Exp Brain Res (2007) 176:1–11
DOI 10.1007/s00221-006-0594-4

RESEARCH ARTICLE

Joint decoding of visual stimuli by IT neurons' spike counts is not improved by simultaneous recording

Britt Anderson · Mark I. Sanderson · David L. Sheinberg

Received: 2 December 2005 / Accepted: 29 May 2006 / Published online: 28 July 2006

Abstract Information about visual stimuli such as objects and faces is represented across populations of neurons of the inferior temporal cortex. Does recording from inferotemporal neurons simultaneously tell you more than recording from them sequentially? Equivalently, are neurons conditionally independent given a stimulus? To evaluate these issues, we recorded from two monkeys during a passive viewing task. Multiple neurons were simultaneously recorded on separate electrodes. From spike counts in 50-ms windows, we computed the mutual information between counts and images for each neuron individually and jointly with other simultaneously recorded neurons. To determine the significance of these values, we shuffled the stimulus labels (to test if there was significant information) or shuffled responses across trials involving the same image (to see if there was synergistic coding). We recorded from 127 pairs of neurons where each neuron individually was visually responsive. Depending on the time window, we found up to ~ 90% of these pairs showed significant information about the visual stimulus. Shuffling across trials failed to show evidence for synergistic coding. In summary, if you were given two of our neuronal responses and asked to guess the stimulus which produced them you could not, in principle, do better with two simultaneously recorded spike counts than with any two spike counts selected randomly from trials of the same type.

Keywords Vision · Neurons · Temporal lobe · Action potentials

Supported by: NIH R01-EY014681, NSF CRCNS 0423031, and the James S. McDonnell Foundation. We appreciate the comments of Dr. Elie Bienenstock.

B. Anderson
Brain Science Program, Brown University,
Box 1953, Providence, RI 02912, USA

M. I. Sanderson · D. L. Sheinberg (✉)
Department of Neuroscience, Brown University,
Box 1953, Providence, RI 02912, USA
e-mail: David_Sheinberg@brown.edu

Introduction

Higher order primate visual neurons respond selectively to complex images (Tanaka 1997) and, individually, their firing rate profiles are informative (Tovee et al. 1993). While response profiles show an overall consistency there is trial-to-trial response variation. Is there message in this ''noise?'' Figure 1 demonstrates the experimental issues.

We can see that one neuron responds most vigorously to the harp and another to the pear, but we can also see that there is clear variability on a trial-by-trial basis. Do downstream neurons take functional advantage of this variability or is it an impediment to stimulus decoding? For the image in the lower left, the maracas, the responses of the two neurons to this image, if any, are smaller in magnitude. Perhaps joint codes, where we consider the pair of neurons as one response, are particularly helpful for disambiguating these types of stimulus-response pairings.

One direct way to address this issue is to compare the information about a stimulus available from collections of neural responses that were recorded simultaneously to collections that were not recorded simultaneously. This direct attack requires that we

