Consistency of Encoding in Monkey Visual Cortex

Matthew C. Wiener, Mike W. Oram, Zheng Liu, and Barry J. Richmond

Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892-4415

Are different kinds of stimuli (for example, different classes of geometric images or naturalistic images) encoded differently by visual cortex, or are the principles of encoding the same for all stimuli? We examine two response properties: (1) the range of spike counts that can be elicited from a neuron in epochs representative of short periods of fixation (up to 400 msec), and (2) the relation between mean and variance of spike counts elicited by different stimuli, that together characterize the information processing capabilities of a neuron using the spike count code. In monkey primary visual cortex (V1) complex cells, we examine responses elicited by static stimuli of four kinds (photographic images, bars, gratings, and Walsh patterns); in area TE of inferior temporal cortex, we examine responses elicited by static stimuli in the sample, nonmatch, and match phases of a delayed match-to-sample task. In each area, the ranges of mean spike counts and the relation between mean and variance of spike counts elicited are sufficiently similar across experimental conditions that information transmission is unaffected by the differences across stimulus set or behavioral conditions [although in 10 of 27 (37%) of the V1 neurons there are statistically significant but small differences, the median difference in transmitted information for these neurons was 0.9%]. Encoding therefore appears to be consistent across experimental conditions for neurons in both V1 and TE, and downstream neurons could decode all incoming signals using a single set of rules.

Key words: coding; visual; cortex; V1; TE; inferior temporal cortex; monkey; natural images; mean–variance relation

Many different kinds of visual stimuli are used in neurophysiological experiments. This raises the question of whether results obtained using one class of stimuli can be expected to hold for others. For example, photographic or naturalistic images might somehow be processed differently from geometric stimuli frequently used in experiments. Previously, we have shown that knowledge of the operating range of the responses of a neuron, along with the linear relation between log(mean) and log(variance) of spike counts elicited by different stimuli (Dean, 1981; Tolhurst et al., 1981, 1983; van Kan et al., 1985; Vogels et al., 1989; Britten et al., 1993; Levine et al., 1996; Bair and O'Keefe, 1998; Gershon et al., 1998; Lee et al., 1998), characterizes the information processing capacity of a neuron using the spike count code (Gershon et al., 1998; Wiener and Richmond, 1998). If different classes of stimuli are encoded differently, responses to those classes of stimuli might have different operating ranges and/or might give rise to a different relation between mean and variance than observed for other stimuli. In this paper, we address this question in primary visual cortex (V1) of awake monkeys. We also examine how behavioral context affects visual responses in area TE of inferior temporal cortex.

We examine operating range and the relation between mean and variance in responses (here, spike counts) that are elicited from V1 by four kinds of stimuli: three kinds of geometric stimuli, i.e., bars, sine-wave gratings, and Walsh patterns, and photographic images, which are often used to study the statistics and processing of natural images (Field, 1987; Atick and Redlich, 1990, 1992; Rolls and Tovee, 1995; Dan et al., 1996; Olshausen and Field, 1996b; Bell and Sejnowski, 1997; van Hateren and Ruderman, 1998; van Hateren and van der Schaaf, 1998; Vinje and Gallant, 2000). In area TE, in which the physical properties of stimuli are integrated with the behavioral context in which they are viewed (Spitzer and Richmond, 1991; Eskandar et al., 1992; Chelazzi et al., 1998; Liu and Richmond, 2000), we examine whether behavioral context (whether an animal is in the sample, nonmatch, or match phase of a delayed match-to-sample task) affects those same response properties. We find significant but small differences in range of mean spike counts elicited from V1 neurons by stimuli of different kinds and, in 10 of 27 neurons, in the relation between log(mean) and log(variance). Although the differences in the mean–variance relation do not affect the ability of the neuron to distinguish among stimuli, the differences in range of mean spike counts make it slightly more difficult to use the responses of the neuron to distinguish among Walsh patterns or photographic images than to distinguish among bar or grating stimuli. Estimates of the information processing capacity of the neuron are consistent across stimulus sets. We find small but significant differences in the largest mean counts that are elicited from neurons in TE, but not in the relation between mean and variance in different behavioral contexts.

MATERIALS AND METHODS

Data collection

V1. Responses were recorded using standard single-electrode techniques from complex cells in primary visual cortex of two awake rhesus monkeys. At the beginning of each trial, a fixation point appeared on the screen. One hundred milliseconds after the monkey fixated the point, a stimulus was flashed on the receptive field of the neuron for 300 msec and then replaced with the background. The monkey was rewarded for fixating within 0.5° of the fixation point from the appearance of the
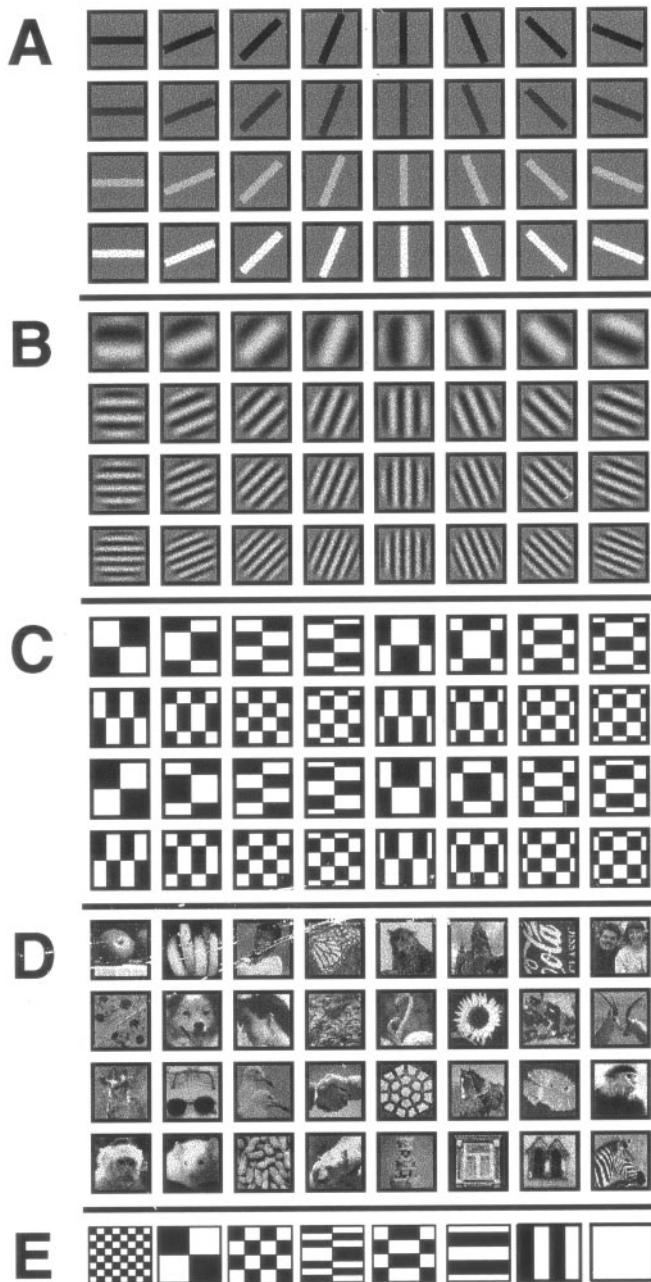
Figure 1. Stimuli used in the experiments. For the experiment in V1, the stimulus set consisted of 32 oriented bars (*A*), 32 sine-wave gratings (*B*), 32 Walsh patterns (*C*), and 32 photographic images (*D*). All stimuli except the bars were equiluminant with the background at 1.2 cd/m^2. Bars appeared at luminances of 0.08, 0.63, 1.78, and 2.19 cd/m^2. The brightest bars had a contrast [measured as (max − min)/(max + min) of luminance] of 87.2%, the gratings had a contrast of 65.4%, the Walsh patterns had a contrast of 96.4%, and the contrasts of the pictures ranged from 56 to 96%. For the experiment in TE, eight Walsh patterns (*E*) were used. The contrast was the same as for the Walsh patterns in the V1 experiments.

fixation point until the stimulus disappeared and was not required to react to the stimulus in any way. After a delay of 300 msec, the next trial began.

Receptive fields were mapped by hand using bar stimuli and were located 1.5–3° from the fovea in one monkey and 5–6° from the fovea in the other. Stimuli were always 3.5° on a side, and they covered the receptive field and part of the surround. The stimuli that were used (Fig. 1) included 32 oriented bars (*A*), 32 sine-wave gratings (*B*), 32 Walsh patterns (*C*), and 32 photographic images (*D*). Although it is still a small

subset of all possible stimuli, this is, to our knowledge, the most extensive set of stimuli used to examine the mean–variance relation in monkey primary visual cortex. For each neuron, each stimulus was presented on a video monitor in randomized order approximately the same number of times; the median number of presentations per stimulus ranged from 8 to 52 (median 14) in different neurons. No significant differences were found between the results from the two monkeys, so we present data from both together.

Spikes were counted in a 300 msec window starting at stimulus onset. We use this period because during normal primate vision, a new image appears on each receptive field one to three times per second because of saccadic eye movement, after which the image is kept nearly still on the retina (compared to saccade velocities).

