Quantifying Neural Coding of Event Timing

Demetris S. Soteropoulos and Stuart N. Baker

Institute of Neuroscience, Newcastle University, UK.

Running title: Neural Coding for Event Timing

Address for correspondence:

Prof Stuart Baker
Institute of Neuroscience,
Newcastle University,
Henry Wellcome Building,
Medical School,
Framlington Place,
Newcastle upon Tyne.
NE2 4HH.
UK.
E-mail: stuart.baker@ncl.ac.uk
Phone: +44 191 222 8206
Fax: +44 191 222 5622

For submission to
Journal of Neurophysiology
Innovative Methodology

Copyright © 2008 by the American Physiological Society.
Abstract

Single neuron firing is often analyzed relative to an external event, such as successful task performance, or the delivery of a stimulus. The peri-event time histogram (PETH) examines how - on average - neural firing modulates before and after the alignment event. However, the PETH contains no information about the single-trial reliability of the neural response, which is important from the point of view of a target neuron. In this paper, we propose the concept of using the neural activity to predict the timing of the occurrence of an event, as opposed to using the event to predict the neural response. We firstly estimate the likelihood of an observed spike train, under the assumption that it was generated by an inhomogeneous gamma process with rate profile similar to the PETH shifted by a small time. This is used to generate a probability distribution of the event occurrence, using Bayes’ rule. By an information theoretic approach, this method yields a single value (in bits) which quantifies the reduction in uncertainty regarding the time of an external event following observation of the spike train. We show that the approach is sensitive to the amplitude of a response, to the level of baseline firing, and to the consistency of a response between trials, all of which are factors that will impact a neuron’s ability to code for the time of the event. The technique can provide a useful means of determining which of several behavioral events a cell encodes best, as well as permitting objective comparison of different cell populations.

Introduction

Many studies align neuronal activity to different stimuli or behavioral events in an attempt to gauge the neuron’s response. The basis of the peri-event time histogram (PETH; Gerstein & Kiang, 1960) is that the average across many trials will be a reasonable representation of the neural response, as the averaging will reduce any activity not time-locked to the event. Subsequently, PETH features such as mean rate, peak rate and time of peak response can then be compared for different stimuli.

More recent work has emphasized that neuronal activity can vary from trial-to-trial in many ways (Baker and Gerstein 2001; Croner et al. 1993; Fortier et al. 1993; Gur and
Variation in *latency, amplitude* and *shape of response* can all lead to erroneous interpretation of trial-averaged measures (Brody 1999; 1998). Several reports have proposed methods which mitigate the effects of such trial-to-trial variability or allow measurements from single trials (Baker and Gerstein 2001; Grun et al. 2003; Nawrot et al. 1999; Nawrot et al. 2003; Pauluis and Baker 1999). For the PETH, variability will tend to smear the response, making it appear smaller and broader than the actual single trial response.

However, trial-to-trial variability is a physiological phenomenon that is interesting in its own right; several reports have attempted to measure this using the Fano factor (Andolina et al. 2007; Fortier et al. 1993; Gur and Snodderly 2006; Kara et al. 2000; Vogel et al. 2005). Variability in response latency and/or amplitude may provide important clues to what a cell is coding for (DiCarlo and Maunsell 2005). Highly reliable firing, time-locked to a behavioral event, may signal the coding of that event. Greater variability may simply be noise, but more likely represents the encoding of another variable; this may not be experimentally observable (an internal ‘hidden state’). It is possible that a neuron could fire reliably following one stimulus, but not for others. This would be important in distinguishing what the cell encodes; the most reliable response would not necessarily be the largest. Trial averaged measures such as the PETH discard valuable information about the trial-to-trial response variability.

An important question is what impact this trial to trial variability has in the information that single neurons can convey to their targets regarding the timing of an external event?

We suggest here a method to measure this information. The method proceeds by comparing single trial responses with the ‘archetypal’ response, estimated by the trial-averaged response in the PETH. A cell is assigned an information value (in bits), representing how well it encodes the timing of a behavioral event or stimulus. This singular information value will be affected by a host of factors such as variability in response latency and response amplitude from trial to trial, as well as response size, baseline rate and the intrinsic variability of neural spiking.

Reliable neurons, which have single trial responses similar to the PETH, return high information values. Rather than attempting to account for any latency or amplitude
changes from trial-to-trial, these are treated as genuine features of the cell’s response as such variability degrades the cell’s ability to represent the time of the alignment event, and leads to smaller information values. The method permits a novel interpretation of cell firing, and can reveal which of several events a neuron codes best.

**Methods**

**General Concept**

In this approach, we invert the usual interpretation of the PETH. Rather than ask when spikes occur relative to a behavioral event, we consider what is the probability that the behavioral event occurred at a given time during the observation window defined by the PETH. Suppose $\delta \tau$ is time within this observation window; typically, $\delta \tau=0\text{ms}$ will be the actual time the event is known to occur by the experimenter. Solely from observing the spike train on a single trial $\mathcal{S}P$, we wish to estimate the probability of event occurrence at time $\delta \tau$, $P(\delta \tau \mid \mathcal{S}P)$.

Direct estimation of this probability is difficult. Much easier is the problem of estimating $P(\mathcal{S}P \mid \delta \tau)$. To do this, we firstly shift the PETH by $\delta \tau$ (Fig. 1B) and then assume that the observed single trial spike train resulted from this underlying rate modulation. We use a model for the spike generating process to estimate the likelihood of the observed spike train, given this rate (the details of the model will be addressed in subsequent sections). This process is repeated for all possible values of $\delta \tau$, producing the likelihood plot $P(\mathcal{S}P \mid \delta \tau)$ of Fig. 1C.

Bayes’ rule can be applied to this situation; it states:

$$P(\delta \tau \mid \mathcal{S}P) = \frac{P(\mathcal{S}P \mid \delta \tau).P(\delta \tau)}{P(\mathcal{S}P)}$$  \hspace{1cm} (1)

Assuming we have no prior information about either the expected shift, or the observed data, $P(\delta \tau)$ and $P(\mathcal{S}P)$ can both be assumed constant. Equation 1 then reduces to
The constant of proportionality can be determined by the requirement that a probability distribution encloses unit area.

Figure 1C shows the shift likelihood distribution for a single trial. The whole procedure is repeated for all trials, producing a family of shift distributions (Fig. 1D, black). Averaging across trials yields an estimate of the mean shift likelihood distribution (Fig. 1D, grey line).

This averaged shift likelihood curve indicates how well the single neuron encodes the time of the behavioral event used to align the PETH. A sharply peaked curve indicates that an external observer could use the spiking to predict the event occurrence with high fidelity. A broad distribution indicates that this cell is of little use in determining the occurrence of the event. We can summarize this by calculating the entropy of the distribution:

\[
P(\delta \tau | SP) \propto P(SP | \delta \tau) \tag{2}
\]

\[
S = -\sum_{\delta \tau} P(\delta \tau | SP) \log_2 P(\delta \tau | SP) \tag{3}
\]

The horizontal bars above \( P(\delta \tau | SP) \) indicate that the mean shift probability distribution is to be used. The information which the cell encodes about the behavioral event can be estimated from the difference between this entropy and that of the uniform prior on \( P(\delta \tau) \). Assuming a maximum shift of \( \pm 300 \) ms and a 1 ms shift resolution, the entropy of the prior is just \( \log_2 (601) \).