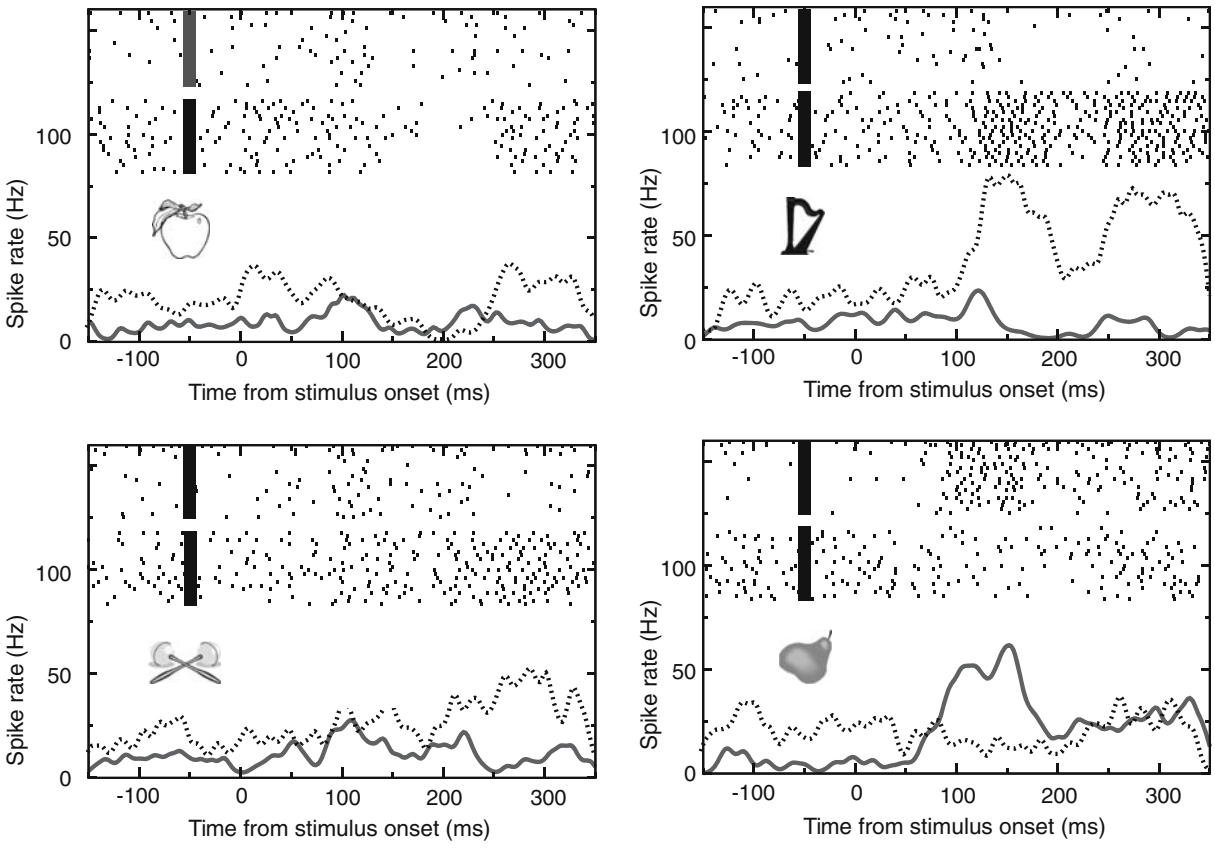

Fig. 1 The rasters and estimated spike density functions are shown for two simultaneously recorded inferotemporal (*IT*) neurons. The *small insets* show a low resolution version of the visual stimulus. Each row (demarcated by a *thick horizontal line*) represents a single trial. The two sets of *tick marks* in rows along the top of each figure are the times of individual spikes for each of the two neurons whose estimated spike density functions are shown below. Note that these two neurons consistently increase their firing rates to two different images (and decrease their background rate to a third). For these stimuli we could probably make a good guess about the stimulus regardless of whether the spike train pair we were given came from the same trial as long as they were of the same image, but what about the maracas (*lower left image*)? For this image any changes in firing are subtler, and it may be that for stimuli and spike responses of this magnitude the correlation between neurons would be helpful for decoding

identify a relevant stimulus for the neurons from which we record and that we record simultaneously from multiple cells with some response to the visual stimuli. As there are many possible choice of visual stimuli available for driving visual neurons (e.g., faces, hands, man-made objects, natural objects, geometric shapes, oriented lines, gabor patches) and many ways of depicting them (real objects, perspectival rendering, color, grayscale, line drawings, static, animated, etc.), the actual visual stimuli chosen gives concrete examples of an investigator's biases about the types of stimuli for which joint coding is important.

In the early visual system, including the retina, thalamus, and primary visual cortex, specific visual stimuli generally reproduce a robust and consistent neural response. However, further along the visual pathways, the neural responses vary with the perceptual state (Sheinberg and Logothetis 1997; Kreiman et al. 2002). It is here, we need to explain varying neuronal responses and perceptions to the same retinal stimulus where the coding across neuronal responses might prove most useful.

Because damage to inferotemporal (IT) structures of the monkey impairs visual encoding and decoding of objects (Gross 1994) this is an anatomical site where interactions among neurons would be expected to carry visual stimulus information. Gawne and Richmond (1993) were among the first to attack this issue and used pairs of neurons' recorded from of a single electrode. Neuronal responses to simple Walsh pattern were filtered and then pooled for a principal components analysis. Responses were then reconstructed from these components. The authors concluded that most of the information carried by these spatially proximate IT neurons about the viewed Walsh patterns was largely redundant. Rolls and

colleagues (Rolls et al. 2003a; Franco et al. 2004) computed the information between neural firing patterns and visual images with different computational methods. In their experiments, monkeys passively viewed a small set of visual images (objects, faces, and shapes) while two to four neurons were recorded from an array of extracellular recording electrodes in monkey area TE. Because most of the neuron pairs came from separate electrodes, the neurons were separated by a greater distance than in the work of Gawne and Richmond (1993). Both papers report that the component of the neuron response that carried most of the information about the stimulus was in the spike rates and not stimulus dependent correlations in firing. Further, they found little redundancy among the neurons in this population. They did not specifically address whether the response were independent given the stimulus.

We recorded from multiple visually responsive neurons of monkey IT while the animal passively viewed complex images known to increase the firing rates of cells in this region. To determine if spike count relations between cells, on a stimulus specific basis, are important for decoding visual stimuli we assessed the mutual information between the stimulus and spike counts from relatively short (50 ms) windows. We determined the statistical significance of this result by comparing it to estimates of the mutual information after shuffling the stimulus labels or after shuffling the spike responses across trials.

Methods

Subjects and behavioral task

The subjects were two male macaque monkeys, E (10.2 kg) and J (9.2 kg) housed and trained in accordance with the policies and procedures set forth in the U.S. Public Health Service Policy on Humane Care and Use of Laboratory Animals and the National Institutes of Health Guide for the Care and Use of Laboratory Animals, as adopted by the Society for Neuroscience in its policy on the use of animals in Neuroscience research. All experiments were approved by the Institutional Animal Care and Use Committee.

Both monkeys had titanium head posts for maintaining head restraint. At a separate surgery both monkeys were fitted with titanium recording chambers (E left side, J both sides). The combined duration of daily test sessions was between 1 and 3 h. Both a Robinson style eye coil system (CNC Engineering; used for J), and an ISCAN RK-726PCI video eye tracking system (ISCAN, Inc., Burlington, MA; used for E) were used to monitor eye movements. Both animals are still alive and therefore more precise anatomical data on recording sites is not available.

The data for our analysis was taken from physiological recording sessions that met the following criteria: at least one recorded cell showed clear visual modulation, at least two neurons had been isolated on separate electrodes (recorded simultaneously), and data were available for a minimum of ten trials for each visual image (range 10–50 repetitions/stimulus). Stimulus sets were typically selected to provide a range of response from all isolated cells. After this screening procedure, we used objective criteria for selecting pairs of visual neurons for subsequent analyses (for details on selections procedure, see Data analysis).

The basic behavioral paradigm was a passive viewing task. Seated in a primate chair, head fixed, the monkey viewed a computer monitor that was approximately 120 cm before him with a screen refresh rate of 100 Hz. At the beginning of the trial, a yellow fix spot (0.3° visual angle) would appear centered on the screen. The monkey had to acquire and hold fixation for 500 ms before the fix spot would disappear. Then, a series of images would be presented *sequentially* at the center of the screen. For different experimental sessions, the stimuli were either 2, 4, or 6° in size. After the last image the fix spot reappeared 6° from center at one of the four randomly selected locations and the monkey had to move his eyes to the fix spot to obtain a fluid reward. During the target presentation intervals the monkeys were allowed to inspect the images, so as not to drastically constrain their natural inclination to inspect the objects. Both animals were also participants in search related studies, so we did not want to overtrain them to maintain tight fixation at the expense of active inspection. For our data analysis, however, we discarded any trial for which the animals' eye position was not directed at the image when it appeared (less than 2% trials were removed). All trials, therefore, included stimuli that appeared within the large receptive field of an IT cell (Rolls et al. 2003b).