TE. Responses were recorded from neurons in visual area TE while a monkey performed a sequential delayed match-to-sample task using eight Walsh patterns (Fig. 1*E*). The monkey touched a contact lever to start each trial; a fixation point appeared at the center of a screen immediately. The monkey was required to fixate within ±5° of this spot for the entire trial. As soon as the monkey fixated the fixation point, the fixation point was replaced by a sample Walsh pattern 8.5° on a side, followed by up to two nonmatching patterns and a repeat of the original (now matching) pattern. When the original stimulus reappeared, the monkey was required to release the bar within 2 sec to receive a reward. Stimuli were displayed for 500–1000 msec, with 300–800 msec between stimuli. Further experimental details can be found in Liu and Richmond (2000). Following Liu and Richmond (2000), we counted spikes from 70 to 470 msec after stimulus onset; the delay allows for response latency in area TE. The median number of presentations per stimulus ranged from 20 to 64 (median 48) in different neurons.

Eye position. Eye position was measured every 8 msec using a magnetic search coil. Trials were divided according to whether or not eye position remained within a square region 6 min of arc on a side (the smallest difference definitely detectable by our eye coil) during the entire trial. Average eye position and amount of eye movement during a trial were unrelated to which stimulus was presented, the group to which the presented stimulus belonged in V1, and behavioral condition in TE (ANOVA; $p > 0.05$).

All work was conducted in accordance with the National Institutes of Health animal care guidelines and approved by the National Institute of Mental Health Animal Care and Use Committee.

Regression analysis

For each neuron, each stimulus produces a sample mean spike count, μ_i, and a sample variance of spike count, σ_i^2, where the subscript i labels stimulus. We fit the line $\log \sigma^2 = b + m \log \mu$ to the set of points (μ_i, σ_i^2). Residuals were weighted by the estimated variance of the logarithm of the variance (which depends on the number of trials available for each stimulus; see below). The logarithmic transformation of means and variances makes the regression residuals more nearly uniform across the range of the mean responses (so that the data more closely conform to the assumptions underlying the regression analysis) and ensures that the model can never predict variances <0 (because the model is equivalent to $\sigma^2 = \mu^b \, e^a$). A model using Fano factors to relate the mean and variance also cannot predict variances <0 but does not make the regression residuals uniform across the range of mean responses and is substantially less compact than the regression model because a separate Fano factor is needed for each stimulus (the factors for different stimuli span an order of magnitude in our data from both V1 and TE). Estimates of log(mean) and log(variance) obtained by taking the logarithm of the sample mean and variance are biased and result in underestimation of the variance. We corrected for the bias using a Taylor series expansion (Kendall and Stuart, 1961); only a few terms are needed for good results.

Estimates of log(mean) and log(variance) of spike count from finite samples are uncertain. Standard regression methods assume that one quantity (the independent variable) is known without uncertainty. To check whether the uncertainty of the mean makes a difference when using real responses, we performed our analyses using both standard regression methods [with log(mean) of spike count as the independent variable] and regression methods designed for data with uncertainty in both variables (Fuller, 1987; Ripley and Thompson, 1987). One advantage of these techniques is that they treat the two variables symmetrically; the same line is obtained no matter which quantity is thought of as the dependent variable and which the independent variable. Both methods require estimates of the uncertainty with which the variance of spike count is known. The sample variance S^2 is distributed as $(\sigma^2/(n-1))\chi^2_{n-1}$,

where χ^2_{n-1} is a chi-squared distribution with $n-1$ degrees of freedom. This distribution has mean σ^2, variance $\sigma^4(2/(n-1))$, and SD $\sigma^2\sqrt{2/(n-1)}$, and for moderate values of n is approximately Gaussian. Thus, we can approximate symmetric points of the distribution by $E[S^2] \pm k\,SD[S^2]$, or $\sigma^2 \pm k\,\sigma^2\sqrt{2/(n-1)}$, where k varies depending on the percentile of the distribution desired (for example, for the 5th and 95th percentiles, $k \approx 1.645$). The logarithms of these points are:

$$\log\left[\sigma^2\left(1 \pm k\sqrt{\frac{2}{n-1}}\right)\right] = \log\sigma^2 + \log\left(1 \pm k\sqrt{\frac{2}{n-1}}\right)$$

$$\approx \log\sigma^2 \pm k\sqrt{\frac{2}{n-1}}\,.$$

Thus, the logarithm of the sample variance has mean approximately $\log\sigma^2$ and variance approximately $2/(n-1)$. Note that the variance of the logarithm of the variance does not depend on the variance (even though the variance of the variance does depend on the variance).

The methods designed to deal with uncertainty in both variables also require estimates of the uncertainty with which the mean of spike count is known. The sample mean of a normal distribution with true mean and variance μ and σ^2 is normally distributed with mean μ and variance σ^2/n; for non-normal distributions, this is an approximation. A calculation similar to the one above for the sample variance shows that the mean and variance of the distribution of the logarithm of the sample mean are approximately $\log\mu$ and $\sigma^2/n\mu^2$, respectively.

We ask whether a model using a single regression line for all stimuli predicts log(variance) of spike count from log(mean) of spike count less well than does a model using a different regression line for each stimulus set or behavioral condition. When ignoring uncertainty in the sample mean, this is simply comparing an analysis of covariance of log(variance) against log(mean) to an analysis of covariance of log(variance) against log(mean) conditioned on stimulus set. We performed the standard analysis of covariance both with and without weights on the basis of estimated variance; the results are nearly identical. Here, we present results calculated using the weights.

Information analysis

Information theory is a statistical approach that deals with the relation between inputs, or stimuli, and outputs, or responses (Shannon and Weaver, 1949; Cover and Thomas, 1991). The entropy of any signal X, $H(X) = -\Sigma_x p(x)\log_2 p(x)$, measured in bits, quantifies the uncertainty of the signal. The conditional entropy $H(R|S)$ measures the uncertainty in a response if the stimulus $s \in S$ is known. The mutual, or transmitted, information between a stimulus and a response, $I(R;S)$, is the reduction in uncertainty about which stimulus has been presented, provided by knowing the response, or vice versa: $I(R;S) = H(S) - H(S|R) = H(R) - H(R|S)$.

Estimating transmitted information requires estimating the conditional response probabilities $p(r|s)$ for each response r (here, the number of spikes elicited) and stimulus s. Reading these values from the response histogram for each stimulus tends to overestimate information. Instead, for V1 neurons, we estimate the conditional response probabilities by a truncated Gaussian distribution with mean calculated from the observed responses and variance predicted using the mean–variance relation (93% consistent at the $p = 0.05$ level; χ^2 test) (Gershon et al., 1998; Wiener and Richmond, 1998). To avoid inaccuracy in our estimates of the means and variances of the logarithms of the sample mean and sample variance, we did not use data sets with fewer than eight trials per stimulus. This method has been shown to give answers comparable to those obtained using a well validated neural network method (Heller et al., 1995; Golomb et al., 1997). Spike count distributions for the TE neurons are not well modeled by the truncated Gaussian distribution ($<50\%$ consistent at the $p = 0.05$ level; χ^2 test), so we omit this calculation.

Transmitted information measures the outcome of a particular experiment; changing the stimuli presented, or even the frequency with which the stimuli are presented, will almost certainly change the transmitted information. The channel capacity of a neuron, the maximum information the neuron can transmit using a particular code and given the reliability of its responses, does not change from experiment to experiment, but estimating it requires knowing the distribution of responses to all possible stimuli, not only to those stimuli presented. Using the relation between log(mean) and log(variance), the mean response to a stimulus determines the variance of responses to that stimulus. Because for V1 neurons the truncated Gaussian is a good model of the distributions of

spike counts elicited by stimuli (see Results), the mean and variance together determine the entire response distribution. Therefore, stimuli that elicit the same number of spikes on average are indistinguishable, and every stimulus can be labeled by the mean number of spikes it elicits. This provides a model of all possible response distributions. Given a range of possible mean responses (a neuron can fire only a finite number of action potentials in any counting window), channel capacity can be estimated using this model by maximizing transmitted information over stimulus presentation probabilities, as described in detail in Gershon et al. (1998).

Analysis of scatter around the regression line

Scatter around a regression line represents variability not explained by the regression. In our regression of log(variance) versus log(mean) of spike count, we know of at least one source of such variability: both means and variances are estimated from samples. The amount of scatter resulting from this measurement problem is determined by the number of trials available for estimating each mean and variance; as the number of trials decreases, the scatter around the regression line increases.

Assuming that the regression is valid, that is, that log(variance) is a linear function of log(mean), the mean residual sum of squares around the log(variance) versus log(mean) regression line is an estimate of the variance of log(variance) of the responses. Estimating means and variances using only a subset of the n points available will cause the sum of squared residuals to increase. Only neurons with a median of at least eight trials per stimulus in the subsampled data sets (so at least 16 trials per stimulus in the full sets) were included in the analysis of scatter around the regression line. We use simulated data to estimate how quickly the sum of squared residuals decreases with increasing numbers of trials per stimulus under the assumption that all of the scatter around the regression line is attributable to finite sample size. The artificial responses have the same number of trials per stimulus and are generated from distributions with the same mean spike count, as observed for each stimulus in the corresponding real neuron. However, in the artificial data, the variance of spike count for each stimulus is calculated from the observed mean of spike count using the regression line relating log(variance) and log(mean), and spike counts are generated by sampling from a truncated Gaussian distribution with the given mean and variance. Thus, in the artificial data, all scatter around the mean–variance regression line arises from sample size effects only.

