LETTER Communicated by Robert Kass

A Method for Selecting the Bin Size of a Time Histogram

Hideaki Shimazaki
shimazaki@ton.scphys.kyoto-u.ac.jp
Shigeru Shinomoto
shinomoto@scphys.kyoto-u.ac.jp
Department of Physics, Kyoto University, Kyoto 606-8502, Japan

The time histogram method is the most basic tool for capturing a time-dependent rate of neuronal spikes. Generally in the neurophysiological literature, the bin size that critically determines the goodness of the fit of the time histogram to the underlying spike rate has been subjectively selected by individual researchers. Here, we propose a method for objectively selecting the bin size from the spike count statistics alone, so that the resulting bar or line graph time histogram best represents the unknown underlying spike rate. For a small number of spike sequences generated from a modestly fluctuating rate, the optimal bin size may diverge, indicating that any time histogram is likely to capture a spurious rate. Given a paucity of data, the method presented here can nevertheless suggest how many experimental trials should be added in order to obtain a meaningful time-dependent histogram with the required accuracy.

1 Introduction

Neurophysiological studies are based on the idea that information is transmitted between cortical neurons by spikes (Johnson, 1996; Dayan & Abbott, 2001). A number of filtering algorithms have been proposed for estimating the instantaneous activity of an individual neuron or the joint activity of multiple neurons (DiMatteo, Genovese, & Kass, 2001; Wiener & Richmond, 2002; Sanger, 2002; Kass, Ventura, & Cai, 2003; Brockwell, Rojas, & Kass, 2004; Kass, Ventura, & Brown, 2005; Brown, Kass, & Mitra, 2004). The most basic and frequently used tool for spike rate estimation is the time histogram method. For instance, one aligns spike sequences at the onset of stimuli repeatedly applied to an animal and describes the response of a single neuron with a peristimulus time histogram (PSTH) or the responses of multiple neurons with a joint PSTH (Adrian, 1928; Gerstein & Kiang, 1960; Gerstein & Perkel, 1969; Abeles, 1982).

The shape of a PSTH is largely dependent on the choice of the bin size. With a bin size that is too large, one cannot represent the time-dependent spike rate. And with a bin size that is too small, the time histogram

 © 2007 Massachusetts Institute of Technology

fluctuates greatly, and one cannot discern the underlying spike rate. There is an appropriate bin size for each set of spike sequences, which is based on the goodness of the fit of the PSTH to the underlying spike rate. For most previously published PSTHs, however, the bin size has been subjectively selected by the authors.

For data points distributed compactly, there are classical theories about how the optimal bin size scales with the total number of data points n. It was proven that the optimal bin size scales as $n^{-1/3}$ with regard to the bar-graph-density estimator (Révész, 1968; Scott, 1979). It was recently found that for two types of infinitely long spike sequences, whose rates fluctuate either smoothly or jaggedly, the optimal bin sizes exhibit different scaling relations with respect to the number of sequences, timescale, and amplitude of rate modulation (Koyama & Shinomoto, 2004).

Though interesting, the scaling relations are valid only for a large amount of data and are of limited use in selecting a bin size. We devised a method of selecting the bin size of a time histogram from the spike data. In the course of our study, we realized that a theory on the empirical choice of the histogram bin size for a probability density function was presented by Rudemo (1982). Although applicable to a Poisson point process, this theory appears to have rarely been used by neurophysiologists in analyses of PSTHs. In the actual procedure of neurophysiological experiments, the number of trials (spike sequences) plays an important role in determining the resolution of a PSTH and thus in designing experiments. Therefore, it is preferable to have a theory that accords with the common protocol of neurophysiological experiments in which a stimulus is repeated to extract a signal from a neuron. Given a set of experimental data, we wish to not only determine the optimal bin size, but also estimate how many more experimental trials should be performed in order to obtain a resolution we deem sufficient.

For a small number of spike sequences derived from a modestly fluctuating rate, the estimated optimal bin size may diverge, implying that by constructing a PSTH, it is likely that one obtains spurious results for the spike rate estimation (Koyama & Shinomoto, 2004). Because a shortage of data underlies this divergence, one can carry out more experiments to obtain a reliable rate estimation. Our method can suggest how many sequences should be added in order to obtain a meaningful time histogram with the required accuracy. As an application of this method, we also show that the scaling relations of the optimal bin size that appears for a large number of spike sequences can be examined from a relatively small amount of data. The degree of the smoothness of an underlying rate process can be estimated by this method. In addition to a bar graph (piecewise constant) time histogram, we also designed a method for creating a line graph (piecewise linear) time histogram, which is superior to a bar graph in the goodness of the fit to the underlying spike rate and in comparing multiple responses to different stimulus conditions.

These empirical methods for the bin size selection for a bar and a line graph histogram, estimation of the number of sequences required for the histogram, and estimation of the scaling exponents of the optimal bin size were corroborated by theoretical analysis derived for a generic stochastic rate process. In the next section, we develop the bar graph (peristimulus) time histogram (Bar-PSTH) method, which is the most frequently used PSTH. In the appendix, we develop the line graph (peristimulus) time histogram (Line-PSTH) method.

2 Optimization of the Bar Graph Time Histogram

We consider sequences of spikes repeatedly recorded from a single neuron under identical experimental conditions. A recent analysis revealed that in vivo spike trains are not simply random, but possess interspike interval distributions intrinsic and specific to individual neurons (Shinomoto, Shima, & Tanji, 2003; Shinomoto, Miyazaki, Tamura, & Fujita, 2005). However, spikes accumulated from a large number of spike trains are in the majority mutually independent and can be regarded as being derived from a time-dependent Poisson point process (Snyder, 1975; Daley & Vere-Jones, 1988; Kass et al., 2005).

It would be natural to assess the goodness of the fit of the estimator $\hat{\lambda}_t$ to the underlying spike rate λ_t over the total observation period T by the mean integrated squared error (MISE),

$$\text{MISE} \equiv \frac{1}{T} \int_0^T E\,(\hat{\lambda}_t - \lambda_t)^2\, dt, \tag{2.1}$$

where E refers to the expectation over different realizations of point events, given λ_t. We begin with a bar graph time histogram as $\hat{\lambda}_t$ and explore a method to select the bin size that minimizes the MISE. The difficulty of the problem comes from the fact that the underlying spike rate λ_t is not known.

A bar graph time histogram is constructed simply by counting the number of spikes that belong to each bin of width Δ. For an observation period T, we obtain $N = \lfloor T/\Delta \rfloor$ intervals. The number of spikes accumulated from all n sequences in the ith interval is counted as k_i. The bar height at the ith bin is given as $k_i/n\Delta$. Figure 1 shows the schematic diagram for the construction of a bar graph time histogram.

Given a bin of width Δ, the expected height of a bar graph for $t \in [0, \Delta]$ is the time-averaged rate,

$$\theta = \frac{1}{\Delta} \int_0^\Delta \lambda_t\, dt. \tag{2.2}$$

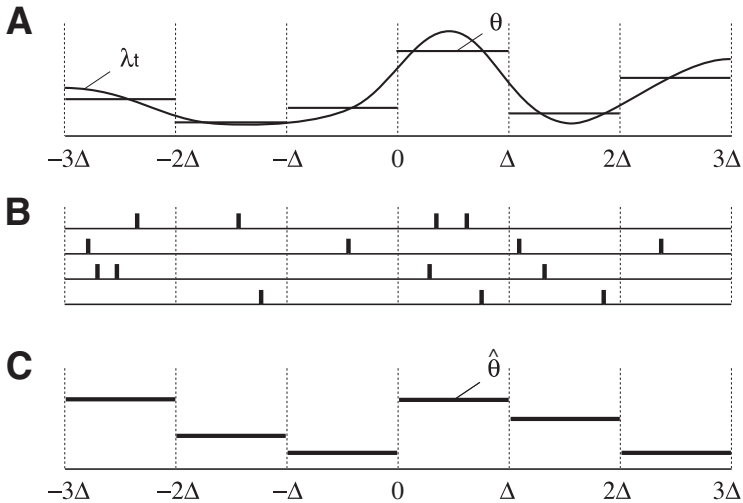

Figure 1: Bar-PSTH. (A) An underlying spike rate, λ_t. The horizontal bars indicate the time-averaged rates θ for individual bins of width Δ. (B) Sequences of spikes derived from the underlying rate. (C) A time histogram for the sample sequences of spikes. The estimated rate $\hat{\theta}$ is the total number of spikes k that entered each bin, divided by the number of sequences n and the bin size Δ.

