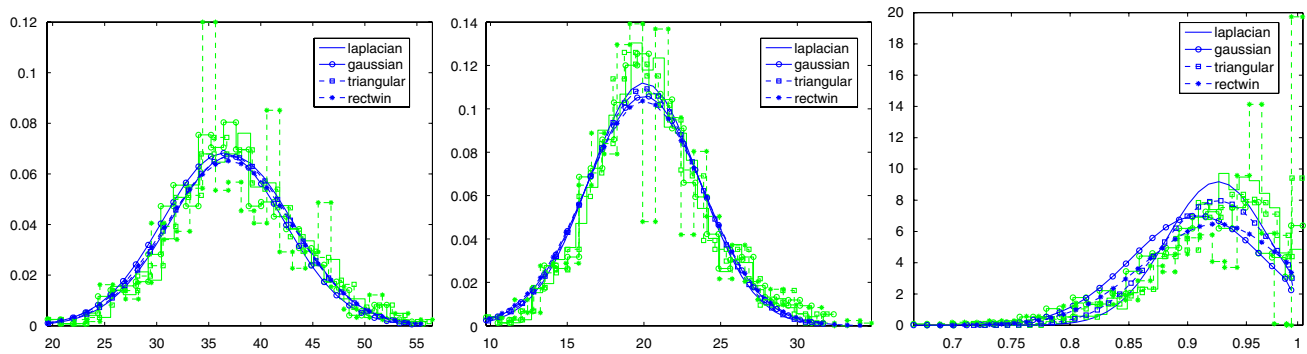
Fig. 4 Estimated pdf of the measures for each kernel considered (*green/gray*) and corresponding fitted Gaussian pdf (*blue/black*). The pdf was estimated by a normalized histogram of the evaluation of the measure with kernel/bin size 2 ms for 1,000 pairs of uncorrelated spike trains with mean firing rate 20 spk/s and jitter noise of 3 ms. (See paradigm in Sect. 4.3 for details)

4.1 Discrimination of difference in firing rate

The first paradigm considered was intended to analyze the characteristics of each measure with regard to the firing rate of one spike train relatively to another of fixed firing rate. The key point was to understand if the measures could be used to differentiate two spike trains of different firing rates. This is important because neurons have been found to often encode information in the spike train firing rates [6, 7, 37]. To simplify matters, all spike trains were simulated as 1-s-long homogeneous Poisson processes. Although this simplification is unrealistic, it allows a first analysis without the introduction of additional effects due to modulation of firing rates in the spike trains. The scenario where the firing rates are modulated over time is considered in the next section. Another important factor in the analysis is the spike train length. Naturally, in this scenario, the discrimination of the measures is expected to improve as the spike train length is increased since more information is available. In practice, however, this value is often smaller than 1 s. Thus, the value was chosen as a compromise between a reasonable value for actual data analysis and good statistical illustration of the properties of each measure.

In our study, simulations were made for each dissimilarity measure utilizing each of the four described kernels. In each case, the analysis was repeated for four kernel sizes, 10, 25, 50, and 100 ms. The kernel sizes used were purposely chosen relatively large since firing rate information can only be extracted at a slower time scale. The results are shown in Fig. 5 in terms of mean values ±1 SD, as estimated from 1,000 randomly generated spike train pairs. For each pair, one of the spike trains was generated at a reference firing rate of 20 spk/s, whereas the firing rate of the other was one of 2.5–40 spk/s, in steps of 2.5 spk/s.

Utilizing the estimated statistics, the discrimination provided by the measures was evaluated in terms of the discrimination index v (Eq. 17) with regard to the results when both spike trains have firing rate 20 spk/s. The results

are shown in Fig. 6. The results for VP and van Rossum's distances reflect the importance of the choice of time scale, materialized in the form of the kernel size selection. Only for the largest kernel size (100 ms) did these two distances behave as we intuitively expected. This is not surprising since discrimination can only occur if the dissimilarity can incorporate an estimation of the firing rate in its evaluation. Even for this kernel size the discriminant index curve shows a small bias towards smaller firing rates. This is natural since the optimal kernel size is infinity, and smaller kernel size tends to result in bias related to the total number of spikes. The discrimination behavior of the CS dissimilarity, however, seems nearly insensitive to the choice of the kernel size. On the other hand, when the firing rate is above the reference the outcome is not the desired. For lower firing rates, the positive discrimination index is due to the presence of the norm of the spike train in the denominator of the definition. One of the most remarkable observations is the consistency of the results for each measure throughout the four kernels. Although there are subtle differences in values they seem to be of importance only for small kernel sizes for which, as pointed out, the results are not significant anyway. Comparing with the results for the CC dissimilarity we verify the resemblance with the CS dissimilarity. Like the latter, the CC dissimilarity also is unable to correctly distinguish increases in firing rate of one spike train with respect to the other.

