Neural Computation, *in press.*

Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli

Alessandro Treves[¶], Stefano Panzeri[§], Edmund T. Rolls[§]
Michael Booth[§] and Edward A. Wakeman[§]

[¶] *SISSA - Programme in Neuroscience, 34013 Trieste, Italy*
[§] *University of Oxford, Department of Experimental Psychology, Oxford OX1 3UD, United Kingdom*

keywords: spiking statistics, firing variability, information coding, single-neuron model, natural stimuli, IT cortex

Running title: Firing rate distributions, and information transmission

Abstract

The distribution of responses of sensory neurons to ecological stimulation has been proposed to be designed to maximize information transmission, which according to a simple model would imply an exponential distribution of spike counts in a given time window. We have used recordings from inferior temporal cortex neurons responding to quasi-natural visual stimulation (presented using a video of everyday lab scenes, and a large number of static images of faces and natural scenes) to assess the validity of this exponential model and to develop an alternative simple model of spike count distributions. We find that the exponential model has to be rejected in 84% of cases (at the $P < 0.01$ level). A new model, which accounts for the firing rate distribution found in terms of slow and fast variability in the inputs which produce neuronal activation, is rejected statistically in only 16% of cases. Finally, we show that the neurons are moderately efficient at transmitting information, but not optimally efficient.

Introduction

The firing rates of single neurons from the inferior temporal visual cortex, in response to a large, natural set of stimuli, typically have a distribution which is graded (continuous from zero response to the maximum), unimodal (has a single peak often close the spontaneous firing rate, or to zero) and has an approximately exponential tail. Such firing rate distributions, usually expressed in terms of the number of spikes in a window of fixed length, are a common observation in many parts of the brain, including the frontal cortex (Abeles et al., 1990) and the hippocampus and related structures (Barnes et al., 1990). Indeed, exponential distributions have been used in formal neural network analyses as a first, simple model of realistic distributions of graded firing rates (Treves and Rolls, 1991).

Part of the special interest of this observation in the case of sensory, e.g. visual, cortices is however in the fact that an **exponential** distribution of spike counts would maximize their entropy (and thus the information a cell would transmit, for a *noiseless discrete* code), under the constraint of a fixed average count (Shannon, 1948)[1]. It has therefore been suggested (Levy and Baxter, 1996; Baddeley,

[1]What is discussed here is the entropy of the spike count distribution. The entropy of the spike *train*, i.e. of the collection of spike emission times, is maximized by Poisson spiking processes.

1996; Baddeley et al., 1997) that neurons might tend to use an exponential distribution of firing rates because it would be the most efficient in metabolic terms, i.e., it would be the most informative once a given metabolic "budget" has been set in terms of the average rate. It remains dubious, though, whether the distribution that maximizes information transmission in the absence of noise would be of any particular value in the presence of noise. Moreover, other distributions would be optimal under somewhat different assumptions. For example, a very different distribution, a **binary** distribution, would instead maximize the *instantaneous rate* at which information about slowly varying stimuli is transmitted by a noisy spiking code, of the type used by neurons in the brain (Panzeri et al., 1996a).

Other approaches to understanding the observed spike count distributions are not based on the notion that such distributions would reflect the optimization of information transmission, but rather on the idea that they would simply result from the intrinsic variability of the underlying process. Spikes are produced in a neuron when a sufficient amount of current enters the soma. (The current rather than the voltage is the relevant activation variable when considering the emission of several spikes (Treves and Rolls, 1991; Koch et al., 1995)). The current is the sum of many synaptic inputs, and if these are weakly correlated, over a large set of natural stimuli, the distribution of the current can be expected by the central limit theorem to be approximately normal, that is Gaussian. Since going from current to spike count involves what is essentially a rectification (a threshold-linear transform), a possible model for the spike count distribution would be a **truncated Gaussian**, with an additional peak at zero. The truncated Gaussian model considers only the mean current over a time window, and a fixed, deterministic, conversion of this value into a spike count, thus it neglects sources of rapid variability. A simple and widely used model that instead emphasizes fast variability is that of Poisson firing. In the Poisson model, spike emission is a stochastic event, with a probability dependent on the instantaneous value of the activation. If this source of fast variability dominates over the slower fluctuations of the mean current, the shape of the resulting spike count distribution is close to a simple **Poisson** distribution.

We use face selective cells (Rolls, 1984) recorded from the macaque inferior temporal visual cortex to show that *none* of these simple models, though attractive in their simplicity, captures satisfactorily the statistics of firing in response to large sets of natural stimuli. On the other hand, we find that a model, the S+F random model, which includes both *slow* and *fast* variability in the activation as generators of the spike count distribution, and takes them both to be roughly normally distributed, comes much closer to an accurate description of the observed statistics. We then discuss why the agreement with this still rather simple model is not, and should not be expected to be, perfect. Finally we measure the efficiency with which the observed distribution transmits information. Although the model we propose as an explanation has nothing to do with optimising efficiency, we do find that the efficiency is moderately high.

Part of the results have been previously published in abstract form (Panzeri et al., 1996b).

Methods

Collection of firing rate distributions

We constructed firing rate probability distributions of 15 visual neurons shown to have face selective responses recorded in the cortex in the anterior part of the superior temporal sulcus of 2 rhesus macaques. All the cells had responses that were face selective in that they responded at least twice as much to the most effective face stimulus as to the most effective non-face stimulus in initial screening (for criteria see Rolls and Tovée (1995)). Spike arrival times were recorded with a Datawave spike acquisition system. During the recording session the monkey was looking at a 5-min video showing natural scenes familiar to the monkey. The video included views of laboratory rooms and equipment as well as people and monkeys typically seen daily by these animals, and was continuously recorded with no breaks or cuts. For approximately 30 % of the video time at least one face was included in the view. During the video, the monkey was not required to maintain any static visual fixation, in an attempt to emulate natural viewing behaviour. As the video was shown more than once when recording some of the cells, a total of

22 datasets (from monkey *ay* 20 datasets, and from monkey *ba* 2 datasets) were available for the analysis. We emphasize that the behaviour of the monkey during consecutive recording sessions from the same cell was not constrained to be the same, and in particular eye movements during the video could vary from one presentation to the next. The neurophysiological methods have been described previously (Rolls and Tovée, 1995).

Histograms of single-cell response probability distributions for the video data were created by dividing the recording time into non-overlapping time bins of a given length, and counting the number of spikes within each bin. The error bars for each histogram are drawn according to the following standard procedure. The fluctuations in the observed frequency in each bin can be thought of as coming from the stochastic process of making random selections among a finite numbers of items, following therefore a distribution close to Poisson. The standard deviation of the observed frequency for each particular spike count is thus estimated as the square root of the observed frequency for that count, independently of the underlying distribution of the count frequencies across different spike counts. We used time windows of length $L = 50$, 100, 200, 400 and 800 ms.

As a further comparison, we analyzed the responses of a second set of 14 face-selective IT neurons recorded from another monkey (*am*), when a set of 65 static visual stimuli (natural images; 23 faces and 42 non-face images) were repeatedly presented to the animal during a visual fixation task. This large set of different stimuli was used to provide evidence on how the neurons would distribute their responses to a wide range of stimuli similar to those that they might see naturally. The stimulus was presented for 500 ms on each trial. This latter set of data has been previously studied using information theoretical techniques (Rolls and Tovée, 1995; Rolls et al., 1997b). In this paper we analyze this data, and we find that with such static images the spike count distribution and other measures are very similar to those obtained with the video data. The fact that similar results on the distribution of spike counts were obtained with two different cell samples, and with very different ways of presenting the stimuli (continuous video vs a large set of static images) corroborates the findings and analyses described in this paper.

Histograms of single-cell firing rate probability distributions for the static stimuli data were created by counting the number of spikes emitted by the cell in a time window L ms long, starting 100 ms after stimulus onset, in response to each of the 65 static stimuli in the sample, for each stimulus presentation. The firing rate probability histograms were calculated on the basis of 380-600 trials for each cell. Error bars for the histograms were calculated as for the video data. We used time windows of length $L = 50$, 100, 200 and 400 ms.

Models of the firing rate distribution of single cells

Exponential distribution. This is the one-parameter distribution for the spike count n

$$P(n) = \frac{1}{1 + (\bar{r}L)} \exp\left(-\lambda n\right), \tag{1}$$

where λ is determined by the the single parameter $\bar{r}$ as follows:

$$\lambda = \ln\left(1 + \frac{1}{\bar{r}L}\right) . \tag{2}$$

The parameter $\bar{r}$ is calculated from the data as the total spike count divided by the total duration of the video. Notice that the exponential distribution (1) maximizes the entropy of the spike count under the constraint of fixed average firing rate (Rieke et al., 1996).

The significance of the agreement between the observed distribution for each dataset and this model distribution is evaluated from the χ^2 statistics. $\chi^2(L)$ is measured for each time window as

$$\chi^2 = \sum_n \frac{[N|P_{obs}(n) - P_{model}(n)| - 1/2]^2}{NP_{model}(n)} \tag{3}$$

where subtracting 1/2 is a procedure known as the Yates correction (or correction for continuity (Fisher and Yates, 1963)). This correction takes properly into account the fact that $NP_{obs}(n)$, unlike $NP_{model}(n)$, can take only integer values, and is useful to ensure that histogram bins with very few expected events $NP_{model}(n)$ do not weigh disproportionately on χ^2. In addition, bins with zero observed events, such as those at the large-n tail of the distribution, are all grouped together with the nearest preceding bin with at least one event. This is again a standard procedure to enable the use of the simple form of χ^2 distribution. From $\chi^2(L)$ we derive, using the relevant probability distribution of χ^2, the likelihood $P(L)$ of the fit for time window L. Since the only free parameter of the probability distribution, $\bar{r}$, is not determined by minimizing χ^2, but directly from the data, and it is common to all 5 window lengths L, we use as the number of degrees of freedom of the fit $df = n_{bins} - 1 - (1/5)$, with n_{bins} the final number of histogram bins.

Binary distribution. A binary distribution of firing rates is clearly not what is found[2], even in other experiments in which static stimuli are used, so there is no point trying to fit it to the observed spike count.

Poisson distribution. This is expressed as

$$P(n) = \exp(-\bar{r}L)\frac{(\bar{r}L)^n}{n!}, \tag{4}$$

and again the only parameter is $\bar{r}$. Since $\bar{r}$ is just the average of the rate $r_i = n_i/L$ which is different in each time bin i, the distribution of n_i across bins will remain approximately Poisson only to the extent that the variability in r_i is minor compared to the fast variability ("within bin" variability) described by the model. With respect to fitting it to the observed distribution, the same considerations and procedures apply as with the exponential model.

