Rapidly Converging Binless Approaches to the Calculation of Information in Spike Trains

J. D. Victor, F. Mechler, and D. S. Reich

Department of Neurology and Neuroscience, Weill Medical College of Cornell University, 1300 York Avenue, New York, NY 10021

ABSTRACT

Estimation of information in spike trains generally consists of several steps: (i) embedding spike trains into a space, (ii) clustering similar spike trains into groups, (iii) using a "plug-in" formula for transinformation based on how these groups relate to the stimuli, and (iv) estimating biases due to small sample size. Traditional approaches use binning as part of the embedding stage (i), with each bin corresponding to a separate dimension. Bins that are too wide lead to underestimates of information since temporal detail is lost, while bins that are too narrow lead to biases associated with extreme undersampling. Metric space methods (Victor & Purpura 1997) avoid the binning problem, but still may underestimate information due to the clustering at stage (ii). Jackknife estimators at stage (iv) can be superior to the standard (Treves & Panzeri 1995) bias correction, but no bias correction is effective when the amount of data is limited.

We present an alternative to stage (iii) that bypasses the difficulties associated with binning and clustering. We use linear, continuous embeddings of spike trains in low-dimensional spaces, and apply an asymptotically unbiased "binless" estimator of differential entropy (Kozachenko & Leonenko, 1987) directly to the embedded spike trains. Information is estimated from the difference between the entropy of the set of all spike trains, and the entropies of the spike trains elicited by each stimulus. In simulations, the rapid convergence properties of the binless entropy estimator lead to marked improvements in information estimates in the regime of limited data.

OVERVIEW

Limited data is not the only hurdle in estimation of entropy and information from neural data. An equally important issue is that the space of spike trains has a peculiar hybrid topology. It has a discrete character, since the number of spikes in any spike train must be an integer. But it also has a continuous character, owing to the continuous character of time. Reducing a spike train to a discrete series of integers via binning destroys this topology, in that small shifts in the time of a spike (that cause a spike to cross a bin boundary) results in as much of a change as moving a spike to an arbitrarily distant bin. The distinguishing feature of the present approach is that it exploits this natural hybrid topology of spike trains, and is thus more likely to be robust and efficient than procedures that ignore it.

These considerations are most relevant to situations in which responses are spike trains of limited duration, elicited by a small set of stimuli presented at discrete times. In other situations (e.g., analysis of extended responses to a rapid sequential presentation of a rich stimulus set), the temporal structure of the stimulus acts to destroy whatever temporal structure might be generated by neural processing. In such situations, bin-based methods work well (Strong et al., 1998), and the topology of the response space is less crucial.

For data quantities typical of laboratory experiments using transiently-presented stimuli, the present approach provides highly effective information estimates. At the expense of a modest decrease in precision, convergence (as a function of amount of data) and accuracy are substantially greater than that of binned approaches. Computations are straightforward and have a solid theoretical underpinning.

REFERENCES

Efron, B. and Tibshirani, R. (1993) An introduction to the bootstrap. Chapman and Hall: New York.

Kozachenko, L.F. and Leonenko, N.N. (1987) Sample estimate of the entropy of a random vector. Problemy Peredachi Informatsii, vol. 23, no. 2, pp. 9-16. Translated as Problems of Information Transmission PRITA9 23, 95-101, 1987.

Miller, G.A. (1955) Note on the bias on information estimates. Information Theory in Psychology; Problems and Methods II-B. 95-100.

Strong, S.P., Koberle, R., Ruyter van Steveninck, R.R. de, and Bialek, W. (1998) Entropy and information in neural spike trains. Phys. Rev. Lett. 80, 197-200.

Treves, A., and Panzeri, S. (1995) The upward bias in measures of information derived from limited data samples. Neural Computation 7, 399-407.

Victor, J.D., and Purpura, K.P. (1997) Metric-space analysis of spike trains: theory, algorithms, and application. Network 8, 127-164

ACKNOWLEDGEMENTS

Thanks to Bruce Knight, Chip Levy, and Liam Paninsky for helpful discussions.