Below we discuss possible methods for calculating \( P(SP | \delta \tau) \), the likelihood of the single trial responses given the PETH. These are not presented as the only possibilities, but rather as reasonable options. If other estimates with better performance and fewer limitations are devised in future, they can easily be integrated with the general concept described above.
(A) Spike Density Likelihood (Poisson Process)

A simple way to estimate $P(SP|\delta \tau)$ is to assume that the spike train results from an inhomogeneous Poisson process, with rate parameter equal to the shifted PETH. For the single trial under consideration, assume that $\lambda_{\delta t}(i)$ represents the probability of spiking per bin estimated from the PETH shifted by $\delta t$. The ends of the shifted PETH were padded by a constant rate, equal to the mean rate of the 50ms-long section of the PETH at the end being padded.

In order to improve the signal-to-noise ratio of the PETH, we calculated it using a modified procedure similar to previous work (Pauluis and Baker 1999; Richmond et al. 1987; Schwartz 1992). The instantaneous firing rate was calculated for each trial as the reciprocal of the ISI, using 1 ms bins; this was averaged across trials. The PETH was smoothed by convolution with a normalized Gaussian kernel (width parameter 10ms; the same methods were used to calculate the PETH for all the methods described below).

If spikes occurred in bins $i \in S$, then the likelihood of that single trial can be estimated by:

$$P(SP | \delta t) = \prod_{i \in S} \lambda_{\delta t}(i) \cdot \prod_{j \notin S} (1 - \lambda_{\delta t}(j))$$  \hspace{1cm} (4)

We note in passing that, like other methods relying on likelihood, this product can be problematic to calculate for large number of bins. The value can become so small as to be outside the floating point range of standard computers. This is easily solved by taking the logarithm of the individual terms, and then summing:

$$\log(P(SP | \delta t)) = \sum_{i \in S} \log(\lambda_{\delta t}(i)) + \sum_{j \notin S} \log(1 - \lambda_{\delta t}(j))$$  \hspace{1cm} (5)
The scaling of $P(SP \mid \delta t)$ is unimportant, as these values will eventually be rescaled to enclose unit area forming $P(\delta t \mid SP)$. Once $\log( P(SP \mid \delta t) )$ is found for all $\delta t$, we therefore add an arbitrary constant to all values, before taking anti-logs. This ensures that the results remain inside the allowed floating point range.

Figure 2 shows an example of the application of this method. From the shifted spiking probability curve $\lambda_{\delta t}(i)$ (Fig. 2A), the per-bin probability is found (Fig. 2B). Forming the product of these probabilities yields the data likelihood for that shift. Compiling for all shifts and normalizing the curve to enclose unit area produces the shift likelihood distribution for that trial (Fig. 2C); averaging across trials yields the mean shift likelihood (Fig. 2D, grey).

The most attractive feature of this procedure is its simplicity and ease of calculation. However, it does not take into account the dependence of spike probability on the spiking history, such as the absolute refractory period. If a spike occurs in bin $i$, the probability of spiking in bin $i+1$ assuming Poisson spiking is simply $\lambda(i+1)$, whereas it is zero for real neuronal data.

(B) Inter-Spike Interval Likelihood

The following section describes a method for estimating $P(SP \mid \delta t)$ based not on the timing of the spikes, but on the observed interspike intervals (ISIs). We assume that the underlying spiking process can be modeled as a gamma distribution (Baker and Gerstein 2001; Baker and Lemon 2000; Stein 1965). This is a more realistic representation of real neural firing than a Poisson process, as it shows a refractory period and can be a better match for observed ISI distributions. The probability of observing an interval $ISI$ given mean firing rate $R$ and gamma order parameter $a$ is:

$$\gamma(I) = \left( aR \right)^a \frac{1}{(a-1)!} ISI^{a-1} e^{-IaR}$$  \hspace{1cm} (6)$$

A gamma process with order $a=1$ is a Poisson process. For a given experimental neuron, the most likely order can be estimated by using the method described briefly below and in Baker & Lemon (2000).
To calculate the likelihood of the spike train data from a single trial \( P(SP \mid \delta t) \), we firstly take the PETH and shift it by a time \( \delta t \) relative to the single-trial spike train. Figure 3ABC shows an example PETH at three different latency shifts (Fig. 3A: -100ms, 3B: 0ms, 3C: +100ms). The PETH is then averaged over the period spanned by each ISI, yielding the mean rate pertaining to that interval \( R_i \) (thick lines in Fig. 3ABC). The ISIs for trial no.1 are shown in Fig. 3ABC by the dashed vertical lines spanning the three plots, and within each ISI, the horizontal lines represent the mean PETH rate during that interval. Just by inspecting the plots the mismatch at the shifted latencies (-100ms, +100ms) is obvious; in both cases short intervals are seen with low rates and large intervals with high rates. Considering the seventh ISI as an example (grey shading Fig. 3C), the corresponding mean PETH rate \( (R_i) \) is very different for the 3 latency shifts (-100, 0 and 100 ms) at 18, 58 and 20Hz respectively. Figure 3DEF shows the three different ISI distributions (gamma order 4) calculated for these underlying rates. The observed interval (15 ms) is much less probable for mean rates of 18 or 20Hz compared to the distribution for a mean rate of 58Hz.

For each interval, the combination of a rate and order estimate defines a probability distribution of intervals, from which the observed interval is drawn. These can be used to quantify the likelihood of observing each interval; the product of these likelihoods yields the data likelihood for that trial, given a PETH shift of \( \delta t \):

\[
P(SP \mid \delta t) = \prod_i \gamma(ISI_i, a, R_i)
\]

This process was repeated for shifts \( \delta t \) from -300 to 300 ms in steps of 1 ms. The shift likelihood distribution for trial 1 is shown in Fig. 3G. The entire procedure was repeated for all available trials (Fig. 3H). Because the method works on inter-spike intervals, it cannot be used when a trial has fewer than two spikes. The shift likelihood curve for such trials was assigned as a uniform distribution. The single-trial shift likelihood distributions were then averaged to give the mean shift likelihood distribution (light grey line in Fig. 3H).

**(C) Spike Density Likelihood (Gamma Process)**

The two methods described above are complementary. The Poisson spike density method is capable of responding to rapid changes in spike density; however, it cannot
account for the dependence of spiking probability on spiking history. By contrast, the
interval density method naturally accounts for the complexity of interval statistics, but it cannot take into account rate changes faster than the length of an interval.

The final method which we propose depends on spike density, but also takes account of spike history. We start by considering the death rate as a function of time since the last spike. This is the probability that a spike will occur at a particular ISI, given that no spike has occurred up until this time (Matthews 1996). The death rate (Fig. 4C) is calculated by dividing the ISI distribution (Fig. 4A: three gamma ISI distributions with a mean ISIs of 25, 50 and 100ms) by the inverse of the cumulative probability function of the ISI distribution (survival plot; Fig. 4B). The marks on each line in Fig. 4C show the probability of a spike occurring after a time equal to the mean ISI for each distribution has passed.

