A Phase Function to Quantify Serial Dependence between Discrete Samples

Ramana Dodla* and Charles J. Wilson
Department of Biology, The University of Texas at San Antonio, San Antonio, Texas

ABSTRACT Auto- and cross-correlation methods, when applied to discrete events, can determine periodicity and correlation times within and between event train sequences. However, if the number of available events for analysis is too few, the correlation techniques yield ambiguous and insufficient results. Here we report a technique based on measurements of phases of event times that could detect the periodicity even among very few discrete data points. The results are demonstrated on in vitro neuronal spike time data, and are found to be highly contrasting when compared with the correlation techniques. The technique could become invaluable, for example, for treating in vivo spike time records that often last very short duration, or for determining short timescales in discrete biophysical experimental data.

Received for publication 25 August 2009 and in final form 9 November 2009.

*Correspondence: ramana.dodla@utsa.edu

Interevent or interspike interval distribution-based parameters such as mean, standard deviation, and coefficient of variation are insensitive to random shuffling of discrete data points, and are strictly useful only for complete characterization of renewal processes (1). Despite their stochastic nature, spike events can have serial dependence, and measures that are based on higher order distributions such as autocorrelation function (or cross-correlation function between pairs of sequences) become useful not only in determining the periodicity among the spikes, but also in finding the correlation time between events of the same train or between pairs of such trains (2). Although correlations based on large data sets pose few difficulties (3,4), brief data sets, such as those from in vivo recordings, would pose considerable difficulty. In data sets that are beset with perceptible nonstationarity, inclusion of large amounts of data for computing the correlation functions biases the results and ignores the serial dependences that exist at much shorter timescales than that of the entire data. The problem of dealing with short experimental data sets is compounded by the tradeoff in deciding an appropriate bin width to be used to count the serial dependence.

We address this problem by introducing the concept of a phase at each spike event in relation to a previous spike event. Such a concept is familiar in neuronal studies for finding phases of spike events with respect to a background activity (5) or a stimulus to determine preferred phases of neuronal firing. Here the role of stimulus-response is played by the spike train pair. Spike events are time-ordered, and are represented as pairs of phases in a two-dimensional space. Then an effective drift of such pairs from their mean is computed. In a state that is dominated by synchrony, such a drift becomes minimum. Because the spike events are now converted to instantaneous phases the drift also becomes oscillatory as the relative time shift between the trains (τ) is increased. Each phase pair contributes to the drift measure at any given τ, thus eliminating the need for binning the data. We term this measure a phase function. First, we formulate the phase function, and then apply it to a very brief experimental data set to find an autophase function. The method reveals the inherent periodicity and the decay constant in the data clearly, enabling further study of the biophysical mechanisms, whereas the correlation method applied to it does not show measurable oscillations. In contrast to the previous methods that quantified dissimilarity between spike trains (6–8), our method quantifies the relative occurrence of spike events in time.

Let $A_1, A_2, A_3, \ldots, A_{N_A}$ be the spike time sequence of the first cell. Let $B_1, B_2, B_3, \ldots, B_{N_B}$ be the spike time sequence of the second cell, and be time-shifted with respect to the first by τ. For illustration, assume that when time-ordered, the spike times t_1, t_2, t_3, t_4, t_5, t_6, …correspond, respectively, to A_1, $B_1, A_2, B_2, A_3, B_3, \ldots$as shown in Fig. 1 *a*. We restrict our analysis to the time window $[\max\{A_1, B_1\}, \min\{A_{N_A}, B_{N_B}\}]$. At t_3, cell 1 and cell 2 can be assumed to have traversed in their orbits for time x_1 since the last spike time t_2. With respect to their individual interspike intervals, the fractions of the time covered by cell 1 and cell 2 are γ_1 and δ_1, where $\gamma_1 = x_1/(A_2 - A_1) = (t_3 - t_2)/(A_2 - A_1)$, $\delta_1 = x_1/(B_2 - B_1) = (t_3 - t_2)/(B_2 - B_1)$. These two points are represented as a phase plane pair (γ_1, δ_1) as shown in Fig. 1 *b*. Similarly, from Fig. 1 *a*, two more pairs are constructed and the corresponding pairs are marked in Fig. 1 *b*. The phase pairs (γ_i, δ_i), $i = 1, \ldots, N$, where N is the maximum number of such constructed pairs, form the basis of our computation of shifts of the instantaneous phases along and vertical to the line drawn from the origin to the average point ($\langle\gamma\rangle$,$\langle\delta\rangle$). The shifts along the average line correspond to the drift of the phases that we will use in defining the phase function. The vertical shifts,

