# Point Process Models of Single-Neuron Discharges

DON H. JOHNSON

*Computer and Information Technology Institute, Department of Electrical and Computer Engineering,*
*Rice University, MS #366, Houston, Texas 77005-1892*

dhj@rice.edu

Action Editor: C. Koch

**Abstract.** In most neural systems, neurons communicate via sequences of action potentials. Contemporary models assume that the action potentials' times of occurrence rather than their waveforms convey information. The mathematical tool for describing sequences of events occurring in time and/or space is the theory of point processes. Using this theory, we show that neural discharge patterns convey time-varying information intermingled with the neuron's response characteristics. We review the basic techniques for analyzing single-neuron discharge patterns and describe what they reveal about the underlying point process model. By applying information theory and estimation theory to point processes, we describe the fundamental limits on how well information can be represented by and extracted from neural discharges. We illustrate applying these results by considering recordings from the lower auditory pathway.

## 1. Introduction

The generic neuron expresses three anatomical features: the dendrites—those portions of the neuron that receive inputs, the soma—the central portion that contains the cell's nucleus and may also receive inputs, and the axon—a tendril that branches and serves as the input to other neurons. Inputs bring signals to a neuron, resulting in a complex symphony of ionic and transmembrane potential changes that can make the soma's excitable membrane produce the crescendo of an action potential. Ignoring much detail, the *action potential*—a pulse-like depolarization of the neuron's transmembrane potential—occurs when the potential exceeds a critical threshold and ionic concentrations achieve supercritical values. The biophysical equations governing action potential generation are the Hodgkin-Huxley equations. Because of the form of these equations, the action potential takes on "a life of its own" as ionic currents flow in concert without much regard to other influences. This action potential propagates, spreading throughout the neuron. If the membrane is active, the action potential is "repeated" because of a propagating depolarization. In particular, most axons are active and are thought to be the main conveyor of neural output to other neurons. If the membrane is passive, the depolarization pulse diffuses, decaying in amplitude and spreading in time. This "internal feedback" of action potentials into the passive dendrites can affect the cell's responsiveness to later inputs.

The action potential thus produced does not, in and of itself, represent information. Action potential waveforms may differ somewhat depending on when previous action potentials occurred and the neuron's general state, but what matters is the occurrence of an action potential, which results in transmitter or modulator release somewhere along the axon's extent. Neuroscientists interested in the *neural code*—what

information does a sequence of action potentials represent—therefore idealize the action potential as an isolated event that has occurred at some instant of time and that propagates to wherever the axon makes contact. This code reflects not only the neuron's input(s) and how the neuron processes them, but also the neuron's biophysical properties. Analyzing neural recordings amounts to teasing apart these components of response. Understanding this decomposition makes analysis of neural data much less ad hoc. Computational neuroscientists want to find models that can mimic neural behavior and thus seek to understand the biophysical mechanisms that underlie neural processing. In part, they seek information from the neural code that can help constrain their models. This article is concerned with the theoretical ramifications of this isolated-event idealization of neural output to the measurement and analysis of the neural code.

The mathematical tool that describes the occurrence of isolated events in time and/or space is *point process theory* (Snyder, 1975). From its application within neuroscience that dates from the 1960s to now, this formalism has grown more powerful. For useful surveys see (Fienberg, 1974; Holden, 1976; Sampath and Srinivasan, 1977). Early applications focused on single sequences of events occurring in time having a very simple structure (Gerstein and Kiang, 1960; Moore et al., 1966; Rodieck et al., 1962). Superposition and deletion models attempted to describe convergence of multiple inputs on a single neuron (Cox, 1962; Cox and Lewis, 1966; Ten Hoopen and Reuver, 1965; Lawrence, 1970; Linebarger and Johnson, 1986). Multiple-event sequences and their possible statistical dependence structures can be analyzed using the theory of *marked* point processes (Snyder, 1975). We concentrate here on techniques for analyzing individual neural output.

## 2. Point Process Theory

### 2.1. Definition of a Point Process

A *regular point process* (Snyder, 1975) is defined so that the probability of an event occurring in the time interval $[t, t + \Delta t)$, is given by

$$\Pr[\text{one event in } [t, t + \Delta t) \mid N_t, \mathbf{w}_t]$$
$$= \mu(t; N_t, \mathbf{w}_t)\Delta t$$
$$\Pr[\text{more than one event in } [t, t + \Delta t) \mid N_t, \mathbf{w}_t]$$
$$= o(t, \Delta t) \tag{1}$$

Here

- $N_t$ is the number of events that have occurred prior to time $t$ (observations are assumed to start at time $t = 0$);
- $\mathbf{w}_t$ is the vector of occurrence times of these $N_t$ events: $\mathbf{w}_t = [w_1, \ldots, w_{N_t}]$; and
- $o(t, \Delta t)$ decreases to zero with decreasing $\Delta t$ faster than linearly:
  $\lim_{\Delta t \to 0} o(t, \Delta t)/\Delta t = 0$.

These equations mean that no more than one event can occur in a sufficiently small interval and that the probability of one event occurring within a small interval is proportional to the interval's duration. The quantities $N_t$ and $\mathbf{w}_t$ describe the *history* of the process, detailing the number and the times at which all events occurred prior to time $t$. These quantities are depicted in Fig. 1. Note that the probabilities are *conditional* probabilities: they depend on the point process's history. This description has been found to be the most convenient form for describing nonstationary, dependent point processes required for realistic models of neural discharge patterns.

The conditional probability of an event occurring during a small time interval equals the product of the interval duration $\Delta t$ and a nonnegative quantity, $\mu(t; N_t, \mathbf{w}_t)$, known as the *intensity* of the point process. The intensity has units of events per second and can be considered the *instantaneous rate* at which events occur. The intensity can depend on time and/or the point process's history. If the intensity depends explicitly on time $t$, the process is nonstationary. Such temporal variations usually represent rate variations that are due to external influences, which have been modified—processed—by the neuron. These external influences can be deterministic, depending directly on a stimulus for example, or could be stochastic, mimicking variations beyond experimental control. The "historical" portion of the intensity describes the detailed interactions of previous events on the current one. In a neuroscientific context, this history could, among other possibilities, represent the neuron's dynamic active-membrane properties and its sensitivity to modulators and transient ionic concentrations, all of which certainly contribute to the neuron's processing function. For example, the discharge rate typically decreases to zero momentarily after an action potential occurs. This refractory effect has long been noted; in the present context, this phenomenon means that the intensity for virtually all accurate point process
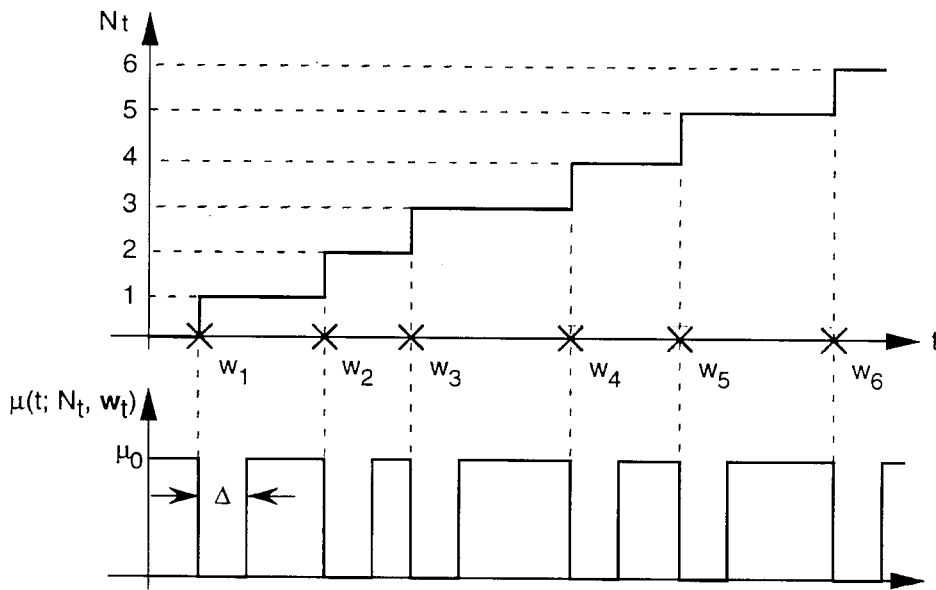
*Figure 1.* The essential quantities that define a point process are noted along side a sample function. $w_n$ is the occurrence time of the $n$th event and $\tau_n = w_n - w_{n-1}$ is the interevent interval. $N_t$ is known as the counting process. The lower trace depicts the corresponding intensity of this process. This intensity depends on immediate process history, with the intensity dropping to zero for a time interval equaling $\Delta$ after each event. This type of intensity means that event occurrence is suppressed for a period of time after each event. This model could be used to model absolute refractory effects. The analytic form of the intensity is given by Eq. (14).

models of neural discharges must contain a historical component.

*Once we have the intensity of a point process, the process's entire statistical structure is specified.* Consequently, a thorough analysis of a discharge pattern must yield the process's intensity function, and this intensity captures the complete structure of the process. In technical terms, deciphering the neural code amounts to measuring the intensity. However, the intensity's temporal variations and dependence on event history can share common biophysical underpinnings; this mathematically neat separation of the intensity into two components does not necessarily translate into separate aspects of neural function. This fact, demonstrated explicitly later, has bad and good consequences. By measuring and modeling the intensity we gain an understanding of the neural code. Even when the rate is constant, a seemingly boring situation from an information processing viewpoint, careful analysis of the intensity's historical component can yield important insights into the neuron's processing function. Unfortunately for the scientist, the intensity is usually related to neural mechanisms in a complicated, not easily deciphered, way.

### 2.2. Examples of Point Processes

*The Poisson Process.* The Poisson process has the simplest structure of any point process. Its intensity does not depend on history: $\mu(t; N_t, \mathbf{w}_t) = \lambda(t)$. The quantity $\lambda(t)$ could be constant, meaning that events occur randomly at a fixed average rate (the process is *stationary*), or it could vary with time, meaning that occurrence rate varies accordingly, or it could vary stochastically, resulting in what is known as a *doubly stochastic Poisson process*. No matter which of these situations applies, event occurrence does not depend statistically on when previous events occur: The Poisson process has no memory. Since a neuron's dynamic membrane properties would most certainly introduce memory into its output, the Poisson process can only approximate neural output.

In point process theory, the Poisson process is akin to white Gaussian noise in random process theory: analytic results about the Poisson process's structure are easy to derive and virtually all other point processes present analytic difficulties. For example, suppose we want to derive the probability that $n$ events occur between the times $t$ and $t + T$, $t, T > 0$. Using the

notation of point process theory, we want an expression for $\Pr[N_{t+T} - N_t = n]$. For small $\Delta T$ with $t$ held fixed,

$$\Pr[N_{t+T+\Delta T} - N_t = n]$$
$$= \Pr[N_{t+T} - N_t = n] \cdot \Pr[N_{t+T+\Delta T} - N_{t+T} = 0]$$
$$+ \Pr[N_{t+T} - N_t = n - 1] \cdot \Pr[N_{t+T+\Delta T} - N_{t+T} = 1]$$
$$+ \Pr[N_{t+T} - N_t = n - 2] \cdot \Pr[N_{t+T+\Delta T} - N_{t+T} = 2]$$
$$+ \cdots$$

The sum arises because, in a Poisson process, the occurrence of events in disjoint time intervals are statistically independent (i.e., event history has no effect on event occurrence). Because no more than one event can occur in a sufficiently small interval, the expression for our probability becomes

$$\Pr[N_{t+T+\Delta T} - N_t = n]$$
$$= \Pr[N_{t+T} - N_t = n] \cdot (1 - \lambda(t+T)\Delta T$$
$$- o(t+T, \Delta T))$$
$$+ \Pr[N_{t+T} - N_t = n - 1] \cdot \lambda(t+T)\Delta T.$$

Rearranging and taking the limit as $\Delta T \to 0$ yields the differential-difference equation

$$\frac{\partial \Pr[N_{t+T} - N_t = n]}{\partial T}$$
$$= -\lambda(t+T)(\Pr[N_{t+T} - N_t = n]$$
$$- \Pr[N_{t+T} - N_t = n - 1]).$$

Solving this equation is easy but tedious; direct substitution shows that the solution to this equation is

$$\Pr[N_{t+T} - N_t = n]$$
$$= \frac{1}{n!} \left( \int_t^{t+T} \lambda(\alpha)\, d\alpha \right)^n \exp\left\{ -\int_t^{t+T} \lambda(\alpha)\, d\alpha \right\},$$
$$n = 0, \ldots \tag{2}$$

We can use this result to determine the process's rate of discharge. We define the *rate of discharge* $\lambda(t)$ to be the limit of the expected number of events in an interval that begins at time $t$ divided by the interval's duration:

$$\lambda(t) = \lim_{\Delta t \to 0+} \frac{\mathcal{E}[N_{t+\Delta t} - N_t]}{\Delta t}.$$

For the Poisson point process, the required expected value equals $\int_t^{t+\Delta t} \lambda(\alpha)\, d\alpha$, and the instantaneous rate equals the intensity as expected. We shall find that the Poisson point process is *unique* in having this property.

We can also derive from Eq. (2) the probability density function (pdf) of the interevent interval $\tau_n$. The probability that no events occur within the next $T$ seconds after the event time $w_{N_t}$ equals the probability that the interevent interval $\tau_{N_t+1}$ exceeds $T$: $\Pr[N_{t+T} - N_t = 0] = \Pr[\tau_{N_t+1} > T]$. Let $p_{\tau_{N_t+1}}(\tau)$ denote the pdf of the $(N_t + 1)^{\text{st}}$ interevent interval:

$$\int_T^\infty p_{\tau_{N_t+1}}(\alpha)\, d\alpha = \exp\left\{ -\int_{w_{N_t}}^{w_{N_t}+T} \lambda(\beta)\, d\beta \right\}.$$

Calculating the derivative with respect to $T$ of this expression yields the interevent interval pdf:

$$p_{\tau_{N_t+1}}(\tau; t) = \lambda(w_{N_t} + \tau) \exp\left\{ -\int_{w_{N_t}}^{w_{N_t}+\tau} \lambda(\beta)\, d\beta \right\}. \tag{3}$$

When the Poisson process is stationary, with $\lambda(t) = \lambda_0$, we have the well-known result that the interevent intervals in a Poisson process are exponentially distributed: $p_\tau(\tau) = \lambda_0 \exp\{-\lambda_0 \tau\}$, $\tau > 0$.