4.2 Discrimination of phase in firing rate modulation

The scenario depicted in the previous paradigm is obviously simplistic. In this case study, an alternative situation is considered in which spike trains must be discriminated through differences in their *instantaneous* firing rates. Spike trains were generated as 1-s-long inhomogeneous Poisson processes with instantaneous firing rate given by sinusoidal waveforms of mean 20 spk/s, amplitude 10 spk/s, and frequency 1 Hz. A pair of spike trains was generated at a time
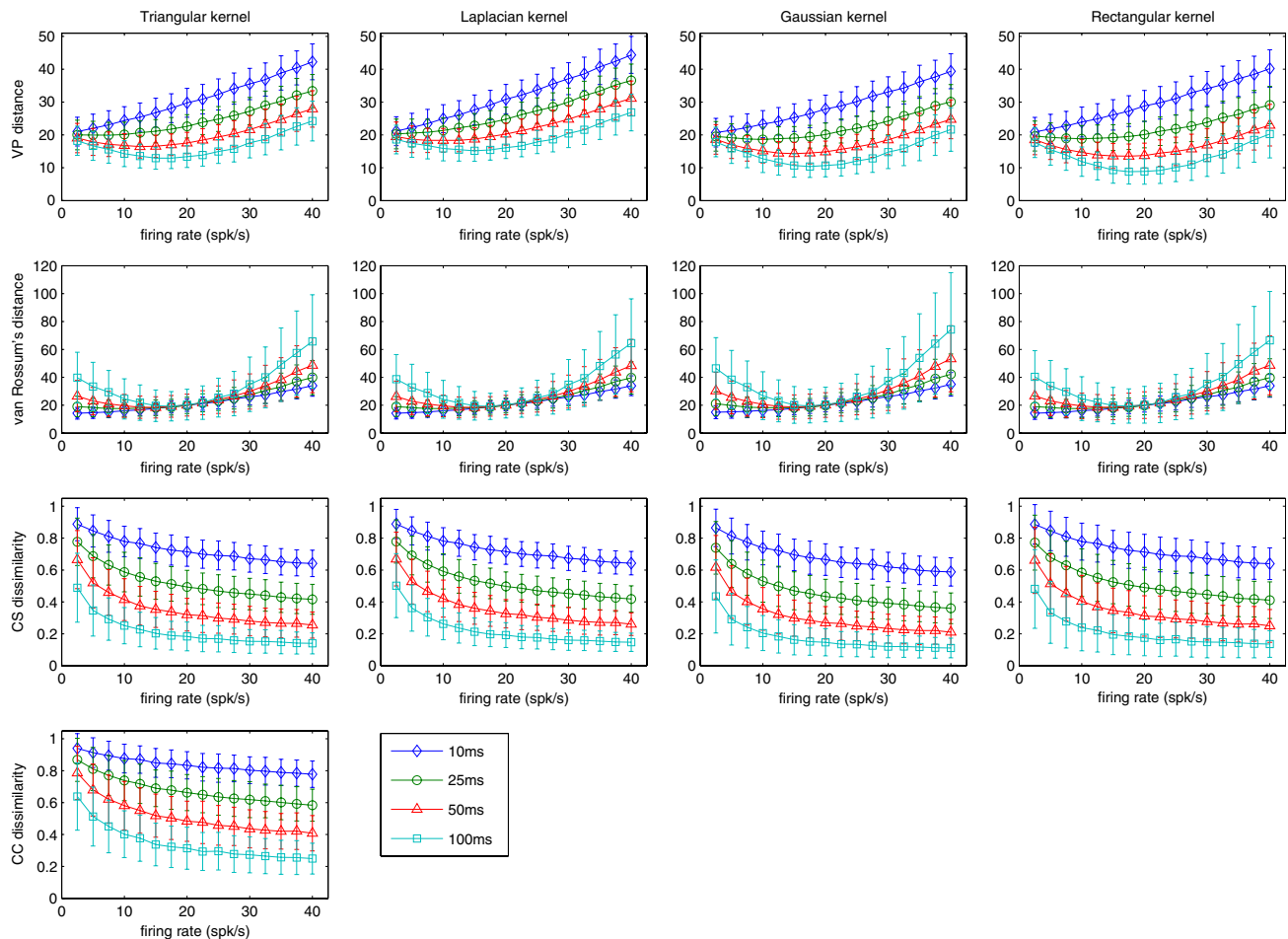
Fig. 5 Value of the dissimilarity measures for each kernel considered as a function of the modulating spike train firing rate. All dissimilarity evaluations are with regard to an homogeneous spike train with average rate 20 spk/s. For each measure and kernel, results are given for four different kernel sizes (shown in the legend) in terms of the measure average value ±1 SD. The statistics of the measures were estimated over 1,000 randomly generated pairs of spike trains

and the phase difference of the sinusoidal waveforms used to modulate the firing rate of each spike train varied from 0° to 360°. The goal was to verify if the measures were sensitive to instantaneous differences in the firing rate as characterized by the modulation phase difference. This too is a simplification of what is often found in practice where firing rates change abruptly and in a non-periodic manner. Nevertheless, the paradigm aims at representing a general situation while simultaneously being restricted to allow for a tractable analysis.

Obviously, the results are somewhat dependent on our choice of simulation parameters. For example, lower mean firing rates would mean that the dissimilarity measures would be less reliable and, hence, have higher variance. This could be partially compensated by increasing the spike train length. However, the above values are an attempt to approximate real data.

The simulation protocol is similar to that of the case analyzed in the previous section. For each phase difference,

we randomly generated 1,000 spike train pairs such that the firing rate modulation of the two spike trains differed by the phase difference and applied the dissimilarity measures using each of the four described kernels. As before, the analysis was repeated for four kernel sizes, 10, 25, 50, and 100 ms. Again, the kernel sizes used were chosen large since firing rate information can only be extracted at a slower time scale. The statistics of the dissimilarity measures are shown in Fig. 7.

