ARTICLE Communicated by Carlos Brody

Sampling Properties of the Spectrum and Coherency of Sequences of Action Potentials

M. R. Jarvis
Division of Biology, California Institute of Technology, Pasadena, California 91125, U.S.A.

P. P. Mitra
Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, U.S.A.

The spectrum and coherency are useful quantities for characterizing the temporal correlations and functional relations within and between point processes. This article begins with a review of these quantities, their interpretation, and how they may be estimated. A discussion of how to assess the statistical significance of features in these measures is included. In addition, new work is presented that builds on the framework established in the review section. This work investigates how the estimates and their error bars are modified by finite sample sizes. Finite sample corrections are derived based on a doubly stochastic inhomogeneous Poisson process model in which the rate functions are drawn from a low-variance gaussian process. It is found that in contrast to continuous processes, the variance of the estimators cannot be reduced by smoothing beyond a scale set by the number of point events in the interval. Alternatively, the degrees of freedom of the estimators can be thought of as bounded from above by the expected number of point events in the interval. Further new work describing and illustrating a method for detecting the presence of a line in a point process spectrum is also presented, corresponding to the detection of a periodic modulation of the underlying rate. This work demonstrates that a known statistical test, applicable to continuous processes, applies with little modification to point process spectra and is of utility in studying a point process driven by a continuous stimulus. Although the material discussed is of general applicability to point processes, attention will be confined to sequences of neuronal action potentials (spike trains), the motivation for this work.

1 Introduction

The study of spike trains is of central importance to electrophysiology. Often changes in the mean firing rate are studied, but there is increasing interest in characterizing the temporal structure of spike trains and the relationships

between spike trains more completely (Gray, König, Engel, & Singer, 1989; Gerstein, Perkel, & Dayhoff, 1985; Abeles, Deribaupierre, & Deribaupierre, 1983). A natural extension to estimating the rate of neuronal firing is to estimate the autocorrelation and the cross-correlation functions. Definitions of these quantities are given in section 2.3. This article discusses the frequency domain counterparts of these quantities. Auto- and cross-correlations correspond to spectra and cross-spectra, respectively. The coherency, which is the normalized cross-spectrum, does not in general have a simple time-domain counterpart.

The frequency domain has several advantages over the time domain. First, often subtle structure can be detected with the frequency domain estimators that is difficult to observe with the time domain estimators. Second, the time domain quantities have problems associated with the sensitivity of the estimators to weak nonstationarity and the nonlocal nature of the error bars (Brody, 1998). These problems are greatly reduced in the frequency domain. Third, reasonably accurate confidence intervals may be placed on estimates of the second-order properties in the frequency domain, which permits the statistical significance of features to be assessed. Fourth, the coherency provides a normalized measure of correlations between time series, in contrast with time-domain cross-correlations that are not normalizable by any simple means.

This article begins by reviewing the population spectrum and coherency for point processes and motivating their use by describing some example applications. Next, direct, lag window, and multitaper estimators of the spectrum and coherency are presented. The concept of degrees of freedom is introduced and used to obtain large sample error bars for the estimators. Many elements of the work discussed in the review section of this article can be found in the references (Percival & Walden, 1993; Cox & Lewis, 1966; Brillinger, 1978; Bartlett, 1966). Most of the material in these references is targeted at either spectral analysis of continuous processes or at the analysis of point processes but with less emphasis on spectral analysis. Building on this framework, corrections, based on a specific model, will be given for finite sample sizes. These corrections are cast in terms of a reduction in the degrees of freedom of the estimators. For a homogeneous Poisson process, the modified degrees of freedom is the harmonic sum of the asymptotic degrees of freedom and twice the number of spikes used to construct the estimate. Modifications to this basic result are given for structured spectra and tapered data. A section is included on the treatment of point process spectra that contain lines. A statistical test for the presence of a line in a background of colored noise is given, and the method for removal of such a line is described. An example application to periodic stimulation is given. Sections 2 through 5 contain the review portion of this article; sections 6 through 12 decribe new results and illustrative applications.

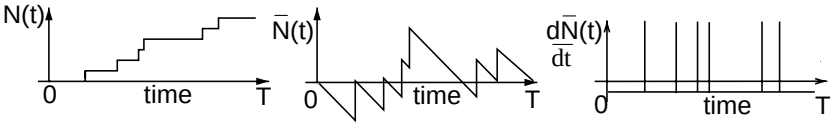

Figure 1: Example illustrating how the processes N, $\overline{N}$, and $d\overline{N}/dt$ relate to each other. The vertical lines in the process $d\overline{N}/dt$ depict delta functions.

2 Population Measures and Their Interpretation

2.1 Counting Representation of a Spike Train. A spike train may be regarded as a point process. If the spike shapes are neglected, it is completely specified by a series of spike times $\{t_i\}$ and the start and end points of the recording interval $[0, T]$. It is convenient to introduce some notation that enables the subsequent formulas to be written in a compact form (Brillinger, 1978). The counting process $N(t)$ is defined as the number of spikes that occur between the start of the interval ($t = 0$) and time t. The counting process has the property that the area beneath it grows as t becomes larger. This is undesirable because it leads to an interval-dependent peak at low frequencies in the spectrum. To avoid this, a process $\overline{N}(t) = N(t) - \lambda t$, where λ is the mean rate, which has zero mean, may be constructed. Note that $d\overline{N}(t) = \overline{N}(t+dt) - \overline{N}(t)$, which is either $1 - \lambda dt$ or $-\lambda dt$ depending on whether there is a spike in the interval dt. Thus $d\overline{N}(t)/dt$ is a series of delta functions with the mean rate subtracted.[1] Figure 1 illustrates the relationship between $N(t)$, $\overline{N}(t)$, and $d\overline{N}(t)/dt$.

2.2 Stationarity. It will be assumed in what follows that the spike trains are second-order stationary. This means that their first and second moments do not depend on the absolute time. In many electrophysiology experiments, this is not the case. In awake behaving studies, the animal is often trained to perform a highly structured task. Nevertheless, it may still be the case that over an appropriately chosen short time window, the statistical properties are changing slowly enough for reasonable estimates of the spectrum and coherency to be obtained. As an example, neurons in primate parietal area PRR exhibit what is known as memory activity during a delayed reach task (Snyder, Batista, & Andersen, 1997). The mean firing rate of these neurons varies considerably during the task but during the memory period is roughly constant. The assumption of stationarity during the memory period is equivalent to the intuitive notion that there is nothing special about 0.75 seconds into the memory period as compared to, say,

[1] A delta function is a generalized function. It has an area of one beneath it but has zero width and therefore infinite height.

0.5 seconds. Second-order stationarity implies that the mean firing rate (λ) is independent of time and that the autocovariance depends on only the lag (τ) and not the absolute time.

Nonstationary spectral methods have been developed that can be used to test for stationarity (Thomson, 2000). Such methods are beyond the scope of this article and may be of limited use when small amounts of point process data are available. Instead we use the following simple rule. If the average rate (peristimulus time histogram) is approximately flat to within the error bars expected based on a Poisson process, then we consider the process to be approximately stationary. Spectral analysis can be applied even if this criterion is not met. In this case, however, confidence intervals should be evaluated using a jackknife method, as this is insensitive to the stationarity assumption and care should be taken when interpreting structure, particularly at low frequencies.

2.3 Definitions. Equations 2.1 to 2.4 give the first- and second-order moments of a single spike train for a stationary process. The spectrum $S(f)$ is the Fourier transform of the autocovariance function ($\mu(\tau) + \lambda\delta(\tau)$),

$$\frac{E\{dN(t)\}}{dt} = \lambda \tag{2.1}$$

$$\frac{E\{d\overline{N}(t)\}}{dt} = 0 \tag{2.2}$$

$$\mu(\tau) + \lambda\delta(\tau) = \frac{E[d\overline{N}(t)d\overline{N}(t+\tau)]}{dtd\tau} \tag{2.3}$$

$$S(f) = \lambda + \int_{-\infty}^{\infty} \mu(\tau)\exp(-2\pi if\tau)d\tau, \tag{2.4}$$

where E denotes the expectation operator.

The autocovariance measures how likely it is that a spike will occur at time $t + \tau$ given that one has occurred at time t. Usually $\mu(\tau)$ is estimated rather than the full autocovariance, which includes a delta function at zero lag.[2] However, in order to take the Fourier transform, the full autocovariance is required. The inclusion of this delta function leads to a constant offset of the spectrum. This offset is an important difference between continuous time processes and point processes. The population coherency $\gamma(f)$ is defined in equations 2.5 to 2.7:

$$\mu_{ab}(\tau) = \frac{E[d\overline{N}_a(t)d\overline{N}_b(t+\tau)]}{dtd\tau} \tag{2.5}$$

[2] When estimating the autocovariance using a histogram method, one generally omits the spike at the start of the interval that would always fall in the bin nearest zero.

$$S_{ab}(f) = \int_{-\infty}^{\infty} \mu_{ab}(\tau) \exp(-2\pi i f \tau) d\tau + \lambda_a \delta_{ab} \tag{2.6}$$

$$\gamma(f) = \frac{S_{12}(f)}{\sqrt{S_{11}(f) S_{22}(f)}}, \tag{2.7}$$

where indices 1 and 2 denote simultaneously recorded spike trains from different cells.

Unlike the spectrum, which is strictly real and positive, the coherency is a complex quantity. The modulus of the coherency, which is known as the coherence, can vary only between zero and one.[3] This makes coherence particularly attractive for detecting relationships between spike trains because it is insensitive to the mean spike rates.

3 Examples and Their Interpretation

Before discussing the details regarding how to estimate the spectrum and coherency, it will be helpful to consider some simple examples.

3.1 Example Population Spectra. For a homogeneous Poisson process of constant rate λ, the autocovariance is simply $\lambda\delta(\tau)$, and hence the spectrum is a constant equal to the rate λ. At the opposite extreme, consider the case where the spikes are spaced by intervals $\Delta\tau$. This is not a stationary process, but if a small amount of drift is permitted, so that over an extended period there is nothing special about a given time, it becomes stationary. The spectrum of this process contains sharp lines at integer multiples of $f = \frac{1}{\Delta\tau}$. Due to the drift, the higher harmonics will become increasingly blurred, and in the high-frequency limit, the spectrum will tend toward a constant value of the mean rate λ. As a final example, consider the case where $\mu(\tau)$ is a negative gaussian centered on zero τ. This form of $\mu(\tau)$ is consistent with the probability of firing being suppressed after firing.[4] The spectrum of this process will be below λ at low frequencies and will go to a constant value λ at high frequencies. Figure 2 illustrates these different population spectra.