If the residual sum of squares in the real data increases less quickly than expected based on the artificial data, then some of the scatter around the line is not caused by sampling. (We do not know or speculate here on the source of this nonsampling variance.) The nonsample variance c can be obtained by solving $k = (aRSS_{part} + c)/(aRSS_{full} + c)$, where $aRSS_{full}$ and $aRSS_{part}$ are the residual sums of squares from regressions from full and subsampled artificial data sets, respectively, and k is the ratio RSS_{part}/RSS_{full} measured from the actual data. The portion p of residual sums of squares attributable to sampling can then be calculated, and the total portion of variance explained is $r^2 + p(1 - r^2)$, where the first term is the usual r^2 from the regression and the second term represents the variability explained by sampling. Simulations show the results of this method to be unbiased.

The rate of change of the residual sum of squares, and therefore the percent of scatter due to sampling, can also be estimated using the formulas given above for uncertainty of the measured sample variance. Tests using simulated data (for which all scatter is attributable to sampling effects) show that an analysis based on the formulas overestimates (by a few percent) the percent of scatter attributable to sampling (we believe this overestimation is attributable to the fact that our data are truncated Gaussians rather than true Gaussians). Consistent with this, for the actual V1 data the estimated percent of scatter attributable to sampling is higher using the formulas than based on the simulations (see Results). Thus, we regard the estimate based on the formulas as an upper bound and the estimate based on simulations as our best guess. Because the TE data are not well modeled by the truncated Gaussian distribution, we omit this analysis for TE.

Principal component analysis

Principal component analysis (Ahmed and Rao, 1975) can be used to compress a data set. The first few principal components of spike train data reflect aspects of the temporal structure of the spike trains (Optican and Richmond, 1987; Richmond and Optican, 1987, 1990; Tovee et al., 1993; Heller et al., 1995; Tovee and Rolls, 1995). In Wiener and Richmond (1999), we showed that the first and second principal components
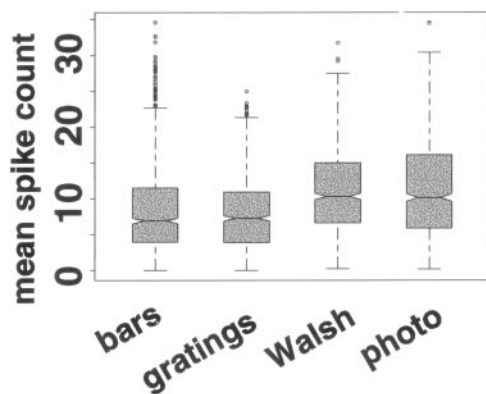
Figure 2. Distribution across 27 V1 neurons of mean responses elicited by bars, gratings, Walsh patterns, and photographic images. The *line* in the *middle* of each *shaded box* shows the median response. The *notch* shows a 95% confidence interval around the median (if two notches do not overlap vertically, the corresponding medians are different at the 5% level). The *bottom* and *top edges* of the *boxes* show the 25th and 75th percentiles, and the *extended whiskers* show 1.5 times the interquartile range. Mean responses outside a range 1.5 times the width of the inter-quartile range from the median are shown as separate *points*. The fifth percentile and median of mean responses elicited by bars and gratings are lower than the fifth percentile and median of mean responses elicited by Walsh patterns and photographic images, although the 95th percentile of mean responses is not distinguishable across the four stimulus sets. Each *shaded box* is based on 896 measurements: the mean response of each of 27 neurons to each of the 32 stimuli in each set. We have no explanation for the greater number of outlier points for bars as opposed to the other stimulus sets.

of neuronal responses obey a version of the mean–variance relation: the logarithm of the variance of each principal component is linearly related to the logarithm of the mean of the first principal component. The regression and analysis of scatter around the regression line can be performed in the same way as described above.

To find the principal components of these data, we low-pass filtered each spike train by convolution with a Gaussian distribution with SD of 5 msec and resampled at 1 msec resolution to create a spike density function. The principal components were calculated by performing singular value decomposition on the matrix of spike density functions.

RESULTS

Our data set included 27 complex cells from V1 (16 from one monkey, 11 from another) and 20 neurons from area TE. As explained in Materials and Methods, we performed the analyses using both standard regression methods and methods designed for situations in which both variables are measured with uncertainty. The results using the two regression methods were very similar. We present the results obtained using standard regression methods. At the end of this section, we compare results using the two regression methods.

Distribution of mean responses

It is well known that different stimuli elicit different numbers of spikes, but this does not require that stimuli from different sets consistently elicit different numbers of spikes. Across the 27 V1 neurons, the stimulus set significantly affected the median of mean spike counts elicited (Fig. 2) (Friedman test; $p < 0.05$); the same was true in 26 of the individual neurons (Kruskal–Wallis test; $p < 0.05$). Stimulus set accounts for 11% [median; interquartile range (iqr), 5–18%] of the variability in spike counts. The least effective photographic images and Walsh patterns used in these experiments elicited mean spike counts larger than those elicited by the least effective bar and grating stimuli; that is, Walsh patterns and photographic images were less likely than bars and
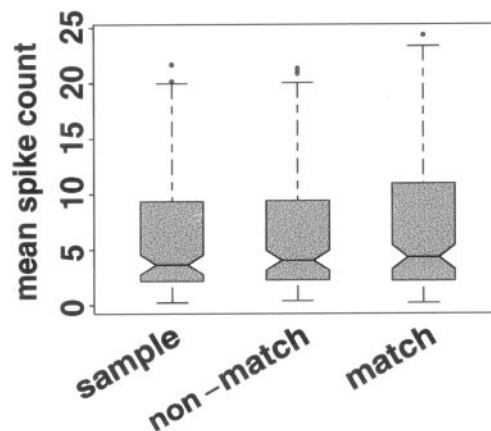

Figure 3. Distribution across 20 TE neurons of mean responses elicited by stimuli in the sample, nonmatch, and match phases of a delayed match-to-sample task. For interpretation of boxplots, see Figure 2. The distributions of responses are not distinguishable across the three behavioral conditions. Each *shaded box* is based on 160 responses: the mean response of each of 20 neurons to each of the eight stimuli when viewed in the indicated phase of the task.

gratings to elicit mean responses near zero. We did not explicitly search for optimal bar or grating stimuli for the neurons we recorded. Figure 2 shows that the median of mean spike counts was between 20 and 27 spikes per second. Reich et al. (2001), using optimal stationary gratings (the stimuli in their experiment most comparable to our stationary stimuli), elicited median firing rates of 23 spikes per second from V1 complex cells. The 75th percentile across neurons of mean spike counts elicited in our experiments is 20 spikes in a 300 msec period, or about 66 spikes per second; the 95th percentile was about 43 spikes per second. Reich et al. (2001) report that the 75th percentile across neurons of mean firing rate was between 40 and 45 spikes per second using an optimal stationary grating, and the 95th percentile was 80 spikes per second (their Fig. 3*D*). Thus the distribution of mean responses that we observed was similar to those obtained when an explicit effort to identify the optimal stimulus was made.

Behavioral context did not significantly affect the median of mean spike counts across TE neurons (Fig. 3) (Friedman test; $p > 0.05$), or in any individual TE neuron (Kruskal–Wallis test; $p > 0.05$). However, behavioral context did affect the largest mean responses elicited by stimuli; the largest mean spike counts in the match condition were larger than the largest responses in the sample and nonmatch conditions (Friedman test on 95th percentile of mean responses; $p < 0.05$).

Consistency of the mean–variance relation across stimulus sets and behavioral conditions

To determine whether a single regression line adequately described responses under different conditions, we examined two models for each set of responses. One model used a single regression line to predict log(variance) of spike count from log(mean) of spike count for all stimuli presented to a particular neuron. The other model used a different regression line to predict log(variance) of spike count from log(mean) of spike count for each stimulus set in V1 or each behavioral condition in TE. If different conditions do give rise to different relations between log(mean) and log(variance), the model using several regressions should predict log(variance) significantly better than the model with a single regression. We test this by comparing the variance of the residuals from the two models. The variance of
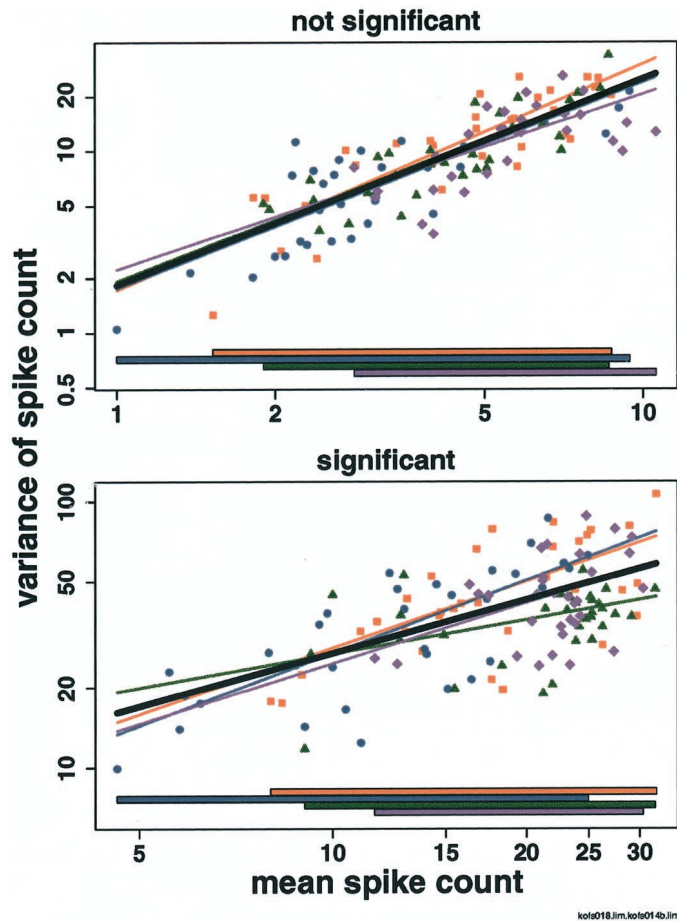
Figure 4. Relation between log(mean) and log(variance) in two V1 neurons: one in which the model using four regression lines is not significantly better than the model using a single regression line (*top*), and one in which it is better (*bottom*). *x-*, *y-axes*, Mean and variance of number of spikes elicited by each stimulus on a logarithmic scale. Note the different scales on the *x-* and *y-*axes in the two panels. Each panel shows log(mean) versus log(variance) for each of four stimulus sets in a single V1 neuron. Oriented bars, *Red squares*; oriented gratings, *blue circles*; Walsh patterns, *green triangles*; photographic images, *purple diamonds*. Stimulus-set specific regression lines are in the corresponding colors, extended to the edges of the plot for visibility. The regression line for all stimuli taken together is shown in *black*. *Colored bars* at the *bottom* show the range of means for each stimulus set. The neurons shown have the median *p* values for the *f* test comparing the model using a single regression line to the model using four regression lines among those neurons for which the *f* test is (*bottom*) and is not (*top*) significant.

the residuals is the residual sum of squares divided by the residual degrees of freedom. A model using several lines has fewer residual degrees of freedom than a model using a single line, so the residual sum of squares must decrease more quickly than the residual degrees of freedom to justify using the additional parameters.