The number of images shown sequentially on a trial, and the time each image remained on the screen varied across experimental sessions, but was consistent within a block of trials (one to six images per trial, 300 or 600 ms duration per image). There were generally 50–100 trials per experimental session. Images were photographs of objects or clip art pictures of musical instruments, toys, and fruits.

Electrophysiology

For measuring neuronal responses, we used the Thomas Recording multi-channel system with a five channel head (Thomas Recording GmbH, Giessen, Germany). In brief, this system employs a five channel microdrive with integrated preamplifiers (26 dB gain). Each of the five quartz coated tungsten/platinum fiber microelectrodes (80 μm diameter quartz, 20 μm diameter platinum/tungsten core) is capable of independent advancement. Signals are subsequently amplified and filtered for the simultaneous recording of both action potentials (500 Hz to 8 kHz) and local field potential (1–100 Hz) (Mountcastle et al. 1991; Eckhorn and Thomas 1993).

Analog signals for spike analyses were sampled at 25 kHz and raw traces were displayed on-line along with on-line automated estimates of spikes as an aid to the investigator. All data from a trial were saved for off line analysis and resorted prior to these analyses. Spike sorting was predominantly manual. Using locally written software, we visually inspected tracings to identify channels with well isolated spikes. We then set amplitude thresholds and limits to select waveform samples. We sometimes employed a *k*-means clustering algorithm on these waveform samples. *k*-Means clustering divides a data set into "*k*" groups that produce the minimum distance to the groups' centers (Lewicki 1998). For spike sorting, the spikes are treated as a vector in a shape space and the spikes sorted into groups based on similarity in shape. For these data, we primarily used this method as an efficient way to remove certain non-action potential artifacts.

Data analysis

We screened all neurons with at least 100 spikes in a data set for visual responsiveness and selectivity. We operationalized visual responsiveness by comparing the number of spikes recorded from a neuron in the period from 80 to 230 ms after stimulus onset (for all stimuli collectively) to the number of spikes in the time period from −150 to 0 ms before stimulus onset (Wilcoxon rank sum test). We operationalized visual selectivity by comparing the number of spikes to different stimuli in the 80–230 ms time window (Kruskal–Wallis test; Matlab, The MathWorks, Natick, MA, USA). The first test detected neurons without regard to stimulus specificity, the latter test protected against the omission of cells responsive to one of a set of stimuli. Figure 2 shows examples of the types of cells for which each test was sensitive. Obviously, cells significant by both measures were also included. While our main analyses focus on pairs of cells we emphasize that cells were selected based on their individual properties not pairwise properties.

For neurons meeting this inclusion criterion, we determined the number of spikes generated by each neuron in each of the several time windows. For the main analysis we used windows of: 0–50, 50–100, 100–150, 150–200, 200–250, and 250–300 ms after the onset of the visual stimulus. We subsequently repeated the analyses for a subset of the data on windows extending out to 550 ms. When computing tests of responsiveness, we compared to the 50 ms immediately preceding stimulus onset.

For each time period, we compared the spike counts to the prestimulus time period by a Wilcoxon Rank Sum Test, and computed the Kruskal–Wallis statistic for each cell as a function of stimulus. We also calculated the mutual information between each cell's spike counts and the images individually, and determined its probability by a permutation test.

As background statistics for each pair of simultaneously recorded cells we recorded whether one, both, or neither, for the time period under study was responsive, selective, or informative as determined from the above tests. We also calculated the Pearson product moment correlation between the spike counts.

Spike counts were determined to 1 ms precision and binned for short time epochs zeroed to the onset of an image. The length of the time bins was usually 50 ms. This gives a spike count value (denoted r_1 or r_2, respectively, the vector value $\mathbf{r} = [r_1, r_2]$) for each neuron. For each time period and for each pair of neurons, we computed a direct estimation of the mutual information jointly (i.e., for $\mathbf{r}$) and individually (i.e., for r_1 and r_2). The difference between this joint mutual information and the sum of the individual mutual informations was the synergy. In addition, we calculated the ΔI, a measure of the coding inefficiency that results from assuming conditional independence of the spike responses. We determined a statistical significance for each comparison by its permutation distribution (see the discussion of "shuffling" below and Table 1). We now consider these information measures and our significance tests in more detail.

An expression for mutual information that demonstrates what is being measured is,

$$I(s;\mathbf{r}) = \sum_{s,\mathbf{r}} P(s,\mathbf{r}) \log \frac{P(s,\mathbf{r})}{P(s)P(\mathbf{r})}. \quad (1)$$

This equation is the average of the logarithm of the ratio of the joint distribution of stimuli and response vector to their product distribution. If the stimuli and