Truncated Gaussian distribution. This model assumes that the spike count reflects an underlying activation variable, the current $h(t)$ flowing into the cell body, which is approximately normally distributed around a mean $\bar{h}$. The fluctuations in h, of width $\sigma_S(L)$, are taken to occur on time scales *slow* with respect to the window length L, so that $h(t) \simeq h_i$ at any time t between iL and $(i+1)L$. The current is taken to translate into a firing frequency through simple rectification, that is the firing rate in a time bin, r_i, is zero if the activation h_i is below a threshold, T, and is linear above the threshold:

$$\begin{aligned} r &= 0 && \text{if} \quad h < T \\ r &= g(h-T) && \text{if} \quad h > T \end{aligned}$$

Such an input-output function is a simple but often reasonable model of current-to-frequency transduction in pyramidal cells (Lanthorn et al., 1984) and at the same time it lends itself easily to analytical treatment (Treves and Rolls, 1991; Treves and Rolls, 1992). We note also that integrate-and-fire models of neuronal firing operate effectively with an activation function which is close to threshold-linear (Amit and Tsodyks, 1991; Treves, 1993). Note that saturation effects in the firing rate are not represented (they could be included, if they turn out to be important, at the price of analytical simplicity, but are shown not to be important in the Discussion). The parameter g is the gain of the threshold-linear transform, and here it is fixed, $g = 1$, so that the activation h is measured, like r, in Hz.

The probability distribution for r is then continuous and identical to the one for h above threshold

$$p(r_i)dr_i = \frac{1}{\sqrt{2\pi}\sigma_S}\exp -\frac{[r_i - (\bar{h}-T)]^2}{2\sigma_S^2}dr_i \tag{5}$$

and

$$P(r_i = 0) = \phi[(T-\bar{h})/\sigma_S] \tag{6}$$

[2]Bursting cells might intuitively be taken to behave as binary units, but in any case we are not aware of any similar experiment reporting binary spike counts, with a cell emitting say either 0 or 3 spikes in a given window.

where we have defined

$$\phi(x) \equiv \int_{-\infty}^{x} \frac{dx}{\sqrt{2\pi}} \exp -(\frac{x^2}{2}) \tag{7}$$

The spike count distribution is a discrete histogram, whereas this model has been derived, above threshold, in the form of a continuous distribution of real-valued frequencies. To convert the continuous probability distribution $p(r)$ of the model into a histogram $P(n)$, we make the rough assumption that if the mean firing rate across the bin, r_i, exactly equals the integer n_i/L, the number of spikes observed will be n_i, whereas if it is slightly above or below it, the number of spikes, depending on when the first spike occurs, may also be observed to be respectively $n_i + 1$ or $n_i - 1$. Specifically,

$$P(n) = \int_{(n-1)/L}^{n/L} dr[Lr - n + 1]p(r) + \int_{n/L}^{(n+1)/L} dr[n - Lr + 1]p(r). \tag{8}$$

$p(r)$ is then discretized, but in much finer size than $1/L$, to evaluate $P(n)$ numerically.

To fit this model to the observed distributions we have now to determine two parameters, σ_S and $h_0 \equiv \bar{h} - T$. These are not measured directly from the data, but obtained by minimizing the discrepancy in the fit, as explained below.

Adding fast fluctuations: the S+F model. The new model we consider can be seen as an extension of the one generating the truncated Gaussian distribution, in the sense of taking into account also fast fluctuations in the activation variable (conceptually similar to the variability in the response to a given stimulus), assumed to be roughly normally distributed like the slow fluctuations (which conceptually may be roughly compared with the different responses produced to each stimulus, which can be thought of as changing in the video and in the world with a time scale usually – but not always – longer than the time window considered). To be clear, "fast" here means fluctuations occurring over time scales shorter than the window L used in counting spikes, and it does not imply any definite assumption as to what is stimulus, for any given cell, and what is noise. In practice, it leads to the distribution for the instantaneous current

$$p(h(t)) = \frac{1}{\sqrt{2\pi}\sigma_F} \exp -\frac{(h(t) - h_i)^2}{2\sigma_F^2} \tag{9}$$

The standard deviation σ_F, like that of the slow fluctuations denoted as σ_S, is measured in Hz, and will be a function of the window length L, as a short window will make moderately fast fluctuations appear slow, and viceversa. The model takes slow and fast fluctuations to be uncorrelated, and we expect their total squared amplitude to be roughly constant across time windows, $\sigma_S^2(L) + \sigma_F^2(L) \simeq$ constant (see Results and Discussion).

It is again assumed that the current results deterministically in a firing frequency, through a simple threshold-linear activation function. Now however the input-output transform is taken to hold at any moment in time, as if an instantaneous firing frequency $r(t)$ could be defined, instead of the discrete events represented by individual spikes:

$$\begin{aligned} r(t) &= 0 \qquad\qquad\quad \text{if} \quad h(t) < T \\ r(t) &= g(h(t) - T) \quad \text{if} \quad h(t) > T. \end{aligned}$$

In the model, the firing rate distribution is then obtained by averaging over the fast noise. The mean response r_i in a given time window is calculated by averaging $r(t)$ over fast fluctuations, and expressing the result as a function of the mean activation h_i during the window:

$$r_i = \sigma_F \left[\frac{1}{\sqrt{2\pi}} \exp -\frac{(h_i - T)^2}{2\sigma_F^2} + \frac{h_i - T}{\sigma_F} \phi\left(\frac{h_i - T}{\sigma_F}\right) \right]. \tag{10}$$

From this equation, and taking into account the normal distribution of slow fluctuations, we obtain the probability distributions of firing rates:

$$p(r_i) = \frac{1}{\sqrt{2\pi}\sigma_S} \exp -\frac{(h_i(r_i) - \bar{h})^2}{2\sigma_S^2} \left[\phi\left(\frac{h_i(r_i) - T}{\sigma_F}\right)\right]^{-1} \tag{11}$$

where $h_i(r_i)$ is determined by inverting eq. (10).

The averaging has the effect of smoothing over the fluctuations that are faster than the time window, and as a result the effective input-output transform, eq. (10), is as illustrated in Fig.1.

Again, the discrete spike count distribution is derived from the continuous firing rate distribution using the procedure (Eq. 8) described above for the truncated Gaussian model.

The fit of this model to the observed distributions involves 3 free parameters, h_0 (again one can see that it is only the difference $\bar{h} - T$ that counts), σ_F and σ_S, which as for the truncated Gaussian model must be obtained by minimization.

Procedures and parameters for the fit

We fit the models to the 5 observed distributions of spike counts in response to the video, simultaneously across the 5 time windows, for each dataset. For the exponential and Poisson models this is straightforward, since the only free parameter can be read off the data directly. For the last two models there are more free parameters, and they must be adjusted so as to optimize the fit. One parameter that we did not allow to vary for different time windows is the difference between the mean input and the threshold, that is h_0. $\sigma_S(L)$ and $\sigma_F(L)$ (for the truncated Gaussian model there is only $\sigma_S(L)$) are, instead, left free to vary independently for each time window. Thus we have $1 + 5$ free parameters for the truncated Gaussian model, and $1 + 5 \times 2$ for the S+F model including fast variability. These are chosen by maximizing a pseudo maximum likelihood cost function, which we construct as

$$C = -\sum_L \log(P(L)) \tag{12}$$

from the probabilities $P(L)$ that, at each time scale, the observed distribution could indeed be generated from the model. $P(L)$ is derived from the $\chi^2(L)$ value calculated as for the exponential model, using now as the number of degrees of freedom $df = (n_{bins}) - 1 - 1 - (1/5)$ for the truncated Gaussian model and $df = (n_{bins}) - 1 - 2 - (1/5)$ for the S+F model, where again we take into account with the $(1/5)$ that h_0 is in common to 5 time windows. Note that the same data are used with all windows L, so the 5 fits are not independent, and a single fit with $df = 11$ would not yield a meaningful P level. When optimizing the fit of the S+F and truncated Gaussian models to the data, we also tried allowing h_0 to take different values for the different time windows. As expected, the fits are better than those shown in Figs. 2 and 3 and Table 1. However, we prefer to rely on the condition in which the parameter h_0 is the same for a given dataset for the different time windows, because h_0 is meant to correspond to the same physical variable across window lengths (i.e., the mean input current over the recording session).

The maximization is performed using standard routines from Press et al. (1992). In particular, within a loop over time scales, for each L either $\sigma_S(L)$ alone or $\sigma_S(L)$ and $\sigma_F(L)$ are first optimized for any given value of h_0 by routine `amoeba`, and then the common h_0 value is optimized by (unidimensional maximization) routines `golden` and `mnbrak`. At the end, the individual levels of significance $P(L)$ are extracted.

Note that for visual inspection, the observed histograms are plotted in their original form, without any grouping.

Exactly the same analysis is applied to the probability distributions of the responses to static stimuli, the only difference being that in this case the single time windows are fitted separately. This is because in this case the time window is locked to the stimulus onset, and therefore the mean activation h_0 was not expected in this case to be constant across different time windows.

Error estimates of the parameters. This can be done, for the S+F model, by calculating the region in parameter space within which C increases by no more than a set amount. We do not report error bars on individual parameters because they would misleadingly indicate that each parameter may be undetermined to that extent. In fact, it turns out that both $\sigma_S(L)$ and $\sigma_F(L)$ largely covary with h_0. h_0 itself has large error bars when it is large negative, whereas it is much better determined if the distribution has a clear peak above zero, which then roughly corresponds to h_0.

Power spectra of the firing rate distributions

Power spectra of the firing rates are computed using standard routines from the train of spikes recorded during each video presentation. We calculated, for every cell, the power spectra using a sampling frequency of either 100, 500 or 1000 Hz, dividing the data into non-overlapping segments each 256 data points long, and windowing with a Bartlett window (Press et al., 1992). The non-overlapping segments used for the spectral analysis covered all the 5 minutes of the video presentation. We produced also a normalized and averaged power spectrum across the datasets, by first normalizing the power spectra of individual cells (to give a total power of 1), and then averaging over the dataset.

Efficiency measure

For each dataset and each window L, we extract from the observed spike count distribution the so called information per spike[3]

$$\chi = \sum_n P(n) \frac{n}{\bar{r}L} \log_2 \frac{n}{\bar{r}L}, \tag{13}$$

a quantity introduced by Bialek et al. (1991) and Skaggs et al. (1993), and which we argue below to be relevant. As shown by Panzeri et al. (1996a), this quantity can range between zero and a maximum,

$$0 \leq \chi \leq \log_2(1/a), \tag{14}$$

where a is the *sparseness* (Treves and Rolls, 1991) of the distribution, defined as

$$a = \frac{(\sum_i r_i/N)^2}{\sum_i (r_i)^2/N} = \frac{1}{\sum_n P(n)(n/\bar{r}L)^2} \tag{15}$$

(N is the number of events, that is time bins). Note that the information per spike, χ, reaches its maximal value of $\log_2(1/a)$ for a binary distribution.