Despite its limitations to describe single-neuron data, the Poisson process provides a starting point for understanding a given point process's properties. The rule of thumb is that if you can't analyze some aspect of a Poisson process's structure, related analysis of a more exact point process description may be difficult, if not impossible.

Note that classical definition of instantaneous rate as the reciprocal of the interevent interval (Adrian, 1928) makes no sense here. For a stationary Poisson process, the reciprocal of the average interval does equal the rate: $1/\mathcal{E}[\tau] = \lambda_0$. However, the expected value of the reciprocal interval $\mathcal{E}[1/\tau]$ does *not* equal this value; in fact, this expected value is infinite. Even more difficulties arise when the process is nonstationary. Calculating the reciprocal interval captures the statistics of the instantaneous rate's *integral* as well as its value (see Eq. (3)). Furthermore, the temporal variations of the rate can be more rapid than its average, which means that event occurrence grossly misrepresents actual rate variations.

*Renewal Processes.*    A *renewal* process results when the intensity's historical dependence amounts to the time at which the last event preceding time $t$ occurred (Cox, 1962): $\mu(t; N_t, \mathbf{w}_t) = \mu(t; N_t, w_{N_t})$. Thus, interevent intervals are statistically *independent* random variables, with the probability of an event

occurring varying with time since the last event. When the intensity does not depend on absolute time, the intensity is a function only of the interval $t - w_{N_t}$ since the last event[1]:

$$\mu(t; N_t, \mathbf{w}_t) = h(t - w_{N_t}).\qquad(4)$$

If the function $h(\cdot)$ is a constant (no dependence on the interval since the last event), a Poisson process results. If this function equals zero for small arguments, then jumps to some nonzero value, $h(\tau) = \mu_0 u(\tau - \Delta)$, we can describe absolute refractory effects with $\Delta$ identified as the refractory interval and $u(\cdot)$ is the unit-step function[2]. After the refractory interval has ended, the constant-valued intensity means that the process has entered a Poisson-like mode, with event occurrence oblivious to event history. The lower panel of Fig. 1 depicts the intensity for this type of renewal process.

This figure also clarifies the somewhat confusing point that the event occurrence rate for this point process does *not* equal $\mu_0$. Because during portions of time the intensity equals zero, the rate of event occurrence must be less that $\mu_0$. To compute the average rate, we can use the implicit formula that the average rate $\bar\lambda$ should equal the long-term average of the intensity. Considering Fig. 1, the area under the intensity equals $\mu_0 T(1 - \bar\lambda\Delta)$. Dividing this equation by $T$ and setting the result equal to $\bar\lambda$ yield $\bar\lambda = \mu_0/(1 + \mu_0\Delta)$. This example demonstrates that the intensity function indicates the *true* rate of discharge immediately following an event while the average rate indicates the overall rate at which events occur. Another conclusion can be drawn: gross measures of event rate—ones that do not take event history into account—do not indicate the true rate. However, we can, in some circumstances, determine underlying quantities from measured ones. Assuming we knew the value of $\Delta$, we could convert the measured average rate into the rate at which events occur once the refractory interval is over. Thus, two points emerge: measures that average over event histories do not truly reflect point process parameters or structure, but such averages can serve as the basis for estimating the intensity.

The intensity for the renewal process, and for other examples as well, is more conveniently expressed in terms of the interevent intervals than in terms of event occurrence times. The sequence of interevent intervals $\tau_n = w_n - w_{n-1}$ (with $w_0 \equiv 0$) is completely equivalent to the sequence of event times: one can be easily calculated from the other. We express a point process's intensity as $\tilde\mu(t; N_t, \tau)$, with $t = \sum_{n=1}^{N_t+1} \tau_n$ and

$\tau_{N_t+1} = t - w_{N_t}$ (the "as yet uncompleted" interval between the last recorded event and the current time):

$$\mu(t; N_t, \mathbf{w}_t) \leftrightarrow \tilde\mu(t; N_t, \tau).$$

Here $\tau$ equals $[\tau_1, \ldots, \tau_{N_t+1}]$, the vector of all previous interevent intervals. The interval-based expression for the intensity of a (possibly nonstationary) renewal process takes the form $\tilde\mu(t; N_t, \tau) = f(t, \tau_{N_t+1})$; if stationary, we have $\tilde\mu(t; N_t, \tau) = h(\tau_{N_t+1})$ (for $N_t \geq 1$; see note 1). While re-expressing a point process's intensity in terms of intervals seems merely to be a notational nicety at this point, we shall find subsequently that it has great utility.

Every stationary renewal process is completely characterized by the probability density function $p_\tau(\cdot)$ of the interval between successive events. This fact follows because the procedure used to derive Eq. (3) can be used on renewal processes. This pdf is related to the process's intensity through the function $h(\cdot)$ defined in Eq. (4) as

$$\begin{aligned} h(\tau) &= \frac{p_\tau(\tau)}{\int_\tau^\infty p_\tau(\alpha)\,d\alpha} \\ p_\tau(\tau) &= h(\tau)\exp\left\{-\int_0^\tau h(\alpha)\,d\alpha\right\} \end{aligned}.\qquad(5)$$

The quantity $h(\cdot)$ is known as the *hazard function* or the *recovery function* associated with the point process. Clearly, this quantity is just the renewal process's intensity. Our example has $h(\tau) = \mu_0 u(\tau - \Delta)$, which corresponds to the interval pdf $p_\tau(\tau) = \mu_0 \exp\{-\mu_0 \cdot (\tau - \Delta)\} u(\tau - \Delta)$.

Calculating the hazard function frequently provides more insight into the process's structure than does the interval distribution. The hazard function $h(\tau)$ is proportional to the probability of an interval equaling $\tau$ given that the interval is at least that long; this interpretation follows directly from Eq. (5). A constant-valued hazard function means that the probability of an event occurring is as likely now as it is later: The process thus has no memory over the range of intervals for which the hazard function, the intensity, is constant. If the hazard function is an increasing function, then the process becomes more "anxious" to produce an event as the interval becomes longer. For example, if the hazard function increases linearly, $h(\tau) = K \cdot \tau$, $K$ a constant, then $p_\tau(\tau) = K\tau \exp\{-K\tau^2/2\}$, a much narrower distribution of intervals than in the Poisson (constant hazard function $\leftrightarrow$ exponential interval pdf)

case. A decreasing hazard function has the opposite interpretation. However, a point process's intensity cannot decrease to zero as $\tau$ approaches infinity; if it did, events could cease to occur, and the process would not be self-consistent[3].

The general form of the hazard function translates directly into the point process's *regularity*[4]: how variable are durations of interevent intervals. We can quantify the regularity of a stationary point process by calculating the interevent interval's *coefficient of variation* $C[\tau]$:

$$C[\tau] = \frac{\sigma[\tau]}{\mathcal{E}[\tau]}. \tag{6}$$

Here, $\sigma[\tau]$ denotes the standard deviation of the interevent interval $\tau$ and $\mathcal{E}[\tau]$ the interval's expected (average) value:

$$\mathcal{E}[\tau] = \int_0^\infty \alpha p_\tau(\alpha)\, d\alpha$$
$$\sigma[\tau] = (\mathcal{E}[\tau^2] - (\mathcal{E}[\tau])^2)^{1/2}.$$

This quantity normalizes the interevent interval variability by the average interval duration. The larger the coefficient of variation, the more irregular the point process. The minimum value of $C[\tau]$ is zero, which occurs when the discharge pattern is periodic (all intervals are equal-length). For a Poisson process, the expected value $\mathcal{E}[\tau]$ equals $1/\lambda_0$ and the standard deviation equals the same value. Thus, its coefficient of variation equals one. Because the Poisson process has no memory, it is, in some sense, a very irregular point process, with events occurring randomly with respect to all others. While it is theoretically possible for the coefficient of variation to exceed unity[5], $C[\tau]$ values near one are usually taken as an indication of irregularity. For the situation where the hazard function increases linearly, $h(\tau) = K\tau$, $\mathcal{E}[\tau] = \sqrt{\pi/2K}$ and $\sigma[\tau] = \sqrt{(2 - \pi/2)/K}$, yielding a coefficient of variation of $\sqrt{4/\pi - 1} = 0.522$. Thus, this point process exhibits moderate irregularity. This, the Poisson, and the periodic examples shown in Fig. 2 demonstrate that point process regularity does indeed correspond to the coefficient of variation and that regularity can be easily visualized. When we include a refractory interval into this model, the coefficient of variation decreases, resulting in a more regular discharge pattern. For example, when the interval density equals an exponential shifted by the refractory interval $\Delta$, the standard deviation of the interevent interval equals the Poisson value, but the
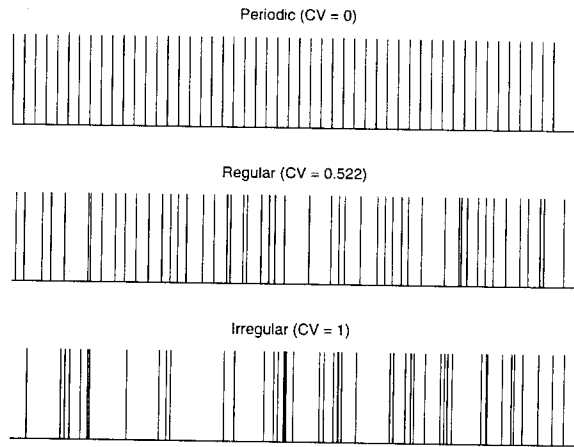


*Figure 2.* The panels depict portions of point processes having the same average rate (equal to one), but with different coefficients of variation. The lower panel shows a Poisson process, the middle has a hazard function equal to $K\tau$ with $K = \pi/2$ (to achieve an average rate of one), and the upper is periodic. The coefficient of variation for each case are noted.

expected interval equals $\Delta + 1/\lambda_0$. In this case, $C[\tau] = 1/(1 + \lambda_0\Delta)$. Thus, an approximately constant hazard function generally corresponds to a irregular point process but becomes increasingly regular as $\lambda_0\Delta$ increases. Generally, increasing hazard functions correspond to more regular discharge patterns than produced by those having approximately constant hazard functions.

*Markov Point Processes.* Events in nonrenewal processes depend not only on when the previous event occurred but also on ones before that. In *Markov* point processes, event occurrence depends on the most recent $k + 1$ events and not on ones prior to those. Expressed in terms of intervals, when an event occurs depends statistically on the values of the most recent $k$ interevent intervals:

$$\bar{\mu}(t; N_t, \tau) = \bar{\mu}\big(t; N_t, [\tau_{N_t+1}, \tau_{N_t}, \ldots, \tau_{N_t-(k-1)}]\big).$$

The number $k$ is known as the *order* of the Markovian dependence structure. A renewal process has order zero: event occurrence depends only on the most recent event. Higher-order dependence means that successive interevent intervals are correlated random variables, which complicates development of concise descriptions of data.

If the spike production mechanism is described "classically"—a synaptic event influences the transmembrane potential of active membrane modeled by

Hodgkin-Huxley $Na^+$ and $K^+$ channels—the equations governing spike production have no state variable that persists beyond the occurrence of an action potential. Thus, spike occurrence does depend on when the previous spike occurs but not on any others, which means that a renewal process model *always* results. When Markovian point processes are needed to describe a spike train, a more complicated channel model will be required, and some persistent variable (or variables) must be part of active membrane's dynamics.

*The Hawkes Process.* This example demonstrates that point processes need not be Markovian. Here, the intensity depends on the output of a linear filter having an input equal to impulses occurring at event times (Hawkes, 1971):

$$\mu(t; N_t, \mathbf{w}_t) = \lambda_0 + \int_0^t g(t - \alpha) \, dN_\alpha.$$

The Stieltjes representation of this convolution integral is needed for technical reasons; it simply equals the summed impulse responses delayed by all event times occurring prior to time $t$:

$$\mu(t; N_t, \mathbf{w}_t) = \lambda_0 + \sum_{i=1}^{N_t} g(t - w_i).$$

Thus, the intensity depends on *all* past event times viewed through the window described by $g(\cdot)$. This model could be used to describe situations in which event occurrence depends on some quantity that varies with the times at which all previous events occur ($g(\cdot)$ has infinite duration) or that varies with events occurring within the most recent $T$ seconds ($g(\cdot)$ is nonzero only over $[0, T]$).

An arbitrary impulse response $g(\cdot)$ cannot be used in this expression. For instance, intensities are always positive quantities, meaning that the summed impulse responses cannot be more negative than $-\lambda_0$. Further restrictions on the impulse response result if we demand the Hawkes process be stationary. Defining $\bar{\lambda}$ as the average occurrence rate, the intensity must satisfy

$$\bar{\lambda} = \lambda_0 + \bar{\lambda} \int_{-\infty}^t g(t - \alpha) \, d\alpha \Rightarrow$$

$$\bar{\lambda} = \frac{\lambda_0}{\left(1 - \int_0^\infty g(\alpha) \, d\alpha\right)}.$$

This result means that the filter's impulse response must satisfy $\int_0^\infty g(\alpha) \, d\alpha < 1$ for the Hawkes process to exist.

Hawkes process has aspects appealing to neuroscience applications. For example, intracellular concentration at any time depends on the time-pattern of all previous event occurrences. The effect of concentration on discharge rate might be well approximated by the integration expressed by the Hawkes process that uses an infinite-duration $g(\cdot)$. In addition, this process could model postevent facilitation or inhibition that persists for a finite time rather than for a given number of events. In this case, a finite-duration $g(\cdot)$ would be used. To employ it means that we must be able to measure the impulse response $g(\cdot)$, an estimation problem that has not been solved to date. If $g(t)$ has infinite support but decreases rapidly, equaling $e^{-t}u(t)$ for example, the properties of the Hawkes process might be well described by a Markovian model: only the most recent $k$ events would approximate well the infinite dependence contained in the Hawkes process.