The total number of spikes k from n spike sequences that enter this bin of width Δ obeys the Poisson distribution,

$$p(k \mid n\Delta\theta) = \frac{(n\Delta\theta)^k}{k!} e^{-n\Delta\theta}, \tag{2.3}$$

whose expectation is $n\Delta\theta$. The unbiased estimator for θ is denoted as $\hat{\theta} = k/(n\Delta)$, which is the empirical height of the bar graph for $t \in [0, \Delta]$.

2.1 Selection of the Bin Size. By segmenting the total observation period T into N intervals of size Δ, the MISE defined in equation 2.1 can be rewritten as

$$\text{MISE} = \frac{1}{\Delta} \int_0^{\Delta} \frac{1}{N} \sum_{i=1}^{N} \left\{ E\,(\hat{\theta}_i - \lambda_{t+(i-1)\Delta})^2 \right\} dt, \tag{2.4}$$

where $\hat{\theta}_i \equiv k_i/(n\Delta)$. Hereafter, we denote the average over the segmented rate $\lambda_{t+(i-1)\Delta}$ as an average over an ensemble of (segmented) rate functions

$\{\lambda_t\}$ defined on an interval of $t \in [0, \Delta]$, as

$$\text{MISE} = \frac{1}{\Delta} \int_0^{\Delta} \left\langle E\,(\hat{\theta} - \lambda_t)^2 \right\rangle dt. \tag{2.5}$$

The expectation E refers to the average over the spike count, or $\hat{\theta} = k/(n\Delta)$, given a rate function λ_t, or its mean value θ.

The MISE can be decomposed into two parts:

$$\text{MISE} = \langle E(\hat{\theta} - \theta)^2 \rangle + \frac{1}{\Delta} \int_0^{\Delta} \left\langle (\lambda_t - \theta)^2 \right\rangle dt. \tag{2.6}$$

The first and second terms are, respectively, the stochastic fluctuation of the estimator $\hat{\theta}$ around the expected mean rate θ and the averaged temporal fluctuation of λ_t around its mean θ over an interval of length Δ.

The second term of equation 2.6 can be decomposed further into two parts:

$$\frac{1}{\Delta} \int_0^{\Delta} \langle (\lambda_t - \langle\theta\rangle + \langle\theta\rangle - \theta)^2 \rangle dt = \frac{1}{\Delta} \int_0^{\Delta} \langle (\lambda_t - \langle\theta\rangle)^2 \rangle\, dt - \langle (\theta - \langle\theta\rangle)^2 \rangle. \tag{2.7}$$

The first term in equation 2.7 represents a mean squared fluctuation of the underlying rate λ_t from the mean rate $\langle\theta\rangle$ and is independent of the bin size Δ, because

$$\frac{1}{\Delta} \int_0^{\Delta} \langle (\lambda_t - \langle\theta\rangle)^2 \rangle\, dt = \frac{1}{T} \int_0^{T} (\lambda_t - \langle\theta\rangle)^2\, dt. \tag{2.8}$$

We define a cost function by subtracting this term from the original MISE,

$$\begin{aligned} C_n(\Delta) &\equiv \text{MISE} - \frac{1}{T} \int_0^{T} (\lambda_t - \langle\theta\rangle)^2\, dt \\ &= \langle E(\hat{\theta} - \theta)^2 \rangle - \langle (\theta - \langle\theta\rangle)^2 \rangle. \end{aligned} \tag{2.9}$$

This cost function corresponds to the risk function in Rudemo (1982, equation 2.3), obtained by direct decomposition of the MISE. The second term in equation 2.9 represents the temporal fluctuation of the expected mean rate θ for individual intervals of period Δ. As the expected mean rate θ is not an observable quantity, we have to replace the fluctuation of the expected mean rate with that of the observable estimator $\hat{\theta}$. Using the decomposition

rule for an unbiased estimator ($E\hat{\theta} = \theta$),

$$\langle E(\hat{\theta} - \langle E\hat{\theta}\rangle)^2\rangle = \langle E(\hat{\theta} - \theta)^2\rangle + \langle(\theta - \langle\theta\rangle)^2\rangle, \tag{2.10}$$

the cost function is transformed into

$$C_n(\Delta) = 2\left\langle E(\hat{\theta} - \theta)^2\right\rangle - \left\langle E(\hat{\theta} - \langle E\hat{\theta}\rangle)^2\right\rangle. \tag{2.11}$$

Due to the assumed Poisson nature of the point process, the number of spikes k counted in each bin obeys a Poisson distribution; the variance of k is equal to the mean. For the estimated rate defined as $\hat{\theta} = k/(n\Delta)$, this variance-mean relation corresponds to

$$E(\hat{\theta} - \theta)^2 = \frac{1}{n\Delta} E\hat{\theta}. \tag{2.12}$$

By incorporating equation 2.12 into equation 2.11, the cost function is given as a function of the estimator $\hat{\theta}$,

$$C_n(\Delta) = \frac{2}{n\Delta}\left\langle E\hat{\theta}\right\rangle - \left\langle E(\hat{\theta} - \langle E\hat{\theta}\rangle)^2\right\rangle. \tag{2.13}$$

The optimal bin size is obtained by minimizing the cost function $C_n(\Delta)$, as

$$\Delta^* \equiv \arg\min_{\Delta} C_n(\Delta). \tag{2.14}$$

By replacing the expectation with the sample spike count, the cost function (see equation 2.13) is converted into this useful recipe:

Algorithm 1: A Method for Bin Size Selection for a Bar-PSTH

i. Divide the observation period T into N bins of width Δ, and count the number of spikes k_i from all n sequences that enter the ith bin.

ii. Construct the mean and variance of the number of spikes $\{k_i\}$ as

$$\bar{k} \equiv \frac{1}{N}\sum_{i=1}^{N} k_i, \text{ and } v \equiv \frac{1}{N}\sum_{i=1}^{N}(k_i - \bar{k})^2.$$

iii. Compute the cost function:

$$C_n(\Delta) = \frac{2\bar{k} - v}{(n\Delta)^2}.$$

iv. Repeat i through iii while changing the bin size Δ to search for Δ^* that minimizes $C_n(\Delta)$.

2.2 Extrapolation of the Cost Function. With the method developed in the preceding section, we can determine the optimal bin size for a given set of experimental data. In this section, we develop a method to estimate how the optimal bin size decreases when more experimental trials are added to the data set: given n sequences, the method provides the cost function for $m(\neq n)$ sequences.

Assume that we are in possession of n spike sequences. The fluctuation of the expected mean rate $\langle(\theta - \langle\theta\rangle)^2\rangle$ in equation 2.10 is replaced with the empirical fluctuation of the time histogram $\hat{\theta}_n$ using the decomposition rule for the unbiased estimator $\hat{\theta}_n$ satisfying $E\hat{\theta}_n = \theta$,

$$\left\langle E(\hat{\theta}_n - \langle E\hat{\theta}_n\rangle)^2\right\rangle = \left\langle E(\hat{\theta}_n - \theta)^2\right\rangle + \left\langle(\theta - \langle\theta\rangle)^2\right\rangle. \tag{2.15}$$

The expected cost function for m sequences can thus be obtained by substituting the above equation into equation 2.9, yielding

$$C_m(\Delta|n) = \left\langle E(\hat{\theta}_m - \theta)^2\right\rangle + \left\langle E(\hat{\theta}_n - \theta)^2\right\rangle - \left\langle E(\hat{\theta}_n - \langle E\hat{\theta}_n\rangle)^2\right\rangle. \tag{2.16}$$

Using the variance-mean relation for a Poisson distribution, equation 2.12, and

$$E(\hat{\theta}_m - \theta)^2 = \frac{1}{m\Delta}E\hat{\theta}_m = \frac{1}{m\Delta}E\hat{\theta}_n, \tag{2.17}$$

we obtain

$$C_m(\Delta|n) = \left(\frac{1}{m} - \frac{1}{n}\right)\frac{1}{\Delta}\left\langle E\hat{\theta}_n\right\rangle + C_n(\Delta), \tag{2.18}$$

where $C_n(\Delta)$ is the original cost function (see equation 2.13) computed using the estimators $\hat{\theta}_n$. By replacing the expectation with sample spike count averages, the cost function for m sequences can be extrapolated with this formula, using the sample mean $\bar{k}$ and variance v of the numbers of spikes, given n sequences and the bin size Δ. The extrapolation method is summarized as algorithm 2:

Algorithm 2: A Method for Extrapolating the Cost Function for a Bar-PSTH

A. Construct the extrapolated cost function,

$$C_m(\Delta|n) = \left(\frac{1}{m} - \frac{1}{n}\right)\frac{\bar{k}}{n\Delta^2} + C_n(\Delta),$$

using the sample mean $\bar{k}$ and variance v of the number of spikes obtained from n sequences of spikes. $C_n(\Delta)$ is the cost function computed for n sequences of spikes with algorithm 1.