We then define our efficiency measure, at each scale L, as

$$\varrho(L) \equiv \frac{\chi}{\log_2(1/a)}. \tag{16}$$

so that ϱ varies between 0 and 1. Further aspects of this measure are considered in the Results and Discussion.

A related efficiency measure, used by Bialek et al. (1991), compares the instantaneous rate of information transmission to the entropy rate. Its expression in the short time limit is as follows:

$$\varrho_B(L) \equiv \frac{\chi}{\log_2(e/\bar{n})}, \tag{17}$$

[3]The quantity χ is, in general, the time derivative of the information (about any correlated variable; here, the time bin of each count) conveyed by the spike count, and divided by the mean firing rate. For a Poisson process with independent spikes, χ acquires the additional and intuitive meaning of average information conveyed by each spike. We use the notation χ for consistency with previous papers (Panzeri et al., 1996a), even at the risk of confusion with the – unrelated – quantity χ^2 appearing in this paper.

and can be applied with reasonable accuracy only to *very* short time windows, so that the mean number of spikes in the window $\bar{n} \ll 1$. ϱ_B also varies between 0 and 1. Note that for very short windows, as the distribution of spike counts in individual bins becomes binary, 0 or 1, $\chi \to \log_2(1/a)$ and $\varrho \to 1$. ϱ_B also tends to 1 but much more slowly (only for very short windows), because $\log_2(e/\bar{n}) \to \log_2(1/a) + \log_2(e)$

Results

Spike count distributions and how they fit simple models

To emulate naturally occurring firing activity, we recorded from face selective cells in the inferior temporal cortex of monkeys who were viewing a video of everyday lab scenes, and whose eye movements, level of attention and motivation were not constrained. The firing rate distributions of the 22 datasets collected in these conditions show the general trend discussed in the introduction, in that they appear graded, unimodal (with the single peak either close to zero or to the spontaneous activity) and with an exponential-like tail. (These datasets form the majority of those also analysed by Baddeley et al. (1997). They included in addition 4 datasets from non-face selective neurons, which we exclude here because the effective stimuli for these datasets were not known.)

We present now the single cell quantitative analysis for the video data, showing that the S+F model accounts for most of the data, while the exponential model does not describe the data satisfactorily, especially at low rates, and the truncated Gaussian and Poisson models do not fit at all. Table 1 gives a summary of how well the 4 models considered fit the observed distributions of spike counts. Considering 22 datasets (from 15 different face cells, some of which were recorded over multiple presentations of the video) and 5 window lengths, there are a total of 110 possible fits. Setting a confidence level of $P = 0.01$, we would expect that, if a true underlying distribution existed and were known to us, it would fit the data at this confidence level always except about once. The 4 simple models are not expected to get this close to the data. In fact, they differ considerably in the extent to which they can explain the observed distributions. The Poisson model is always rejected, and the truncated Gaussian almost always (for 96.3% of the cases). The exponential model is rejected for 83.6% of the cases. The S+F model is rejected for only 15.4% of the cases: of the 22 datasets at five time scales, 13 give acceptable fits at all time scales, 3 are rejected for one window length, 4 are rejected at 2, 2 at 3 window lengths, and for no dataset is the fit to be rejected at 4 or all 5 window lengths. The rejections are more concentrated at shorter window lengths: 5 at 50 ms, 4 at 100 ms, 6 at 200 ms, 2 at 400 ms, and none at 800 ms. We should remember that the sharp separation between slow and fast fluctuations implied in the S+F model is clearly a simplification, particularly when considering multiple time scales, in each of which the separation is made at a different value. Nevertheless for 13 datasets the fit is acceptable at *all* window lengths, and for all datasets it is acceptable for at least 2 time window lengths. We can conclude that the agreement between the data and the S+F model is not just qualitative, but is quantitatively a good fit.

We show the response probability distributions for 9 examples of the 22 available datasets, in Figs. 2 (for a time window of 100 ms) and 3 (for a time window of 400 ms). These 9 datasets and two windows are representative of the quality of the results in all 22 datasets and 5 windows, as indicated by the similar percent of model rejections when considering only these 18 cases: S+F = 16.6% ; exp = 83.3%; Poisson = 100%; Gaussian = 100% . An exponential curve fitted to the mean rate of each dataset is shown by the dashed line in Figs. 2 and 3. It is clear, particularly for the longer window, that many of the cells have a poor fit, with too few spikes at very low rates, and too many spikes at a slightly higher rate. This is confirmed by the statistical analyses using the χ^2 goodness of fit test, as shown in Table 1. In Figs. 2 and 3 we also show the fit of the S+F model to the observed rate distributions (solid line). The fits look much more acceptable. Again, this is confirmed by the statistical analyses using the χ^2 goodness of fit test, as quantified in Table 1. The Poisson and truncated Gaussian models, which give much worse fits than the exponential model, are not shown. In particular, these last two models tend to tail off at high spike counts much faster than the real data. In addition, the Poisson model constrains the distribution to have a peak value at a non zero count (in fact, at the mean count), and this is much worse than constraining it to have the peak always at zero, as in the case with the exponential model.

To check that the general shape of the distribution is not especially dependent on either watching a video, or on the particular video, we performed a similar analysis on the spike count distributions of the second population of 14 neurons responding to 65 natural static visual stimuli, presented during a visual fixation task. We present in Fig. 4 three representative firing rate distributions of three different cells responding to static stimuli. (The time window is 100 ms in Fig. 4a and 400 ms in Fig. 4b). It is evident that the main features of the distribution (namely the gradedness, the unimodality and the exponential-like tail) are found also in this case. Table 1 summarizes, for the experiment with static stimuli, the results of the statistical analysis using the χ^2 test for the fits of the observed distributions of spike counts to the various models. Considering 14 cells and 4 time windows, there are a total of 56 possible fits. As before, the Poisson model was always rejected, and the truncated Gaussian almost always (for 89.2 % of cases). (The level of statistical significance used throughout was $P < 0.01$). The exponential model was rejected for 75% of cases. The S+F model was rejected only in one case (corresponding to 1.8%, reasonably within the acceptable range of rejections). We note that, as the number of trials was smaller than in the video case, the different models were rejected less often for the experiment with static stimuli, but the number of trials was nevertheless high enough (380-600) to test adequately the fit of the firing rate distributions (see Fig. 4), and to rule out all of the models analyzed apart the S+F one (see Table 1). A further appreciation of the fact that the exponential distribution is not a very good fit to the observed firing rate distribution, especially at low rates, is provided by showing the average of the rate distribution across cells (possible by normalising the mean rate of each cell to 1). This graph is shown as Fig. 8a of (Rolls et al., 1997b). Further, the graph for the video data averaged across cells and plotted in the same way appeared very similar to that Figure, and for that reason is not reproduced here.

We conclude that, despite its shortcomings which, as discussed below, we ascribe mainly to its simplicity, our hypothesis, that is the S+F model, accounts for most of the data satisfactorily, and in any case better than the hypothesis that these distributions of firing rates tend to be maximum-entropy exponential distributions. The Poisson model has little to do with the data. The truncated Gaussian model, whose sole difference with the S+F model is that it does not take fast variability into account, also gives very poor fits.

Parameters of the S+F model from the fits and power spectra

We present the results for the parameters of the best fits to the S+F model only for the video data. The reason for this is that the most interesting point that can be made from studying the parameters, namely the dependence of the parameters σ_F and σ_S on the duration of the time window, can be studied only for the video data, where the mean activation h_0 was constant across the time windows and therefore the different time windows can be fitted simultaneously[4]. For the video data, the parameters extracted from the fits for the S+F model take values in the range from about 10 to 100 Hz. The mean activation h_0 (expressed in units of firing rate) varies from about -60 to +30 Hz across different datasets. Significant relationships were found between the parameters σ_F and σ_S describing, at different time scales, the amplitudes of the fast and slow fluctuations underlying each particular dataset. In particular, the proportion of fast fluctuations increased logarithmically with the length of the time window (see Fig. 6). This probably reflects that fact that for natural scenes the power in the temporal spectrum of the intensity of a pixel decreases approximately as the inverse of the temporal frequency (Dong and Atick, 1995).

If the fluctuations can in fact be divided to a good approximation between fast and slow, as the simple model posits, and if the parameters from the fit were to precisely quantify their amplitudes, one

[4]For the static image data, either multiple non-overlapping short windows are taken for each trial, e.g. 8 consecutive 100ms windows, but then they would refer to different phases in the responses; or the data analysed is not the same across windows, and therefore h_0 is not common to different window lengths.

would expect the total power to remain constant, irrespective of where the division between slow and fast falls,

$$\sigma_S^2(L) + \sigma_F^2(L) = \sigma_{TOT}^2 = \text{constant.} \tag{18}$$

To check to what extent this occurs for each dataset, Fig. 5 shows the relation between the slow and fast variability parameters, with the points pertaining to the same dataset linked by lines. If Eq. 18 were to be satisfied, all curves would describe arcs of a circle. One sees that for most datasets the expected relation is not far from the observed one, considering that each of the 5 points on the curve comes from a different χ^2 minimization. This is a useful consistency check of the S+F model, and in particular indicates that the parameters σ_S and σ_F may indeed be associated with slow and fast fluctuations.

Given this association, the observation that in Fig. 5 most datasets lie between approximately 30^o and 60^o has further interesting implications. These are made clearer in Fig. 6, where the proportion of the power assigned by the fit to fast fluctuations,

$$F/(S+F) = \frac{\sigma_F^2(L)}{\sigma_S^2(L) + \sigma_F^2(L)} \tag{19}$$

is plotted against the time scale of the window used (on a log scale; each dashed line represent a different dataset). If each of the 4 octaves included between our 5 time scales were to contribute an equal power to the fluctuations, the curves joining points from the same dataset would be straight upward lines. This appears indeed to be the general trend, as emphasized by the bold line in Fig. 6 which gives the average over the 22 datasets. Moreover, most of the curves start at values of $F/(S+F)$ around 0.2 and end at values around 0.6-0.7. From these observations, we conclude that as the length of the time window increases, the contribution of the fast fluctuations increases approximately logarithmically. Given that the total variance of the slow and fast fluctuations is approximately constant, this implies that the contribution of the slow fluctuations decreases logarithmically. This is what would be expected if the sources of the slow and fast fluctuations were evenly distributed across time scales over a wide range, that extends from below our shortest window of 50 ms to above the longest one of 800 ms. In statistics this trend is often referred to as the $1/f$ law and it is theorized to underly many random processes. Note that if the power of fluctuations were distributed as $p(f)df \propto df/f$, then the fraction of power between two frequencies f_1 and $f_2 = f_1/2$ differing by an octave would be proportional to $\log(f_1) - \log(f_2) = \log 2$, that is constant, as we do approximately find. An underlying basis of this may simply be that there is no intrinsic time scale which is characteristic of the real images being seen in the video, in which some images, or elements of images, last for a long time, and others for a shorter time (see Dong and Atick (1995) for a quantitative analysis of the statistics of natural images).