Supported by NIH NEI EY9314.

BINLESS ESTIMATES OF ENTROPY

Entropy estimates of one-dimensional distributions

Let $p(x)$ be a continuous probability density on the real line . Our immediate goal is to estimate the differential entropy of $p(x)$, defined as

$$H_{diff} = -\int_{-\infty}^{\infty} p(x)\log_2 p(x)dx \quad , (1)$$

from a finite sample of observations $x_1, \ldots, x_N$ drawn according to $p(x)$. We seek an estimate that depends continuously on the individual observations, and exploits the continuous nature of p. The continuity assumption for p means that within a sufficiently small neighborhood of x_j, we can approximate p by a locally uniform distribution of density $p(x_j)$. In this neighborhood, observations are approximately distributed as a uniform Poisson process. From local Poisson statistics, one can determine $q(x, l)$, the probability that, after N-1 other samples have been drawn according to p, the nearest neighbor to a sample x is at a distance of least l :

$q(x, l) = \exp[-2l(N-1)p(x)]$. (2)

The expected value of $\log_2(l)$ can now be related to p. This leads to the estimate

$$H_{diff} \approx -\sum_{j=1}^{N} \frac{1}{N}\log_2 p(x_j) \approx \frac{1}{N}\sum_{j=1}^{N}\log_2 l_j + \log_2[2(N-1)] + \frac{g}{\ln(2)} \quad , (3)$$

where l_j is the observed distance from x_j to its nearest neighbor and

$$g = -\int_0^{\infty} e^{-v}\ln v dv$$

is the Euler-Mascheroni constant (≈0.5772156649). Eq. 3, essentially the one-dimensional case of eq. 2 of Kozachenko and Leonenko (1987). It was shown by those authors to be asymptotically unbiased provided that p obeys certain integrability conditions.

Entropy estimates of multi-dimensional distributions

The above analysis readily extends to multidimensional distributions $p(x)$, where x is a point in an r-dimensional Euclidean space. The relationship between $p(x)$ and the nearest-neighbor distribution $q(x, l)$ near must be modified to take into account the volume associated with a change in nearest-neighbor distance from l to $l+\Delta l$, which is proportional to the volume of an r-dimensional spherical shell of radius l. and thickness Δl . Following along the same lines as the one-dimensional analysis above, we find