B. Search for Δ_m^* that minimizes $C_m(\Delta|n)$.

C. Repeat A and B while changing m, and plot $1/\Delta_m^*$ versus $1/m$ to search for the critical value $1/m = 1/\hat{n}_c$ above which $1/\Delta_m^*$ practically vanishes.

It may come to pass that the original cost function $C_n(\Delta)$ computed for n spike sequences does not have a minimum or has a minimum at a bin size comparable to the observation period T. Because of the paucity of data, one may consider carrying out more experiments to obtain a reliable rate estimation. The critical number of sequences n_c above which the cost function has a finite bin size Δ^* may be estimated in the following manner. With a large Δ, the cost function can be expanded as

$$C_n(\Delta) \sim \mu \left(\frac{1}{n} - \frac{1}{n_c} \right) \frac{1}{\Delta} + u \frac{1}{\Delta^2}, \tag{2.19}$$

where we have introduced n_c and u, which are independent of n. The optimal bin size undergoes a phase transition from the vanishing $1/\Delta^*$ for $n < n_c$ to a finite $1/\Delta^*$ for $n > n_c$. The inverse optimal bin size is expanded in the vicinity of n_c as $1/\Delta^* \propto (1/n - 1/n_c)$. We can estimate the critical value $\hat{n}_c$ (see Figure 3) by applying this asymptotic relation to the set of $\hat{\Delta}_m^*$ estimated from $C_m(\Delta|n)$ for various values of m:

$$\frac{1}{\Delta_m^*} \propto \left(\frac{1}{m} - \frac{1}{\hat{n}_c} \right). \tag{2.20}$$

The minimum number of sequences required for the construction of a Bar-PSTH is estimated from $\lceil \hat{n}_c \rceil$. It should be noted that there are cases that the optimal bin size exhibits a discontinuous divergence from a finite value. Even in such cases, the plot of $\{1/m, 1/\Delta^*\}$ could be useful in exploring the discontinuous transition from nonvanishing values of $1/\Delta^*$ to practically vanishing values.

In the opposite extreme, with a sufficiently large number of spike sequences, our method selects a small bin size. It is known that the optimal bin size exhibits a power law scaling with respect to the number of sequences n. The exponent of the scaling relation depends on the smoothness of the underlying rate (Koyama & Shinomoto, 2004). Given a large number of spike sequences, the method for extrapolating the cost function can also be used to estimate the optimal bin size for a larger number of spike sequence $m > n$, and further estimate the scaling exponent representing the smoothness of the underlying rate.

2.3 Theoretical Analysis on the Optimal Bin Size. To verify the empirical methods, we obtain the theoretical cost function of a Bar-PSTH directly from a process with a known underlying rate. Note that this theoretical cost function is not available in real experimental conditions in which the underlying rate is not known.

The estimator $\hat{\theta} \equiv k/(n\Delta)$ is a uniformly minimum variance unbiased estimator (UMVUE) of θ, which achieves the lower bound of the Cramér-Rao inequality (Blahut, 1987; Cover & Thomas, 1991),

$$E(\hat{\theta} - \theta)^2 = \left[-\sum_{k=0}^{\infty} p\,(k \mid n\Delta\theta)\, \frac{\partial^2 \log p\,(k \mid n\Delta\theta)}{\partial\theta^2} \right]^{-1} = \frac{\theta}{n\Delta}. \tag{2.21}$$

Inserting this into equation 2.9, the cost function is represented as

$$\begin{aligned} C_n(\Delta) &= \frac{\langle\theta\rangle}{n\Delta} - \langle(\theta - \langle\theta\rangle)^2\rangle \\ &= \frac{\mu}{n\Delta} - \frac{1}{\Delta^2}\int_0^{\Delta}\int_0^{\Delta} \phi\,(t_1 - t_2)\; dt_1 dt_2, \end{aligned} \tag{2.22}$$

where $\mu = \langle\theta\rangle$ is the mean rate, and $\phi(t_1 - t_2) = \langle(\lambda_{t_1} - \mu)(\lambda_{t_2} - \mu)\rangle$ is the autocorrelation function of the rate fluctuation, $\lambda_t - \mu$. To obtain the last equation, we used

$$\begin{aligned} \langle(\theta - \langle\theta\rangle)^2\rangle &= \left\langle \left\{ \frac{1}{\Delta}\int_0^{\Delta} (\lambda_t - \mu)\, dt \right\}^2 \right\rangle \\ &= \frac{1}{\Delta^2}\int_0^{\Delta}\int_0^{\Delta} \langle(\lambda_{t_1} - \mu)(\lambda_{t_2} - \mu)\rangle\, dt_1 dt_2. \end{aligned} \tag{2.23}$$

The cost function with a large bin size can be rewritten as

$$\begin{aligned} C_n(\Delta) &= \frac{\mu}{n\Delta} - \frac{1}{\Delta^2}\int_{-\Delta}^{\Delta} (\Delta - |t|)\phi(t)\, dt \\ &\sim \frac{\mu}{n\Delta} - \frac{1}{\Delta}\int_{-\infty}^{\infty} \phi(t)\, dt + \frac{1}{\Delta^2}\int_{-\infty}^{\infty} |t|\phi(t)\, dt, \end{aligned} \tag{2.24}$$

based on the symmetry $\phi(t) = \phi(-t)$ for a stationary process. Equation 2.24 can be identified with equation 2.19 with parameters

$$n_c = \frac{\mu}{\int_{-\infty}^{\infty} \phi(t)\, dt} \tag{2.25}$$

$$u = \int_{-\infty}^{\infty} |t|\phi(t)\,dt. \tag{2.26}$$

For the process giving the correlation of the form $\phi(t) = \sigma^2 e^{-|t|/\tau}$ and the process giving a gaussian correlation $\phi(t) = \sigma^2 e^{-t^2/\tau^2}$ used in the simulation in the next section, the critical numbers are obtained as $n_c = \mu/2\sigma^2\tau$ and $n_c = \mu/\sigma^2\tau\sqrt{\pi}$, respectively.

Based on the same theoretical cost function, equation 2.22, we can also derive scaling relations of the optimal bin size, which is achievable with a large number of spike sequences. With a small Δ, the correlation of rate fluctuation $\phi(t)$ can be expanded as $\phi(t) = \phi(0) + \phi'(0_+)|t| + \frac{1}{2}\phi''(0)t^2 + O\left(|t|^3\right)$, and we obtain an expansion of equation 2.22 with respect to Δ,

$$C_n(\Delta) = \frac{\mu}{n\Delta} - \phi(0) - \frac{1}{3}\phi'(0_+)\Delta - \frac{1}{12}\phi''(0)\Delta^2 + O(\Delta^3). \tag{2.27}$$

The optimal bin size Δ^* is obtained from $dC_n(\Delta)/d\Delta = 0$. For a rate that fluctuates smoothly in time, the correlation function is a smooth function of t, resulting in $\phi'(0) = 0$ due to the symmetry $\phi(t) = \phi(-t)$. In this case, we obtain the scaling relation

$$\Delta^* \sim \left(-\frac{6\mu}{\phi''(0)n}\right)^{1/3}. \tag{2.28}$$

For a rate that fluctuates in a zigzag pattern, in which the correlation of rate fluctuation has a cusp at $t = 0$ ($\phi'(0_+) < 0$), we obtain the scaling relation

$$\Delta^* \sim \left(-\frac{3\mu}{\phi'(0_+)n}\right)^{1/2}, \tag{2.29}$$

by ignoring the second-order term. Examples that exhibit this type of scaling are the Ornstein-Uhlenbeck process and the random telegraph process (Kubo, Toda, & Hashitsume, 1985; Gardiner, 1985).