Editor: Edward H. Egelman.

doi: 10.1016/j.bpj.2009.11.003

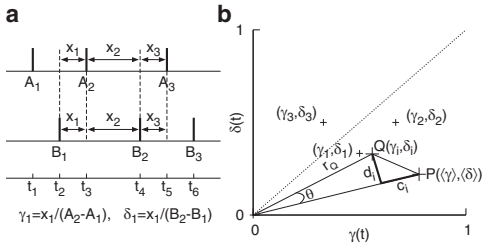

FIGURE 1 Converting spike trains into phase pairs. (*a*) Two brief time-ordered spike trains are illustrated. At a given spike event, its relative time shift from the most recent spike event is divided by the interspike interval to convert to its instantaneous phase. The phase pair (γ_1, δ_1) computed at time t_3 is illustrated. (*b*) The three phase pairs of panel *a* are shown in the phase plane. The value *Q* is a typical phase plane point, and *P* is the point representing the average values of γ- and δ-values. The shift $c_i(\tau)$ that depends on τ represents a drift and is used in defining the phase function.

d_i (not discussed here), correspond to a diffusion of phases and thus quantify the dissimilarity between the spike trains.

The average quantities of the phase components are defined as $\langle\gamma\rangle = 1/N\sum_{i=1}^{N}\gamma_i$ and $\langle\delta\rangle = 1/N\sum_{i=1}^{N}\delta_i$. The construction of the drift parameter is shown in Fig. 1 *b*. The shifts along the line *OP* indicate the drift affected by a typical point $Q(\gamma_i, \delta_i)$. We mark these shifts $c_i(\tau)$. The quantity $c_i(\tau)$ can be determined by subtracting the intercept that *OQ* makes on the line *OP* ($r_Q \cos\theta$) from the length of *OP* (i.e., r_P). Thus, $c_i(\tau) = r_P - r_Q \cos\theta$. The scalar product of the vectors $\overrightarrow{OQ}$ and $\overrightarrow{OP}$ is given by $r_Q r_P \cos\theta = \gamma_i\langle\gamma\rangle + \delta_i\langle\delta\rangle$. Thus, $c_i(\tau) = r_P - (\gamma_i\langle\gamma\rangle + \delta_i\langle\delta\rangle)/r_P$, where $r_P = \sqrt{\langle\gamma\rangle^2 + \langle\delta\rangle^2}$. We see that $c_i(\tau)$ is proportional to the sum of the components with one of the components scaled by the ratio of the average phases. The value $c_i(\tau)$ measures the shift of this quantity from the sum of the squares of the average phases. Hence, $c_i(\tau)$ could rightly be called a drift component. Now we define a dimensionless drift parameter for the phase pairs as $C(\tau) = 1/N\sum_{i=1}^{N}|c_i(\tau)|$. As γ_i and δ_i are confined between 0 and 1, the average quantities $\langle\gamma\rangle$ and $\langle\delta\rangle$ will also be each confined to the interval [0, 1]. Hence, the maximum $c_i(\tau)$ and consequently $C(\tau)$ can assume is $\sqrt{2}$. Thus, we define a normalized phase function [$\psi(\tau)$] based on the distance parameter $C(\tau)$ as $\psi(\tau) = C(\tau)/\sqrt{2}$. The value $\psi(\tau)$ can range between 0 and 1.

The phase function depends on the drift of the instantaneous phases from the average phases, and is plotted as a function of τ (both positive and negative) for realizing the serial dependence in the data. We term the phase function an autophase function if it is computed by relatively shifting a sequence with respect to itself, and a cross-phase function if it is computed between two different sequences. A steady phase function for a range of τ-values indicates absence of serial dependence between the spike trains in that region.