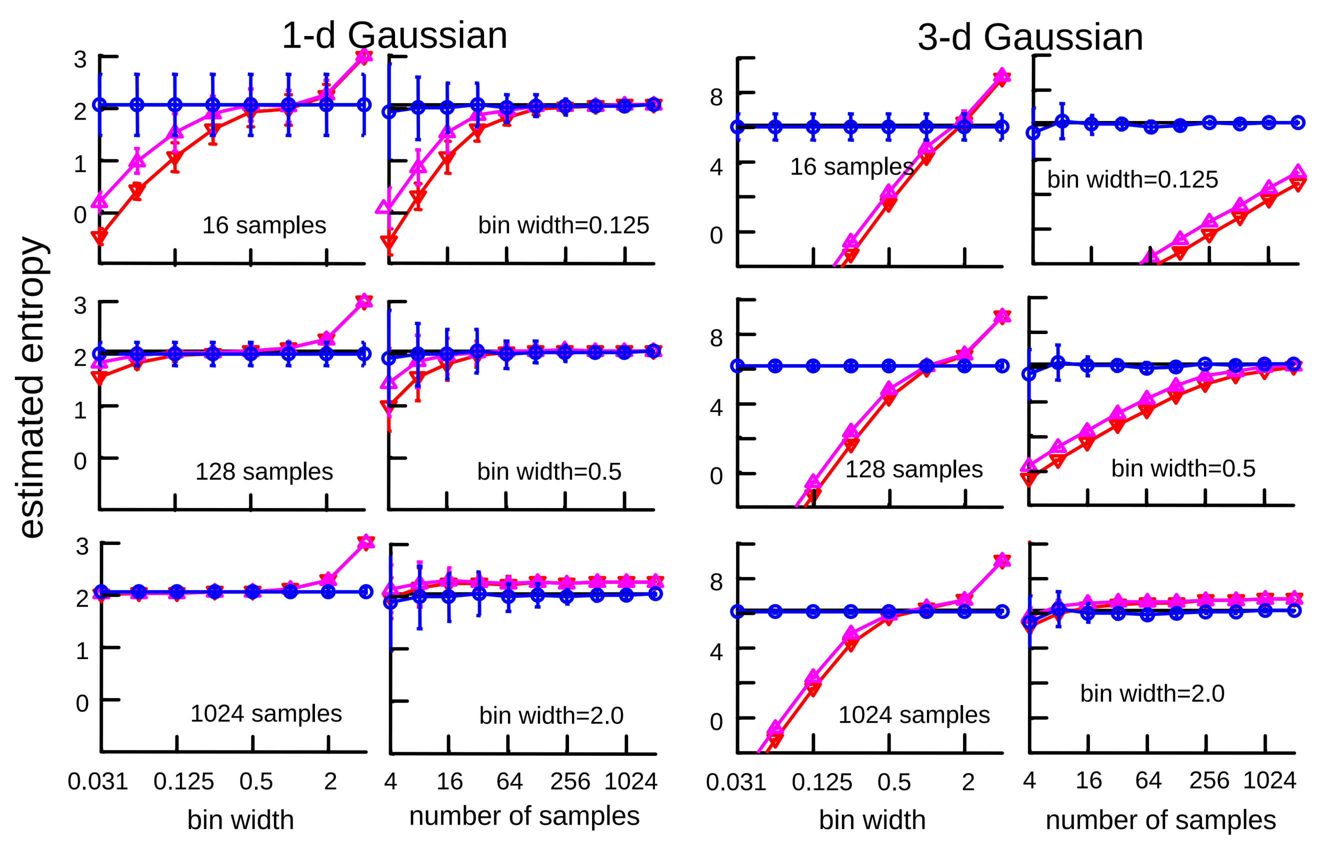
$$q(x, l) = \exp[-\frac{S_r l^r (N-1)p(x)}{r}] \qquad \text{where} \quad S_r = \frac{r p^{r/2}}{\Gamma(\frac{r}{2}+1)}$$

is the surface area of a unit r-dimensional spherical shell (S_1=2, S_2=2π, S_3=4π, …). This leads to

$$H_{diff} \approx -\sum_{j=1}^{N}\frac{1}{N}\log_2 p(x_j) \approx \frac{r}{N}\sum_{j=1}^{N}\log_2(l_j) + \log_2[\frac{S_r(N-1)}{r}] + \frac{g}{\ln(2)} \quad , (4)$$

which is eq. 2 of Kozachenko and Leonenko (1987).

SIMULATIONS: ENTROPY ESTIMATES



Comparison of estimators of differential entropy from a finite number of samples. Binned estimators with either the **classical** (eq. 10) or **jackknife** bias corrections depend strongly on bin width and converge much more slowly with increasing sample size than **binless** estimators, but have somewhat less scatter.

BINLESS ESTIMATES OF INFORMATION

Information estimates in a Euclidean space

Consider a discrete set of symbols $a_1, \ldots, a_s$, presented with probabilities $q_1, \ldots, q_s$, resulting in outputs x characterized by conditional probability densities

$p_k(x) = p(x \mid a_k)$

in a Euclidean space of dimension r. The transmitted information is given by

$$I = H_{diff} - \sum_{k=1}^{s} q_k H_{diff}(x \mid a_k) \quad , (5)$$

where H_{diff} is the differential entropy for the (unconditional) density $p(x)$, and $H_{diff}(x \mid a_k)$ is the differential entropy for the conditional density $p_k(x) = p(x \mid a_k)$. Applying eq. 4 leads to an expression for transmitted information that rests on a comparison between the minimum distance between spike trains elicited by the same symbol, and the unrestricted minimum distances:

$$I \approx \frac{r}{N}\sum_{j=1}^{N}\log_2(\frac{l_j}{l_j^*}) - \sum_{k=1}^{s}\frac{N_k}{N}\log_2\frac{N_k-1}{N-1} \quad . (6)$$

Here N_k is the number of presentations of the kth stimulus ($N_k = q_k N$), l_j is (as before) the minimum distance between the observation x_j and any other observation, and l_j^* is the minimum distance between the observation x_j and any other observation elicited by the same stimulus.