The analysis of these results with the discrimination index v with respect to the statistics of each measure at zero phase is depicted in Fig. 8. In this paradigm, the maximum value of the measures was desired to occur at 180°, with a monotonically increasing behavior for phase differences smaller and monotonically decreasing for phase differences greater. As Fig. 8 shows, all measures performed satisfactorily using any of the four kernels and at any kernel size. The CS dissimilarity has the best discrimination with the discrimination index reaching 0.8,
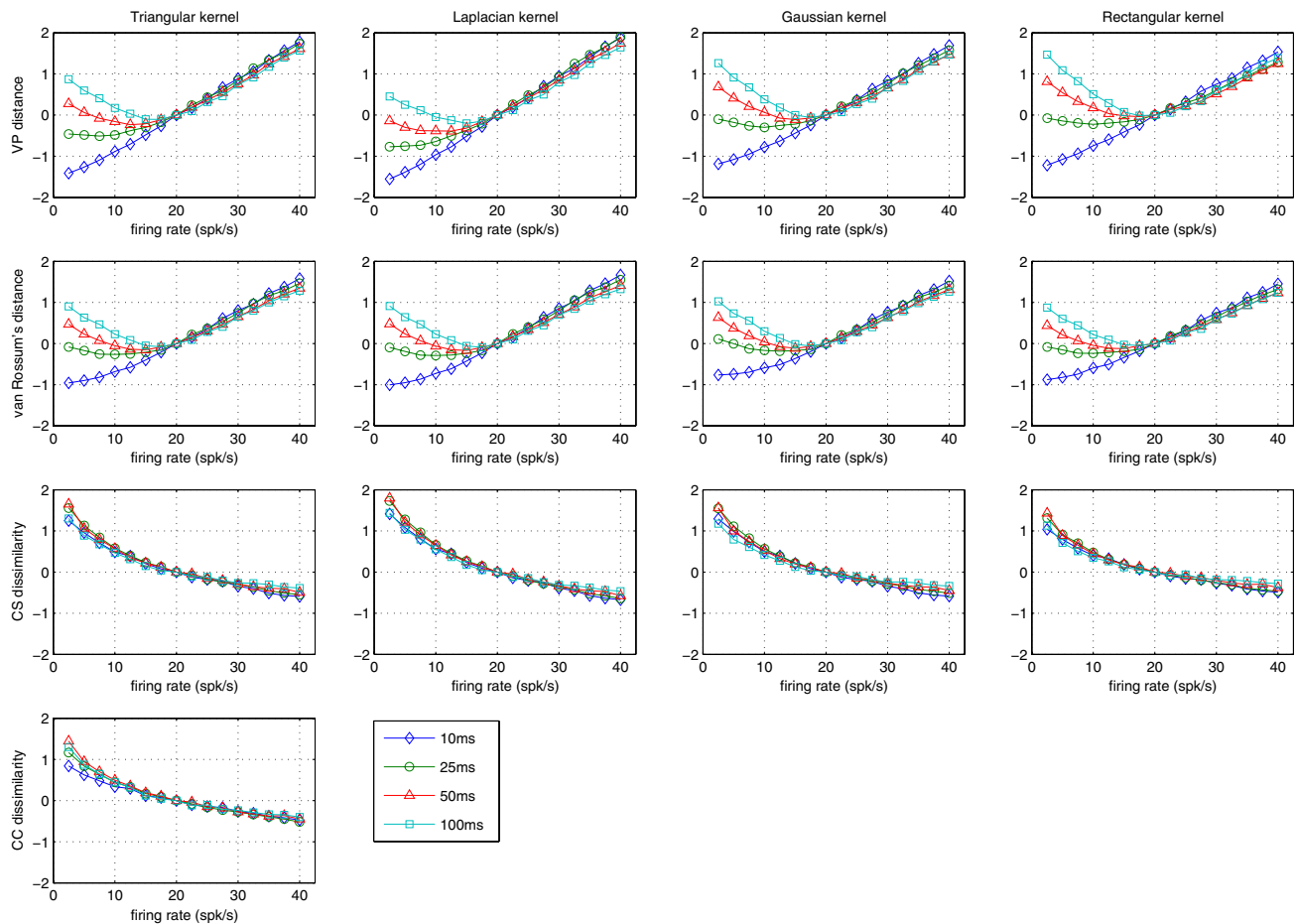
Fig. 6 Discriminant index of the dissimilarity measures for each kernel as a function of the modulating spike train firing rate. See the results in Fig. 5 for reference. The different curves are for different kernel sizes (shown in the legend)

compared to a maximum value of 0.65 for the second best. On the other end, overall the CC-based dissimilarity performed the worse. Comparing with the CS dissimilarity (which differs only because the spike times are not quantized) we verify once again the disadvantages of doing binning. With regard to the effect of each kernel, the Gaussian kernel consistently yields the best discrimination for the same kernel size. Conversely, the Laplacian and rectangular kernels seem to perform the worst, although this observation is largely measure-dependent. As expected, and similarly to the previous paradigm, the best discrimination is obtained for the largest kernel size since it yields a better estimation of the intensity function. It is noteworthy, however, that in this paradigm the kernel size cannot be chosen too large; otherwise, the intensity function would be over smoothed, thus reducing the differentiation between phases and decreasing the discrimination performance. This phenomenon was observed when we attempted a kernel size of 250 ms (not shown).