### 2.3.   Generating Point Processes

Interestingly, the natural way to generate a point process, whether stationary or nonstationary, is to compute the sequence of interevent *intervals* rather than the times of occurrence. The less desirable technique, based on Eq. (1), is to define an interval $\Delta t$ and flip a biased coin as each such interval passes. The probability of a spike (the coin comes up heads) would equal the product of the intensity and the interval duration. Two problems with this approach are evident. First of all, only the "tail probability" of the uniform distribution would be considered, which is the most difficult aspect to check of any random number generator. For this approach to work, the maximum probability of a discharge must be significantly less than one, meaning that most of the coin flips have a very low probability of coming up heads. Checking that a uniform random number generator accurately produces small probabilities of, say, $10^{-6}$ can be challenging. A second problem is that *many* random numbers would be needed to produce each event. Thus, much computation (random number generation) is spent *not* creating events.

The more elegant, computationally efficient approach rests on the conditional pdf of an interval given the lengths of the previous intervals (Ozaki, 1979). Simple manipulation of the point process's defining equations yields a closed form expression for its

conditional interval distribution:

$$p_{\tau_{n+1}|\tau_n,\ldots,\tau_1}(\tau_{n+1} \mid \tau_n, \ldots, \tau_1)$$
$$= \tilde{\mu}(\tau_{n+1}; n, \tau)\exp\left\{-\int_0^{\tau_{n+1}} \tilde{\mu}(\alpha; n, \tau)\,d\alpha\right\}.$$

To generate the point process, use the inverse distribution function technique (Devroye, 1986). The key to this technique is that the random variable $U = P_X(X)$, the probability *distribution* function of a random variable applied to the random variable itself, is a uniformly distributed (over (0, 1]) random variable. Consequently, to generate $X$, generate the uniform random variable $U$ and evaluate the inverse distribution function $X = P_X^{-1}(U)$. This technique can also be used for conditional distribution function: $P_{X|Y}(X \mid Y = y)$ is also uniformly distributed for each conditioning value. To generate a random variable having this conditional distribution, we use $X = P_{X|Y}^{-1}(U \mid Y = y)$. Applying this technique to the problem at hand, we first note that the desired conditional distribution function equals

$$P_{\tau_{n+1}|\tau_n,\ldots,\tau_1}(\tau_{n+1} \mid \tau_n, \ldots, \tau_1)$$
$$= 1 - \exp\left\{-\int_0^{\tau_{n+1}} \tilde{\mu}(\alpha; n, \tau)\,d\alpha\right\}.$$

Thus, to generate the sequence of interevent intervals, we use

$$E_{n+1} = \int_0^{\tau_{n+1}} \tilde{\mu}(\alpha; n, \tau)\,d\alpha, \qquad (7)$$

where $E_{n+1}$ is an exponentially distributed random variable (unit parameter)[6]. Denoting the integral by $k(\tau_{n+1}, \tau)$, the input-output relation of the system that generates interevent intervals from a sequence of statistically independent, exponentially distributed random variables is $\tau_{n+1} = k^{-1}(E_{n+1}, \tau)$, and it depends on process history and on temporal variations as shown in Fig. 3. Thus, to generate interevent intervals, first generate a sequence of exponentially distributed random variables $\{E_n\}$, *one for each interevent interval*. With these serving as the input, the integral of Eq. (7) is computed for successive values of $\tau_n$ until it equals the input value. This computation is relatively easy to perform and, when compared to the "coin-flipping" method, trades computation (numerical evaluation of the integral) for random variable generation (flipping coins). In my experience, more accurate and
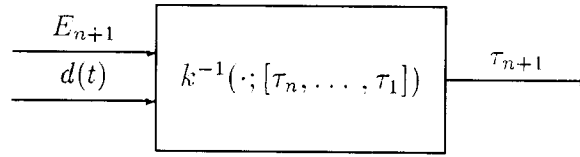


*Figure 3.* The interval $\tau_{n+1}$ between events $w_{n+1}$ and $w_n$ result from the diagrammed system, where $E_{n+1}$, a sequence of statistically independent random variables, and $d(t)$, representing a time-varying drive, serve as the inputs.

faster simulations result from the computation-based approach.

In some cases, the integral can be evaluated analytically, resulting in an explicit expression for each interevent interval. Consider a stationary renewal process whose intensity is given by $\tilde{\mu}(\tau_n) = \lambda_0 u(\tau_n - \Delta)$. The required integral of the intensity is

$$E_{n+1} = \begin{cases} 0, & \tau_{n+1} < \Delta \\ \lambda_0 \cdot (\tau_{n+1} - \Delta), & \tau_{n+1} \geq \Delta \end{cases}.$$

Consequently, the interevent interval is simply given by $\tau_{n+1} = E_{n+1}/\lambda_0 + \Delta$. The relation of this process to a Poisson process is thus made explicit: it is created by extending Poisson intervals by the refractory interval duration.

Periodic point processes can be generated by a related technique. Instead of letting the left side of Eq. (7) equal a random variable, set it equal to one. Now an event occurs whenever the integral of the intensity equals this value, and a periodic sequence of events occurs when the intensity equals a constant. By using more general intensities that are time varying and have serial dependencies, a whole class of deterministic point processes can be generated. Furthermore, the rate of such point processes is now well defined: it equals the intensity.

The fundamental point process generation algorithm expressed by Eq. (7) has an interesting interpretation: this exact result essentially represents an "integrate-and-fire" description for discharge pattern generation. Here, the quantity integrated is *not* a membrane potential but the intensity of the point process, which contains all of the previous-event dependencies and temporal variations contained in the process. Furthermore, the "firing threshold" is not fixed but instead is random, equaling an exponential random variable. We can construe our result here to be the accurate integrate-and-fire model, whereas the usual one only approximates Hodgkin-Huxley-like mechanisms.

Furthermore, the accurate one *always* applies, whether or not the point process arose from a neural discharge pattern.

## 3. Information Representation by a Point Process

In sensory systems, especially, and in neural information processing systems more generally, we are ultimately concerned with how a neuron's discharge pattern represents information and what fidelity that representation provides. From the point process viewpoint, information-bearing signals or indicators are expressed by the intensity's temporal variations. Conceptually, the intensity consists of two components: an *intrinsic* component determined by spike generation mechanisms and an *extrinsic* component that represents a signal. The intrinsic component includes the absolute refractory interval and interspike interval dependencies that reflect how one discharge influences the timing of another. In modeling terms, the intrinsic component determines the point process model: whether it is Poisson, renewal, or Markov, for example. This component does not convey information because it does not directly depend on external sources. The intensity's extrinsic component conveys information; this component explicitly depends on a signal (or several signals). Temporal rate variations, stimulus responses in the context of sensorineural systems, arise because of temporal variations in the extrinsic component. In most cases, the extrinsic component and the external environment are related deterministically and invertably (no information loss in the transformation from stimulus to rate). The inherent stochastic nature of the point process and the intensity's intrinsic component place limits on the "information expressiveness" of a discharge pattern. In this section, we review measures of this expressiveness and their fundamental properties.

Two measures can be used to assess the expressiveness: information theoretic measures (capacity and cutoff rate) and mean-squared estimation error. The first is broad and attempts to define the ultimate information representation ability of a discharge pattern's extrinsic component (Cover and Thomas, 1991). Investigators have attempted to measure the capacity of a single neuron (Rieke et al., 1993). The second measure is more specific and characterizes how well an optimal signal processing system can extract temporal extrinsic intensity variations beyond those induced by spike generation mechanisms (Bialek et al., 1991;

Johnson and Swami, 1983; Mark and Miller, 1992; Miller, 1985). The theoretical understanding of both measures in the point process case is immature, with the Poisson process being the only one for which broad results exist. In the Poisson case, the intrinsic discharge properties are the simplest, but because of refractory effects this model can only approximate neural discharge patterns.

To employ either measure, we model a point process's intensity as a stochastic process. The intensity is (usually) nonlinearly related to the signal that the point process represents. For example, in the auditory system, auditory-nerve fiber discharges are related to a bandpass-filtered version of an acoustic signal. In auditory system modeling, we should therefore use a bandpass model for the input's power spectrum. Mathematically, let $X_t$ be a stationary process that describes extrinsic information; the intensity of the point process is related to $X_t$ in a causal way: $\mu(t; N_t, \mathbf{w}_t) = F[X_s, s \leq t; N_t, \mathbf{w}_t]$. We must assume that the functional $F[\cdot]$ is known, with the notation meant to express the presence of extrinsic and intrinsic components. The intensity's dependence at time $t$ on the input's past expresses the possible presence of filtering (linear or nonlinear) of the signal.

### 3.1. Information Capacity

The information capacity (or simply the capacity) of a point process is the maximal mutual information between the intensity $\mu(t; N_t, \mathbf{w}_t)$ and the sequence of events expressed by the counting process $N_t$ over the interval $[0, T]$, with the maximum obtained by searching over all possible intensities that meet certain constraints (Brémaud, 1981, pp. 180–186). Typical constraints are a maximal rate, a minimal rate, an average rate, and a bandwidth limit. Letting $\mathcal{L}$ represent constraint set, the capacity is abstractly expressed by

$$C = \lim_{T \to \infty} \max_{X_t} \max_{\mu(t; N_t, \mathbf{w}_t) \in \mathcal{L}} \frac{1}{T} I_T[\mu(t; N_t, \mathbf{w}_t); N_t].$$
(8)

The details of this calculation are well beyond the scope of this article. Let it suffice that the capacity is known in interesting cases of point processes and constraint sets. When we constrain the intensity's maximum, minimum, and average values—$\lambda_{\min} \leq \mu(t; N_t, \mathbf{w}_t) \leq \lambda_{\max}$ and $\mathcal{E}[\mu(t; N_t, \mathbf{w}_t)] = \bar{\lambda}$—the capacity of *any* Markov point process corresponds to that of a Poisson process

and is given by Davis (1980) and Kabanov (1978)

$$
C = \begin{cases}
\lambda_{min} \left[ \dfrac{1}{e} \left( \dfrac{\lambda_{max}}{\lambda_{min}} \right)^{\lambda_{max}/(\lambda_{max}-\lambda_{min})} \right. \\
\qquad \left. - \ln \left( \dfrac{\lambda_{max}}{\lambda_{min}} \right)^{\lambda_{max}/(\lambda_{max}-\lambda_{min})} \right], \\
\qquad\qquad\qquad\qquad \bar{\lambda}_{max} - \lambda_{min} > \lambda^\circ \\
(\bar{\lambda} - \lambda_{min}) \ln \left( \dfrac{\lambda_{max} - \lambda_{min}}{\bar{\lambda} - \lambda_{min}} \right), \\
\qquad\qquad\qquad\qquad \bar{\lambda} - \lambda_{min} < \lambda^\circ,
\end{cases}
$$

where capacity has units of nats/s (multiply by $\log_2 e$ to obtain bits/s) and

$$
\lambda^\circ = \lambda_{min} \left[ \dfrac{1}{e} \left( \dfrac{\lambda_{max}}{\lambda_{min}} \right)^{\lambda_{max}/(\lambda_{max}-\lambda_{min})} \right].
$$

The formal proof does not suggest whether the Poisson process *uniquely* achieves this capacity; it does state that Markov point processes cannot have better capacities than the Poisson. Conceptually, additional constraints on the intrinsic component—that model refractory effects, for example—could be incorporated in the capacity calculation. The rate constraints only generally confine the extrinsic component: the encoded signal maximizing capacity is a random telegraph wave, a stochastic square wave that switches between maximal and minimal rates. This intensity has infinite bandwidth, with the average holding times at extremal rates tending to zero but with their ratio equaling a specific constant ($e$).

Note that $\lambda_{min}$ and $\lambda_{max}$ proscribe the stationary intensity's *instantaneous* extremal values, not its minimal and maximal average rates nor its peak rates achieved by transient stimuli. The minimum rate represents a background uninvolved in representing information. The larger this value relative to the maximum rate, the smaller the capacity. In auditory-nerve fibers, the minimum is zero (rates less than spontaneous values can be obtained) and the maximum about 700 spikes/s (Johnson, 1980). When $\lambda_{min} = 0$, the capacity expressions simplify considerably:

$$
C = \begin{cases}
\dfrac{\lambda_{max}}{e}, & \bar{\lambda} > \lambda^\circ \\
\bar{\lambda} \ln \left( \dfrac{\lambda_{max}}{\bar{\lambda}} \right), & \bar{\lambda} < \lambda^\circ
\end{cases}
$$

and $\lambda^\circ = \lambda_{max}/e = 0.37\lambda_{max}$. Thus, for large average rates, a point process's capacity is determined by its maximal rate; for averages rates less than $\lambda_{max}/e$, the capacity decreases from this maximum as average rate decreases (see Fig. 4).

