

Short communication

Entropy analysis of neuronal spike train synchrony

 Yoshinao Kajikawa^{a,*}, Troy A. Hackett^{a,b}
^a Department of Psychology, Vanderbilt University, 301 Wilson Hall, 111 21st Avenue South, Nashville, TN 37203, USA

^b Department of Hearing & Speech Science, Vanderbilt University, TN, USA

Received 7 February 2005; received in revised form 20 April 2005; accepted 22 April 2005

Abstract

Although vector strength (VS) and the Rayleigh tests are widely used to quantify neuronal firing synchrony to cyclic events, their use is valid only for singly peaked, unimodal distributions. In this report, we propose a new method to quantify synchrony, applicable to both unimodal and multimodal distributions. We also propose a statistical test to examine temporal structure under a null hypothesis of no synchrony. © 2005 Elsevier B.V. All rights reserved.

Keywords: Neurophysiology; Spike train; Synchrony; Entropy; Vector strength; Statistics

1. Introduction

The synchronization of neuronal firing to a repeating external event is an important aspect of neuronal coding strategy often quantified by VS (Drew and Doucet, 1991; Goldberg and Brown, 1969). The great utility of VS is that one can examine the significance of synchrony by the Rayleigh test, where greater synchrony yields a larger value of VS. However, vector-based analyses are limited to unimodal temporal firing patterns, because vectors of multimodal peaks cancel each other and diminish VS. Thus, it is apparent that a more robust measure of synchrony is needed in such a situation.

A solution to this problem comes from regarding synchrony as a large departure from a random, uniform pattern. Such randomness can be quantified by entropy regardless of the type of distribution, and can be tested in spike trains with random patterns created by shuffling inter-spike intervals, ISI (Werner and Mountcastle, 1963). In this report, we propose an entropy-based analysis to evaluate neural synchrony and test its significance by comparing with those evaluated for ISI-shuffled spike trains. The results indicate that this method can be broadly applied to assess

the synchrony of both unimodal and multimodal neural events.

2. Quantification and test of synchrony

Assume that a spike train was recorded during cyclic stimulation with cycle length, T . The probability distribution of the spike timing in a cycle, $\text{Pr}(0 < t < T)$, is obtained by normalizing a period histogram of spike timing relative to cyclic stimulation. Then, VS and E are calculated as, $\text{VS} = \left| \sum_{t=0}^T \text{Pr} \exp(2\pi i t / T) \right|$ and $E = - \sum_{t=0}^T \text{Pr} \log_2(\text{Pr})$, respectively. For a uniform (flat) distribution, $\text{VS} = 0$ and $E_{\max} = \log_2(N)$, where N is the number of bins in the period histogram. For a completely phase-locked spike train (meaning all spikes occurred at the same timing), $\text{VS} = 1$ and $E = 0$. Thus, VS and E change in opposite directions and their possible maximal values disagree. To enable a measure using entropy to change in parallel to, and to the same extent as VS, we propose a new measure, $D = 1 - E/E_{\max}$. For a unimodal Pr , D becomes 0 or 1 at the same time as VS. However, when Pr exhibits multiple peaks in a cycle period VS may diminish, but not D . For example, the vectors of two well phase-locked peaks separated by $T/2$ cancel each other out, resulting in $\text{VS} = 0$, while D may take a value between 0 and 1. Note that VS contains phase information through $\exp(2\pi i t / T)$, whereas E contains

* Corresponding author. Tel.: +1 615 322 7491; fax: +1 615 343 8449.
E-mail address: yoshi.kajikawa@vanderbilt.edu (Y. Kajikawa).

only Pr , which represents histogram magnitude with no phase information. Thus, neither E nor D holds any phase information. If Pr is multimodal, then not only is it invalid to quantify synchrony with VS , but it is also less informative with respect to peak phases.