3.2 Example Population Coherency. The population coherency of two homogeneous Poisson processes is zero. In contrast, if two spike trains are equal, then the coherence is one and the phase of the coherency is zero at all frequencies. If two spike trains are identical but offset by a lag $\Delta\tau$, then the coherence will again be one, but the phase of the coherency will vary linearly with frequency with a slope proportional to $\Delta\tau$ and given by $\phi(f) = 2\pi f \Delta\tau$.

[3] Some authors define coherence as the modulus squared of the coherency.

[4] This need not necessarily correspond to the biophysical refractive period; it could arise instead from a characteristic integration time.

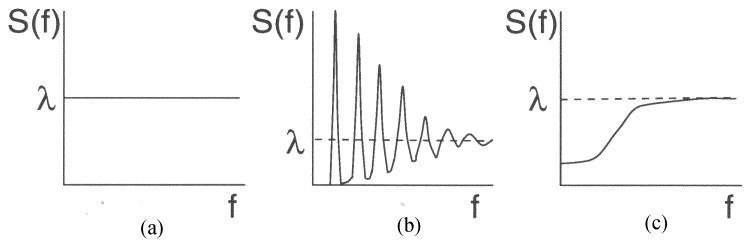

Figure 2: Example population spectra for different types of underlying process. (a) Homogeneous Poisson process with rate λ. (b) Regularly spaced spikes with jitter. (c) Spike trains in which the probability of firing is suppressed immediately after firing.

4 Estimating the Spectrum

Section 3 demonstrated that the population spectrum may provide insights into the nature of a spike train. In this section, the question of how to estimate the spectrum from a finite section of data will be introduced. In what follows, the population quantity λ in the definition of $\overline{N}(t)$ is replaced by a sample estimate $N(T)/T$.

4.1 Direct Spectral Estimators.

4.1.1 Definition. A popular, though seriously flawed, method for estimating the spectrum is to take the modulus squared of the Fourier transform of the data $d\overline{N}(t)$. This estimate is known as the periodogram and is the simplest example of a direct spectral estimator. More generally, a direct spectral estimator is the modulus squared of the Fourier transform of the data, but with the data being multiplied by an envelope function $h(t)$, known as a taper (Percival & Walden, 1993). Equations 4.1 through 4.3 define the direct estimator. On substituting $\overline{N}(t)$ into equation 4.2, a form amenable to implementation on a computer is obtained (see equation 4.4). In this form, the Fourier transform may be computed rapidly and without the need for the binning of data. Note that equation 4.3 results in $h(t)$ scaling as $1/\sqrt{T}$ as the sample length is altered. This ensures proper normalization of the Fourier transformation as sample size varies:

$$I^D(f) = |J^D(f)|^2 \tag{4.1}$$

$$J^D(f) = \int_0^T h(t)e^{-2\pi ift}d\overline{N}(t) \tag{4.2}$$

where,

$$\int_0^T h(t)^2 dt = 1 \tag{4.3}$$

$$J^D(f) = \sum_{j=1}^{N(T)} h(t_j) e^{-2\pi i f t_j} - \frac{N(T)H(f)}{T}, \tag{4.4}$$

and $H(f)$ is the Fourier transform of the taper.

The direct estimator suffers from bias and variance problems, described below, and is of no practical relevance for a single spike train sample.

4.1.2 Bias. It may not be immediately apparent why the above procedure is an estimate of the spectrum, especially when one is permitted to multiply the data by an arbitrary, albeit normalized, taper. The relation between $I^D(f)$ and the spectrum may be obtained by taking the expectation of equation 4.1.

$$E\{I^D(f)\} = E\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t)h(t')e^{-2\pi i f(t-t')} d\overline{N}(t) d\overline{N}(t')\right\}. \tag{4.5}$$

Assuming that the integration and expectation operations may be interchanged and substituting equation 2.3 yields[5]

$$E\{I^D(f)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t)h(t')e^{-2\pi i f(t-t')}\{\mu(t-t') + \lambda\delta(t-t')\}dtdt', \tag{4.6}$$

which may be rewritten in the Fourier domain as

$$E\{I^D(f)\} = \int_{-\infty}^{\infty} S(f')|H(f-f')|^2 df'. \tag{4.7}$$

The expected value of the direct estimator is a convolution of the true spectrum and the modulus squared of the Fourier transform of the taper. The normalization condition on the taper (see equation 4.3) is equivalent to the requirement that the kernel of the convolution has unit area underneath it. Sharp features in the true spectrum will be thus be smeared by an amount that depends on the width of the taper in the frequency domain. If the taper is well localized in the frequency domain, the expected value of the direct estimate is close to the true spectrum, but if the taper is poorly localized, then the expected value of the direct estimator will be incorrect, that is, the direct spectral estimator is biased. There is a fundamental level beyond which the bias cannot be reduced due to the uncertainty relation forbidding simultaneous localization of a function in the time and frequency domains below a given limit. Since the maximum width of the taper is T, the minimum

[5] For the moment, we assume that the population quantity λ is known. This is, of course, not the case in practice, and one employs the estimate $N(T)/T$ as stated before. The effect of this extra uncertainty is given in equation 4.8.

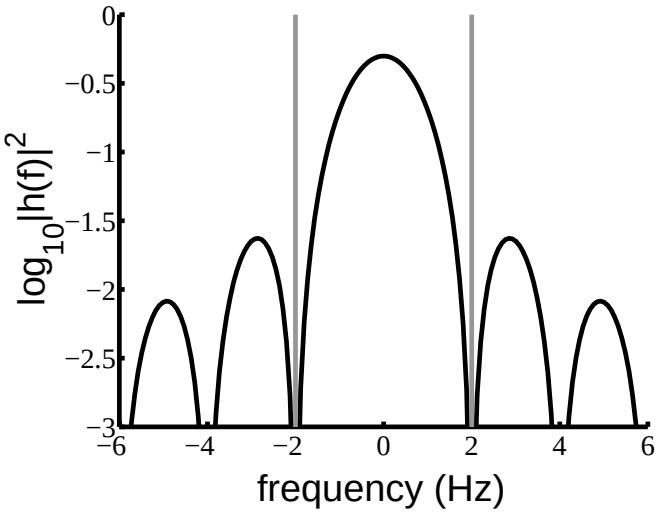

Figure 3: Smoothing kernel $|H(f)|^2$. This is the expected direct estimate of the spectrum in the case of a population spectrum that has a delta function (very sharp feature) at the center frequency. A rectangular taper of length 0.5 second was used. Solid vertical lines are drawn at $\pm$ the Raleigh frequency.

frequency spread is $1/T$, which is known as the Raleigh frequency. Figure 3 shows the smoothing kernel for a rectangular taper and a T of 0.5 second. Note that this kernel has large side lobes, which is the primary motivation for using tapering.

In the above argument, equation 2.3 was used in spite of the appearance of the population quantity λ rather than the sample estimate $N(T)/T$ for which equation 4.5 was defined. A more careful treatment, which includes this correction, leads to an additional term at finite sample sizes in the expectation of the direct spectral estimator at low frequencies. The full expression is

$$E\{I^D(f)\} = \int_{-\infty}^{\infty} S(f')|H(f-f')|^2 df' - |H(f)|^2 S(0)/T. \tag{4.8}$$

In the case of the periodogram, where $h(t) = 1/\sqrt{T}$, the effect is clear since in this case, $J^D(0) = 0$ and hence $I^D(0) = 0$ for any set of spike times and any T.

4.1.3 Asymptotic Variance. In the previous section it was shown that provided the taper is sufficiently local in frequency, the expected value of the direct spectral estimator will be close to the true spectrum. However, the fact that the estimate is on average close to the true spectrum belies a serious problem with direct spectral estimators: the estimates have very large

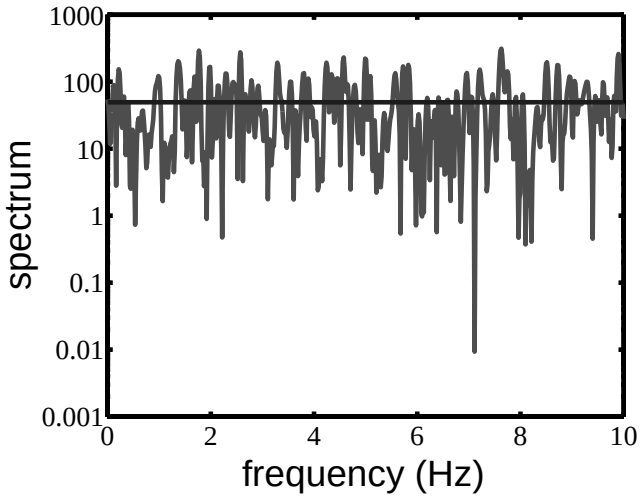

Figure 4: An example of a direct spectral estimate. A 40% cosine taper was used. A sample of duration 20 seconds was drawn from a homogeneous Poisson process with a constant rate of 50 Hz. The population spectrum for this process, shown by the solid horizontal line, is flat. The direct spectral estimate is clearly noisy, although on average the correct spectrum is obtained.

fluctuations about this mean. The underlying source of this problem is that one is attempting to estimate the value of a function at an infinite number of points using a finite sample of data. The problem manifests itself in the fact that direct spectral estimators are inconsistent estimators of the spectrum.[6] In fact it may be shown that under fairly general assumptions, the estimates are distributed exponentially (or equivalently as $S(f)\chi_2^2/2$) for asymptotic sample sizes (i.e., $T \to \infty$) (Brillinger, 1972). Figure 4 illustrates that direct spectral estimators are noisy and untrustworthy, a fact emphasized by the observation that the χ_2^2 distribution has a standard deviation equal to its mean. In the next three sections, methods for reducing the variance of direct spectral estimators using different forms of averaging will be discussed.