The definition of capacity (Eq. (8)) does not directly express a communications model and does not necessarily provide an accurate measure of how well information is expressed by a point process. An alternate measure, the cutoff rate $R_0$, does. The conceptual communications model underlying its calculation is that one of a set of $q$ discrete symbols is transmitted every $T$ seconds (the symbol interval) using a symbol-specific intensity waveform. Computing the cutoff rate amounts to finding the optimal waveform set that conforms to a set of constraints while ensuring reliable
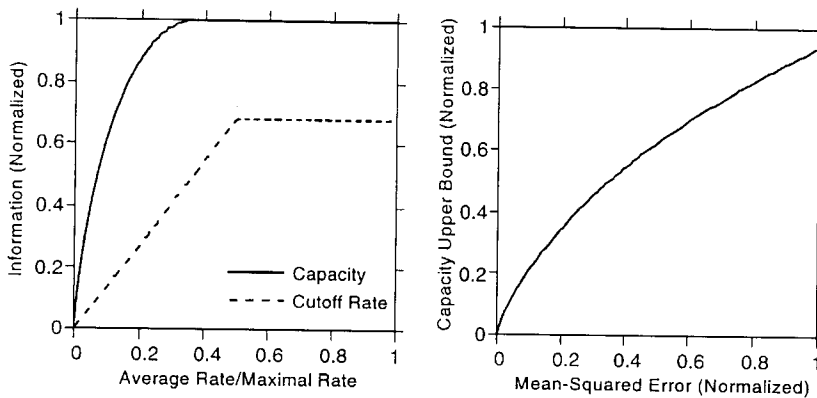


*Figure 4.* The relation between information measures and rate or mean-squared error are depicted. In each case, the minimal rate $\lambda_{min}$ equals zero, and the vertical scales have been normalized by $\lambda_{max}/e$, the maximal capacity of a point process. In the left panel, capacity and cutoff rate for Poisson processes are plotted against $\bar{\lambda}/\lambda_{max}$. The relation between a normalized upper bound on capacity and the mean-squared error (normalized by its maximum) is shown in the right panel. In addition to $\lambda_{min} = 0$, $\bar{\lambda} = \lambda_{max}/2$ for this calculation, and the intensity was that of a random telegraph wave. This latter plot indicates that capacity increases even as the mean-squared error becomes worse.

communication (the represented symbol can be determined without incurring rampant errors). The cutoff rate is smaller than the capacity[7], and it provides a more accurate view of the information bearing capability than the capacity. The cutoff rate for the Poisson process is given by Davis (1980) and Snyder and Rhodes (1980)

$$R_0 = -\ln\left(\frac{1}{q} + \frac{q-1}{q}e^{-d^2/2}\right),$$

where $d^2$ is a measure of the minimum pairwise distance between waveforms used to represent the $q$ symbols. The expressions for this measure are quite complicated, but if we restrict ourselves to the case of vanishingly small symbol intervals and an infinite number of symbols (to approximate the transmission of a continuous-valued waveform in continuous time), the cutoff rate becomes

$$R_0 = \begin{cases} \frac{1}{2}(\bar{\lambda} - \lambda_{\min})\left(\dfrac{\lambda_{\max}^{1/2} - \lambda_{\min}^{1/2}}{\lambda_{\max}^{1/2} + \lambda_{\min}^{1/2}}\right), \\ \qquad 0 \le \bar{\lambda} - \lambda_{\min} \le \dfrac{\lambda_{\max} - \lambda_{\min}}{2} \\ \frac{1}{4}\left(\lambda_{\max}^{1/2} - \lambda_{\min}^{1/2}\right)^2, \\ \qquad\qquad\qquad\qquad \text{otherwise.} \end{cases}$$

Again, when $\lambda_{\min} = 0$, we have that $R_0 = \bar{\lambda}/2$ when the average rate is less than half the maximum, and $R_0 = \lambda_{\max}/4$ when the average is greater (Fig. 4).

## 3.2.  Minimum Mean-Squared Error

Employing the minimum mean-squared error measure amounts to finding the optimal system that extracts the estimate of $X_t$ from the point process that has the smallest mean-squared error $\epsilon^2$. In the Poisson case, the only component present is the extrinsic one, and the resulting mean-squared error characterizes the limits to which a neuron's output can be used to extract an estimate of the signal driving the neuron's rate. Achieving this limit would require the neural pathway receiving the neuron's projections to operate as an optimal estimator.

The minimum mean-squared error estimator is well-known to be the conditional expected value of the information signal when provided with point process observations: $\hat{X}^{\text{MMSE}}(t) = \mathcal{E}[X_t \mid N_t, \mathbf{w}_t]$. Note that we can obtain an upper bound on mean-squared error by assuming that no observations are made but that we still minimize $\epsilon^2$. In this case, $\hat{X}(t) = \mathcal{E}[X_t]$, and

$\epsilon_{\max}^2 = \mathcal{V}[X_t]$. When we do make observations, the optimal estimator is nonlinearly related to the point process and is difficult to find, even in the case of a Poisson process (Snyder, 1975, Sec. 6.5.2). Instead of explicitly finding the estimator, lower bounds on the minimum mean-squared error are known (Snyder, 1975, Sec. 6.5.4). Evaluation of this bound usually requires computation, with few cases extant allowing analytic evaluation.

Simpler expressions for the error and explicit formulation of the estimator become possible if we consider the Poisson process and restrict the estimator to be a linear filter (i.e., a Wiener filter). Here, the intensity is proportional to a positive-valued extrinsic signal $X_t$. Calculation of the mean-squared error reveals that it depends on the power spectrum of the input $X_t$, in terms of both the bandwidth and the shape of the power spectrum. When the input has a first-order, low-pass power spectrum having bandwidth $B$, the mean-squared estimation error normalized by the maximal error is given by

$$\bar{\epsilon}^2 = \frac{\bar{\lambda}B}{\mathcal{V}[\lambda]}\left(\sqrt{1 + \frac{2\mathcal{V}[\lambda]}{\bar{\lambda}B}} - 1\right).$$

Thus, the normalized error depends only on $B/(\mathcal{V}[\lambda]/\bar{\lambda})$, the ratio between the bandwidth and the intensity's variance to average ratio. When the intensity's amplitude is constrained so that $\lambda_{\min} \le \lambda(t) \le \lambda_{\max}$, the variance can be bounded[8] as $\mathcal{V}[\lambda(t)] \le \bar{\lambda}(\lambda_{\max} - \bar{\lambda})$. Consequently, $\mathcal{V}[\lambda(t)]/\bar{\lambda} \le \lambda_{\max} - \bar{\lambda}$, and this bound can be used to lower bound $\bar{\epsilon}^2$ by expressing it in terms of $B/(\lambda_{\max} - \bar{\lambda})$. When $\lambda_{\min} = 0$, an intensity attaining this bound is the random telegraph wave. As shown in Fig. 5 for the case of Poisson process having a random telegraph wave intensity, this normalized error increases as the bandwidth/rate ratio increases, nearly reaching the saturation value when the bandwidth equals the average rate (about 75% of $\epsilon_{\max}^2$).

Furthermore, note that accurate estimates, arbitrarily defined to occur when the normalized squared error is less than 10%, occurs when the intensity's bandwidth is less than a small fraction (about 0.01) of the average rate. This conclusion means that the stimulus representation capability of single auditory-nerve fibers is quite limited. For fibers responding to low frequencies, the intensity's bandwidth is about 300 Hz, and the average rate does not exceed 200 spikes/s. Thus, $B/\bar{\lambda} = 1.5$, which corresponds to a normalized error of at least 79%. The intensity bandwidth that can be
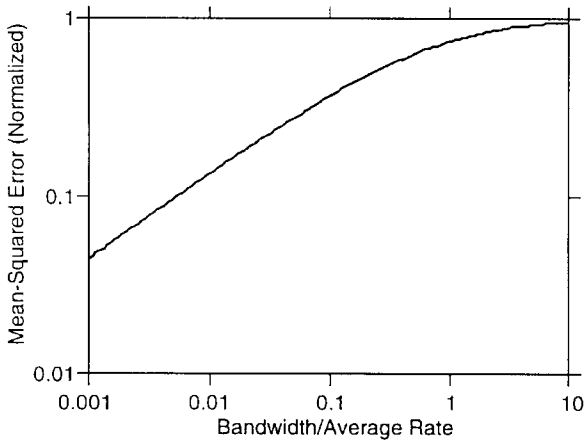
*Figure 5.* The normalized mean-squared error produced by a Wiener filter is plotted against the ratio of bandwidth and the average rate. In this example, the intensity equaled a symmetric random telegraph wave having $\lambda_{\min} = 0$, $\bar{\lambda} = \lambda_{\max}/2$, and $\mathcal{V}[\lambda(t)] = \bar{\lambda}^2$. For all bounded intensities having a first-order power spectrum, this curve serves to lower bound the mean-squared error.

well-represented by auditory-nerve fiber discharges is about 20 Hz.

Features of this result apply generally. Efforts to use Wiener filters to "read the neural code" (Rieke et al., 1993) can only succeed when the input bandwidth is much less than the average rate. When the bandwidth is comparable to or larger than the average rate, estimating the input from a single neuron's output can only produce unacceptably large errors. Our preliminary calculations of the lower bound on the optimal mean-squared error do not indicate that substantial improvements result when we use optimal (nonlinear) estimators.

### 3.3.   Relationships Between Capacity and Mean-Squared Error

To gain a more realistic evaluation of capacity, Shamai and Lapidoth (1993) imposed bandwidth constraints on the intensity. While only being able to calculate upper and lower bounds on the Poisson process capacity, they found a relationship between their upper bound and the mean-squared estimation error. This relation depends in a complex way on the maximal and minimal intensity constraints, but the general form of this relation can be exemplified by the special case $\bar{\lambda} = \lambda_{\max}/2$ and $\lambda_{\min} = 0$, wherein (Fig. 4)

$$\tilde{C} \leq \frac{e}{2} \frac{\bar{\epsilon}^2}{2 - \bar{\epsilon}^2} \ln \frac{2}{\bar{\epsilon}^2},$$

with $\tilde{C}$ denoting the capacity normalized by its maximal value $\lambda_{\max}/e$. Somewhat surprisingly, the capacity is small when the mean-squared estimation error is small, and large when the errors are worse. Thus, capacity does *not* assess the information bearing capabilities of a neuron's output; large capacities can be achieved only when the neuron's extrinsic intensity component *cannot* be well estimated by any system.

This surprising result can be traced to the definition of capacity (Eq. (8)) for a point process: this definition (and many other cases as well) does not correspond to a communication problem. The mutual information measures how well a system's output is associated with its input, with the larger the mutual information, the greater the association. In information theory, we most often take the system to be a communications channel. In discrete channel models, the input represents the finite set of communication symbols and the output the received and decoded symbol. Hence, maximizing mutual information by varying input probability assignments does indeed measure how well the output expresses the input. In the point process case, maximal association between a point process's intensity and the event pattern does not explicitly include extracting the intensity (or the extrinsic input determining the intensity). Thus, capacity does not directly reflect the ability to extract information from a point process. The cutoff rate portrays more accurately than the capacity the ability to extract information represented by a point process. Its properties rest on the assumption that the intensity is selected carefully to maximize information transmission. In general, this ideal situation does not hold since process must represent a wide variety of signals, not carefully selected ones.

The mean-squared estimation error is a more direct, more meaningful measure. Unfortunately, no general result for the estimation error is known. In the case where the intensity is derived from a stationary Gaussian process, a lower bound for the mean-squared error has been derived (Snyder, 1975, Sec. 6.5.4). Using this bound or the results described here, they indicate that single neurons can represent signals with fidelity only if the signals vary slowly compared the neuron's average response rate. Many of these results apply to Poisson processes; it is not known how they generalize to point processes that model neural discharges better. For any particular discharge pattern, understanding its intensity model, particularly its intrinsic structure, determines the neuron's information carrying capability.

## 4.   Analyzing a Point Process

Before the data analyst starts work on any empirical point process, he or she *must* pointedly question the data. The ultimate goal of an analysis technique is to discern what the point process's intensity might be. Employing any specific technique, such as measuring the interevent interval distribution with the interval histogram for example, makes implicit assumptions that the analyst must explicitly check. By measuring the interval histogram, the analyst assumes that the point process's statistical structure—its intensity—does not change with time, only with process history. The validity of this assumption must be verified. Stationarity considerations also enter into situations where the data are knowingly nonstationary. For example, to measure a cell's response to a stimulus, we employ averaging techniques that average the response obtained over a succession of stimulus presentations. Implicit in this averaging is the assumption that the responses do not vary statistically from trial to trial; said another way, the stimulus or the response is assumed not to influence succeeding responses. These kinds of assumptions may not hold for the data at hand. One is tempted to view these unwanted temporal variations as contaminating a more pristine model for the data. As ideal as we would *want* data to be, experimental artifacts or truly variable responses inherent in the system must be addressed *before* attempting to measure the intensity—to decipher the neural code.

### 4.1.   Testing for Temporal Variations

Unfortunately, techniques for assessing nonstationarity are not well developed. From a statistical viewpoint, so many aspects of a point process can vary with time—mean rate, interevent interval variance, interevent interval correlation, and interevent interval probability distribution to name a few—that developing a single, all encompassing test seems difficult at best (Tsao, 1984). Work has been described for analyzing trends in point processes (Cox and Lewis, 1966, Ch. 3). We describe here an outline of a crude procedure for testing whether a point process maintains a statistically similar character or not (Johnson et al., 1986).

A stationary point process consists of a sequence of identically distributed random variables—interevent intervals—that may have a complex dependence structure, such as that evidenced by Markov or Hawkes processes. Thus, we consider a point process as a

stochastic sequence $\{\tau_n\}_{n=1}^N$ where the index is the interval's sequential number. At a minimum, the average value of these intervals should be constant if the process is stationary. We thus form estimates $\bar{\tau}(m)$ of the average interval at various points in the dataset and try to check whether these estimates obey the "rules" of empirical averages:

$$\bar{\tau}(m) = \frac{1}{L} \sum_{n=mL+1}^{(m+1)L} \tau_n.$$

Here, we average over $L$ successive intervals to achieve each estimate of the average interval duration. *If* the process is stationary, the mean and variance of this quantity are given by

$$\mathcal{E}[\bar{\tau}(m)] = \mathcal{E}[\tau_n] \equiv \mathcal{E}[\tau]$$
$$\mathcal{V}[\bar{\tau}(m)] = \frac{\mathcal{V}[\tau_n]}{L} \sum_{l=-(L-1)}^{L-1} \left(1 - \frac{|l|}{L}\right) \rho_\tau(l),$$
$$\mathcal{V}[\tau_n] \equiv \mathcal{V}[\tau].$$

Here, $\mathcal{V}[\cdot]$ means the variance of the indicated quantity, which is the square of the standard deviation that appears in the expression for the coefficient of variation (Eq. (6)), and $\rho_\tau(l)$ indicates the *correlation coefficient* between intervals separated by $l$ intervening intervals:

$$\rho_\tau(l) = \frac{\mathcal{E}[\tau_n \tau_{n+l}] - (\mathcal{E}[\tau])^2}{\mathcal{V}[\tau]}.$$

If the process were renewal, the correlation coefficients would all equal zero save for $\rho_\tau(0)$, which *always* equals one. The standard deviation of the short-term average would then equal $\sigma[\tau]/\sqrt{L}$. This quantity is usually used in stationarity tests despite the possibility of significant interevent interval correlation. Estimating correlation coefficients can be made only if the process is stationary, and we don't know that yet. Essentially, preliminary analysis assumes the correlation, if present, is small.