It is interesting in this respect to analyze the standard direct measure of the variability at different time scales, that is the power spectrum of the *spike trains* of the cells responding to the video. The power spectra for our set of cells do show an approximate $1/f$ behaviour at low frequencies, typically up to 4-8 Hz (see Fig. 7). This is completely consistent with the S+F model, in which the slow fluctuations of the activation are found to decrease in the same way, but on a much wider frequency range. The power spectra tend to level off to a constant value instead of tailing away. In fact this is just an artifact intrinsic to extracting power spectra from spike trains (in which each isolated spike is essentially a delta function with a flat Fourier transform), rather than from a variable changing continuously in time, as discussed e.g. by Bair et al. (1994). Note that both the initial $1/f$ trend and the ensuing flatness of the power spectra at high frequencies are still visible when sampling at 1000 Hz, in which case all spikes, recorded at 1 ms resolution, indeed appear as delta functions. The high frequency, almost flat portion of the spectra is therefore mainly telling us that we are looking at spikes, discrete events. The most informative part is the low frequency end, where the shape we find is interestingly different from that observed in other (and very different) experiments, for example by Bair et al. (1994). The parameters extracted from the S+F model are not affected by the fact that neuronal output is in the form of spikes. Their values are consistent with the shape of power spectra at low frequencies, but provide informative evidence over a wider frequency range.

Quantifying the efficiency of the distributions

We measured the information efficiency ϱ introduced under Methods and discussed below for our 22 video datasets, and also for the data obtained with static images in a visual fixation task (Rolls and Tovée, 1995; Rolls et al., 1997a; Rolls et al., 1997b). The important difference with the video experiment is, of course, that each time window (for the static image data we choose windows of 12, 25, 50, 100, 200 and 400 ms, starting at 100 ms after presentation of the image) corresponds to a trial with a single stimulus, and trials with the same stimulus are repeated a number of times (4-10) to obtain reliable estimates of the firing rate. In this situation, the part of the activation that is constant across trials with the same stimulus can be taken to be the *signal*, while the part that varies can be considered *noise*. The firing statistics of each cell can therefore be related to the transmission of the signal, that is information about which image was shown. Meaningful transmission of information involves populations, not single cells, but to the extent that the information in the firing rates of different cells is approximately independent, or additive, as found in the inferior temporal cortex (Rolls et al., 1997a), the information transmitted by a population of N_{cells} cells over a short time can be estimated as

$$I_{POP} \simeq I_{cell} \times N_{cells}. \tag{20}$$

Over times short with respect to the mean interspike interval the mutual information contributed by a single cell is well approximated by its time derivative (Bialek et al., 1991; Skaggs et al., 1993)

$$I_{cell}(t) \simeq t \times \frac{dI_{cell}}{dt} = t \sum_i r_i \log_2(r_i/\bar{r}) \equiv t\bar{r}\chi. \tag{21}$$

The quantity χ, defined as the ratio between the time derivative of the information and the mean firing rate of the cell, is usually called mutual information per spike, and is a measure of the part of the variability due to the signal (since each r_i is the mean firing rate response to each different signal). It is a much simpler quantity to measure than the full mutual information conveyed by each cell, because it requires only the measurement of the mean responses and not of the distribution of the responses around their means.

To check that, as indicated by formula (21), a measure of χ can replace the more complicated measure of I_{cell}, we calculated, for the data with static images, also the mutual information directly from the neuronal responses, using methods described in (Panzeri et al., 1996a; Rolls et al., 1997b) which are not based on any short time assumption. We found that the true mutual information measured in this way is in very precise agreement with that using the short time approximation (21) for all the 14 cells for times up to 25-40 ms, and for the cells with lower firing rates, for time windows up to 50-100 ms. This shows that (21) correctly quantifies the true initial rate of information transmission. For times longer than 50-100 ms the true mutual information saturates rapidly (Tovée et al., 1993; Panzeri et al., 1996a; Treves et al., 1996a), and the linear approximation implied by (21) becomes progressively less precise. But what is important is that the information per spike χ in eq. (21) is almost constant at different poststimulus times (see (Treves et al., 1996b; Tovée and Rolls, 1995)), so that the linear rate of information transfer implied by (21) can be estimated with reasonable accuracy also by measuring the mean rates r_i entering χ in longer windows. We note that the information per spike is a valid indicator of the information that single neurons provide in short times even when the independence assumption in Eq. 20 is not valid (see e.g. Rieke et al. (1996)). Fig. 8 indeed shows that, for our static image data, χ can be measured with reasonable approximation also from long windows. The information per spike for each cell varies without major monotonic trends upwards or downwards, and as an average across cells it is fairly constant. This finding establishes the validity of χ, measured also from long windows, as a quantifier of the amount of information transmitted about which static image was shown. In contrast, the entropy of the spike count distribution, which is the quantifier considered by (Levy and Baxter, 1996; Baddeley, 1996), not only has little to do with information, but also varies dramatically with the length of the window over which it is measured.

The information per spike was then compared to its theoretical maximum value among distributions of responses with the same sparseness, to derive a measure of efficiency. The maximum value is $\chi = \log_2(1/a)$, where a is the sparseness, as explained in the Methods, and the maximum is attained by a binary distribution in which the mean response $r_i = 0$ for a fraction $1 - a$ of the images. Fig. 9 (the full line correspond to the average of static image data) shows that the efficiency measure ϱ lies in the range $0.5 - 0.7$ for most of the cells and most of the windows, with only a slight upward trend for the shorter windows. In particular, with a time window of 25 ms, the average ϱ is 0.58; with a window of 50 ms, the average ϱ is 0.56; and for 100, 200 and 400 ms, it is still 0.55. The upward trend for shorter windows (the average value for 12 ms is 0.65) is partly due to the fact that when the window is short enough to include only zero or one spike, and very rarely two or more, the distribution of spike counts on each trial becomes essentially binary, and also the distribution of mean responses to each image approaches, though much more slowly, a binary shape.

Therefore the efficiency measure ϱ does indeed quantify, in a fairly stable way across time scales, the extent to which the distribution of firing rates over the chosen time interval is efficient at conveying information about which image was shown. For comparison, the alternative efficiency measure ϱ_B, which gauges the information per spike on an entropy not an information scale, and which as defined in Methods can only be measured over very short intervals, is found to be on average $\varrho_B = 0.106$ for 6 ms windows, and 0.122 for 12 ms. (This means that, for the considered time windows and set of cells, the information actually transmitted by the cells is about 9 times smaller than their entropy.)

The efficiency measure was extracted also from the video data. In contrast with the static image data, in the video experiment each time bin, whatever the window length, corresponds in principle to a different image, and each response is therefore recorded for a single trial. The measure χ quantifies, strictly speaking, the information conveyed about "which time bin" the spike count corresponds to, rather than the information about "which image" from a fixed set was shown. Still, for neurons such as the inferior temporal cortex cells, located in the ventral visual stream and presumably dedicated to analyses of visual scenes aimed at the perception of objects and individuals, it is reasonable to assume that most of the signals being coded persist for quite some time, say a fraction of a second. In this case, when using relatively long time windows such as 400 ms or 800 ms, we can consider most of the variability occurring at time scales shorter than the window as being mainly "noise", and the slower variability as being genuine "signal". For the longer of our windows we are then in a situation close to that of the static image experiment, except that we a have a single "trial" for each image, and many more images. The measure χ is not too sensitive to the number of trials per image or to the number of images available, and can be used as a quantifier of the transmission of meaningful information. For shorter windows we have to keep in mind, instead, that χ will quantify also the transmission of what is reasonably defined as noise.

Fig. 9 (dashed line) shows the average across the dataset of the efficiency ϱ extracted from a measure of χ from the video data, for all datasets and all 5 window lengths. For the two longer windows, 400 and 800 ms, the average values obtained for across datasets were 0.64 and 0.59 respectively, and for the different datasets varied in the range $0.5 - 0.7$. For progressively shorter windows ϱ increases smoothly, and at $L = 50$ ms is for many datasets around 0.9. Note that the upward trend for short windows is in part due to the fact noted above that χ captures also the transmission of noise. In part it is due to the fact that with a single "trial" per image the distribution of responses, which coincides in this case with the spike count distribution, is forced to become binary for short intervals, so that χ saturates to its maximum value.

The conclusion that we want to emphasise from these efficiency measures is the indication that when an appropriate measure of efficiency is used, the firing rate distributions we observe, both with the video data and with the static image data, are not optimal but nevertheless quite efficient (50% to 70% efficient) at conveying information quickly.

Discussion

Neurons in the cortex of the anterior part of the superior temporal sulcus of the macaque have characteristic firing rate distributions, with an approximately exponential tail during natural visual stimulation. The overall conclusion we reach at the end of the Discussion on the basis of the results is that the observed spike count (or equivalently firing rate) distribution of single neurons to natural stimuli can be largely accounted for in terms of normally distributed inputs with a "slow" component that may reflect the different stimuli, and a "fast" component that may reflect noise, to a neuron with a threshold for firing. Although this conclusion was reached with temporal lobe visual cortex cells, the same simple model and explanation could account for the shape of the firing rate distributions found in many parts of the mammalian brain (Panzeri et al., 1997).

The S+F model does not need to assume the details of real neuronal spiking dynamics. It also does not need to assume how individual synaptic inputs are summed, by just taking the activation, $h(t)$, to fluctuate around its mean value $\bar{h}$ due to a multitude of effects, at different time scales. The model does assume that both the fast and the slow fluctuations have an approximately normal distribution. This is the important simplifying assumption behind our model. It can be seen that in order for this simple situation to apply, synaptic inputs must be uncorrelated, or only weakly correlated, with the synaptic strengths with which they are weighted in the summed current. In fact, it has recently been shown by Settanni and Treves (in preparation) how to compute, for a simple feed-forward neuronal network model, the modified firing rate distribution that results from a correlation, due to associative learning, between synaptic inputs and synaptic weights. The modifications appear in any case to be small, and the uncorrelated approximation used in this paper is a good approximation in all those cases in which there is no special factor inducing a prominent correlation between a set of stimuli and the synaptic weights. In particular, the shape predicted by the S+F model might be expected to hold to a good approximation when the inputs to the cells, drawn from a large "ecological" set, are uncorrelated *with each other*, and therefore presumably also with their synaptic weights. In this situation the outputs of the population are also expected to be largely independent of each other, which tends to be the case when neurons are coding for objects in a high dimensional space, that is in a world of many objects (Rolls et al., 1997a; Rolls and Treves, 1998).