The scaling relations, equation 2.28 and equation 2.29, are the generalization of the ones found in Koyama and Shinomoto (2004) for two specific time-dependent Poisson processes.

3 Application of the Method to Spike Data

In this section, we apply the methods for the bar graph (peristimulus) time histogram (Bar-PSTH) and the line graph (peristimulus) time histogram (Line-PSTH) to (1) spike sequences (point events) generated by the simulations of time-dependent Poisson processes and (2) spike sequences recorded

from a neuron in area MT. The methods for a Line-PSTH are summarized in the appendix.

3.1 Application to Simulation Data. We applied the method for estimating the optimal bin size of a Bar-PSTH to a set of spike sequences derived from a time-dependent Poisson process. Figure 2A shows the empirical cost function $C_n(\Delta)$ computed from a set of n spike sequences. The empirical cost function is in good agreement with the theoretical cost function computed from the mean spike rate μ and the correlation $\phi(t)$ of the rate fluctuation, according to equation 2.22. In Figure 2D, a time histogram with the optimal bin size is compared with those with nonoptimal bin sizes, demonstrating the effectiveness of optimizing the bin size.

Algorithm 2 provides an estimate of how many sequences should be added for the construction of a Bar-PSTH with the resolution we deem sufficient. In this method, the cost function of m sequences is obtained by modifying the original cost function, $C_n(\Delta)$, computed from the spike count statistics of n sequences of spikes. In subsequent applications of this extrapolation method, the original cost function, $C_n(\Delta)$, was obtained by averaging over the initial partitioning positions.

We applied this extrapolation method for a Bar-PSTH to a set of spike sequences derived from the smoothly regulated Poisson process. Figure 3A depicts the extrapolated cost function $C_m(\Delta|n)$ for several values of m, computed from a given set of $n(= 30)$ spike sequences. Figure 3B represents the dependence of an inverse optimal bin size $1/\Delta_m^*$ on $1/m$. The inverse optimal bin size, $1/\Delta_m^*$, stays near 0 for $1/m > 1/\hat{n}_c$ and departs from 0 linearly with m for $1/m \leq 1/\hat{n}_c$. By fitting the linear function, we estimated the critical number of sequences $\hat{n}_c$. In Figure 3C, the critical number of sequences $\hat{n}_c$ estimated from a smaller or larger n is compared to the theoretical value of n_c computed from equation 2.25. It is noteworthy that the estimated $\hat{n}_c$ approximates the theoretical n_c well even from a fairly small number of spike sequences, with which the estimated optimal bin size diverges.

We also constructed a method for selecting the bin size of a Line-PSTH, which is summarized as algorithm 3 in the appendix. Figure 4 compares the optimal Bar-PSTH and the optimal Line-PSTH obtained from the same set of spike sequences, demonstrating the superiority of the Line-PSTH to the Bar-PSTH in the sense of the MISE. In addition, the Line-PSTH is suitable for comparing multiple time-dependent rates, as is the case for filtering methods. The extrapolation method for a Line-PSTH is summarized as algorithm 4 in the appendix.

With the extrapolation method for a Bar-PSTH (see algorithm 2), one can also estimate how much the optimal bin size decreases (i.e., the resolution increases) with the number of spike sequences. Figure 5A shows log-log plots of the optimal bin sizes Δ_m^* versus m with respect to two rate processes that fluctuate either smoothly according to the stochastic process characterized by the mean spike rate μ and the correlation of the rate

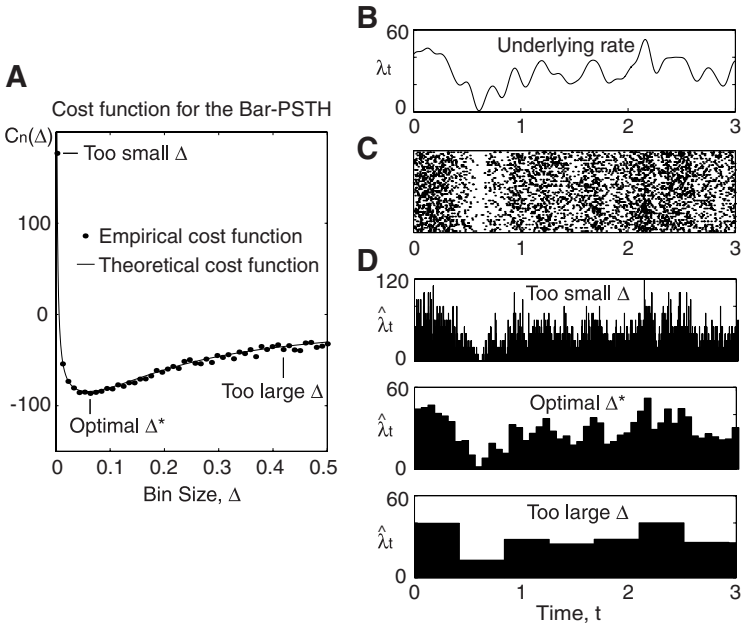

Figure 2: Construction of the optimal Bar-PSTH from spike sequences. (A) Dots: An empirical cost function, $C_n(\Delta)$, computed from spike data according to algorithm 1. Here, 50 sequences of spikes for an interval of 20 [s] are derived from a time-dependent Poisson process characterized by the mean rate μ and the correlation of the rate fluctuation $\phi(t) = \sigma^2 e^{-t^2/\tau^2}$, with $\mu = 30$ [spikes/s], $\sigma = 10$ [spikes/s], and $\tau = 0.1$ [s]. Solid line: The theoretical cost function computed directly from the underlying fluctuating rate using equation 2.22. (B) The underlying fluctuating rate λ_t. (C) Spike sequences derived from the rate. (D) Time histograms made using three types of bin sizes: too small, optimal, and too large.

process $\phi(t) = \sigma^2 e^{-t^2/\tau^2}$ or in a zigzag pattern according to the Ornstein-Uhlenbeck process characterized by the mean rate μ and the correlation of the rate process $\phi(t) = \sigma^2 e^{-|t|/\tau}$. These plots exhibit power law scaling relations with distinctly different scaling exponents. The estimated exponents (-0.34 ± 0.04 for the smooth rate process and -0.56 ± 0.04 for the zigzag rate process) are close to the exponents of $m^{-1/3}$ and $m^{-1/2}$ that were obtained analytically as in equations 2.28 and 2.29 (see also Koyama & Shinomoto, 2004). In this way, we can estimate the degree of smoothness of the underlying rate from a reasonable amount of spike data.

With the extrapolation method for a Line-PSTH (see algorithm 4 in the appendix), the scaling relations for a Line-PSTH can be examined in a

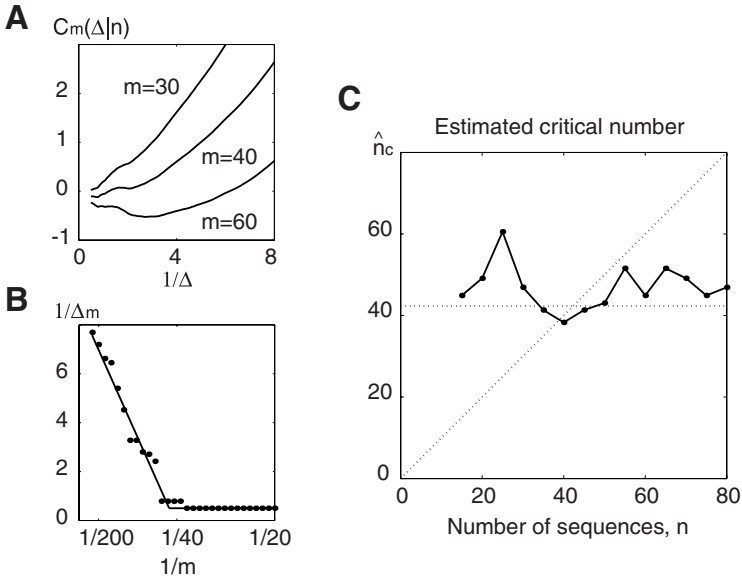

Figure 3: Extrapolation method. (A) Extrapolated cost functions $C_m(\Delta|n)$ of a Bar-PSTH for several values of m, computed from 30 sequences of spikes derived from a Poisson process characterized by the mean rate μ and the correlation of the rate fluctuation $\phi(t) = \sigma^2 e^{-t^2/\tau^2}$, with $\mu = 30$, $\sigma = 2$, and $\tau = 0.1$. The modestly fluctuating rate with $\sigma = 2$ is used as compared with the rate with $\sigma = 10$ used in Figure 1. (B) The inverse optimal bin size $1/\Delta_m^*$ plotted against $1/m$. A solid line is a linear function fitted to the data. (C) The critical number of sequences $\hat{n}_c$ extrapolated from a smaller or larger number of spike sequences. The horizontal axis represents the number of spike sequences n used to obtain the extrapolated cost function $C_m(\Delta|n)$. The vertical axis represents an extrapolated critical number $\hat{n}_c$. The dashed lines represent a theoretical value n_c computed using equation 2.25 and a diagonal.