If the process were stationary, the average interval computed over the entire dataset should not deviate "significantly" from the short-term averages $\bar{\tau}(m)$. Because of the Central Limit Theorem, the sum in the expression for the short term average can be well-approximated as having a Gaussian probability distribution. Exploiting this approximation, we can establish *confidence intervals* about the dataset average for the short-term averages. The confidence intervals require the variance $\sigma^2[\tau] = \mathcal{V}[\tau]$, which we estimate
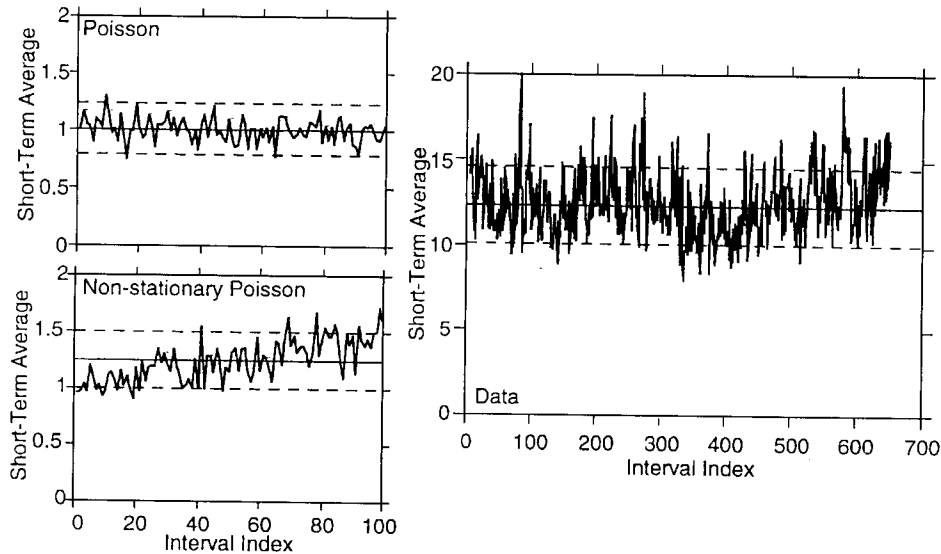
*Figure 6.*    Application of the stationarity test is portrayed on the left for simulated data and on the right for actual measurements. The "Poisson" case is a stationary Poisson process having rate one. The "nonstationary Poisson" case was also Poisson, but having a rate that increased linearly very slowly. The data were single-unit measurements taken from auditory-nerve fibers stimulated by a continuous tone. In each case, 100 intervals comprised the short-term average, and these averages are shown by the jagged solid lines as a function of the index of the average. The confidence intervals are shown as dashed lines on either side of the solid-line mean, with $\sigma$ equaling the standard deviation estimated from the entire dataset divided by the square-root of the number of terms in the short-term average (10). The right panel shows 100-interval averages for data recorded from single auditory-nerve fibers. The means and standard deviations calculated for these examples are: Poisson, 1.00 and 0.99; nonstationary Poisson, 1.25 and 1.29; data, 12.6 and 11.1. The number of times the short-term averages exceed the confidence interval is 6 in the Poisson case and 19 in the nonstationary Poisson case. The data clearly exceed the confidence interval far more than would be consistent with a stationary point process model.

from the entire dataset. *Assuming* a large number of blocks of length $L$ are present in the dataset, the long-term average $\bar{\tau}$ will be more statistically accurate (the variance is smaller) than the short-term averages. We thus use the long-term average as our baseline and establish confidence intervals as $\bar{\tau} \pm k\sigma[\tau]/\sqrt{L}$, where the constant $k$ controls the degree of confidence. For example, variations *beyond* plus or minus one standard deviation ($k = 1$) about the mean of a Gaussian random variable occurs 32% of the time. Thus, we should expect the short-term averages to exceed this confidence interval at this rate and remain *consistent* with a stationarity assumption. If $k = 2$, the excess probability drops dramatically to 4.6%. Thus, with a confidence interval equaling $\bar{\tau} \pm 2\sigma[\tau]/\sqrt{L}$, we should expect about 5% of the short-term averages to exceed these bounds. This confidence interval is typically used in applications.

In Fig. 6, we illustrate the application of the stationarity test. In the stationary (Poisson) case shown there, the confidence interval was exceeded at a rate of 6/100, which roughly corresponds to what we would expect from the previously described theory. In the nonstationary case, however, the exceedence rate was 19/100,

a clear discrepancy from consistency with a stationary model. We accept stationarity in the first case, but not in the second. Stationarity analysis of actual measurements is also shown. These recordings illustrate different kind of discrepancy from stationarity than that of the linearly increasing interval example. Here, the averages are much too random, and exceed the confidence interval far too often. A more detailed analysis revealed that these data are best modeled as fractal-intensity renewal processes (Kumar and Johnson, 1993).

## 4.2.    The Interval Histogram and Hazard Function

The interval histogram estimates the probability density function of interevent interval durations. When the process is stationary and renewal, the interval histogram suffices to characterize the process. When dependencies are present, the interval histogram no longer produces as important a measure for data analysis. Later we discuss how interval histogram analysis changes when we have a Markov point process.

Algorithmically, the interval histogram computes the fraction of measured interevent intervals that lie in a given bin having binwidth $\delta$, then normalizes by the binwidth so that the estimate "integrates" to one[9]. Letting $I_{l,\delta}(\cdot)$ denote the indicator function,

$$I_{l,\delta}(\tau) \equiv \begin{cases} 1, & l\delta \leq \tau < (l+1)\delta \\ 0, & \text{otherwise} \end{cases},$$

the interval histogram computation is expressed by Johnson (1978)

$$\text{INT}(l) = \frac{1}{N\delta} \sum_{n=0}^{N-1} I_{l,\delta}(\tau_n), \qquad (9)$$

where $N$ is the total number of intervals entering into the computation. The expected value of the histogram clearly equals $\int_{l\delta}^{(l+1)\delta} p_\tau(\tau)\, d\tau/\delta$. If the true interval distribution is essentially constant over all length-$\delta$ bins, this expected value equals the interval pdf evaluated at $l\delta$: $\mathcal{E}[\text{INT}(l)] \approx p_\tau(l\delta)$.

Assuming a renewal point process, we can easily analyze the variance of the interval histogram. Because of the indicator function, the individual terms in Eq. (9) correspond to Bernoulli random variables where the probability each equals one is given by $p = \int_{l\delta}^{(l+1)\delta} p_\tau(\tau)\, d\tau$. The variance of a sum of $N$ such independent random variables equals $Np(1-p)$. Assuming that $p \ll 1$, we have that the variance of the interval histogram equals

$$\mathcal{V}[\text{INT}(l)] = \frac{\int_{l\delta}^{(l+1)\delta} p_\tau(\tau)\, d\tau}{N\delta^2} \approx \frac{p_\tau(l\delta)}{N\delta}.$$

To assess this variance better, we consider the histogram's coefficient of variation, which measures the percentage error of its estimate of the interval density:

$$C[\text{INT}(l)] = 1 \Big/ \sqrt{N \int_{l\delta}^{(l+1)\delta} p_\tau(\tau)\, d\tau}$$
$$\approx [N\delta p_\tau(l\delta)]^{-1/2}.$$

Thus, the percentage error increases as the density's value decreases: we can more accurately estimate the histogram's modes than we can its valleys and tails. Note that the argument of the square root is the expected number of intervals in a bin. Setting a criterion value for the coefficient of variation—10% say—translates into statistically significant estimates in those bins containing at least 100 intervals. An example histogram is shown in Fig. 7.

Selecting the binwidth creates a tension in accurately estimating the interval distribution with the interval histogram. First of all, we want the binwidth to be small enough so that we are, to a good approximation, estimating the value of the density rather than its integral over a bin. This desire amounts to a sampling condition: The interval density should vary slowly over a binwidth. On the other hand, if we choose too small a binwidth, the statistical errors can lead to inaccurate measurements: We must have enough intervals lying within a bin to achieve statistically reliable estimates. Thus, what might be called a sampling theorem for interval histograms states that density variations can be captured at a particular interval *only* if a sufficient amount of data are available at that interval. If not, sampling accuracy has been sacrificed to statistical inaccuracy[10]. As a practical matter, we never have enough data; often the statistical constraint is difficult to overcome, and detail in the interval distribution goes unmeasured.

Because the hazard function and the interval pdf are directly related, we estimate the hazard function using a discrete version of Eq. (5):

$$\text{HAZ}(l) = \frac{\text{INT}(l)}{\sum_{k=l}^{\infty} \text{INT}(k)\delta}.$$

The theoretical formula demands that the numerator equal the value of the interval distribution, not its integral. The interval histogram yields this value only at small binwidths, which can produce statistically inaccurate estimates when we don't have enough data. The denominator, on the other hand, is well estimated by the interval histogram because it *does* equal the density's integral over several bins no matter what binwidth we chose.

Calculating the coefficient of variation for the hazard function is analytically difficult because the hazard function is the ratio of two random, correlated quantities. I develop a series of approximations that allows us to develop an expression for $C[\text{HAZ}(l)]$ when it is small. First of all, when a random variable's coefficient of variation is small, it also approximately equals the coefficient of variation of the random variable's reciprocal[11]. Because of this fact, we consider the coefficient of variation of this expression's reciprocal since it is simpler to calculate:

$$C[\text{HAZ}(l)] \approx C\left[\delta + \frac{1}{\text{INT}(l)} \sum_{k=l+1}^{\infty} \text{INT}(k)\delta\right].$$
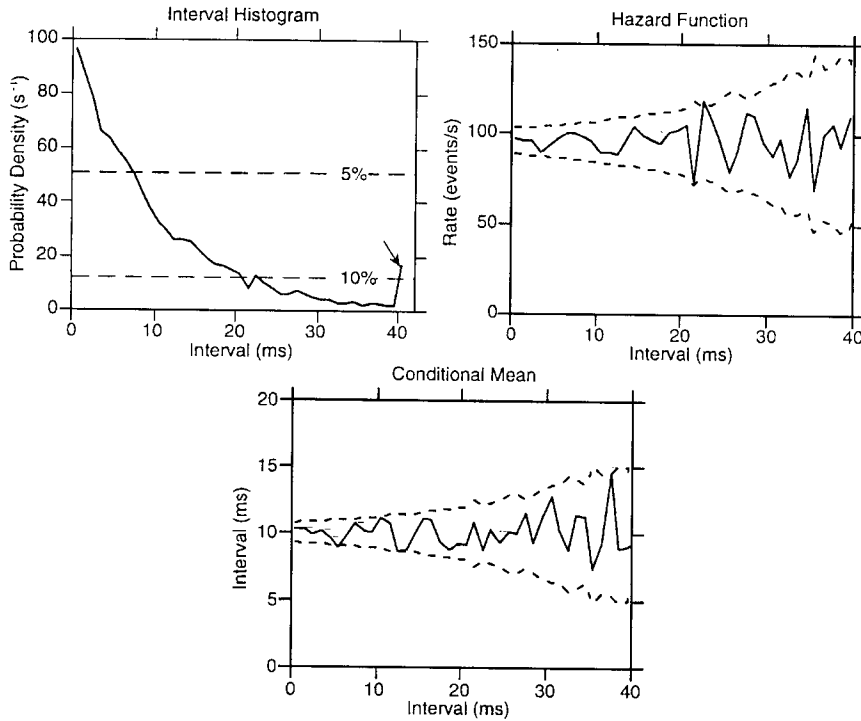
*Figure 7.* From a simulated stationary Poisson process (rate equaling 100 events/s), 7,500 events were taken, and the interval histogram, hazard function, and conditional mean computed. The binwidth in all cases equaled 1 ms.

The horizontal dashed lines on the interval histogram panel correspond to the indicated percentage errors (as quantified by the coefficient of variation). The arrow points at the *overflow bin*: the bin reserved for intervals longer than those taken into account by the interval histogram calculation. Thus, this bin "keeps the books straight" by indicating that longer intervals are present, thereby maintaining the unit-area normalization of the histogram calculation.

The dashed lines on the hazard function panel indicate $\pm 2\sigma$ about the hazard function's average value. Note how the error increases as the interval increases: peaks in the rightmost portion of the hazard function are not statistically significant while much smaller peaks to the left can be. This average value in the hazard function estimate equaled 95.6 events/s, which does not equal the process's average value. This discrepancy can be traced to the borderline choice for histogram binwidth: it is not small enough to result in interval histograms that yield pointwise instead of integrated estimates, but using smaller binwidths would have compromised the statistical accuracy.

The solid line in the bottom panel shows the conditional mean, and the dashed lines the $\pm 2\sigma$ criterion curves that would apply if the events were well described as a renewal process. As the conditional mean should for a Poisson process, it lies within the criterion curves.

We next approximate the variance of the reciprocal of a random variable having a small coefficient of variation as $\mathcal{V}[1/X] \approx \mathcal{V}[X]/(\mathcal{E}[X])^4$ (see note 11). Using this approximation, we find that

$$C[\text{HAZ}(l)] \approx \sqrt{\frac{1}{N\delta p_\tau(l\delta)} - \frac{1}{N\delta \sum_{k=l}^{\infty} p_\tau(k\delta)}}$$

$$= C[\text{INT}(l)][1 - h(l\delta)\delta]^{1/2}. \qquad (10)$$

This expression is meaningless unless $h(l\delta)\delta \ll 1$, which defines the situations under which our approximations ultimately apply. In such situations, the coefficient of variation for the hazard function estimate approximately equals that for the interval histogram, with both becoming large for the same intervals. How-

ever, as opposed to the interval histogram, the value of the hazard function is usually not small when these large values occur, which accentuates the errors.

