

# Decoding Spike Trains Instant by Instant Using Order Statistics and the Mixture-of-Poissons Model

Matthew C. Wiener and Barry J. Richmond

Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland 20892-4415

In the brain, spike trains are generated in time and presumably also interpreted as they unfold in time. Recent work (Oram et al., 1999; Baker and Lemon, 2000) suggests that in several areas of the monkey brain, individual spike times carry information because they reflect an underlying rate variation. Constructing a model based on this stochastic structure allows us to apply order statistics to decode spike trains instant by instant as spikes arrive or do not. Order statistics are time-consuming to compute in the general case. We demonstrate that data from neurons in primary visual cortex are well fit by a mixture of Poisson processes; in this special case, our computations are substantially faster. In these data, spike timing contributed information beyond that available from the spike count throughout the trial. At the end of the trial, a decoder based on the mixture-of-Poissons model correctly decoded about three times as many trials as expected by chance, compared with approximately twice as many as expected by chance using the spike count only. If our model perfectly described the spike trains, and enough data were available to estimate model parameters, then our Bayesian decoder would be optimal. For four-fifths of the sets of stimulus-elicited responses, the observed spike trains were consistent with the mixture-of-Poissons model. Most of the error in estimating stimulus probabilities is attributable to not having enough data to specify the parameters of the model rather than to misspecification of the model itself.

**Key words:** visual cortex; information; coding; modeling; statistics; timing

## Introduction

Because responses are presumably interpreted as they unfold in time, our goal is to decode spike trains instant by instant depending on whether a spike arrived. If spike trains are thought of as words in a language, then decoding—figuring out what stimulus elicited a particular observed spike train—can be thought of as looking up words in a neural dictionary. Ideally this dictionary should allow us, at any point in time, to translate a spike train into the best possible guess of which stimulus elicited it.

One approach to constructing a neural dictionary is to make as few assumptions as possible. Another approach, which we take here, is to model spike trains using (and checking) certain assumptions about their structure. Recent work shows that in spike trains from three areas of the monkey brain (the lateral geniculate nucleus, primary visual cortex (V1), and primary motor cortex) spike times appear to have been thrown down at random, with probabilities determined by the firing rate profile over time, the peristimulus time histogram (PSTH) (Oram et al., 1999, 2001; Baker and Lemon, 2000). Oram et al. (1999) and Baker and Lemon (2000) simulated stochastic (but non-Poisson) spike trains with relations among spikes that were indistinguishable

from those observed in experimental data. Using these models, it is in principle possible to determine how likely each stimulus is to have elicited every possible spike train. However, compiling a dictionary in this way would require simulating very large data sets, which, although less difficult than gathering enough experimental data, would still be prohibitively time-consuming.

Here we show that for spike trains with the stochastic structure described by Oram et al. (1999), we can use order statistics (Arnold et al., 1992), instead of simulation, to calculate directly how likely each stimulus is to have elicited each spike train. Order statistics give the probabilities of individual spike times in trains with this stochastic structure based on the spike count and the firing rate profile—exactly the parameters used by Oram et al. (1999). Using order statistics, we can update the stimulus probabilities at each instant depending on whether a spike arrives. Once we have these probabilities, we can guess, for example, that the stimulus with the highest probability elicited the spike train. This is something like looking up a word letter by letter and discarding words that do not begin with the letters seen so far.

Although much faster than simulation, decoding using order statistics is, in the general case, still more time-consuming than we would like. However, calculating order statistics is much simpler and faster in the special case in which the spike trains are considered as arising from a mixture of a small number of Poisson processes. We show that our data are in fact well fit by such a model. The recognition of this structure in the data substantially simplifies and speeds decoding of single-neuron spike trains instant by instant.

## Materials and Methods

**Decoding.** Decoding, that is, guessing which stimulus elicited a particular spike train, has two steps. We first estimate the probability that each

Received Sept. 27, 2002; revised Dec. 26, 2002; accepted Dec. 26, 2002.

This work was supported by the National Institute of Mental Health (NIMH) and IRP. We thank Dr. John Hertz (Nordita), Dr. Peter Latham (University of California, Los Angeles, CA), Dr. Zheng Liu (NIMH), Dr. Jeff Sachs (Merck Research Laboratories), and Dr. Yasuko Sugase-Miyamoto (NIMH) for comments on this manuscript.

Correspondence should be addressed to Barry J. Richmond, Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Building 49, Room 1B-80, Bethesda, MD 20892-4415. E-mail: bjr@ln.nimh.nih.gov.

M. C. Wiener's present address: RY84-202, Applied Computer Science and Mathematics Department, Merck Research Laboratories, Rahway, NJ 07065. E-mail: matthew\_wiener@merck.com.

Copyright © 2003 Society for Neuroscience 0270-6474/03/232394-13\$15.00/0

stimulus elicited the spike train and then choose a stimulus based on those estimated probabilities. Here we minimize expected error by choosing the most probable stimulus. If some mistakes are more important than others, a decision rule minimizing an appropriate cost function could be used.

With the decision rule fixed, decoding a spike train as it unfolds in time requires estimating the probability  $p_s(t_j|R_j)$ : the probability, estimated at time  $t_j$ , that stimulus  $s$  elicited the response  $R_j = (r_1, \dots, r_j)$ , where each  $r_i$  is 1 if a spike occurred at time  $i$  and 0 if no spike occurred in time bin  $i$ . The presence or absence of a spike is measured in 1 msec bins; time  $t_i$  is the time in the  $i$ th bin. Any bin width can be used as long as there can be at most one spike in any bin. All of our calculations could be recast in continuous time in a straightforward manner.

Using Bayes' rule, estimating  $p_s(t_j|R_j)$  can be transformed into estimating the probability that a spike occurs in the next bin (which may depend on the history of the spike train):

$$p_s(t_j|R_j) = \frac{p_s(t_{j-1}|R_{j-1})P(r_j = 1|s, R_{j-1})}{Z}, \quad (1)$$

if a spike is fired in bin  $j$ , or

$$p_s(t_j|R_j) = \frac{p_s(t_{j-1}|R_{j-1})(1 - P(r_j = 1|s, R_{j-1}))}{Z} \quad (2)$$

if no spike is fired in bin  $j$ .  $P(r_j = 1|s, R_{j-1})$  is the probability that the a spike appears in the time bin  $j$  ( $r_j = 1$ ) if stimulus  $s$  has been presented and the response through time  $t_{j-1}$  is  $R_{j-1}$ . We use  $P$  for the probability of response given stimulus and  $p$  for probability of stimulus given response to create a typographical distinction.  $Z$ , here and throughout, is a normalizing term; that is, it is the sum, across the appropriate variable, usually the stimulus  $s$ , of terms in the numerator of whatever fraction it appears in.

Using Equations 1 and 2, we proceed bin by bin, updating the stimulus probabilities depending on whether a spike occurs in each bin. For efficiency, responses can be decoded in small jumps rather than instant by instant. The probability  $P(r_{j+1} \dots r_{j+k}|s)$  for any response can be calculated using the distribution of first spike times,  $P(\tau_1|s)$ :

$$P(r_{j+1} \dots r_{j+k} = 0, r_{j+k} = 1|s) = P(\tau_1 = t_{j+k}|s), \quad (3)$$

$$P(r_{j+1} \dots r_{j+k} = 0|s) = 1 - \sum_{k'=1}^k P(\tau_1 = t_{j+k'}|s), \quad (4)$$

where the response through time bin  $j$  is assumed known. Below we will show how to calculate the distribution of first spike times (Arnold et al., 1992).

We avoid overfitting using three-way cross-validation: the trials were divided into three subsets, with two-thirds used to fit the model and the remaining third used to test decoding. Each third of the data was used twice to fit and once to test; we averaged over the three splits.

**Order statistics for spike trains.** By considering each spike as the "next first spike," the distribution of first spike times  $P(\tau_1 = t_j|s)$  was calculated for each stimulus using Equations 3 and 4.

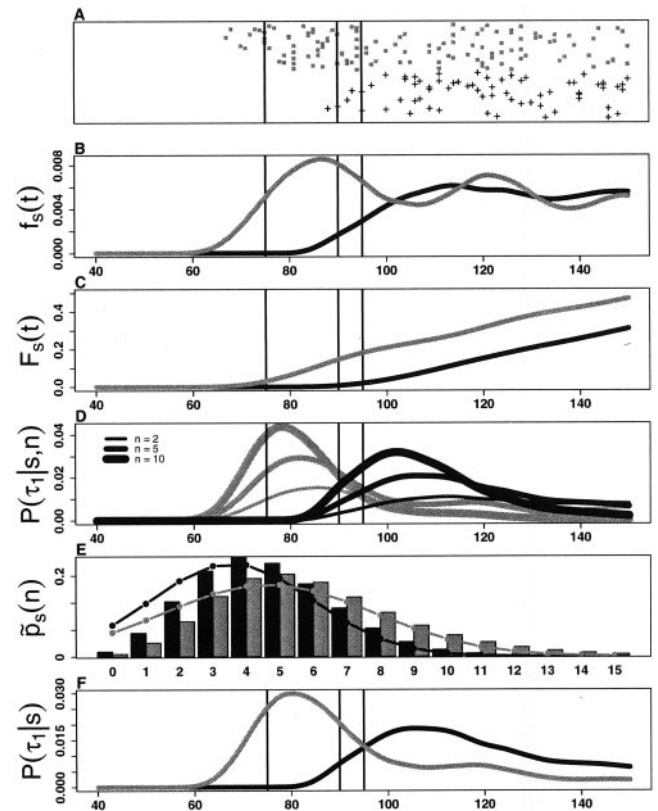
For a spike train with  $n$  spikes, the  $(n, k)$  order statistic describes the distribution of the  $k$ th of  $n$  draws, after those draws are sorted; in our case, this will be  $P(\tau_k|s, n)$ , the time of the  $k$ th of  $n$  spikes in a train elicited by stimulus  $s$ :

$$P(\tau_k = t_j|s, n) = \frac{n!}{(k-1)!(n-k)!} F_s^{k-1}(t_j) f_s(t_j) [1 - F_s(t_j)]^{n-k}. \quad (5)$$

When we focus on the first of  $n$  spikes, i.e.,  $k = 1$ , Equation 5 simplifies to describe the  $(n, 1)$  or first order statistic  $P(\tau_1|s, n)$ :

$$P(\tau_1 = t_j|s, n) = n f_s(t_j) [1 - F_s(t_j)]^{n-1}. \quad (6)$$

In Equations 5 and 6,  $f_s(t_j)$  is the normalized spike density function or firing rate profile over time;  $F_s(t_j)$  is the corresponding cumulative firing



**Figure 1.** Construction of count-weighted order statistics. Labels correspond to equations in Materials and Methods. *x*-Axis (except *E*), Time from stimulus onset (milliseconds), truncated 150 msec after stimulus onset for visibility (although the measured responses extend to 300 msec after stimulus onset); vertical lines show 75, 90, and 95 msec after stimulus onset. *x*-Axis (*E*), Spike count. *y*-Axes: *A*, trials; *B–F*, probability. *A*, Rasters of responses to two stimuli (black, gray throughout). Each row of dots represents a trial; each dot shows a spike time. *B*, Firing rate profiles,  $f_s(t)$ , estimated using local regression (Loader, 1999) with a data window including 10% of the data. *C*, Cumulative firing rate profiles,  $F_s(t)$  (they do not reach 1 because the *x*-axis is truncated). *D*, Order statistics,  $P(\tau_1|s, n)$ , for  $n = 2, 5$ , and  $10$  (Eq. 1). *E*, Count histograms,  $\bar{p}_s(n)$ , with mixture-of-Poissons approximations used to mitigate the effects of using a small data set. *F*, Count-weighted order statistics,  $P(\tau_1|s)$  (Eq. 7).

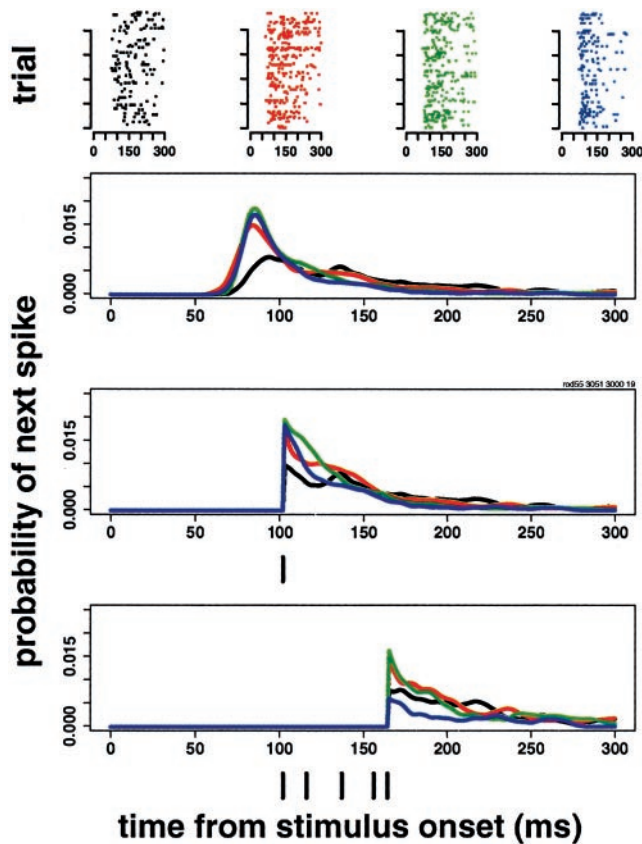
rate profile; and the factor  $[1 - F_s(t_j)]^{n-k}$  is the probability of  $n - k$  spikes arriving after time  $t_j$ . As noted above, we discretize time into 1 msec bins (the same precision with which spike times were recorded).

Figure 1 illustrates estimating the quantities in Equation 6 from data for spike trains elicited by two stimuli (coded black, gray throughout). One stimulus elicits spikes beginning  $\sim 25$  msec earlier than the other. For each stimulus  $s$ , the firing rate profile is  $f_s(t)$  (Fig. 1*B*).  $F_s(t)$  (Fig. 1*C*) is the corresponding cumulative probability. From  $f_s(t)$  and  $F_s(t)$ , we can calculate  $P(\tau_1|s, n)$  for any spike count  $n$  (Eq. 6; Fig. 1*D*). As  $n$  increases, the weight of the first order statistic shifts left; i.e., the first spike is likely to occur earlier.

To avoid requiring the decoder to know in advance how many spikes  $n$  will arrive in a particular trial, we average the first order statistics  $P(\tau_1|s, n)$ , weighted with the expected distribution of spike counts  $\bar{p}_s(n)$  for stimulus  $s$  (Fig. 1*E*); we use  $\bar{p}_s(n)$  to distinguish from  $p_s(t)$ , estimated probability of stimulus  $s$  at time  $t$ . This count-weighted first order statistic,  $P(\tau_1|s)$ , gives the expected probability of the first spike occurring at time  $t$  for spike density  $f_s(t)$  and the expected distribution of spike counts for stimulus  $s$ :

$$P(\tau_1 = t_j|s) = \sum_{n=1}^{N_{\max}} \bar{p}_s(n) P(\tau_1 = t_j|s, n). \quad (7)$$

In other words, the count-weighted first order statistic gives the distribution of the next interspike interval. Note that  $P(\tau_1|s)$  sums to  $1 - \bar{p}_s(0)$ ; the term corresponding to no more spikes ( $n = 0$ ) is left out.



**Figure 2.** Probabilities of first spike times calculated at different times during a trial. The four rastergrams in the first row show responses of four stimuli (taken, for this example, from one of the neurons in the experiment using 16 stimuli). Each row represents a trial; each dot in a row represents the time of a spike in that trial. The colors of the rasters correspond to the colors used in the graphs below. The bottom three panels show the first spike time probabilities calculated at stimulus onset, 102 msec after stimulus onset, and 164 msec after stimulus onset; under each plot, vertical lines show the times of spikes in the train being decoded (which comes from the stimulus whose responses are shown in black). The distribution of next spike times is initially similar for the red, green, and blue stimuli, but later in the trial, when the blue stimulus fires much less than the red or green stimulus, the probabilities diverge.

The second order statistic conditioned on the first spike time is calculated as the first order statistic of a restricted response  $\tilde{f}_s$ , created by left-truncating  $f_s(t)$  at the time of the previous spike and renormalized to sum to 1. The  $(n, k + 1)$  order statistic can be calculated as the  $(n - k, 1)$  order statistic of the remaining response (Arnold et al., 1992).

Order statistics show a Markov property such that future spike times depend on past spike times only through the number of spikes already observed and the most recent spike time. This dependence on the number of spikes already observed means that the process is not a renewal process. Figure 2 shows first order statistics calculated at the beginning of a trial and after the first and fifth spikes in the response.

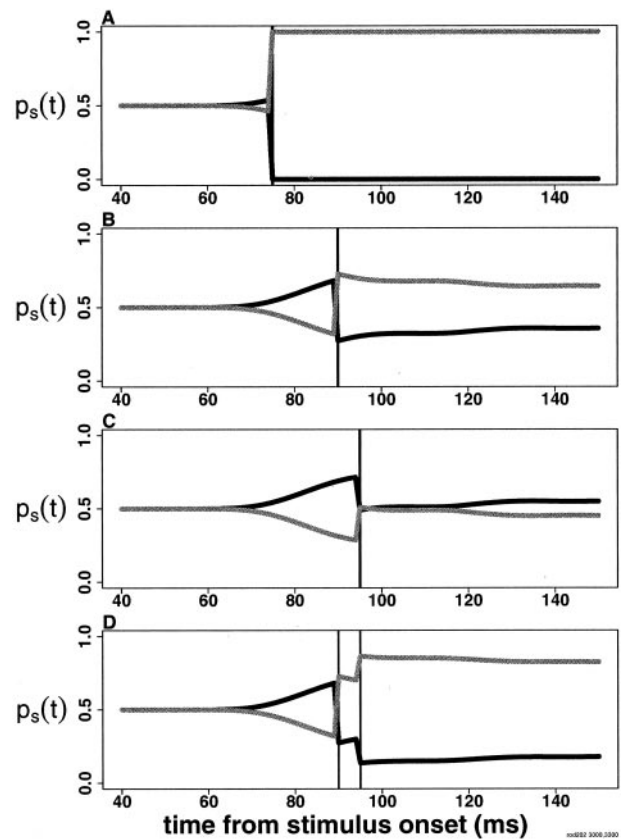
Figure 3 shows the decoding of several spike trains for the two-stimulus example of Figure 1.

*Simplifying order statistics using a mixture of Poisson distributions.* Equation 7 can be used to calculate the count-weighted first order statistic for any distribution of spike counts. If the spike count distribution is a mixture of a finite number of Poisson distributions, our calculations become much simpler.

For a Poisson process with time-varying mean rate  $\lambda_s f_s(t)$ , the probability that the first spike occurs at time  $j$  simplifies to:

$$P(\tau_1 = t_j | s, \lambda_s) = \lambda_s f_s(t_j) e^{-\sum_{k=0}^{j-1} \lambda_s f_s(t_k)}, \quad (8)$$

where  $f_s(j)$  is the spike density function (normalized PSTH) for stimulus  $s$ .  $\lambda_s$  is a rate multiplier. Because we use discrete time bins, the first term



**Figure 3.** Decoding responses for the example of Figure 1. Labels correspond to equations in Materials and Methods. Black and gray are as in Figure 1. *x*-Axis, Time (milliseconds); vertical lines, 75, 90, and 95 msec after stimulus onset. *y*-Axes, Probability. *A*, Stimulus probabilities,  $p_s(t)$ , from decoding a spike train with a single spike at 75 msec, when only one stimulus ever elicits spikes. *B*, Decoding a spike train with a single spike at 90 msec. The probability of a black stimulus rises between 70 and 90 msec after onset, because early spikes are more likely from the gray stimulus, and none appear. *C, D*, Interpretation of a spike 95 msec after stimulus onset depends on whether there was an earlier spike. The decoding algorithm does not look into the future, so *B–D* are identical until 90 msec after stimulus onset, and *B* and *D* are identical until 95 msec after stimulus onset.

in Equation 8, which represents the probability of a spike at time  $t$ , can be  $>1$  for some combinations of  $f_s$  and  $\lambda_s$ . To avoid this problem, we use the approximation:

$$P(\tau_1 = t_j | s, \lambda_s) = [1 - e^{-\lambda_s f_s(t_j)}] e^{-\sum_{k=0}^{j-1} \lambda_s f_s(t_k)}, \quad (9)$$

because  $e^{-\lambda_s f_s(t_j)}$  is the probability of no spike occurring in time bin  $j$  and is guaranteed to be between 0 and 1.

Equation 8 calculates the sum of count-specific first order statistics  $P(\tau_1 | s, n)$  for each  $n$ , weighted with the appropriate Poisson probabilities. This simplifies Equation 7 in the case of a simple Poisson distribution of spike counts. However, the distributions of stimulus-elicited spike counts from several brain areas are not adequately modeled by a Poisson distribution (Baddeley et al., 1997; Gershon et al., 1998; Wiener et al., 2001).

A mixture of a finite number of Poisson distributions is more flexible than a single Poisson distribution for modeling stimulus-elicited spike count distributions and almost as simple. To model the distribution of spike counts elicited by stimulus  $s$  as a mixture of Poisson distributions with mean rates  $\lambda_{s,i}$  with weights summing to 1, we add the first order statistics for each mean rate  $\lambda_{s,i}$  weighted by the probability  $p(\lambda_{s,i})$  of observing that rate:

$$P(\tau_1 = t_j | s) = \sum_i p(\lambda_{s,i}) P(\tau_1 = t_j | s, \lambda_{s,i}). \quad (10)$$



In each mixture, the constituent Poisson processes share a single firing rate profile  $f_s(t)$  and differ only in the rate by which the profile is multiplied. Thus the model is separable: estimation of the firing rate profile does not interact with estimation of the distribution of mean counts. This model provides a good description of our data (see Results).

The calculations in Equations 8 and 10 are conceptually identical to those in Equations 6 and 7. However, because the sum in Equation 6 can be taken analytically in the case of a Poisson distribution, the calculations using a mixture of Poisson distributions are approximately an order of magnitude faster for our data sets. The speedup comes because we calculate order statistics for only a few Poisson means rather than for many spike counts. The two methods differ slightly in how they deal with the discretization of time into bins (the approximation in Eq. 9 is unnecessary when using Eqs. 6, 7). Below we will show that the effect of this difference on the estimated stimulus probabilities is negligible.

**Model parameters on subintervals.** The decoding procedures outlined above iteratively treat each new spike as the first spike in a shorter response. This requires estimating model parameters for the shorter response from the parameters for the full model. For the multiple Poisson formulation, this is simple and quick (whereas in the general case, it is computationally intensive). The rate function of a Poisson process on a subinterval is the original rate function restricted to that interval. The weights  $p(\lambda_{s,i}|s)$  are continuously updated using Bayes' rule, because each weight is the probability with which we believe the spike train being decoded comes from the particular process. The Poisson processes differ only by a multiplier of the rate, so our evidence of which process the train comes from is the number of spikes we have seen:

$$p(\lambda_{s,i})(j) = \frac{P(n(j)|s, \lambda_{s,i})p(\lambda_{s,i})(0)}{\sum_i P(n(j)|s, \lambda_{s,i})p(\lambda_{s,i})(0)}, \quad (11)$$

where  $p(\lambda_{s,i})(0)$  denotes the weights at the beginning of the trial.

**Generating surrogate data.** To compare the performance of our decoder on recorded data with its performance on data with the structure for which the decoder was designed, we simulate spike trains using a procedure similar to that used for decoding. We calculate the distribution  $P(\tau|s, 1)$  of next spike times (above) and randomly select a spike time. The simulation also proceeds iteratively: after each spike time is chosen, the distribution for the next is calculated. The probabilities in Equations 7 and 10 do not sum to 1, because they omit the probability of no additional spikes. For implementation, the probability that no more spikes occur in the interval is assigned to an additional bin. When this bin is selected, we consider the spike train under construction to be complete. By construction, such trains have the structure specified by the model.

Surrogate spike trains matched to observed data are generated using the observed spike density and spike count distribution for each stimulus. The same number of trials observed is generated for each stimulus.

**Incorporating a refractory period.** Real spike train data often depart from the simple mixture-of-Poissons model by becoming either less or more likely to fire for a short time after a spike (exhibiting a refractory or rebound period, respectively). To adjust, we depart slightly from classical order statistics and replace the spike density function  $f(t)$  with  $f(t)r(t - \tau)$ , where  $\tau$  is the time of the previous spike.

We estimate  $r$  by comparing the observed interspike interval distribution with the interspike interval distribution of a set of surrogate trains. Because interval distribution is related to spike count (more spikes mean shorter intervals), we actually generate more surrogate trains than needed and subselect to match the observed spike count distribution. If the interspike interval in the generated data is significantly different from the observed interspike interval ( $\chi^2$  test,  $p < 0.05$ ), then  $r(1)$  is set to  $g_{\text{obs}}(1)/g_{\text{gen}}(1)$ , where  $g_{\text{obs}}$  and  $g_{\text{gen}}$  are the interspike interval distributions in the observed and generated data. A new set of trials is simulated with this new refractory period (in which the frequency of interspike intervals of length 1 now matches the frequency in the observed data), and the process is repeated for subsequent intervals until the interspike interval distribution of the generated data is indistinguishable from the interspike interval distribution of the experimental data.

**Estimating the spike density function from data.** The spike density function measures rate variation over time. To estimate the spike density function for a particular stimulus, we create a histogram at 1 msec precision (the experimental sampling rate) across trials for that stimulus and smooth using local regression (Loader, 1999) with a window using 10% of the data. The spike density functions obtained using this window and the optimal smoothing window determined by generalized cross-correlation (Craven and Wahba, 1979; Loader, 1999) are similar: median correlation, 0.98; interquartile range (iqr), 0.95–0.99. The smoothed histogram is normalized so that its sum across bins is 1.

**Fitting a mixture of Poisson distributions to observed spike counts.** The probability of drawing a particular count  $n$  from a mixture of Poisson distributions is the weighted sum of the probabilities of drawing the count  $n$  from each of the individual Poisson distributions in the mixture:

$$\bar{P}_s(n) = \sum_{i=1}^k p(\lambda_{s,i}) \frac{e^{-\lambda_{s,i}} \lambda_{s,i}^n}{n!}, \quad (12)$$

where  $k$  is the number of distributions in the mixture;  $\lambda_{s,i} \geq 0$  is the mean of the  $i$ th distribution in the mixture for stimulus  $s$ ; and the weights  $p(\lambda_{s,i})$  are positive and sum to 1.

For each stimulus  $s$ , the parameters  $\lambda_{s,i}$  and  $p(\lambda_{s,i})$  are estimated by maximizing the log likelihood of the observed spike counts given the corresponding model. We seek the most parsimonious description of each distribution. Thus, for each stimulus, we first find the best single Poisson distribution ( $k = 1$ ). If the observed distribution could reasonably have come from the fit Poisson ( $\chi^2$  test,  $p > 0.05$ ), we use this model; otherwise, we fit a mixture of two Poisson distributions and again check for consistency. In general, if the data are inconsistent with a mixture of  $k$  Poisson distributions, we fit a mixture with  $k + 1$  Poisson distributions. In this work, we did not require mixtures of more than five Poisson distributions (see Results).

**Information calculations.** Transmitted information (Shannon and Weaver, 1949; Cover and Thomas, 1991) is defined as  $I(R, S) = \sum_{r,s} p(r, s) \log(p(s|r)/p(s))$ , where the stimulus probabilities  $p(s)$  are taken from the experiment, and  $p(s|r)$  is calculated using Equation 1 or 2. Using response models avoids some estimation problems associated with small data sets (Panzeri and Treves, 1996; Golomb et al., 1997) and yields estimates of information comparable with those using other validated methods (Gershon et al., 1998; Wiener and Richmond, 1998).

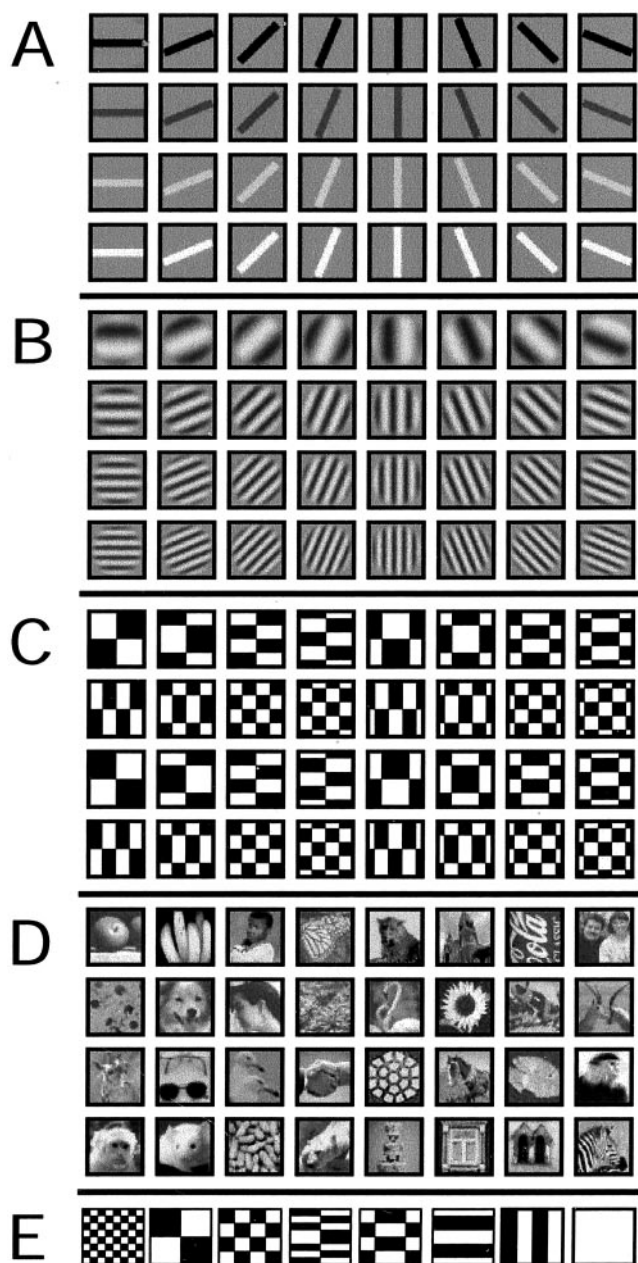
**Data sets.** We decoded responses recorded from monkey primary visual cortex in two previously reported experiments (Kjaer et al., 1997; Wiener et al., 2001). In each experiment, responses were recorded using standard single-electrode techniques from complex cells in primary visual cortex of awake rhesus monkeys. At the beginning of each trial, a fixation point appeared on a screen. One hundred milliseconds after the monkey fixated, a stimulus was flashed on the receptive field of the neuron for 300 msec and then replaced with the background. The monkey was not required to react to the stimulus. If the monkey fixated within  $\sim 2^\circ$  of the fixation point during the entire period from the appearance of the fixation point until the stimulus disappeared, it was rewarded with a drop of liquid when the stimulus disappeared. If the monkey shifted its gaze further than  $2^\circ$  from the fixation point, the trial was aborted, and the monkey received no reward. There was a delay of 300 msec between trials, during which the monkey was not required to fixate.

In one set of experiments (Wiener et al., 2001), 128 stimuli were shown: 32 oriented bars (Fig. 4A), 32 sine-wave gratings (Fig. 4B), 32 Walsh patterns (Fig. 4C), and 32 photographic images (Fig. 4D). In another experiment (Kjaer et al., 1997), 16 Walsh patterns were used (Fig. 4E).

**Computation.** The calculations presented in this paper were performed in the R statistical computing environment (Ihaka and Gentleman, 1996).

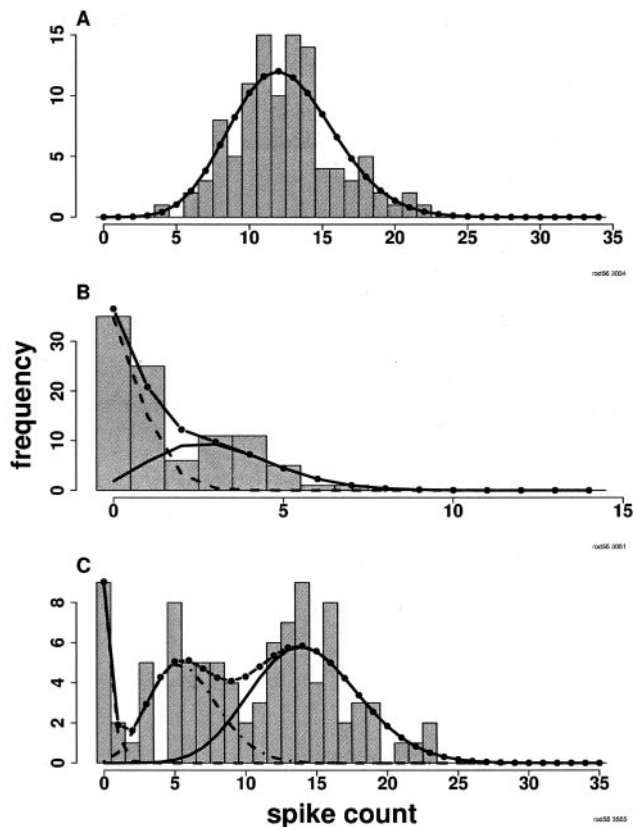
## Results

Our original analyses were performed on data from the 29 neurons from the experiment with 16 stimuli (Kjaer et al., 1997). In a single experiment (recording from a single neuron), each stimu-



**Figure 4.** Stimuli used in the experiments. In one set of experiments (Wiener et al., 2001), the stimulus set consisted of 32 oriented bars (*A*), 32 sine-wave gratings (*B*), 32 Walsh patterns (*C*), and 32 photographic images (*D*). In the other set of experiments, 16 Walsh patterns were used (the 8 shown in *E* and their contrast-reversed counterparts).

lus was presented approximately the same number of times; the number of presentations per stimulus varied from neuron to neuron, ranging from 19 to 230 (median, 42). The mixture-of-Poissons model made decoding sufficiently fast that we were able to decode data from the 17 neurons from the experiment with 128 stimuli (Wiener et al., 2001). In these experiments, the median number of presentations per stimulus ranged from 8 to 52 (median 14) in different neurons; each stimulus was presented approximately the same number of times. Below we will show that decoding results using the two methods on the data from the experiment with 16 stimuli were very similar. Except where stated otherwise, the results in this paper were obtained using the mixture-of-Poissons model.

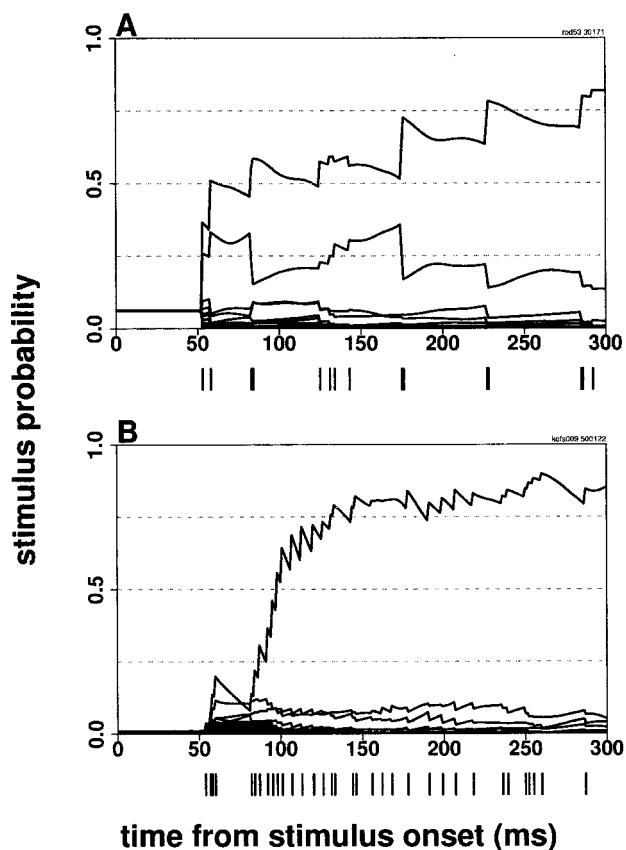


**Figure 5.** Spike count distributions fit by mixtures of Poisson distributions. The *top*, *middle*, and *bottom* panels show spike count distributions fit by a single Poisson distribution, a mixture of two Poisson distributions, and a mixture of three Poisson distributions, respectively. In each panel, the *histogram* (gray bars) shows the observed distribution of spike counts. The *dots* connected by *lines* show the fitted values, and the *solid* and *dashed* lines show the component Poisson distributions. *A*, Single Poisson distribution with mean of 12.4. *B*, Mixture of two Poisson distributions with means of 0.4 (weight, 0.56) and 3.1 (weight, 0.44). *C*, Mixture of three Poisson distributions with means of 0.2 (weight, 0.11), 5.8 (weight, 0.31), and 14.4 (weight, 0.58). Note the different scales on the *x*- and *y*-axes of the three panels.

#### Mixtures of Poisson distributions for spike count

We examined the mixture-of-Poissons model for 2636 spike count distributions in two different sets of V1 data (Kjaer et al., 1997; Wiener et al., 2001); 50.8% of the spike count distributions were adequately fit with a single Poisson distribution, 39.4% with a mixture of two Poisson distributions, 7.4% with three, 2.0% with four, and 0.4% with five. No distribution required a mixture of more than five distributions. Examples of mixture models with one, two, and three components are shown in Figure 5.

The modeled distributions are, by design, smoother than the observed distributions. Various spike counts are more or less likely in the model than in the data. Spike counts that are very frequent in the data may be assigned less probability in the model, and spike counts not actually observed in the data may be assigned nonzero probability (as is the case for a count of four in Fig. 5*A,C*). Fitting a mixture of Poisson distributions to the observed spike count is, among other things, a form of smoothing (meaning we believe that if we gathered enough data we would encounter a response with four spikes elicited by the stimulus corresponding to Fig. 5*C*). As expected when smoothing, the means of the modeled distributions are very close to the observed means (median difference across all neurons, 0% of the measured mean; iqr,  $-3$  to  $+4\%$ ), but the variances are lower (median difference,  $-5\%$  of the measured variance; iqr,  $-24$  to  $+9\%$ ).



**Figure 6.** Decoding responses of neurons in monkey V1 in an experiment using 16 stimuli (Kjaer et al., 1997) (top) or 128 stimuli (Wiener et al., 2001) (bottom). *x*-Axis, Time from stimulus onset. *y*-Axis, Stimulus probabilities. Stimulus probabilities from decoding single spike trains (black vertical lines below each panel). Each line shows the probability assigned by the decoding algorithm to a single stimulus. *A*, Decoding a response in an experiment using 16 stimuli (Kjaer et al., 1997). The top line shows the probability of the stimulus that elicited the spike train (Fig. 4E, second Walsh pattern from the left). *B*, Decoding a response in an experiment using 128 stimuli (Wiener et al., 2001). The top line shows the probability of the stimulus that elicited the spike train (Fig. 4C, first row, second Walsh pattern from the left).

In the Poisson decoding, we model the distribution of mean spike counts in any subinterval of the trial (from any particular time to the end of the trial) on the basis of the distribution of mean spike counts in the entire trial [ $p(\lambda_{s,i})$  at the start of the trial] and the firing rate modulation over time (i.e., how many of the spikes are expected to have occurred by now; see Eq. 11). If we use the mixture-of-Poissons model to predict, on five successively shrinking subintervals (beginning 50, 100, 150, 200, and 250 msec after stimulus onset), the variance of spike counts elicited by each stimulus and take a linear regression of  $\log(\text{predicted variance})$  versus  $\log(\text{observed variance})$ , we find a negative intercept (median across 46 neurons,  $-0.13$ ; iqr,  $-0.25$  to  $-0.03$ ) and a slope near 1 (median,  $0.98$ ; iqr,  $0.95$ – $1.02$ ). The  $r^2$  values of the regressions are high (median,  $0.94$ ; iqr,  $0.91$ – $0.96$ ). Below we will examine the effect on decoding accuracy of using the estimated subinterval spike count distributions compared with using spike count distributions directly measured on the subintervals.

### Decoding

Figure 6 shows an example of the development of stimulus probabilities over time (i.e., the probability that each stimulus elicited the observed spike train) when decoding using the mixture-of-Poissons method. Figure 6A shows an example from an experi-

**Table 1.** Probability of most probable (guessed) stimulus when spike trains are decoded using the mixture-of-Poissons model

| Experiment        | Probability of guessed stimulus [median (iqr)] | Margin of victory [median (iqr)] |
|-------------------|--|----------------------------------|
| 16 stimuli        |  |                                  |
| Correct guesses   | 0.29 (0.20–0.42)                               | 0.08 (0.03–0.20)                 |
| Incorrect guesses | 0.22 (0.16–0.32)                               | 0.04 (0.02–0.11)                 |
| 128 stimuli       |  |                                  |
| Correct guesses   | 0.11 (0.08–0.19)                               | 0.03 (0.01–0.07)                 |
| Incorrect guesses | 0.10 (0.07–0.15)                               | 0.02 (0.01–0.05)                 |

Shows the probability (at the end of the counting window, 300 msec after stimulus onset) of the most probable stimulus in the experiments with 16 and 128 stimuli. Also shown is the margin by which the highest probability was greater than the second-highest probability (the margin of victory). Probabilities are presented separately for trials in which the stimulus with the highest probability did elicit the trial (correctly decoded trials) and for those in which the trial was elicited by some other stimulus (incorrectly decoded trials). Results using order statistic decoding for neurons from the experiment with 16 stimuli are nearly identical.

ment in which 16 Walsh patterns were shown (Kjaer et al., 1997), and Figure 6B shows an example from an experiment in which 128 stimuli were shown (Wiener et al., 2001). The spike trains decoded are shown underneath each panel. The stimulus probabilities at each time depend only on the spike train observed up to that time; the decoding algorithm does not look into the future. Although the probabilities change more abruptly when spikes arrive, they change even when no spikes are fired. Thus the absence of spikes, as well as their presence, is informative (cf. Sherlock Holmes's dog that did not bark).

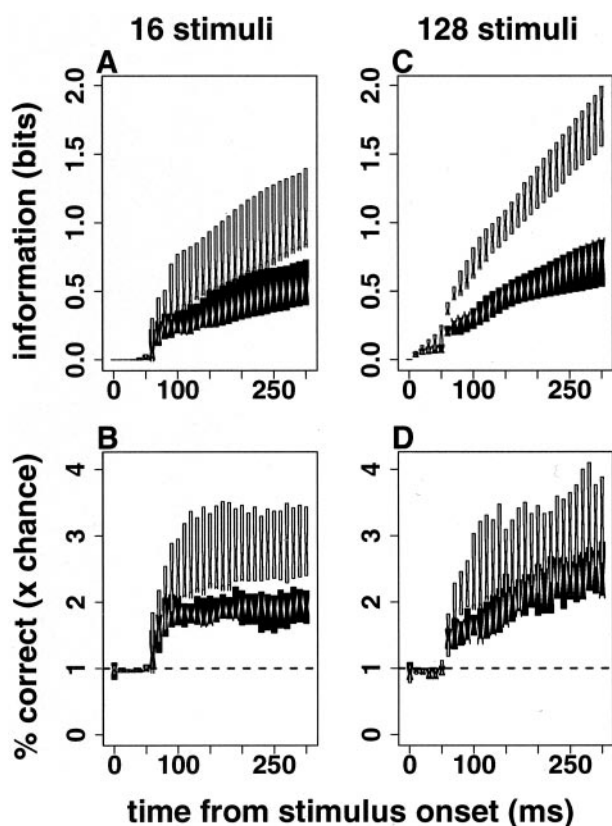
Figure 6, *A* and *B*, shows correctly decoded trials in which one stimulus is far more probable than all others. In these cases, the probability of the guessed stimulus is high. Table 1 summarizes the distribution of the maximum estimated probability, that is, the probability of the guessed stimulus, at the end of our decoding window (300 msec after stimulus onset). It also shows the difference between the probabilities of the most probable and second most probable stimulus (the “margin of victory” over the “runner-up”). The winning probabilities are higher in the experiments with 16 stimuli than in the experiments with 128 stimuli, because the total probability ( $p = 1$ ) is divided among fewer stimuli.

Intuitively, it should become easier to distinguish among stimuli as the probability of one stimulus rises and the probabilities of others fall. We evaluate the decoding in two ways: by measuring the amount of information obtained from the spike train and also by seeing how well the decoding lets us guess which stimulus elicited the observed response. As stated in Materials and Methods, in all our calculations we avoided overfitting using three-way cross-validation: the available trials were divided into three subsets, with two-thirds used to fit the model and the remaining third used to test decoding. Each third of the data was used twice to fit and once to test; the results we present are averaged over the three sets.

Information theory (Shannon and Weaver, 1949) shows that the amount of information in a set of spike trains depends on how they are interpreted, that is, on the nature of the neural code. Figure 7, *A* and *C*, shows that for both sets of experiments, decoding using only the number of spikes in a response (dark boxes) yields substantially less information than considering both spike count and spike timing using the mixture-of-Poissons model (light boxes).

Does this extra information provided by paying attention to timing actually help us decode more accurately? It is natural to guess that the stimulus with the highest probability elicited the observed spike train (as it did in both examples in Fig. 6). Figure

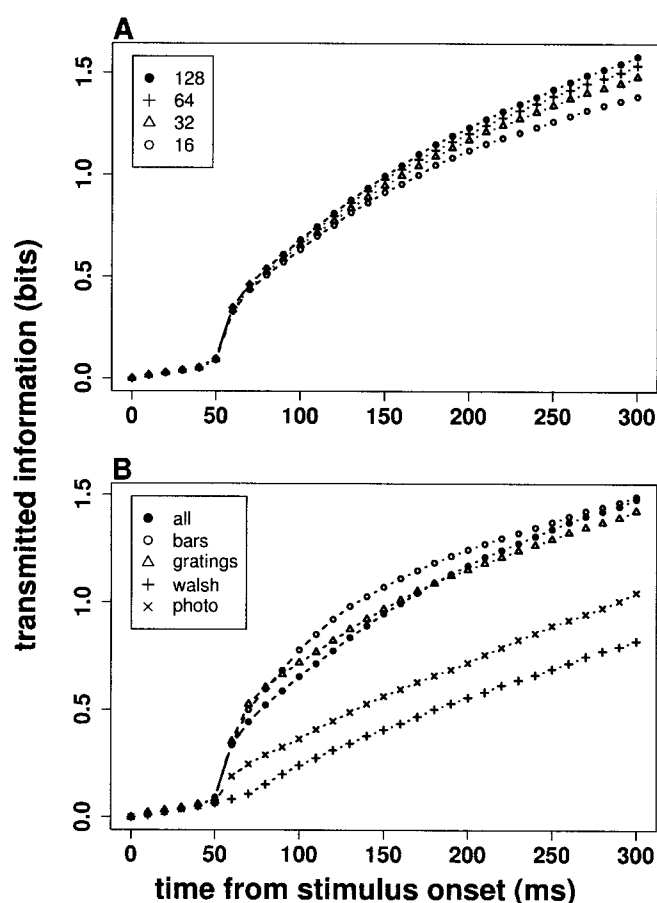




**Figure 7.** Decoding using timing (through the mixture-of-Poissons model) is more effective than decoding using spike count alone. *A, C*, Distribution across 29 (*A*) or 17 (*C*) neurons of information transmitted by the neuronal responses (spike trains) about which stimulus was shown, using expanding windows starting at stimulus onset. *Dark boxes*, Spike count only; *light boxes*, spike count and timing together. *B, D*, Distribution across 29 (*B*) or 17 (*D*) neurons of the percent of trials correctly decoded by guessing the stimulus with highest probability of having elicited the observed response. For comparability, figures are presented as multiples of the percent of trials that would be correctly decoded by chance (*B*,  $1/16 = 6.25\%$ ; *D*,  $1/128 = 0.78\%$ ). The *dotted line* shows the percent of trials correctly decoded by chance. We avoided overfitting using three-way cross-validation: the available trials were divided into three subsets, with two-thirds used to fit the model and the remaining third used to test decoding. Each third of the data was used twice to fit and once to test; the results shown here are averaged over the three sets. *Boxes*, Median; *line and indentation at center*, interquartile range: *bottom* and *top*. If *notches* do not overlap, corresponding medians are different ( $p < 0.05$ ). Whiskers have been eliminated, and the *dark boxes* have been made wider, for visibility.

7, *B* and *D*, shows the distribution (across 29 and 17 neurons, respectively) of the percent of trials correctly decoded, that is, for which the highest-probability stimulus did elicit the response. For comparability between the two experiments, the figures are shown as multiples of the percent of trials that could be correctly decoded simply by guessing ( $1/16 = 6.25\%$ , and  $1/128 = 0.78\%$ ). In both experiments, more trials were correctly decoded using both spike count and timing via the mixture-of-Poissons method than using spike count alone.

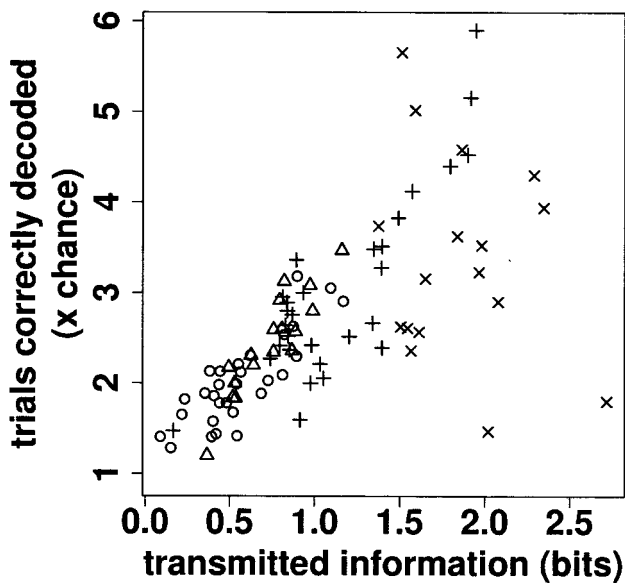
When timing is taken into account, the amount of information transmitted by the neuronal responses is larger in the experiment with 128 stimuli than in the experiment with 16 stimuli (Fig. 7*A, C*). This difference could be because one experiment uses more stimuli than the other, or it could be because the larger experiment used stimuli of several different kinds. To check whether the number of stimuli was the crucial factor, we randomly selected subsets of 64, 32, and 16 stimuli, allowing stimuli of all four kinds to enter each set. Figure 8*A* shows that this has very little effect on the amount of information transmitted by the



**Figure 8.** Differences in stimulus set, rather than in the number of stimuli, account for differences in the results of the two sets of experiments. *x*-Axis, Time from stimulus onset (milliseconds). *y*-Axis, Transmitted information (bits). *A*, Information transmitted by responses of a single neuron to different numbers of stimuli: 128 (*solid circles*), 64 (*plus symbols*), 32 (*open triangles*), and 16 (*open circles*). For all except the full set of 128 stimuli, the *lines* represent the median value of transmitted information over 100 random samples (from the full set of 128) of the appropriate number of stimuli. Transmitted information drops only slightly with the number of stimuli. *B*, Information transmitted by responses of the neuron to random subsamples of 32 stimuli (*filled circles*; same as *triangles* in *A*) and to subsets of 32 stimuli consisting of bars (Fig. 4*A*, *open circles*), gratings (Fig. 4*B*, *open triangles*), Walsh patterns (Fig. 4*C*; *plus symbols*), and photographic stimuli (Fig. 4*D*; *times symbols*). Much less information is transmitted by responses to Walsh patterns or photographic stimuli than by responses to bars, gratings, and random subsamples.

responses. Figure 8*B* compares the amount of information transmitted by responses to the randomly selected groups of 32 stimuli with the amount of information transmitted by the four subsets of 32 stimuli consisting of the bars, gratings, Walsh patterns, and photographic images, respectively (Fig. 4*A–D*). Responses to the Walsh patterns and photographs used in these experiments transmit less information than do the responses to bars and gratings.

Wiener et al. (2001), in an analysis using spike count alone, have previously shown that the difference in information transmission in neuronal responses to stimuli of the four different kinds is primarily attributable to the fact that responses to Walsh patterns and photographs have a smaller dynamic range than responses to bars and gratings. This new result shows that the difference persists in an analysis that takes into account spike timing as well. The persistent difference in results obtained using different kinds of stimuli reinforces a point made by Wiener et al. (2001): conclusions based on experiments using a narrow set of



**Figure 9.** Relation across neurons between transmitted information and percent of trials correctly decoded for individual neurons. *x*-Axis, Transmitted information (bits). *y*-Axis, Percent of trials correctly decoded (as a multiple of the number of trials that would be correctly decoded by chance). Each symbol shows transmitted information and the percent of trials correctly decoded 300 msec after stimulus presentation for a single neuron. The percent of trials correctly decoded increases with increasing information transmitted, except when decoding using the mixture-of-Poissons method in the experiment with 128 stimuli. *Circles*, Decoding using the spike count in the experiment with 16 stimuli; *triangles*, decoding using the spike count only in the experiment with 128 stimuli; *plus symbols*, decoding using the mixture-of-Poissons method in the experiment with 16 stimuli; *times symbols*, decoding using the mixture-of-Poissons method in the experiment with 128 stimuli.

stimuli may not generalize well to experiments using a larger set. The need for a large number of stimuli unfortunately conflicts with the need for many trials per stimulus in an experiment with a limited number of trials.

### Relation between decoding success and transmitted information

Intuitively, we expect transmitted information and decoding success to be related: if more information is transmitted by the responses, more responses should be correctly decoded. Above (Fig. 7) we showed that in both sets of experiments (with 16 and 128 stimuli), transmitted information and percent of trials correctly decoded both increase as the trials progress. Figure 9 shows that, for individual neurons from the two sets of experiments, more information transmitted generally corresponds to a greater portion of trials correctly decoded.

### Effect of refractory period

Because we estimate the spike count distribution from data, we incorporate the effect of refractory period on spike count. We found that the effect of any additional refractory or rebound period on the number of trials correctly decoded is very small: 0.2% fewer trials are correctly decoded when the refractory period is included (median across neurons; iqr,  $-2.7$  to  $2.2\%$ ). The effect on individual stimulus probabilities was also small: in the experiments using 16 stimuli, the median correlation between the estimated stimulus probabilities obtained when taking the additional refractory effects into account and those obtained when not taking the additional refractory effects into account was 0.997 (median across all trials in 29 neurons), and the fifth per-

centile of correlations was 0.96. The median correlation in the experiments using 128 stimuli was 0.999 (median across all trials in 17 neurons), and the fifth percentile was 0.99. Except where otherwise noted, the results throughout this study were calculated using the refractory period.

### Decoding signal versus decoding noise

When there are no differences in timing among responses elicited by different stimuli, we would expect the decoder to reduce to a spike count decoder. To check this, we simulated two sets of responses with identical spike density functions (normalized PSTH) but different distributions of spike count (Poisson distributions with means of 4 and 10). The only nonrandom differences in timing arise from the differences in spike count. A trial with seven spikes should, at the end of the trial, be assigned with probability 0.4 to the Poisson distribution with mean 4 and with probability 0.6 to the Poisson distribution with mean 10. In fact, depending on the data used to create the Poisson models, different spike trains of length 7 will be assigned different probabilities, but we expect the assigned probabilities to converge to  $p = 0.4$  and 0.6. Figure 10A shows that when spike timing is known to be random, decoding on the basis of timing (using the mixture-of-Poissons model; *white boxes*) converges to the correct answer more slowly than the optimal decoding using spike count alone (*black boxes*).

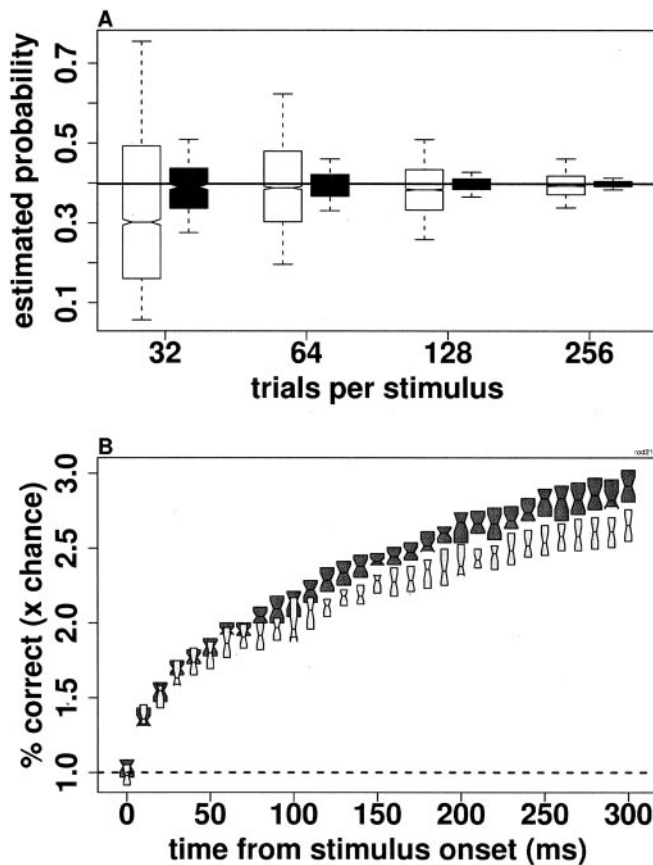
A similar effect can be observed in surrogate data generated to match the spike count distribution seen in one of the neurons from the experiment with 16 stimuli but with spikes equally likely at any time (i.e., with a flat PSTH) for all stimuli. Here again, spike timing is purely random and carries no information about which stimulus generated the response. Figure 10B shows that more trials are correctly decoded using the spike count alone than using the full mixture-of-Poissons model. Again, attempting to decode using random timing degrades performance.

In the cases just presented, the spike density function is (by construction) of no help in decoding but still must be estimated. When a small amount of data is available, random spike time variations cause differences in the estimates of the spike density functions for different stimuli, and these spurious differences interfere with decoding. The random differences in the spike density functions become smaller when more data are available to estimate them (although true differences in spike density functions can be detectable even with modest amounts of data). The need for data to eliminate spurious correlations would only increase if more complicated timing features were to be taken into account. Thus it is important to avoid trying to decode on the basis of complicated timing features that can be accounted for as stochastic consequences of the spike count and simpler timing features.

### Comparing different decoding variants

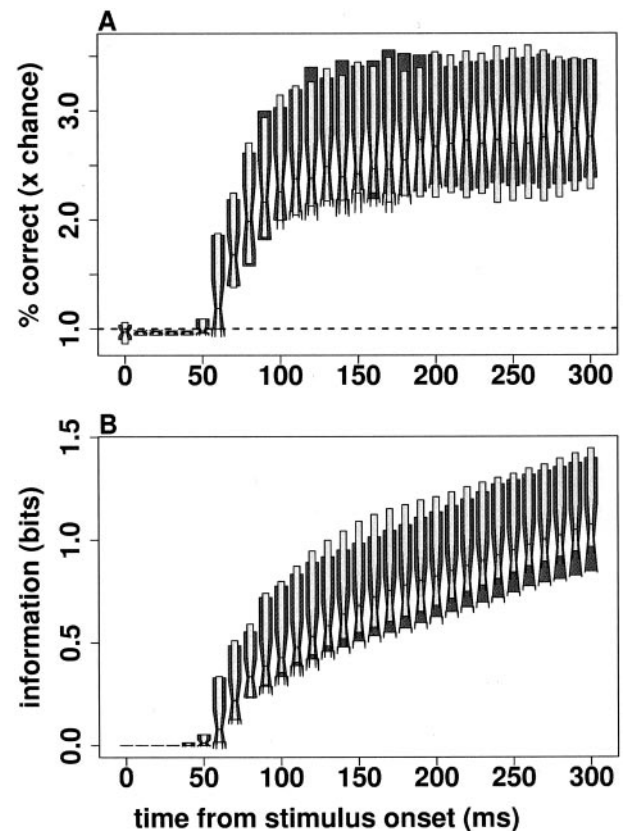
Trains from the neurons from the experiment with 16 stimuli were decoded using both the order statistic method (with the spike count distributions represented using mixtures of Poisson distributions) and the simplified mixture-of-Poissons method. The correlation between the two sets of probabilities at the end of the trial (that is, 300 msec after stimulus onset) was 0.997 (median across neurons; iqr, 0.992–0.999). This verifies that the two methods give nearly identical results (up to certain issues related to time discretization near the end of a trial). As a result, the same stimulus was guessed by the two methods in 95% of the cases (median across neurons; iqr, 94–98%). It may seem odd that different stimuli could be guessed when the correlation between





**Figure 10.** Decoding using the mixture-of-Poissons method is inferior to spike count decoding when there are no timing differences among responses to different stimuli. *A*, Mixture-of-Poissons decoding converges to the correct answer less quickly than does spike count decoding in a two-stimulus example with identical PSTH for both stimuli. *x*-Axis, Number of trials per stimulus. *y*-Axis, Distribution of estimated probabilities that a train with seven spikes arose from a homogeneous Poisson process with a mean of 4 spikes rather than from one with a mean of 10 spikes. The *left* (open) box in each set shows results using the mixture-of-Poissons decoding method. The *right* (filled) box shows results when decoding on the basis of spike count alone. The simulated data were generated by a homogeneous Poisson process with a mean of 4 for one stimulus and a homogeneous Poisson process with mean of 10 for the other stimulus. The mean number of spikes for each process was held fixed while the number of trials changed. Similar results were obtained using inhomogeneous Poisson processes as long as the rates varied in time in the same way (i.e., the normalized PSTHs were the same for the two processes). The *left* box in each set represents 2500 values, the results of decoding 500 randomly generated test trains using models derived from five different sets of simulated data. The *right* box represents 250 values, the result of decoding the response “seven spikes” according to the models derived from 250 sets of simulated data (using additional sets of simulated data does not change the results). Interpretation of *box plots*: The *line* at the center of each *box* shows the median estimated probability, and the *notch* shows a 5% confidence interval for the median. The *bottom* and *top* of the *box* show the 25th and 75th percentiles, and the *bottom* and *top* *whiskers* show the 5th and 95th percentiles. *B*, *x*-Axis, Time from stimulus onset (milliseconds). *y*-axis, Number of trials correctly decoded (as a multiple of chance) using the spike count only (*dark boxes*) and the mixture-of-Poissons method (*light boxes*). Each *box plot* represents results from 25 artificial data sets with spike distributions modeled on those from one of the neurons from the experiment with 16 stimuli and no differences in the timing of stimuli as above. Interpretation of *box plots* is as above (whiskers removed for visibility).

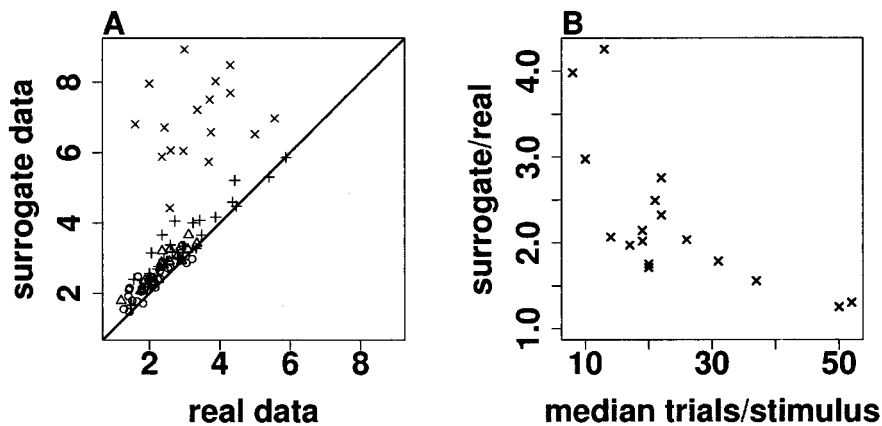
probabilities is so high. Table 1 shows that the difference between the probability of the winning stimulus and the next most probable stimulus can be quite small; thus small discrepancies can change the guessed stimulus in a small number of trials. As would be expected, the differences in the largest and second-largest probabilities in trials when the two methods made the same guess were  $\sim 5$  times as large as the differences in trials when the two



**Figure 11.** Decoding results using the mixture-of-Poissons model to estimate spike count distributions on subintervals (*dark boxes*) are similar to those obtained when directly measuring spike count distributions on subintervals (*light boxes*). *A*, The portion of trials correctly decoded is similar for the two methods. *x*-Axis, Time from stimulus onset (milliseconds). *y*-axis, Percent of trials correctly decoded (as a multiple of chance). The *dotted line* shows the approximate percent of trials correctly decoded by chance. *B*, Using measured distributions results in higher transmitted information than using estimated distributions. *x*-Axis, Time from stimulus onset (milliseconds). *y*-Axis, Transmitted information. *Light boxes* are narrower for visibility only. This graph uses the data from the experiment with 16 stimuli (Kjaer et al., 1997). Interpretation of *box plots* is as in Figure 7.

methods made different guesses. We did not decode the experiments with 128 stimuli using the order statistic method because of the computational burden.

As previously presented, when using the mixture-of-Poissons model, we have modeled the distribution of mean spike counts in any subinterval of the trial (from any particular time to the end of the trial) on the basis of the distribution of mean spike counts in the entire trial [ $\lambda_{s,t}$  and  $p(\lambda)$  at the start of the trial] and the firing rate modulation over time (Eq. 11). It is possible, at substantial expense in both computation time and additional model complexity, to avoid this approximation by measuring the distribution of spike counts in each subinterval and finding a new mixture of Poisson distributions to model each distribution. Despite the additional model complexity and computation, decoding using measured subinterval distributions gives results similar, but not identical, to those obtained when using estimated subinterval distributions: the correlation between the two sets of probabilities at the end of trials was 0.88 (iqr, 0.83–0.94), and the same stimulus was guessed in 75% of the trials (median across neurons; iqr, 65–82%). Approximately the same number of trials were correctly decoded using the two methods (Fig. 11*A*), although using the measured distributions yielded slightly more transmitted information (Fig. 11*B*).



**Figure 12.** Performance of the decoder on real data and matched surrogate data. *A*, Percent of trials correctly decoded (in multiples of chance) for the real data (*x*-axis) and for matched surrogate data (*y*-axis). Results are shown both when decoding using the spike count only (circles, 16 stimuli; triangles, 128 stimuli) and when decoding using the mixture-of-Poissons method (plus symbols, 16 stimuli; times symbols, 128 stimuli). Each symbol shows results for a single neuron, and each neuron is represented twice, once for decoding using the spike count and once for decoding using the mixture-of-Poissons method. The figures for surrogate data are medians across 10 artificial data sets. Real data are decoded nearly as well as surrogate data when using the spike count code and even when using the mixture-of-Poissons method for the experiment with 16 stimuli. However, for data from the experiment with 128 stimuli, the decoder performs much better on surrogate data than real data. *B*, The performance of the decoder depends on the number of trials per stimulus available. *x*-Axis, Median number of trials per stimulus. *y*-Axis, Ratio of the number of trials correctly decoded for surrogate data to the number of trials correctly decoded in real data. Only data from the experiment with 128 stimuli decoded using the mixture-of-Poissons method are shown (times symbols are used for consistency with *A*).

Decoding using the order statistic method or the mixture-of-Poissons method with directly measured spike count distributions on subintervals takes substantially longer (by approximately an order of magnitude) than using the mixture-of-Poissons model when estimating the count distributions on subintervals. The results of this section show that the extra time produces little change in decoding.

### How good is the mixture-of-Poissons model?

#### Time-rescaling goodness-of-fit test

Barbieri et al. (2001) and Brown et al. (2002) presented a method for checking the goodness of fit of spike-train models. Their method is based on the time-rescaling theorem, which shows that any point process can be transformed, by a suitable rescaling of time, into a homogeneous Poisson process. To apply their method to our data, we randomly chose two-thirds of the trains to estimate the model, withholding one-third of the trains to use as a test set. Each model was tested using the Kolmogorov–Smirnov test presented by Brown et al. (2002), with  $p = 0.95$ . In the neurons from the experiment with 128 stimuli, the test spike trains were consistent with the model for 109 stimuli (median across neurons; iqr, 104–112). In the neurons from the experiment with 16 stimuli, the test spike trains were consistent with the model for 11 stimuli (median across neurons; iqr, 10–13). When the refractory–rebound effect was ignored, the results were very similar: the test spike trains were still consistent with the model for 108 of 128 stimuli (median across neurons; iqr, 104–111) and in 11 of 16 stimuli (median across neurons; iqr, 9–13). Overall, 82% of stimuli were consistent with the model when the refractory period was included in the model, and 78% were still consistent when the refractory period was ignored. Thus the mixture of Poisson distributions with different mean rates, and therefore different temporal structures, seems to account for much of the observed temporal structure of the spike trains. This may explain

why our decoding results with or without the refractory period are so similar.

#### Number of trials correctly decoded

Another way to examine how well our model describes the data is to compare the performance of the decoder on real data with its performance on artificially generated surrogate data that match the real data in many ways and are known to have the structure specified by the model. If our model describes the data well, we would expect to decode real spike trains nearly as well as we decode artificially generated spike trains. If our model describes the data poorly, we would expect to decode artificial spike trains much more accurately than real spike trains.

We generated cell-matched surrogate data as described in Materials and Methods. Figure 12*A* compares the percent of trials correctly decoded in real and surrogate data, both for the spike count code and using the mixture-of-Poissons model. When decoding using the spike count only, we correctly decode 15% (median across neurons; iqr, 7–20%) more trials from surrogate data than from real data in the experiments with 16 stimuli (circles) and 18% (median; iqr, 10–23%) more in the experiments with 128 stimuli (triangles). The decoder does nearly as well when decoding data from the experiments with 16 stimuli using the mixture-of-Poissons method (plus symbols): we correctly decode 17% (median; iqr, 7–23%) more trials from surrogate data than from real data. However, when the mixture-of-Poissons method is used with data from the experiments with 128 stimuli (times symbols), many more trials are correctly decoded from surrogate data than from real data (median, 100% more; iqr, 75–150%). Figure 12*B* shows that the performance of the mixture-of-Poissons decoder is related to the amount of data available: the more data available to estimate the model parameters in the first place, the more closely the number of trials correctly decoded in real data approaches the number correctly decoded in surrogate data.

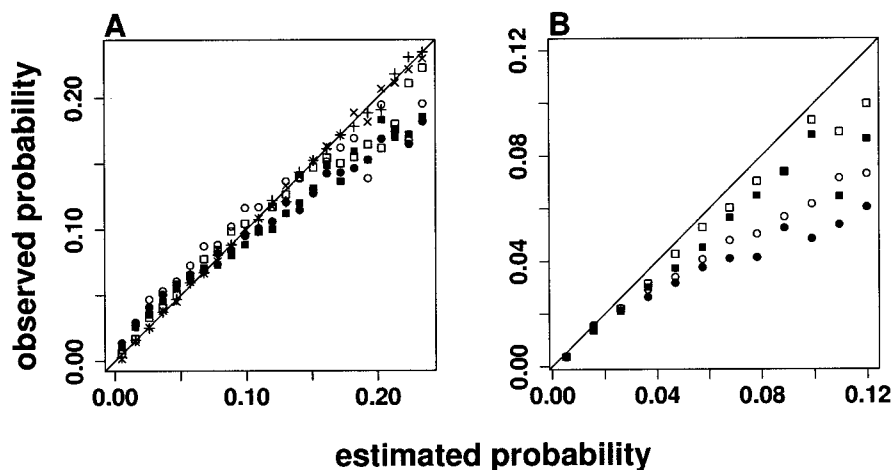
To check whether the number of stimuli involved (and therefore the complexity of the problem) contributes to the difference between the two sets of results, we created 16-stimulus neurons from 128-stimulus neurons by taking the responses from 16 randomly chosen stimuli and discarding the rest. Twenty-five sub-neurons were created from each of the two neurons with the most trials per stimulus (50 and 52). In these artificial neurons, 16% more trials are correctly decoded from surrogate data than from the real data (median; iqr, 11–20%), not significantly different from that seen for the neurons from the experiment that actually had only 16 stimuli (Kruskal–Wallis test,  $p = 0.7$ ). There is still no significant difference if we try to minimize the influence of the number of trials per stimulus by comparing the results for the subsampled data only with results from the seven neurons from the experiment with 16 stimuli with between 40 and 60 trials per stimulus (Kruskal–Wallis test,  $p = 0.2$ ). Thus correctly choosing among 128 stimuli requires more data to parameterize the model than correctly choosing among 16 stimuli.

### Accuracy of individual stimulus probabilities: model error and estimation error

The most stringent possible test of the performance of the decoder assesses the accuracy of all the estimated stimulus probabilities. For example, if we look at all the stimuli assigned probabilities near  $p = 0.1$  in various trials, we would hope that those stimuli were actually the stimuli that elicited those trials 10% of the time. This is essentially the same test we might use to evaluate the quality of a weather-forecasting service: it should rain on half the days with a 50% chance of rain, one-fourth of the days with a 25% chance of rain, and so on. To test this, we bin the estimated probabilities (we have one estimated probability for each stimulus for each trial) and then ask how often the stimuli represented in each bin really did elicit the corresponding trial.

Departure from accurate prediction can come from two sources. The first possible source of error is model misspecification: the mixture-of-Poissons model may not capture the true structure of the spike trains. The second possible source of error is difficulty fitting the model on the basis of the amount of data available. We will call these model error and estimation error. To separate these errors, we must examine how accurately the model estimates stimulus probabilities in three different kinds of data: the real data; artificial data, generated according to the model, with the same number of trials per stimulus as the real data; and artificial data, generated according to the model, with many more trials per stimulus than available in the real data. Model error can be examined by comparing results from the real data and the smaller set of artificial data: because the two sets have equal numbers of trials per stimulus, any difference in our ability to decode must derive from the fact that one set of data is known to be consistent with the model, whereas the other is not. Estimation error can be examined by comparing results from the larger and smaller artificial data sets: because both artificial data sets are known to be consistent with the model, any difference in our ability to decode must derive from difference in the amount of data available.

Figure 13A shows that, for the data from the experiment with 16 stimuli, the predicted stimulus probabilities are close to the true stimulus probabilities. Small predicted stimulus probabilities tend to be slightly too large, and large predicted stimulus probabilities tend to be slightly too small. This is true whether we decode using the spike count only (*filled squares*) or using the full mixture-of-Poissons model (*filled circles*), although the difference is smaller for spike count only. The same effect is seen in artificial data with the same number of trials per stimulus as seen in the real data, although it is substantially less pronounced (*open squares*, *open circles*). This indicates that some of our error is attributable to the model not quite fitting the data. However, in artificial data with many more trials per stimulus than in our actual data (500 trials per stimulus), the predicted probabilities are almost exactly equal to the observed probabilities both when the spike trains are decoded using the spike count only (*plus symbols*) and when they are decoded using the full model (*times*



**Figure 13.** Accuracy of estimated stimulus probabilities. *x*-Axis, Estimated stimulus probability. *y*-Axis, Observed probability. Observed probabilities are calculated by binning the estimated probabilities (one for each stimulus for each trial) and then asking how often the stimuli represented in each bin really did elicit the corresponding trial. If our model described the data perfectly, and we had sufficient data to estimate the model, the estimated and observed probabilities would be identical. In both panels, the number of probabilities in each bin drops sharply as the estimated probability increases. *A*, Results from 29 neurons from the experiment with 16 stimuli. *Filled symbols* show results for the real data; *open symbols* show results for artificial data with the same number of trials per stimulus as the real data; *times* and *plus symbols* show results for very large artificial data sets (500 trials per stimulus). *Squares* and *plus symbols* show results when decoding using the spike count only; *circles* and *times symbols* show results when decoding using the full mixture-of-Poissons model. *B*, Results from the two neurons with the most (50) trials per stimulus from the experiment with 128 stimuli. Symbols are as in *A*. Results for extremely large artificial data sets are not included because of the computational burden. Note different scales in the two panels.

*symbols*). The excellent agreement between estimated and actual stimulus probabilities obtained on large artificial data sets serves as confirmation that the decoder works as it is supposed to and also indicates that part of the error seen with real data sets is attributable to difficulty correctly estimating the model parameters (the spike density functions and the mixtures of Poissons for spike count distributions) using the amount of data available.

Figure 13B is similar to Figure 13A but for data from the two neurons with the most (50) trials per stimulus in the experiments with 128 stimuli. Here the predicted probabilities are almost always lower than the true probabilities, whether decoding using the spike count only or using the mixture-of-Poissons model. Because of the amount of computation involved, we do not include results for extremely large artificial data sets in Figure 13B, but the results in Figure 13A imply that the decoder works properly, so for a sufficiently large artificial data set, the probabilities would lie along the identity line. Figure 13A shows that most of our misestimation of probabilities can be ascribed to limited sample size rather than to mis-specification of the model (i.e., the distance between the *filled* and *open symbols* is smaller than the distance between the *open symbols* and the identity line), particularly for the full mixture-of-Poissons model. Excluding the three smallest probability bins (in which the error is extremely small), estimation error accounts for approximately three-fifths of the total estimation error when decoding using spike count only and for almost four-fifths of the error when decoding using the full mixture-of-Poissons model. The fact that the decoder estimates probabilities less well in the data with 128 stimuli than in the data with 16 stimuli again shows that more data are required to accurately solve the more difficult problem.

## Discussion

In this paper we present the order statistic model of spike trains, which is based on the observation that in several brain areas, single-neuron spike trains are almost entirely described as sto-



chastic samples from the firing rate profile (Oram et al., 1999, 2001; Baker and Lemon, 2000). In accordance with these observations, in the order statistic model, individual spike times are informative because they reflect underlying rate variation. This model requires estimating only the spike count distribution, the spike density function, and the interspike interval distribution from data. All of these can be reasonably estimated using amounts of data gathered in virtually every experiment. Our results show that for decoding problems with many conditions, decoding performance is quite good with amounts of data frequently collected, but very large amounts of data may be needed for optimal decoding.

The order statistic model is general and can be used with any estimate of the spike count distribution, modeled or measured. The mixture-of-Poissons model gives up some of this generality, modeling each spike count distribution as a mixture of Poisson distributions and the spike trains as instances of a mixture of Poisson processes. In exchange, the Poisson model speeds up the calculation of order statistics; we calculate distributions of next spike times for a few (up to five) Poisson means instead of for many (up to 50) spike counts (Eqs. 8, 10 vs 6, 7). For our data, decoding using the Poisson model was approximately an order of magnitude faster, and the extra generality of the order statistic model was not needed: >98% of the spike count distributions were fit with mixtures of three or fewer Poisson distributions, and all were fit by mixtures of five or fewer. A mixture of Poisson processes may or may not fit a particular data set. Other structures identified in particular data sets might help simplify calculating order statistics in a similar way, although the details of the calculation would be different.

Any decoding method must decide which aspects of a response are important, that is, carry information that can be used to distinguish among stimuli, and which can be ignored. Ignoring spike timing that does carry information reduces decoding accuracy (Fig. 7), but so does paying attention to timing that does not carry information (Fig. 10). When spike timing is taken into account, there are so many distinct responses that a great deal of data is needed to determine whether apparent correlations between stimulus and response are reliable or merely (as in the examples of Fig. 10) sampling artifacts. Thus decoding using timing is difficult precisely because timing presents great opportunities for encoding. Here we avoid the difficulties of directly counting how often each response is elicited by using a model that captures the structure of observed responses, including high-order correlations among spike times (Oram et al., 1999, 2001). This model leads to simple decoding methods that are much more effective than decoding based on spike count alone.

The decoding scheme described here assumes that the response being decoded was generated by one of the stimuli, and in that case, we checked in a small amount of data from area TE of monkey temporal lobe that the decoder works even when spikes occur between trials. We can also treat the no-stimulus, or inter-trial, condition as just another experimental condition. Even in this case, the decoding algorithm assumes that the time of stimulus onset is known. Several groups have suggested that eye movement or other causes of large-scale change in the visual field cause neural responses that can be used as synchronizing or reset signals (Sobotka and Ringo, 1997; Sobotka et al., 1997; Huang and Paradiso, 2000; Greschner et al., 2002).

The question of response alignment is related to the question of response latency. Once responses are aligned on stimulus onset, systematic differences in response latency are reflected in the firing rate profile over time and possibly also in the spike count

distribution. Figures 1 and 3 show examples, based on real data, of constructing order statistics and decoding real data with a latency difference. Rhythmic firing phase-locked to stimulus onset would also be reflected in the PSTH. Rhythmic firing not phase-locked to stimulus onset would not show up in the PSTH but might be accounted for by the refractory-rebound term. Any additional underlying statistical structure different from that assumed by the model might require additional modeling work to accommodate.

### Other models and methods

The models presented here can be estimated using modest amounts of data, because they assume that spike trains have a certain stochastic structure. This contrasts with the “direct method,” which makes no assumptions about the form of the neural code but requires very large data sets, because it directly counts how often each stimulus elicits each response. The direct method has been used to calculate information transmission rates in the fly (Strong et al., 1998), in anesthetized monkeys (Reinagel and Reid, 2000), and in isolated retina (Nirenberg et al., 2001). However, in awake monkeys the amount of data gathered has not been sufficient to estimate information transmitted by spike timing; only calculations using the spike count in single bins have been possible (Reich et al., 2001). Thus, at least in awake monkeys, the direct method cannot currently substitute for finding some principle allowing us to compactly describe spike trains, that is, a model.

Others researchers have modeled spike trains using variants of Poisson processes. Lansky and Vaillant (2000) and Lansky et al. (2001) model the responses of hippocampal place cells using mixtures of two Poisson processes. Kass and Ventura (2001) model firing probability as a product of time from stimulus onset and a refractory term depending on the time of the last spike (i.e., an inhomogeneous Poisson process with a refractory period); they estimate the two components together using a generalized linear model, whereas we estimated them separately. Barbieri et al. (2001) model spike trains by characterizing interspike intervals using gamma and inverse Gaussian distributions (generalizing the exponential distribution of intervals arising from a Poisson process). They use time rescaling to account for inhomogeneities in firing rate, inducing a Markov dependence similar to that seen in the order statistic and mixture-of-Poissons models presented here. One possible use of statistical models such as the mixture-of-Poissons model and the others described above is to provide a benchmark against which to evaluate spike trains produced by biophysical models that address the cellular mechanisms of neuronal information processing.

Another way to take timing into account when decoding is to sequentially decode the spike count in different time bins. In the limit of bins that can contain at most one spike, the decoder simply asks, at each time, “Did a spike occur now? . . . now? . . . now?” This is equivalent to decoding using order statistics or the mixture-of-Poissons model only if the responses in all bins are independent. The Markov dependence built into the order statistic and mixture-of-Poissons models seems to incorporate most of the correlations observed in data (Oram et al., 1999).

We decode to determine which of a categorical set of stimuli was presented, without regard to stimulus characteristics such as bar or grating orientation. Other groups have developed approaches to reconstruct continuously varying signals (Gabbiani and Koch, 1996; Rieke et al., 1997; Brown et al., 1998; Schwartz and Moran, 2000; Wessberg et al., 2000; Manwani et al., 2001; Serruya et al., 2002). An appropriate discretization of the contin-

uous signal might allow our methods to be applied to these problems. This might require finding a time scale on which the PSTH gave information about the signal (that is, how long a section of signal should be treated as a single stimulus in our formulation), as well as imposing probabilistic continuity conditions (as by Brown et al., 1998).

### Conclusion

The results of Oram et al. (1999) suggest that, at least in V1 and the lateral geniculate nucleus, individual spike times are random and carry information only by reflecting underlying rate variation. Other features of precise spike timing can be predicted from spike count and the PSTH and therefore cannot carry information unavailable from these coarser measures. In this context, order statistics (Arnold et al., 1992), whether calculated directly or in special cases such as a mixture of Poisson processes, formalizes the intimate connection between spike timing and spike count (Wiener and Richmond, 1999; Oram et al., 1999). Because spike count distributions are often non-Poisson (Baddeley et al., 1997; Gershon et al., 1998), single inhomogeneous Poisson models cannot be expected to match features of precise timing. Therefore, the mixture-of-Poissons model is a more appropriate null hypothesis when searching in neuronal responses for timing relations that are unexpected, and which therefore may carry unique information.

### References

- Arnold BC, Balakrishnan N, Nagaraja HN (1992) A first course in order statistics. New York: Wiley.
- Baddeley R, Abbott LF, Booth MCA, Sengpiel F, Freeman T, Wakeman EA, Rolls ET (1997) Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc R Soc Lond B Biol Sci* 264:1775–1783.
- Baker SN, Lemon RN (2000) Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. *J Neurophysiol* 84:1770–1780.
- Barbieri R, Quirk MC, Frank LM, Wilson MA, Brown EN (2001) Construction and analysis of non-Poisson stimulus-response models of neural spiking activity. *J Neurosci Methods* 105:25–37.
- Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA (1998) A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J Neurosci* 18:7411–7425.
- Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM (2002) The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput* 14:325–346.
- Cover TM, Thomas JA (1991) Elements of information theory. New York: Wiley.
- Craven P, Wahba G (1979) Smoothing noisy data with spline functions. *Numerische Mathematik* 31:377–403.
- Gabbiani F, Koch C (1996) Coding of time-varying signals in spike trains of integrate-and-fire neurons with random threshold. *Neural Comput* 8:44–66.
- Gershon ED, Wiener MC, Latham PE, Richmond BJ (1998) Coding strategies in monkey V1 and inferior temporal cortices. *J Neurophysiol* 79:1135–1144.
- Golomb D, Hertz J, Panzeri S, Treves A, Richmond BJ (1997) How well can we estimate the information carried in neuronal responses from limited samples? *Neural Comput* 9:649–665.
- Greschner M, Bongard M, Rujan P, Ammermuller J (2002) Retinal ganglion cell synchronization by fixational eye movements improves feature estimation. *Nat Neurosci* 5:341–347.
- Huang X, Paradiso MA (2000) Stimulus structure and expectation reflected in the delayed responses of macaque V1 neurons. *Soc Neurosci Abstr* 26:559.
- Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *J Comput Graphical Stat* 5:299–314.
- Kass RE, Ventura V (2001) A spike-train probability model. *Neural Comput* 13:1713–1720.
- Kjaer TW, Gawne TJ, Hertz JA, Richmond BJ (1997) Insensitivity of V1 complex cells to small shifts in the retinal image of complex patterns. *J Neurophysiol* 78:3187–3197.
- Lansky P, Vaillant J (2000) Stochastic model of the overdispersion in the place cell discharge. *Biosystems* 58:27–32.
- Lansky P, Fenton AA, Vaillant J (2001) The overdispersion in activity of place cells. *Neurocomputing* 38:1393–1399.
- Loader C (1999) Local regression and likelihood. New York: Springer.
- Manwani A, Steinmetz PN, Koch C (2001) The impact of spike timing variability on the signal-encoding performance of neural spiking models. *Neural Comput* 14:347–367.
- Nirenberg S, Carcieri SM, Jacobs AL, Latham PE (2001) Retinal ganglion cells act largely as independent encoders. *Nature* 411:698–701.
- Oram MW, Wiener MC, Lestienne R, Richmond BJ (1999) The stochastic nature of precisely timed spike patterns in visual system neuronal responses. *J Neurophysiol* 81:3021–3033.
- Oram MW, Hatsopoulos NG, Richmond BJ, Donoghue JP (2001) Synchrony in motor cortical neurons provides direction information that is redundant with the information from coarse temporal response measures. *J Neurophysiol* 86:1700–1716.
- Panzeri S, Treves A (1996) Analytical estimates of limited sampling biases in different information measures. *Network* 7:87–107.
- Reich DS, Mechler F, Victor JD (2001) Formal and attribute-specific information in primary visual cortex. *J Neurophysiol* 85:305–318.
- Reinagel P, Reid RC (2000) Temporal coding of visual information in the thalamus. *J Neurosci* 20:5392–5400.
- Rieke F, Warland D, de Ruyter van Steveninck RR, Bialek W (1997) Spikes: a Bradford book. Cambridge, MA: MIT.
- Schwartz AB, Moran DW (2000) Arm trajectory and representation of movement processing in motor cortical activity. *Eur J Neurosci* 12:1851–1856.
- Serruya MD, Hatsopoulos NG, Paninski L, Fellows MR, Donoghue JP (2002) Brain-machine interface: instant neural control of a movement signal. *Nature* 416:141–142.
- Shannon CE, Weaver W (1949) The mathematical theory of communication. Urbana, IL: University of Illinois.
- Sobotka S, Ringo JL (1997) Saccadic eye movements, even in darkness, generate event-related potentials recorded in medial sputum and medial temporal cortex. *Brain Res* 756:168–173.
- Sobotka S, Nowicka A, Ringo JL (1997) Activity linked to externally cued saccades in single units recorded from hippocampal, parahippocampal, and inferotemporal areas of macaques. *J Neurophysiol* 78:2156–2163.
- Strong SP, Koberle R, van Steveninck RDR, Bialek W (1998) Entropy and information in neural spike trains. *Phys Rev Lett* 80:197–200.
- Wessberg J, Stambaugh CR, Kralik JD, Beck PD, Laubach M, Chapin JK, Kim J, Biggs SJ, Srinivasan MA, Nicolelis MAL (2000) Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* 408:361–365.
- Wiener MC, Richmond BJ (1998) Using response models to study coding strategies in monkey visual cortex. *Biosystems* 48:279–286.
- Wiener MC, Richmond BJ (1999) Using response models to estimate channel capacity for neuronal classification of stationary visual stimuli using temporal coding. *J Neurophysiol* 82:2861–2875.
- Wiener MC, Oram MW, Liu Z, Richmond BJ (2001) Consistency of encoding in monkey visual cortex. *J Neurosci* 21:8210–8221.