similar manner. Figure 5B represents the optimal bin sizes computed for two rate processes that fluctuate either smoothly or in a zigzag pattern. The estimated exponents are -0.24 ± 0.04 for the smooth rate process and -0.50 ± 0.05 for the zigzag rate process. The exponents obtained by the extrapolation method are similar to the analytically obtained exponents, $m^{-1/5}$ and $m^{-1/2}$ respectively (see equations A.12 and A.13). Note that for the smoothly fluctuating rate process, the scaling relation for the Line-PSTH is $m^{-1/5}$, whereas the scaling relation for a Bar-PSTH is $m^{-1/3}$. In contrast, for the rate process that fluctuates jaggedly, the exponents of the scaling relations for both a Bar-PSTH and a Line-PSTH are $m^{-1/2}$.

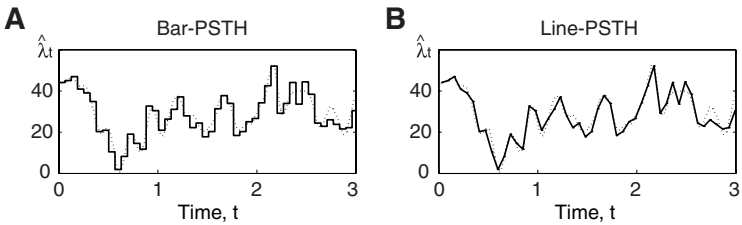

Figure 4: Comparison of the optimal Bar-PSTH and the optimal Line-PSTH. The spike data used are the same as the data used in Figure 2. (A) The optimal Bar-PSTH constructed from the spike sequences derived from an underlying fluctuating rate process (dashed line). (B) The optimal Line-PSTH constructed from the same set of spike sequences.

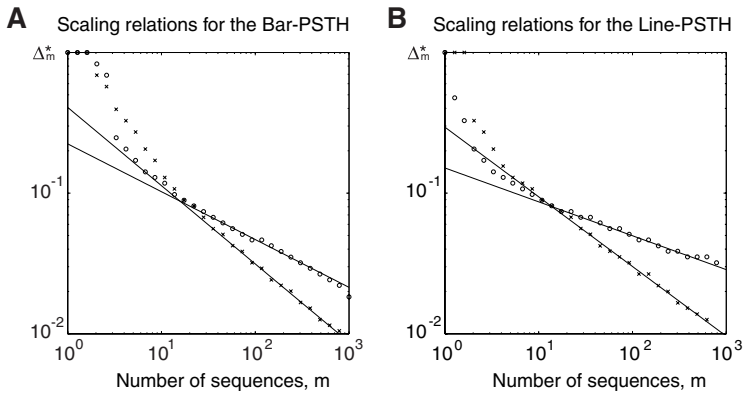

Figure 5: Scaling relations for the Bar-PSTH and the Line-PSTH. The optimal bin sizes Δ_m^* are estimated from the extrapolated cost function $C_m(\Delta|n)$, computed from 100 sequences of spikes. The optimal bin sizes Δ_m^* versus m is shown in log-log scale. Two types of rate processes were examined: one that fluctuates smoothly, which is characterized by the correlation of the rate fluctuation $\phi(t) = \sigma^2 e^{-t^2/\tau^2}$, and the other that fluctuates jaggedly, which is characterized by the correlation of the rate fluctuation $\phi(t) = \sigma^2 e^{-|t|/\tau}$. The parameter values are the same as the ones used in Figure 2. (A) Log-log plots of the optimal bin size with respect to m for a Bar-PSTH. Lines are fitted to the data in an interval of $m \in [50, 500]$, whose regression coefficients, -0.34 ± 0.04 and -0.56 ± 0.04, correspond to the scaling exponents for a Bar-PSTH. (B) Log-log plots of the optimal bin size with respect to m for a Line-PSTH. Lines are fitted to the data in an interval of $m \in [50, 500]$, whose regression coefficients, -0.24 ± 0.04 and -0.50 ± 0.05, correspond to the scaling exponents for a Line-PSTH.

3.2 Application to Experimental Data. We also applied our method for optimizing a Bar-PSTH to publicly available neuronal spike data (Britten, Shadlen, Newsome, & Movshon, 2004). The details of the experimental methods are described in Newsome, Britten, and Movshon (1989) and Britten, Shadlen, Newsome, & Movshon (1992). It was reported that the neurons in area MT exhibit highly reproducible temporal responses to identical visual stimuli (Bair & Koch, 1996). The estimation of a time-dependent rate by a PSTH is, however, sensitive to the choice of the bin size. Therefore it would be preferable for such an appraisal of the reproducibility to be tested with our objective method of optimizing the bin size. To make a reliable estimate of the optimal bin size from the spike sequences of a short observation period, we took an average of the cost functions computed under different partitioning positions.

We examine here the data recorded from a neuron under the repeated application of a random dot visual stimulus with 3.2% of coherent motion (w052, nsa2004.1, Britten et al., 2004). The method for a Bar-PSTH was applied to the data, with close attention paid to how the optimal bin size changes with the number of spike sequences sampled. Figure 6 depicts the results for the first $n = 5$, the first $n = 20$ and the total $n = 50$ sequences of spikes.

The optimal bin size practically diverges ($\Delta^* = 1000$ [ms]) for the $n = 5$ sequences, implying that with this small amount of data, a discussion about the time-dependent response does not make sense, as far as we rely on the histogram method. Even in this stage, it is possible to extrapolate the cost function by means of algorithm 2. The critical number of trials, above which the optimal bin size is finite, was estimated as $\hat{n}_c \approx 12$. By increasing the number of spikes to $n = 20$, we obtained the optimal bin size $\Delta^* = 33$ [ms], implying that a discussion about the neuronal response is possible based on this amount of data. The critical number of trials estimated from this data set ($n = 20$) is $\hat{n}_c \approx 10$. The optimal bin size for the total 50 sequences is $\Delta^* = 23$ [ms]. The critical number of sequences estimated from the total 50 sequences is $\hat{n}_c \approx 12$.

Algorithm 1 confirms that the temporal rate modulation can be discussed with a sufficiently large number of sequences, such as $n = 20$ or $n = 50$. With algorithm 2, three sets of sequences with $n = 5, 20$, and 50 suggest the critical number of sequence to be $\hat{n}_c \approx 10$–12.

4 Discussion

We have developed a method for selecting the bin size, so that the Bar-PSTH (see section 2) or the Line-PSTH (see the appendix) best represents the (unknown) underlying spike rate. The suitability of this method was demonstrated by applying it to not only model spike sequences generated by time-dependent Poisson processes, but also real spike sequences recorded from cortical area MT of a monkey.