The hazard function shown in Fig. 7 illustrates this point: for small intervals, statistical variations in the estimate are small, but increase dramatically for larger intervals. For an exponential interval distribution having parameter $\lambda$,

$$C[\text{INT}(l)] = \frac{e^{\lambda l\delta/2}}{(N\delta\lambda)^{1/2}}$$

$$C[\text{HAZ}(l)] = C[\text{INT}(l)] (1 - \lambda\delta)^{1/2}.$$

The coefficient of variation in the hazard function is thus proportional to that of the interval histogram. In

the interval histogram case, the exponentially increasing percentage error is due to the exponentially decreasing histogram: The errors are constant. In contrast, because the hazard function is a constant, the same exponential increase is due to exponentially increasing errors. In general, care must be taken in the tails of the interval distribution when computing either quantity; in addition, statistical significance is lost whenever the interval histogram's value falls below a threshold value defined by an error criterion. Note that while the hazard function's coefficient of variation can be estimated from the interval histogram (Eq. (10)), we usually do not know what the hazard function should be. Consequently, we don't know what curve about which to center the confidence interval.

### 4.3.  Measuring Interevent Dependence

Understanding the statistical interdependence of interevent intervals cannot be stressed enough if we desire as complete a characterization as possible of a point process. This dependence structure, expressed by the historical component of the intensity, describes how events interact with each other, which places strong constraints on mechanisms that might produce the measurements. The simplest nontrivial dependence structure is the Markovian, wherein dependence extends back a fixed *number* of previous event occurrence times; more remote events than that do not affect the current event's timing. It is important to realize that Markovian dependence extends across a variable time epoch prior to each event; fixed-duration dependence structures cannot be modeled with Markovian models. A Hawkes process could be constructed for such situations. Because of difficulties in dealing with the Hawkes process, a Markovian dependence structure is usually assumed first, the number and type of dependence measured, then simulated event patterns compared with the original data as a consistency test.

A point process has Markov *order q* if the conditional interval distribution depends on no more than the $q$ previous intervals.

$$p_{\tau_n \mid \tau_{n-1}, \ldots, \tau_1}(\tau_n \mid \tau_{n-1}, \ldots, \tau_1)$$
$$= p_{\tau_n \mid \tau_{n-1}, \ldots, \tau_{n-q}}(\tau_n \mid \tau_{n-1}, \ldots, \tau_{n-q}), \quad n > q \quad (11)$$

A renewal process, which has no dependence structure, has Markov order zero. To determine the Markov order, one could compute the *conditional interval histogram* or, equivalently, the *conditional hazard function*, from

data (Johnson et al., 1986): the usual interval histogram is computed, but only from those intervals that are preceded by a sequence of intervals equaling predefined values. A succession of histograms or hazard functions, each of which is conditioned on a greater number of previous intervals, would need to be computed and compared to determine when the addition of conditioning interval no longer changed the histogram. By estimating the conditional hazard function, we are directly estimating the intensity. Because of this fact, we have found conditional hazard functions to provide greater insight into the point process than conditional interval histograms.

As a practical matter, the probability of a sequence of intervals precisely equaling some set of values is zero. Instead, the conditioning intervals must be selected from those lying within a sequence of interval ranges. In our experience, the ranges must be greater than the binwidth used in the interval histogram itself; if not, so few interval sequences lie in the conditioning interval ranges that statistically significant histograms cannot be computed from limited data. For example, consider the easily analyzed case of a Poisson process. If it has a rate of 100 events/s and we use a 1 ms binwidth, demanding at least 10% significance (coefficient of variation) in the usual interval histogram means that we need somewhat more than 7,400 events to obtain that significance level for intervals having durations less than or equal to two time constants of the actual interval distribution. If we computed the first-order conditional interval histogram based on a conditioning value lying within a 1 ms range, we would need 100 times that amount of data to satisfy the same significance criterion. If instead, we use a 10 ms conditioning range, a value much too large for a histogram binwidth, we need about 55,000 events, a large but manageable number. A second-order histogram requires that much more data, etc. Unfortunately, for even small orders, *much* data—on the order of exponential in $q + 1$—are required to yield statistically significant estimates.

The statistical concerns can be mitigated to some degree by a less statistically demanding measure based on the computation of *conditional mean* (Johnson et al., 1986; Rodieck et al., 1962). A clear consequence of Eq. (11) is that the conditional expected value also has the same Markovian property.

$$\mathcal{E}[\tau_n \mid \tau_{n-1}, \ldots, \tau_1] = \mathcal{E}[\tau_n \mid \tau_{n-1}, \ldots, \tau_{n-q}], \quad n > q.$$
$$(12)$$

Just as clear is the conclusion that this property is necessary, but not sufficient, for the point process to be Markovian. To estimate the $q$th-order conditional mean, one simply computes the average value of all intervals that are preceded by a particular sequence of intervals:

$$\text{CM}(\tau_1, \ldots, \tau_q) = \frac{1}{N(\tau_1, \ldots, \tau_q)}$$

$$\times \sum_{n=q+1}^{N} \tau_n I(\tau_{n-1} = \tau_1; \cdots; \tau_{n-q} = \tau_q).$$

Here, $I(\tau_{n-1} = \tau_1; \cdots; \tau_{n-q} = \tau_q)$ denotes the indicator function that the previous interval sequence $\{\tau_{n-1}, \ldots, \tau_{n-q}\}$ equals the conditioning sequence $\{\tau_1, \ldots, \tau_q\}$, and $N(\tau_1, \ldots, \tau_q)$ denotes the number of intervals preceded by this sequence. Computationally, no sequence of data will equal a specified set of values. To compute the conditional mean, we form bins and specify equality in the conditioning process by the occurrence of intervals lying in a sequence $\{l_1, \ldots, l_q\}$ of bins:

$$\text{CM}(l_1\delta, \ldots, l_q\delta)$$

$$= \frac{1}{N(\tau_1, \ldots, \tau_q)} \sum_{n=q+1}^{N} \tau_n I_{l_1; \cdots; l_q, \delta}(\tau_{n-1}; \cdots; \tau_{n-q}).$$

Estimation of the conditional mean still has a statistical complexity of order exponential in $q$, but for small orders the data demands are much less than those of the conditional interval histogram. The most commonly computed conditional mean occurs for $q = 1$: the first-order conditional mean.

Figure 7 demonstrates that the first-order conditional mean for a simulated Poisson process is constant. This result means that the average interval duration does not depend on the value of the conditioning interval, consistent with the fact that the process has no first-order dependence. The error bounds were derived from a renewal process assumption so as to define a null-hypothesis statistical test. The estimate of the expected value $\mathcal{E}[\tau_n]$ was first computed. If renewal, the theoretical first-order conditional mean should equal this value. The statistical error in this equality is based on simple considerations of computing averages: $\mathcal{V}[\text{CM}(\tau_1)] = \mathcal{V}[\tau]/N(\tau_1)$. The variance of intervals is calculated from the entire dataset in the obvious way. As the numerator of this expression is fixed, the variation of the error bound curves evident in Fig. 7 is due to the denominator. This quantity, which equals the number of intervals lying in $l\delta \leq \tau_1 < (l+1)\delta$, is precisely

what the interval histogram computes. Thus, the error curves for the first-order conditional mean will be proportional to the reciprocal of the interval histogram when they are computed with the same binwidth.

### 4.4. The PST Histogram

The PST (post-stimulus or peri-stimulus time) histogram measures the temporal variations in the rate of discharge by averaging the response to a periodically repeating stimulus (examples are shown in Fig. 9). The onset of each stimulus presentation serves as the histogram's time origin. Assume that the time axis is quantized into bins of length $\delta$, and that the stimulus is presented every $M$ bins with the first stimulus presented at time $t = 0$. The count of the events in the $m$th bin $[m\delta, (m+1)\delta)$ is denoted by $c(m)$. The PST histogram computation averages this sequence across stimulus presentations (Johnson, 1978):

$$\text{PST}(m) = \frac{1}{K\delta} \sum_{k=0}^{K-1} c(m + kM).$$

Here $K$ denotes the number of stimulus presentations. Normalized in this way, the units of a PST histogram are events/s.

For even the simplest point process, the expected value of the PST histogram contains many surprises. First, assume that the binwidth is small so that the intensity definition of Eq. (1) can be used. This assumption means that more than one event rarely occurs within a single bin. Under this condition, the expected value of the sequence $c(m)$ is $\mathcal{E}[\mu(m\delta; N_{m\delta}, \mathbf{w}_{m\delta})]\delta$, where the expected value is computed with respect to event history. When the intensity's time dependence repeats with period $M\delta$, the expected value of a PST histogram equals this quantity divided by $\delta$. Thus, the PST histogram corresponds to averaging over all event histories that lead to or prevent an event occurring at time $m\delta$ after the stimulus presentation:

$$\mathcal{E}[\text{PST}(m)] = \mathcal{E}[\mu(m\delta; N_{m\delta}, \mathbf{w}_{m\delta})]. \qquad (13)$$

Thus, this expected value does *not* depend solely on the intensity's temporal variation. The *only* point process that expresses this temporal dependence directly is the Poisson because its intensity does not depend on event history. For all other point processes, including renewal ones, *the PST histogram represents stimulus-related temporal variations averaged across the point process's history.* This expected value is

extremely difficult to evaluate. Consequently, the PST histogram is a complicated function of the process's stimulus-related nonstationarity and the inherent "non-stationarity" induced by the process's statistical structure (Johnson and Swami, 1983). Examples of this phenomenon are presented in succeeding sections.

## 5. Application to Discharge Patterns in the Auditory Pathway

### 5.1. Analyzing Auditory-Nerve Fiber Discharges

Measurement of the conditional mean from single auditory-nerve fibers has long shown that, to a very good approximation, that a renewal process model can be used (Gaumond et al., 1982; Kiang et al., 1965). Typical is the conditional mean in Fig. 8, which shows no statistically significant deviations from a constant. Measured hazard functions typically show what appears to be an exponentially increasing quantity, eventually becoming a constant after about 20 ms (Fig. 8). Detailed analysis of the hazard function can be found in Li and Young (1993). The constant portion of the hazard function means that the associated interval histogram has an exponential tail as in the Poisson process. Thus, the interspike interval pdf resembles a delayed exponential, with the hazard function approximated as a delayed step. A common model for constant-rate auditory-nerve fiber recordings is

$$\tilde{\mu}(\tau) = \mu_0 u(\tau - \Delta), \qquad (14)$$

where $\Delta$ is the absolute refractory interval (deadtime) of the discharge pattern. Typically, $\Delta$ is in the range 0.7–1 ms. Consequently, the intensity of the process varies as a function of time depending on when discharges occur (see Fig. 1). This stationary model can be used when spontaneous activity or the responses to high-frequency stimuli are measured.

When the discharge rate varies with time, a nonstationary model must be used. A *nonstationary* renewal process is defined by an intensity of the form

$$\mu(t; N_t, \mathbf{w}_t) = f(t, t - w_{N_t}).$$

This characterization is too general to motivate data analysis techniques. A separable form of this intensity has proven to describe auditory-nerve fiber recordings quite well: The time-varying part—the extrinsic component that depends on absolute time—is assumed to be separate from the refractory (history-dependent) part (the intrinsic component) (Gaumond et al., 1982; Johnson and Swami, 1983; Siebert and Gray, 1963):

$$\mu(t; N_t, \mathbf{w}_t) = d(t) \cdot r(t - w_{N_t}).$$

For theoretical reasons, we must assume that the recovery function attains a maximum value of one at long interspike intervals. The *drive* $d(t)$ models the extrinsic component of the intensity, which is directly related to the stimulus, the signal conveyed by the renewal point process. For example, when a low-frequency tone—a sound pressure proportional to $\sin(2\pi f t)$—serves as the stimulus, the drive has been modeled as
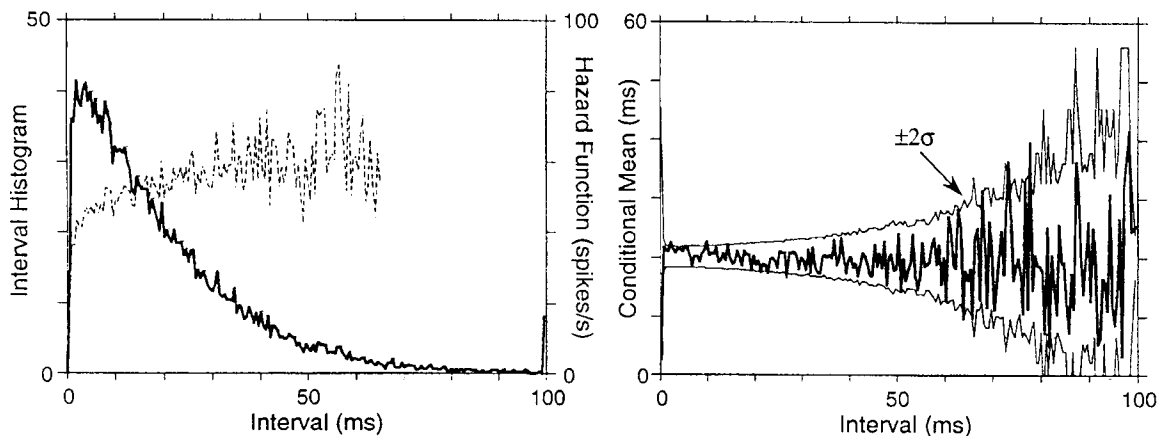


*Figure 8.* The left panel shows the interval histogram (solid line) with the hazard function (dashed line) superimposed upon it. Because the interval histogram estimates a probability density function, its vertical scale has units of probability/s. The conditional mean (solid line) is shown on the right as well as ±2σ curves. These histograms were computed from data collected from a single auditory-nerve fiber responding to a long-duration high-frequency tone. Based on this kind of analysis, the renewal model for auditory-nerve fiber discharge pattern is justified.