We are interested in using this method for non-homogeneous gamma processes, where the rate will not necessarily be constant during the ISI. To accommodate changes in rate that fall between two spikes, we suggest expressing the death rate not as a function of time after the previous spike, but as a function of the integral of the spike density since the last spike:

\[
P(spike(t + T)) = f \left( \int_{t}^{t+T} \lambda(\tau) d\tau \right)
\]

where \(t\) is the time of the last spike (Barbieri et al. 2001; Kass and Ventura 2001). When rate is constant, this is just a rescaling of the abscissa (Fig. 4D). However, if rate changes, making the spiking probability a function of the accumulated density naturally incorporates the rate changes. Integrating the rate is analogous (in a very simplistic way) to the way the cell membrane integrates arriving potentials until a threshold is reached and the cell fires. Such a non-leaky integrate and fire neuron is an over simplified model of neural firing, but can generate firing patterns matching that of a gamma process (Tuckwell 1988)
Processes with different rates will obviously have a different probability of a spike occurring after the rate integral reaches the same value (as indicated by the marks on the lines in Fig. 4D). However, for the same order of the gamma process, the functions plotted in Fig. 4D differ only by a scaling factor. If they are rescaled by dividing by the mean rate, they overlie perfectly (Fig. 4E). Gamma processes with different orders have functions with different shapes, some examples of which are shown in Fig. 4F.

We thus calculated the probability of spiking at time $T$, when the previous spike occurred at time $t$, as:

$$P(\text{spike at } (t+T)) = f\left(\int_t^{t+T} \lambda(\tau)d\tau\right) \cdot \lambda(t+T)$$

(9)

where $f$ is calculated by the double-scaling of the death rate for a gamma process described above, and $\lambda(t)$ is estimated from the shifted PETH. The likelihood $P(\text{SP} | \delta t)$ can then be estimated by multiplying the probabilities of the observed data (analogous to Equation 4).

Figure 4GHI shows the performance of the spike density method for three different shifts $\delta t$ for a single trial. Figure 4G1H1I1 are the corresponding spike densities, while Fig. 4G2H2I2 represent the accumulated spike density since the last spike. Figure 4G3H3I3 is a plot of the probability of observing a spike at every time point, while Fig. 4G4H4I4, is the probability of the observed data in each bin (i.e. $P(\text{spike})$ when a spike occurred and $1-P(\text{spike})$ when no spike occurred). The grey rectangles are to focus the attention for a region of increased rate for the three shifts. As can be seen in the case of the 100ms shifts in either direction (Fig. 4GI) the probability of spiking given the shifted spike density is low because the spikes are now over an area of low spike density. As the likelihood of each $\delta \tau$ is a product of the probabilities of the observed data, this will cause $\delta \tau=100\text{ms}$ and $\delta \tau=100\text{ms}$ to have a much lower likelihood than when $\delta \tau=0\text{ms}$.
Bias Estimation and Significance Testing

In order to estimate a P<0.05 significance level for the measured information about event timing, a shuffled dataset (Fig. 5B) was generated from the original data (Fig. 5A). The ISIs of each trial were shuffled; the first spike was chosen randomly so that

\[ 0 \leq ISI_0' \leq ISI_0 + ISI_{last} \]  

where \( ISI_0' \) is the first incomplete interval of the shuffled trial, and \( ISI_0, ISI_{last} \) are the first and last incomplete intervals of the real trial (‘incomplete intervals’ are the ISIs which fell only partially within the timeframe of the actual trial, i.e. the time period before the first spike, and after the last).

The timing information for the shuffled data was calculated as described above (Fig 5C). The entire process was repeated 100 times, yielding a distribution of the information which would be measured if the cell coded nothing above chance levels about the occurrence of the behavioral event (Fig. 5D). If the information value found from unshuffled data (dotted line, Fig. 5D) was larger than 95/100 of the values from shuffled data then the cell can be considered to convey significant information about the event.

Information is defined as the difference in entropy before and after an observation is made. Our estimate of entropy before observation of the spike train is not biased, since we calculate it analytically as \( \log_2 \) (number of bins). However, our estimate of entropy after observation is biased: due to finite sampling, the experimental estimate of shift likelihood will not be exactly equal to a uniform distribution, even if the cell codes nothing about the event timing. As a result, our information estimates will be biased upwards (Panzeri and Treves 1996). In order to correct for this, the mean information derived from the shuffled data should be subtracted from the information determined from the original data. For the rest of the paper, \( I \), will be used to refer to the information estimate without this bias subtraction. However for the majority of the simulations presented here, as the number of trials was very large this correction
would have made negligible difference. In experimental data with smaller numbers of trials, however, it could be necessary. A small number of shuffles for bias estimation (n~10 to 20) should be sufficient; whereas around 100 are required for significance estimation.

Another source of bias arises from the fact that the estimate of the underlying rate change comes from the PETH, which is calculated from summing the single trial data. If only small numbers of trials are available, chance fluctuations in cell spiking will produce noise in the PETH, to which the spike trains will necessarily best align with zero-lag. This will lead to a shift-probability distribution with a weak central peak, even in the absence of task related firing, producing an elevated information estimate. In order to avoid this effect, in all results reported here the PETH used to estimate the likelihood for the n’th trial was calculated using all trials except the n’th.

**Optimal order estimation**

Both the inter-spike interval method, and the gamma process spike density method, require an estimate to be made of the order of the gamma process using to model the spike train. For experimental data, we have here followed the procedure suggested by Baker & Lemon (2000) to do this. Briefly, the instantaneous firing was estimated by convolving the spike train with a Gaussian kernel with width parameter $\sigma$ equal to the modal ISI. Inhomogeneous gamma processes were then generated using this rate, but with different orders. The resulting ISI histograms were compared with the ISI distribution of the original data, and the one giving the best fit chosen as the optimal order to describe that cell. Some problems with this method have been identified (Gerstein 2004), which may lead to errors in order estimation. This is most likely when the firing rate is changing; to minimize such errors, data for order fitting should be taken from a period where the rate is relatively stationary.

It is known that the regularity of neural spiking can modulate during task performance; this can be quantified using the measure of irregularity developed by (Davies et al. 2006). Such regularity changes imply that estimates of order taken from one part of the task will not necessarily be valid for another. In the present paper, we
do not account for this – which would be possible using the methods of Davies et al (2006) - although we do quantify the effect that errors in assumed order will produce.

**Simulated Data**

Simulated data of known properties were generated to assess the behavior of the analysis method. In these simulations, the single trial firing rate increased in a stepwise fashion from a baseline rate to a higher level for a defined time, before returning to the baseline. Using this rate profile, inter-spike intervals were generated from a gamma distribution of a particular order, following the decimation method described in Baker and Gerstein 2001. The order of the gamma was fixed at $a=4$, and 100 trials were generated, unless otherwise stated.