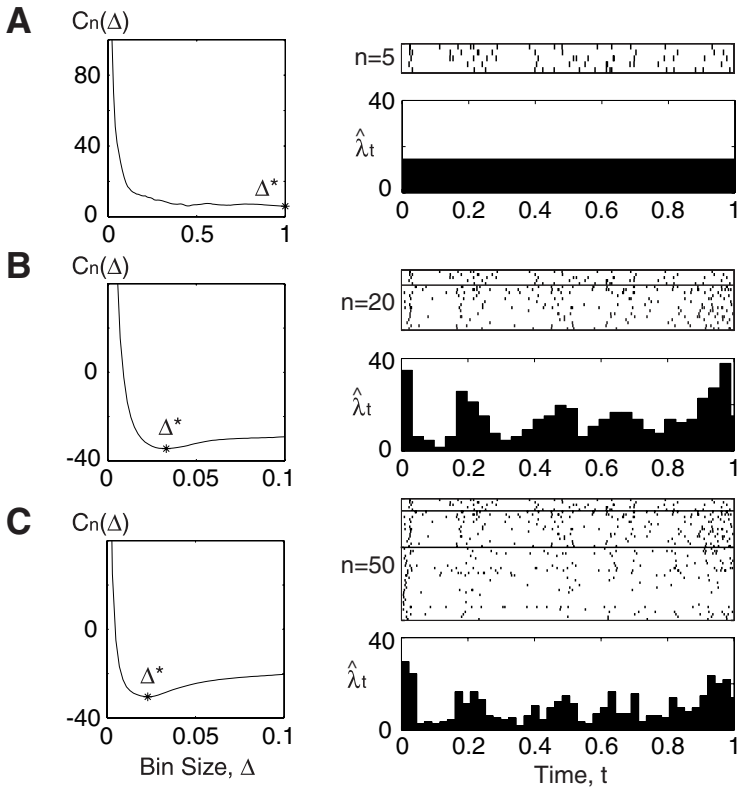

Figure 6: The Bar-PSTHs for the spike sequences recorded from a MT neuron (w052 in nsa2004.1; Britten et al., 2004). Left: The cost function (averaged over different partitioning positions). Right: Spike sequences sampled and the optimal Bar-PSTH for the first 1 [s] of the recording period 2 [s]. (A) The first 5 spike sequences. (B) The first 20 spike sequences that include 5 sequences of A. (C) The complete 50 spike sequences that include 20 sequences of B.

For a small number of spike sequences derived from a modestly fluctuating rate, the cost function does not have a minimum ($C_n(T) < C_n(\Delta)$ for any $\Delta < T$), or has a minimum at a bin size Δ that is comparable to the observation period T, implying that the time-dependent rate cannot be captured by means of the time histogram method. Our method can nevertheless extrapolate the cost function for any number of spike sequences and suggest how many sequences should be added in order to obtain a meaningful time histogram with the resolution we require. The model data and real data illustrated that the optimal bin size may diverge ($\Delta^* = O(T)$), and even under such conditions, our extrapolation method works reasonably well.

In this study, we adopted the mean integrated squared error (MISE) as measuring the goodness of the fit of a PSTH to the underlying spike rate. There were studies of the density estimation based on other standards such as the Kullback-Leibler divergence (Hall, 1990) and the Hellinger distance (Kanazawa, 1993). To our knowledge, however, there are yet no practical algorithms based on these standards that optimize the bin size solely with raw data. It is interesting to explore practical methods based on other standards to the estimation of the time-dependent rate and compare them with the MISE criterion.

Use of regular (equal) bin size is a constraint particular to our method. The regular histogram is suitable for stationary time-dependent rates. In practical applications, however, the rate does not necessarily fluctuate in a stationary fashion but can change abruptly during a localized period of time. In such a situation, a histogram with variable bin width can better fit the underlying rate than a regular histogram can. In order to make a better estimate of the underlying rate in the sense of the MISE, it is desirable to develop an adaptive method that adjusts the bin size over time. It is also desirable to develop the optimization method for the Bar-PSTH and the Line-PSTH into a method for higher-order spline fittings that can be compared with filtering methods. Nevertheless, as long as the Bar-PSTHs or the Line-PSTHs are used as conventional rate-estimation tools, our method for selecting the bin size should be used for their construction.

Appendix: Optimization of the Line Graph Time Histogram

A line graph can be constructed by simply connecting top centers of adjacent bar graphs of the height $k_i/(n\Delta)$. Figure 7 schematically shows the construction of a line graph time histogram. For the same set of spike sequences, the optimal bin size (window size) of a Line-PSTH is, however, different from that of a Bar-PSTH. We develop here a method for selecting the bin size for a Line-PSTH.

The expected heights of adjacent bar graphs for intervals of $[-\Delta, 0]$ and $[0, \Delta]$ are

$$\theta^- \equiv \frac{1}{\Delta}\int_{-\Delta}^{0} \lambda_t \, dt,$$

$$\theta^+ \equiv \frac{1}{\Delta}\int_{0}^{\Delta} \lambda_t \, dt.$$

The expected line graph L_t in an interval of $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$ is a line connecting top centers of these bar graphs,

$$L_t = \frac{\theta^+ + \theta^-}{2} + \frac{\theta^+ - \theta^-}{\Delta} t. \tag{A.1}$$

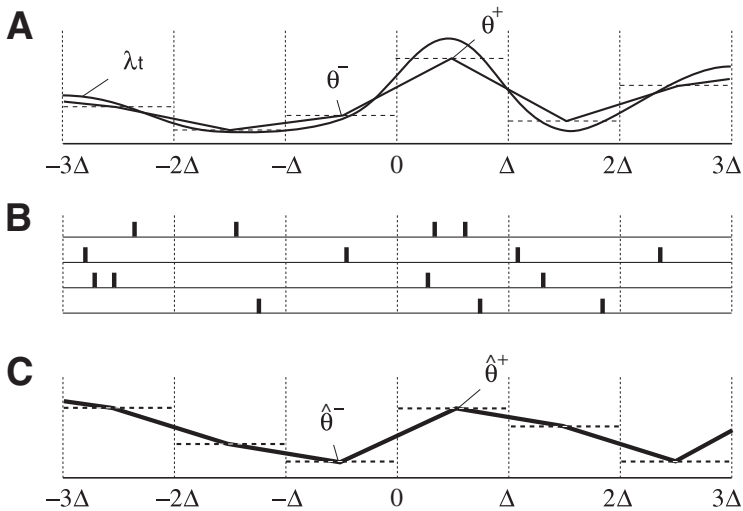

Figure 7: Line-PSTH. (A) An underlying spike rate, λ_t. The horizontal bars indicate the original bar graphs θ^- and θ^+ for adjacent intervals $[-\Delta, 0]$ and $[0, \Delta]$. The expected line graph L_t is a line connecting the two top centers of adjacent expected bar graphs $(-\frac{\Delta}{2}, \theta^-)$ and $(\frac{\Delta}{2}, \theta^+)$. (B) Sequences of spikes derived from the underlying rate. (C) A time histogram for the sample sequences of spikes. The empirical line graph $\hat{L}_t$ is a line connecting the two top centers of adjacent empirical bar graphs $(-\frac{\Delta}{2}, \hat{\theta}^-)$ and $(\frac{\Delta}{2}, \hat{\theta}^+)$.

The unbiased estimators of the original bar graphs are $\hat{\theta}^- = k^-/(n\Delta)$ and $\hat{\theta}^+ = k^+/(n\Delta)$, where k^- and k^+ are the numbers of spikes from n spike sequences that enter the intervals $[-\Delta, 0]$ and $[0, \Delta]$, respectively. The empirical line graph $\hat{L}_t$ in an interval of $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$ is a line connecting two top centers of adjacent empirical bar graphs,

$$\hat{L}_t = \frac{\hat{\theta}^+ + \hat{\theta}^-}{2} + \frac{\hat{\theta}^+ - \hat{\theta}^-}{\Delta} t. \tag{A.2}$$

A.1 Selection of the Bin Size. The time average of the MISE (see equation 2.1) is rewritten by the average over the N segmented rates,

$$\text{MISE} = \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} \left\langle E\left(\hat{L}_t - \lambda_t\right)^2 \right\rangle dt. \tag{A.3}$$

The MISE can be decomposed into two parts:

$$\mathrm{MISE} = \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \left\langle E(\hat{L}_t - L_t)^2 \right\rangle dt + \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \left\langle (\lambda_t - L_t)^2 \right\rangle dt. \tag{A.4}$$

The first term of equation A.4 is the stochastic fluctuation of the empirical linear estimator $\hat{L}_t$ due to the stochastic point events, which can be computed as

$$\frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \left\langle E(\hat{L}_t - L_t)^2 \right\rangle dt = \frac{2}{3}\left\langle E(\hat{\theta}^+ - \theta^+)^2 \right\rangle. \tag{A.5}$$

Here we have used the relations, $E(\hat{\theta}^+ - \theta^+)(\hat{\theta}^- - \theta^-) = 0$ and $\left\langle E(\hat{\theta}^+ - \theta^+)^2 \right\rangle = \left\langle E(\hat{\theta}^- - \theta^-)^2 \right\rangle$.