4.2 Trial Averaging. If a number of trials (N_T) is available, then the variance of the direct estimator may be reduced by trial averaging,

$$I^{DT}(f) = \frac{1}{N_T} \sum_{n=1}^{N_T} I_n^D(f), \tag{4.9}$$

where $I_n^D(f)$ is the direct spectral estimate based on the nth trial.

[6] Inconsistent estimators have a finite variance even for an infinite length sample.

In the large T limit, taking the average entails summing N_T independent samples from a χ^2_2 distribution, the result of which is distributed as $\chi^2_{2N_T}$. The reduction in variance is inversely proportional to the number of trials corresponding to a reduction in standard deviation, which is the familiar factor of $1/\sqrt{N_T}$.

At first sight, it appears one would be getting something for nothing by breaking a single section of data into N_T segments and treating them as separate trials. This is not the case. The reason is that if the data are segmented into short length samples, there is loss of frequency resolution proportional to the inverse of segment length. Lag window and multitaper estimators use the information from these independent estimates without artificially segmenting the data.

4.3 Lag Window Estimates. A powerful property of the frequency domain is that unless two frequencies are very close together, direct estimates of the spectrum of a stationary process at different frequencies are nearly uncorrelated. This property arises when the covariance between frequencies falls off rapidly. If the true spectrum varies slowly over the width of the covariance, then the large sample covariance of a direct spectral estimator is

$$\mathrm{cov}\{I^D(f_1), I^D(f_2)\} \simeq E\{I^D(\overline{f})\}^2 \left| \int_{-\infty}^{\infty} h(t)^2 e^{-2\pi i \Delta f t} dt \right|^2 \tag{4.10}$$

where $\overline{f} = (f_1 + f_2)/2$ and $\Delta f = f_1 - f_2$.

For $\Delta f = 0$, this expression reduces to the previously mentioned result that the variance of the estimator is equal to the square of the mean. For $\Delta f \gg 1/T$, $|\int_{-\infty}^{\infty} h(t)^2 e^{-2\pi i \Delta f t} dt|^2 \to 0$, since $h(t)^2$ is a smooth function with extent T. This implies that $\mathrm{cov}\{I^D(f_1), I^D(f_2)\} \approx 0$ for $|f_1 - f_2| \gg 1/T$. The approximate independence of nearby points means that if the true spectrum varies slowly enough, then closely spaced points will provide several independent estimates of the same underlying spectrum. This is the motivation for the lag window estimator, which is simply a smoothed version of the direct spectral estimator (Percival & Walden, 1993). The lag window estimator is defined as

$$I^{LW}(f) = \int_{-\infty}^{\infty} K(f - f') I^D(f') df' \tag{4.11}$$

where

$$\int_{-\infty}^{\infty} K(f) df = 1. \tag{4.12}$$

Averaging over trials may be included by using the trial-averaged direct spectral estimate I^{DT} (see equation 4.9) in place of I^D in the above expression. It is assumed that $K(f)$ is a smoothing kernel with reasonable properties.

4.3.1 Bias. The additional smoothing of the lag window kernel modifies the bias properties of the estimator from those expressed in equation 4.8. The expected value of the lag window estimator is given by

$$E\{I^{LW}(f)\} = \int_{-\infty}^{\infty} K(f-f')|H(f'-f'')|^2 S(f'')df'df'' - \frac{S(0)}{T}\int_{-\infty}^{\infty} K(f-f')|H(f')|^2 df'. \tag{4.13}$$

4.3.2 Asymptotic Variance. The large sample variance of this estimator is readily obtained using equation 4.10:

$$\text{var}\{I^{LW}(f)\} = \frac{\xi}{N_T} E\{I^{LW}(f)\}^2 \tag{4.14}$$

where

$$\xi = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(f)K(f')|\mathcal{H}(f-f')|^2 df df' \tag{4.15}$$

and

$$\mathcal{H}(f) = \int_{-\infty}^{\infty} h(t)^2 e^{-2\pi i f t} dt. \tag{4.16}$$

Equation 4.14 includes the reduction in variance due to trial averaging. $1/\xi$ can be interpreted as the effective number of independent estimates beneath the smoothing kernel, as demonstrated by the following qualitative argument. If Δf is the frequency width of the smoothing kernel $K(f)$ and δf is the frequency width of the taper $\mathcal{H}(f)$, then since $K(f) \sim 1/\Delta f$, it follows that $\xi \sim 1/(\Delta f)^2 \int_{\Delta f}\int_{\Delta f} |\mathcal{H}(f-f')|^2 df df'$ and hence that $\xi \sim \delta f/\Delta f$.

4.4 Multitaper Estimates. While the lag window estimator is based on the idea that nearby frequencies provide independent estimates, the estimation is not very systematic, since one should be able to decorrelate explicitly nearby frequencies from the knowledge of the correlations introduced by a finite window size. This is achieved in multitaper spectral estimation. The basic idea of multitaper spectral estimation is to average the spectral estimates from several orthogonal tapers. The orthogonality of the tapers ensures that the estimates are uncorrelated for large samples (consider substituting $h_1(t)h_2(t)$ for $h(t)^2$ in equation 4.10). A critical question is the choice of a set of orthogonal tapers. A natural choice are the discrete prolate spheroidal sequences (dpss) or Slepian sequences, which are defined

by the property that they are maximally localized in frequency. The dpss tapers maximize the spectral concentration defined as

$$\lambda = \frac{\int_{-W}^{W} |H(f)|^2 df}{\int_{-\infty}^{\infty} |H(f)|^2 df}, \tag{4.17}$$

where in the time domain $h(t)$ is strictly confined to the interval [0,T].

For given values of W and T, a finite number of tapers have concentrations (λ) close to one and therefore have well-controlled bias. This number, known as the Shannon number, is $2WT$. This sets an upper limit on the number of independent estimates that can be obtained for a given amount of spectral smoothing.

A direct multitaper estimate of the spectrum is defined as

$$I^{MT}(f) = \frac{1}{K} \sum_{k=0}^{K-1} I_k^D(f). \tag{4.18}$$

The eigenspectra I_k^D are direct spectral estimates based on tapering the data with the kth dpss function. As previously, trial averaging can be included by using I^{DT} rather than I^D. More sophisticated estimates involve adaptive (rather than constant) weighting of the data tapers (Percival & Walden, 1993). Multitaper spectral estimation has been recently shown to be useful for analyzing neurobiological time series, both continuous processes (Mitra & Pesaran, 1999) and spike trains (Pesaran, Pezaris, Sahani, Mitra, & Andersen, 2000).

4.4.1 Bias. The bias for the multitaper estimate is given by equation 4.8 but with $|H(\cdot)|^2$ replaced by an average over tapers $\frac{1}{K} \sum_{k=0}^{K-1} |H_k(\cdot)|^2$.

4.4.2 Asymptotic Variance. The asymptotic variance of the multitaper estimator, including trial averaging, is given by

$$\mathrm{var}\{I^{MT}(f)\} = \frac{1}{N_T K} E\{I^{MT}(f)\}^2 \tag{4.19}$$

4.5 Degrees of Freedom. At this point it is useful to introduce the concept of the degrees of freedom (ν_0) of an estimate. The degrees of freedom is twice the number of independent estimates of the spectrum. Degrees of freedom is a useful concept as it permits the expressions for the variance of the different estimators to be written in a common format:

$$\mathrm{var}\{I^X(f)\} = \frac{2E\{I^X(f)\}^2}{\nu_0}, \tag{4.20}$$

where

X	D	DT	LW	MT
ν_0	2	$2N_T$	$2N_T/\xi$	$2N_TK$

Degrees of freedom is also a useful framework in which to cast both finite size corrections and the confidence limits for the spectra and coherence.

The variance of estimators of the spectrum can be estimated using internal methods such as the bootstrap or jackknife (Efron & Tibshirani, 1993; Thomson & Chave, 1991). Jackknife estimates can be constructed over trials or over tapers. If ν_0 is large (> 20), then the theoretical and jackknife variance are in close agreement. If distributional assumptions can be validly made about the point process, theoretical error bars have an important advantage over internal estimates since they enable the understanding of different factors that enter into the variance in order to guide experimental design and data analysis. However, jackknife estimates are less sensitive to failures in distributional assumptions, and this provides them with statistical robustness.

It is conventional to display spectra on a log scale because taking the log of the spectrum stabilizes the variance and leads a distribution that is approximately gaussian.

4.6 Confidence Intervals. The expected values of the estimators and their variance have been discussed for several different spectral estimators, but it is desirable to put confidence intervals on the spectral estimates rather than standard deviations.

As mentioned in section 4.2, the averaging of direct spectral estimates from different trials yields, in the large sample limit, estimates that are distributed as $\chi^2_{2N_T}$. In general for the other estimates, a well-known approximation (Percival & Walden, 1993) is to assume that the estimate is distributed as $\chi^2_{\nu_0}$. Confidence intervals can then by placed on estimates on the basis of this $\chi^2_{\nu_0}$ distribution. The confidence interval applies for the population spectrum $S(f)$ and is obtained from the following argument:

$$P\left[q_1 \le \chi^2_{\nu_0} \le q_2\right] = 1 - 2p, \tag{4.21}$$

where P indicates probability, q_1 is such that $P[\chi^2_{\nu_0} \le q_1] = p$, and q_2 is such that $P[\chi^2_{\nu_0} \ge q_2] = p$. It follows that

$$P\left[q_1 \le \nu_0 I^X(f)/S(f) \le q_2\right] = 1 - 2p. \tag{4.22}$$

Hence an approximate $100\% \times (1 - 2p)$ confidence interval for $S(f)$ is given by

$$P\left[\nu_0 I^X(f)/q_2 \le S(f) \le \nu_0 I^X(f)/q_1\right]. \tag{4.23}$$

For large ν_0 (> 20) these confidence intervals do not differ substantially from those based on a gaussian (± 2 standard deviations) but at small ν_0, the difference can be substantial as for these values the $\chi^2_{\nu_0}$ distribution has long tails.

4.7 High-Frequency Limit. The population spectrum goes to a constant value equal to the rate λ in the high-frequency limit. In practice, spectra calculated from a finite sample will go to a value close to λ, but fluctuations in the number of spikes in the interval will lead to an error in this estimate. For a given sample, the spectrum will go to the value given by

$$I(f \to \infty) = \frac{1}{N_T K} \sum_{k=0}^{K-1} \sum_{n=1}^{N_T} \sum_{j=1}^{N_n(T)} h_k(t_j^n)^2, \tag{4.24}$$

where t_j^n is the jth spike in the nth trial and $N_n(T)$ is the total number of spikes in the nth trial. In the case of direct and lag window estimators, the averaging over tapers need not be performed.