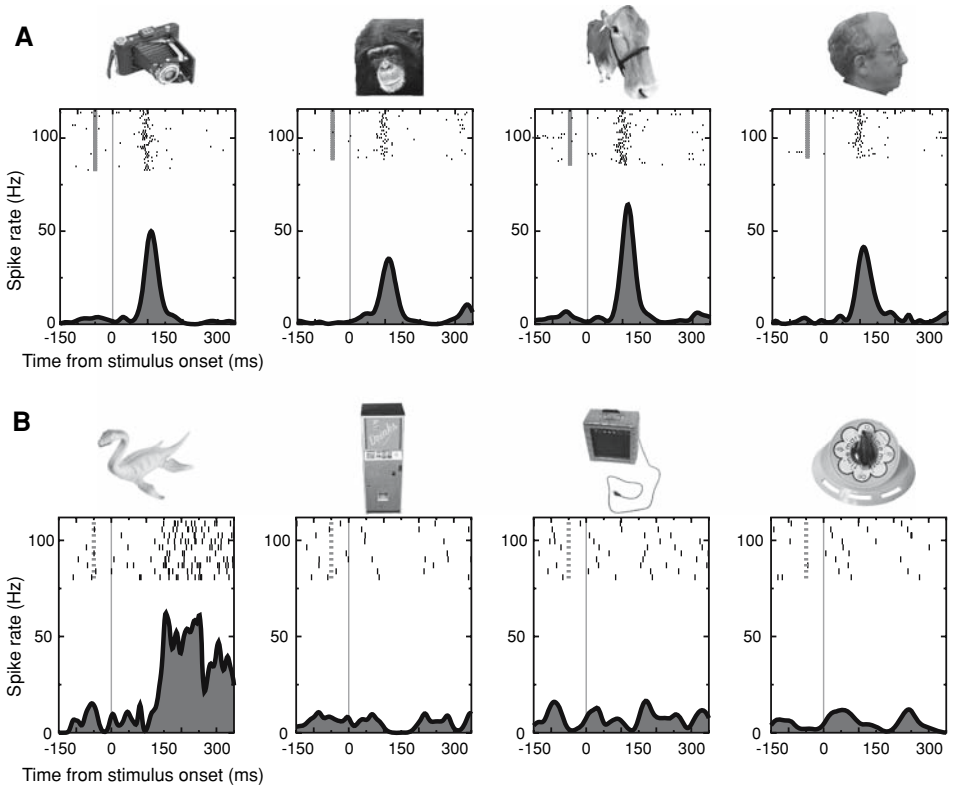

Fig. 2 Raster plots for two separate neurons from two different experimental sessions. These examples demonstrate the two notions of "visual responsiveness" used for selection. The *top figure* illustrates responses from a cell exhibiting robust, but generally nonspecific activation. The *lower panel* shows data from a cell that responds selectively to only one of the presented images

Table 1 The two different permutation strategies are illustrated

Raw data			Label shuffle (MI)			Trial shuffle (synergy and Δ)		
Image	N1	N2	Image	N1	N2	Image	N1	N2
Apple	*2*	*5*	*Banana*	*2*	*5*	*Apple*	1	*5*
Banana	**9**	**3**	*Apple*	**9**	**3**	**Banana**	*7*	**3**
Banana	*7*	*4*	**Banana**	*7*	*4*	*Banana*	**9**	*4*
Apple	1	4	Apple	1	4	Apple	*2*	4

Four faux trials are shown with a stimulus listed for each accompanied by a spike count measure. Different font shapes are used for each trial. In the middle set of three columns we show the label shuffling strategy. Note that the spike count pairs are the same, but that the labels of the stimuli have been shuffled. Sometimes a rate pair will be shifted to a stimulus of the same name and sometime the stimulus names will be different. In the right set of three columns we illustrate the trial shuffling strategy. Here the stimulus names are left in place and one of the spike count vectors is rearranged. Note that counts are only shuffled within a stimulus set. That is apple counts always end up with apple labels. It is not necessary to shuffle both spike counts since permutating both spike count vectors is equivalent to some permutation of one of the counts and a reordering of the trials

response are independent then these two distributions will be equal and every term in the sum will be zero and thus the mutual information will be zero. The direct method of computing mutual information estimates these quantities by counting the observed occurrences and dividing by the number of observations. Mutual information is symmetric in its arguments so that $I(s; \mathbf{r}) = I(\mathbf{r}; s)$.

Finite sample effects upwardly bias mutual information estimates making it difficult to directly compare raw values (Miller 1955). That is, even if one constructs independent stimuli and response relations, in a finite sample, the calculated value will not equal zero. There are equations that compute a corrected value based on the number of response classes and trials. Also, there are more involved ways of computing the mutual information (Paninski 2003). Still, none of these approaches completely solves the difficulty of estimating the actual mutual information. We chose therefore to focus not on the absolute value calculated from the data, but whether the observed value was more extreme than expected by chance.

If we randomly shuffle the labels for the images relative to the spike counts then when we calculate

Eq. 1, our numerator inside the logarithm will just be a chance association of stimuli and images. Because of finite sampling we will still compute a positive number. If we do this a large number of times, we will generate a distribution of values for the chance association of stimuli and spike counts within the experimental data. If our empirically recorded value is larger than 95% of this distribution, we can conclude that we would have had a 5% chance or less of seeing a value this extreme. We used 200 permutations. An introduction to permutation tests is provided by Good (1994). Our permutation strategy is shown schematically in Table 1.

If our neurons jointly code information about the stimulus then we expect to have more information about the stimulus from considering the joint distribution than from each neuron separately. That is, we believe that,

$$I(s;\mathbf{r}) > I(r_1;s) + I(r_2;s).$$

We can quantify this advantage by using the synergy defined as,

$$\text{Synergy}(r_1,r_2) \triangleq I(s;\mathbf{r}) - I(r_1;s) - I(r_2;s). \quad (2)$$

The above equation can also be written (see Appendix) as,

$$\text{Synergy}(r_1,r_2) = \langle I(r_1;r_2|s)\rangle_s - I(r_1;r_2). \quad (3)$$

The use of synergy has been developed by Brenner et al. (2000) (see also Scheidman et al. 2003). Equation 3 illustrates some of the implications of neuronal correlations for synergistic coding. First, the second term in Eq. 3 is the mutual information between the spike counts of the two neurons for all the stimuli. This can be thought of as a ''super''-correlation since it quantifies any predictive relationship between these two measures, linear or nonlinear. As mutual information is non-negative, any relationship between the two neurons will only decrease the opportunity for synergistic coding. The first term in Eq. 3 is the mutual information between the two neurons conditioned on the stimulus. It measures how much one neuron's spike counts tells us about the other neuron's given a particular stimulus. If the two neurons are conditionally independent then this will be zero for all stimuli and the average over all the stimuli (indicated by the brackets) will also be zero. So, regardless of the correlation between the neurons, if the two neurons are conditionally independent, there can be no synergistic coding, there can only be redundancy, which will be determined by the correlation between spike counts over all (e.g., if both neurons independently fire a little to one stimulus and a lot to another).