$d(t) = \mu_0 \exp\{a \sin(2\pi f t + \phi)\}$ (Johnson, 1974). The instantaneous rate is thus *modulated* by the drive. Decoding the neural code thus becomes equivalent to *demodulating* the observed process so as to separate the information-bearing component—$d(t)$—from the "carrier", the recovery function.

Enabled by the relative simplicity of the intensity for auditory-nerve fiber discharges, we can predict what the PST histogram will be in simple situations. The expected value of the PST histogram for this case equals

$$\mathcal{E}[\text{PST}(m)] = d(m\delta)\mathcal{E}\left[r\left(m\delta - w_{N_{m\delta}}\right)\right]. \qquad (15)$$

The quantity we want to extract, $d(m\delta)$, is multiplied by a quantity that is always less that one and varies implicitly on $d(\cdot)$. The total effect is nonlinear: the expected value depends in some fashion on the drive, and the PST histogram equals the drive times this expected value. Analytic computation of this expected value is difficult in most cases, requiring that simulations be performed to assess the effect of the recovery function on the PST histogram (Johnson and Swami, 1983). Results of this analysis are shown in Fig. 9. The nonlinearity is evident in the left column, where the only variation of the drive is its amplitude. The
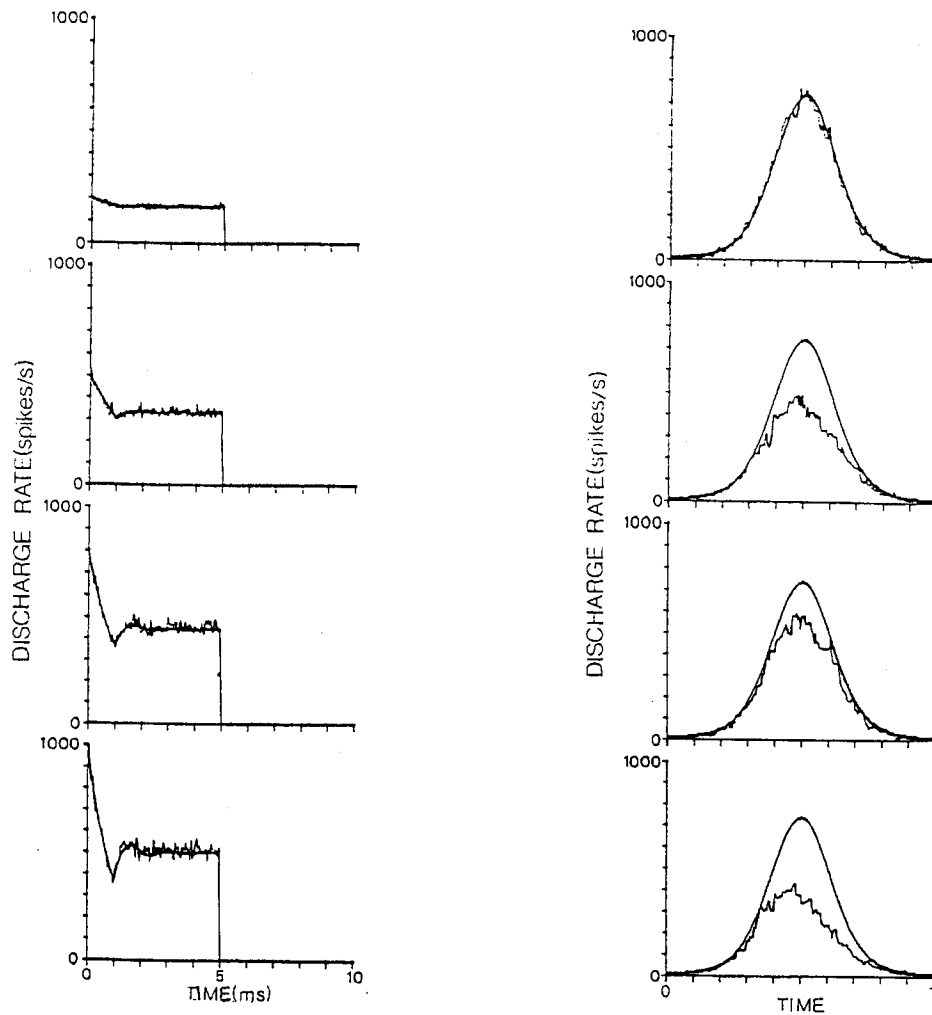


*Figure 9.* Shown as jagged lines are PST histograms calculated from simulated renewal processes (the recovery function was a delayed step; see Eq. (14) having an absolute refractory interval of $\Delta = 1$ ms) and as solid lines the theoretical predictions of Eq. (15). The left column displays PST histograms of a simulated point process where the drive function is a square pulse of 5 ms duration. The amplitude of the pulse increases down the column, taking on values of 200, 500, 800, and 1000 spikes/s. In the right column, the PST histograms of a simulated point process whose drive had the form $50\exp\{-2.7\cos(2\pi t/T)\}$. One period of the PST histogram is shown. From top to bottom, the frequency of each simulated drive was 0.1, 0.5, 1.0, and 3.0 kHz. In all cases, smooth lines indicate the actual discharge rate and jagged lines measured PST histograms. Taken from Johnson and Swami (1983).

effect of the recovery function is to introduce a transient into the PST histogram, which can be traced to the sudden change in drive amplitude. The initial portion of the PST histogram is a decaying exponential whose time constant equals the reciprocal of the drive's amplitude $\mu_0$. The "steady-state" value of the response can be shown to be $\mu_0/(1 + \mu_0\Delta)$.

Researchers have developed techniques for estimating $d(t)$ from spike trains (Johnson and Swami, 1983; Mark and Miller, 1992; Miller, 1985). These techniques apply estimation theoretic procedures to the discharge pattern, accumulating an estimate of a renewal process's time-varying component. These procedures are too complicated to describe here; they all assume the separable model for the intensity, and exploit it to the fullest.

### 5.2. Modeling Discharge Patterns of Lateral Superior Olive Neurons

Figure 10 shows the first-order conditional mean for data recorded from single units in the lateral superior olive (LSO)[12]. The monotonically decreasing trend for the conditional mean demonstrated there was consistently found in all LSO unit activity. Clearly, these data require an intensity-based description that has an order of at least one: the data *cannot* be modeled as a renewal point process. Estimates of the second-order conditional expected value indicated that no significant second-order dependence was present (Johnson et al., 1986). Hence, a first-order Markovian model seems to describe these discharge patterns well.

To estimate the intensity of a first-order Markov point process, the *conditional hazard function* must be estimated. Derived in the obvious way from the

conditional interval histogram, this histogram is computed by creating a histogram only from those intervals preceded by intervals of a narrow range of durations. When applied to LSO discharge patterns, a surprisingly simple structure emerged (Johnson et al., 1986): the effect of varying the conditioning interval is to *shift* the conditional hazard function along the horizontal axis. Consequently, the intensity of stationary activity recorded from most LSO units can be described by the simple relation

$$\tilde{\mu}(\tau_n; \tau_{n-1}) = \mu_0 r(\tau_n - s(\tau_{n-1})), \qquad (16)$$

where $r(\cdot)$ is the basic hazard (intensity) shape that is being shifted and $s(\tau_{n-1})$ is the shift of this shape with preceding interval. The fundamental equation defining the generation of interevent intervals becomes $E_{n+1} = R(\tau_{n+1} - s(\tau_n))$, where $R(\cdot)$ is a function whose derivative is $r(\cdot)$. The generation equation for a point process that describes LSO data becomes

$$\tau_{n+1} = s(\tau_n) + R^{-1}(E_{n+1}). \qquad (17)$$

Several intriguing conclusions can be gleaned from this relationship.

- The shifting function determines the point process's statistical dependence structure (the correlation between successive interevent intervals).
- The first-order conditional expected value of the interevent interval is found to be $\mathcal{E}[\tau_n \mid \tau_{n-1}] = s(\tau_{n-1}) + K$, $K$ a constant. Thus, measuring the conditional mean amounts to measuring the shifting function $s(\cdot)$.
- The intervals in the point process are generated by a rather simple difference equation.
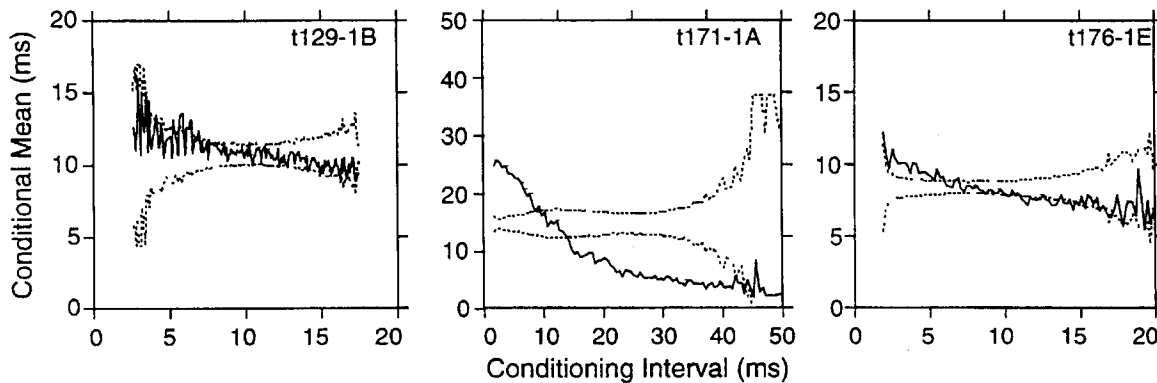


*Figure 10.* The estimate of the first-order conditional expected value depicted here was computed from stimulated activity recorded from three single units located in the LSO. The error bounds are clearly exceeded, from which one would conclude that these data must be described by a model that has Markov order of at least one.
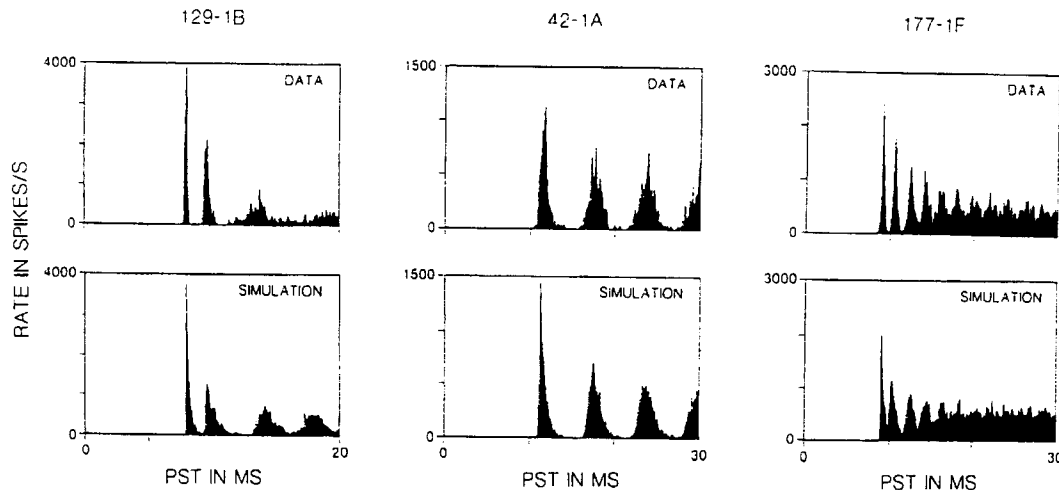
*Figure 11.*    The top row displays PST histograms of discharge patterns recorded from single LSO units responding to repeated presentations of a tone burst. The bottom row contains PST histograms gleaned from simulations that used detailed idealizations of the drive and the recovery function. The drive was constant for the leftmost histogram and rose slowly in a linear fashion for the remaining two. Taken from Johnson et al. (1986).

Thus, this relationship entirely describes the statistical structure of the point process, which reveals the underlying simplicity of the point process.

The PST histogram discrepancies from the Poisson ideal are dramatic when LSO responses are considered (Johnson et al., 1986). Figure 11 shows the measured and simulated response of several LSO units to high-frequency tone bursts. The measured response to the step-like stimulus displays an oscillatory response that can exhibit unequally spaced variations. The simulations made use of dependence measurements obtained during the later sustained portion of the tone-burst response to form estimates of the recovery function. The drive in all simulations contained *no* oscillations; comparatively constant functions were used for the drive. The close agreement is evident. We can conclude that *significant variations of measured PST histograms can be due to the statistical structure of the point process, not to the drive.* A non-Poisson dependence structure will *always* induce some sort of "transient" response when suddenly applied stimuli are used. Thus, only detailed analysis can separate the intrinsic intensity components from the extrinsic.

## 6.   Summary

Point processes, the ubiquitous model of neural discharge patterns, are completely characterized by their intensities. The structure of the intensity reveals that

event occurrence depends on two fundamental components:

- A component whose temporal variations depend on external sources, not the point process itself, and
- A structural component that expresses how events influence each other.

This logical separation does not mean they are easily teased apart. For example, changing event rate by increasing the first component makes events occur more closely together, which centers the structural component about a different operating point. Furthermore, external stimuli can influence *both* components.

Because we idealize the information-bearing aspect of a neuron's output to be the time-of-occurrence of action potentials, to decipher the neural code expressed by a discharge pattern means to determine the intensity of an underlying point process model. In my work, I follow a specific routine in all our preliminary analyses with a view toward determining the form of a point process model's intensity capable of describing the data. These steps, exemplified here in analyzing auditory-nerve fiber and LSO recordings, are

1. *Isolate a Statistically Stationary Segment in a Recording.* The stationarity test described here must be considered a minimal one. Developing robust tests of stationarity is a research topic that has

not yet found a satisfactory answer. Whatever test is used, succeeding steps depend on focusing on statistically stable regions of the discharge pattern. These regions need not be contiguous. For example, we analyzed LSO responses to a gated stimulus during identical temporal poststimulus epochs (Johnson et al., 1986). Finding no stationary segments need not stop analysis. We have found that careful analysis of temporal rate variations can be revealing, and assist the researcher in applying more complicated intensity models (Kumar and Johnson, 1993).