In V1, we asked whether a model using four regressions, one for each stimulus set, predicted log(variance) of spike count from log(mean) of spike count significantly better than a model using a single regression for all four data sets. In 17 of 27 neurons, the two models were statistically indistinguishable (Fig. 4, *top*), but in 10 of 27 neurons, the reduction in sum of squared errors did justify using the extra parameters (*f* test; $p < 0.05$) (Fig. 4, *bottom*). Even in the neurons in which the change was significant, however, the increase in percentage of variance explained was
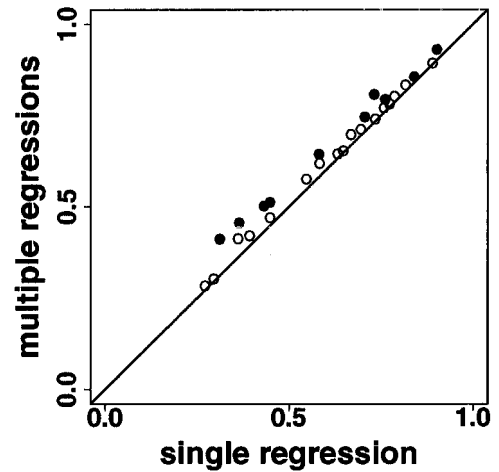

Figure 5. Portion of variance explained (r^2) increases for a model using a separate log(variance) versus log(mean) regression for each stimulus set as compared to a model using a single log(variance) versus log(mean) regression for all stimuli, for 27 V1 neurons. The *horizontal* and *vertical axes* show the r^2 values for the single-regression and multiple-regression models. *Filled circles* represent neurons for which the improvement in prediction of log(variance) from log(mean) is significant (*f* test, $p < 0.05$), and *open circles* represent those for which it is not.

small (Fig. 5, Table 1). Across all stimuli from all V1 neurons, the model using a single regression explained about two-thirds of the variance (median r^2, 0.65; iqr, 0.44–0.76), and the model using four regressions explained only slightly more (median r^2, 0.65; iqr, 0.49–0.79). Across all neurons, the median increase in r^2 was 0.03 (iqr, 0.01–0.05); for only those neurons in which the four-regression model predicted variance significantly better than the single-regression model, the median increase in r^2 was 0.06 (iqr, 0.04–0.07). We will show below that these small changes in predicted variance do not affect the ability of the neuron to distinguish among different stimuli. Therefore, for each neuron only a single regression line is needed to describe the relation between log(mean) and log(variance) of spike count for stimuli of all four kinds. Across 27 V1 neurons, the median intercept of the single regression line is 0.6 (iqr, 0.4–0.8), and the median slope is 1.1 (iqr, 1.0–1.2).

Although in 10 of 27 neurons, using four regression lines predicts variance significantly better than using a single regression line, naturalistic (photographic) stimuli are not consistently treated differently from the geometric stimuli. Examining models using two regression lines, one for stimuli from one set and another for stimuli from the other three sets, we found that bars, gratings, Walsh patterns, and photographic images were distinguishable from all other stimuli in 4, 7, 8, and 5 of the 10 neurons, respectively. There was no clear pattern to which stimulus sets were distinguishable from others in individual neurons.

In area TE, we asked whether a model using three regression lines, one each for the sample, nonmatch, and match task conditions, predicted log(variance) from log(mean) significantly better than a model using a single regression line for all task conditions together (Fig. 6). In 19 of the 20 neurons, the improvement in prediction did not justify using the extra parameters (Table 2), and 1 of 20 neurons is expected to show an effect at the $p = 0.05$ level by chance. Thus, we conclude that for neurons in TE, only a single regression line is needed to describe the responses in all three behavioral contexts. Across 20 TE neurons, the median

Table 1. Summary of r^2 for regression of log(variance) versus log(mean) for V1 neurons

	Minimum	First quartile	Median	Third quartile	Maximum
r^2, single regression, 0–300 msec	0.27	0.44	0.65	0.76	0.90
r^2, four regressions, 0–300 msec	0.28	0.49	0.65	0.79	0.93
Change (all)	0.00	0.01	0.03	0.05	0.10
Change ($p < 0.05$)	0.02	0.04	0.06	0.07	0.10
r^2, single regression, 0–150 msec	0.22	0.44	0.67	0.76	0.91
r^2, single regression, 151–300 msec	0.34	0.52	0.66	0.81	0.92

Columns show the minimum, first quartile, median, third quartile, and maximum values across 27 V1 neurons. First row, Distribution of r^2 for the model using a single regression for all four stimulus sets, with counting window 0–300 msec after stimulus onset. Second row, Distribution of r^2 for the model using a separate regression for each stimulus set, with counting window 0–300 msec after stimulus onset. Third row, Distribution of increase in r^2 when going from the model using a single regression to the model using four regressions. Fourth row, Same as third row, but only for the 10 of 27 neurons for which the four-regression model predicts log(variance) from log(mean) significantly better than the single-regression model. Fifth and sixth rows, Same as first row, but in two disjoint pieces of the original counting window: 0–150 msec after stimulus onset (fifth row) and 151–300 msec after stimulus onset (sixth row).

intercept of the regression line is 0.3 (iqr, 0.0–0.6), and the median slope is 1.3 (iqr, 1.2–1.5).

It has been reported that instability of eye position during presentation of an optimal moving bar increases response variability in V1 neurons (Gur et al., 1997). If this happened for all stimuli generally, it might affect the mean–variance relation. For the 16 V1 neurons with 13 or more trials per stimulus, we sorted the data for each stimulus into trials during which eye position was very stable and those during which it was less stable (see Materials and Methods). Each stimulus was represented by two points: one from those trials during which eye position was more stable, and one from those trials during which eye position was less stable. There was no systematic increase in variance when eye position was less stable (variance increased in 51% of stimuli and decreased in 49%). In each of the 16 neurons, we calculated log(variance) versus log(mean) regressions for all trials taken together and for the two subsets individually. Only stimuli with five or more trials in each subset were included in this analysis (median number of stimuli included 108 of 128; iqr, 72–126). As expected given that no systematic change in variance was observed with eye movement, the sum of squared errors was indistinguishable whether a single regression line was used for trials in both subsets or a separate regression line was used for each subset ($p > 0.05$; f test). Thus, only a single regression line is needed to describe the mean–variance relation in both subsets (Fig. 7). This is consistent with the findings of Bair and O'Keefe (1998) in area MT. In our data, the trials with more stable fixation show greater scatter around the regression line than the trials with less stable fixation because there were fewer such trials: 30% of trials in each neuron were contained in a region 6 min of arc on a side (median; iqr, 26–36%). (Scatter around the regression line is discussed later in Results.) For completeness, we separated trials with more and less stable fixation for the 15 of 20 TE neurons for which there were sufficient numbers of trials per stimulus in both conditions; as in V1, only a single regression was needed to describe the mean–variance relation in both subsets.

Transmitted information and channel capacity

Transmitted information measures how well an observer can guess which stimulus elicited any particular observed response (here, spike count). For each V1 neuron, transmitted information was calculated for all stimuli together and for the four stimulus sets individually. The information for a particular stimulus set was calculated using the relation between mean and variance measured from stimuli from that set only; the information for all

stimuli together was estimated twice, once using the model with a single regression for all stimuli and once using the model with a separate regression for each of the four stimulus sets. The information for all stimuli together was nearly identical no matter which model was used (difference of 0.6% median; iqr, 0.1–1.2%), even for the 10 of 27 neurons for which the model with four regressions predicted variance significantly better than the model with a single regression (difference of 0.9% median; iqr, 0.2%–1.1%). However, the information that was transmitted about the individual stimulus sets varied a great deal (Fig. 8) (Friedman test; $p < 0.001$). The differences in information are attributable not to differences in the relation between mean and variance for the stimulus sets (which make very little difference in information when all stimuli are considered), but rather to the different mean responses elicited by stimuli of different kinds (Fig. 2).

The least effective Walsh patterns and photographic images do not elicit mean responses as small as those elicited by the least effective bar and grating stimuli, whereas the most effective stim-
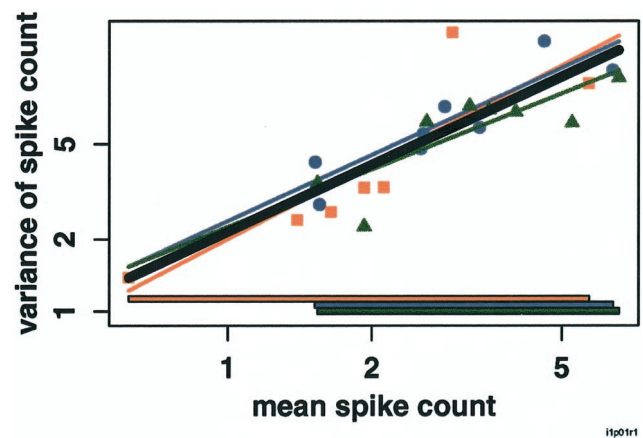

Figure 6. Responses obey a single mean–variance relation across behavioral conditions in a TE neuron. *x-, y-axes,* Mean and variance of number of spikes elicited by each stimulus on a logarithmic scale. Log(mean) versus log(variance) for each of three task conditions in a single TE neuron. Sample, *Red squares;* nonmatch, *blue circles;* match, *green triangles.* Regression lines for individual conditions are shown in the corresponding colors, extended to the edges of the plot for visibility. The regression line for all conditions together is shown in *black. Colored bars* at the *bottom* show the range of means for each task condition (sample, nonmatch, and match). The neuron shown has the median p value for the f test comparing the model using a single regression line to the model using three regression lines.

Table 2. Summary of r^2 for regression of log(variance) versus log(mean) for TE neurons

	Minimum	First quartile	Median	Third quartile	Maximum
r^2, single regression, 70–470 msec	0.51	0.70	0.85	0.90	0.97
r^2, three regressions, 70–470 msec	0.55	0.77	0.86	0.92	0.98
Change (all)	0.001	0.01	0.02	0.04	0.11
r^2, single regression, 70–270 msec	0.55	0.74	0.78	0.91	0.96
r^2, single regression, 271–470 msec	0.18	0.74	0.87	0.92	0.97

Columns show the minimum, first quartile, median, third quartile, and maximum values across 20 TE neurons. First row, Distribution of r^2 for the model using a single regression for all three behavioral conditions, with counting window 70–470 msec after stimulus onset. Second row, Distribution of r^2 for the model using a separate regression for each behavioral condition, with counting window 70–470 msec after stimulus onset. Third row, Distribution of increase in r^2 when going from the model using a single regression to the model using three regressions. Fourth and fifth rows, Same as first row, but in two disjoint pieces of the original counting window: 70–270 msec after stimulus onset (fourth row) and 271–470 msec after stimulus onset (fifth row).
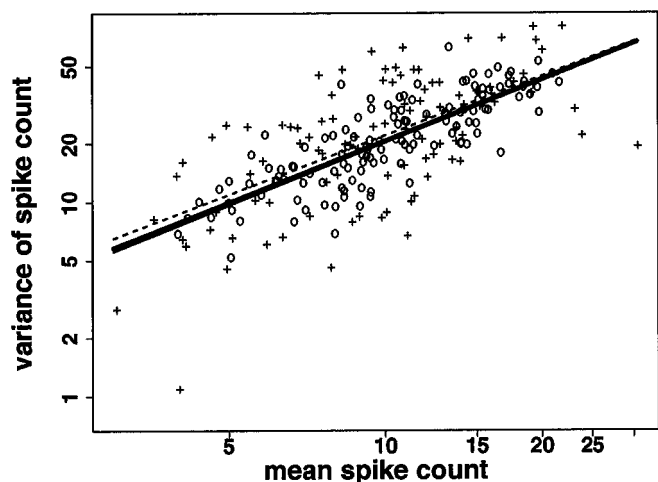

Figure 7. Responses obey a single mean–variance relation for more and less stable fixation. *x*-, *y-axes*, Mean and variance of number of spikes elicited by each stimulus on a logarithmic scale. Log(mean) versus log-(variance) for trials during which eye position remained within a square 6 min of arc on a side during the entire trial (+), and for trials during which eye position ranged over a larger region (○) in a single V1 neuron. Regression lines for the separate conditions are shown, using a *dashed line* for the trials with more stable fixation, and a *solid line* for the trials with less stable fixation. The *solid line* for trials with more stable fixation merges with the regression line for both conditions taken together, which is shown with a *thicker solid line*. The greater scatter around the line for trials with more stable fixation (+) than for trials with less stable fixation (○) is explained by the fact that there were fewer trials with more stable fixation: 30% of trials in each neuron were contained in a region 6 min of arc on a side (median; iqr, 26–36%). The neuron shown has the median *p* value for the *f* test comparing the model using a single regression line to the model using two different regression lines.

uli from each group elicit similar responses. This means that mean responses to Walsh patterns and photographic images are, in effect, crowded into a smaller range than mean responses to bars and stimuli. Because response variance grows with response mean, responses to Walsh patterns and photographic images are on average also more variable. Thus, individual responses to photographic images and Walsh patterns are less informative about which stimulus was presented than individual responses to bars and gratings.

Transmitted information describes the outcome of a particular experiment. Channel capacity, which depends on the mean–variance relation and the range of possible mean responses (Gershon et al., 1998; Wiener and Richmond, 1998), is a more robust
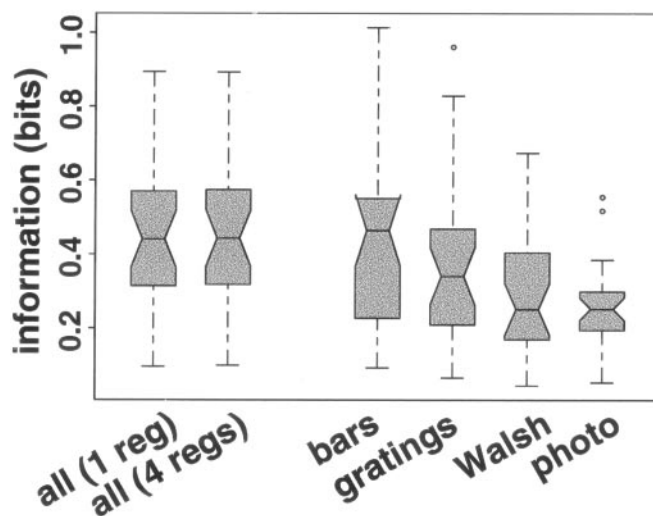

Figure 8. Information transmitted by spike count about which of the stimuli was presented, for all stimuli together and for each of the stimulus sets separately, for neurons in V1. The *vertical axis* shows transmitted information measured in bits. Each boxplot shows the distribution across 27 V1 neurons of information for the set of stimuli indicated on the *x* axis: all stimuli together (using either a single regression for all four stimulus sets or using a separate regression for each of the four stimulus sets), bar stimuli alone, gratings alone, Walsh patterns alone, and photographic images (*photo*) alone. For interpretation of boxplots, see Figure 2. Responses become less informative as the range of distinct means becomes smaller and the distributions of responses become more variable.

measure of the information-processing capability of a neuron. For each V1 neuron, we calculated channel capacity on the basis of the mean–variance relation and dynamic range estimated from all stimuli together, and on the basis of the mean–variance relation and dynamic range estimated for each stimulus set separately. Because a single regression describes the mean–variance relation for all four stimulus sets, channel capacity depends mostly on the estimate of the range of possible mean responses. Here, we assume that the minimum possible mean response is zero (which in some cases requires extrapolating the mean–variance relations beyond the range of observed mean responses) and the maximum possible mean response is 25% larger than the largest observed mean response (Fig. 9, *first box* in each set). We have shown previously (Wiener and Richmond, 1998) that estimates of channel capacity change relatively slowly with changes in the maximum mean response considered; here, if we assume the maximum possible mean is only 10% larger than the largest observed mean
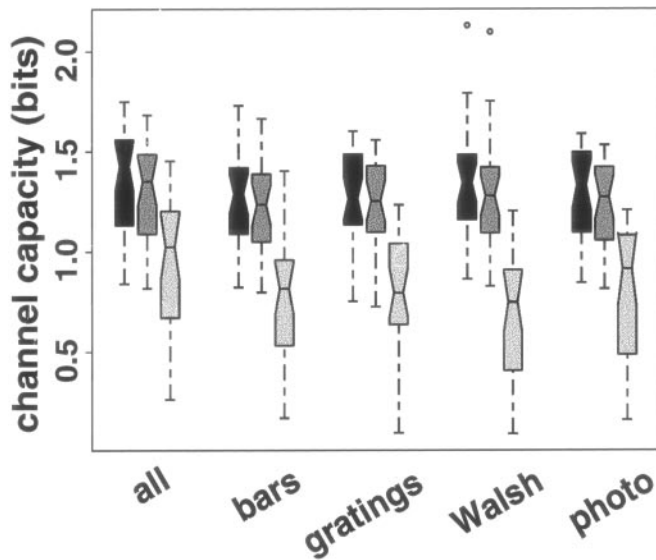
Figure 9. Channel capacity depends on the lowest allowed mean response more than on stimulus set or highest allowed mean response in neurons in V1. The *left* (*darkest*) *box* in each set shows the distribution across 27 V1 neurons of channel capacity if the range of allowable mean responses extends from 0 (no spikes ever elicited) to 1.25 times the maximum observed mean response. The *middle box* shows channel capacity if the range extends from 0 to 1.1 times the maximum observed mean response. The *right* (*lightest*) *box* shows channel capacity if the range extends from the minimum observed mean response to 1.25 times the maximum observed mean response. Changing the minimum allowed mean response has a much greater effect than changing the maximum allowed mean response. For interpretation of boxplots, see Figure 2.

response, estimates of channel capacity drop by only 3.8% (median; iqr, 3.4–4.5%) (Fig. 9, *second box* in each set). This insensitivity to the maximum mean is attributable to the fact that responses with higher means are more variable than responses with lower means, so allowing larger means yields diminishing returns. The estimates for the different groups are indistinguishable no matter which upper bound is used (Friedman test; $p >$ 0.05). Correspondingly, because responses with smaller means are less variable, channel capacity is much more sensitive to the smallest mean response allowed. If we assume that the minimum achievable mean response is equal to the minimum observed mean response (Fig. 9, *third box* in each set), the different estimates of channel capacity are lower than the previous estimates by 0.50 bits (median; iqr, 0.43–0.62; paired t test, $p \ll 0.01$) and are no longer statistically indistinguishable from one another (Friedman test; $p < 0.05$), the estimates using only the Walsh patterns having dropped more than the other estimates.

These results show the importance of the assumed minimum achievable mean response for estimates of channel capacity. Even when the same relation between mean and variance is used, a change in the assumed minimum achievable mean can change the estimate of channel capacity dramatically (Fig. 9, comparison between the *two left columns* and the *right column* in each set). The smaller the smallest observed mean response, the less dramatic the effect will be. Therefore, it is important in experiments seeking to examine the information-processing capabilities of a neuron to use a range of stimuli that elicit the largest possible range of mean responses from a stimulus and, in particular, to include stimuli that elicit few spikes as well as those that elicit many.

We cannot use these methods to calculate information or channel capacity for the TE neurons, because we do not have a
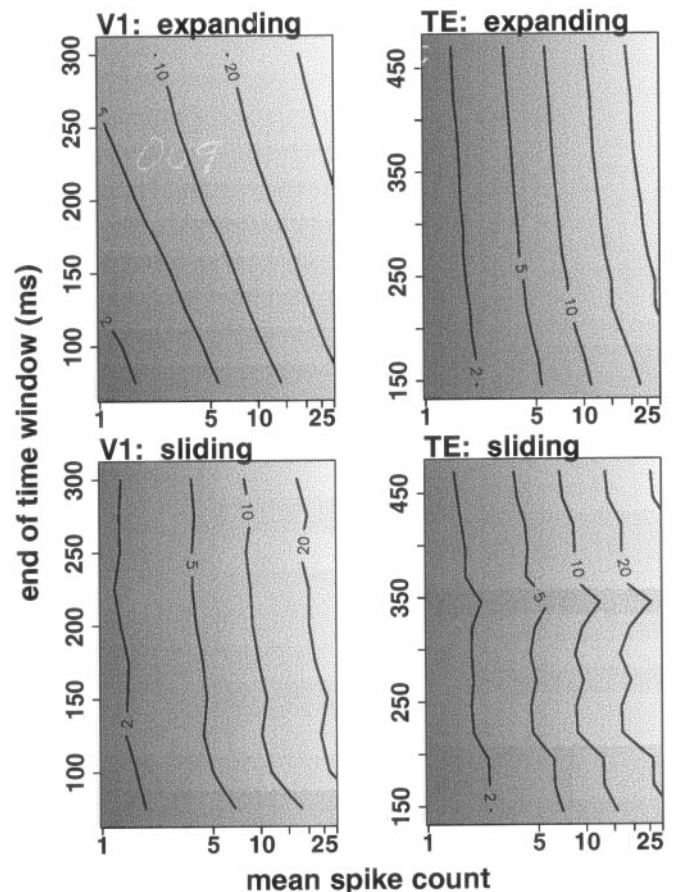

Figure 10. Consistency of the mean–variance relation in different counting windows. The *left* and *right panels* show results for neurons from V1 and TE, respectively. The *top panels* show results for windows expanding from a fixed start time (stimulus onset in V1, 70 msec after stimulus onset in TE), and the *bottom panels* show results for sliding 50-msec-wide windows starting at different times. In each panel, the *horizontal axis* shows mean spike count (on a logarithmic scale), the *vertical axis* shows the end of the counting window at 25 msec intervals (windows start at stimulus onset in V1, and 70 msec after stimulus onset in TE), and *gray scale* represents the variance (lighter means higher variance). Contours representing variances of 2, 5, 10, 20, and 40 spikes per second squared are also shown. The mean–variance relation at a particular time corresponds to a horizontal slice through the plot. *V1*, The variance associated with a particular mean spike count is larger in longer counting windows (with fixed starting point). This increasing variance is also seen in sliding windows. *TE*, The variance associated with a particular mean spike count is larger in longer counting windows, but variance increases more slowly than in V1. Again, the increase is seen in sliding windows as well.

good model for the spike count distributions from the TE neurons. However, the fact that both the range of mean responses and the relation between mean and variance are identical across the three behavioral conditions suggests that the information content of responses in the three conditions will be similar.

The mean–variance relation in different counting windows

Our focus in this paper is whether the mean–variance relation for spike counts in a particular counting window is consistent across different stimulus sets in V1 and across different behavioral conditions in TE. Above, we have come to the conclusion that for a particular counting window (0–300 msec after stimulus onset in V1, 70–470 msec after stimulus onset in TE), any differences in the mean–variance relation are sufficiently small as to not affect

information transmission. Figure 10 shows the mean–variance relation over time in representative neurons from V1 (*left*) and TE (*right*). The horizontal axis shows mean response, the vertical axis shows the end of the counting window, and gray scale and contours show the variance predicted for each mean response by the mean–variance relation. In both V1 and TE, the variance associated with a particular mean spike count increases as the window expands. Variance increases more rapidly in V1 neurons than in TE neurons.

Although the relation between log(mean) and log(variance) changes over time (Fig. 10), the explanatory power of the relation does not change much; the r^2 values for the regressions are similar for the full period analyzed and for the two half-periods (Tables 1, 2). In expanding windows, the number of neurons for which using multiple regressions is justified [that is, for which the model using multiple regressions predicts log(variance) from log(mean) significantly (f test; $p < 0.05$) better than the model using a single regression] is similar to that found in the full window: 9–11 of 27 neurons in V1, and 0, 1, or 2 of 20 neurons in TE. In sliding windows, up to 14 (in V1) or 4 (in TE) neurons show differences among the mean–variance relations. However, in both expanding and sliding windows in the V1 neurons, the small differences found among the relations between mean and variance of spike count in different windows did not affect information transmission; the amount of information found in the responses was nearly identical, whether a single regression or multiple regressions were used to estimate variance from mean. (As for the main counting window, we cannot explicitly calculate information for TE neurons using our model because the spike count distributions are not well-modeled by a truncated Gaussian distribution.)

Analysis of scatter around the regression line

Although a single regression line can be used to predict log(variance) from log(mean) across stimulus sets in each V1 neuron, the prediction is not perfect; substantial scatter around the regression lines remains (Fig. 4). In the V1 neurons examined here, the regression explained 65% of the variability of measured variance (median r^2; iqr, 0.44–0.76). Thus 35% (median; iqr, 24–56%) of the variability in the V1 neurons is seen as scatter around the regression lines. As explained below, we estimate that about two-thirds of this scatter can be attributed to sample size effects.

The amount of scatter around the regression line relating log(mean) and log(variance) depends in part on the number of trials per stimulus that are used to estimate the means and variances; the more trials per stimulus, the less scatter. As explained in Materials and Methods, we estimated the measurement effect of sample size (number of trials per stimulus) on residual sum of squares around the regression line using artificial data in which log(mean) and log(variance) are exactly linearly related. In such artificial data, any change in residual sum of squares can be attributed only to the measurement effect of sample size. If the residual sums of squares do not increase as rapidly in regressions subsampling the real data, we can conclude that some of the scatter around the regression line has other sources. In the V1 neurons, 70% (median; iqr, 58–78%) of the residual sums of squares remaining after regression can be attributed to the measurement effects of sampling. (Using the formulas rather than simulation, which gives an upper bound, the median percent of scatter attributable to sampling is 75%, with iqr 63–89%.) Thus, the regression relating log(mean) and log(variance) is not only consistent across stimulus sets in V1 neurons; it is actually better than it looks, because about two-thirds of the
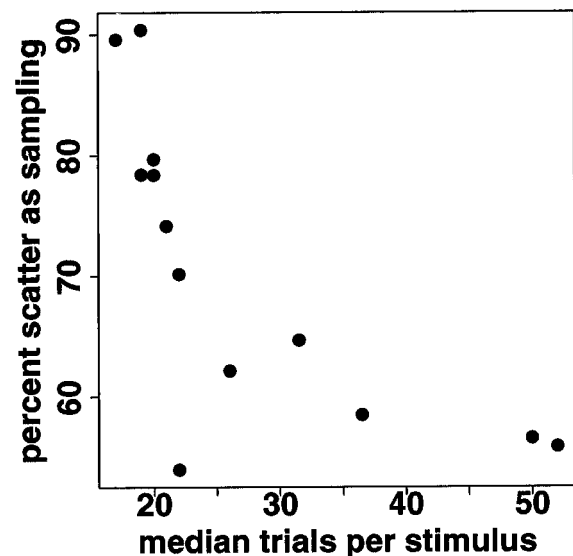

Figure 11. The amount of scatter around the regression line that can be attributed to the measurement effect of sample size decreases with increasing numbers of trials per stimulus. Each *point* shows the result for one V1 neuron. x-axis, Median number of trials per stimulus; y-axis, percent of scatter around the regression line attributable to sampling.

scatter around the regression line is attributable to limited sample size. As expected, the more trials available for estimating the variance, the smaller the percent of scatter attributable to the measurement effect of sample size (Fig. 11). When both the predictive power of the mean response and the measurement effect of sample size are taken into account, only 13% (median; iqr, 7–15%) of response variance in V1 neurons remains to be explained by other factors.