Information estimates for spike trains: stratification by spike count

Eq. 6 cannot be applied directly to neural data, since it requires that spike trains are represented by quantities lying within a Euclidean space of a particular dimension r. To describe a spike train containing n spikes requires n parameters – effectively one for each spike time. Thus, the set of spike trains of finite duration naturally breaks into a hierarchy of spaces: an n-dimensional space for the spike trains containing n spikes. We then break the transmitted information into two kinds of contributions:

$$I = I_{count} + \sum_{n=1}^{\infty} p(d(x)=n) I_{timing}(n) \quad , (7)$$

where $p(d(x)=n)$ is the probability that a response x contains exactly n spikes, I_{count} is the contribution due to the number of spikes elicited by each stimulus, and $I_{timing}(n)$ is the contribution due to the distribution of spike times of all responses containing n spikes.

We use the binless approach (eq. 6) to estimate $I_{timing}(n)$ and the usual plug-in estimates for discrete data to estimate I_{count}. That is,

$$I_{timing}(n) \approx \frac{r}{N(n)}\sum_{j=1}^{N(n)}\log_2(\frac{l_j}{l_j^*}) - \sum_{k=1}^{s}\frac{N(n,a_k)}{N(n)}\log_2\frac{N(n,a_k)-1}{N(n)-1} \quad , (8)$$

Interesting form! Second term is ~ maximum information if responses are fully segregated; first term is reduction due to interpenetration of responses. No γ, no S_r.

where r is the embedding dimension for n-element spike trains (see below), the j-summation is over all $N(n)$ spike trains containing exactly n spikes, $N(n, a_k)$ is the number of observed joint occurrences $N(n, a_k)$ in which a stimulus a_k elicits a response containing n spikes, and the minimum distances l_j and l_j^* consider only stimuli that elicit exactly n spikes. I_{count} can be estimated in terms of these quantities and the known stimulus probabilities q_k, by the "plug-in" estimate:

$$I_{count} \approx -\sum_{n=0}^{n_{max}}\sum_{k=1}^{s}\frac{N(n,a_k)}{N}\log_2 N(n,a_k) + \sum_{n=0}^{n_{max}}\frac{N(n)}{N}\log_2 N(n) + \sum_{k=1}^{s} q_k\log_2 q_k + I_{bias} \quad . (9)$$

The bias in the estimate of I_{count} can be estimated by the classical correction for entropy estimates (Miller 1955; Carlton, 1969; Treves and Panzeri, 1995)

$$I_{bias} = -\frac{(s-1)(n_{max}-1)}{2N\ln 2} \qquad (10)$$

or the jackknife estimate (Efron and Tibshirani 1993). The bias in the timing term of eq. 7 is asymptotically zero. This is because the binless estimate for $I_{timing}(n)$ and the naïve estimate for $p(d(x)=n)$ are mutually uncorrelated and each is asymptotically unbiased (Kozachenko and Leonenko 1987).

Information estimates for spike trains: an embedding

To embed the n-element spike trains into a Euclidean space, we first apply a monotonic time-warping transformation $t(t)$ so that so that the transformed spike times are approximately equally spaced in the interval [-1,1]. This transformation allows creation of approximately independent coordinates via the Legendre polynomials P_h, which are orthogonal on [-1,1]. The hth embedding coordinate c_h maps a spike train x_j containing n spikes at times $t_1, \ldots t_n$ into

$$c_h(x_j) = \sqrt{2h+1}\sum_{k=1}^{n} P_h(t(t_k)) \qquad (11).$$

By virtue of this normalization, if the n spike times within each spike train x_j are drawn at random from the pool of spike times, the mean-squared value of the hth coordinate of an spike train with n spikes will be n. Moreover, for spike times are drawn at random, coordinate values will be uncorrelated.