**RESULTS**

**Method Comparison/Limitations**

Figure 6A compares the information values obtained using the three methods to estimate $P(\text{SP} \mid \delta t)$ described above. These have been applied to data simulated with step increases in rate (20 to 60 Hz, duration of step 200ms) and various values of the gamma order parameter. For gamma order 1 (corresponding to Poisson spike trains), all methods produced the same information value. However, as the gamma order increased, so did the information estimated using the two methods which took account of the non-Poisson nature of the firing. This increase is a desirable property: it reflects the increased information about event time which can be carried by a regular spike train compared with an irregular one. By contrast, the method which assumed Poisson statistics yielded the same information value throughout, underestimating the true value when the assumption of Poisson firing was violated. Since experimental spike trains are typically more regular than a Poisson process, this makes the Poisson spike density method inapplicable to experimental data. We will therefore not consider it any further.

The gamma spike density and gamma interval methods produced very similar results in most situations which we have examined (Fig. 6A and other data not shown). Differences can occur, however, in situations where the rate modulates over a similar
timescale to the ISI duration. An example of such a circumstance is shown in Fig. 6B, as a raster display with PETH overlain. The rate modulates from 0 to 10Hz in an oscillatory fashion, with a 4 Hz cycle time. In such cases, averaging the rate over an ISI fails to capture the time dependence of the spike density. Figure 6C shows the shift likelihood distributions for the gamma spike density likelihood method and Fig. 6D for the ISI likelihood method. The ISI method yielded a substantially lower information value (0.11 bits) compared to the gamma spike density method (1.85 bits). This occurred because the ISI method necessarily averages the spike density over each ISI. Rapid fluctuations in density are thus smoothed out, and the cell is assigned an erroneously low value of information.

Because the ISI method averages rate over the ISI duration, it is less sensitive to noise. This is illustrated in Fig. 6E which plots the $I_r$ values measured from simulated data with no modulation in rate – $I_r$ values would ideally be zero in this case, although a small bias will be present as described in Methods. Although both methods produce small I values, they are consistently smaller for the ISI method. However, this difference is so small that it is unlikely to be of importance for experimental data.

The remainder of this paper will focus on the performance of the gamma spike density method, as it both takes account of the non-Poisson nature of neural firing and also can account for rapid changes in spike density.

**Bias**

With finite data, the method presented here will lead to a non-zero raw information estimate $I_r$, even if a cell does not modulate its activity systematically around a task event. Any noise fluctuations in the shift probability estimate will lower its entropy relative to a uniform distribution, leading to a positive bias in the information. We have used a shuffling method to estimate and correct for this. The dependence of the bias on various parameters is illustrated in Fig. 7; these curves were calculated from simulated data with no actual modulation in rate.

Figure 7A presents the effect of smoothing kernel width and firing rate (10, 20 and 40Hz) on bias when the trial numbers are low (10 trials). Unsurprisingly, at little or no
smoothing the bias is high and decreases as the smoothing increases. The arrowhead in Fig. 7A is the maximum ordinate scale of Fig. 7B, which is based on PETHs with 50 trials. Similar curves as in Fig. 7A are seen, but with this larger number of trials even with little or no smoothing the bias is much smaller. The arrowhead in Fig. 7B shows the maximum ordinate scale of Fig. 7C, which is based on PETHs with 100 trials. The same trend is seen for bias vs smoothing as in Fig 7AB, but the value of the bias is now even smaller. The decrease in bias with increased smoothing/trial number is because random fluctuations in the PETH (arising from the stochastic nature of spiking) are decreased.

In the simulations of Fig. 7ABC, the number of trials and the noise fluctuations in the PETH are interdependent. To explore the effect of noisy fluctuations in the PETH in more detail, we first simulated data with a constant rate. The known flat rate profile was then corrupted by the addition of different levels of Gaussian noise. Figure 7D shows how noise in the PETH affected the bias estimate for different baseline rates (data simulated as gamma process order 4, 50 trials, maximum value of $\delta t=100\text{ms}$). As the level of noise increased so did the bias, although for the same noise CV, the bias was greater for higher baseline rates. This can be explained by referring to Fig. 4D. For the same gamma order, as the mean rate increases, the death rate plot becomes steeper around the region of unitary rate integral. This means that small changes in the rate integral will result in greater changes in the probability of observing a spike for higher rates at different $\delta \tau$. Random fluctuations in the PETH will thus result in greater modulation in the shift likelihood distribution for higher rates and hence a higher bias.

The regularity of a spike train will also affect the bias estimate. This is explored in Fig. 7EFG, where regularity is systematically changed by altering the order of the gamma process used to simulate the data. As spiking regularity increases so does the bias, although for 100 trials the absolute size of these effects are small. This can be explained by referring to Fig. 4F. As gamma order increases, and firing becomes more regular, the death rate plot becomes steeper around the region of unitary rate integral. This means that for the same change in the rate integral, there will be a greater change in the probability of observing a spike as $\gamma$ increases. For a given PETH, a spike train
of higher order will show greater changes in likelihood at different $\delta \tau$ thus having lower entropy, which will result in a higher bias. The arrowhead in Fig. 7E is the maximal ordinate scale of Fig. 7F and similarly the arrowhead in Fig. 7F is that for Fig. 7G.

At first sight, there may appear to be a contradiction between Fig. 7EFG and Fig. 7D: in Fig. 7D, bias is largest for high baseline rates, whereas in Fig. 7EFG, bias is greatest at low rates. The explanation for this is that, for the same number of trials, the noise fluctuations in the PETH (measured by CV of bin counts) are greater for lower baseline rates. This effect combines with that shown in Fig. 7D to produce the relationships seen in Fig. 7EFG.

For the remainder of the Results section we will only show the $I_r$ values, based on estimates on 100 trials. Ignoring the bias correction greatly reduces the computation time, and as made clear by Fig. 7C the bias is in fact negligible for this trial number.

**Smoothing Kernel**

A key aspect of our method is that the PETH is used as an estimate of the underlying rate modulation. By smoothing the PETH, a better estimate can be produced. Increasing the width of the smoothing kernel is advantageous as it reduces spurious variability in the PETH. This reduces the bias in the information measures (Fig. 7ABC). However, smoothing also blurs any genuine relationship between the cell’s firing and the behavioral event, reducing the measured information. Figure 8A shows how increased smoothing alters the $I_r$ value. In these simulations, the rate increased to double the baseline rate during a 200 ms-long step. The information value dropped with increasing kernel width; the effect was greatest for high rates. In practice, a value for smoothing will need to be chosen heuristically to compromise between over-smoothing (reducing information, Fig. 8A) and under-smoothing (increasing bias, Fig. 7ABC).

**Maximum Shift ($\delta \tau$)**
One arbitrary parameter of the present method is the maximum shift $\delta \tau$ used. As the maximum shift is increased, so the entropy of the uniform prior will rise (upper dotted line in Fig. 8B). If a cell codes for a behavioral event, then in most circumstances shifts above a certain duration will all have very low probability (there are some exceptions to this, however, e.g. oscillatory responses, Fig. 6B). The entropy of the mean shift probability distribution will therefore increase little with further increases in the maximum shift. This is shown by the other lines in Fig. 8B (200ms step in rate, baseline rate and rate during the step labeled on the curves), which rise less steeply at large shifts than the prior entropy. Since the information is the difference between the posterior shift distribution entropy and the entropy of the uniform prior, the measured information will rise with larger values of the maximum shift. For this reason, our method cannot give an assumption-free estimate of information, but depends on how uncertain event timing is initially assumed to be. Practically, comparison between different cell populations must use the same maximum shift to be valid.