Under an assumption that two neurons are conditionally independent, the first term is zero, except for the estimation bias, and the second term is positive due to estimation bias and, potentially, a correlation between the neurons across stimuli. By shuffling our trials within a stimulus class, we can estimate the value for each of these terms due to the above effects. A function of random variables is itself a random variable so we can subtract each from the other and have a distribution on their difference. While Eq. 3 gives the rationale, Eq. 2 proves simpler for estimation since the second and third terms do not change when we permute across trials, but respect stimulus class, and so we only have to recalculate one value.

There has been discussion recently over the preferred information theoretic quantity to calculate for neuronal measures (Scheidman et al. 2003; Latham and Nirenberg 2005). Therefore, we also calculated the delta I (denoted ΔI). This quantity, whose rationale is extensively discussed in the two citations, can be thought of as a measure of distance. The rationale is to conceive of all stimuli as being ''encoded.'' The average code length is the product of the stimulus probabilities and the length of the codewords that encode them. It can be shown that for any distribution $P(x)$ there exists a code book (a dictionary of signals to codewords) that achieves a minimum expected code length and that the length of the codeword for each signal is a function of its probability ($\sim \log 1/P(x)$). If we built a code book using a wrong distribution (denoted $Q(x)$) we would have code words of length $\sim \log 1/Q(x)$ and this would change our average code length. How big a penalty would we pay? It can be shown (Cover and Thomas 1991) that the extra length we would pay is bounded from above by

$$D(P||Q) + 1,$$

where

$$D(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$$

For computing ΔI, we compare the penalty between using an optimal code based on the true distribution of responses for each stimulus $P(r_1, r_2|s)$ with one that assumes conditional independence $P(r_1|s)P(r_2|s)$. Nirenberg and Latham (2003) prefer to state the formula for ΔI in terms of the probability of the stimulus given the response rather than the measured probability of the response given the stimulus because of their coding

interpretation. However, the one can be derived from the other by the application of Bayes Rule. The result is:

$$\Delta I = \sum_{r_1, r_2, s} P(s, \mathbf{r}) \log \frac{P(s|\mathbf{r})}{P_{\text{ind}}(s|\mathbf{r})}, \quad (4)$$

and $P_{\text{ind}}(\cdot)$ assumes conditional independence so that the joint probability density of the stimulus given the vector of rates can be decomposed as the product of the probability of the stimulus given each spike rate alone. Empirical estimates of this quantity are also subject to bias, so to compute whether the experimentally observed number is unusual we compare the ΔI to the actual data to values from the randomly shuffled data. If there is indeed no importance of correlations in spike rates within a stimulus, then our observed value is just one of the many equally probable results we could have received from the numerous permutations of the trial labels. The probability of having observed a greater or equal value by chance is quantified by seeing what proportion of the shuffle values is greater than our empirical result.

Results

For the 50–100 ms time bin, 117 individual neurons satisfied the screening criteria, yielding 127 cell pairs for analysis (the number of cells and pairs was slightly lower for other time bins with lower spike counts, since we eliminated pairs in which either of the cells did not respond at all in a given time window). Note that all pairs refer to cells recorded simultaneously from two separate electrodes.

Observed firing rates were extremely heterogenous. As expected, the overall mean spike counts for individual cells varied as a function of time from stimulus onset (see Fig. 3).

Mutual information between images and individual neurons became obvious by the 100–150 ms time window (see Fig. 4). The pattern was similar for pairs of visually responsive neurons where information began to increase 50–100 ms after stimulus onset. The majority of pairs showed mutual information values significantly increased above chance, reaching nearly 90% of the population by 150 ms (see Fig. 6).

To better resolve the time frame on which IT neurons become informative about the stimulus, we repeated the mutual information analysis with 10-ms increments. The rise in the number of pairs showing significant mutual information began about 80 ms, was definite by 90 ms and peaked in the region of 150 ms.

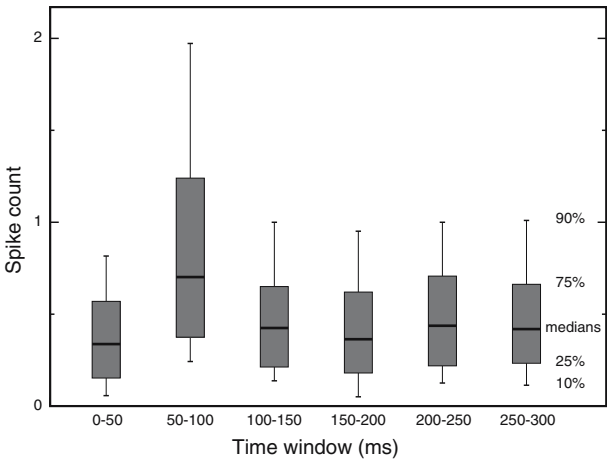

Fig. 3 For each neuron we counted the number of spikes in 50 ms time windows. We then took the mean of these counts for each neuron for all stimuli and all trials. The figure displays the median and percentiles for the mean spike counts for all neurons across time

Since the time bins were so short this was essentially a 0, 1 binary code. These results are in general agreement with recently published data from a collection of individually recorded cells (Hung et al. 2005)

The likelihood that a cell pair had significant information about the stimulus was not, in general, specifically related to whether the neurons were correlated in their firing rates. For example, in the 100–150 ms window 36 of 39 pairs with significant Pearson