4.3 Discrimination of synchronous firings

In this scenario we consider that spike trains are to be differentiated based on the synchrony of neuron firings. More precisely, spike trains are deemed distant (or dissimilar) with regard to the relative number of synchronous spikes. That is, dissimilarity measures are expected to be inversely proportional to the probability of a spike co-occur with a spike in another spike train. Unlike the previous two case studies where differences in firing rate were analyzed, this case puts the emphasis of analysis in the role of each spike. Thus, since the time scale of analysis is much more fine, the precision of a spike time has increased relevance. On the other hand, it must be noted that in this case we consider synchrony in a more general sense than the usual concepts of precision and reliability often utilized in temporal coding studies [1, 19, 31], and on which most of the previous comparisons on spike train metrics have focused [4, 18, 28]. Rather, synchrony refers here to spike trains with *correlated* spike firings. This is a more general
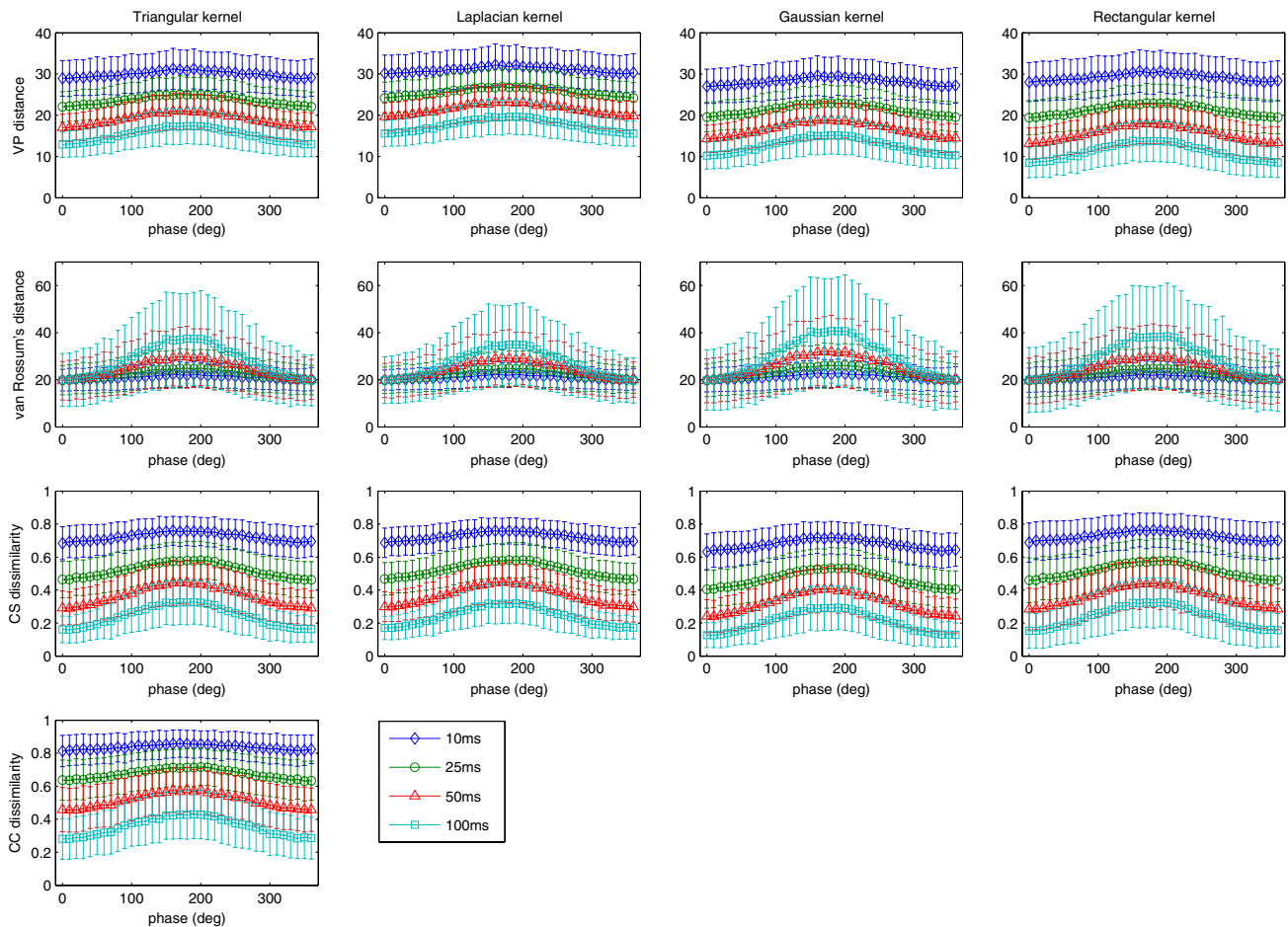
Fig. 7 Value of the dissimilarity measures for each kernel in terms of the phase difference of the firing rate modulation. Like in the previous paradigm, results are shown for each measure, kernel, and four different kernel sizes (shown in the legend) in terms of the measure average value ±1 SD. The statistics were estimated over 1,000 randomly generated pairs of spike trains

paradigm which allows one to obtain spike trains similar to those utilized in the previous studies if one considers the cases with high correlation and low noise [4, 19]. The advantages of this approach is that we can study the asymptotic behavior of the measures over a wide range of synchrony and noise characteristics.

To generate spike trains with a given synchrony, the multiple interaction process (MIP) model was used [38, 39]. In the MIP model a reference spike train is first generated as a realization of a Poisson process. The spike trains are then derived from this one by copying spikes with probability ε. The operation is performed independently for each spike and for each spike train. Put differently, ε is the probability of a spike co-occurring in another spike train, and therefore controls what we refer to as synchrony. It can also be shown that ε is the count correlation coefficient [38]. The resulting spike trains are Poisson processes. By generating the reference spike train with firing rate $\varepsilon\lambda$ it is ensured that the derived spikes trains have firing rate λ. To make the simulation more realistic,

jitter noise was added to each spike time to recreate the variability in spike times often encountered in practice, thus making the task of finding spikes that are synchronous more challenging. Jitter noise was generated as independent and identically distributed zero-mean Gaussian noise.