**Step Magnitude**

Intuitively, the more that a cell changes its firing rate associated with an event, the more reliably it encodes the event’s timing. Figure 8C quantifies how the size of rate change alters the raw information measured using the present method. Each line shows results for a different baseline rate; the abscissa gives the ratio between the firing rate during a 200 ms event-locked step rate increase/decrease and this baseline. As expected, the information about the event increases with the size of the rate step for both rate increases and decreases. In the case of rate decreases, as the rate during the step approaches zero, the I value tends to level out. However, the absolute size of the rate step is not the only important factor. The points marked with circles in Fig. 8C both relate to rate increases of 30 Hz. However, a rise from 10 Hz to 40 Hz provides 1.74 times more information than when rate changes from 20 to 50 Hz (1.31 vs 0.75 bits respectively). This is because a 25 ms inter-spike interval (relating to a 40 Hz response) is less likely to occur during a 10 Hz firing rate epoch than a 20 ms interval (relating to a 50 Hz response) during a 20 Hz firing rate epoch. This shift likelihood distributions for the 10 to 40 Hz step are thus narrower, raising the information measured.
Figure 8C demonstrated that the information encoded about event time depends on the size of the unit’s rate increase. In experimental data, the amplitude of rate modulation is often non-stationary from trial-to-trial (Baker et al, 2001). It is interesting to know how much such rate variation will affect the information estimates. This is examined in Fig. 8D for step rate increases. Data were simulated by choosing the amplitude of the rate step for each trial from a normal distribution. The mean rate corresponded to that given in the figure key; the standard deviation of the normal distribution was chosen to yield a coefficient of variation from 0.1 to 0.5, which is plotted on the abscissa. Response amplitude variation reduced the measured information values; this decrease was more pronounced for higher firing rates.

If there were concern that response amplitude varies substantially between trials, it would be possible to extend our method by assuming that the spike density on a single trial is a scaled version of the PETH:

\[ \lambda(t) = \alpha \cdot PETH(t) \]  

(11)

The data likelihood could then be estimated as a two dimensional function of both shift \( \delta t \) and response amplitude scaling \( a \). The shift probability would then be given by the marginal, found by integrating the two dimensional distribution over \( a \). Whether this additional computational complexity is warranted will depend on the nature of the experimental data.

**Step Width**

The duration of rate modulation also influences how well a cell codes the timing of an event; this is investigated by Fig. 8E. The raw information rises for the four rate profiles investigated as the duration of the rate step increases, whether it is a step increase (doubling of baseline rate) or decrease (rate drops to zero). However, information levels off for widths around 150 ms, and there is then no further increase. So long as the response is wide enough unambiguously to include ISIs from the period of elevated rate (for rate increases), then increasing the number of these intervals does not substantially change the mean shift probability distributions. Similarly for rate step decreases, as long as the width of the response is comparable to ISIs from the baseline period further increases will have little effect. Essentially, our
method is sensitive to rate changes, and is not much affected by periods of constant rate in the PETH.

The effect of varying response duration from trial to trial is shown in Fig. 8F. Here, the mean response duration was 300 ms; the actual duration used for each trial was selected from a normal distribution, with coefficient of variation between 0.1 and 0.5 (shown on the abscissa). Variability in the response duration decreased the information measured; data with higher firing rates were greater affected.

**Spiking Regularity**

Figure 9 shows a series of simulations carried out to assess the effect of spiking regularity of the information values, and also what happens with this regularity varies from trial to trial, and when miscalculated.

To determine the effect of spiking regularity on event encoding, we measured the information from spike trains simulated with the same rate profiles but different orders of gamma process. In all cases, increased regularity (higher gamma order) led to higher information (Fig. 9A). This is because, as the regularity of the process increases, the intervals within the response region become less likely to arise from the baseline period and vice versa.

An important assumption of the present method is that the ISIs follow a gamma distribution. The appropriate order of this distribution must be determined from the spike train before the analysis can be performed. Figure 9B shows how the measured event information is affected by incorrectly estimating the gamma order. Each line on this plot relates to simulations carried out using a certain order of gamma process; the abscissa shows the order assumed for the analysis. Information is normalized by the information measured when the assumed order coincides with the actual order used to simulate the data. Overestimating the order led to a corresponding overestimate of information; likewise, assuming an order below the true value underestimated the information.
For experimental data, it is most likely that order will be underestimated. If rate is non-stationary, this will lead to higher ISI irregularity than expected if rate were constant. We recently published a method capable of estimating spiking regularity when rate is non-stationary (Davies et al. 2006). Although generally effective, in the presence of very rapid rate changes, or serially correlated ISIs, this can also overestimate the irregularity of discharge. Errors in order determination for experimental data are thus always likely to err towards assuming the spike train is more irregular than it actually is, leading to a lower assigned gamma order than the actual underlying value. This would result in an underestimate of the information carried by an experimental spike train. Our method is thus likely to yield conservative information estimates in the face of errors in order estimation.

To examine the effects of trial to trial variability in the spike regularity, PETHs with a rate step were generated in which the spiking regularity, determined by the order of the gamma process used to simulate the data, was different for each trial (but constant within a trial). The gamma order $a$ determines the coefficient of variation (CV) of the ISIs, according to

$$CV = \frac{1}{\sqrt{a}}$$

A mean desired gamma order $\bar{a}$ was selected, and the corresponding ISI coefficient of variation $\bar{CV}$ found. For a given trial, the ISI CV was selected from a normal distribution with mean $\bar{CV}$, and standard deviation $k\bar{CV}$. This was then transformed back to a gamma process order, which was used to simulate data for that trial. The parameter $k$ is the coefficient of variation of the ISI CV, and was varied from 0.1 to 0.5

Figure 9CD shows how the measured information is affected by trial to trial variability in regularity. Data are shown for simulations with either a low (order 2, Fig. 9C) or high (order 6, Fig. 9D) average regularity. Each line shows the results for a different rate profile; the abscissa of these plots is the parameter $k$ referred to above. In all cases, the information estimate was robust to across-trial variation in spiking regularity.
So far, we have assumed that even if spiking regularity varies from trial to trial, it remains constant within a single trial. It has recently been shown, however, that regularity can modulate in a task-dependent manner (Davies et al 2006). Figure 9EF investigates the sensitivity of the method to regularity changes within the trial. In these simulations, the baseline period was generated with a gamma order of 2 (Fig 9E) or 6 (Fig. 9F); the order during the rate step ranged from 1 to 16 in different simulations, but within a simulation did not vary from trial-to-trial. In all cases, information was estimated assuming that the order remained constant and equal to the baseline order throughout. Results for four rate profiles are shown. The measured information values increases monotonically as the response region becomes more regular.

**Latency Variability**

To encode an event effectively, a cell must respond in a fixed way from trial to trial. Figure 10 examines how variability in the latency of response from trial to trial affects the information measure. Data were simulated with a 200 ms rate step, in which the response latency in each trial was chosen at random from a normal distribution. The different lines of Fig. 10A show results for different baseline and response rates; in all cases, the rate doubled in the response region. As expected, information decreases with increasing jitter; this is more pronounced for the high rate responses, which had higher information to begin with. The circle on each line represents the latency standard deviation at which information had fallen to 50% of the value with no latency jitter. A latency standard deviation of 60 ms leads to at least a 50% fall in information for all the rate steps tested.