For the phase function to exist, the data must allow at least two phase pairs to be computed. Thus, for an autophase function, the minimum number of spike times required is three. The leading and lagging spike times in the spike trains outside the computational window cannot form phase pairs because a phase cannot be defined for them. However, the serial dependence in them can be captured at appropriate values of τ that would bring them into the computational window.

The autophase function for very small nonzero time-lag can be computed by noting that the phase pairs for $\tau \to 0^+$ will assume values alternately between (0, 0) and (1, 1). Thus, $\langle\gamma\rangle = \langle\delta\rangle = 0.5$, and consequently $C(\tau \to 0^+)$ becomes $1/\sqrt{2}$. Hence, $\psi(\tau \to 0^+) = 0.5$. Similarly, we can see that $\psi(\tau \to 0^-)$ will also become 0.5. However, at $\tau = 0$, $\gamma_i = \delta_i = 1$ for all i, leading to $\psi(\tau = 0) = 0$ and thus, the autophase function is discontinuous at $\tau = 0$. Like an autocorrelation function, the autophase function can also become discontinuous for completely periodic spike trains. This does not, however, affect the emergence of periodicity seen in the functional dependence of $\psi(\tau)$. Thus, proximity of the phase function to zero does not have significance in the emerging oscillations in $\psi(\tau)$. At large τ, an asynchronous state may be assumed if the phase pairs become completely random. We estimate $\psi(\tau)$ at large τ for an asynchronous state that has the phase pairs homogeneously and continuously distributed on the line *OP* between 0 and 1 (Fig. 1 *b*). Without loss of generality, the relation between the phases can be derived from the case of two regular spike trains that are phase-shifted: $\gamma_i = \delta_i$, and $\langle\gamma\rangle = \langle\delta\rangle = 0.5$. These values can be used to compute the asymptotic value of $C(\tau)$,

$$C(\tau\to\infty) = \frac{1}{N}\sum_{i=1}^{N}\sqrt{2}\,|\langle\gamma\rangle - \gamma_i| = 2\sqrt{2}\int_0^{0.5}(0.5-x)dx = 0.25\sqrt{2},$$

thus, $\psi(\tau \to \infty) = 0.25$.

We now demonstrate the above method using a very brief data of 3–7 spike times. The data are the spike times recorded from rat globus pallidus neurons in vitro. Cell attached recordings were made from brain slices taken from a p-17 rat in artificial cerebro-spinal fluid at a temperature of 32°C. The physiological experimental protocols followed are identical to those reported in Deister et al. (9). The neurons in general fire rhythmically, but irregularly. The irregularities displayed by the cells are quite complex and thus, offer a challenge to the methods that seek to quantify serial dependence in the data. The data consists of the spike times and is presented in Fig. 2 *a* as spike times versus interspike intervals (ISIs). The autophase function (APF) for this data is shown in Fig. 2 *b*. As predicted, the APF is zero for $\tau = 0$, and is near 0.5 for very small τ. The APF profile clearly captured the inherent periodicity in the data with the first peak at 0.154 s, and the second peak at 0.313 s. The height of the peaks decreased gradually. When an

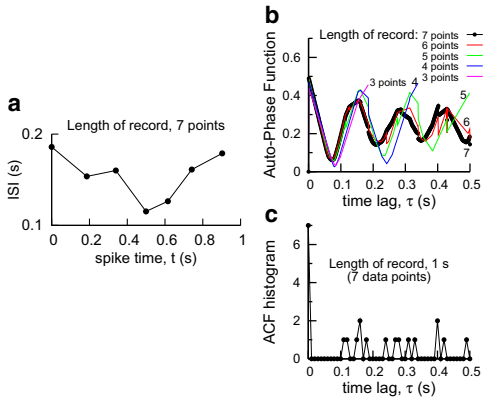

FIGURE 2 (*a*) One-second spike train data recorded from rat globus pallidus neurons. The data consists of seven uneven spike times displayed here as spike times versus interspike intervals (ISIs). (*b*) The autophase function computed for the data in panel *a* plotted as a function of the time-lag, τ, already displays clear and measurable oscillations and decay time constant. Results for even briefer data (using 6, 5, 4, and 3 points counted from $t = 0$) are also shown. (*c*) The conventional autocorrelation function (ACF) histogram for data in panel *a* as a function of time-lag does not yet display measurable oscillations.