For each combination of synchrony and jitter standard deviation, 1,000 spike train pairs were generated, and the dissimilarity measures in terms of the four different kernels were evaluated. All spike trains were 1-s-long and the firing rate 20 spk/s, the same reasons mentioned for the previous paradigms. The kernel size for the results shown was 2 ms. The kernel size was chosen small since in this scenario the characterizing feature is synchronous firings. Obviously, one could search over the kernel size, which would provide information about the inherent timescale of the spike firings [1, 28, 40]. However, it is not done here since it would add a third dimension to our experiment and significantly complicate the analysis of the results. The results are shown in Fig. 9. The discrimination index v shown in Fig. 10 quantifies the discrimination with regard
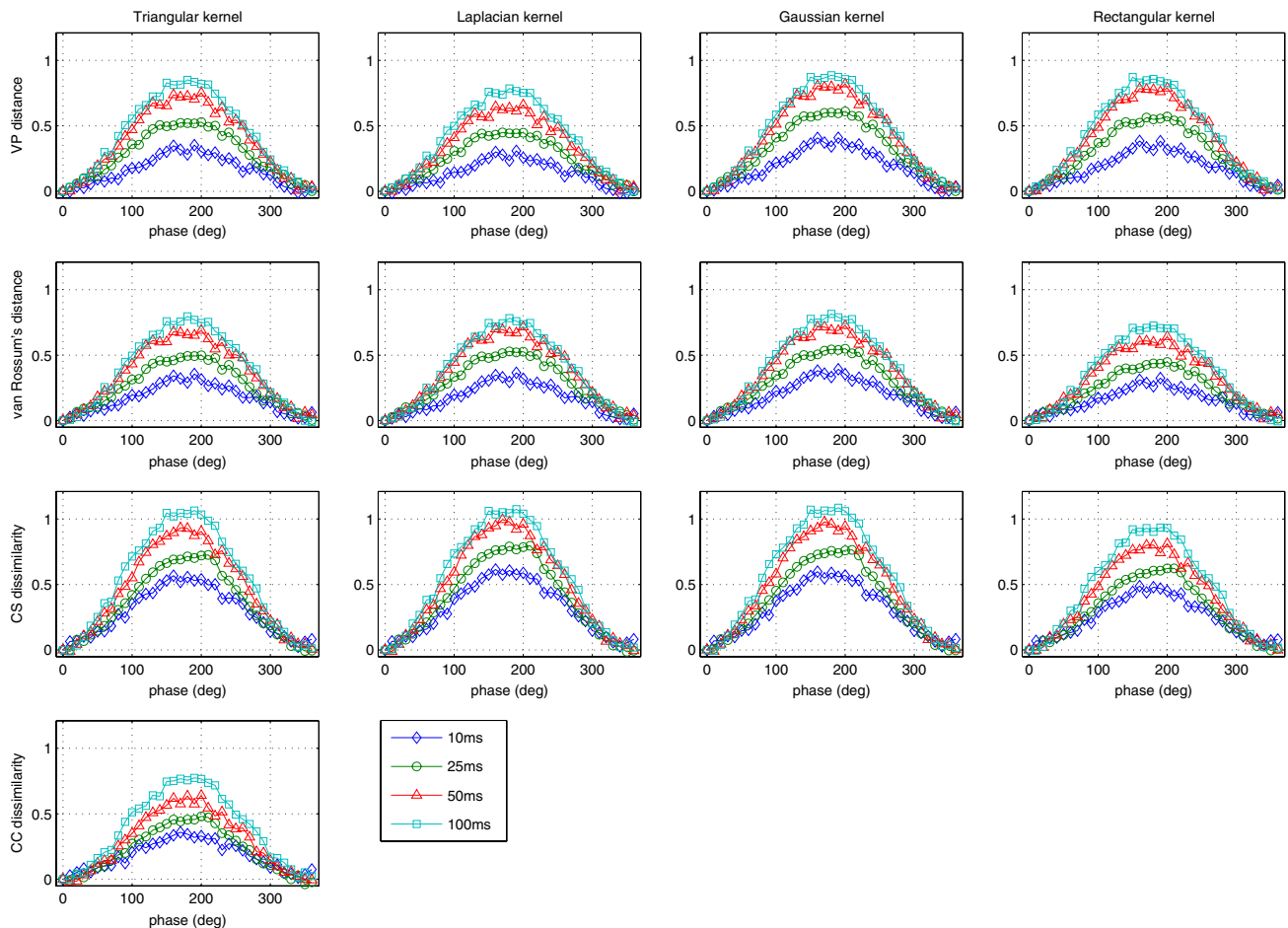
Fig. 8 Discriminant index of the dissimilarity measures for each kernel in terms of the phase of the firing rate modulation as given by Fig. 7. The different curves are for different kernel sizes (shown in the legend)

to the distribution of the measure values without synchrony. Simply put, the goal was to find which measure would better improve its discrimination as the synchrony is increased.

From Fig. 10, the CS and CC dissimilarities have notably better discrimination ability than VP and van Rossum's distance. A similar result had been obtained by [28] in a specific dataset, and which we show here to be a general properties of metrics of this form. In spite of the formulation of the VP and van Rossum distances as a coincidence detectors, these results show the importance of the normalization in d_{CS} and d_{CC} for measuring synchrony. Basically, while the VP and van Rossum distances measure the overall dissimilarity, the CS and CC dissimilarities normalize by the norm of the spike trains, providing a measure of dissimilarity "per spike," which more closely matches the concept of synchrony as the probability of synchronous spikes. The results also reveal that the CS dissimilarity is more consistent than the CC dissimilarity since its discrimination decreases in a more graded manner with the presence of variability in the synchronous spike

times (even for the same kernel function). This is due to the time quantization effect associated with binning in the CC dissimilarity. The VP and van Rossum's distances have comparable discrimination ability. Comparing the measurements in terms of the kernel functions, it was found that the Laplacian kernel provides the best results, followed by the triangular kernel. This was to be expected since this kernel "rewards" perfectly synchronous spikes and heavily penalizes others, thus focusing on synchronous firings and minimizing the interference of uncorrelated spikes in the distance. Nevertheless, the advantage between different kernels is small.