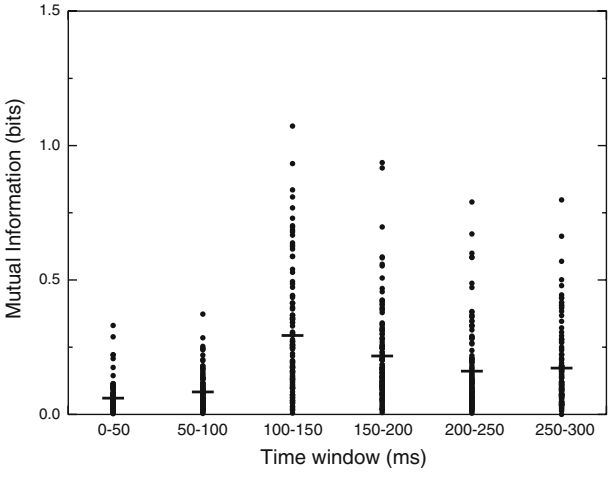

Fig. 4 The mutual information for individual neurons is shown for each time window. Each *filled circle* is the direct mutual information for one cell and the *horizontal tick marks* the mean information for all cells for that time window. Note that the positive values at the very earliest time window reveal the positive bias for mutual information from direct estimation and are the reason we emphasize the proportion of the population of cell pairs showing a statistically significant increase in information rather than absolute values of information

correlations had significant mutual information, but 72 of 82 neuron pairs without significant Pearson correlations also had significant mutual information. Although this goes against intuition, correlation between neurons does not contribute to the potential for joint coding. In Eq. 3, the second term reflects the correlation between neurons across all stimuli. Since mutual information is always nonnegative this second term always reduces the overall value. It is only when neurons are correlated within a stimulus category that there can be synergistic coding. As an extreme example, if both neurons fired the exact same random number of spikes to all stimuli, they would be completely correlated and yet nothing would be learned about the stimulus from considering the neurons individually or jointly.

The likelihood that a cell pair had significant information about the stimulus was also not particularly related to their firing rates. We coded each cell pair based on whether none, one, or both cells had a mean spike count above the median for each 50 ms time bin between 0 and 250 ms. Figure 5 shows a plot of the statistical significance of the mutual information value subdivided by the firing rates. In general, there is no dramatic difference among the three groups, showing that even the cells with relatively lower firing rates, if they are visually selective, can contribute information about the stimuli. Subdividing the data this way did not show any trends for either the synergy or ΔI measures (data not shown).

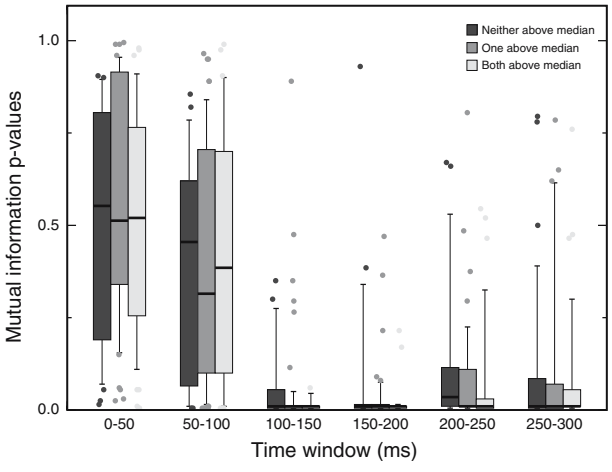

Fig. 5 The box and whisker plot shows the relationship between the observed statistical significance of each cell pair, the time period after stimulus onset, and whether neither, one, or both cells had activity above the median firing rate. The *dark lines* show the median values, the *boxes* demarcate the 25 and 75%, and the *lines* demarcate the 10 and 90% values. Extreme values are shown as *filled circles*. In general, the likelihood of observing significant mutual information for a cell pair is determined by the time period and not the relative firing rates of the cells involved

These data sets came from a situation where we tried to select stimuli that had differential effects on the firing of individual cells. Therefore, it is not remarkable that a high percentage of pairs in this selected set showed that spike counts provided information about the stimulus. The more important question is whether the joint spike count, obtained simultaneously, was more informative than that which could have been gotten from each neuron independently.

For this comparison, we compared the proportion of neuron pairs exceeding a 95% cut-off when we calculated the synergy (see Eq. 2) and the ΔI respecting the stimulus labels but shuffling the counts across trials. This shuffle revealed only chance proportions of the population showing increased synergy or ΔI amounts (see Fig. 6), except perhaps for the last time period. We therefore reanalyzed a subset of the data where the images remained on the screen for 600 ms so that we could extend the analysis windows further out from the onset of the stimulus. These data are shown in Fig. 7 and demonstrate no long-term trend.

Do we only see significant joint information (or synergy) if each of the constituent cells was individually informative? We looked at the proportion of the population of cells pairs subdivided by whether both, one or neither of the cells was significantly informative for that time window. There seemed to be, as expected, a trend for pairs with two individually significant cells to show joint significance when shuffling labels. For example in the 100–150 time window there were 80 pairs with two individually significant neurons and all 80 were jointly significant. Of the 34 pairs with only one of the pair informative about the stimulus 25 achieved statistical significance jointly. For the seven pairs were neither cell was individually informative three pairs were jointly informative. This was in contrast to the measure of synergy where the handful of positive tests were scattered across the three subdivisions (2, 1, 0). A similar pattern was found for later windows as well (data not shown).

Discussion

It is well established that visual neurons are not independent; they do fire together more often than predicted by chance. The value of stimulus induced correlations for neuronal information processing is generally accepted. But are the correlations between neurons, above and beyond that induced by a stimulus, informative?