This expression yields a value that is typically very close to the sample estimate of the mean rate.[7] Significant departures from this high-frequency limit are of interest when interpreting the spectrum, as these indicate enhancement or suppression relative to a homogeneous Poisson process.

4.8 Choice of Estimator, Taper, and Lag Window. The preceding section discussed the large sample statistical properties of direct, lag window, and multitaper estimates of the spectrum. The choice of which estimator to use remains a contentious one (Percival & Walden, 1993). The multitaper method is the most systematic of the estimators, but the lag window estimators should perform almost as well for spike train spectra that have reasonably small dynamic ranges.[8] However, it is possible to construct spike trains with widely different timescales, which can possess a large, dynamic range. In addition, the multitaper technique leads to a simple jackknife procedure by leaving out one data taper in turn. A further important property of the multitaper estimator is that it gives more weight to events at the edges of the time interval and thus ameliorates the arbitrary downweighting of the edges of the data introduced by single tapers.

When using the lag window estimator, there are many choices available for both the taper and the lag window. The choice of taper is generally not critical provided that the taper goes smoothly to zero at the start and end of the interval. A rectangular taper has particularly large side lobes

[7] It is exactly the sample estimate of the mean rate for a rectangular taper.

[8] Dynamic range is a measure of the variation in the spectrum as a function of frequency and is defined as $10 \log_{10} \left(\frac{\max_f S(f)}{\min_f S(f)} \right)$.

in the frequency domain, which can lead to significant bias. The choice of lag window is also usually not critical; typically a gaussian kernel will be satisfactory.

5 Estimating the Coherency

Sample coherency, which may be estimated using equation 5.1, may be evaluated using any of the previously mentioned spectral estimators. The principal difference is that the direct estimator, in terms of which the other estimators are expressed, is given by equations 5.2 and 5.3 rather than 4.1 and 4.2:

$$C(f) = I_{12}^X / \sqrt{I_{11}^X I_{22}^X} \tag{5.1}$$

$$I_{12}^D(f) = J_1^D(f) J_2^D(f)^* \tag{5.2}$$

$$J_a^D(f) = \int_{-\infty}^{\infty} h(t) e^{-2\pi i f t} d\overline{N}_a(t), \tag{5.3}$$

where the $\overline{N}_1(t)$ and $\overline{N}_2(t)$ are simultaneously recorded spike trains from two different cells and X denotes the type of spectral estimator. Possible choices of estimator X include D direct, DT trial-averaged direct, LW lag window, or MT multitaper.

Lag window and multitaper coherency estimates may be constructed by substituting $I_{12}^D(\cdot)$ in place of $I^D(\cdot)$ in equations 4.11 and 4.18. The estimates are biased over a frequency range equal to the width of the smoothing, although the exact form for the bias is difficult to evaluate.

5.1 Confidence Limits for the Coherence. The treatment of error bars is somewhat different between the spectrum and the coherency, since the coherency is a complex quantity. Usually one is interested in establishing whether there is significant coherence in a given frequency band. In order to do this, the sample coherence should be tested against the null hypothesis of zero population coherence. The distribution of the sample coherence under this null hypothesis is

$$P(|C|) = (\nu_0 - 2)|C|(1 - |C|^2)^{(\nu_0/2-2)} \qquad 0 \le |C| \le 1. \tag{5.4}$$

A derivation of this result is given in Hannan (1970). In outline, the method is to rewrite the coherence in such a way that it is equivalent to a multiple correlation coefficient (Anderson, 1984). The distribution of a multiple correlation coeffient is then a known result from multivariate statistics. In the case of coherence estimates based on lag window estimators, the appropriate ν_0 may be used, although this is only approximately valid because this method of derivation assumes integer $\nu_0/2$.

It is straightforward to calculate a confidence level based on this distribution. The coherence will exceed $\sqrt{1-p^{1/(\nu_0/2-1)}}$ only in $p \times 100\%$ of experiments. In addition, it is notable that the quantity $(\nu_0/2-1)|C|^2/(1-|C|^2)$ is distributed as $F_{2(\nu_0-2)}$ under this null hypothesis. It is useful to apply a transformation to the coherence before plotting it, which aids in the assessment of significance. The variable $q = \sqrt{-(\nu_0-2)\log(1-|C|^2)}$ has a Raleigh distribution with density $p(q) = qe^{-q^2/2}$. This density function does not depend on ν_0 and furthermore has a tail that closely resembles a gaussian. For certain values of a fitting parameter β,[9] a further linear transformation $r = \beta(q-\beta)$ leads to a distribution that closely resembles a standard normal gaussian for $r > 2$. This means that for $r > 2$, one can interpret r as the number of standard deviations by which the coherence exceeds that expected under the null hypothesis.

5.2 Confidence Limits for the Phase of the Coherency. If there is no population coherency, then the phase of the sample coherency is distributed uniformly. If there is population coherency, then the distribution of the sample phase is approximately gaussian provided that the tails of the gaussian do not extend beyond a width 2π. An approximate 95% confidence interval for the phase (Rosenberg, Amjad, Breeze, Brillinger, & Halliday, 1989; Brillinger, 1974) is

$$\hat{\phi}(f) \pm 2\sqrt{\frac{2}{\nu_0}\left(\frac{1}{|C(f)|^2}-1\right)}, \tag{5.5}$$

where $\hat{\phi}(f)$, the sample estimate of the coherency phase, is evaluated using $\tan^{-1}\{Im(C)/Re(C)\}$.

6 Finite Size Effects

In the preceding sections, error bars were given for estimators of the spectrum and the coherence. However, these error bars were based on large sample sizes (they apply asymptotically as $T \to \infty$). Neurophysiological data are not collected in this regime, and particularly in awake behaving studies where data are often sparse, corrections arising at small T are potentially important. In order to estimate the size of these corrections, a particular model for the point process is required. The model studied was chosen primarily for its analytical tractability while still maintaining a nontrivial spectrum.

The model and the final results are presented here, but the details of the analysis are reserved for the appendix. The model is a doubly stochastic inhomogeneous Poisson process with a gaussian rate function. A specific realization of a spike train is generated from the model in the following

[9] A reasonable choice for β is 23/20.

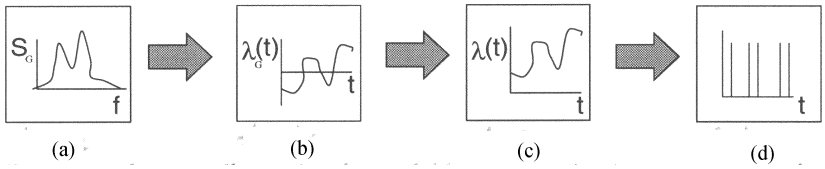

Figure 5: Schematic illustrating the model for which finite size corrections to the asymptotic error bars will be evaluated. (a) A spectrum $S_G(f)$ is defined. (b) A realization $\lambda_G(t)$ is drawn from a gaussian process with this spectrum. (c) The mean rate λ is added to $\lambda_G(t)$ to obtain $\lambda(t)$. (d) This rate function is used to generate a realization of an inhomogeneous Poisson process, yielding a set of spike times.

manner. First, a population spectrum $S_G(f)$ is specified. From this, a realization of a zero mean gaussian process $\lambda_G(f)$ is generated. A constant λ, the mean rate, is then added to this realization. This function is then considered to be the rate function for an inhomogeneous Poisson process. A realization of this inhomogeneous Poisson process is then generated. A schematic of the model is shown in Figure 5.

Technically this is not a valid process because the rate function $\lambda(t)$ may be negative. However, if the area underneath the spectrum is small enough, then the fluctuations about the mean rate seldom cross zero, and corrections due to this effect are negligible. In addition, large violations of this area constraint have been tested by Monte Carlo simulation, and the results still apply to a good approximation.

An important feature of this model is that the population spectrum of the spike trains is simply the spectrum of the inhomogeneous Poisson process rate function plus an offset equal to the mean rate.[10] The spectrum of the rate function is a positive real quantity, and therefore within this model the population spectrum cannot be less than the mean rate at any frequency. Intuitively, the reason is that the process must be more variable than a homogeneous Poisson process at all frequencies.

To make the nature of the result clear, a simplified version is given in equation 6.1. This version is for the particular case of a homogeneous Poisson process (which has a flat population spectrum) and a rectangular taper:[11]

$$\mathrm{var}\{I^X(f)\} = \lambda^2 \left[\frac{2}{\nu_0} + \frac{1}{N_T T \lambda} \right], \tag{6.1}$$

where λ is the mean rate.

[10] This result does not depend on the gaussian assumption.

[11] The expression also holds approximately for the multitaper estimate provided all tapers up to the Shannon limit are used.

A sample-based estimate of $N_T T\lambda$ is the total number of spikes over all trials. It is to be noted that finite-size effects reduce the degrees of freedom. This result implies that there is a point beyond which additional smoothing does not decrease the variance further, and this point is approximately when ν_0 is equal to twice the total number of spikes. The full result is given in equations 6.2 to 6.7:

$$\mathrm{var}\{I^X(f)\} = \frac{2E\{I^X(f)\}^2}{\nu(f)} \tag{6.2}$$

$$\frac{1}{\nu(f)} = \frac{1}{\nu_0} + \frac{C_h^X \Phi(f)}{2TN_T E\{I^X(f)\}^2}, \tag{6.3}$$

where

$$C_h^X = \begin{cases} \int_0^1 f(t)^4 dt & \text{if} \quad X = \text{LW, D or DT} \\ \frac{1}{K^2} \sum_{k,k'} \int_0^1 f_k(t)^2 f_{k'}(t)^2 dt & \text{if} \quad X = \text{MT} \end{cases} \tag{6.4}$$

$$f(t/T) = \sqrt{T} h(t) \tag{6.5}$$

and

$$\Phi(f) = \lambda_{hf} + 4[E\{I^X(f)\} - \lambda_{hf}] + 2[E\{I^X(0)\} - \lambda_{hf}] + [E\{I^X(2f)\} - \lambda_{hf}] \tag{6.6}$$

$$\lambda_{hf} = E\{I^X(f \to \infty)\}. \tag{6.7}$$

C_h^X is a constant of order unity, which depends on the taper. When a taper is used to control bias, some of the spikes are effectively disregarded, and this has an effect on the size of the correction. The function $f(t)$ has the same form as the taper $h(t)$ but is defined for the interval $[0, 1]$. C_h^X is the integral of the fourth power of f and obtains its minimum value of one for a rectangular taper. It is typically between 1 and 2 for other tapers. In the multitaper case, cross-terms between tapers are included.