2. *Determine the Order of the Dependence Structure.* The complexity of the intensity increases with how many previous discharges affect the current one. This number is most effectively measured with the conditional mean. If the first-order conditional mean is a constant (within statistical error bounds as in Fig. 8), a renewal model can be employed. If not a constant (as in Fig. 10), higher order conditional means should be calculated. Unfortunately, the higher the order, the greater the demands on the amount of data needed. This order-determination problem has been attacked in the time series analysis literature in the guise of non-Gaussian Markovian processes. The techniques developed to date also require much data (Kumar and Johnson, 1990).

3. *Measure the Hazard Function According to the Measured Markovian Order.* The resulting conditional hazard functions reveal the statistical structure of your data. These measurements provide the fodder for developing a parsimonious model that succinctly captures the dominant aspects of the discharge pattern.

4. *Simulate a Point Process Having an Intensity Derived from Data.* We use simulation to confirm that our approximations to the intensity are indeed self-consistent: Do histograms measured from data and simulation agree well (within expected statistical error)? We have found that demanding agreement often leads to important model refinements (Johnson et al., 1986).

Auditory stimuli frequently evoke transient responses. We have shown responses produced by ideal auditory-nerve fibers (Fig. 9) and LSO neurons (Fig. 11) when an amplitude-gated tone is presented. Because the PST histogram presents a measurement that mixes the time-dependent and structural components in an complicated way (Eq. (13)), we simulate responses to transient rate changes and compare the PST histogram with this change in an attempt to understand how the intensity's intrinsic properties affect the response. As we have shown, such effects pervade PST histograms measured from tone-burst responses throughout the auditory system. To measure the time-varying component, which reveals what the neuron is representing about the stimulus, we resort to (tedious) simulation: guessing a drive function, simulating the response, measuring the PST histogram, and making appropriate corrections. We have found that the intensity's event-history component can also undergo transient changes, which complicates this procedure. No statistical procedure for directly measuring the intensity's components from the PST histogram has yet been found.

Ideally, having an intensity that completely describes a neuron's discharge pattern would directly reveal neural function with respect to the stimulus condition and neural state under study. In our work with computational neuron models (Zacksenhouse et al., 1995), we have not found this situation to be clear-cut. On one hand, requiring a computational model to produce an LSO-like discharge pattern having an intensity indistinguishable from that measured does place tight constraints on the model's biophysical mechanisms and on the model neuron's geometry-dependent aspects (Zacksenhouse et al., 1995). On the other, the relationship between the intensity components and the model neuron components is quite complicated, and working backward from data to model parameters does not seem straightforward. For example, producing an interval histogram having a specific form frequently involves setting synaptic conductances, synaptic time constants, and active model parameters in concert. We infer computational models by performing simulations to produce discharge patterns that, when analyzed by the same methods applied to data, are indistinguishable. Refining a computational model to meet constraints placed by point-process analysis is tedious, frequently requiring adjustment of several parameters to approximate measured quantities such as the hazard function. Be that as it may, careful analysis of a neuron's discharge pattern represents the first step in determining what computational models will work. For example, if the data are inconsistent with a renewal model, classic $Na^+$-$K^+$ Hodgkin-Huxley models cannot work. Future work will certainly elucidate how a point process's statistical structure constrains computational models.

The intensity description of point processes does formally define what the neural code of a single neuron is. What the general form of Eq. (1) suggests is that the code represents extrinsic and intrinsic influences. For sensory systems, the extrinsic components are related to the stimulus, and affects the time-dependent components of the intensity. Intrinsic ones are related to the neuron's biophysics. Because of the nonlinear nature of these mechanisms, they also reflect the extrinsic influences as well. Teasing these components apart *must* take into account the discharge pattern's statistical properties, which are completely captured by the measured intensity. We have shown that measurements of the hazard function, particularly during stationary data segments, open the door to determining what the intrinsic and extrinsic influences might be. Much work remains in developing systematic procedures for measuring the intensity that best describes any given dataset.

Stimuli and more abstract information are represented by the intensity's extrinsic component. The capability of a single neuron to represent information depends on the intrinsic component. From considerations of the capacity, the point process maximizing this quantity is the Poisson, the one having the simplest intrinsic structure. Assuming this result applies to measures of the quality of information representation, the Poisson process represents information with greater fidelity than any point process. As we have seen, capacity calculations do not accurately portray how well signals can be extracted from point process observations. The cutoff rate, which provides a better picture, is about a factor of two smaller than the capacity. The more direct mean-squared error measure seems better. Calculations made on the Poisson process indicates that errors must increase as the bandwidth of the extrinsic component increases (for processes having fixed maximal and minimal rates). As a rule of thumb, this bandwidth, when normalized by the average rate, must be less than 0.1 to obtain reasonable fidelity in extracting information from a *single* neuron's discharge pattern, be the measurement artificial (made by investigators) or natural (made by neural systems to which the neuron projects). For example, a neuron responding to external inputs that has an average response rate of 100 spikes/s can only represent information accurately when the largest input bandwidth is less than about 10 Hz. For the auditory system, this frequency range is infrasonic; neural *groups* would seem essential for representing signals that can be heard.

Depending on how the "numbers work out", this conclusion would seem to apply to other neural systems as well.

## Notes

1. For technical reasons, we cannot declare such renewal processes to be stationary: having an intensity not depend explicitly on $t$ is necessary for stationarity, but not sufficient [Cox and Lewis, 1966, Ch. 4]. In most practical cases, this detail does not impact data analysis.

2. The unit-step function has the definition $u(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$.

3. More precisely, a valid hazard function's integral over zero to infinity must be infinite.

4. The term *regular* used here must not be confused with the same word used in the definition expressed by Eq. (5). There, *regular* refers to the proportionality of event occurrence probability to $\Delta t$, a completely different concept than the usage here of interevent interval variability. Which term applies should be clear from context. In addition, the current usage is more prevalent in neuroscience; the previous appears rarely.

5. As an example of a density that can yield a coefficient of variation greater than one, consider the log-normal random variable. Its pdf has the form $p_X(x) = \exp\{-((\log x - m)/\sigma)^2/2\}/\sqrt{2\pi\sigma^2 x^2}$ for $x$ positive. Here, $C[x] = (e^{\sigma^2} - 1)^{1/2}$.

6. Note that exponentially distributed random variables $E$ can be generated from uniform ones $U$ (distributed over $(0, 1]$) by $E = -\ln U$.

7. For the discrete memoryless channel model and the additive Gaussian noise communications channel model so prevalent in the communications literature, cutoff rate and capacity are equal.

8. $\mathcal{E}[\lambda^2] = \int_{\lambda_{\min}}^{\lambda_{\max}} \lambda^2 p(\lambda)\, d\lambda \leq \lambda_{\max} \int_{\lambda_{\min}}^{\lambda_{\max}} \lambda p(\lambda)\, d\lambda = \lambda_{\max}\bar{\lambda}$. Thus, $\mathcal{V}[\lambda] = \mathcal{E}[\lambda^2] - \bar{\lambda}^2 \leq \lambda_{\max}\bar{\lambda} - \bar{\lambda}^2$.

9. This normalization is chosen so that $\sum_l \text{INT}(l)\delta = 1$.

10. This tradeoff suggests an alternative method for estimating the interval histogram: vary the binwidth so that (roughly) a constant number of intervals fall into each bin. We sample more finely (smaller binwidth) where the interval distribution is large, less so where it is small. This approach, known as the nearest neighbor method (Devroye, 1987), does not significantly improve statistical accuracy.

11. Represent a positive-valued random variable as its mean plus a zero-mean perturbation: $X = m + \epsilon$. The variance of $X$, $\sigma^2$, thus equals the variance of $\epsilon$. When coefficient of variation of $X$ is small, $1/X = \frac{1}{m} \cdot \frac{1}{1+\frac{\epsilon}{m}} \approx \frac{1}{m} \cdot (1 - \frac{\epsilon}{m}) = \frac{1}{m} - \frac{\epsilon}{m^2}$. Thus, the variance of the reciprocal approximately equals $\sigma^2/m^4$,

and the coefficient of variation equals $\sigma/m$. Intuitively, a small percentage error equates to an equal percentage error in the reciprocal.

12. The cat lateral superior olive is an anatomically homogeneous nucleus located in the brainstem. Depending on definitional details, cells in this nucleus are second or third order neurons, receiving inputs that originate from both ears.

# References

Adrian ED (1928) The Basis of Sensation. Christopher, London.

Bialek W, Rieke F, de Ruyter, van Steveninck RR, Warland D (1991) Reading a neural code. *Science* 252:1852–1856.

Brémaud P (1981) Point Processes and Queues. Springer-Verlag, New York.

Cover TM, Thomas JA (1991) Elements of Information Theory. Wiley, New York.

Cox DR (1962) Renewal Theory. Methuen, London.

Cox DR, Lewis PAW (1966) The Statistical Analysis of Series of Events. Methuen, London.

Davis MHA (1980) Capacity and cutoff rate for Poisson-type channels. *IEEE Trans. Info. Th.* IT-26:710–715.

Devroye L (1986) Non-uniform Random Variate Generation. Springer-Verlag, New York.

Devroye L (1987) A Course in Density Estimation. Birkhauser, Boston.

Fienberg SE (1974) Stochastic models for single neuron firing trains. A survey. *Biometrics* 30:399–427.

Gaumond RP, Molnar CE, Kim DO (1982) Stimulus and recovery dependence of cat cochlear nerve fiber spike discharge probability. *J. Neurophysiol.* 48(3):856–873.

Gerstein GL, Kiang NYS (1960) An approach to the quantitative analysis of electrophysiological data from single neurons. *Biophysical J.* 1:15–28.

Hawkes AG (1971) Spectra of some self-exciting mutually exciting point processes. *Biometrika* 58:83–90.

Holden AV (1976) Models of the Stochastic Activity of Neurones. Vol. 12 of Lecture Notes in Biomathematics. Springer-Verlag, Berlin.

Johnson DH (1974) The Response of Single Auditory-Nerve Fibers in the Cat to Single Tones: Synchrony and Average Discharge Rate. Ph.D. thesis, MIT, Cambridge, MA.

Johnson DH (1978) The relationship of post-stimulus time and interval histograms to the timing characteristics of spike trains. *Biophysical J.* 22:413–430.

Johnson DH (1980) The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J. Acoust. Soc. Am.* 68(4):1115–1122.

Johnson DH, Swami A (1983) The transmission of signals by auditory-nerve fiber discharge patterns. *J. Acoust. Soc. Am.* 74:493–501.

Johnson DH, Tsuchitani C, Linebarger DA, Johnson M (1986) The application of a point process model to the single unit responses of the cat lateral superior olive to ipsilaterally presented tones. *Hearing Res.* 21:135–159.

Kabanov YM (1978) The capacity of a channel of the Poisson type. *Theory Prob. and Applications* 23:143–147.

Kiang NYS, Watanabe T, Thomas EC, Clark LF (1965) Discharge Patterns of Single Fibers in the Cat's Auditory Nerve. MIT Press, Cambridge, MA.

Kumar AR, Johnson DH (1990) A distribution-free model order estimation technique using entropy. *Circuits, Systems, and Signal Processing* 9:31–54.

Kumar AR, Johnson DH (1993) Modeling and analyzing fractal point processes. *J. Acoust. Soc. Am.* 93:3365–3373.

Lawrence AJ (1970) Selective interaction of a stationary point process and a renewal process. *J. Appl. Prob.* 7:483–489.

Li J, Young ED (1993) Discharge-rate dependence of refractory behavior of cat auditory-nerve fibers. *Hearing Res.* 69:151–162.

Linebarger DA, Johnson DH (1986) Superposition models of the discharge patterns of units in the lower auditory system. *Hearing Res.* 23:185–198.

Mark KE, Miller MI (1992) Bayesian model selection and minimum description length estimation of auditory-nerve discharge rates. *J. Acoust. Soc. Am.* 91:989–1002.

Miller MI (1985) Algorithms for removing recovery-related distortion from auditory-nerve discharge patterns. *J. Acoust. Soc. Am.* 77:1452–1464. Also see Erratum. *J. Acoust. Soc. Am.* 79:570, 1986.

Moore GP, Perkel DH, Segundo JP (1966) Statistical analysis and functional interpretation of neuronal spike data. *Ann. Rev. Physiol.* 28:493–522.

Ozaki T (1979) Maximum likelihood estimation of Hawkes' self-exciting point processes. *Ann. Inst. Statist. Math.* 31, Part B:145–155.

Rieke F, Warland D, Bialek W (1993) Coding efficiency and information rates in sensory neurons. Europhysics Letters 22:151–156.

Rodieck RW, Kiang NYS, Gerstein GL (1962) Some quantitative methods for the study of spontaneous activity of single neurons. *Biophysical J.* 2:351–368.

Sampath G, Srinivasan SK (1977) Stochastic Models for Spike Trains of Single Neurons. Vol. 16 of Lecture Notes in Biomathematics. S Levin, Ed. Springer-Verlag, New York.

Shamai (Shitz) S, Lapidoth A (1993) Bounds on the capacity of a spectrally constrained Poisson channel. *IEEE Trans. Info. Th.* 39:19–29.

Siebert WM, Gray PR (1963) Random process model for the firing pattern of single auditory neurons. In *MIT Quart. Prog. Rep.* 271:241–245. MIT Res. Lab. Electronics.

Snyder DL (1975) Random Point Processes. Wiley, New York.

Snyder DL, Rhodes IB (1980) Some implications for the cutoff-rate criterion for coded direct-detection optical communication systems. *IEEE Trans. Info. Th.* IT-26:327–338.

Ten Hoopen M, Reuver HA (1965) Selective interaction of two recurrent processes. *J. Appl. Prob.* 2:286–292.

Tsao YH (1984) Tests for stationarity. *J. Acoust. Soc. Am.* 75:486–498.

Zacksenhouse M, Johnson DH, Williams J (1995) Computer simulation of LSO discharge patterns. Unpublished manuscript.