5 Conclusion

In this paper we compare several binless spike train measures presented in the literature for their discrimination ability. Given the wide use of these measures in spike trains analysis, classification and clustering, this study provides a systematic evaluation of the discrimination
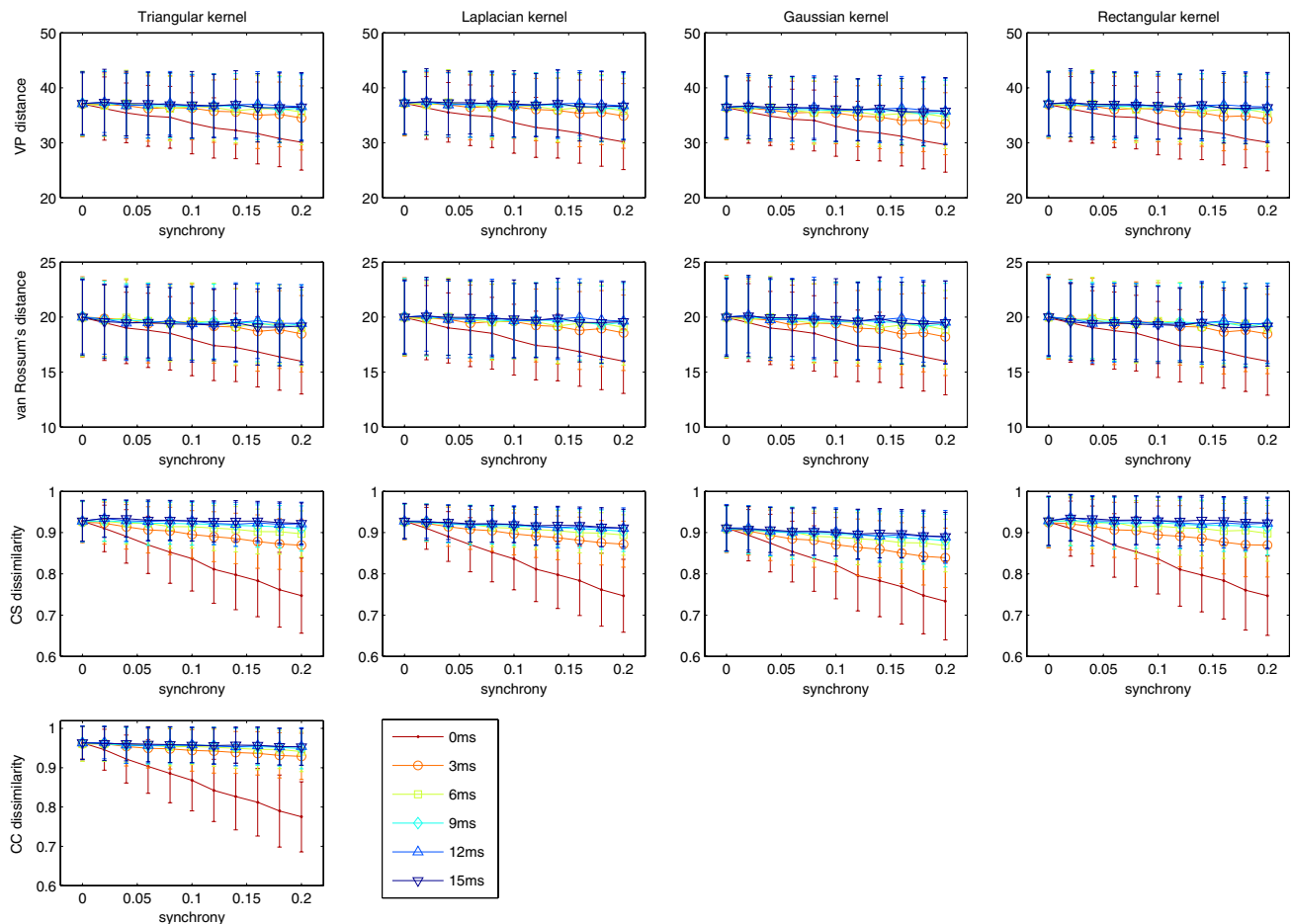
Fig. 9 Value of the dissimilarity measure for each kernel as a function of the synchrony among spike trains. The statistics were estimated over 1,000 randomly generated pairs of spike trains simulated with MIP model and average firing rate 20 spk/s. The kernel size was 2 ms. The different curves show the results under different levels of jitter standard deviation, with the value given in the legend

characteristics, fundamental for understanding the behavior of each measure and deciding which might be more appropriate taking the intended aim into consideration. Accordingly, the measures were compared in three experiments with the information for discrimination contained in average firing rates, instantaneous firing rates, and synchrony, covering a broad spectrum of spike encoding theories. These experiments were designed to recreate potential hypothesis testing scenarios that one may want to test in practice using real spike trains, although they are unavoidably simplified approximations.

The results reveals that no single measure performs the best or consistently throughout all three paradigms. For instance, although the VP and van Rossum distances have better discrimination in the constant firing rate paradigm, they are outperformed in the synchrony-based discrimination task by the CS and CC dissimilarities. On the other hand, the results of the latter measures are not at all consistent in the first paradigm, mostly because of their instability for a small number of spikes. Nevertheless, all measures performed consistently and comparably in the second paradigm, in terms of modulation of the instantaneous firing rates.