Equation 6.6 describes how the finite-size correction depends on the structure of the spectrum. $\Phi(f)$ is the sum of four terms. The first term is the only one present for a flat spectrum. The second term is a correction that depends on the spectrum at the frequency being considered. The next two terms depend on the spectrum at zero frequency and the spectrum at twice the frequency being considered. The final three terms all depend on the difference between the spectrum at some frequency and the high-frequency limit. Equation 6.6 applies provided that the spike train is well described by the model. However, this is not necessarily the case, and a suppression of the spectrum, which cannot be described by the model, often occurs at low

frequencies.[12] In the event that there is a significant suppression of the spectrum, $\Phi(f)$ may become small or even negative. To avoid this, a modified form for $\Phi(f)$ that prevents this may be used:

$$\begin{aligned}\Phi(f) = \; & \lambda_{hf} + 4\max([E\{I^X(f)\} - \lambda_{hf}], 0) \\ & +2\max([E\{I^X(0)\} - \lambda_{hf}], 0) \ldots \\ & + \max([E\{I^X(2f)\} - \lambda_{hf}], 0). \end{aligned} \tag{6.8}$$

The modification to the result is somewhat ad hoc, so Monte Carlo simulations of spike trains with enforced refractory periods have been performed to test its validity. These simulations demonstrated that although the correction derived using equation 6.8 was significantly different from that obtained from the Monte Carlo simulations in the region of the suppression, equation 6.8 provided a pessimistic estimate in all cases studied. This increases confidence that applying finite-size corrections using equation 6.8 will provide reasonable error bars for small samples.

Equation 6.3 gives the finite-size correction in terms of a reduction in ν_0. The new $\nu(f)$ may be used to put confidence intervals on the results, as described in section 4.6, although the accuracy of the χ^2_ν assumption will be reduced. In the case of coherence, an indication of the correction to the confidence level can be obtained by using the smaller of the two $\nu(f)$ from the spike train spectra to calculate the confidence level using equation 5.4. In all cases, if the effect being observed achieves significance only by an amount that is of the same order as the finite-size correction, then it is recommended that more data be collected.

7 Experimental Design

Often it is useful to know in advance how many trials or how long a time interval one needs in order to resolve features of a certain size in the spectrum or the coherence. To do this, one needs to estimate the asymptotic degrees of freedom ν_0. This depends on the size of feature to be resolved α, the significance level for which confidence intervals will be calculated p, and the fraction of experiments that will achieve significance $\mathcal{P}$. In addition, the reduction in the degrees of freedom due to finite size effect depends on the total number of spikes N_s and also C_h (see section 6).

An estimate of v_0 may be obtained in two stages. First, α,p and $\mathcal{P}$ are specified and used to calculate degrees of freedom ν. Second, the asymptotic degrees of freedom ν_0 is estimated using ν, N_s, and C_h. The feature size $\alpha = (S-\lambda)/\lambda$ is the minimum size of feature that the experimenter is content to resolve. For example, a value of 0.5 indicates that where the population

[12] Note that any spike train spectra displaying significant suppression below the mean firing rate can immediately rule out the inhomogeneous Poisson process model.

spectrum exceeds 1.5λ, the feature will be resolved. The significance level should be set to the same value that will be used for calculating the confidence interval for the spectrum, typically 0.05. For a given p, there is some probability $\mathcal{P}$ that an experiment will achieve significance. To calculate ν, one begins with a guess ν_g. Then q_1 is chosen such that $P\left[\chi^2_{\nu_g} \geq q_1\right] = p/2$. On the basis of this, one then evaluates $\mathcal{P}, = 1 - \Phi\left[q_1/(1+\alpha)\right]$ where Φ is the cumulative $\chi^2_{\nu_g}$ distribution.[13] If $\mathcal{P}$, is equal to the specified fraction $\mathcal{P}$, then $\nu = \nu_g$; otherwise a different ν_g is chosen. This procedure is readily implemented as a minimization of $(\mathcal{P} - \mathcal{P},(\nu_g))^2$ on a computer. Having obtained ν, one can estimate ν_0 using

$$\frac{1}{\nu_0} = \frac{1}{\nu} - \frac{C_h\,[1+4\alpha]}{2N_s\,[1+\alpha]^2}, \tag{7.1}$$

where the 4α is omitted from the numerator if $\alpha < 0$.

Figure 6 illustrates example design curves generated using this method. These curves show the asymptotic degrees of freedom as a function of feature size for different total numbers of spikes.

The existence of a region bounded by vertical asymptotes implies that as long as the total number of measured spikes is finite, modulations in the spectrum below a certain level cannot be detected no matter how much the spectrum is smoothed. These curves may be used to design experiments capable of resolving spectral features of a certain size.

In the case of the coherence, one calculates how many degrees of freedom are required for the confidence line to lie at a certain level as described in section 5.1.

8 Line Spectra

One of the assumptions underlying the estimation of spectra is that the population spectrum varies slowly over the smoothing width (W for multitaper estimators). While this is often the case, there are situations in which the spectrum contains very sharp features, which are better approximated by lines than by a continuous spectrum. This corresponds to periodic modulations of the underlying rate, such as when a periodic stimulus train is presented. In such situations, it is useful to be able to test for the presence of a line in a background of colored noise (i.e., in a locally smooth but otherwise arbitrary continuous population spectrum). Such a test has been previously developed, in the context of multitaper estimation, for continuous processes

[13] These formulas apply for $\alpha > 0$. If $\alpha < 0$, then $P\left[\chi^2_{\nu_g} \leq q_1\right] = p/2$ and $\mathcal{P}, = \Phi\left[q_1/(1+\alpha)\right]$ should be used.

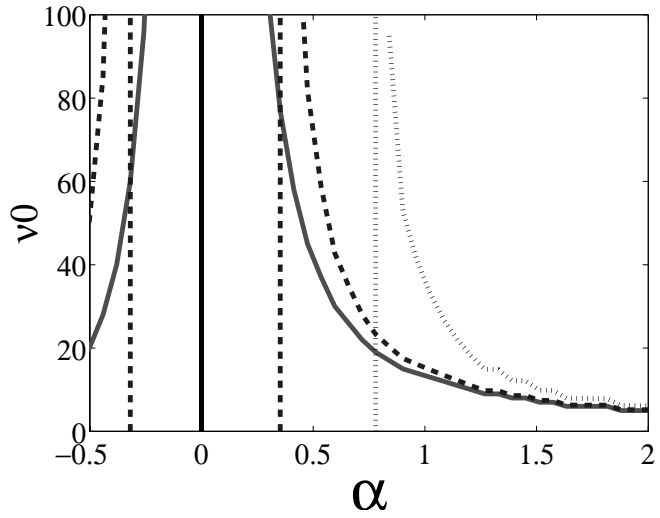

Figure 6: Example design curves for the case when $p = 0.05$, $\mathcal{P} = 0.5$, and $C_h = 1.5$. The three curves correspond respectively to $N_s = \infty$ (solid line), $N_s = 100$ (dashed line), and $N_s = 20$ (dotted line).

(Thomson, 1982). In the following section the analogous development for point processes is presented.

8.1 F-Test for Point Processes. A line in the spectrum has an exactly defined frequency, and consequently the process $N(t)$ has a nonzero first moment. The natural model in the case of a single line is given by

$$E\{dN(t)\}/dt = \lambda_0 + \lambda_1 \cos(2\pi f_1 t + \phi). \tag{8.1}$$

A zero mean process ($\overline{N}$) may be constructed by subtraction of an estimate of $\lambda_0 t$. Provided that the product of the line frequency(f_1) and the sample duration(T) is much greater than one, the sample quantity $N(T)/T$ is an approximately unbiased estimate of λ_0. The resultant zero mean process $\overline{N}$ has a Fourier transform that has a nonzero expectation:

$$J_k(f) = \int_{-\infty}^{\infty} h_k(t) e^{-2\pi i f t} d\overline{N}(t) \tag{8.2}$$

$$E\{J_k(f)\} = c_1 H_k(f - f_1) + c_1^* H_k(f + f_1), \tag{8.3}$$

where,

$$c_1 = \lambda_1 e^{i\phi}/2. \tag{8.4}$$

In the case where $f > 0$ and $f_1 > W$,

$$E\{J_k(f)\} \simeq c_1 H_k(f - f_1). \tag{8.5}$$

The estimates of $J_k(f_1)$ from different tapers provide a set of uncorrelated estimates of $c_1 H_k(0)$. It is hence possible to estimate the value of c_1 by complex regression:

$$\hat{c}_1 = \frac{\sum_k J_k(f_1) H_k(0)}{\sum_k |H_k(0)|^2}. \tag{8.6}$$

Under the null hypothesis that there is no line in the spectrum ($c_1 = 0$), it may readily be shown that $E\{\hat{c}_1\} = 0$ and $\text{var}\{\hat{c}_1\} = S(f_1)/\sum_k |H_k(0)|^2$. The residual spectrum,[14] which has the line removed, may be estimated using

$$\hat{S}(f) = \frac{1}{K} \sum_k |J_k(f) - \hat{c}_1 H_k(f - f_1)|^2. \tag{8.7}$$

In the large sample limit, the distributions of both $\hat{c}_1$ and $\hat{S}(f_1)$ are known (Percival & Walden, 1993) and may be used to derive

$$\frac{|\hat{c}_1|^2 \sum_k |H_k(0)|^2 (K-1)}{\sum_k |J_k(f_1) - \hat{c}_1 H_k(0)|^2} \doteq F_{2,2(K-1)}, \tag{8.8}$$

Where $\doteq$ denotes "is distributed as."