The second term of equation A.4 is the temporal fluctuation of λ_t around the expected linear estimator L_t. We expand the second term by inserting $\mu = \langle \theta^- \rangle = \langle \theta^+ \rangle$ and obtain

$$\begin{aligned}
\frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \left\langle (\lambda_t - L_t)^2 \right\rangle dt &= \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \left\langle (\lambda_t - \mu)^2 \right\rangle dt \\
&\quad - \frac{2}{\Delta}\int_{-\Delta/2}^{\Delta/2} \langle (\lambda_t - \mu)(L_t - \mu) \rangle dt \\
&\quad + \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \left\langle (L_t - \mu)^2 \right\rangle dt \\
&= \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \left\langle (\lambda_t - \mu)^2 \right\rangle dt \\
&\quad - \left\{ 2\left\langle (\theta^+ - \mu)\,(\theta^0 - \mu) \right\rangle + 2\left\langle (\theta^+ - \mu)\,\theta^* \right\rangle \right\} \\
&\quad + \left\{ \frac{2}{3}\left\langle (\theta^+ - \mu)^2 \right\rangle + \frac{1}{3}\left\langle (\theta^+ - \mu)(\theta^- - \mu) \right\rangle \right\},
\end{aligned} \tag{A.6}$$

where

$$\theta^0 \equiv \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \lambda_t \, dt,$$

$$\theta^* \equiv \frac{2}{\Delta^2}\int_{-\Delta/2}^{\Delta/2} t\lambda_t \, dt.$$

To obtain the above equation, we have used relations, $\langle\theta^-\rangle = \langle\theta^+\rangle = \langle\theta^0\rangle = \mu$, $\langle\theta^*\rangle = 0$, $\langle(\theta^+ - \mu)^2\rangle = \langle(\theta^- - \mu)^2\rangle$, $\langle(\theta^+ - \mu)(\theta^0 - \mu)\rangle = \langle(\theta^- - \mu)(\theta^0 - \mu)\rangle$, and $\langle(\theta^+ - \mu)\theta^*\rangle = -\langle(\theta^- - \mu)\theta^*\rangle$.

The first term of equation A.6 represents a mean squared fluctuation of the underlying rate and is independent of the choice of the bin size Δ (see equation 2.8). We introduce the cost function by subtracting the variance of the underlying rate from the original MISE:

$$\begin{aligned} C_n(\Delta) &\equiv \mathrm{MISE} - \frac{1}{T}\int_0^T (\lambda_t - \mu)^2\, dt \\ &= \frac{2}{3}\left\langle E(\hat{\theta}^+ - \theta^+)^2\right\rangle - 2\left\langle(\theta^+ - \mu)(\theta^0 - \mu)\right\rangle - 2\left\langle(\theta^+ - \mu)\theta^*\right\rangle \\ &\quad + \frac{2}{3}\left\langle E(\theta^+ - \mu)^2\right\rangle + \frac{1}{3}\left\langle(\theta^+ - \mu)(\theta^- - \mu)\right\rangle. \end{aligned} \tag{A.7}$$

The first term of equation A.7 can be estimated from the data, using the variance-mean relation, equation 2.12. The last four terms of equation A.7 are the covariances of the expected rates $\theta^-, \theta^+, \theta^0$, and θ^*, which are not observables. We can estimate them by using the covariance decomposition rule for unbiased estimators:

$$\begin{aligned} \langle(\theta^+ - \langle\theta^+\rangle)(\theta^p - \langle\theta^p\rangle)\rangle &= \langle E[(\hat{\theta}^+ - \langle E\hat{\theta}^+\rangle)(\hat{\theta}^p - \langle E\hat{\theta}^p\rangle)]\rangle \\ &\quad - \langle E[(\hat{\theta}^+ - \theta^+)(\hat{\theta}^p - \theta^p)]\rangle, \end{aligned} \tag{A.8}$$

where p denotes $-, +, 0$, or $*$. The first term of equation A.8 can be computed directly from the data, using the relation $\langle E\hat{\theta}^-\rangle = \langle E\hat{\theta}^+\rangle = \langle E\hat{\theta}^0\rangle = \mu$, and $\langle E\hat{\theta}^*\rangle = 0$. Unlike the Bar-PSTH, however, the mean-variance relation for the Poisson statistics is not directly applicable to the second term. We suggest estimating these covariance terms from multiple sample sequences, as in algorithm 3:

Algorithm 3: A Method for Bin Size Selection for a Line-PSTH

i. Divide the observation period T into $N+1$ bins of width Δ.
Count the number of spikes $k_i(j)$ that enter ith bin from jth sequence.
Define $k_i^{(-)}(j) \equiv k_i(j)$ and $k_i^{(+)}(j) \equiv k_{i+1}(j)$ $(i = 1, 2, \ldots, N)$.
Divide the period $[\Delta/2, T - \Delta/2]$ into N bins of width Δ.
Count the number of spikes $k_i^{(0)}(j)$ that enter ith bin from jth sequence. In each bin, compute $k_i^{(*)}(j) \equiv 2\sum_\ell t_i^\ell(j)/\Delta$, where $t_i^\ell(j)$ is the time of the ℓth spike that enters the ith bin, measured from the center of each bin.

ii. Sum up $k_i^{(p)}(j)$ for all sequences: $k_i^{(p)} \equiv \sum_{j=1}^n k_i^{(p)}(j)$, where $p = \{-, +, 0, *\}$.

Average those spike counts with respect to all the bins: $\bar{k}^{(p)} \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{(p)}$.
Compute covariances of $k_i^{(+)}$ and $k_i^{(p)}$:

$$c^{(+,\,p)} \equiv \frac{1}{N}\sum_{i=1}^{N}\left(k_i^{(+)} - \bar{k}^{(+)}\right)\left(k_i^{(p)} - \bar{k}^{(p)}\right).$$

Compute covariances of $k_i^{(+)}(j)$ and $k_i^{(p)}(j)$ with respect to sequences and average over time:

$$\bar{c}^{(+,\,p)} \equiv \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n-1}\sum_{j=1}^{n}\left(k_i^{(+)}(j) - \frac{k_i^{(+)}}{n}\right)\left(k_i^{(p)}(j) - \frac{k_i^{(p)}}{n}\right).$$

Finally, compute

$$\sigma^{(+,\,p)} \equiv \frac{c^{(+,\,p)}}{(n\Delta)^2} - \frac{\bar{c}^{(+,\,p)}}{n\Delta^2}.$$

iii. Compute the cost function:

$$C_n(\Delta) = \frac{2}{3}\frac{\bar{k}^{(+)}}{(n\Delta)^2} - 2\sigma^{(+,\,0)} - 2\sigma^{(+,\,*)} + \frac{2}{3}\sigma^{(+,\,+)} + \frac{1}{3}\sigma^{(+,\,-)}.$$

iv. Repeat i through iii while changing Δ to search for Δ^* that minimizes $C_n(\Delta)$.

A.2 Extrapolation of the Cost Function. As in the case of the Bar-PSTH, the cost function for any m sequences of spike trains can be extrapolated using the variance-mean relation for the Poisson statistics,

$$C_m\left(\Delta|n\right) = \frac{2}{3\Delta}\left(\frac{1}{m} - \frac{1}{n}\right)\left\langle E\hat{\theta}_n^{+}\right\rangle + C_n\left(\Delta\right), \tag{A.9}$$

where $C_n\left(\Delta\right)$ is the original cost function (see equation A.7). With the original cost function $C_n\left(\Delta\right)$, we can easily estimate a cost function for m sequences with algorithm 4:

Algorithm 4: A Method for Extrapolating the Cost Function for a Line-PSTH

A. Construct the extrapolated cost function,

$$C_m(\Delta|n) = \frac{2}{3}\left(\frac{1}{m} - \frac{1}{n}\right)\frac{\bar{k}^{(+)}}{n\Delta^2} + C_n(\Delta),$$

where $C_n(\Delta)$ is the cost function for the line-graph time histogram computed for n sequences of spikes with the algorithm 3.