We have shown that in some of the V1 neurons a model using a separate regression line for each stimulus set predicts log(variance) from log(mean) significantly better than a model using a single line for all four stimulus sets, although the improvement is small. In the neurons for which the lines differed most, slightly more scatter could be attributed to sampling when residuals were calculated as deviation from the four lines individually rather than from a single line. In this group-by-group analysis, the percent of scatter attributable to sampling was 73% (median; iqr, 70–80%), which, when combined with the predictive power of the mean response, left only 7% (median; iqr, 5–12%) of the scatter to be explained by other factors. In five V1 neurons, there were sufficient trials to analyze scatter around the regression line separately for trials during which fixation was very stable and trials during which fixation was less stable (see Materials and Methods). Analyzing scatter separately for these two subsets resulted in very little additional variance explained.

Because spike count has been shown to influence spike timing (Oram et al., 1999; Wiener and Richmond, 1999), it is natural to wonder whether these results about the scatter around the line relating log(mean) and log(variance) of spike count carry over in some way to results for timing. Wiener and Richmond (1999) showed that the logarithms of variances of principal components of neural responses are related to the logarithm of the mean of the first principal component. The first principal component is highly correlated with spike count, so we do not examine it further here. The second principal component indicates whether the spikes in a response tend to come early or late in the response (Optican and Richmond, 1987; Wiener and Richmond, 1998). As
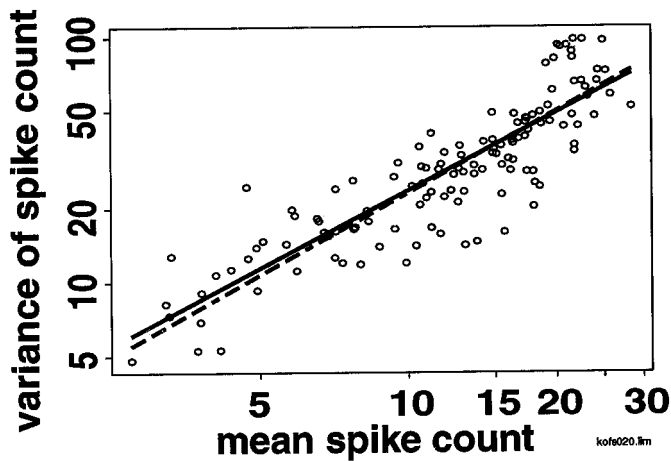
Figure 12. Regression lines calculated with and without uncertainty in the mean are similar. The *x*- and *y-axes* show, on a logarithmic scale, the means and variances of spike counts elicited from a single neuron by different stimuli. The regression *lines* calculated using standard regression methods (*solid*) and methods taking into account uncertainty in estimates of both variables (*dashed*) are shown. Ignoring the uncertainty in estimates of the sample mean biases the slope toward 0, but the effect is small. The cell shown has the median change in slope among V1 neurons.

in Wiener and Richmond (1999), the r^2 values for the regression of log(variance) of the second principal component against log-(mean) of the first principal component are lower than for the regression of log(variance) versus log(mean) of spike count: $r^2 = 0.12$ (median; iqr, 0.03–0.21) across the 27 V1 neurons. Scatter around the regression line depends chiefly on the number of trials from which each mean and variance is estimated (see Materials and Methods). Therefore, we expect that the scatter around the regression line relating log(variance) of the second principal component to log(mean) of the first principal component should be similar to the scatter around the regression line relating log(variance) and log(mean) of spike count. Across the 27 V1 neurons, 64% (median; iqr, 52–75%) of the scatter around the regression line relating log(variance) of the second principal component to log(mean) of the first principal component is attributable to sampling effects, leaving 32% of the variability to be explained by other factors (median; iqr, 26–42%). This means that most of the variability in a low-frequency measure of response timing is related to average spike count, just as is the variability of spike count itself.

Different regression methods give similar results

When regression methods taking into account uncertainty in both variables are used, the sums of squared residuals in the *x* direction (around the logarithms of the means) are much smaller than the sums of squared residuals in the *y* direction (around the logarithms of the variances), by a factor of 33 (median; iqr, 14–110) in the V1 neurons and by a factor of 13 (median; iqr, 7–24) in the TE neurons. This suggests that taking into account uncertainty in the mean should have a relatively small effect and differences between results using the two regression methods will be small.

To assess the practical effect of the uncertainty of estimates of mean response, we used both standard regression methods and methods designed for data with uncertainty in both variables (Fuller, 1987; Ripley and Thompson, 1987). Estimates of the slope using these methods are larger than those predicted using standard regression methods, by 10% (median; iqr, 7–17%) in the V1 neurons and by 6% (median; iqr, 3–12%) in the TE neurons.

However, the intercepts also change, and the combined effect in the range of data available is quite small (median difference of predicted variance −0.2, iqr, −0.9 to 2.2), with the nonstandard regression tending to estimate lower variances than the standard regression for low mean spike count (Fig. 12).

The main result of this paper, that the relation between log-(mean) and log(variance) is consistent across multiple stimulus sets in V1 neurons and across behavioral conditions in TE neurons, still holds when regression methods accounting for uncertainty in both variables are used. Furthermore, the amount of scatter around the regression line that can be attributed to the measurement effect of sample size is quantitatively similar whether the regression methods account for the uncertainty in estimates of the logarithm of the mean or ignore it. In these V1 data, the measurement effect of sample size accounts for 70% (median; iqr, 58–78%) of the sum of squared residuals around the regression line when standard regression methods are used (as reported above), and 67% (median; iqr, 61–80%) when uncertainty in the sample mean is taken into account.

DISCUSSION

We have examined spike-count coding in single neurons in monkey primary visual cortex. We find that the previously observed linear relation between log(mean) and log(variance) is sufficiently consistent across a wide range of stationary black-and-white images (including photographic images); for practical purposes, there is no reason to use more than a single relation. In particular, the relation between mean and variance is not systematically different for photographic images than for simple geometric stimuli. In area TE, the relation between mean and variance of spike count and the distributions of mean responses to stimuli presented in the sample, nonmatch, and match phases of a delayed match-to-sample task are statistically indistinguishable (Figs. 3, 6), consistent with the results of McAdams and Maunsell (1999). The variance associated with a given mean increases with the length of the counting window (Fig. 10). We do not know now the reasons for this change in the mean–variance relation over time, but correlations between firing rates at different times can cause such an effect.

The relation between mean and variance is not, however, the only factor affecting the ability of a neuron to transmit information. In our experiments in V1 neurons, photographic images and Walsh patterns elicited larger mean responses than bar and grating stimuli (Fig. 2). As a consequence, estimates of transmitted information depend on which stimuli are presented (Fig. 8). If four different researchers had conducted four different experiments using our four stimulus sets, there might be controversy over how much information neurons in V1 "really" transmit. Estimates of channel capacity based on results from the different stimulus sets are more consistent than estimates of transmitted information (Fig. 9), because they depend only on two fundamental statistical properties of the responses: the relation between mean and variance and the range of allowed mean responses. To characterize the information processing capacity of a neuron, it is important to elicit the largest possible range of mean responses, particularly low mean responses. Thus, it is important to use as large and varied a stimulus set as possible in neurophysiological experiments.

We have also shown for V1 neurons that the relation between mean and variance is better than it looks; approximately two-thirds of the scatter around the log(variance) versus log(mean) regression line is attributable to the measurement effect of sample

size. Although we cannot perform such a quantitative analysis for the TE neurons (because the spike count distributions in TE are not well modeled by truncated Gaussian distributions, as are those in V1), the variances of spike count distributions from TE neurons are measured with uncertainty, and this must contribute to the scatter around the regression lines in TE.

Comparison with other studies

Previous studies of whether naturalistic stimuli are encoded with less variability than other stimuli reached contradictory conclusions. Rieke et al. (1995) reported that frogsong-like noise elicited less variable responses than pure noise from frog auditory neurons. In fly H1 visual neurons, de Ruyter van Steveninck et al. (1997) reported that naturally moving stimuli elicited less variable responses than constant stimuli. However, when Warzecha and Egelhaaf (1999) studied the H1 neuron, they reported that the variability was the same across the two conditions. Our results in monkey visual cortex agree more closely with the results in fly H1 visual neuron of Warzecha and Egelhaaf (1999) than with those of de Ruyter van Steveninck et al. (1997).

In another study, Warzecha et al. (2000) reported that in fly H1 neuron, variance of spike count depends very little on mean spike count except for very low values of the mean. In contrast, we find a strong relation between mean and variance of spike count across the entire range of observed means for neurons from monkey V1 and TE. Warzecha et al. (2000) exclude onset transients from their analysis; this may contribute to the difference between the two sets of results.