The null hypothesis may be tested using this relation; if it is rejected, the line can be removed using equation 8.7 to estimate the residual spectrum. It is worth noting that although relation 8.8 was derived for large samples, the test is remarkably robust as the sample size is decreased. Numerical tests indicate that the tail of the F distribution is well reproduced even in situations where there are as few as five spikes in total.

8.2 Periodic Stimulation. A common paradigm in neurobiology where line spectra are particularly important is that of periodic stimulation. When a neuron is driven by a periodic stimulation of frequency f_1, the spectrum may contain lines at any of the harmonics nf_1. Provided that $f_1 > 2W$, the analysis of section 8.1 applies with each harmonic being separately tested for significance.

The first moment of the process, which has period $1/f_1$, is given by equation 8.9 and may be estimated using $\hat{c}_n$:

$$\lambda(t) = \lambda_0 + \sum_n \lambda_n \cos(2\pi n f_1 t + \phi_n). \tag{8.9}$$

[14] It is also possible to estimate a residual coherency. In order to do this, one uses a residual cross-spectrum $\hat{S}_{xy}(f) = \frac{1}{K} \sum_k (J_k^x(f) - \hat{c}_1^x H_k(f - f_1))^* (J_k^y(f) - \hat{c}_1^y H_k(f - f_1))$, together with the residual spectra to evaluate the usual expression for coherency.

where $\lambda_n = 2|c_n|$, $\phi_n = \tan^{-1}\{Im(c_n)/Re(c_n)\}$, the sum is taken over all the significant coefficients.

This rate function $\lambda(t)$ is the average response to a single stimulus or impulse response. The coefficients c_n are the Fourier series representation of $\lambda(t)$.

8.3 Error Bars. It is possible to put confidence intervals on both the modulus and the phase of the coefficients $\hat{c}_n$. For large samples (more than 10 spikes), the real and imaginary parts of $\hat{c}_n$ are distributed as independent gaussians, each with standard deviation $\sigma_n = \sqrt{S(nf_1)/(2\sum_k |H_k(0)|^2)}$. For $c_n/\sigma_n > 3$, the distribution of $|\hat{c}_n|$ is well approximated by a gaussian centered on $|c_n|$ and with standard deviation σ_n. In addition, the estimated phase angle ($\hat{\phi}_n$) is also almost gaussian with mean ϕ_n and standard deviation $\sigma_n/|c_n|$. Approximate error bars or confidence intervals may be obtained using a sample-based estimate of σ_n, $\hat{\sigma}_n = \sqrt{\hat{S}(nf_1)/(2\sum_k |H_k(0)|^2)}$.

Estimating error bars for the impulse response function is more involved due to their nonlocal nature (if one of the Fourier coefficients is varied the impulse response function changes everywhere). It is therefore of interest to estimate a global confidence interval, defined as any interval such that the probability of the function crossing the interval anywhere is some predefined probability. A method for estimating a global confidence band is detailed in Sun and Loader (1994) and outlined here. First, a basis vector $\Phi(t)$ is constructed:

$$\Phi(t) = \begin{bmatrix} \hat{\sigma}_1 \cos(2\pi f_1 t) \\ \vdots \\ \hat{\sigma}_N \cos(2\pi f_N t) \\ \hat{\sigma}_1 \sin(2\pi f_1 t) \\ \vdots \\ \hat{\sigma}_N \sin(2\pi f_N t) \end{bmatrix} \tag{8.10}$$

where N is the total number of harmonics.

The elements of this vector have unit variance, and a standard approximation, which relies on σ_n being gaussian, may be applied.

$$P(sup|\lambda(t) - E\{\lambda(t)\}| > c||\Phi(t)||) \leq 2(1 - N(c)) + (k/\pi)e^{-c^2/2} \tag{8.11}$$

Where *sup* is the maximum value of its operand, $||\Phi(t)||$ denotes the length of vector $\Phi(t)$, $N(c)$ is the cumulative standard normal distribution, and k is a constant. k may be evaluated by constructing the $2 \times N$ matrix $X(t) = [\Phi(t)\ d\Phi(t)/dt]$, forming its QR decomposition (Press, Teukolsky, Vetterling, & Flannery, 1992) and then evaluating $k = \int_0^T |R_{22}(t)/R_{11}(t)|dt$.

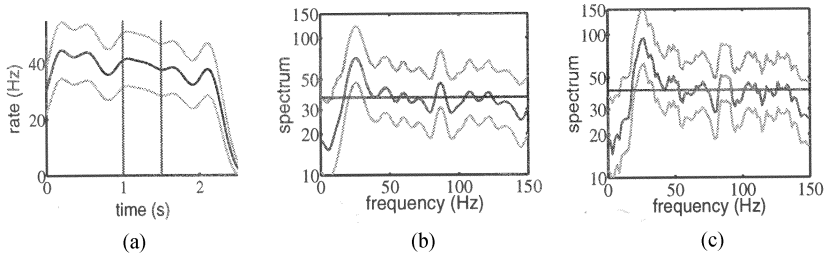

Figure 7: (a) Gaussian kernel (100 ms width) smoothed firing rate with 2σ error bars based on a stationarity assumption. The vertical lines indicate the period over which the spectrum was calculated. A light is flashed at time zero, and the spectrum is evaluated over the interval when the monkey is required to remember the target location. (b) The spectrum evaluated over this interval using a lag window estimator with a 40% cosine taper and a gaussian lag window of width 3.5 Hz. 95% confidence limits are shown with the finite size correction included (this typically resulted in a decrease in $\nu(f)$ from about 50 to 36). The horizontal line indicates the high-frequency limit. (c) The same spectrum evaluated using a multitaper estimator. A bandwidth (W) of 5 Hz was used allowing five tapers. Both estimators have the same degrees of freedom.

Confidence intervals for the residual spectrum are calculated in the usual manner (using χ^2_ν), although at the line frequencies, the interval is slightly broadened due to the loss of two degrees of freedom incurred by estimation of c_n. Section 11 contains an example application of the methods described in this section.

9 Example Spectra

Figure 7 is a spectrum calculated from data collected from a single cell recorded from area PRR in the parietal cortex of an awake behaving monkey during a delayed memory reach task (Snyder et al., 1997). The spectrum is calculated over an interval of 0.5 second during which the firing rate is reasonably stationary (the average firing rate lies within the error bars) and is averaged over five trials. The spectrum shows two features that achieve significance. There is enhancement of the spectrum in the frequency band 20–40 Hz, indicating the presence of an underlying broadband oscillatory mode in the neuronal firing rate. In addition, there is suppression of the spectrum at low frequencies. As discussed previously, a suppression of this sort is consistent with an effective refactory period during which the neuron is less likely to fire. Care must be taken at low frequencies since at frequencies comparable to the smoothing width, the spectrum is particularly sensitive to any nonstationarity in the data.

It may be useful for the practitioner to review parameter choice and the relative merits of multitaper and lag window estimators. First, there is the

choice of time interval. This was chosen to be as short as possible, consistent with resolving the structure of interest. Choice of bandwidth is a subject on which a considerable literature exists and which is beyond the scope of this article (Loader, 1999). However, in practice, it is usually straightforward to choose a reasonable bandwidth by trial and error. A balance should be sought between undersmoothing, where the structure is not clear due to the high variance of the estimator, and oversmoothing, where the structure is blurred out because of excessive frequency averaging. One of the advantages of the multitaper method over the lag window method is that once the time interval and bandwidth have been set, the calculation of the spectrum is automatic and, in the sense defined in section 4.4, optimal. If a lag window estimate is used, then the trade-off between bias and variance is not formalized and one has to choose the taper and kernel smoother separately (see section 4.8). Although both the lag window and the multitaper estimates reveal the same structure in the example, this may not have been the case if the spectrum had a larger dynamic range. Whichever method is used, exploratory data analysis is always recommended when choosing the time interval and bandwidth.

10 Example Coherency

To illustrate the estimation of coherency, simulated spike trains were generated from a coupled doubly stochastic Poisson process. For a given trial, a pair of rate functions was drawn from a gaussian process. The realizations share a coherent mode that is linearly mixed into the rates of both cells. These coupled rate functions are then used to draw a realization of an inhomogeneous Poisson process for each cell independently. Using this method, 15 trials of duration 0.5 second were generated. The coherent mode was set such that the population coherence was a gaussian of height 0.35 and standard deviation 5 Hz centered on 20 Hz. The phase of this mode was set to 180 degrees. Figure 8 indicates that this coherent mode is reasonably estimated.

11 Example Periodic Stimulation

An example of an analysis of a periodic stimulus paradigm is shown in Figure 9. The datum is a single cell recording collected from the barrel cortex of an awake behaving rat during periodic whisker stimulation at 5.5 Hz (Sachdev, Jenkinson, Melzer, & Ebner, 1999). There is a single trial of duration 50 seconds. During the trial, the average firing rate, estimated using a 2 second kernel, is constant to within the error bars expected for a Poisson process.