Two details

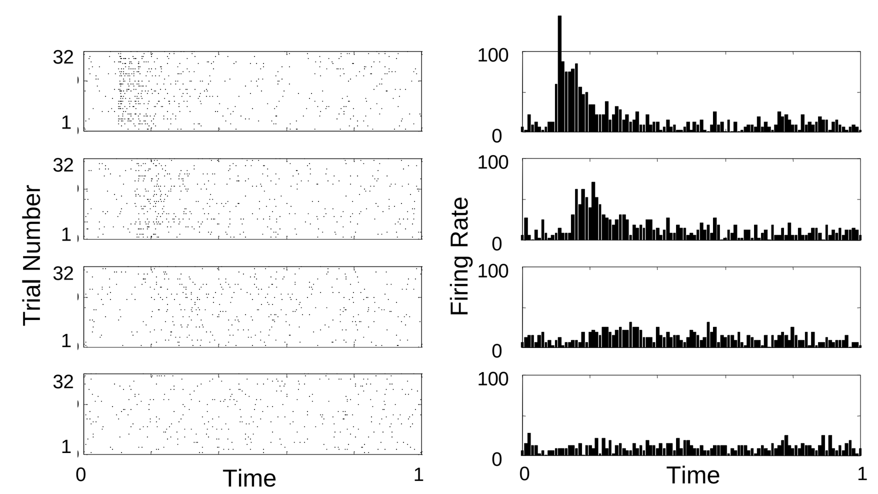
In estimating $I_{timing}(n)$, terms in eq. 8 will be undefined if the embedding results in nearest-neighbor distances of 0. Eq. 8 will also contain undefined terms if some stimuli elicit only one spike train containing n spikes, since those spike trains will have no nearest neighbors within their class from which to calculate l^*.