One expectation of the method is that in cases with trial to trial latency variability, the peak in the single trial shift probability distributions would be related to the response latency within the trial. Figure 10B shows a raster and PETH of a 20-40 Hz rate step with a latency standard deviation of 50 ms.

Figure 10C is a cluster plot of the actual response latency versus the maximum likelihood shift. There is substantial and significant (P<0.01, shuffle test) correlation
between the two. However, the variation in estimated shift was greater than in actual response latency (standard deviation 70 vs 50 ms), and the slope of the regression line was smaller than unity (0.75). This is due to variations in spike timing, which mean that spikes do not occur exactly at the moment of a rate change. Even with no trial-to-trial latency variation, maximum likelihood shifts show some dispersion (e.g. Figs 4H, 12D).

To explore this correlation further, rate steps of different amplitude (both increase and decrease) were simulated with a latency standard deviation of 60ms. The changes in correlation between estimated and actual response latency are plotted versus the size of the response in Fig. 10D. As expected, if there is no response (value 1 on the abscissa) the $r^2$ value is zero. As the difference between the baseline rate and the step rate increases, the $r^2$ value increases.

One assumption made by our method in its current form is that the PETH is a reasonable estimate of the single-trial response of a neuron. Where there is considerable latency variability from trial to trial, the PETH will be smeared compared with the actual single trial response. If the downstream decoder reading the neuron’s discharge had knowledge of this single trial response, it could obtain more information about event timing from the cell than would be estimated by our algorithm. Such differences are however likely to be small; for example, for the situation illustrated in Fig. 10B, the information estimate assuming the actual single trial response is 1.14bits, compared with 1.03 bits when using the PETH. Estimating the true single trial response profile would be a complex procedure for a downstream decoder, and it is by no means certain that the higher information value given by making this assumption is more valid than assuming that the trial-averaged response is representative.

Experimental Data

Figure 11 shows the results of the method applied to two real neurons recorded from the supplementary motor area (SMA) of a monkey performing a precision grip task
with the hand contralateral to the recording site. Full details of the task are given in a previous publication (Soteropoulos and Baker 2006; 2007). Briefly, the animal held both hands on home pad switches. A ‘go cue’ signaled that a movement should be made. The animal then lifted one hand from the home pad (the ‘reach’ event) and gripped the levers of a precision grip manipulandum between finger and thumb (the ‘squeeze’ event, indicating first lever movement). Following a hold period, the levers were released to obtain a reward.

This task is an interesting one in the present context, because three very different events (go cue, reach and squeeze) occur in close temporal proximity. This is illustrated in Fig. 11A, which presents histograms of two events compiled relative to the third. A neuron which modulates its firing relative to one of these events will necessarily also do so relative to the other two. It is therefore difficult to determine which event the cell is actually encoding.

The neuron shown in Fig 11B had a robust modulation in activity around all three events, as judged from the PETH and raster displays. Figure 11C plots the single trial and mean shift probability distributions for each event. In each case, there was a clear peak at zero shift, although this was visibly narrower when aligned to the squeeze event. Calculation of the information yielded values of 1.8, 2.27 and 2.87 bits relative to go cue, reach and squeeze events respectively. This cell therefore seems best related to the lever squeeze. In this case, the baseline firing rate and response amplitude are the same for each response (they are compiled from the same spikes, but merely aligned to different events). The larger information value for the squeeze event can therefore be interpreted unambiguously as activity being better time-locked to that event than to the others.

The neuron illustrated in Fig. 11D was recorded simultaneously with that in Fig. 11B, but had a much smaller rate modulation. The greatest rate occurred in the PETH aligned to the squeeze event, although there was considerable variability. The event information analysis (Fig. 11E) produced low information values for all three events, indicating that this cell did not encode event occurrence with high fidelity.
Application to Estimation of Latency Jitter

A further use for the present method is presented in Figure 12. A 200 ms-long rate step from 40 to 80 Hz was simulated, with latency jitter of 100 ms SD (normal distribution). Raster and PETH displays for these data are presented in Fig. 12A. Because of the latency jitter, the PETH no longer represented the step change in rate present on single trials, but was smeared. Figure 12B shows data simulated with no latency jitter, but with an underlying single-trial rate profile similar to the PETH of Fig. 12A. These datasets are indistinguishable on the basis of their PETHs. However, in one the PETH is an accurate representation of the single trial response, in the other the unit varies its response from trial to trial. The shift likelihood distribution was substantially more peaked for the cell with constant response than for the cell with jitter (Fig 12CD); this resulted in a larger information value for the reliable neuron. The simulation and analysis of Fig. 12BD was repeated 50 times. Figure 12E presents a histogram of the information measured from these simulations (mean: 1.05 bits) where there was no latency jitter, allowing comparison of the information with that determined from the data of Fig. 12AC (dotted line in Fig. 12E). The cell shown in Fig. 12AC had consistently smaller information values and coded for 47% of the information that could have been coded if there was not trial to trial variability.

The procedure followed in Fig. 12 could be applied to experimental data: the information obtained from a simulation based on the experimentally derived PETH can be compared with the information measured directly from the experimental spike train. Any mismatch will determine whether trial-by-trial response variability occurs, and if so quantify how great this is. This is demonstrated in Fig. 13 for the example cell shown in Fig. 11B. Figure 13A is the raster and PETH plot of the cell aligned to squeeze, while Fig. 13B is a similar display for the neuron simulated to have the same PETH. Figure 13CD show the shift likelihood distributions for the real (Fig. 13C) and simulated spike trains (Fig. 13D). The shift likelihood distributions are much narrower and better aligned for the simulated neuron, giving it an Iᵣ value of 3.97 bits, compared to 2.87 for the real cell. Figure 13E presents a histogram of the Iᵣ values from 100 simulations, in which the mean Iᵣ was 3.93 bits. We can therefore conclude that although the experimentally recorded cell did code the occurrence of the squeeze...
event well, trial to trial variability in its response meant that it coded for only 73\% of the information that could have been coded if its response had been consistent.

**Discussion**

In this paper, we have introduced a novel approach to spike train analysis. When recording from a single neuron, the experimenter has precise knowledge of the timing of external events such as stimulus delivery or task performance. These well defined events are often used to align neural data such as in the PETH. The present work, by contrast, views the behavioral event as uncertain in its timing, and quantifies how accurately its occurrence could be determined from the spike train. This probably matches better the perspective of downstream neurons, whose only information about external events comes from the spike trains which they receive.

**Methodological Considerations**

As with any information theoretic measure, the basic analysis of Fig. 1 yields information values which are upwardly biased (Fig. 7). We propose a shuffle-based technique to correct for this bias, and to determine whether the measured information is greater than that expected by chance (Fig. 5). In more conventional stimulus-response paradigms, shuffle correction has been shown to be an inappropriate means of correcting for bias due to limited data sampling (Treves & Panzeri, 1996), since it alters the number of empty bins in response frequency histograms. However, in the present situation, the shift likelihood distribution is used to calculate the posterior entropy. Whilst limited data sampling can lead to spurious peaks in the shift likelihood distribution, this is effectively measured using a shuffle approach. The shuffle correction for information bias is thus likely to be accurate in this case.