The other simplifying assumption in the S+F model is a minor one, the threshold-linear input-output transfer function. It is conceivable that using a more complicated transfer function which tries to model more realistically neuronal output could result, if amenable to treatment, in slightly better fits. This is not the point, however, since the purpose of the S+F model is precisely to show that a reasonable approximation to the observed distribution arises already from ingredients as simple as those that comprise the model. It is important to note that, due to the smoothing effect of fast noise, the effective input-output transform (10) is, unlike the instantaneous current-to-frequency transform, *supralinear* in a range around threshold, as shown in Fig. 1. It is this supralinearity which tends to convert the Gaussian tail of the activation distribution into an exponential tail of the firing rate distribution, without producing a full exponential distribution with a mode at zero. We note that the effect of fast noise in conjunction with a threshold-linear activation function, tends to produce the characteristic shape of the lower part of a sigmoid activation function. Further, there is no hint, in the observed distributions, of the need for a rounding or saturation in the upper portion of the transfer function. In fact a fully sigmoidal transfer function, with a saturation level within the normal firing range, would tend to produce a sharper cut in the tail of the spike count distributions, quite unlike the long exponential tails observed. This implies that the neurons are operating below saturation at a very high level of firing, and this is consistent with the fact that inferior temporal cortex neurons rarely fire above 100 spikes/s.

Are the fits really adequate?

The S+F model did in general fit the data well, and much better than any other model (see Table 1). The 15% of cases (for the video data) in which the fit was not good at $P < 1\%$ is mainly related to the simplicity of the model. The fits would be even better if single time scales were fitted on their own. Using the simultaneous fits, however, allows us to assign an understandable meaning to the mean activation h_0, which otherwise would be a "dummy" fit parameter that might be manipulated in order to overfit (there would be a different, meaningless h_0 at each scale). The fits could be improved by making the model less simple in some of the following ways: i) The threshold-linear transform is oversimplified, especially when applied to several different time scales at once. ii) The partition into fast (effectively instantaneous) and slow (very long lasting) fluctuations is artificial and certainly too radical. iii) Neglecting the spiking nature of neuronal outputs is likely to cause distortions, especially at the shortest windows and in those histogram bins with zero or very few spikes. iv) The conversion of the continuous probability density into a model histogram of spike counts is only a rough approximation.

All these sources of inaccuracy were nevertheless accepted in order to keep our formulation of the basic hypothesis, the S+F model, as simple and intuitive as possible. If a proposal comes forth for an elegant way to remove any of these inaccuracies, it will be interesting to see whether this produces a significant improvement in the fits.

Entropy is no substitute for information

The literature on the efficient coding of natural images, inspired by the work by Barlow (1961), see also Barlow (1989), is typically limited to a discussion of *entropy*. In particular, Shannon's theorem about exponential distributions, invoked by Levy and Baxter (1996) and Baddeley (1996), is only about maximizing their entropy. It is obvious that to understand the efficiency with which firing rates might encode information, one has to refer to what is being coded, that is to use measures of *mutual information*. Ultimately, estimating mutual information requires the repetition of the same visual stimuli over many trials, which takes the experiment somewhat beyond the realm of "natural", or ecological, situations. The necessity for multiple trials is easily understood by remembering that entropy is just a measure of total variability, while mutual information effectively subtracts from the total variability the variability still present when the message being transmitted is kept fixed. It can in fact be calculated as the entropy minus the average conditional entropy. Whether the observed distributions have nearly *minimal* (Olshausen and Field, 1996) or *maximal* entropy (Baddeley, 1996) under the appropriate choice of constraints, may not be very relevant to their information efficiency. Entropy is not an appropriate measure irrespective of whether it is calculated from the spike count or, more sensibly, by binning firing rates in bins whose width is dependent on the variability of the rate itself. As our fits suggest, it is likely that the origin of the distribution is much simpler than requiring a complex process to obtain a fully exponential distribution. Further, as indicated in the Results, the information transmitted about the static visual stimuli may be as much as 9 times smaller than the entropy of the firing (which defines the maximal information that could be transmitted), and therefore we conclude that the precise shape of the firing rate distribution does not appear to be accounted for by the need to maximize the information transmission rate.

When analysing the efficiency of the observed distributions we should not only discard entropy in favour of information, but also consider that the information important for brain processing is that conveyed by *populations* of cells, not by single cells. We should consider, then an indicator of the information provided by populations of cells, that can be extracted from recordings of single cells, as discussed next.

Measures of information efficiency

We have used the information per spike (Bialek et al., 1991; Skaggs et al., 1993) as a simple single-cell measure of transmitted information, and we regard it as indicative of the information transmitted by a population, at least in the context of object and face coding by IT cells (Rolls et al., 1997a). The results of measuring the information per spike from the static image data indicate that it is valid to extract this measure even from long windows, even though the responses of these temporal cortex cells are not constant over the window (Rolls and Tovée, 1995), and strictly speaking the notion that this is the average "information in one spike" does not apply to long windows that can contain many spikes. Dividing the information per spike by its maximum possible value (attained, among distributions of fixed sparseness, by a binary distribution) we have obtained the measure of efficiency ϱ. The constraint on the sparseness is a meaningful one, in the light of the importance of this parameter in determining e.g. memory capacity (Treves and Rolls, 1991). This is therefore a relevant measure of information efficiency, although other measures may also be useful to help understand to what extent the actual distribution of firing rates found could encode information efficiently. A different measure was used by Bialek et al. (1991) and Rieke et al. (1996), where essentially the denominator in Eq. 16 is the entropy per spike instead of the maximum information per spike. Both are obviously acceptable definitions, with the latter, ϱ_B, describing more the extent to which the cell uses its biophysical potential, and ours, ϱ, that to which it exploits its information processing capacity (at fixed sparseness). For our static image data ϱ_B turns out to be fairly low, around 0.1, while ϱ is in the range 0.5 to 0.7.

Returning to the video data, ϱ can still be taken as a measure of information efficiency, provided that the variability in the firing of a cell that occurs on time scales shorter than the sampling window is considered to be noise, while that on slower time scales is regarded to carry, to a great extent, the signal. This assumption is somewhat arbitrary, but one should note that ultimately, what is signal and what is noise is indeed an arbitrary decision dependent on what one wants to analyze. When watching a video that includes faces in motion, the head motion may be regarded as noise in a system that analyzes face identity, and as signal in a system that tries to measure head motion. In any case, over reasonably long windows ϱ turns out to be, also for the video data, in the $0.5 - 0.7$ range.

The use of the ϱ measure of efficiency indicates then that the distributions found are not maximally efficient, as was suggested for exponential distributions (Levy and Baxter, 1996; Baddeley, 1996), using an argument which would not in any case apply when information is transmitted in a noisy way (noisy in the sense that the number of spikes in the window is not fixed given the stimulus). Instead, our analysis indicate that the efficiency, on at least the ϱ scale, is intermediate, with typical values of $0.5 - 0.7$.

Variability at different time scales

Analyses of the spiking statistics of cortical neurons in behaving animals typically show large variability, as quantified e.g. by the coefficient of variation (standard deviation over mean) of interspike intervals. Softky and Koch (1993) have pointed out that the observed variability is inconsistent with the view that cortical neurons can be reduced to integrate-and-fire oscillators receiving a barrage of uncorrelated stochastic inputs from other cells firing at slowly varying or quasi-stationary mean rates. They are thereby led to suggest that synchronized inputs that occur without any preset periodicity are important in causing cortical neurons to fire. A lot of attention has been devoted to checking the validity of the assumption that neurons can be modelled as integrate-and-fire units, and of the assumption that afferent inputs are really uncorrelated. Less attention has been devoted to the assumption that afferent inputs are from cells firing at quasi-stationary rates (or at least varying more slowly than the time scale of the analysis) (Bair et al., 1994; vanVreeswijk and Sompolinsky, 1996).

Our results do not address directly the issue of the variability in the synaptic inputs to the cell being recorded, but, through our model fitted to the real data, do address the variability of the activation h, which may be taken to reflect the underlying variability of the inputs the cell integrates. If the model is essentially valid, it yields an estimate of the proportion of the variability in h which occurs at time scales faster than any given analysis window, as shown in Fig. 6. This appears to be a sizable proportion, for our datasets. What it indicates is that, even in experiments in which external correlates, such as visual images, are stationary or slowly varying, one should be open to the possibility that many neuronal signals may occur much more rapidly, and need not be synchronized to contribute to the variability in the spiking statistics of the receiving neurons. Thus the results in this paper challenge the view that a precise synchronization of the inputs to a neuron is necessary to account for its firing rate distribution (Softky, 1994).

With respect to the distribution of variance at different time scales, our results are indirect, being based on the estimate of the fit parameters $\sigma_F(L)$ and $\sigma_S(L)$, but are broadly compatible with a simple $1/f$ distribution as a function of frequency. While we do not want to attach any special significance to this particular distribution, which has elicited considerable theoretical interest, we note that even the general trend towards it is obscured if one measures instead the standard quantity, the power spectrum of the raster plot. This is because the directly observable events that enter the power spectrum, the spikes, distort by their all or none nature the time course of the underlying variable, the activation current. Using the fit procedure to the S+F model is one way to circumvent this distortion.

Conclusion

This analysis accounts for the observed spike count distributions of single cells under ecological conditions in terms of normal distributions of neuronal activations. A normal distribution holds, of course, for any variable that can be thought of as the sum of many unrelated terms, none of which is dominant. Here the variable h, the total current into the soma, is the approximately linear (Rall and Segev, 1987) sum of many synaptic inputs multiplied by the corresponding weights. Our model implies the assumption that individual synaptic inputs are not correlated with their synaptic weights. That assumption is not, however, critical. In fact, analytical work on the modification of the model distribution brought about by reasonable degrees of correlation between inputs and weights shows that such modifications are very minor (Settanni and Treves, in preparation).