B. Search for Δ_m^* that minimizes $C_m(\Delta|n)$.

C. Repeat A and B while changing m, and plot $1/\Delta_m^*$ versus $1/m$ to search for the critical value $1/m = 1/\hat{n}_c$ above which $1/\Delta_m^*$ practically vanishes.

A.3 Scaling Relations of the Optimal Bin Sizes. We can obtain a theoretical cost function of a Line-PSTH directly from the mean spike rate μ and the correlation $\phi(t)$ of the rate fluctuation. According to the mean-variance relation based on the Cramér-Rao (in)equality (see equation 2.21), the cost function, equation A.7, is given by

$$\begin{aligned} C_n(\Delta) = &\frac{2\mu}{3n\Delta} - \frac{2}{\Delta^2}\int_0^{\Delta}\int_{-\Delta/2}^{\Delta/2}\left(1 + \frac{2t_2}{\Delta}\right)\phi(t_1 - t_2)\, dt_1 dt_2 \\ &+ \frac{2}{3\Delta^2}\int_0^{\Delta}\int_0^{\Delta}\phi(t_1 - t_2)\, dt_1 dt_2 + \frac{1}{3\Delta^2}\int_0^{\Delta}\int_{-\Delta}^{0}\phi(t_1 - t_2)\, dt_1 dt_2. \end{aligned} \tag{A.10}$$

By expanding the correlation of the rate fluctuation,

$$\phi(t) = \phi(0) + \phi'(0_+)|t| + \frac{1}{2}\phi''(0)t^2 + \frac{1}{6}\phi'''(0_+)|t|^3 + \frac{1}{24}\phi''''(0)t^4 + O\left(|t|^5\right),$$

we obtain

$$\begin{aligned} C_n(\Delta) = &\frac{2\mu}{3n\Delta} - \phi(0) - \frac{37}{144}\phi'(0_+)\Delta + \frac{181}{5760}\phi'''(0_+)\Delta^3 \\ &+ \frac{49}{2880}\phi''''(0)\Delta^4 + O(\Delta^5). \end{aligned} \tag{A.11}$$

Unlike the Bar-PSTH, the line graph successfully approximates the original rate to first order in Δ, and therefore the $O(\Delta^2)$ term in the cost function vanishes for a Line-PSTH.

The optimal bin size Δ^* is obtained from $dC_n(\Delta)/d\Delta = 0$. For a rate process that fluctuates smoothly in time, the correlation function is a smooth function, resulting in $\phi'(0) = 0$ and $\phi'''(0) = 0$ due to the symmetry $\phi(t) = \phi(-t)$, and we obtain the scaling relation

$$\Delta^* \sim \left(\frac{1280\mu}{181\phi''''(0)\, n} \right)^{1/5}. \tag{A.12}$$

The exponent $-1/5$ for a Line-PSTH is different from the exponent $-1/3$ for a Bar-PSTH (see equation 2.28).

If the correlation of rate fluctuation has a cusp at $t = 0$ ($\phi'(0_+) < 0$), we obtain the scaling relation,

$$\Delta^* \sim \left(-\frac{96\mu}{37\phi'(0_+)n} \right)^{1/2}, \tag{A.13}$$

by ignoring the higher-order terms. The exponent $-1/2$ for a Line-PSTH is the same as the exponent for a Bar-PSTH (see equation 2.29).

Acknowledgments

We thank Shinsuke Koyama, Takeaki Shimokawa, and Kensuke Arai for helpful discussions. We also acknowledge K. H. Britten, M. N. Shadlen, W. T. Newsome, and J. A. Movshon, who have made their data available to the public. This study is supported in part by Grants-in-Aid for Scientific Research to S.S. from the Ministry of Education, Culture, Sports, Science and Technology of Japan (16300068, 18020015) and the 21st Century COE Center for Diversity and Universality in Physics. H.S. is supported by the Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists.

References

Abeles, M. (1982). Quantification, smoothing, and confidence-limits for single-units histograms. *Journal of Neuroscience Methods, 5*(4), 317–325.

Adrian, E. (1928). *The basis of sensation: The action of the sense organs*. New York: Norton.

Bair, W., & Koch, C. (1996). Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Computation, 8*(6), 1185–1202.

Blahut, R. (1987). *Principles and practice of information theory*. Reading, MA: Addison-Wesley.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual-motion: A comparison of neuronal and psychophysical performance. *Journal of Neuroscience, 12*(12), 4745–4765.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (2004). Responses of single neurons in macaque MT/V5 as a function of motion coherence in stochastic dot stimuli. *Neural Signal Archive*, nsa2004.1. Available online at http://www.neuralsignal.org.

Brockwell, A. E., Rojas, A. L., & Kass, R. E. (2004). Recursive Bayesian decoding of motor cortical signals by particle filtering. *Journal of Neurophysiology, 91*(4), 1899–1907.

Brown, E. N., Kass, R. E., & Mitra, P. P. (2004). Multiple neural spike train data analysis: State-of-the-art and future challenges. *Nature Neuroscience, 7*(5), 456–461.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.

Daley, D., & Vere-Jones, D. (1988). *An introduction to the theory of point processes*. New York: Springer-Verlag.

Dayan, P., & Abbott, L. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

DiMatteo, I., Genovese, C. R., & Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika, 88*(4), 1055–1071.

Gardiner, C. (1985). *Handbook of stochastic methods: For physics, chemistry and the naural sciences*. Berlin: Springer-Verlag.

Gerstein, G. L., & Kiang, N. Y. S. (1960). An approach to the quantitative analysis of electrophysiological data from single neurons. *Biophysical Journal, 1*(1), 15–28.

Gerstein, G. L., & Perkel, D. H. (1969). Simultaneously recorded trains of action potentials—analysis and functional interpretation. *Science, 164*(3881), 828–830.

Hall, P. (1990). Akaikes information criterion and Kullback-Leibler loss for histogram density-estimation. *Probability Theory and Related Fields, 85*(4), 449–467.

Johnson, D. H. (1996). Point process models of single-neuron discharges. *Journal of Computational Neuroscience, 3*(4), 275–299.

Kanazawa, Y. (1993). Hellinger distance and Akaike information criterion for the histogram. *Statistics and Probability Letters, 17*(4), 293–298.

Kass, R. E., Ventura, V., & Brown, E. N. (2005). Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology, 94*(1), 8–25.

Kass, R. E., Ventura, V., & Cai, C. (2003). Statistical smoothing of neuronal data. *Network-Computation in Neural Systems, 14*(1), 5–15.

Koyama, S., & Shinomoto, S. (2004). Histogram bin width selection for time-dependent Poisson processes. *Journal of Physics A—Mathematical and General, 37*(29), 7255–7265.

Kubo, R., Toda, M., & Hashitsume, N. (1985). *Nonequilibrium statistical mechanics*. Berlin: Springer-Verlag.

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature, 341*(6237), 52–54.

Révész, P. (1968). *The laws of large numbers*. New York: Academic Press.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics, 9*(2), 65–78.

Sanger, T. D. (2002). Decoding neural spike trains: Calculating the probability that a spike train and an external signal are related. *Journal of Neurophysiology, 87*(3), 1659–1663.

Scott, D. W. (1979). Optimal and data-based histograms. *Biometrika, 66*(3), 605–610.

Shinomoto, S., Miyazaki, Y., Tamura, H., & Fujita, I. (2005). Regional and laminar differences in in vivo firing patterns of primate cortical neurons. *Journal of Neurophysiology, 94*(1), 567–575.

Shinomoto, S., Shima, K., & Tanji, J. (2003). Differences in spiking patterns among cortical neurons. *Neural Computation, 15*(12), 2823–2842.

Snyder, D. (1975). *Random point processes*. New York: Wiley.

Wiener, M. C., & Richmond, B. J. (2002). Model based decoding of spike trains. *Biosystems, 67*(1–3), 295–300.

Received November 17, 2005; accepted September 27, 2006.