The limitation of D is that there is no conventional test to examine its significance. To perform such test, the distribution of D under the null hypothesis of no synchrony is required for comparison. Fortunately, the synchrony in a spike train depends on the serial sequence of ISI. A train obtained after shuffling the serial order of ISI degrades such synchrony, if present (Moore et al., 1966; Werner and Mountcastle, 1963), but preserves the firing rate and ISI distribution. Accordingly, we randomized the serial sequence of ISI and calculated its cumulative summation, which represented the timing of spikes in null hypothetical spike train. For each shuffled spike train obtained in this way, we estimated null hypothetical D . By performing this procedure 1000 times, we obtained a distribution of 1000 null hypothetical values of D . Then, we compared D of the original spike train with this distribution (Fig. 1 right). The inverse of the number of times shuffled gives the resolution of the P value of the null hypothesis. The same strategy was also used to test VS along with the Rayleigh test (Fig. 1 center).

Fig. 1 shows several examples of spike trains. Those in (A) and (B) were generated by repetition of unimodal and bimodal conditioning distributions, respectively. Both spike trains had one phase-locked peak near the beginning of the cycle (Fig. 1A and B left). Even though the peak in (A) was broader than the first one in (B), VS was much smaller in (B) due to the second peak. Consequently, the Rayleigh test concluded that only spike train (A) had significant synchrony, regardless of the rich temporal pattern of (B). In contrast, the D value of spike train (B) was only slightly smaller than that of (A). Additionally, the distributions of the null hypothetical D values obtained from shuffled spike trains for both (A) and (B) were greatly reduced compared to those of the original spike trains (Fig. 1A and B right). Thus, the test of D using shuffled spike trains concluded that both spike trains had significant temporal pattern, that is, synchrony. The spike train in (C) was obtained by shuffling the ISI of the spike train in (B), thus ISI distributions were identical. The temporal pattern was clearly degraded by shuffling. Synchrony was not significant by either the Rayleigh test or test of D with further shuffling. In panel (D), the multimodal spike train was representative of many units recorded from primary auditory cortex during cyclic auditory stimulation. Like the spike train in (B), the Rayleigh test was not significant, whereas the test of D was

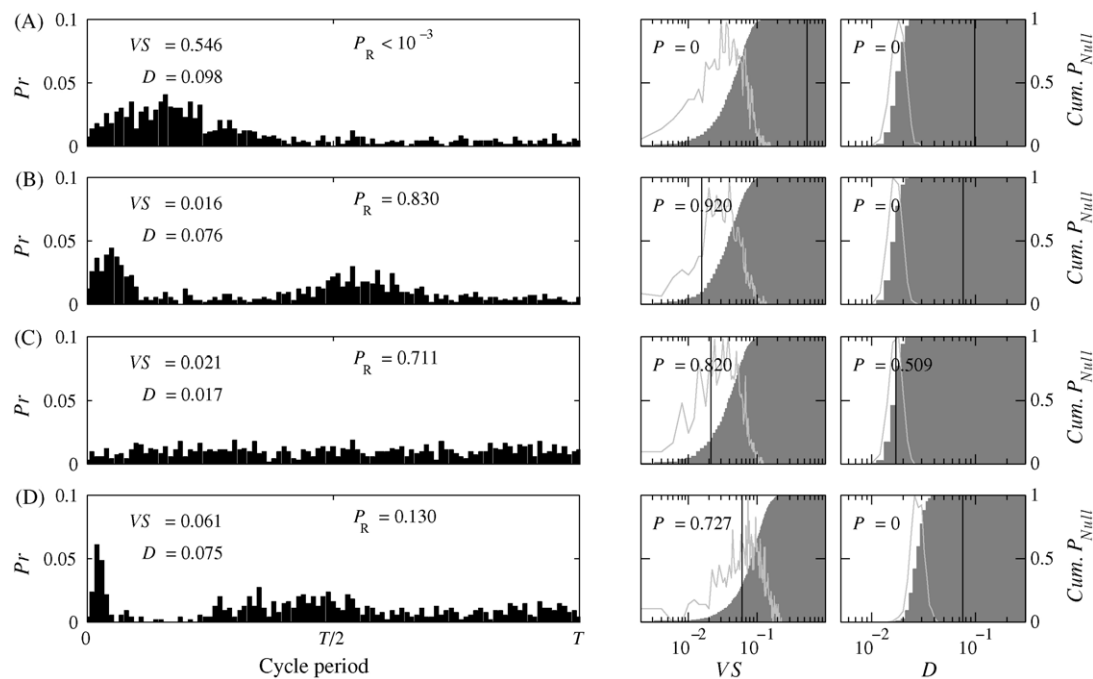


Fig. 1. (A and B) Simulated spike trains modeled with single (A) or double (B) Gaussian peaks within a cycle, with an arbitrary cycle length. (C) A spike train obtained by shuffling the serial ISI sequence of the spike train in (B). Each spike train in (A–C) had 750 spikes in 30 cycles. (D) Multiunit response to trains of upward FM sweeps (linearly changing from 2 to 10 kHz for 30 ms) repeated at 1 Hz ($T = 1$ s), recorded from marmoset auditory cortex. Left panels show Pr , the period histograms of spike trains normalized by the number of spikes. Horizontal axis spans full cycle length, and 1 s in (D). Insets indicate, D , VS and P value of the Rayleigh test. Center and right panels show the cumulative probability distribution of VS or D under null hypothesis, calculated from 1000 ISI-shuffled spike trains (gray area), distribution of null hypothetical VS or D underlying cumulative distribution (light gray curve), and VS or D of the original spike trains (vertical black line), respectively. These panels are plotted on the x -axis of VS or D by a logarithmic scale. The vertical height of the point at which the vertical black line crosses the border of the gray area, or equivalently the areas under the light gray line left of the black line, corresponds to the significance level of synchrony. Insets show P values above the critical level for rejection of the null hypothesis.

significant, highlighting the applicability of this method to neurophysiological data.

The center panels in Fig. 1 show the distributions of null hypothetical VS obtained from identical shuffled spike trains used for test of D . Even with this procedure, VS of the original spike train (B) and (D) were in the middle of null hypothetical VS, and determined to be not significant. Interestingly, the P values obtained for VS using shuffled spike trains were similar to those of the Rayleigh test, confirming the validity of acquiring the null hypothetical distribution of the parameter by means of shuffled spike trains.

We also examined synchrony of simulated spike trains modeled by unimodal period histograms differing extensively in background and peak heights with the Rayleigh test and D . Both generally concluded same results, except for cases where fewer spikes were included or where the peak was less prominent. In these cases, even when synchrony was significant by the Rayleigh test, it was not significant with D , suggesting a lower rate of type 1 error by the test of D .

3. Discussion

Motivation for the present study came from a need for a new measure of synchronization not based on vector expression. The proposed measure, D , quantifies synchronization, based on entropy. In the evaluation of this new measure, it became clear that D is robust not only when VS works properly, but also when VS does not, which may occur with a wide variety of physiological response patterns.

Lebedev et al. (1994), used the standard deviation, σ , of samples phase angles from that of mean vector as a measure of synchrony. The parameter they estimated $(\sigma_{\text{uniform}} - \sigma)/\sigma_{\text{uniform}}$, where σ_{uniform} is the maximum possible σ , derived from a distribution with uniform spike dispersion, could vary from 0 to 1 in parallel with VS. Although this was a reasonable measure of synchrony, he discarded data with multimodal temporal patterns from their analysis because of inherent difficulties with vector-based analysis.

There is a robust statistical test for uniformity of circular variable distribution of any type, called Rao's spacing test (Rao, 1976). The Rao's statistic, U_n , is an accumulation of deviations of arc length intervals from an expected interval for uniform probability distribution for a given number of observations. U_n is compared with an a priori derived U_n distribution. Importantly, however, both Rayleigh and Rao's tests assume circular variables positioned on circular circumference to be independent from each other, like the direction of a random walk (Greenwood and Durand, 1955; Rao, 1976). This point is often neglected when it has been applied to neurophysiological data.