The estimated impulse response function $\hat{\lambda}(t)$ is seen to have two distinct sharp peaks, outside of which the response does not differ significantly from zero. The moduli of the Fourier coefficients are significant out to $n = 25$. This

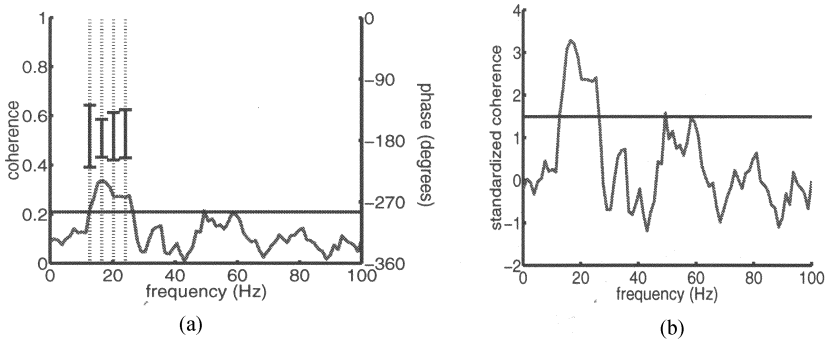

Figure 8: (a) Coherence (left axis) and phase of the coherency (right axis). Fifteen trials of 0.5 second duration were simulated using a doubly stochastic Poisson process as described in the text. A multitaper estimator with a smoothing width of 7 Hz was used. Finite size corrections were used and resulted in 25% reduction in the degrees of freedom. A horizontal line has been drawn at the 95% confidence level under the null hypothesis of no coherency. Where the null hypothesis is rejected, the phase of the coherency is estimated and shown with an approximate 95% confidence interval. (b) The standardized coherence is a transformation that maps the null distribution onto an approximately standard normal variate (as described in section 5.1). The estimated coherence at 20 Hz would therefore lie at three standard deviations if there were no population coherence.

automatically sets the smoothing of $\hat{\lambda}(t)$ as structure on a timescale of less than $1/(25 \times 5.5) = 7$ ms does not achieve significance. Note that the coefficients are enhanced at multiples of 6 (approximately 33 Hz), which comes from having two peaks in the time domain $\lambda(t)$ separated by approximately 30 ms. The phase of the coefficients closely follows a straight line, but there is a small, periodic deviation from this line, which is again at index multiples of 6. The gradient of the straight line depends on the time delay of the response. The residual spectrum was calculated by first evaluating a multitaper estimate from which the significant harmonics were removed. This spectrum had a bandwidth of 1.5 Hz, chosen to avoid overlap of the harmonics leading to the multitaper estimate being undersmoothed. A further smoothing was performed using a lag window.[15] The resultant spectrum displays a slight but significant suppression relative to a Poisson process out to almost 200 Hz. Such a spectrum is characteristic of a short timescale refractive period. The residual spectrum is particularly useful because rate nonstationarity has

[15] The previous theory developed for lag window estimators applies to this hybrid estimator with $|H(\cdot)|^2$ replaced by $\frac{1}{K}\sum_{k=0}^{K-1}|H_k(\cdot)|^2$ in equation 4.13 and $|\mathcal{H}(\cdot)|^2$ replaced by $\frac{1}{K}\sum_{k,k'=0}^{K-1}|\mathcal{H}_{kk'}(\cdot)|^2$ in equation 4.15.

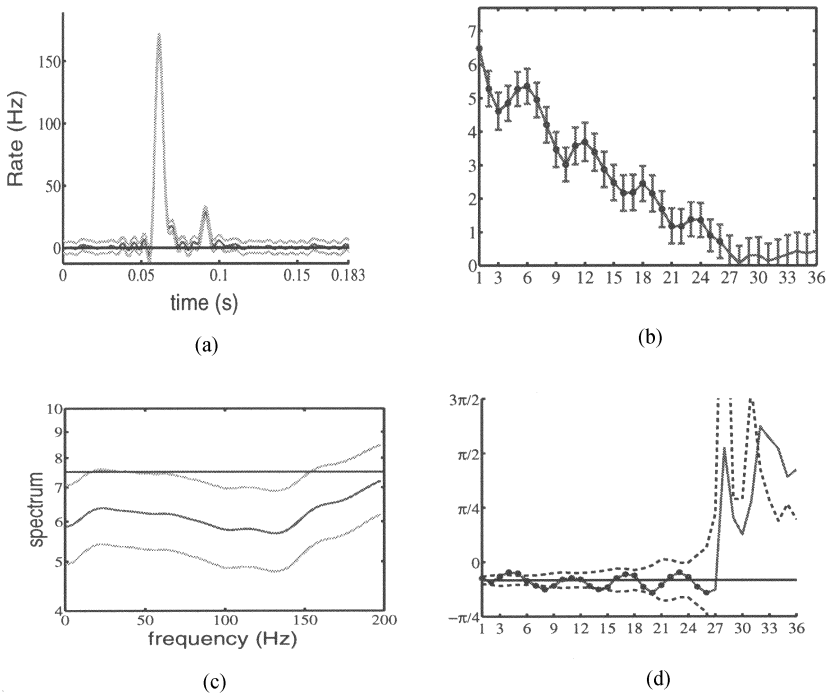

Figure 9: Response to a periodic stimulation of frequency 5.5 Hz. (a) Impulse response function with global 95% confidence interval. (b) $|\hat{c}_n|$ versus index n with 95% confidence interval. Dots indicate points that achieved significance in the F-test. (c) Residual spectrum with finite-size-corrected confidence interval. A multitaper spectrum with 100 tapers and a bandwidth of 1.5 Hz was used initially to avoid overlap of harmonics. This spectrum was then further smoothed using a gaussian lag window with standard deviation 9 Hz. (d) The coefficient phases $\hat{\phi}_n$ (in radians) versus index n after subtraction of a fitted straight line of gradient $2\pi/3 \pm 0.01$. The black dashed lines are a 95% confidence interval about zero.

been removed. As was the case for the spectrum, the choice of bandwidth for the residual spectrum depends on the observed structure in the data. The large bandwidth of 9 Hz was chosen because the residual spectrum is very flat and can therefore tolerate a high degree of smoothing.

12 Summary

It is our belief that spectral analysis is a fruitful and underexploited analysis technique for spike trains. In this article, an attempt has been made to collect

Table 1: Basic Direct Spectral Estimator in Terms of Which the Other Estimators Can Be Written.

$J_a^{kn}(f) = \int_0^T h_k(t) e^{-2\pi i f t} d\overline{N}_a^n(t)$
$I_{ab}^{kn}(f) = J_a^{kn}(f) J_b^{kn*}(f)$

Notes: For clarity, the superscript D on the direct spectral estimate has been omitted. The index n labels trials, index k labels tapers, and indices a and b label cells.

Table 2: Estimators and Large Sample Degrees of Freedom ν_0 of Estimates of the Spectrum ($ab = 11$).

X	$I_{ab}^X(f)$	Equation Number	ν_0
D	$I_{ab}^{01}(f)$	4.1	2
DT	$\frac{1}{N_T}\sum_{n=1}^{N_T} I_{ab}^{0n}(f)$	4.9	$2N_T$
LW	$\frac{1}{N_T}\sum_{n=1}^{N_T}\int_{-\infty}^{\infty} K(f-f') I_{ab}^{0n}(f') df'$	4.11	$2N_T/\xi$
MT	$\frac{1}{N_T K}\sum_{n=1}^{N_T}\sum_{k=0}^{K-1} I_{ab}^{kn}(f)$	4.18	$2N_T K$

Note: The indices on the I_{ab}^{kn} are as follows: ab label the cells from which the estimates are constructed, k labels the taper, and n labels the trial.

the machinery necessary for performing spectral analysis on spike train data into a single document. Starting from the population definitions, the statistical properties of estimators of the spectrum and coherency have been reviewed. Estimation methods for both continuous spectra and spectra that contain lines have been included. In addition, new corrections to asymptotic error bars have been presented that increase confidence in applying spectral techniques in practical situations where data are often sparse. Tables 1 to 5 summarize the important formulas. Matlab software implementing the methods discussed in this article is available online from http://www.vis.caltech.edu/~WAND/.

Appendix: Derivation of Finite Size Correction

Following is an outline derivation of the finite-size corrections described in section 6. First, the characteristic functionals (Bartlett, 1966) for the processes $\overline{N}$ and the inhomogeneous Poisson process rate $\lambda(t)$ are related:

$$C_{\overline{N}}(\theta(t)) = E\left\{\exp\left(i\int_0^T \theta(t) d\overline{N}\right)\right\}$$

Table 3: Main Formulas Required for Estimating Spectral Error Bars.

	Equation	Equation Number	Comment
Variance	$\text{var}\{I^X_{aa}(f)\} = \frac{2E\{I^X_{aa}(f)\}^2}{\nu(f)}$	4.20	Use ν_0 for asymptotic or $\nu(f)$ if using finite size correction
Degrees of freedom	$\frac{1}{\nu(f)} = \frac{1}{\nu_0} + \frac{C^X_h \Phi(f)}{2TN_T E\{I^X(f)\}^2}$	6.3	See text for definitions of C^X_h and $\Phi(f)$
Confidence $(1-2p)\times 100\%$	$\left[\nu I^X(f)/q_2, \nu I^X(f)/q_1\right]$	4.23	q_1 s.t $P[\chi^2_\nu \le q_1] = p$ q_2 s.t $P[\chi^2_\nu \ge q_2] = p$

Note: Refer to section 4 for additional information.

Table 4: Main Formulas Required for Coherency Estimation.

	Equation	Equation Number	Comment
Coherency	$C^X(f) = \frac{I^X_{ab}}{\sqrt{I^X_{aa} I^X_{bb}}}$	5.1	
Distribution for coherence	$P(\lvert C\rvert) = (\nu-2)\lvert C\rvert(1-\lvert C\rvert^2)^{(\nu/2-2)}$	5.4	Under null hypothesis $\gamma = 0$
Confidence for phase	$\hat{\phi}(f) \pm 2\sqrt{\frac{2}{\nu}\left(\frac{1}{\lvert C(f)\rvert^2} - 1\right)}$	5.5	Approximately 95%

Note: Refer to section 5 for additional information.

Table 5: Main Formulas Required for the Detection and Removal of a Line from the Spectrum.