The conclusion we make is that the reasonable fit between the S+F model and the spike count distributions of inferior temporal cortex cells is consistent with the possibility that there is no special optimization principle or purpose to the firing rate distributions found. We note that such principles *might* include information efficiency, minimum or maximum entropy, synchrony, speed, stationarity, or the formation of cell assemblies. Evidence for the validity of those principles has to be found in other types of analyses, while the analysis of spike count histograms of cortical cells responding to natural stimuli has nothing to say in favour of any of them. However, we noted above that the information transmitted about the static visual stimuli may be as much as 9 times smaller than the entropy of the firing (which defines the maximal information that could be transmitted), and therefore we are led to conclude that the precise shape of the firing rate distribution does not appear to be accounted for by anything related to the need to maximize the information transmission rate.

With the new efficiency measure introduced here, we show that temporal cortex visual neurons responding to large sets of static (Rolls and Tovée, 1995; Rolls et al., 1997a; Rolls et al., 1997b) and dynamic visual stimuli are able to encode information quite efficiently, with an efficiency in the order or 0.5 - 0.7. However, the fact that the efficiency is not close to 1 (optimal) again indicates that the precise form of the firing rate distribution found is probably not produced simply in order to maximize the efficiency of information transmission.

Acknowledgements

We are grateful to Roland Baddeley and Bill Bialek for interesting discussions. Partial support was from the Medical Research Council PG8513790; the cooperation between Oxford and SISSA was funded by the Human Capital and Mobility Programme of the European Community. S. Panzeri is supported by a EC Marie Curie Research Training Grant ERBFMBICT972749. M.Booth is supported by a Wellcome Trust research studentship.

References

Abeles, M., Vaadia, E. and Bergman, H. (1990). Firing patterns of single units in the prefrontal cortex and neural network models, *Network* **1**: 13–25.

Amit, D. J. and Tsodyks, M. V. (1991). Quantitative studies of attractor neural networks retrieving at low spikes rates: I. substrate - spikes, rates and neuronal gain, *Network* **2**: 259–273.

Baddeley, R. J. (1996). An efficient code in v1?, *Nature* **381**: 560–561.

Baddeley, R. J., Abbott, L. F., Booth, M., Sengpiel, F., Freeman, T., Wakeman, E. A. and Rolls, E. T. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes, *Proc. R. Soc. Lon. Ser. B* **264**: 1775–1783.

Bair, W., Koch, C., Newsome, W. and Britten, K. (1994). Power spectrum analysis of bursting cells in area mt in the behaving monkey, *J. Neurosci.* **14**: 2870–2892.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages, *in* W. Rosenblith (ed.), *Sensory Communication*, MIT Press, Cambridge, pp. 217–234.

Barlow, H. B. (1989). Unsupervised learning, *Neural Comp.* **1**: 295–311.

Barnes, C. A., McNaughton, B. L., Mizumori, S. J. Y., Leonard, B. W. and Lin, L.-H. (1990). Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing, *in* J. Storm-Mathisen, J. Zimmer and O. P. Ottersen (eds), *Understanding the brain through the hippocampus*, Vol. 83 of *Progress in Brain Research*, Elsevier Science, chapter 21, pp. 287–300.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R. and Warland, D. (1991). Reading a neural code, *Science* **252**: 1854–1857.

Dong, D. W. and Atick, J. J. (1995). Statistics of natural time-varying images, *Network* **6**: 345–358.

Fisher, R. A. and Yates, F. (1963). *Statistical tables: for biological, agricultural and medical research*, Longman.

Koch, C., Bernander, O. and Douglas, R. (1995). Do neurons have a voltage or a current threshold for action potential initiation?, *J. Comp. Neurosci.* **2**: 63–82.

Lanthorn, T., Storm, J. and Andersen, P. (1984). Current-to-frequency transduction in ca1 hippocampal pyramidal cells: slow prepotentials dominate the primary range firing, *Exp. Brain Res.* **53**: 431–443.

Levy, W. B. and Baxter, R. A. (1996). Energy efficient neural codes, *Neural Comp.* **8**: 531–543.

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images, *Nature* **381**: 607–609.

Panzeri, S., Biella, G., Rolls, E. T., Skaggs, W. E. and Treves, A. (1996a). Speed, noise, information and the graded nature of neuronal responses, *Network* **7**: 365–370.

Panzeri, S., Booth, M., Wakeman, E. A., Rolls, E. T. and Treves, A. (1996b). Do firing rate distributions reflect anything beyond just chance?, *Society for Neuroscience abstracts* **22**: 1124.

Panzeri, S., Rolls, E. T., Treves, A., Robertson, R. G. and Georges-Francois, P. (1997). Efficient encoding by the firing of hippocampal spatial view cells, *Society for Neuroscience abstracts* **23**: 195.4.

Press, W. M., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical recipes in C*, Cambridge University Press, Cambridge.

Rall, W. and Segev, I. (1987). Functional possibilities for synapses on dendrites and dendritic spines, *in* G. Edelman and J. Cowan (eds), *Synaptic Function*, John Wiley, New York, pp. 605–636.

Rieke, F., Warland, D., de Ruyter van Steveninck, R. R. and Bialek, W. (1996). *Spikes: exploring the neural code*, MIT Press, Cambridge MA.

Rolls, E. T. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces, *Human Neurobiology* **3**: 209–222.

Rolls, E. T. and Tovée, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex, *J. Neurophysiol.* **73**: 713–726.

Rolls, E. T. and Treves, A. (1998). *Neural Networks and Brain Function*, Oxford University Press, Oxford.

Rolls, E. T., Treves, A. and Tovée, M. J. (1997a). The representational capacity of the distributed encoding of information provided by populations of neurons in the primate temporal visual cortex, *Exp. Brain Res.* **114**: 149–162.

Rolls, E. T., Treves, A., Tovée, M. J. and Panzeri, S. (1997b). Information in the neuronal representation of individual stimuli in the primate temporal visual cortex, *J. Comp. Neurosci.* **4**: 309–333.

Shannon, C. E. (1948). A mathematical theory of communication, *AT&T Bell Labs. Tech. J.* **27**: 379–423.

Skaggs, W. E., McNaughton, B. L., Gothard, K. and Markus, E. (1993). An information theoretic approach to deciphering the hippocampal code, *in* S. Hanson, J. Cowan and C. Giles (eds), *Advances in Neural Information Processing Systems*, Vol. 5, Morgan Kaufmann, San Mateo, pp. 1030–1037.

Softky, W. (1994). Submillisecond coincidence detection in active dendritic trees, *Neuroscience* **58**: 13–41.

Softky, W. and Koch, C. (1993). The highly irregular firing of cortical cells is incosistent with temporal integration of random epsp's, *J. Neurosci.* **13**: 334–350.

Tovée, M. J. and Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex, *Visual Cognition* **2**: 35–58.

Tovée, M. J., Rolls, E. T., Treves, A. and Bellis, R. J. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex, *J. Neurophysiol.* **70**: 640–654.

Treves, A. (1993). Mean-field analysis of neuronal spike dynamics, *Network* **4**: 259–284.

Treves, A., Barnes, C. A. and Rolls, E. T. (1996a). Quantitative analysis of network models and of hippocampal data, *in* T. Ono, B. L. McNaughton, S. Molotchnikoff, E. T. Rolls and H. Nishijo (eds), *Perception, Memory and Emotion: Frontier in Neuroscience*, Elsevier, Oxford, pp. 567–579.

Treves, A. and Rolls, E. T. (1991). What determines the capacity of autoassociative memories in the brain, *Network* **2**: 371–397.

Treves, A. and Rolls, E. T. (1992). Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network, *Hippocampus* **2**: 189–200.

Treves, A., Skaggs, W. E. and Barnes, C. A. (1996b). How much of the hippocampus can be explained by functional constraints?, *Hippocampus* **6**: 666–674.

vanVreeswijk, C. and Sompolinsky, H. (1996). Chaos in neural networks with balanced excitatory and inhibitory activity, *Science* **274**: 1724–1726.

Probability Model	Rejections ($P < 0.01$) (video data)	Rejections ($P < 0.01$) (static stimuli)
Exponential	83.6 % (92/110)	75 % (42/56)
Poisson	100 % (110/110)	100 % (56/56)
Gaussian	96.3 % (106/110)	89.2 % (50/56)
S+F Model	15.4 % (17/110)	1.8 % (1/56)

Table 1: Percentage of rejections ($P < 0.01$) of the fits to the various models.

Figure caption

Fig. 1: Activation functions in the model. The solid line (——) is the instantaneous threshold linear activation function (with $g = 1$). The dashed line (– – – –) is an example of an "effective" activation function obtained integrating over fast noise, using eq. 10, with $\sigma_F = 50Hz$. The chain line (— · —) is another example of effective activation function ,eq. 10, obtained with $\sigma_F = 10Hz$.

Fig. 2: Probability histograms for 9 datasets (video data). The time window is 100 ms long. Each graph plots the probability on the y axis of the number of spikes, in a 100 ms time bin, given in the x axis. The histogram is the neurophysiological data from the neuron, with the standard deviation shown; the dashed line shows an exponential distribution fitted to the mean firing rate of the cell; and the solid line shows the fit of the S+F model described in the text.

Fig. 3: probability histograms for 9 datasets (video data). The time window is 400 ms long. Conventions as in Fig. 2.

Fig. 4: probability histograms for 3 cells (responding to static stimuli). **(a):** The time window is 100 ms long. **(b):** The time window is 400 ms long. Conventions as in Fig. 2.

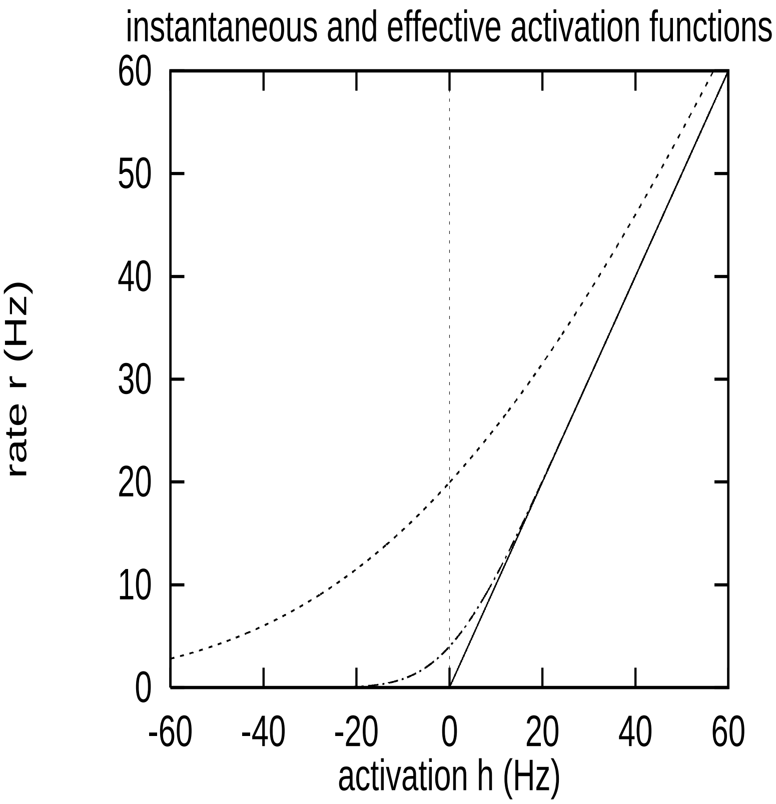
Fig. 5: Ordinate: The standard deviation of the fast noise, $\sigma_F(L)$. Abscissa: The standard deviation of the slow noise, $\sigma_S(L)$. Units in Hz (video data).