The significance estimation consumes substantial processing time: for a PSTH of 1000 ms duration, 50 trials and max $\delta t$ of 300ms, the shuffling approach took on average 1.6 hrs (100 shuffles) on a PC with 2Ghz Centrino Processor with 1GB RAM, while for 20 trials the time taken was 0.7 hrs (100 shuffles). We expect that with increasing processor speeds, estimation of the significance using shuffling will
become less time consuming, and not therefore be a substantial barrier to use of this technique.

In order to explore the behavior of this method, we have used simulated data whose properties are known. The simulations used a simple step change in rate throughout, since this was completely defined by two variables – the duration and amplitude of the step. Although real neuronal responses are far more complicated, the trends observed from rate steps should be generally applicable to more complex response profiles. If anything, responses with more complex profiles will penalize misalignment between a single trial spike train and the PETH more, resulting in a more sharply peaked shift likelihood distribution and a higher information value.

The method as presented in this paper represents neural spike generation as a gamma process. This is a better model than the simpler Poisson process, as it allows for a refractory period and can match the ISI histograms of real neurons well (Baker & Lemon, 2000; Stein, 1965). An important assumption is that successive ISIs are independent (as with all stochastic point processes). Real neurons often exhibit serial correlation in their ISIs, but this is usually generated mainly by the effects of rate modulation. Thus short intervals are most likely to be followed by short intervals, as they will probably be drawn from a region of high firing rate. It is unclear whether experimental data exhibit serial ISI correlations if the effects of rate modulation are corrected for (Davies et al 2006).

There is no reason why the present method should be restricted to situations where a gamma distribution appropriately models experimentally observed ISIs. Other studies have made use of different distribution functions (e.g. Weibul distribution, Rayleigh, beta, lognormal) with success. A particular instance where a simple gamma distribution may be inappropriate is where experimental ISIs are bimodally distributed (Bhumbra and Dyball 2004). So long as an appropriate model can be formed which generates a likelihood of observing a given spike train at a particular underlying firing rate profile, information about event encoding can be calculated.

Figures 8,9,10 show that the information measured is affected by several variables such as response amplitude and order as well as by non-stationarities in the data. This
is not necessarily a disadvantage. The method provides an estimate of information carried about event timing, on the assumption that the ‘decoder’ believes the average neural response to match the observed PETH. Since the PETH is, by definition, that part of the neural response which occurs reliably relative to the event in question, this is a reasonable baseline assumption. If there is some a priori reason to assume a more complex decoding scheme, that could probably be incorporated into the method for information estimation. However, in the absence of this, it is not unreasonable that the method should return lower values of information, corresponding to less efficient coding of an event, if a cell varies its response from trial to trial.

**Importance of Trial-to-Trial Variability**

Neuronal variability is an under-studied topic, but it is important because the brain must carry out processing accurately on a single trial. A PETH only gives the mean response of a neuron, it does not represent how reliable that response is. An implicit assumption of many studies is that the brain averages over many neurons with similar responses in a single trial, in the same way that the PETH averages over one neuron across many trials. However, because response variation is often correlated between cells from trial to trial (Baker et al. 2001), these two situations are not comparable (Lee et al. 1998; Shadlen and Newsome 1998).

The method which we describe has many possible applications. Most importantly, it conflates many different variables which might affect how well a cell can encode event timing into a single, principled summary measure, in defined units (bits). This can be used to determine which event is best encoded by a cell (Fig. 11), or whether the fidelity of event encoding changes from one condition to another (e.g. as a result of learning). It also allows objective comparisons to be made between different units. Because it is expressed in bits, the information encoded about event timing can be directly compared with the information encoded about other possible features, such as the type of stimulus, or the nature of the behavioral response. In a recent paper, Arabzadeh et al (2006), have presented a method for estimating the information provided by the neural spike train regarding the type of stimulus at different times relative to stimulus onset. The authors also explored how uncertainty in the stimulus time would affect the information regarding stimulus identity; not surprisingly, as the
temporal resolution of stimulus time decreases so does the information regarding stimulus identity. The method presented here allows an estimation of this uncertainty; using this estimate and employing the method proposed by Arabzadeh et al (2006), it might become feasible to estimate how much information is actually available about which stimulus (or event) occurred. This is just one example of where our paradigm could be integrated with previous formalism.

In most situations, the present method is not suited to extract specific parameters of the neural response from the spike trains. Response amplitude and latency are best measured from a PETH; response variability can be measured by other techniques (e.g. Fano factor). As such, our method forms a useful complement to existing techniques, not a substitute for them. However, using the technique described in Figs. 12-13, it is possible to estimate how much the encoded information was degraded by response latency variation. This could be a useful way of measuring response variability, as it expresses the measurement on a functionally meaningful scale (i.e. the consequences of this variation for impaired downstream decoding).

As multiple neuron recordings are becoming more common, the method presented here could also allow probing of the impact of neuronal variability at the population level. A population of neurons would allow a better estimate for the occurrence time of an event than a single neuron. To apply this method to simultaneous recordings, the shift likelihood distributions for the same trial across neurons should first be multiplied and then averaged across trials to give the mean shift likelihood distribution. In some preliminary simulations, we found that as the number of neurons increases so does the information, but correlation between the neurons limits the information increase. Further discussion of this potential extension to the method are however beyond the scope of this paper.

Acknowledgements

Funded by The Wellcome Trust. DS was also funded by the Wingate Foundation, the Leventis Foundation, the Cambridge Commonwealth Trust, and Christ’s College, Cambridge.
Figure Legends

Figure 1. Schematic of Analysis Method Concept.
A, raster plot (100 trials) and PETH of a simulated spike train. B, PETH for three different shifts relative to spike train of first trial. C, shift likelihood distribution for the spike train shown in B. D, overlain shift likelihood distributions for all trials, with the averaged shift likelihood distribution (grey line).

Figure 2. Schematic for Spike Density Likelihood Method (Poisson).
A, Smoothed PETH of simulated data with rate step from 20 to 60 Hz. The PETH is also the spike density function (right hand ordinate axis). Vertical dashed grey lines indicate the time of spikes for the first trial B, Probability plot of spike train for trial one, given spike density in A and $\delta\tau$ of 0 ms. C, shift likelihood distribution for spike train in B. D, overlain shift likelihood distributions for all trials, with the averaged shift likelihood distribution (grey line).

Figure 3. Schematic for Interspike Interval Likelihood Method.
A, Shifted PETH ($\delta\tau$ : -100ms, rate step 20 to 60Hz). Vertical dashed grey lines indicate the time of spikes for the first trial. Dark horizontal lines are the mean rate for the duration of the ISI. B, same as A but with no shift in the PETH. C, Same as A but PETH shifted 100ms in opposite direction ($\delta\tau$ : 100ms). Note that for shifted PETHs in A,B the mean rates corresponding to short ISIs are low. D, ISI distribution with a mean rate corresponding to the mean rate of ISI no.7 (18Hz) for the PETH in A. Arrow indicated the probability of observing a 7ms ISI from this distribution. E, same as D but distribution mean rate is that corresponding to ISI no. 7 from unshifted PETH shown in B. F, same as E but distribution mean rate is that corresponding to ISI no. 7 from shifted PETH shown in C. Under each distribution the probability of seeing a 7ms ISI is shown. G, shift likelihood distribution for spike train in ABC. H, overlain shift likelihood distributions for all trials, with the averaged shift likelihood distribution (grey line).