	Equation	Equation Number	Comment
Complex amplitude of line	$\hat{c}_1 = \frac{\sum_k J_k(f_1) H_k(0)}{\sum_k \lvert H_k(0)\rvert^2}$	8.6	
F-test to access the significance of a line	$\frac{\lvert\hat{c}_1\rvert^2 \sum_k \lvert H_k(0)\rvert^2 (K-1)}{\sum_k \lvert J_k(f_1) - \hat{c}_1 H_k(0)\rvert^2} \doteq F_{2,2(K-1)}$	8.8	Null $c_1 = 0$
Residual spectrum	$\hat{S}(f) = \frac{1}{K}\sum_k \lvert J_k(f) - \hat{c}_1 H_k(f-f_1)\rvert^2$	8.7	

Note: Refer to section 8 for additional information.

$$= E_\lambda \left\{ \exp\left(\int_0^T \lambda(t) b(\theta(t)) dt \right) \right\} \tag{A.1}$$

$$b(\theta(t)) = \exp\left[i\theta(t) - \frac{i}{T} \int_0^T \theta(t') dt' \right]. \tag{A.2}$$

Under the gaussian process assumption for $\lambda(t)$, this integral may be done:

$$C_{\overline{N}} = \exp\left[\frac{1}{2} \int_0^T \int_0^T b(t) \Lambda(t, t') b(t') dt dt' + \overline{\lambda} \int_0^T b(t) dt \right] \tag{A.3}$$

$$\Lambda(t, t') = E_\lambda \{ (\lambda(t) - \overline{\lambda})(\lambda(t') - \overline{\lambda}) \}. \tag{A.4}$$

Note that $\overline{\lambda}$ denotes the mean rate. Taking the log of the characteristic functionals now yields the following relation between the resultant cumulant functionals:

$$K_{\overline{N}} = lnE \left\{ \exp(i \int_0^T \theta(t) d\overline{N}) \right\} = \frac{1}{2} \int_0^T \int_0^T b(t) \Lambda(t, t') b(t') dt dt' + \overline{\lambda} \int_0^T b(t) dt. \tag{A.5}$$

Next, $\theta(t)$ is chosen appropriately and substituted into $K_{\overline{N}}$. The form for $\theta(t)$, which is required to obtain the covariance of multitaper estimators, is

$$i\theta(t) = \theta_1 h_k(t) e^{-2\pi i f_1 t} + \theta_2 h_k(t) e^{2\pi i f_1 t} + \theta_3 h_{k'}(t) e^{-2\pi i f_2 t} + \theta_4 h_{k'}(t) e^{2\pi i f_2 t}. \tag{A.6}$$

Substituting into the cumulant functional for $\overline{N}$ yields

$$K_{\overline{N}} = lnE\{\exp(\theta_1 J_k^D(f_1) + \theta_2 J_k^{D*}(f_1) + \theta_3 J_{k'}^D(f_2) + \theta_4 J_{k'}^{D*}(f_2))\}, \tag{A.7}$$

where J_k^D is the Fourier transform of the data tapered by a function indexed by k. Application of the cumulant expansion theorem (Ma, 1985) then leads to

$$K_{\overline{N}} = E\{\exp(\theta_1 J_k^D(f_1) + \theta_2 J_k^{D*}(f_1) + \theta_3 J_{k'}^D(f_2) + \theta_4 J_{k'}^{D*}(f_2)) - 1\}_C. \tag{A.8}$$

This may then be differentiated and set to zero:

$$K_{lmno} = \left. \frac{\partial K_{\overline{N}}}{\partial \theta_1^l \partial \theta_2^m \partial \theta_3^n \partial \theta_4^o} \right|_{\theta_1 = \theta_2 = \theta_3 = \theta_4 = 0} = E\{J_k^{Dl}(f_1) J_k^{Dm*}(f_1) J_{k'}^{Dn}(f_2) J_{k'}^{Do*}(f_2)\}_C. \tag{A.9}$$

Moments of the estimators may be expressed in terms of these cumulant derivatives. The expressions are simplified by the fact that all cumulant derivatives that have indices summing to an odd number are zero because $\overline{N}$ is a zero mean process:

$$E\{I_k^D(f)\} = K_{1100} \tag{A.10}$$

$$\mathrm{var}\{I^{MT}(f)\} = \frac{1}{K^2}\sum_{k=0}^{K-1}\sum_{k'=0}^{K-1}\mathrm{cov}\{I_k^D(f), I_{k'}^D(f)\} \tag{A.11}$$

$$\mathrm{cov}\{I_k^D(f), I_{k'}^D(f)\} = K_{1010}K_{0101} + K_{1111} + K_{1001}K_{0110}. \tag{A.12}$$

The problem has now been reduced to that of calculating these derivatives within the model. This is done by substituting the expression for $\theta(t)$ into the right-hand side of equation A.5. Considerable algebra then leads to the following exact result:

$$K_{lmno} = K_{lmno}^A + K_{lmno}^B, \tag{A.13}$$

where

$$K_{lmno}^A = \frac{1}{2}\sum_{l_i,m_i,n_i,o_i}\frac{l!m!n!o!}{\Pi l_i!\Pi m_i!\Pi n_i!\Pi o_i!}\left[\frac{-H_1(f_1)}{T}\right]^{l_2+l_4}\left[\frac{-H_1(f_1)^*}{T}\right]^{m_2+m_4}\cdots \left[\frac{-H_1(f_2)}{T}\right]^{n_2+n_4}\left[\frac{-H_1(f_2)^*}{T}\right]^{o_2+o_4} I_{l_3,m_3,n_3,o_3}^{l_1,m_1,n_1,o_1} \tag{A.14}$$

where $\sum_i l_i = l$ and cases where $l_1 + l_2 = l$ or $l_3 + l_4 = l$ are excluded (and also for n, m, o):

$$I_{l_3,m_3,n_3,o_3}^{l_1,m_1,n_1,o_1} = \int_\infty^\infty S_\lambda(f)H_{l_1+m_1+n_1+o_1}[f_1(l_1-m_1)+f_2(n_1-o_1)-f]\cdots H_{l_3+m_3+n_3+o_3}^*[f_1(l_3-m_3)+f_2(n_3-o_3)-f]df, \tag{A.15}$$

where $S_\lambda(f)$ is the spectrum of the gaussian process and H_l is

$$H_l(f) = \int_{-\infty}^\infty h^l(t)\exp(-2\pi ift)dt \tag{A.16}$$

$$H_0(f) = T\exp(-i\pi fT)sinc(\pi fT) \tag{A.17}$$

$$K^B_{lmno} = \bar{\lambda} \sum_{p=0}^{l} \sum_{q=0}^{m} \sum_{r=0}^{n} \sum_{s=0}^{o} \begin{bmatrix} l \\ p \end{bmatrix} \begin{bmatrix} m \\ q \end{bmatrix} \begin{bmatrix} n \\ r \end{bmatrix} \begin{bmatrix} o \\ s \end{bmatrix} H_{p+q+r+s}$$
$$\times [f_1(p-q) + f_2(r-s)] \cdots \left[\frac{-H_1(f_1)}{T}\right]^{(l-p)}$$
$$\times \left[\frac{-H_1(f_1)^*}{T}\right]^{(m-q)} \left[\frac{-H_1(f_2)}{T}\right]^{(n-r)} \left[\frac{-H_1(f_2)^*}{T}\right]^{(o-s)} . \quad \text{(A.18)}$$

The preceding result is somewhat cumbersome but readily evaluated computationally for a given spectrum. The expression simplifies greatly when only frequencies above the smoothing width are considered and many of the terms may be neglected. Restricting attention to the second-order properties, there are only a few remaining dominant terms. Terms from K_{1001} lead to the previously discussed asymptotic results, but there are corrections that arise from the term K_{1111}. Assuming that the population spectrum varies slowly over the width of the tapers leads to the result given by equations 6.2 through 6.7. The validity of this assumption has been tested computationally and was found to be very accurate even for spectra with sharp peaks.

Acknowledgments

We thank C. Buneo, R. Sachdev, and F. F. Ebner for providing example data sets, C. Loader for help with the calculation of global error bars, and D. R. Brillinger and D. J. Thomson for comments that substantially improved the manuscript. M. J. is grateful to R. A. Andersen for both his continued support of theoretical work in his lab and his careful reading of the manuscript. M. J. acknowledges the generous support of the Sloan Foundation for Theoretical Neuroscience.

References

Abeles, M., Deribaupierre, F., & Deribaupierre, Y. (1983). Detection of single unit responces which are loosely time-locked to a stimulus. *IEEE Transactions on Systems Man and Cybernetics, 13*, 683–691.

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*. New York: Wiley.

Bartlett, M. S. (1966). *An introduction to stochastic processes*. Cambrigde: Cambridge University Press.

Brillinger, D. R. (1972). The spectral analysis of stationary interval functions. In *Proceedings of the Sixth Berkley Symposium on Mathematical Statistics and Probability* (1:483–513).

Brillinger, D. R. (1974). *Time series*. New York: Holt, Rinehart and Winston.

Brillinger, D. R. (1978). *Developments in statistics vol1*, chapter Comparative aspects of the study of ordinary time series and of point processes, pages 33–129. Orlando, FL: Academic Press.

Brody, C. D. (1998). Slow covariations in neuronal resting potentials can lead to artefactually fast cross-correlations in their spike trains. *Journal of Neurophysiology, 80*, 3345.

Cox, D. R., & Lewis, P. (1966). *The statistical analysis of series of events*. London: Chapman and Hall.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. London: Chapman and Hall.

Gerstein, G. L., Perkel, D. H., & Dayhoff, J. E. (1985). Cooperative firing activity in simultaneously recorded population of neurons—detection. *Journal of Neuroscience, 4*, 881–889.

Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual-cortex exhibit intercolumnar synchronization which reflects global stimulus properties. *Nature, 338*, 334–337.

Hannan, E. J. (1970). *Multiple time series*. New York: Wiley.

Loader, C. R. (1999). *Local regression and likelihood*. New York: Springer-Verlag.

Ma, S. K. (1985). *Statistical mechanics*. Singapore: World Scientific Publishing.

Mitra, P. P., & Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophysical Journal, 76*, 691–708.

Percival, D. B., & Walden, A. T. (1993). *Spectral analysis for physical applications*. Cambridge: Cambridge University Press.

Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., & Andersen, R. A. (2000). *Temporal structure in neuronal activity during working memory in macaque parietal cortex*. Unpublished manuscript.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes*. Cambridge: Cambridge University Press.

Rosenberg, J. R., Amjad, A. M., Breeze, P., Brillinger, D. R., & Halliday, D. M. (1989). The Fourier approach to the identification of functional coupling between neuronal spike trains. *Prog. Biophys. Molec. Biol., 53*, 1–31.

Sachdev, R., Jenkinson, E., Melzer, P., & Ebner, F. F. (1999). Dual electrode recordings from the awake rat barrel cortex. *Somatosensory and Motor Research, 16*, 163–192.

Snyder, L. H., Batista, A. P., & Andersen, R. A. (1997). Coding of intention in the posterior parietal cortex. *Nature, 386*, 167–169.

Sun, J., & Loader, C. R. (1994). Simultaneous confidence bands for linear regression and smoothing. *Annals of Statistics, 22*, 1328–1345.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of IEEE, 70*, 1055–1996.

Thomson, D. J. (2000). *Nonlinear and nonstationary signal processing*. Cambridge: Cambridge University Press.

Thomson, D. J., & Chave, A. D. (1991). *Advances in spectrum analysis and array processing*. Englewood Cliffs, NJ: Prentice Hall.

Received March 6, 2000; accepted August 26, 2000.