One of the most important findings in this study was that in some situations the measures did not perform consistently with our expectations for the experiment. This was observed with all the measures in the first paradigm. The results for the VP or van Rossum's distance are inconsistent for small kernel sizes (Fig. 6). Since the paradigm required sensitivity to firing rate, this was to be expected, but the results alert for the need to select an appropriate kernel size. The results became consistent once the kernel size was at least equal to the average inter-spike interval (50 ms in this case), which indicates that for these tasks one should choose a kernel size at least inversely proportional to the maximum firing rate. On the other hand, using the CS or CC dissimilarities the results in the first paradigm were inconsistent regardless of the kernel size (Fig. 6).
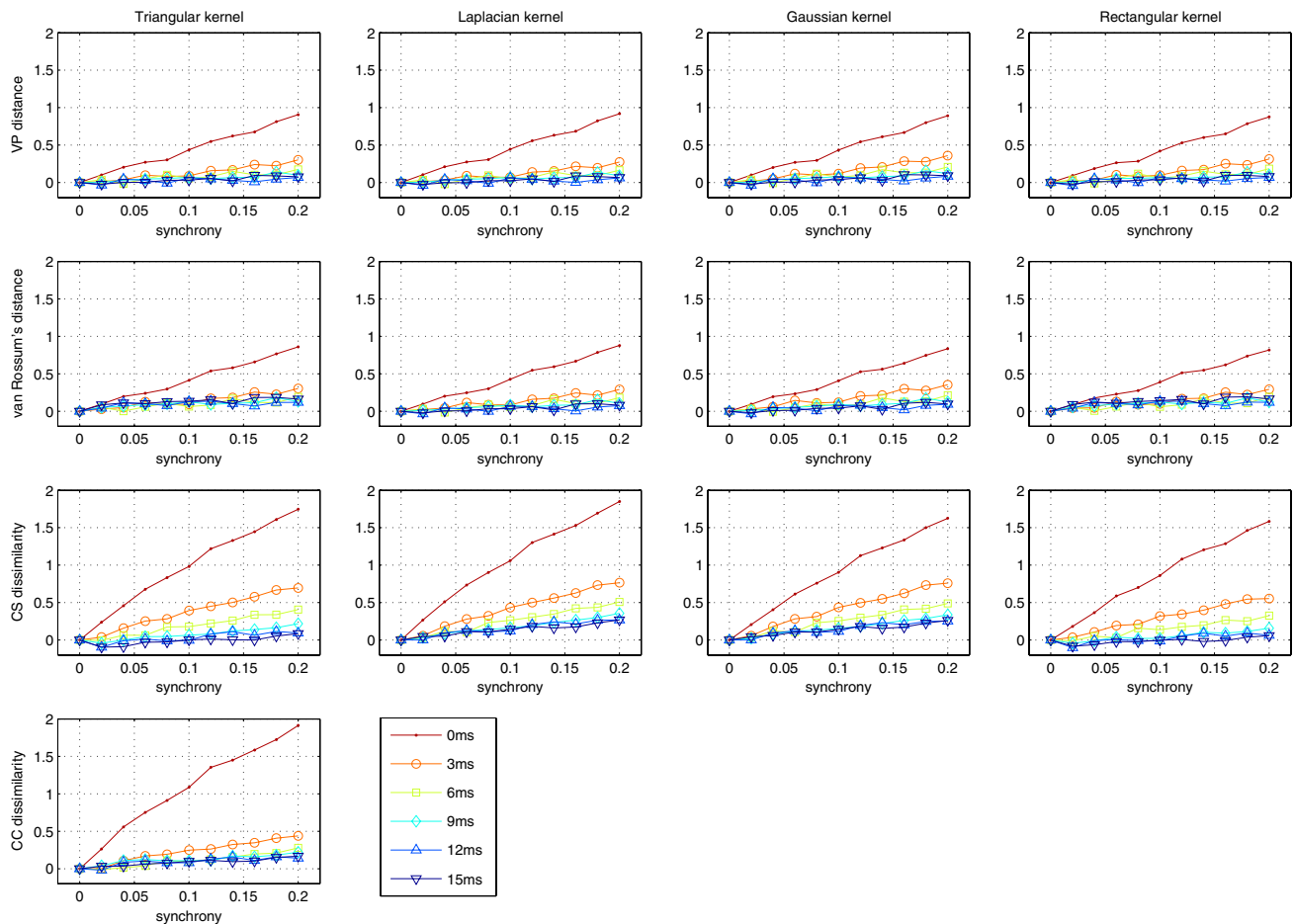
Fig. 10 Discriminant index of the dissimilarity measures for each kernel in terms of the synchrony between the spike trains as given by Fig. 9. The different curves are for different standard deviations (shown in the legend) of the jitter noise added to the synchronous spikes

Although the normalization proved helpful in the third paradigm (and it is beneficial for small firing rates in the first paradigm), it causes the dissimilarities to continue decreasing even as the firing rate of the second spike train keeps increasing above that of the reference spike train. Even though one could argue that the first paradigm is perhaps the least representative of a practical situation, based on these results we recommend caution when attempting to utilize these measures to quantify the information in spike trains, as these outcome may be severely underestimated.

An intriguing but not entirely surprising result is that, although the VP and van Rossum distances yield quite different results at times, as noticed clearly in Figs. 5 and 7, their discrimination was comparable in all paradigms (see Figs. 6, 8, 10). The similarity in their definition and between their values had already been noted [16, 27], but it is shown here to translate as well in terms of discrimination ability. However, it must be remarked that in all the paradigms the spike trains were modeled as realizations of Poisson processes, and therefore we cannot infer if this still

holds for spike trains from more general point process models.

More than a direct comparison of the measures in their published form, we considered also the effect of the kernel function utilized. Hence, we briefly summarized how each measure can accommodate different kernel functions in their evaluation. Nevertheless, the results reveal that the dependence of the measures on a specific kernel is minor. Still, the Gaussian kernel performs the best for firing rate paradigms, whereas the Laplacian kernel performed the best in the synchrony paradigm. On the other extreme, the rectangular kernel performed the worst.

Finally, the results depict the importance of *binless* spike train measures. As stated earlier, the only difference between the CS dissimilarity evaluated with the rectangular kernel and the CC dissimilarity is the time quantization incurred with binning. Comparing the results in these two situations in Figs. 8 and 10 shows that small improvements in discrimination and robustness to jitter noise were achieved in the first and second cases, respectively, by utilizing the spike times directly.