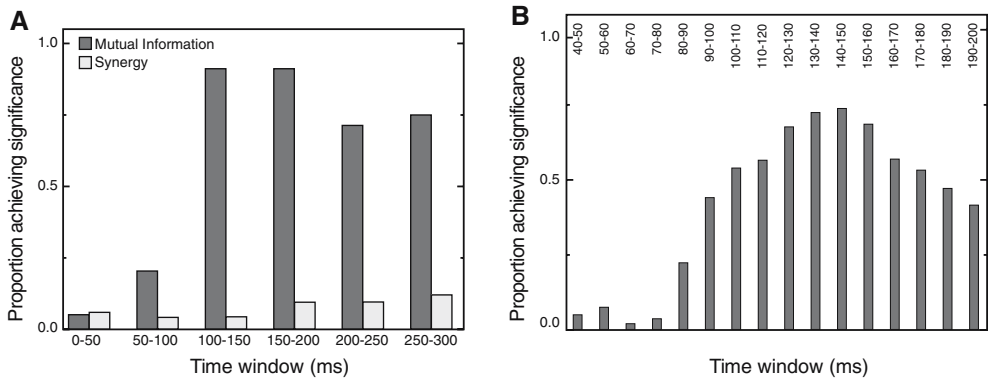

Fig. 6 **a** The *dark gray bars* are the results for mutual information and the *light gray bars* are for the synergy. Results for the ΔI are similar to the synergy. The height of the bar shows the proportion of neuron pairs, for each of six time periods that exceeded 95% of their permutation distribution. Permuting the stimulus labels had a large effect on mutual information beginning at 50 ms post stimulus onset, whereas shuffling the trial labels within a stimulus set had little to no effect on the proportion of neuron pairs showing significant synergy at any time period. **b** In order to more finely resolve the onset of changes in mutual information as a function of the time of stimulus onset we show the proportion of neuron pairs with significant mutual information in 10 ms increments from 80 ms after stimulus onset to 200 ms

Ultimately, the proper question, if our interest is in understanding neural information processing, is not whether neurons fire in a correlated fashion, but whether this joint response tells us more about the stimulus than we could learn from the neurons independently. As yet, there has not been a convincing demonstration that correlated firing in the visual areas of the cerebral cortex carries substantially more information than the spike rates of single neurons considered independently. There is a suggestion that information may exist in

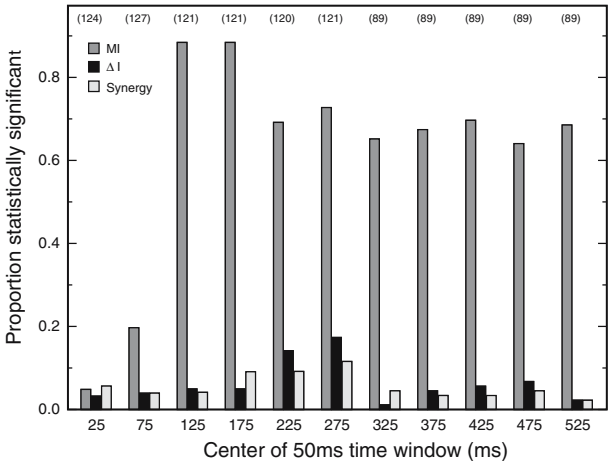

Fig. 7 The proportion of neuronal pairs achieving statistical significance for different information measures subdivided by time period (bin sizes were 50 ms). The numbers in parentheses at the top of each set of bars are the total number of available pairs for each time period. The numbers drop for the later times because we had fewer data sets where the images were displayed for 600 ms. The basic results are repeated here. Mutual information for neuron pairs becomes significant early whereas neither the synergy or ΔI measures show proportions of the population achieving results greater than expected by chance

population codes of the visual thalamus (Dan et al. 1998) but there is little evidence elsewhere. For the IT cortex, a higher order visual area involved in object recognition, there are very few published data collected from multiple neurons recorded simultaneously on separate electrodes.

Recently, Aggelopoulos et al. (2005) evaluated the mutual information between jointly recorded spike rates from IT neurons in monkeys performing a simple visual discrimination task and failed to find substantial information in an estimate of stimulus dependent synchrony. Most of the information was in the joint spike rates. Hirabayashi and Miyashita (2005) recorded in monkey IT during a face versus nonface decision task while recording with a single tetrode. These authors report stimulus specific correlations in the firing of neurons, but the information gained from the correlation code was less than one-tenth that learned from the spike rates.

We add to this body of negative findings by showing that in IT, the information for a joint rate response is essentially what one would expect if the two neurons, given the stimulus, were treated as independent. As always, our negative findings may be the result of our specific stimulus design and task protocol. For instance, our task was neither attentionally demanding, nor did it employ complex and ambiguous visual stimuli. If correlated neuronal activity is selectively employed in these circumstances then we would not have detected it with this protocol. The likelihood of finding a synergy code may be a function of cellular proximity. Our recording method uses electrodes separated by hundreds of microns and is therefore recording a different collection of cells than reported by Hirabayashi and

Miyashita (2005) and Gawne and Richmond (1993), who collected cells with a single recording electrode. We do not have clear data relating cellular separation to our information measures, although improved imaging of in vivo electrode tip position may soon help address this issue.

As we used spike counts in short time windows as our outcome measure, and as it is logically possible for a coincidence code to be independent of changes in rate, our use of spike counts could miss an association between neurons. However, we note that even when we used a 10 ms window, a window in which there were usually only one or no spikes, we still failed to find cooperative coding. So although there is no accepted threshold for neuronal ''synchrony'', we believe that a 10 ms temporal window is within the accepted range for this kind of code.

In conclusion, we wish to emphasize what we did and did not find. There was usually more information, in absolute terms, for pairs of neurons than in either one of the constituent neurons. That is, there was a population code. What we did not find was that there was more to learn about the stimulus from considering the joint spike count on a trial-by-trial basis. A practical consequence is that deductions drawn from sequentially recorded neurons might prove sufficient for studies on population codes (e.g., Hung et al. 2005). Even if neurons were independent encoders there are still benefits from recording multiple neurons simultaneously. Simultaneous recordings can provide substantial efficiencies. In addition, we recorded from multiple neurons with very low firing rates that were visually selective, not because that was our intent, but because they happened to be on electrodes other than the ones where the physiologist was attending, and which were the focus of a day's experimental session. While one could, in principle, get a sample from such cells with a single electrode there is a psychological bias to record cells with high spike rate responses. By advancing multiple electrodes to a region of visual activity, one can still focus on one or two electrodes in a conventional fashion while leaving others to record less active, but potentially informative, visual cells. It may be that these low firing rate cells are exactly the ones where timing issues are of greatest importance.