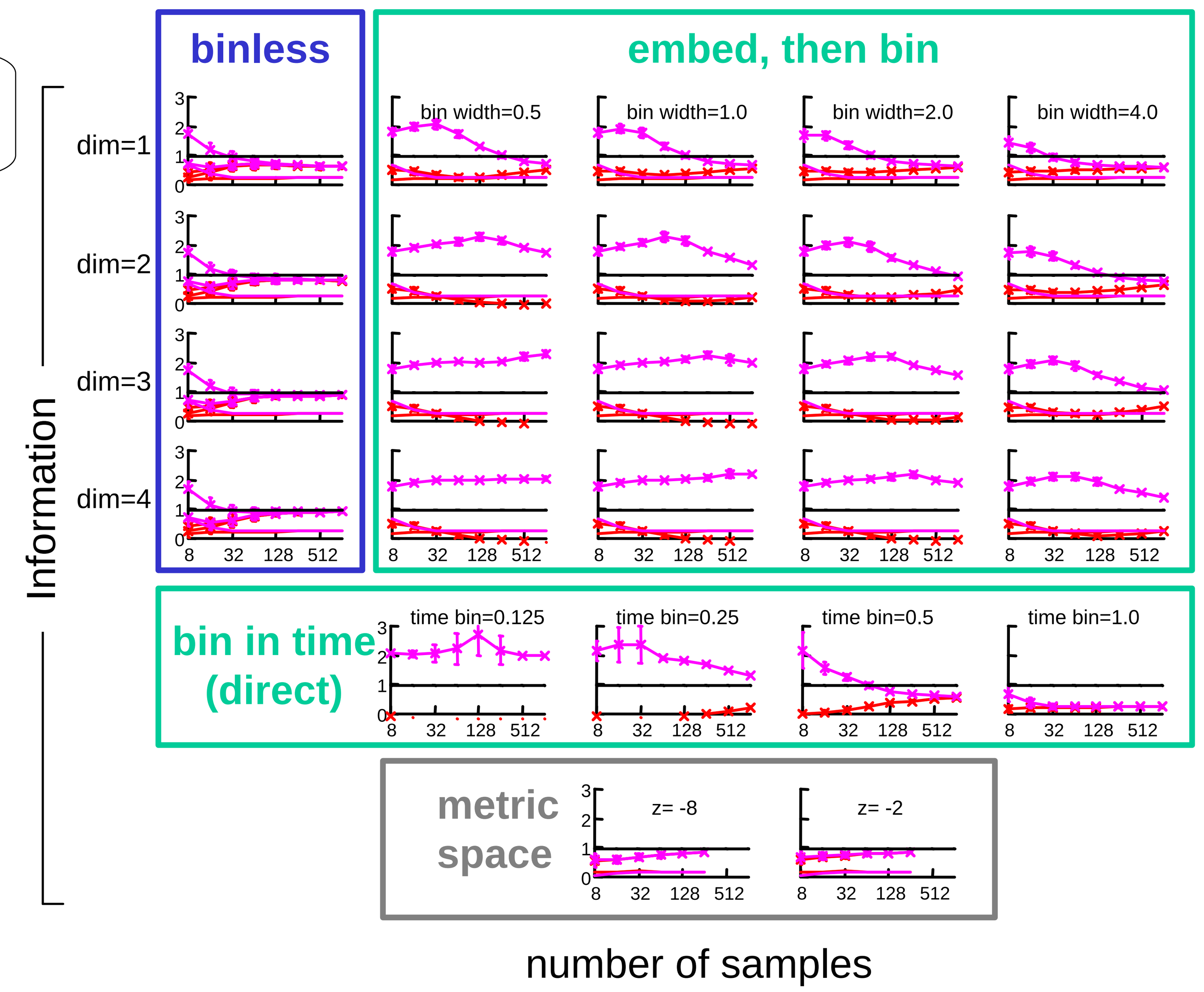
If there are "zero distances", we partition the set of spike trains containing exactly n spikes into disjoint sets $C_n, Z_{n,1}, Z_{n,2}, \ldots, Z_{n,b(n)}$ such that (i) C_n contains all of the spike trains that are not at a distance of zero from any other spike train and (ii) each $Z_{n,m}$ contains a maximal subset of spike trains that are at a distance of zero from each other. The subsets $Z_{n,m}$ are then used to refine the partition (eq. 7) of information into discrete and continuous components, $Z_{n,b(n)}$.

Singletons arise when eq. 8 applied to C_n to estimate $I_{continuous}$ but one or more spikes train in C_n are the sole representatives (among the n-spike responses in C_n) of the responses to its stimulus. Since this eventuality is a direct consequence of having a limited amount of data, we consider two ways of extrapolating to what the dataset might plausibly consist of, if we had additional data. One extreme is that additional observations would yield identical responses to this singleton. The other extreme is that additional observations would indicate that the observed singleton response is completely uninformative. As illustrated below, both arms of this bracketing strategy yield well-defined estimates of I_{timing}, and these rapidly converge as the number of samples increases.

SIMULATIONS: INFORMATION ESTIMATES

We applied the **binless** method summarized by eqs. (7), (8), and (9) to simulated spike trains (modulated Poisson processes) shown here, along with **binned** procedures, including embedding (eq. 11) followed by binning and time-domain binning, and the **metric-space** method of Victor and Purpura (1997).





Information estimates are debiased with either the **classical** (eq. 10) or **jackknife** bias corrections, and the spike count contribution to transmitted information (where applicable) is indicated by the colored lines without data points. For the **binless** estimators, the bracketing strategy described under "**Two details**" (above) leads to separate upper and lower estimates. They converge to an asymptotic value (horizontal black line) much more rapidly than the **binned** estimators. This is especially true for higher embedding dimensions. **Metric-space** estimators are rapidly convergent but slightly downwardly biased. Similar results were found in other simulations (not shown), including (a) Poisson spike trains differing only in rate, (b) sinusoidally modulated Poisson spike trains differing only in modulation phase, and (c) non-Poisson spike trains.