References

1. Victor JD, Purpura KP (1996) Nature and precision of temporal coding in visual cortex: a metric-space analysis. J Neurophysiol 76(2):1310–1326
2. Victor JD, Purpura KP (1997) Metric-space analysis of spike trains: theory, algorithms, and application. Netw Comp Neural Syst 8:127–164
3. Wohlgemuth S, Ronacher B (2007) Auditory discrimination of amplitude modulations based on metric distances of spike trains. J Neurophysiol 97:3082–3092. doi:10.1152/jn.01235.2006
4. Houghton C (2009) Studying spike trains using a van Rossum metric with a synapse-like filter. J Comp Neurosci 26:149–155. doi:10.1007/s10827-008-0106-6
5. Brown EN, Kass RE, Mitra PP (2004) Multiple neural spike train data analysis: state-of-the-art and future challenges. Nature Neurosci 7:456–461. doi:10.1038/nn1228
6. Dayan P, Abbott LF (2001) Theoretical neuroscience: computational and mathematical modeling of neural systems. MIT Press, Cambridge
7. Rieke F, Warland D, van Steveninck RR, Bialek W (1999) Spikes: exploring the neural code. MIT Press, Cambridge. ISBN 0-262-18174-6
8. Mainen ZF, Sejnowski TJ (1995) Reliability of spike timing in neocortical neurons. Science 268(5216):1503–1506. doi:10.1126/science.7770778
9. Hatsopoulos NG, Ojakangas CL, Paninski L, Donoghue JP (1998) Information about movement direction obtained from synchronous activity of motor cortical neurons. Proc Natl Acad Sci 95(26):15706–15711
10. Wagner H, Brill S, Kempter R, Carr CE (2005) Microsecond precision of phase delay in the auditory system of the barn owl. J Neurophysiol 94(2):1655–1658. doi:10.1152/jn.01226.2004
11. Eggermont JJ (2006) Properties of correlated neural activity clusters in cat auditory cortex resemble those of neural assemblies. J Neurophysiol 96(2):746–764. doi:10.1152/jn.00059.2006
12. Riehle A, Grün S, Diesmann M, Aertsen A (1997) Spike synchronization and rate modulation differentially involved in motor cortical function. Science 278(5345):1950–1953. doi:10.1126/science.278.5345.1950
13. Grün S, Diesmann M, Aertsen A (2002) Unitary Events in multiple single-neuron activity. I. detection and significance. Neural Comp 14(1):43–80
14. Grün S, Diesmann M, Aertsen A (2002) Unitary Events in multiple single-neuron activity. II. nonstationary data. Neural Comp 14(1):43–80
15. Grün S, Diesmann M, Grammont F, Riehle A, Aertsen A (1999) Detecting unitary events without discretization of time. J Neurosci Methods 93(1):67–79. doi:10.1016/S0165-0270(99)00126-0
16. van Rossum MCW (2001) A novel spike distance. Neural Comp 13(4):751–764
17. Schreiber S, Fellous JM, Whitmer D, Tiesinga P, Sejnowski TJ (2003) A new correlation-based measure of spike timing reliability. Neurocomp 52–54:925–931. doi:10.1016/S0925-2312(02)00838-X
18. Kreuz T, Haas JS, Morelli A, Abarbanel HDI, Politi A (2007) Measuring spike train synchrony. J Neurosci Methods 165(1):151–161. doi:10.1016/j.jneumeth.2007.05.031
19. Hunter JD, Milton JG (2003) Amplitude and frequency dependence of spike timing: Implications for dynamic regulation. J Neurophysiol 90(1):387–394. doi:10.1152/jn.00074.2003
20. Aronov D, Reich DS, Mechler F, Victor JD (2003) Neural coding of spatial phase in V1 of the macaque monkey. J Neurophysiol 89(6):3304–3327. doi:10.1152/jn.00826.2002
21. Aronszajn N (1950) Theory of reproducing kernels. Trans Am Math Soc 68(3):337–404
22. Houghton C, Sen K (2008) A new multineuron spike train metric. Neural Comp 20(6):1495–1511. doi:10.1162/neco.2007.10-06-350
23. Victor JD (2005) Spike train metrics. Curr Opin Neurobiol 15(5):585–592. doi:10.1016/j.conb.2005.08.002
24. Fellous J-M, Tiesinga PHE, Thomas PJ, Sejnowski TJ (2004) Discovering spike patterns in neuronal responses. J Neurosci 24(12):2989–3001. doi:10.1523/JNEUROSCI.4649-03.2004
25. Paiva ARC, Rao S, Park I, Príncipe JC (2007) Spectral clustering of synchronous spike trains. In: Proceedings of the IEEE international joint conference on neural networks, IJCNN-2007, Orlando, FL, USA
26. Toups JV, Tiesinga PHE (2006) Methods for finding and validating neural spike patterns. Neurocomputing 69(10–12):1362–1365
27. Schrauwen B, Van Campenhout J (2007) Linking non-binned spike train kernels to several existing spike train distances. Neurocomputing 70(7–8):1247–1253. doi:10.1016/j.neucom.2006.11.017
28. Wang L, Narayan R, Graña G, Shamir M, Sen K (2007) Cortical discrimination of complex natural stimuli: can single neurons match behavior? J Neurosci 27(3):582–589. doi:10.1523/JNEUROSCI.3699-06.2007
29. Abeles M (1991) Corticonics: neural circuits of the cerebral cortex. Cambridge University Press, New York
30. Shadlen MN, Newsome WT (1994) Noise, neural codes and cortical organization. Curr Opin Neurobiol 4(4):569–579
31. de Ruyter van Steveninck RR, Lewen GD, Strong SP, Koberle R, Bialek W (1997) Reproducibility and variability in neural spike trains. Science 275:1805–1808
32. Papoulis A (1965) Probability, random variables, and stochastic processes. McGraw-Hill, New York
33. Reiss R-D (1993) A course on point processes. Springer, New York
34. Paiva ARC, Park I, Príncipe JC (2009) A reproducing kernel Hilbert space framework for spike train signal processing. Neural Comp 21(2):424–449. doi:10.1162/neco.2008.09-07-614
35. Pekalska E, Duin RPW (2005) The dissimilarity representation for pattern recognition: foundations and applications. World Scientific, Singapore. ISBN 9-812-56530-3
36. Duda RO, Hart PE, Stork DG (2000) Pattern classification, 2nd edn. Wiley Interscience
37. Adrian ED (1928) The basis of sensation: the action of the sense organs. W. W. Norton & Co., New York
38. Kuhn A, Aertsen A, Rotter S (2003) Higher-order statistics of input ensembles and the response of simple model neurons. Neural Comp 15(1):67–101
39. Kuhn A, Rotter S, Aertsen A (2002) Correlated input spike trains and their effects on the response of the leaky integrate-and-fire neuron. Neurocomputing 44–46:121–126. doi:10.1016/S0925-2312(02)00372-7
40. Narayan R, Graña G, Sen K (2006) Distinct time scales in cortical discrimination of natural sounds in songbirds. J Neurophysiol 96(1):252–258. doi:10.1152/jn.01257.2005