Appendix

Demonstration that

$$I(s;\mathbf{r}) - I(r_1;s) - I(r_2;s) = \langle I(r_1;r_2|s)\rangle_s - I(r_1;r_2).$$

Proof

$$\begin{aligned}
&I(s;\mathbf{r}) - I(r_1;s) - I(r_2;s)\\
&\quad\stackrel{a}{=}\sum_{r_1,r_2,s} P(s,r_1,r_2)\log\frac{P(s,r_1,r_2)}{P(s)P(r_1,r_2)}\\
&\qquad-\sum_{r_1,s} P(r_1,s)\log\frac{P(r_1,s)}{P(r_1)P(s)}\\
&\qquad-\sum_{r_2,s} P(s,r_2)\log\frac{P(s,r_2)}{P(s)P(r_2)}\\
&\quad\stackrel{b}{=}\sum_{r_1,r_2,s} P(s,r_1,r_2)\log\frac{P(s,r_1,r_2)}{P(s)P(r_1,r_2)}\\
&\qquad+\sum_{r_1,r_2,s} P(r_1,r_2,s)\log\frac{P(r_1)P(s)}{P(r_1,s)}\\
&\qquad+\sum_{r_2,r_1,s} P(s,r_2,r_1)\log\frac{P(s)P(r_2)}{P(s,r_2)}\\
&\quad\stackrel{c}{=}\sum_{r_1,r_2,s} P(s,r_1,r_2)\ \log\frac{P(s,r_1,r_2)P(r_1)P(s)P(s)P(r_2)}{P(s)P(r_1,r_2)P(r_1,s)P(s,r_2)}\\
&\quad\stackrel{d}{=}\sum_{r_1,r_2,s} P(s,r_1,r_2)\left[\log\frac{P(r_1,r_2|s)}{P(r_1|s)P(r_2|s)}+\log\frac{P(r_1)P(r_2)}{P(r_1,r_2)}\right]\\
&\quad=\sum_{s} P(s)\sum_{r_1,r_2} P(r_1,r_2|s)\log\frac{P(r_1,r_2|s)}{P(r_1|s)P(r_2|s)}\\
&\qquad-\sum_{r_1,r_2} P(r_1,r_2)\log\frac{P(r_1,r_2)}{P(r_1)P(r_2)}\\
&\quad\stackrel{e}{=}\langle I(r_1;r_2|s)\rangle_s - I(r_1;r_2)
\end{aligned}$$

where

a this follows from the definitions.

b this expands $P(r_1, s)$ as $\sum_{r_2} P(r_1, r_2, s)\ldots$, and pulls the minus signs inside the logarithms inverting their arguments.

c combines like terms.

d terms are recombined and we use the relationship that $P(a|\,b) = P(a,\,b)/P(b)$.

e the result of applying the definitions for conditional mutual information and mutual information.

References

Aggelopoulos NC, Franco L, Rolls ET (2005) Object perception in natural scenes: encoding by inferior temporal cortex simultaneously recorded neurons. J Neurophysiol 93:1342–1357

Brenner N, Strong S, Koberle R, Bialek W, deRuytervan Steveninck R (2000) Synergy in a neural code. Neural Comput 12:1531–1552

Cover TM, Thomas JA (1991) Elements of information theory. Wiley, New York

Dan Y, Alonso JM, Usrey WM, Reid RC (1998) Coding of visual information by precisely correlated spikes in the LGN. Nat Neurosci 1:501–507

Eckhorn R, Thomas U (1993) A new method for the insertion of multiple microprobes into neural and muscular tissue, including fiber electrodes, fine wires, needles and microsensors. J Neurosci Meth 49:175–179

Franco L, Rolls ET, Aggelopoulos NC, Treves A (2004) The use of decoding to analyze the contribution to the information of the correlations between the firing of simultaneously recorded neurons. Exp Brain Res 155:370–384

Gawne TJ, Richmond BJ (1993) How independent are the messages carried by adjacent inferior temporal cortical neurons? J Neurosci 13:2758–2771

Good PI (1994) Permutation tests: a practical guide to resampling methods for testing hypotheses. Springer, Berlin Heidelberg New York

Gross CG (1994) How inferior temporal cortex became a visual area. Cereb Cortex 5:455–469

Hirabayashi T, Miyashita Y (2005) Dynamically modulated spike correlation in monkey inferior temporal cortex depending on the feature configuration within a whole object. J Neurosci 25:10299–10307

Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. Science 310:863–866

Kreiman G, Fried I, Koch C (2002) Single-neuron correlates of subjective vision in the human medial temporal lobe. Proc Natl Acad Sci USA 99:8378–8783

Latham PE, Nirenberg S (2005) Synergy, redundancy, and independence in population codes, revisited. J Neurosci 25:5195–5216

Lewicki MS (1998) A review of methods for spike sorting: the detection and classification of neural action potentials. Network 9:R53–R78

Miller GA (1955) Note on the bias on information estimates. In: Quastler H (ed) Information theory in psychology; problems and methods. Free press, Glencoe, pp 95–100

Mountcastle V, Reitboeck H, Poggio G, Steinmetz M (1991) Adaptation of the Reitboeck method of multiple microelectrode recording to the neocortex of the waking monkey. J Neurosci Meth 36:77–84

Nirenberg S, Latham PE (2003) Decoding neuronal spike trains: how important are correlations? Proc Natl Acad Sci USA 100:7348–7353

Paninski L (2003) Estimation of entropy and mutual information. Neural Comput 15:1191–1253

Rolls ET, Aggelopoulos NC, Franco L, Treves A (2003a) Information encoding in the inferior temporal visual cortex: contributions of the firing rates and the correlations between the firing of neurons. Biol Cybern 90:19–32

Rolls ET, Aggelopoulos NC, Zheng F (2003b) The receptive fields of inferior temporal cortex neurons in natural scenes. J Neurosci 23:339–348

Scheidman E, Bialek W, Berry II MJ (2003) Synergy, redundancy, and independence in population codes. J Neurosci 23:11539–11553

Sheinberg DL, Logothetis NK (1997) The role of temporal cortical areas in perceptual organization. Proc Natl Acad Sci USA 94:3408–3413

Tanaka K (1997) Mechanisms of visual object recognition: monkey and human studies. Curr Opin Neurobiol 7:523–529

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