Earlier, we presented an analysis of the mean–variance relation in TE neurons from a monkey performing a delayed match-to-sample task similar to the one in this paper (Gershon et al., 1998). The slopes reported for 14 of 19 of those neurons were <1, in contrast to the results presented here, where the slopes for 19 of 20 neurons are >1. The earlier experiment restricted eye movement more than the TE experiment described here (gaze was required to remain within 1° of the fixation point, as opposed to 5° here), but restricting the current data to trials in which gaze remained within 1° of the fixation point did not significantly change our results. We are not certain why the slopes differ between the two data sets. Similarly, whereas in Gershon et al. (1998) we found that the truncated Gaussian model provided a sufficiently good fit for some purposes to spike count distributions from TE neurons, in the data here <50% of the spike count distributions are consistent with a truncated Gaussian model (χ^2 test; $p < 0.05$). We are not certain why the current data are not well modeled by a truncated Gaussian.

In the experiments described here, we used only stationary black-and-white stimuli. Stimuli involving color or motion might give rise to a different mean–variance relation, as suggested by Croner and Albright (1999). In addition, we chose the stimulus presentation length in V1 (300 msec) to approximate the time between saccades during free viewing. However, our paradigm does not duplicate the correlations in time-varying images (Dong and Atick, 1995). Thus, the relation between mean and variance of spike count when images are brought onto receptive fields by eye movements might be different from the relation observed here.

Implications for neural coding

We have shown that assuming the existence of stimuli that elicit on average very small numbers of spikes, or elicit no spikes at all, results in significantly larger estimates of channel capacity than does assuming that the experiment has revealed the smallest achievable mean response. The large effect of very small responses on channel capacity suggests a link to theories of sparse coding in which few neurons should respond to any particular stimulus (Rolls and Tovee, 1995; Olshausen and Field, 1996a; Vinje and Gallant, 2000).

Stimulus features can be encoded not only by spike count but also by spike timing, although the nature and time scale of that encoding remain the subject of debate (Heller et al., 1995; Victor and Purpura, 1996; Buracas et al., 1998; Sugase et al., 1999; Reinagel and Reid, 2000). We have previously shown (Wiener and Richmond, 1999) that high-frequency components of timing (principal components beyond the fourth) have very low signal-to-noise ratios and can therefore be expected to carry very little information. Although we have not directly examined temporal encoding here, we have shown that the variability of a low-frequency measure of spike timing (the second principal component of the responses) is closely linked to the average spike count of a set of neuronal responses. This suggests that much of the information that could be carried by spike timing is linked to spike count. Principal component analysis is a linear method and may not efficiently detect some aspects of timing. However, we and others have shown that highly nonlinear details of millisecond-precision spike timing are closely related to spike count and its variability (Oram et al., 1999; Richmond et al., 1999; Baker and Lemon, 2000). Thus, two different lines of reasoning point toward the conclusion that consistency of spike-count encoding implies consistency of many aspects of spike timing across stimulus sets in V1 neurons and across behavioral contexts in TE neurons.

Conclusion

We have shown that responses elicited by different kinds of stimuli from neurons in V1 and responses elicited by stimuli shown in different behavioral contexts from neurons in TE share statistical properties that are important in determining how much stimulus-related information the responses contain. A single regression characterizes the relation between log(mean) and log(variance) of spike count in different behavioral contexts in TE. Although the difference between a model with a single regression and a model using a separate regression for each stimulus set was significant in some V1 neurons, the differences in predicted variances were so small as to be irrelevant for decoding, even when comparing naturalistic and geometric stimuli. The advantage of consistent encoding is its simplicity; downstream neurons (and researchers) can process or decode all signals from a neuron in the same way, without needing to determine which encoding rules are in effect.

REFERENCES

Ahmed N, Rao KR (1975) Orthogonal transforms for digital signal processing. Berlin: Springer.
Atick JJ, Redlich AN (1990) Towards a theory of early visual processing. Neural Comput 2:308–320.
Atick JJ, Redlich AN (1992) What does the retina know about natural scenes? Neural Comput 4:196–210.
Bair W, O'Keefe LP (1998) The influence of fixational eye movements on the response of neurons in area MT of the macaque. Vis Neurosci 15:779–786.
Baker SN, Lemon RN (2000) Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. J Neurophysiol 84:1770–80.
Bell AJ, Sejnowski TJ (1997) The "independent components" of natural scenes are edge filters. Vision Res 37:3327–3338.
Britten K, Shadlen M, Newsome W, Movshon J (1993) Responses of neurons in macaque MT to stochastic motion signals. Vis Neurosci 10:1157–1169.
Buracas G, Zador A, DeWeese M, Albright T (1998) Efficient discrim-

ination of temporal patterns by motion-sensitive neurons in primate visual cortex. Neuron 20:959–969.

Chelazzi L, Duncan J, Miller EK, Desimone R (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. J Neurophysiol 80:2918–2940.

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

Croner LJ, Álbright TD (1999) Segmentation by color influences responses of motion-sensitive neurons in the cortical middle temporal visual area. J Neurosci 19:3935–3951.

Dan Y, Atick JJ, Reid RC (1996) Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. J Neurosci 16:3351–3362.

Dean AF (1981) The variability of discharge of simple cells in the cat striate cortex. Exp Brain Res 44:437–440.

de Ruyter van Steveninck RR, Lewen GD, Strong S, Koberle R, Bialek W (1997) Reproducibility and variability in neural spike trains. Science 275:1805–1808.

Dong DW, Atick JJ (1995) Statistics of natural time-varying images. Network 6:345–358.

Eskandar EN, Richmond BJ, Optican LM (1992) Role of inferior temporal neurons in visual memory. I. Temporal encoding of information about visual images, recalled images, and behavioral context. J Neurophysiol 68:1277–95.

Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am A 4:2379–2394.

Fuller WA (1987) Measurement error models. Wiley series in probability and mathematical statistics. New York: Wiley.

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.

Gur M, Beylin A, Snodderly DM (1997) Response variability of neurons in primary visual cortex (V1) of alert monkeys. J Neurosci 17:2914–2920.

Heller J, Hertz JA, Kjaer TW, Richmond BJ (1995) Information flow and temporal coding in primate pattern vision. J Comput Neurosci 2:175–193.

Kendall M, Stuart A (1961) The advanced theory of statistics, Vol 2, Inference and relationship, pp 4–6. London: Hafner.

Lee D, Port NL, Kruse W, Georgopoulos AP (1998) Variability and correlated noise in the discharge of neurons in motor and parietal areas of primate cortex. J Neurosci 18:1161–1170.

Levine MW, Cleland BG, Mukherjee P, Kaplan E (1996) Tailoring of variability in the lateral geniculate nucleus of the cat. Biol Cybern 75:219–227.

Liu Z, Richmond BJ (2000) Response differences in monkey TE and perirhinal cortex: stimulus association related to reward schedules. J Neurophysiol 83:1677–1692.

McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J Neurosci 19:431–441.

Olshausen BA, Field DJ (1996a) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–609.

Olshausen BA, Field DJ (1996b) Natural image statistics and efficient coding. Network 7:333–339.

Optican LM, Richmond BJ (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. J Neurophysiol 57:162–178.

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.

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.

Richmond BJ, Optican LM (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. J Neurophysiol 57:147–161.

Richmond BJ, Optican LM (1990) Temporal encoding of two-dimensional patterns by single units in primate visual cortex. II. Information transmission. J Neurophysiol 64:370–380.

Richmond BJ, Hertz JA, Gawne TJ (1999) The relation between V1 neuronal responses and eye movement-like stimulus presentations. Neurocomputing 26–27:247–254.

Rieke F, Bodnar DA, Bialek W (1995) Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. Proc R Soc Lond B Biol Sci 262:259–65.

Ripley BD, Thompson M (1987) Regression techniques for the detection of analytical bias. Analyst 112:377–383.

Rolls ET, Tovee MJ (1995) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. J Neurophysiol 72:713–726.

Shannon CE, Weaver W (1949) The Mathematical Theory of Communication. Urbana, IL: Illinois UP.

Spitzer H, Richmond BJ (1991) Task difficulty: ignoring, attending to, and discriminating a visual stimulus yield progressively more activity in inferior temporal neurons. Exp Brain Res 83:340–348.

Sugase Y, Yamane S, Ueno S, Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. Nature 400:869–73.

Tolhurst DJ, Movshon JA, Thompson ID (1981) The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. Exp Brain Res 41:414–419.

Tolhurst DJ, Movshon JA, Dean AF (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Res 23:775–785.

Tovee MJ, Rolls ET (1995) Information encoding in short firing rate epochs by single neurons in primate temporal visual cortex. Visual Cognition 2:35–58.

Tovee MJ, Rolls ET, Treves A, Bellis RP (1993) Information encoding and the responses of single neurons in the primate temporal visual cortex. J Neurophysiol 70:640–654.

van Hateren JH, Ruderman DL (1998) Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. Proc R Soc Lond B Biol Sci 265:2315–2320.

van Hateren JH, van der Schaaf A (1998) Independent component filters of natural images compared with simple cells in primary visual cortex. Proc R Soc Lond B Biol Sci 265:359–366.

van Kan PLE, Scobey RP, Gabor AJ (1985) Response covariance in cat visual cortex. Exp Brain Res 60:559–563.

Victor JD, Purpura KP (1996) Nature and precision of temporal coding in visual cortex: a metric-space analysis. J Neurophysiol 76:1310–1326.

Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287:1273–1276.

Vogels R, Spileers W, Orban GA (1989) The response variability of striate cortical neurons in the behaving monkey. Exp Brain Res 77:432–436.

Warzecha A-K, Egelhaaf M (1999) Variability in spike trains during constant and dynamic stimulation. Science 283:1927–1930.

Warzecha A-K, Kretzberg J, Egelhaaf M (2000) Reliability of a fly motion-sensitive neuron depends on stimulus parameters. J Neurosci 20:8886–8896.

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.