Figure 4. Schematic for Spike Density Likelihood Method (Gamma).
A, Example ISI distributions for three different rates, with gamma order 4. B, Inverse cumulative probability (survival) for ISI distributions shown in A. C, Death rate for distributions shown in A. These are formed by dividing the distributions in A by those in B. The markers on each line show the mean ISI value in each case. D, death rate plots as (C), with the abscissa rescaled as integral of the rate. E, as (D), but death rates are rescaled by dividing by the mean spike density of the ISI distributions shown in A. F, Scaled death rate plots for different orders of the gamma distribution. Key in (A) applies to (A-F). G1, Spike density for shifted PETH ($\delta \tau$:100ms). G2, Spike density integral for each time bin (1ms wide). G3, Probability of spiking at each time bin. G4, Probability of observation (spike or no spike) for each time bin. The product of all bins in (G4) will form the likelihood at the specific $\delta \tau$ for this trial. H, as (G) but for $\delta \tau=0$ms. I, as (G) but $\delta \tau=-100$ms. The grey shading in G3, H3, I3 draws attention to the response region in the spike train. For the non-zero $\delta \tau$, the probability of spiking is much smaller during this region. The spike train for this trial is shown under G4, H4, I4.

**Figure 5. Bias Estimation**

A, raster plot of neuron activity. B, raster plot of neuron shown in (A) after random shuffling of the ISIs. C, Single trial and mean shift likelihood distributions for shuffled data. D, distribution of information values calculated from 100 random shuffles. The dotted line shows the information value calculated from the original raster plot in A.

**Figure 6. Comparison between methods.**

A, Raw information values from the three different methods for estimating spike train likelihood, for different orders of gamma process. For each point on the abscissa the same spike train (20 to 60 Hz rate step) was used as input to each methods. B, Raster and PETH with oscillatory activity. C, Shift likelihood distributions for (B) by using the spike density likelihood method (gamma). The mean of the distributions is shown in grey. D, Same as (C) but using the ISI likelihood method. E, Bias in $I_r$, calculated using the spike density likelihood (gamma) method (black line) and the ISI method (dotted line) for PETHs with different baseline rates with no rate step. Value of $n$...
given in ordinate label in (A,E) shows that values are the mean over this number of simulations. This convention will be maintained for all other figures.

**Figure 7. Effect of smoothing, order, trial number and rate on bias.**

A, Effect of smoothing on PETHs (with no rate step), measured at three different baseline firing rates, for 10 trials per simulation. The horizontal dotted line represents the maximum of the ordinate axis in (B). A Gaussian kernel of unit area was used for smoothing with different standard deviations which are shown on the abscissa. Each point is the average of 10 simulations. B, as (A) but 50 trials were used. The dotted line represents the size of the ordinate axis in (C). C, as (B) but 100 trials were used. D, Effect of noise, quantified by coefficient of variation (CV) on the bias, for four different basal rates. E, Effect of spiking regularity on bias for different baseline rates. Ten trials per PETH used. F, as E but 50 trials per PETH used. G, as E but 100 trials per PETH used. Key for (A) applies to (ABCEFG).

**Figure 8. Influence of Different Response Parameters on Information Values.**

A, effect of smoothing kernel width on the I_r value. Response rate was double baseline in all cases. B, effect of different maximum shifts on the entropy of the prior and posterior shift distributions, using simulated data with different rates in the baseline and response regions. C, effect of response size. Abscissa shows the ratio of the response rate to the baseline rate. Circles mark points corresponding to a rate increase of 30 Hz above baseline. D, effect of trial-to-trial variations in the response size for different rate steps. Spiking regularity was fixed with gamma order 4. E, effect of different response duration. F, effect of trial-to-trial variations in the response duration for different rate steps.

**Figure 9. Influence of Spiking Regularity on Information Values.**

A, effect of spiking regularity (gamma order) for different rate steps. B, effect of incorrect estimation of gamma order. Each line shows results for data simulated with a particular gamma order. Abscissa shows the gamma order assumed for the analysis. Information values are expressed as a fraction of that determined when the correct order was assumed. C,D, effect of trial-to-trial variation in spike regularity (measured by ISI CV). Results are shown for simulations with low mean regularity (gamma order 2, A) and high mean regularity (gamma order 6, B). E, F, effect of changes in
spiking regularity in the response step. In all cases, the step width was 300ms. Results are shown for simulations with low mean regularity (gamma order 2, gray shading, E) and high mean regularity (gamma order 6, gray shading, F), which was the order assumed by the analysis. The abscissa gives the order used to simulate the response step. Key for (A) applies to all other plots.

Figure 10. Effect of Latency Variability from Trial to Trial.
A, Effect of latency variability from trial to trial for different rate step profiles. The abscissa shows the standard deviation of the latency variability. The circles on each plot show the latency variability at which the I_r value had dropped to 50% of its value with no latency variability. B, PETH and raster of simulated cell with a latency standard deviation of 50ms and rate step from 20 to 40Hz. C, Cluster plot of the actual latency vs most likely shift. Each dot represents a single trial. D. Effect of different rate profiles on the correlation coefficient. Latency variation was 60ms.

Figure 11. Implementation of analysis on experimental spike trains.
For entire figure, the left column represents data aligned to go-cue, middle column to reach and right hand column to squeeze events. (B-C) relate to discharge of one neuron, (D-E) to the discharge of a different cell that was simultaneously recorded. A, Distribution of 3 different behavioral events relative to each other. B, D, PETH (top) and raster plot (bottom) of the activity of a single neuron recorded from supplementary motor area (SMA). C, E, single trial shift likelihood distributions (black) and averaged shift likelihood distributions (gray).

Figure 12. Comparison of Responses with and without Latency Variation
A, PETH and raster plot for trials with baseline rate of 40 Hz, and a response rate of 80Hz (duration 200 ms, gamma order 4). Response latency varied randomly from trial to trial (normal distribution, SD 100 ms). B, PETH and raster for data simulated to have a rate profile the same as the PETH in (A), with no latency variation from trial to trial. C, D, shift likelihood distributions for single trials (black), and mean shift likelihood distribution (gray), corresponding to the data illustrated in (A,B). E, distribution of information estimated from 50 simulations as in (B). The information measured from (A) is shown as a dotted line.
Figure 13. Comparison of Simulated with Experimental Data
A, PETH and raster plot for experimentally recorded neuron shown in Fig. 11B, aligned to ‘squeeze’ event. B, PETH and raster for data simulated to have a single-trial rate profile the same as the PETH in (A). C, D, shift likelihood distributions for single trials (black), and mean shift likelihood distribution (gray), corresponding to the data illustrated in (A,B). E, distribution of information estimated from 100 simulations as in (B). The information measured from (A) is shown as a dotted line.

Reference List