Independence between variables contradicts the nature of the neuronal spike train, which often exhibits some dependence on its history, like an interaction between successive ISIs (Perkel et al., 1967). For example, firing rates during two halves of a cycle period may have a negative correlation, as

when high firing rate is followed by low firing rate and vice versa in individual cycles. However, such a conditional effect may not be necessarily reflected in a periodogram, but in an orderly sequence of ISI. Thus, degradation of serial order by shuffling ISI in the spike train gives us appropriate control of the data to obtain the null structure of spike train.

Briefly, we propose some applications of this technique. In terms of uniformity or randomness of distribution, entropy can be calculated for the distribution of any variable whose feature relies on its serial sequence, and is not limited to circular variables. Accordingly, the analysis proposed in the present report can be applicable to other distribution functions, like a peri-stimulus time histogram or temporal correlation function (Perkel et al., 1967). It is trivial to obtain the functions of shuffled spike train to examine the significance of its temporal structure. While we have focused on analyses of spike trains, the analysis is also applicable to other point process, such as miniature synaptic events (Fatt and Katz, 1952), which are often analyzed by a time histogram (Miledi and Thies, 1971). In ecological literature, E , E/E_{max} and D are used to represent species diversity, evenness, and redundancy, respectively (Legendre and Legendre, 1998).

One problem of using D is determination of the number of bins to construct the period histogram. Certainly, the binning procedure must maintain the temporal pattern of the distribution. However, if the observations included in the spike train are too few, it can cause a serious bias of entropy (Panzeri and Treves, 1996). Further study is needed to solve this problem. Given an adequate number of spikes, however, the technique proposed in the present report works sufficiently.

Note added in proof

After the manuscript has been accepted, authors realized that Nawrot et al. (2003) used a parameter, termed η in their study, similar to D proposed in the present study for different purpose.

Acknowledgements

We thank Mark Laubach, Eyal Kimchi, William D'Angelo for helpful comments on the manuscript. This work was supported by R01 NIH/NIDCD R01 04318 to T.A.H.

References

- Drew T, Doucet S. Application of circular statistics to the study of neuronal discharge during locomotion. *J Neurosci Meth* 1991;38:171–81.
- Fatt P, Katz B. Spontaneous subthreshold activity at motor nerve endings. *J Physiol* 1952;117:109–28.
- Goldberg JM, Brown PB. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J Neurophysiol* 1969;32:613–36.

- Greenwood JA, Durand D. The distribution of length and components of the sum of n random unit vectors. *Ann Math Stat* 1955;26:233–46.
- Lebedev MA, Denton JM, Nelson RJ. Vibration-entrained and pre-movement activity in monkey primary somatosensory cortex. *J Neurophysiol* 1994;72:1654–73.
- Legendre P, Legendre L. *Numerical ecology*. 2nd English ed. Amsterdam: Elsevier; 1998.
- Miledi R, Thies R. Tetanic and post-tetanic rise in frequency of miniature end-plate potentials in low-calcium solutions. *J Physiol* 1971;212:245–57.
- Moore GP, Perkel DH, Segundo JP. Statistical analysis and functional interpretation of neuronal spike data. *Annu Rev Physiol* 1966;28:493–522.
- Nawrot MP, Aertsen A, Rotter S. Elimination of response latency variability in neuronal spike trains. *Biol Cybern* 2003;88:321–34.
- Panzeri S, Treves A. Analytical estimates of limited sampling biases in different information measures. *Network: Comput Neural Syst* 1996;7:87–107.
- Perkel DJ, Gerstein GL, Moore GP. Neuronal spike trains and stochastic point processes I. The single spike train. *Biophys J* 1967;7:391–418.
- Rao JS. Some tests based on arc-length for the circle. *Sankhya: Indian J Stat* 1976;Ser. B 38:329–38.
- Werner G, Mountcastle VB. The variability of central neural activity in a sensory system, and its implications for the central reflection of sensory events. *J Neurophysiol* 1963;26:958–77.