exponential function is used to fit the peak decay from the maximum, the decay time constant is found to be 0.79 s. Apart from the discontinuity at $\tau = 0$, APF displays discontinuous jumps when the spike time of one train (in Fig. 2 *a*) goes past the spike time of the other train as much bigger or much smaller ISIs are encountered. These jumps get smoothed over when number of data points is increased, but do not disrupt the periodicity displayed by the data as is demonstrated. The asymptotic value of the APF at large τ is estimated as 0.247, which is very close to the analytical prediction. On the other hand, the autocorrelation function computed from the data set (Fig. 2 *c*) did not show any measurable profile. Such brief data is also not sufficient to statistically obtain an ISI histogram. As more points are added to the data set (Fig. 3), the phase function showed different decay times indicating the dominance of the irregularity at longer timescales over that at shorter timescales.

Thus, the phase function provides a tool to explore systematically the periodicity and the time constants in much shorter segments of discrete event data than a correlation function can. An autophase function was obtained using event trains as brief as three data points. Similar results for a cross-phase function can be obtained between spike trains. The phase associated with each spike event that we have introduced is similar in concept to such phases defined with respect to a background activity (5), or in phase response curves (10), but we have advanced this idea to the case where the phases are computed between spike events, and isolated the drift from the distance. This method could also become useful in real-time computation of correlations (11) with very few data points. An implementation of the method in Mathematica (Wolfram Research, Champaign, IL) and MATLAB (The MathWorks, Natick, MA) is provided in the Supporting Material.

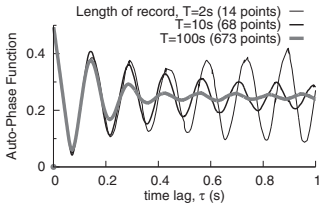

FIGURE 3 Autophase function as a function of time-lag for experimental records of different lengths. The decay time constant of the function is affected by inclusion of longer records.

SUPPORTING MATERIAL

Two implementation routines are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)01725-1.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health/National Institute of Neurological Disorders and Stroke (grant No. NS47085). Computational support was provided by Texas Advanced Computing Center/the University of Texas at Austin and Computational Biology Initiative at the University of Texas Health Science Center at San Antonio/the University of Texas at San Antonio.

REFERENCES and FOOTNOTES

1. Moore, G. P., D. H. Perkel, and J. P. Segundo. 1996. Statistical analysis and functional interpretation of neuronal spike data. *Annu. Rev. Physiol.* 28:493–522.
2. Gerstein, G. L., and N. Y.-S. Kiang. 1960. An approach to the quantitative analysis of electrophysiological data from single neurons. *Biophys. J.* 1:15–28.
3. Palm, G., A. M. H. J. Aertsen, and G. L. Gerstein. 1988. On the significance of correlations among neuronal spike trains. *Biol. Cybern.* 59:1–11.
4. Brown, E. N., R. E. Kass, and P. P. Mitra. 2004. Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.* 7:456–461.
5. Moran, A., H. Bergman, …, I. Bar-Gad. 2008. Subthalamic nucleus functional organization revealed by Parkinsonian neuronal oscillations and synchrony. *Brain.* 131:3395–3409.
6. Victor, J. D., and K. P. Purpura. 1996. Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J. Neurophysiol.* 76:1310–1326.
7. van Rossum, M. C. W. 2001. A novel spike distance. *Neural Comput.* 13:751–763.
8. Kreuz, T., J. S. Haas, …, A. Politi. 2007. Measuring spike train synchrony. *J. Neurosci. Methods.* 165:151–161.
9. Deister, C. A., C. S. Chan, …, C. J. Wilson. 2009. Calcium-activated SK channels influence voltage-gated ion channels to determine the precision of firing in globus pallidus neurons. *J. Neurosci.* 29:8452–8461.
10. Danzl, P., R. Hansen, …, J. Moehlis. 2008. Partial phase synchronization of neural populations due to random Poisson inputs. *J. Comput. Neurosci.* 25:141–157.
11. Dragoi, G., and G. Buzsáki. 2006. Temporal encoding of place sequences by hippocampal cell assemblies. *Neuron.* 50:145–157.