Fig. 6: Ordinate: The fraction of the variability that is due to fast noise, $F/(S + F)$. Abscissa: The time window on a log scale. Dashed lines represent single datasets; the bold line is the average across datasets (video data).

Fig. 7: The average across datasets of the normalized power spectrum of the spike train for each recording session. The sampling frequency is 200 Hz (video data).

Fig. 8: The average across the population (+/- s.d.) of the information per spike $\chi(L)$ for 14 cells tested with 65 static images of natural scenes of faces plotted as a function of the time bin length L. The time axis is on a log scale.

Fig. 9: The encoding efficiencies $\varrho(L)$ for 14 cells tested with 65 static images of natural scenes of faces plotted as a function of the time bin length L. (solid line; +/- s.d.). The encoding efficiencies $\varrho(L)$ for all 22 datasets recorded with the video (dashed line; +/- s.d.). The time axis is on a log scale.



instantaneous and effective activation functions
rate r (Hz)
60
50
40
30
20
10
0
-60
-40
-20
0
20
40
60
activation h (Hz)

Figure 1:

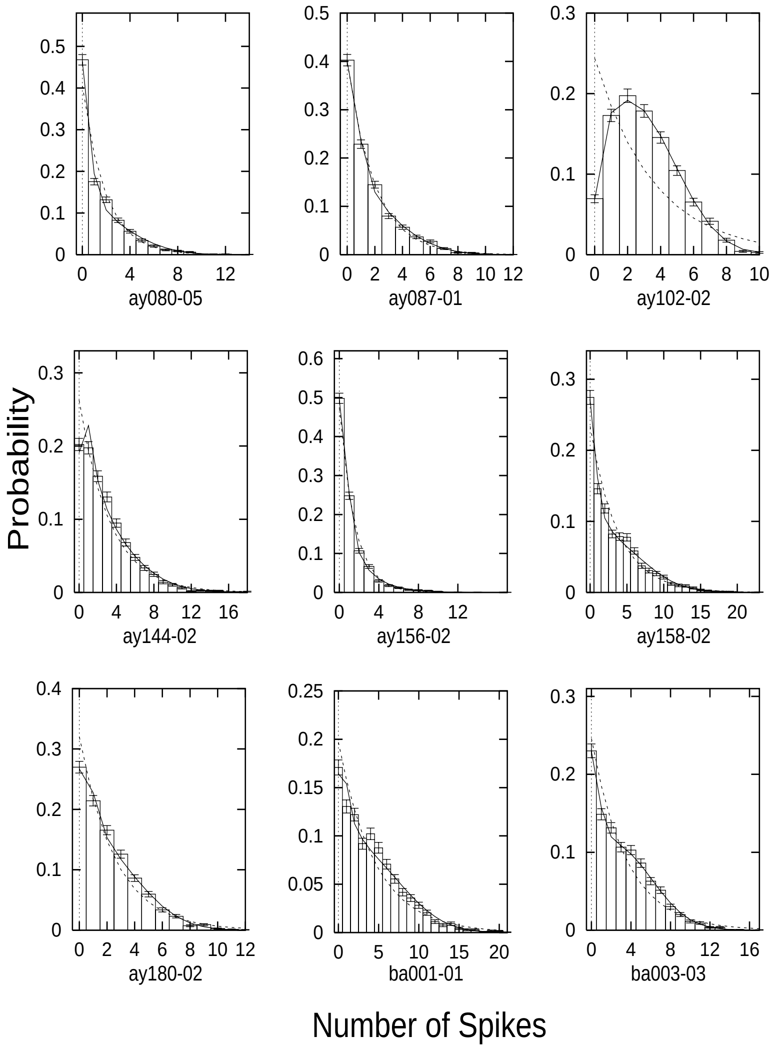
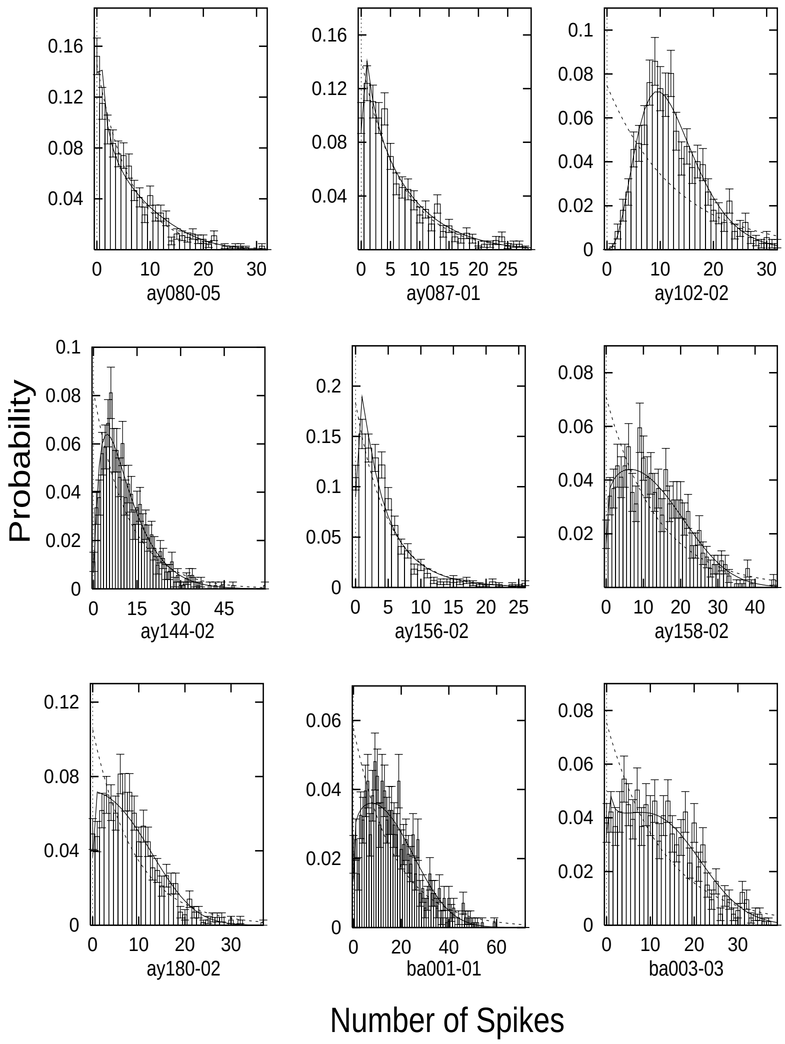
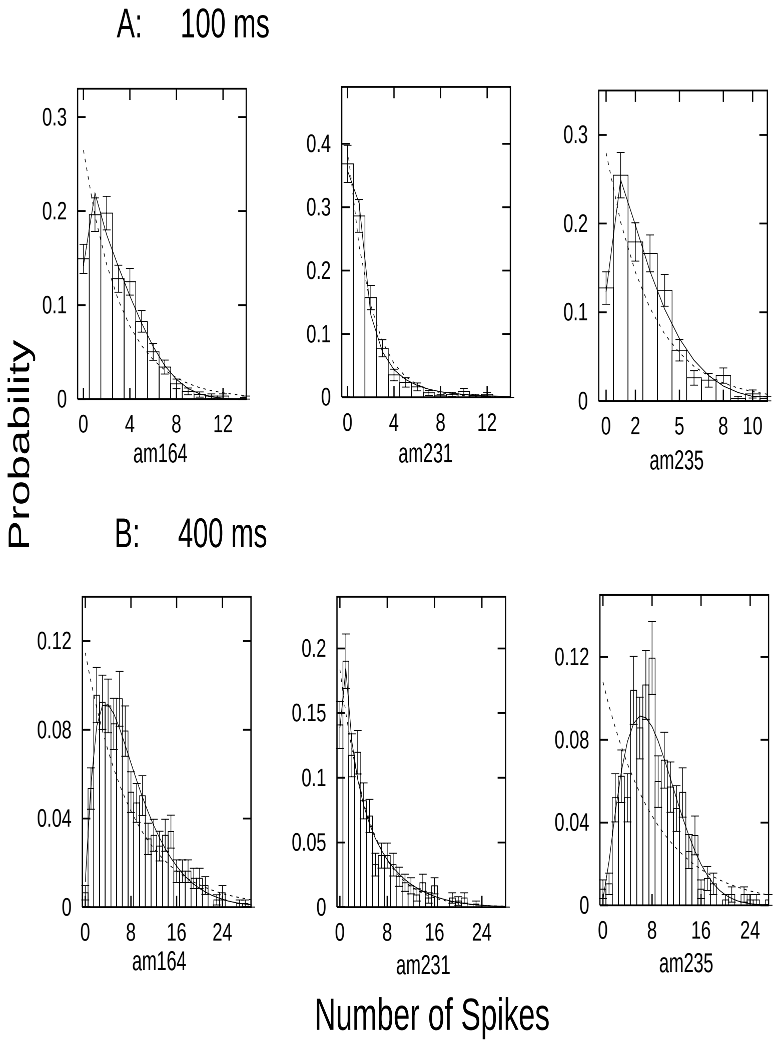

Probability
ay080-05
ay087-01
ay102-02
ay144-02
ay156-02
ay158-02
ay180-02
ba001-01
ba003-03
Number of Spikes

Figure 2:



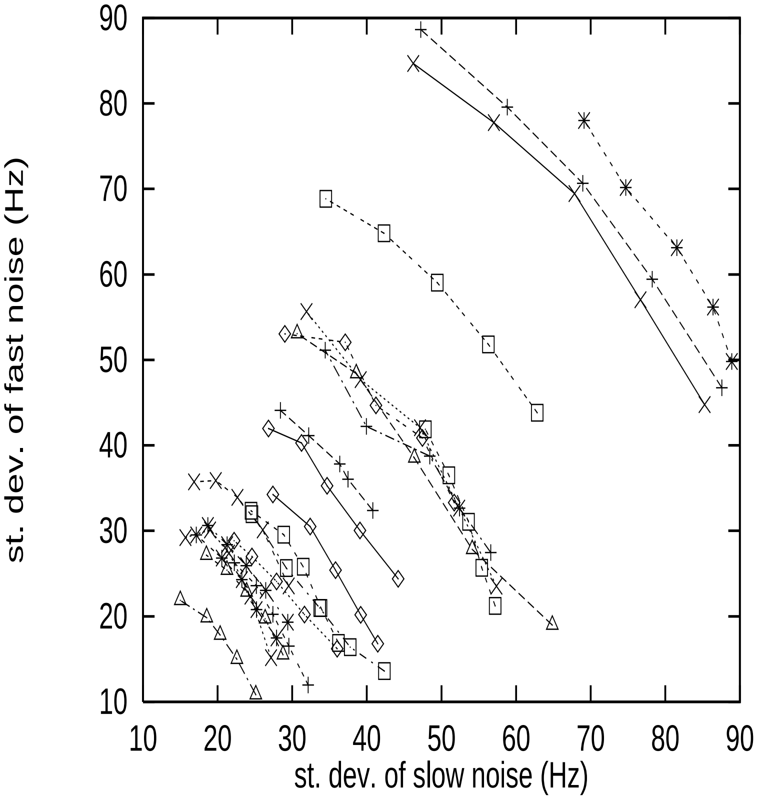
0.16
0.12
0.08
0.04
0
10
20
30
ay080-05
0.16
0.12
0.08
0.04
0
5
10
15
20
25
ay087-01
0.1
0.08
0.06
0.04
0.02
0
0
10
20
30
ay102-02
Probability
0.1
0.08
0.06
0.04
0.02
0
0
15
30
45
ay144-02
0.2
0.15
0.1
0.05
0
0
5
10
15
20
25
ay156-02
0.08
0.06
0.04
0.02
0
10
20
30
40
ay158-02
0.12
0.08
0.04
0
0
10
20
30
ay180-02
0.06
0.04
0.02
0
0
20
40
60
ba001-01
0.08
0.06
0.04
0.02
0
0
10
20
30
ba003-03
Number of Spikes

Figure 3:



A: 100 ms
B: 400 ms
Probability
am164
am231
am235
Number of Spikes

Figure 4:



st. dev. of fast noise (Hz)
st. dev. of slow noise (Hz)
10
20
30
40
50
60
70
80
90

Figure 5:

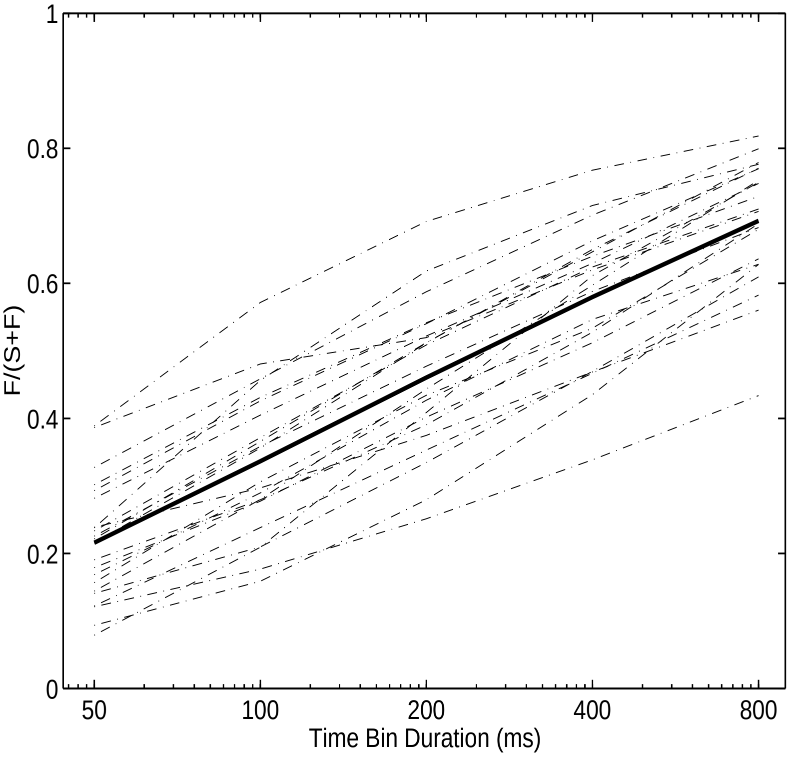


Figure 6:

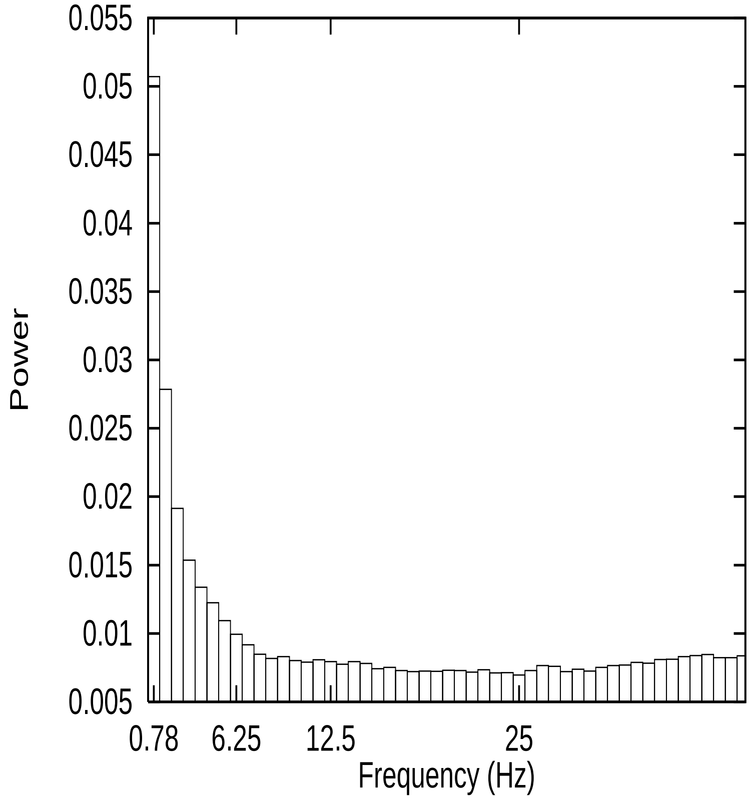

0.055
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
Power
0.78
6.25
12.5
25
Frequency (Hz)

Figure 7:

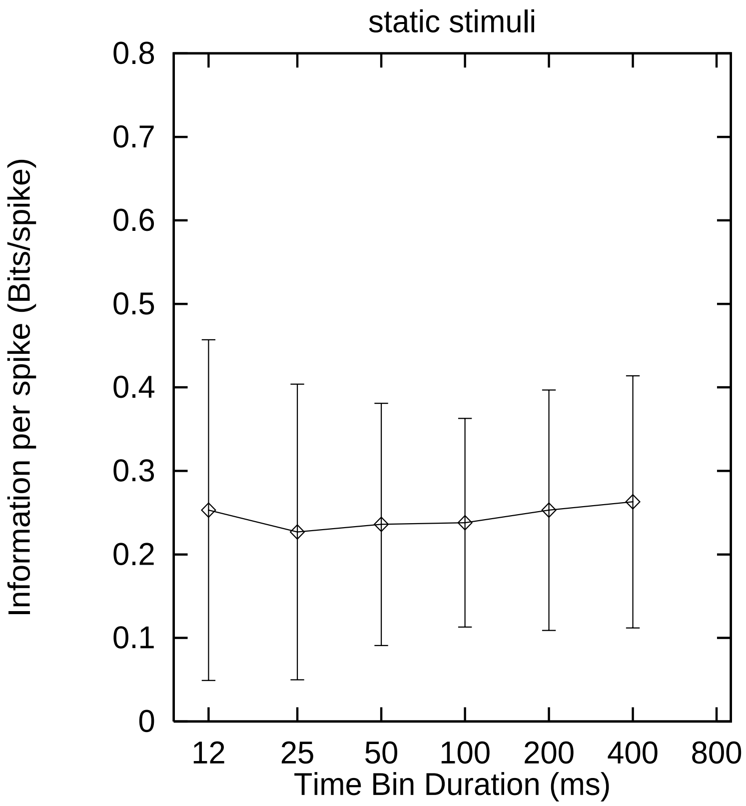

static stimuli
Information per spike (Bits/spike)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
12
25
50
100
200
400
800
Time Bin Duration (ms)

Figure 8:

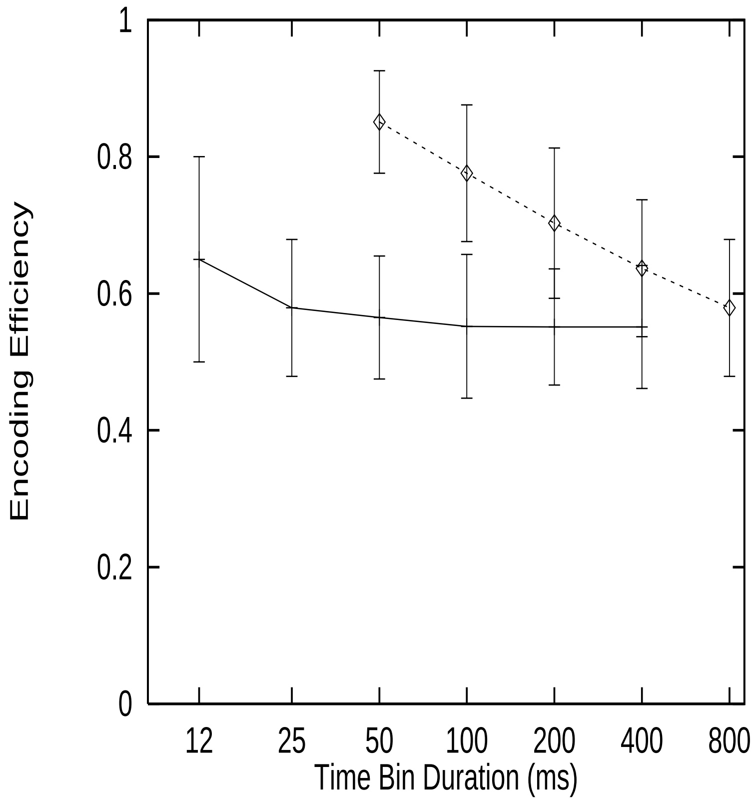

Encoding Efficiency
1
0.8
0.6
0.4
0.2
0
12
25
50
100
200
400
800
Time Bin Duration (ms)

Figure 9: