JOURNAL OF NEUROPHYSIOLOGY
Vol. 73, No. 2, February 1995. *Printed in U.S.A.*

Sparseness of the Neuronal Representation of Stimuli in the Primate Temporal Visual Cortex

EDMUND T. ROLLS AND MARTIN J. TOVEE
University of Oxford, Department of Experimental Psychology, Oxford, OX1 3UD, England

SUMMARY AND CONCLUSIONS

1. To analyze the selectivity and the sparseness of firing to visual stimuli of single neurons in the primate temporal cortical visual area, neuronal responses were measured to a set of 68 visual stimuli in macaques performing a visual fixation task. The population of neurons analyzed had responses that occurred primarily to faces. The stimuli included 23 faces, and 45 nonface images of real-world scenes, so that the function of this brain region could be analyzed when it was processing natural scenes.

2. The neurons were selected to meet the previously used criteria of face selectivity by responding more than twice as much to the optimal face as to the optimal nonface stimulus in the set. Application of information theoretic analyses to the responses of these neurons confirmed that their responses contained much more information about which of 20 face stimuli had been seen (on average 0.4 bits) than about which (of 20) nonface stimuli had been seen (on average 0.07 bits).

3. The sparseness of the representation of a scene or object provided by each of these neurons (which can be thought of as the proportion of stimuli to which the neuron responds, and which is fundamental to understanding the network operation of the system) can be defined as

$$a = (\sum_{i=1,n} r_i/n)^2/\sum_{i=1,n}(r_i^2/n)$$

where r_i is the firing rate to the ith stimulus in the set of n stimuli. The sparseness has a maximal value of 1.0. It was found that the sparseness of the representation of the 68 stimuli by each neuron had an average across all neurons of 0.65. This indicates a rather distributed representation.

4. If the spontaneous firing rate was subtracted from the firing rate of the neuron to each stimulus, so that the changes of firing rate, i.e., *the responses* of the neurons, were used in the sparseness calculation, then the "response sparseness" had a lower value, with a mean of 0.33 for the population of neurons, or 0.60 if calculated over the set of faces.

5. Multidimensional scaling to produce a stimulus space represented by this population of neurons showed that the different faces were well separated in the space created, whereas the different nonface stimuli were grouped together in the space.

6. The information analyses and multidimensional scaling provided evidence that what was made explicit in the responses of these neurons was information about which face had been seen. Information about which nonface stimulus had been seen was not made explicit in these neuronal responses. These procedures provide an objective and quantitative way to show what is "represented" by a particular population of neurons.

7. The response sparseness value obtained shows further that this population provides a distributed representation of information about which face is being seen. This type of distributed representation is very efficient for fine *discriminations* between the members of a stimulus set: in this case, faces.

INTRODUCTION

The visual pathways project by a number of corticocortical stages from the primary visual cortex until they reach the temporal lobe visual cortical areas (Baizer et al. 1991; Maunsell and Newsome 1987; Rolls 1991, 1992a; Seltzer and Pandya 1978). Neurons with different types of sensitivity to visual stimuli tend to be found in different parts of these temporal cortical areas (Baylis et al. 1987). In some areas neurons respond to stimulus properties such as shape, orientation, texture, and color (Baylis et al. 1987; Tanaka et al. 1991), and in some other areas, especially areas in the cortex in the superior temporal sulcus, up to 20% of the neurons with visual responses have selectivity for faces (Bruce et al. 1981; Desimone 1991; Desimone and Gross 1979; Desimone et al. 1984; Gross et al. 1985; Perrett et al. 1982; Rolls 1981a,b, 1984, 1992a,b). Some of the temporal cortical areas provide a representation of objects and faces that is relatively invariant with respect to retinal position, size, rotation, and even view, and such invariant representations form appropriate inputs to associative neuronal networks in structures to which the temporal cortical areas project, such as the hippocampus and amygdala (see, e.g., Rolls 1992a–c; Treves and Rolls 1994). Consistent with this, lesions of the inferior temporal visual cortex impair the ability of monkeys to respond to objects irrespective of changes in size, lighting, and viewing angle (Weiskrantz and Saunders 1984).

An important question then arises, of how the information about objects and faces is represented by the activity of temporal cortical neurons. Important issues are how selective and "information bearing" (Suga 1989) the neurons are for different classes of stimulus, such as face versus nonface; how selective or information bearing the neurons are for individual items within a class; whether the neurons use "local" or grandmother cell encoding, with strong or even great selectivity of a single neuron for a particular object in the environment (Barlow 1972), or fully distributed representations in which all the neurons participate (Churchland and Sejnowski 1992; Hinton et al. 1986), or sparse representations in which encoding by a sparse ensemble is used (Rolls and Treves 1990; Treves and Rolls 1991). Sparse ensemble encoding might have a number of advantages, such as showing continuous generalization as the nature of the

input changes, showing graceful degradation if the network is incompletely formed or damaged, and providing for large numbers of representations to be stored. The sparseness of the encoding is an important parameter in setting the number of memories that can be stored in a network (Rolls and Treves 1990; Treves and Rolls 1991). It is the aim of the work described here to provide experimental evidence on these issues. In previous work on the selectivity of the responses of temporal cortical neurons, on the cells with responses selective for faces, it has been shown that their responses are selective in that they respond more than twice as much to the best face as to the best nonface stimulus (from a set including a wide range of gratings, simple geometric stimuli, and complex 3-D objects), and in that this difference is statistically significant (see Baylis et al. 1985; Rolls 1984, 1992a). In fact, the majority of the neurons in the cortex in the superior temporal sulcus classified as showing responses selective for faces responded much more specifically than this. For 1/4 of these neurons, their response to the most effective face was >5 times as large as to the most effective nonface stimulus, and for 25% of these neurons, the ratio was greater than 10:1. The degree of selectivity shown by different neurons studied is illustrated in Fig. 6 of Rolls (1992a), and the criteria for classification as face selective are elaborated further by Rolls (1992a). The responses to faces are excitatory, sustained and are time locked to the stimulus presentation with a latency of between 80 and 160 ms. The cells are typically unresponsive to auditory or tactile stimuli and to the sight of arousing or aversive stimuli. These findings indicate that explanations in terms of arousal, emotional, or motor reactions, and simple visual feature sensitivity or receptive fields, are insufficient to account for the selective responses to faces and face features observed in this population of neurons (Baylis et al. 1985; Perrett et al. 1982; Rolls and Baylis 1986). Observations consistent with these findings have been published by Desimone et al. (1984), who described a similar population of neurons located primarily in the cortex in the superior temporal sulcus, which responded to faces but not to simpler stimuli such as edges and bars or to complex nonface stimuli (see also Gross 1992; Gross et al. 1985).

In a previous investigation of whether these neurons respond differently to different faces (and so could provide information useful for face identification), we have shown that in many cases (77% of 1 sample), these neurons are sensitive to differences between faces (as shown by analyses of variance) (Baylis et al. 1985). However, each neuron does not respond only to one face. Instead, each neuron has a different relative response to each of the members of a set of faces. To quantify how finely these neurons were tuned to the faces of particular individuals, a measure derived from information theory, the breadth of tuning metric developed by Smith and Travers (1979), was calculated. This is a coefficient of entropy (H) for each cell that ranges from 0.0, representing total specificity to one stimulus, to 1.0, which indicates an equal response to the different stimuli.[1] The breadth of tuning of the majority of the neurons analyzed was in the range 0.8–1.0. It was thus clear from this and other quantitative measures of the tuning of these face-responsive neurons that they did not respond only to the face of one individual, but that, instead, typically each neuron responded to a number of faces in the stimulus set (which included 5 different faces) (Baylis et al. 1985). Such evidence shows that only across a population or ensemble of such cells is information conveyed that would be useful in making different behavioral responses to different faces. This information about which individual was being seen was very significant, as shown by the finding that the number of standard deviations that separated the response to the most effective from that to the least effective face in the set (a measure analogous to detectability, d', in signal detection theory) was for many neurons greater than 1.0 (Baylis et al. 1985).

The aims of the investigation described here were as follows. First, we wished to obtain further evidence on how selective the face-responsive cells are for faces, by measuring their responses across a very large set of face (23) and nonface (45) stimuli; and by quantifying what is being represented by applying information theoretic analyses to measure how much information was available in their responses about the faces in the set and about the nonface stimuli in the set. We note that if this analysis shows that there is information about the faces in the set, then this provides evidence that the neuronal responses represent information that could be used to identify which of the faces had been seen. To measure how much information is available in the responses of these neurons, we used the methods developed by Optican and Richmond (1987), Optican et al. (1991), and Tovee et al. (1993), following in detail the methods described by Tovee et al. (1993). Second, we wished to measure the sparseness of the representation provided by these visual neurons and to do this with a sparseness measure that can be applied directly to analyses of the storage capacity of networks of neurons (Treves and Rolls 1991, 1994).

This investigation is one of a series (Hornak et al. 1995; Rolls 1992a, 1994a,b) designed to investigate the normal functions of the temporal lobe visual cortical areas, and how damage to these brain regions may underlie the perceptual deficits found in patients with disruption of function of these regions.

METHODS

Recording techniques

The activity of single neurons was recorded with glass-insulated tungsten microelectrodes [after Merrill and Ainsworth (1972), but without the platinum plating] in two alert macaque monkeys (*Macaca mulatta,* weight 3.0 kg) seated in a primate chair with the use of techniques that have been described previously (Rolls et al. 1976, 1990). All procedures, including preparative and subsequent ones, were carried out in accordance with the Guidelines for the Use of Animals in Neuroscience Research approved by the Society for Neuroscience, and were licensed under the UK Animals (Scientific Procedures) Act 1986. The action potentials of single cells were amplified with the use of techniques described previously (Rolls et al. 1979), converted into digital pulses with the use of the trigger circuit of an oscilloscope, and analyzed on-line with the use of a MicroVaxII computer. The computer collected peri-

[1] $H = -k\Sigma_{i=1,n} p_i \log p_i$, where H is the breadth of responsiveness, k is the scaling constant (set so that $H = 1.0$ when the neuron responds equally well to all stimuli in the set of size n), and p_i is the response to stimulus i expressed as a proportion of the total response to all the stimuli in the set.

stimulus rastergrams of neuronal activity for each trial and displayed, printed, and stored each trial, as well as computed the peristimulus time histogram by summing trials of a given type. Eye position was measured to an accuracy of 0.5° with the search coil technique (Judge et al. 1980), and steady fixation of a position on the monitor screen was ensured by use of a (blink version of a) visual fixation task. The timing of the task is described below. The stimuli were static visual stimuli presented at the center of the video monitor placed at a distance of 1.0 m from the eyes. A full-size face image typically subtended 17° in the visual field. The fixation spot position was at the center of the screen. The monitor was viewed binocularly, with the whole screen visible to both eyes.

VISUAL FIXATION TASK. Each trial started at −500 ms (with respect to the onset of the test image) with a 500-ms warning tone to allow fixation of the fixation point, which appeared at the same time. At −100 ms the fixation spot was blinked off so that there was no stimulus on the screen in the 100-ms period immediately preceding the test image. The screen in this period, and at all other times including the interstimulus interval, was set at the mean luminance of the test images. At 0 ms the tone was switched off, and the test image was switched on for 500 ms. At the termination of the test stimulus, the fixation spot reappeared, and then after a random interval in the range 150–3,350 ms, it dimmed, to indicate that licking responses to a tube in front of the mouth would result in the delivery of fruit juice. The dimming period was 500 ms, and after this, the fixation spot was switched off, and reward availability terminated 500 ms later. [A diagram of the timing of this task is provided by Tovee et al. (1994) and Tovee and Rolls (1994).] The monkey was required to fixate the fixation spot in that if he licked at any time other than when the spot was dimmed, saline instead of fruit juice was delivered from the tube; in that the dimming was by so little that it could only be detected if the monkey fixated the spot; and in that if the eyes moved by >0.5° from *time 0* until the start of the dimming period, then the trial was aborted. (When a trial aborted, a high-frequency tone sounded for 0.5 s, no reinforcement was available for that trial, and the intertrial interval was lengthened from 8 to 11 s.)

Stimuli and stimulus presentation

Visual stimuli were stored in digital form on a computer disk and displayed on a monochrome video monitor with the use of a video framestore (Advanced Electronic Design 512). The resolution of these images was 256 wide by 256 high with 256 gray levels. The monitor provided maximum and minimum luminances of 292 and 11 cd/m^2, respectively, and was adjusted internally and by use of a lookup table for linearity to within 3% with the use of a photometer. The computer randomized the order of presentation of these stimuli, switched the stimuli on and off for each trial, and synchronized its data collection so that the stimulus was turned on at the start of the 21st bin of the peristimulus time histogram.

The set of 68 stimuli used in these experiments was as follows. There were 23 full-size face stimuli similar to those illustrated by Baylis et al. (1985). Thirteen of these were human faces (9 full face, 3 profile, 1 back of head), and 10 were monkey faces (8 full view, 2 profile). The 45 nonface stimuli were digitized from photographs of natural scenes and were kindly provided by Dr. D. Tolhurst as part of a set on which quantitative image processing analysis has been performed. The scenes included woodland, countryside, and foods but tended to exclude manmade objects. A small number (4) of the set of non-(whole)face stimuli included small representations of people in the scene, and because it was found that some of the neurons responded to the people in these scenes, care was taken to keep this subset separate when appropriate, and such stimuli were not included in the nonface subset used in the information theoretic analyses. Figure 1 shows examples of some of the face and nonface stimuli used. The mean gray level of all the stimuli was set to ~127, and the standard deviation of the intensity levels was set to be approximately the same for the different stimuli. In the intertrial interval the luminance of the screen was set to a gray level of 127, so that there was no mean luminance change at *time 0* when the stimulus was turned on.

When the visual stimuli were being presented on the video monitor, 1 subset of 20 visual stimuli was used at a time. The computer randomized the sequence in which the members of the subset were presented, and after it had presented the sequence once, it restarted the set with another random sequence. The computer was allowed to repeat the set at least five times to provide sufficient data for the information theoretic analysis. We checked, and confirmed, that the relative responses to the different stimuli were constant throughout the testing period.

Procedure

As tracks were made into the cortex in the superior temporal sulcus, the responses of each neuron were measured to a standard digitized set of stimuli of different faces and of nonface stimuli (Baylis et al. 1985). If a neuron responded to 1 or more of the faces, but to none of the nonface stimuli in the set, then 20 further digitized and real 3-D nonface stimuli were shown, to determine whether the response of the neuron was selective for faces. The criterion, the same as that used previously (Baylis et al. 1985) was that the response to the optimal face stimulus should be more than twice as large as to the optimal nonface stimulus, and that this difference should be significant.

Profile of responses to the set of stimuli

For each neuron, the mean firing rate to each stimulus in a 500-ms period starting 100 ms after stimulus onset (by which time all the neurons had started to respond) was calculated for the response profiles. The spontaneous rate for each neuron was also obtained in the 200-ms period preceding the onset of the visual stimulus. For some profiles the response of the neuron, that is the firing rate to a stimulus minus the spontaneous firing rate, was plotted.

Sparseness of the representation

The sparseness, a, of the representation of a scene or object provided by these neurons can be defined and was calculated as

$$a = \frac{\langle \eta \rangle^2}{\langle \eta^2 \rangle}$$

where $\langle \cdot \rangle$ denotes an average over the statistical distribution characterizing the firing rate η of a cell to the set of input stimuli, or

$$a = [\sum_{i=1,n}(r_i/n)]^2 / \sum_{i=1,n}(r_i^2/n)$$

where r_i is the firing rate to the ith stimulus in the set of n stimuli. The sparseness has a maximal value of 1.0. This is a measure of the extent of the tail of the distribution, in this case of the firing rates of the neuron to each stimulus. A low value indicates that there is a long tail to the distribution, equivalent in this case to only a few neurons with high firing rates. If these neurons were binary (either responding with a high firing rate, or not responding differently from the spontaneous rate), then a value of 0.2 would indicate that 20% of the neurons had high firing rates and 80% had 0 firing rates. In the more general case of a continuous distribution of firing rates, the sparseness measure, a, still provides a quantitative measure of the length of the tail of the firing rate distribution (Treves and Rolls 1991). This measure of the sparseness of the representation of a set of stimuli by a single neuron

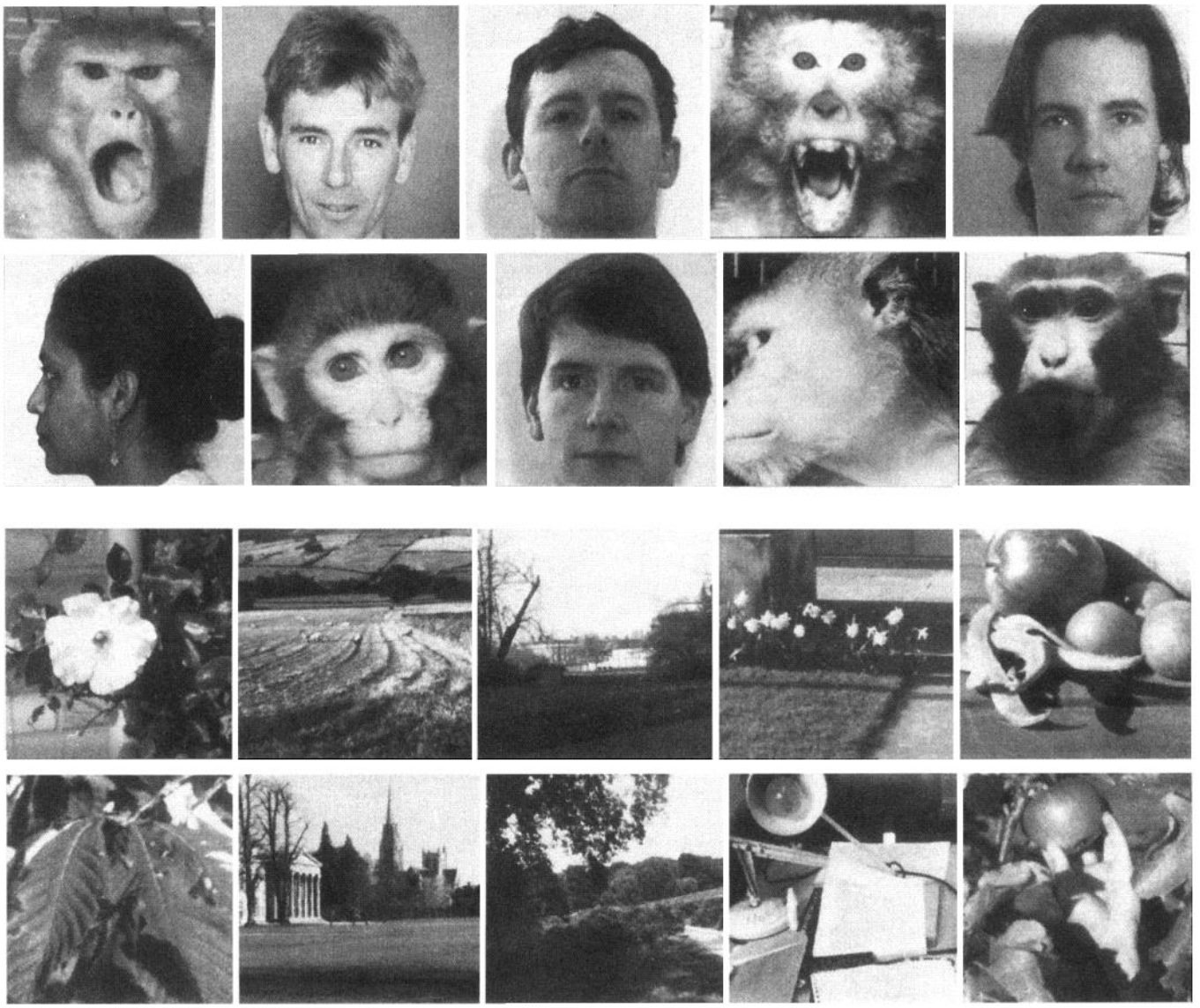

FIG. 1. Examples of the 23 face and 45 nonface stimuli used.

has a number of advantages. One is that it is the same measure of sparseness that has proved to be useful and tractable in formal analyses of the capacity of neural networks that use an approach derived from theoretical physics (see Rolls and Treves 1990; Treves 1990; Treves and Rolls 1991). A second is that it can be applied to neurons that have continuously variable (graded) firing rates, and not just to firing rates with a binary distribution (e.g., 0 or 100 spikes/s) (Treves and Rolls 1991). A third is that it makes no assumption about the form of the firing rate distribution (e.g., binary, ternary, exponential, etc.) and can be applied to different firing rate distributions (Treves and Rolls 1991). Fourth, it makes no assumption about the mean and the variance of the firing rate. Fifth, the measure does not make any assumption about the number of stimuli in the set and can be used with different numbers of test stimuli. Its maximal value is always 1.0, corresponding to the situation when a neuron responds to all the stimuli in a set of stimuli. We note that use of this measure of sparseness in neurophysiological investigations has the advantage that the neurophysiological findings then provide one set of the parameters useful in understanding theoretically (Rolls and Treves 1990; Treves and Rolls 1991) how the system operates.

As described in RESULTS, we have also used a measure of the response sparseness, in which the spontaneous firing rate was subtracted from the firing rate (and the response was clipped to 0). This corresponds to the intuition of some neurophysiologists that it is changes from the spontaneous firing rate that are important, and that is why we show it. However, we note that this response sparseness measure does have problems if the neuron decreases its firing rate below the spontaneous rate for some stimuli, in which case it may be more appropriate to calculate the responses as changes from the lowest firing rate to any stimulus. For these reasons, we place more emphasis here on the sparseness measure a as defined in the previous paragraph on the basis of the absolute firing rates. That measure, a, also has the advantage that in models of neuronal networks, it is the absolute firing rate of the input to each synapse that must be considered when quantifying measures such as the capacity of the network and the interference between stimuli (Rolls and Treves 1990; Treves and Rolls 1991).

Information analysis

The principles of the information theoretic analysis were similar to those developed by Richmond and Optican (1987) and Optican and Richmond (1987), except that we applied a correction proce-

dure for the limited number of trials similar to but not identical to that developed by Optican et al. (1991), in a way made clear below and used by Tovee et al. (1993) and Tovee and Rolls (1994). A novel aspect of our data analysis is that we investigated how much information was available in short epochs of the spike train.

Representation of neural responses

The raw data, expressed as the poststimulus occurrence times of individual spikes, recorded with 1-ms resolution, was smoothed by convolution with a Gaussian kernel with a σ of 5 ms (corresponding to low-pass filtering with −3-dB cutoff at 240 Hz). The smoothed data were then quantized, for some of the analyses, into bins of width 10 ms. As a result, the numbers of spikes in each bin i, giving the components x_i of a multidimensional response vector, were real rather than integers. At least 5 such time series were collected for each stimulus in the stimulus set.

Extraction of the principal components of the variance

To analyze the temporal course of the response, a certain number B of consecutive bins (typically $B = 40$) were selected to form a basis in the space of response vectors. In most cases we chose the bins to cover either of the two poststimulus time periods 0–400 or 150–550 ms. Labeling the bins with the subscript i (e.g., $i = 1, 40$), the response vector during a particular trial k is denoted as x_i^k. The covariance matrix of the responses recorded from a particular cell is

$$C_{ij} = \sum_k (x_i^k - \bar{x}_i)(x_j^k - \bar{x}_j)/N$$

where N is the total number of trials for that cell, and $\bar{x}_i$ is the average number of spikes, over those N trials, falling in bin i. Eigenvalues and eigenvectors of the covariance matrix were extracted with the use of a standard algorithm. The eigenvectors, labeled $l = 1, \ldots, B$ in order of decreasing eigenvalue, form a new basis for the response space. Each subtracted response vector (i.e., with the average response vector subtracted out) is now expressed in this new basis via multiplication by the appropriate orthogonal matrix. The resulting coefficients c_l^k [sometimes called the Karhunen-Loeve transform of the vector $(x_i^k - \bar{x}^k)$] are uncorrelated, in the sense that their covariance matrix is, by construction, diagonal. Note that although the eigenvectors of C_{ij}, the principal components, are normalized to unit length, they are plotted in the figures after multiplication by the square root of the corresponding eigenvalue, to bring out their relative contribution to the variance.

Means and statistical significance of the coefficients

The coefficients c_i^k were then averaged over the trials with the same stimulus. (In the analyses described in RESULTS, 20 stimuli were typically included in the analysis.) The means c_i^s represent the loading of stimulus s on component i, and their statistical significance can be tested with the use of the bootstrap method (Richmond and Optican 1987). A scrambled stimulus-response pairing is obtained by assigning random selections of actually recorded responses to new stimulus labels, while maintaining the same number of trials of each stimulus as in the authentic pairing. The coefficients c^s_i are then calculated for each pseudostimulus label, with the number of labels high enough that the distribution of coefficients approach, for each component, a continuous distribution (we used 100 times as many labels for the bootstrap distribution as experimental stimuli). The statistical significance of the original c_i^s can now be tested against the null hypothesis distribution that the relationship between stimulus and response be purely random. We note that if there are 40 coefficients, then the response space is 40-dimensional.

Raw information measures

If S denotes the set of all stimuli s, and R the set of responses, the average information contained in the responses of a particular cell about that set of stimuli is defined as

$$I(S, R) = \sum_{s \subset S} \sum_{r \subset R} P(s,r) \log_2 \frac{P(s,r)}{P(s)P(r)}$$

where P is the probability of occurrence of a particular event. In evaluating the information content from the data recorded, we consider different definitions of the response set. In one case, responses are simply quantified by the number of spikes within a preset time period (a unidimensional measure). In another case, we consider a multidimensional response space spanning the first Q (typically $Q = 1$, 2, or 3) of the principal components extracted with the algorithm above. If, for example, three principal components are included, the particular response relative to trial k is defined as the triplet (c_1^k, c_2^k, c_3^k). Although the set of stimuli can be discrete (it is in the present experiment), R is generally a continuum (in the 2nd case, a Q-dimensional vector) space. Because in practice one has to evaluate the expression for I by performing a sum rather than an integral, R needs to be quantized. We perform this second quantization following a procedure similar to the first. Consider first the spike count case. The original data are represented by the number of spikes n^k recorded in trial k within the prescribed window, minus the average over all trials. The range of the data is set as $(-\Delta n, \Delta n)$, with Δn the largest between the maximum and absolute value of the minimum recorded for that cell. This range is divided into a preselected number D of bins (we ultimately used $D = 15$). $D - 2$ bins have width $dn = 2\Delta n/(D - 3)$, with the second bin centered around $-\Delta n$, the third shifted by dn, and so on. The first and last bins cover the two semi-infinite intervals at the extremes. If we consider now only the trials relative to a given stimulus s, a smoothing procedure is applied by convolving the individual values n with a Gaussian kernel of width the standard deviation $\sigma_n(s)$ of the values relative to the same stimulus. The result, normalized by dividing by the total number of trials, is quantized into the bins defined above, the area within each bin being used as an estimate of the joint probability $P(s, r)$, where r corresponds to one of the response bins. Summing over all stimuli gives $P(r) = \Sigma_{s \subset S} P(s, r)$. In the case, instead, of the information contained in the Q principal components, the original data are represented by the Q-dimensional set of coefficients c^k. Note that the average response has already been subtracted out. One repeats the same procedure Q times to find the ranges spanned by the coefficients for each component, divides the ranges into bins, smooths the distribution of values recorded for each stimulus, and finds $P(s, r)$. Each of the D^Q bins of the response space is now the product of independently determined bins for each of the Q components, and the relative probability is obtained by multiplying the estimated probabilities of the response falling into each unidimensional interval. As noted by Optican and Richmond (1987), it would be preferable to estimate $P(s, r)$ by using the more time-consuming procedure of recalculating separate principal components for each stimulus, because these are in general different from those calculated over the whole stimulus set. However, this simplification in the construction of the estimator $P(s, r)$ is of minor import when compared with the distortions produced by limited sampling (see below).

Subtracted information measures

The procedure introduced so far for estimating the probability $P(s, r)$ of a particular response is rather simple. Slightly more sophisticated estimators can be used (see Fukunaga 1972; Optican and Richmond 1987) that attempt to take into account the limited number of trials available, in practice, for each stimulus. Ulti-

mately, however, the fact that the probability distribution for each response is extracted from a small sample remains a problem, which cannot be solved by adjusting the kernel used for smoothing, nor by a clever quantization into bins. This can be seen by using again a bootstrap procedure similar to the one used to test the significance of principal component coefficients (Optican et al. 1991). One generates a scrambled stimulus-response pairing, with the number of pseudostimulus labels now equal to the number of real stimuli, and calculates the information I_0 contained in the responses about their randomly paired pseudostimuli. I_0 should be zero, whereas it turns out, in practice, to be of the same order as the original I. This is because the use of probability distributions based on a limited number of trials per stimulus unavoidably biases upward an information estimate, by producing random fluctuations in the distributions of responses to different stimuli, which are then picked up by the information measure as being related to the stimuli.

The most straightforward remedy is to subtract I_0 from I, to obtain a measure $I_s = I - I_0$, which averages zero, even for small samples, in the absence of a stimulus-response causal relationship. This is one of the simplest of several correction procedures for sampling bias in information estimates that have been reported in the literature (see Optican et al. 1991, and references therein). It should be noted that only few of these procedures are applicable to data for which the a priori distribution is unknown, and that the relative validity of different methods is strongly dependent on the data distribution itself (Macrae 1971). The latter of these considerations, especially, makes us tend to prefer the simplest and most conservative procedure (i.e., $I_s = I - I_0$) (see further Treves and Panzeri 1994), rather than the one that consists in subtracting, instead of the "fraction of spurious information," its square, i.e., $I'_s = I[1 - (I_0/I)^2]$, as proposed by Optican et al. (1991). We note that subtraction of the square (which is obviously less than the fraction itself) leaves an estimate I'_s often closer to I than to I_s. Optican et al. (1991) argue in favor of subtracting the square on the basis of computer simulations (and data from 2 cells in the striate cortex), which seem to indicate a faster convergence to the asymptotic "correct" value. We note, however, that those computer simulations, based on different distributions than the unknown ones characterizing real data, could lead to inappropriate conclusions with small sample sizes (i.e., numbers of trials), although of course both correction procedures approach the same asymptotic correct value as the sample size increases. We note that, because the upward bias of I increases with the dimensionality of the response space, the use of some subtracted estimate is particularly important when calculating the information contained in the first Q principal components, with $Q > 1$.

Because on the basis of the above points it was not clear which correction procedure should be used, we specifically investigated which correction procedure may be most appropriate for data obtained from real neurons (Tovee et al. 1993, Fig. 1). Use of the correction procedure $I_s = I - I_0$ (which we term *correction procedure 2* for reference), tended to underestimate the information when there are few (3–10) presentations of each stimulus, but was more likely to be accurate with 15–60 presentations of each stimulus. In contrast, the measure $I'_s = I[1 - (I_0/I)^2]$ (which we will refer to as *correction 1*), underestimated less than *correction 2* with small numbers of presentations of each stimulus (3–7), but did provide somewhat of an overestimate of the information about the stimulus for all other numbers of stimulus presentations in the range 10–60. In the light of these findings, we present the data later with the results calculated with both corrections, knowing that *correction 2* is more conservative and has now received further theoretical justification (Treves and Panzeri 1994). *Correction 1* does allow comparison with other data (Optican et al. 1991), but it must be borne in mind that use of this correction may tend to overemphasize the proportion of information available in principal components higher than 1.

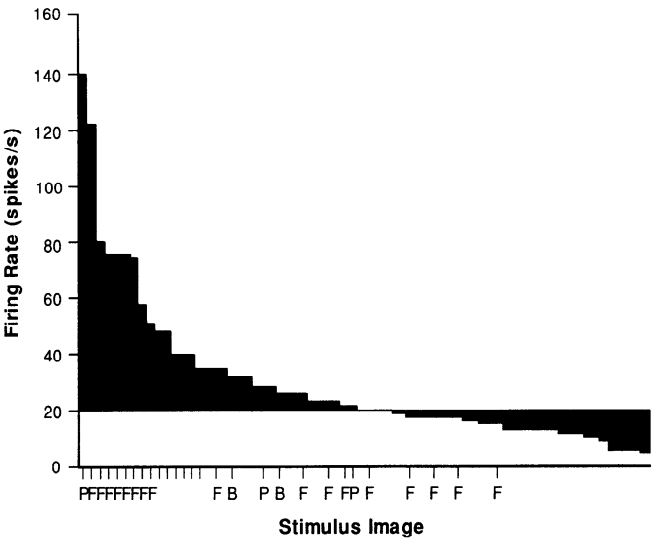

FIG. 2. Distribution of firing rates to all 68 stimuli for 1 neuron (*AM243*). The firing rate of the neuron is shown on the ordinate, the spontaneous firing rate of the neuron was 20 spikes/s, and the histogram bars are drawn to show changes of firing rate from the spontaneous rate (i.e., neuronal responses) produced by each stimulus. Stimuli that are faces are marked F, or P if they are in profile. B refers to images of scenes that included either a small face within the scene, sometimes as part of an image that included a whole person, or other body parts, such as hands or legs. The nonface stimuli are unlabeled.

It should be noted, finally, that an information estimate based on a quantized response set tends to grow with the size D^Q of the set for D small, until it saturates once the width of the D bins along each dimension becomes negligible with respect to the standard deviation of the relative coefficients for each stimulus. We used $D = 15$ [similar to the value of 12 used by Optican and Richmond (1987)] after checking that no marked increase in I_s resulted from using larger D values.

RESULTS

It was possible to complete the analyses with the use of 68 stimuli (23 face and 45 nonface) for 14 neurons recorded in 2 monkeys. To confirm that this is a representative set of this type of cell, we compared the results (for a subset of stimuli) with a larger set of 25 cells, as described at the end of RESULTS. Examples of rastergrams and spike-density functions for this type of neuron have been shown previously in Fig. 3 of Tovee et al. 1993 and are therefore not reproduced here.

The distribution of firing rates to all 68 stimuli for 1 neuron (*AM243*) is shown in Fig. 2*A*. The spontaneous firing rate of this neuron was 20 spikes/s, and the histogram bars indicate the change of firing rate from the spontaneous value produced by each stimulus. Stimuli that are faces are marked F, or P if they are in profile. B refers to images of scenes that included either a small face within the scene, sometimes as part of an image that included a whole person, or other body parts, such as hands or legs. The nonface stimuli are unlabeled. The neuron responded best to two of the faces, had some response to some of the other faces, and had little or no response and sometimes a small decrease of

firing rate below the spontaneous firing rate to the nonface stimuli.

Histograms showing for the same cell (*AM243*) the firing rate distribution produced by the 23 face stimuli are shown in Fig. 3*A*, and of the 45 nonface stimuli are shown in Fig. 3*B*. It is evident that the cell had almost no response to the nonface stimuli (the spontaneous rate was 20 spikes/s), but for face stimuli the responses ranged from a very large increase in firing rate to two of the stimuli to a decrease in firing rate to some of the faces (Fig. 3*A*). The mean response (defined as change of firing rate from the spontaneous) to faces was in fact 27 to faces and to nonfaces was −3 spikes/s. The figures show that this neuron did not respond to all faces and fail to respond to all nonfaces. Instead, it had little response to nonfaces, and had a wide range of responses to faces. This is typical of neurons with face-selective responses and means that an ensemble of such neurons could convey information about which face had been shown (Baylis et al. 1985; Rolls 1984, 1992a,b), a possibility considered quantitatively below.

To provide an indication of how individual neurons differed from each other, similar data for another neuron, *AM231*, (recorded on another track on a different day) are shown in Figs. 4 and 5. The cell had a spontaneous firing rate of 3.7 spikes/s. Figure 4 shows the firing rate profile of the cell to the different stimuli, and Fig. 5, *A* and *B*, the firing rate histograms to the face and nonface stimuli, respectively. The cell had a mean response of 1.1 spikes/s to the nonface stimuli, with only one nonface stimulus producing a response >10 spikes/s. In contrast, the mean response to the face images was 9.6 spikes/s, with many of the faces producing responses of >11 spikes/s.

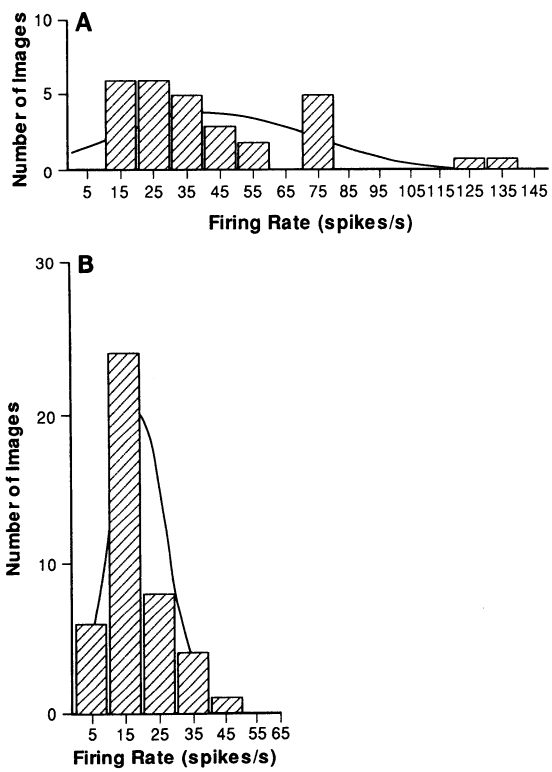

FIG. 3. Histograms showing the firing rate distribution produced by the face stimuli in the set (*A*), and by the 45 nonface stimuli in the set (*B*), for the same neuron, *AM243*, as that shown in Fig. 2. The abscissa shows the firing rate of the neuron in spikes/s. (The spontaneous firing rate was 20 spikes/s.)

Examples of the raw response profiles of cells are important to gauge the type of encoding used by these neurons, and we therefore include profiles from two more cells in Fig. 6. *AM236* (Fig. 6*A*) was quite finely tuned, with very large responses to two faces, and a wide distribution of responses to other faces. Most of the nonfaces produced only a small change in firing rate, which in some cases was an increase, and in other cases a decrease, from the spontaneous rate. *AM240* (Fig. 6*B*) tended to respond much more to most of the faces than to any nonface, and most of the Body (B) images did not activate the neuron.

To examine quantitatively the extent to which these neurons conveyed information about which face was shown, we performed principal component and information theory analysis on the responses of these 14 neurons to 20 faces in the stimulus set. The results are shown in Fig. 7. These neurons conveyed on average 0.39 bits of information per cell about which face had been shown. In line with earlier results (Tovee et al. 1993), little information was carried in the temporal patterns of firing of the cells, in that most of the information was in the first as compared with the higher principal components.

These data can be compared with the data in Fig. 8, which shows that these cells conveyed little information about 20 nonface stimuli in the set (mean = 0.069 bits for the firing rate). (These 20 stimuli were chosen at random from the nonface stimuli in the set. Twenty were chosen to make the information measure as powerful as for the 20 face stimuli.)

It is further shown in Fig. 9 that even more information per cell is obtained when 10 face and 10 nonface stimuli are included in the set. There is more information here partly because including nonfaces and faces resulted in a wide distribution of firing rates entering the analysis.

To provide evidence on the nature of the representation that is used for encoding information by these cells, we return to the issue of how distributed versus how local or sparse the representation is. We used the measure of sparseness of the representation provided by a cell described in METHODS. To provide an indication of how to interpret this, we note that the sparseness for the cell shown in Figs. 2 and 3 over all 68 images was 0.58, over the 23 faces only was 0.66, and over the nonfaces only was 0.81. For the cell shown in Figs. 4 and 5, over all 68 images the sparseness was 0.49, over the 23 faces only was 0.65, and over the nonfaces only was 0.68.

We show in Fig. 10 the distribution of sparseness values for the population of 14 cells for which sparseness measures were available over the full set of 23 face and 45 nonface images. The mean sparseness was 0.65 ± 0.16 (mean ± SD).

The sparseness shown in Fig. 10 is that calculated with the use of the absolute firing rates of the neurons. If the spontaneous rate is subtracted, then a measure can be obtained that reflects more the changes in firing rate produced by the set of stimuli. (In the calculation of this response sparseness, negative response rates were clipped to 0. This is in line with the point that the responses of these neurons generally consisted of increases in firing rate.) To provide an indication of how to

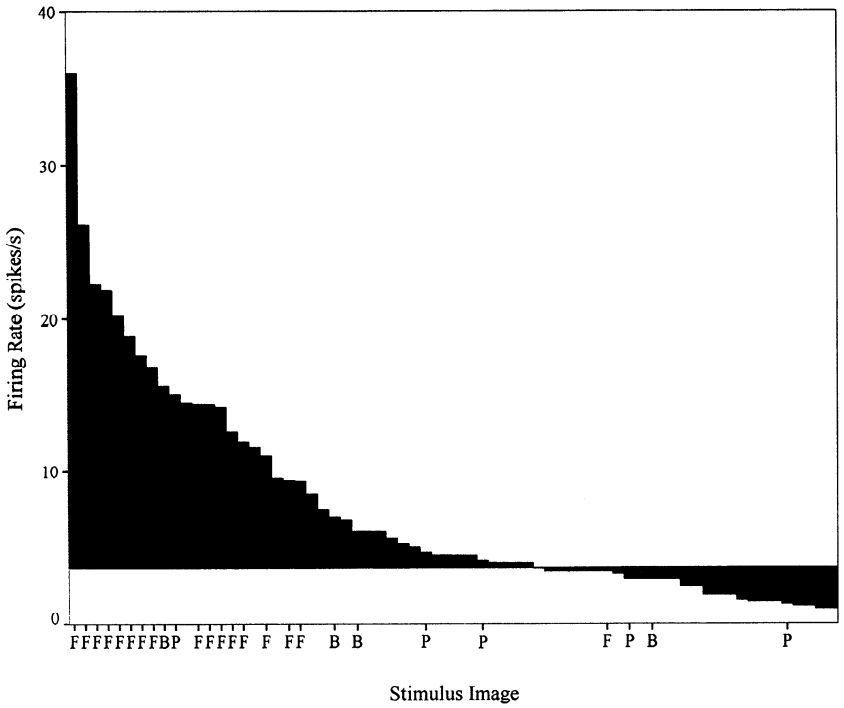

FIG. 4. Distribution of firing rates to all 68 stimuli for another neuron (*AM231*). Conventions as in Fig. 2. The spontaneous firing rate was 3.7 spikes/s.

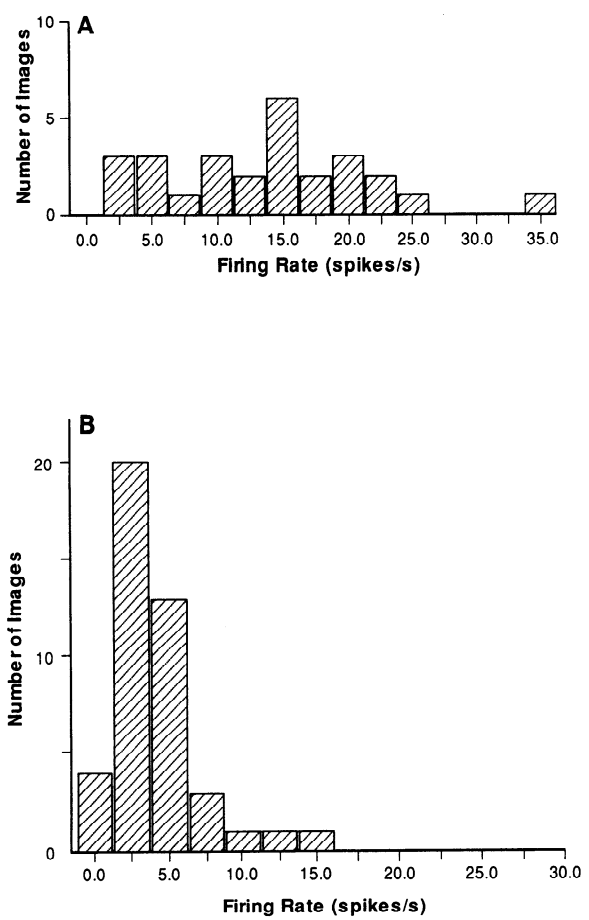

FIG. 5. Histograms showing the firing rate distribution produced by the face stimuli in the set (*A*), and by the 45 nonface stimuli in the set (*B*), for the same neuron, *AM231,* as that shown in Fig. 4. The abscissa shows the firing rate of the neuron in spikes/s. (The spontaneous firing rate was 3.7 spikes/s.)

interpret this, we note that the response sparseness for the cell shown in Figs. 2 and 3 over all 68 images was 0.49, over the 23 faces only was 0.55, and over the nonfaces only was 0.21. For the cell shown in Figs. 4 and 5, over all 68 images the response sparseness was 0.31, over the 23 faces only was 0.57, and over the nonfaces only was 0.33.

The response sparseness indexes of the different cells are shown in Fig. 11. They are lower than the sparseness values based on the absolute firing rates. The mean of the response sparseness over all 68 stimuli was 0.33 ± 0.22.

The mean sparseness across the population of cells to the 23 faces was 0.82 ± 0.12, with the mean response sparseness being 0.60 ± 0.24. The mean sparseness to the 45 nonfaces was 0.80 ± 0.10, with the mean response sparseness being 0.35 ± 0.20.

To provide evidence on what is represented by this population of cells, we performed a cluster analysis to indicate how the neurons clustered or categorized the stimuli. First, the stimulus correlation matrix was calculated, showing the correlation between every pair of stimuli in the responses of the population of neurons. Cluster analysis was then performed on this stimulus correlation matrix (with the use of SPSS, hierarchical cluster analysis, average linkage between groups, and interval measures). The result is shown in Fig. 12. The first division separated three of the profile views of faces from the other stimuli. Next, *face 306,* the juvenile monkey face, was separated from the remaining stimuli. Then many of the remaining face stimuli appeared, with different clusterings between the different faces. Within the clusters of 21 stimuli shown on the right of the dendrogram, 19 were faces, 1 was leaves, and 1 was fruit. The next group of five stimuli included all the hands in the set, providing an indication that some information about hands was provided by this population of neurons. Last, all the remaining stimuli to the left of the hands cluster were nonface stimuli and were all clustered quite close together.

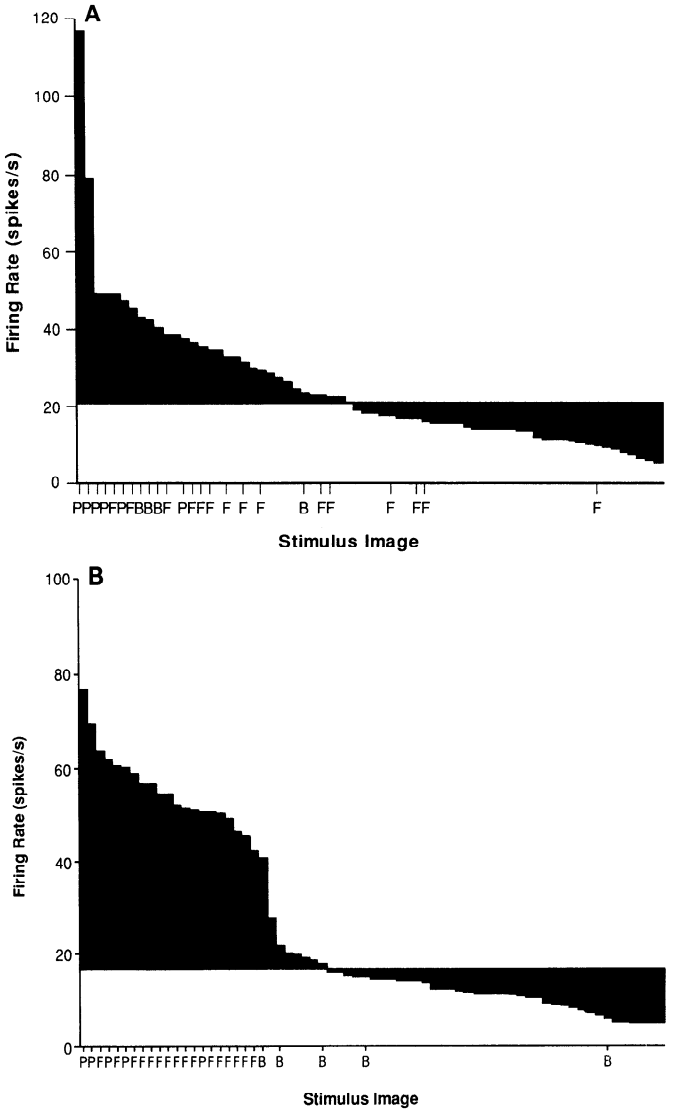

FIG. 6. Distribution of firing rates to all 68 stimuli for 2 more neurons, (*A*) *AM236* with a spontaneous firing rate of 21 spikes/s, and (*B*) *AM240* with a spontaneous firing rate of 16.7 spikes/s. Conventions as in Fig. 2.

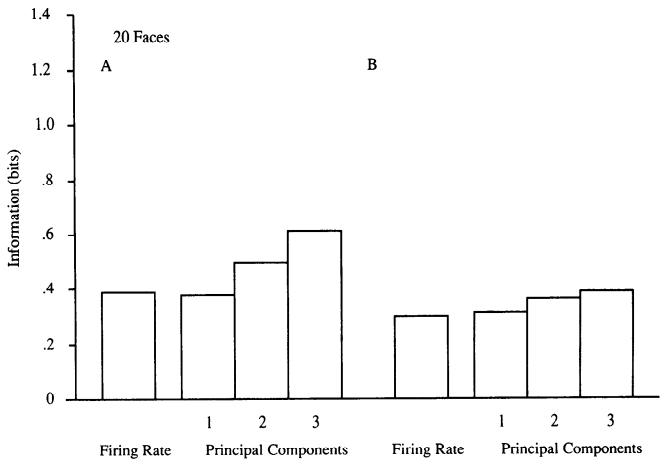

FIG. 7. Mean information available in the firing rate of each of the 14 neurons about the 20 faces in the stimulus set. For comparison, the cumulative information available in principal *component 1,* principal *components 1* and *2,* and principal *components 1–3* is also shown. (Because of the operation of the bootstrap correction procedure, the cumulative information estimate does not always increase as principal components higher than the 1st are added.) *A*: *correction 1* applied. *B*: *correction 2* applied.

Another way of analyzing how this population of cells classifies face and nonface stimuli is to perform multidimensional scaling on the stimulus correlation matrix (Hasselmo et al. 1989; Young and Yamane 1992). (SPSS MDS with Euclidean distances and ratio measures was used.) We found that two dimensions accounted for 96% of the variance. (One dimension accounted for 82%, and 3 for 99%.) We show the 2-dimensional space for the 68 stimuli in Fig. 13. It is shown that the different face stimuli (F or F-P) were well spread out in the space, consistent with the hypothesis that the population of neurons provided evidence that enabled the face stimuli to be separated from each other. It is of interest that in *dimension 1,* the profile views of faces (F-P) were separated from the full-face views (monkey vs. human faces). In contrast, the nonface stimuli (x in Fig. 13) were not well spread out in the space but were mostly grouped together at the low end of *dimension 1*. Where there is some loading of the nonface stimuli on *dimension 1,* it was interesting that these stimuli included the hands in the nonface set. Interestingly, in some of the 45 members of the non-(whole) face set, small parts of the scene contained small representations of people or faces, and these stimuli tended also to be well represented in the space, spread out in their own part of *dimension 2* (stimuli labeled B). The human and monkey faces were intermixed in the space, so that the representation provided by these neurons was of differences between faces that would be useful in face identification, rather than a classification into whether the faces were monkey or human.

Last, we analyzed how rapidly the representation of information about faces and nonfaces became available in the responses of this population of neurons. To do this, we calculated the information that was available in the responses of the neurons about which stimulus had been shown in epochs of different duration, all starting at *time 0,* the onset of the stimuli. This is shown in Fig. 14. It is clear that the informa-

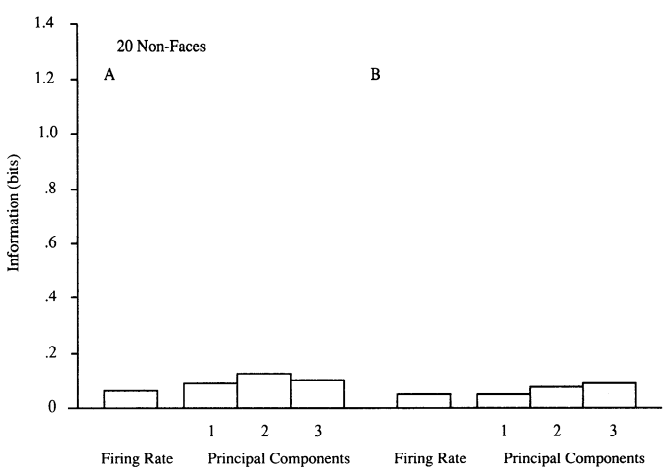

FIG. 8. Mean information available in the firing rate of each of the 14 neurons about 20 of the nonfaces in the stimulus set. For comparison, the information available in principal *components 1–3* is also shown. *A*: *correction 1* applied. *B*: *correction 2* applied.

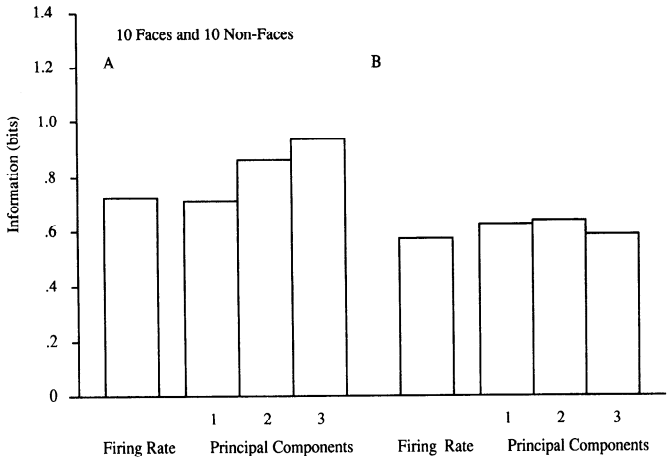

FIG. 9. Mean information available in the firing rate of each of the 14 neurons about 10 of the faces and 10 of the nonfaces in the stimulus set. For comparison, the information available in principal *components 1–3* is also shown. *A*: *correction 1* applied. *B*: *correction 2* applied.

tion accumulates rapidly once the neurons have started to respond (which is typically at 80 ms), and that by 150 ms poststimulus (that is in effectively a 70-ms period of firing), the information available is ~50% of the information that can be obtained by taking a 600-ms epoch.

The recording sites of these neurons in the cortex in the superior temporal sulcus are shown in Fig. 15.

To demonstrate that the population of neurons analyzed with the large stimulus set of 68 stimuli was representative of this type of neuron recorded in the cortex in the superior temporal sulcus, we performed the additional analyses. First, for a larger sample of 25 cells, we calculated the sparseness over a subset of the 68 stimuli consisting of 20 face stimuli, and compared the sparseness values for the original 14 cells with that for the larger group of 25 cells for the same stimulus set. The sparseness value for the 14 cells with the 20 faces was 0.83 ± 0.11; and for the 25 cells was 0.83 ± 0.10. The subset of 14 cells, on which it was possible to run the

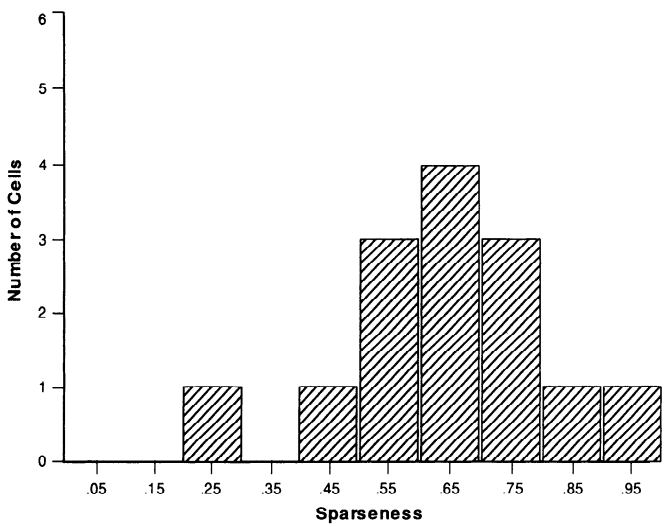

FIG. 10. Distribution of sparseness values for the population of cells. The stimulus set over which the sparseness was calculated was the full set of 23 face and 45 nonface images. The mean sparseness was 0.63 ± 0.15 (mean ± SD).

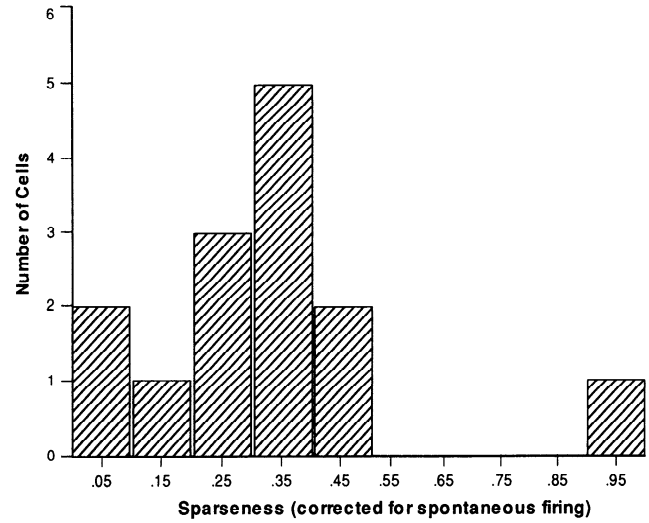

FIG. 11. Distribution of response sparseness values for the population of cells. The stimulus set over which the sparseness was calculated was the full set of 23 face and 45 nonface images. The mean of the response sparseness was 0.33 ± 0.22. The response sparseness is the sparseness calculated over the distribution of response rates, i.e., the firing rate to a stimulus minus the spontaneous rate.

very large set of 68 stimuli, was also typical of the larger population of 25 cells in that the average information (with the use of *correction 2*, and rate-based information) about 20 faces from the subset of 14 cells was 0.44 ± 0.13 bits, and from the 25 cells was 0.42 ± 0.12 bits.

DISCUSSION

First, the results described here provide evidence on the nature of the selectivity of this (or any other) population of temporal cortical neurons with a much more extensive stimulus set than has been available previously. The responses of the cells illustrated in Figs. 2–6 indicate that these cells have a wide and graded range of firing rates to different faces and that, even among a very wide range of nonface stimuli, the majority produce almost no response, and none or at most one or two produces more than a small response. The implication of this is that this population of neurons provides a representation that is useful for distinguishing between different faces, and not between nonface stimuli. The 14 cells analyzed with this very large set of 68 stimuli is representative of cells of this type, as shown by the comparisons with a larger set of 25 cells at the end of RESULTS, and by the fact that the tuning of the cells described here to a subset of the stimuli was similar to that found previously with samples of these cells (Baylis et al. 1985, 1987; Rolls 1992a,b, 1994a,b).

Second, the information theoretic analysis described here provides a firm quantitative basis for this conclusion. As shown in Fig. 7, the average information provided by the firing rate of each neuron about which of 20 of the faces in the set was shown was 0.39 bits. In contrast, the average information about which of 20 nonface stimuli had been shown was 0.069 bits, as shown in Fig. 8.

Third, the results showed that most of the information about which face had been seen was present in the firing rate of the neuron, and that taking into account temporal

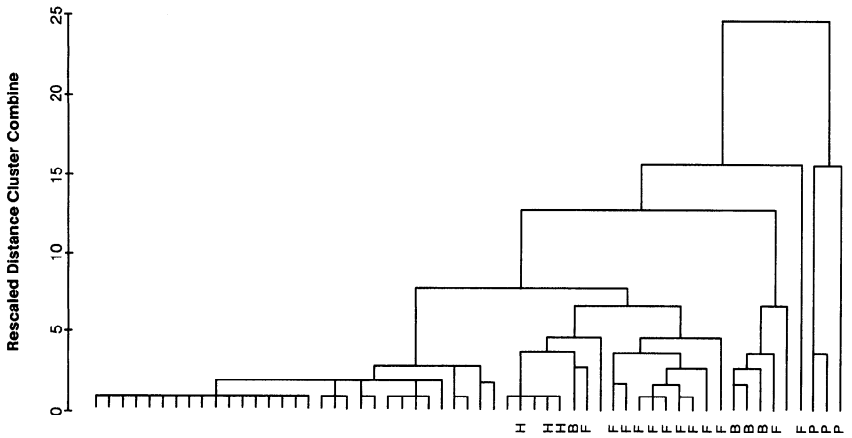

FIG. 12. Cluster analysis to show how the population of cells categorized the set of 68 stimuli (see text). P, face, profile; F, face, front view; B, body parts, including small faces (see text); H, hands.

variation in the spike train to provide temporal encoding did not provide much further information. This is shown in Figs. 7–9, in which there is little more information available when principal *components 2* and *3* are included (which reflect information not in the firing rate), than when only the first principal component (which reflects the firing rate) is taken. This is especially the case when the more conservative *correction 2* is applied (Figs. 7–9*B*) than when *correction procedure 1* is applied (Figs. 7–9*A*). This finding is consistent with our earlier observations that much of the information about which stimulus has been seen is available in the firing rate of the neuron (Rolls and Tovee 1994a; Tovee et al. 1993; Tovee and Rolls 1994).

Fourth, the results show that a considerable amount of information about which stimulus has been seen is present in the responses of these neurons in the temporal cortex. It is shown that, when the information about 10 face and 10 nonface stimuli was calculated, the mean value for each neuron was 0.7 bits.

Fifth, the sparseness of the representation of the 68 stimuli provided by these neurons (taking the absolute firing rate) had an average of 0.65 ± 0.22. If the spontaneous firing rate was subtracted from the firing rate of the neuron to each stimulus, so that the responses of the neurons were used in the sparseness calculation, then the "response sparseness" had an even lower value, with a mean of 0.33 for the population of neurons. If these neurons were binary (either responding with a high firing rate, or not responding differently from the spontaneous rate), then the latter value of 0.33 would signify that each neuron was active for 33% of the stimuli. If we consider just the subset of 23 face stimuli, then the response sparseness was 0.60, signifying (in a simplified binary case) that each neuron was active for 60% of the face stimuli. This latter value is consistent with earlier measurements using the breadth of tuning of these neurons on a much smaller set (5) of face stimuli, which gave breadths of tuning of these neurons that were mainly in the range 0.8–1.0 (Baylis et al. 1985). (In a simple binary case with 5 stimuli, if the neurons were active to 3 of the stimuli, the breadth of tuning is 0.7.)

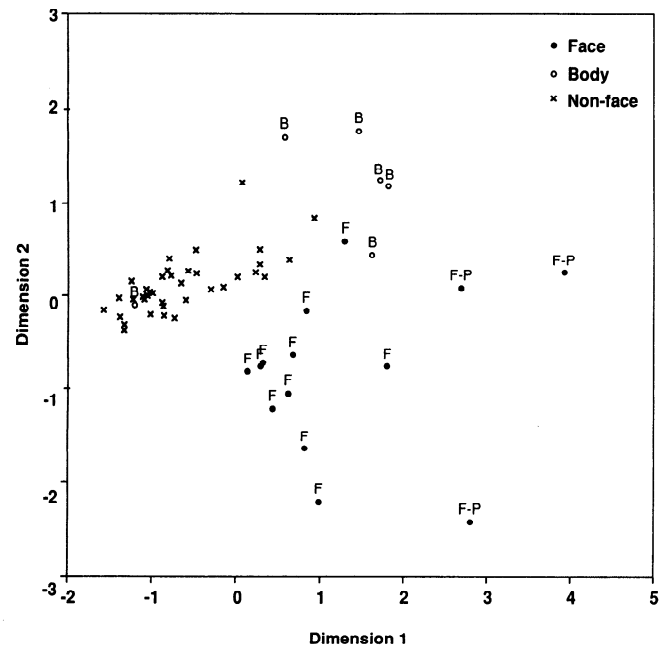

FIG. 13. Two-dimensional space achieved by multidimensional scaling on the stimulus correlation matrix for the 68 stimuli (see text). Labels as in Fig. 12, except F-P, face, profile.

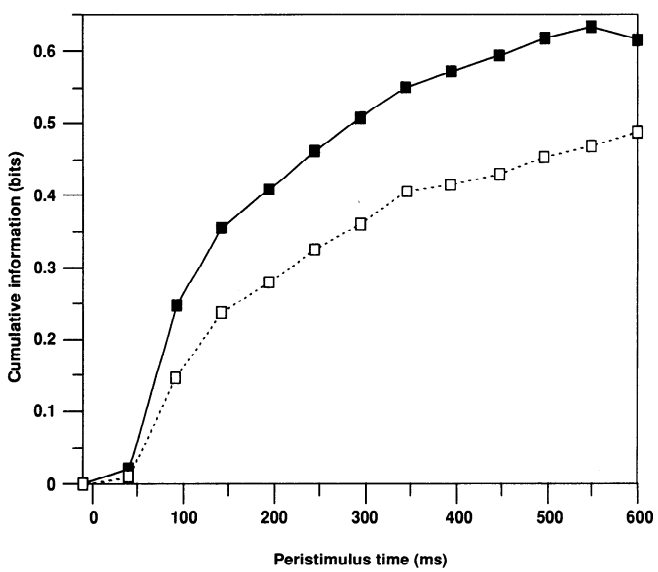

FIG. 14. Mean of the cumulative information available from each cell in different time epochs starting at *time 0*, the onset of the stimuli. Solid squares: correction 1; open squares: correction 2.

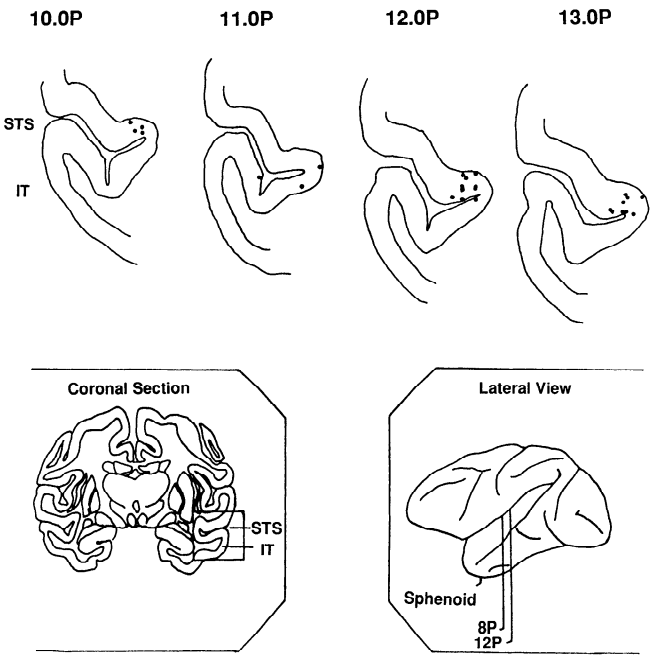

FIG. 15. Recording sites shown on coronal sections of the neurons included in this study. Positions of the coronal sections are shown on a lateral view of the macaque brain. Distances refer to millimeter posterior (P) to the sphenoid reference plane (see text). STS, superior temporal sulcus; IT, inferior temporal cortex.

The mean response sparseness of 0.60 of this population of face-selective neurons indicates that, within the class faces, these neurons implement distributed encoding, with activity in any neuron in the population occurring to approximately one-half of the stimuli in the set. One might comment that, because of the similarity between different faces, these neurons find it difficult to respond more specifically than this. But a more interesting point is that this very distributed encoding allows the maximum information about a set of stimuli to be provided by a population of neurons, provided of course that they do not have the same profile of responsiveness to the set of stimuli. Such a representation would be ideal for *discrimination,* because the maximum information suitable for comparing fine differences between different stimuli would be made available across the population (if 50% were active to each stimulus). However, a representation as distributed as this would not be appropriate for a memory system, in which the aim is to store a large number of memories. In an associative memory with neurons with continuously variable firing rates, such as the autoassociative memory believed to be implemented in the hippocampus (Rolls 1989), we have shown (Treves 1990; Treves and Rolls 1991, 1994) that the maximum number $p_{\max}$ of firing patterns that can be (individually) retrieved is proportional to the number C^{RC} of (associatively) modifiable *RC* synapses per cell, by a factor that increases roughly with the inverse of the sparseness a of the neuronal representation. Approximately

$$p_{\max} \simeq \frac{C^{RC}}{a \ln (1/a)} k$$

where k is a factor that depends weakly on the detailed structure of the rate distribution, on the connectivity pattern, etc., but is roughly in the order of 0.2–0.3. Thus, in a memory system [and similar considerations apply to a pattern associator (Rolls and Treves 1990) such as are suggested to be implemented in the amygdala and orbitofrontal cortex, and in corticocortical backprojections (Rolls 1989, 1992a,c; Treves and Rolls 1994)], one parameter limiting the number of memories that can be stored and retrieved correctly is the number of modifiable synapses received by each neuron (in the order of 5,000–15,000), and another is the sparseness of the representation, a. Thus, in memory systems, optimal performance requires a sparse representation, and this is consistent with values for sparseness in the hippocampus that are in the order of 0.01–0.04 (Barnes et al. 1990; Cahusac et al. 1989; Jung and McNaughton 1993; Leonard and McNaughton 1990; O'Mara et al. 1994; Rolls and O'Mara 1993; Rolls et al. 1989; see Treves and Rolls 1994). It is therefore proposed that these fundamentally different constraints account for the different sparsenesses of representations found in the high-order sensory cortices such as the temporal cortical areas described here, and in memory systems such as the hippocampus. In the sensory cortex a relatively distributed representation may be used to optimize discriminative ability. In memory systems, much more sparse representations may be used to maximize the number of memories that can be stored. [We note that many of these cells have other properties that make them suitable for discrimination between faces, including invariance with respect to size, spatial frequency, translation, and even in some cases view (see Rolls 1992a, 1994a; Rolls and Tovee 1994b).]

This implies that in the pathways by which information is transferred to memory systems, there should be a system that increases the sparseness of the representation. For the hippocampus the dentate granule cells may contribute to this function (Rolls 1989; Treves and Rolls 1992, 1994). It is suggested that the lateral nucleus of the amygdala, which receives sensory input from the temporal cortical visual areas, may perform this function and then pass a sparse representation to the actual associative (e.g., visual-taste) neurons in other parts of the amygdala.

This discussion on the sparseness of representation provided by these face-selective cells runs counter to the possibility that they are very specifically tuned and provide a cardinal or "grandmother cell" type of very sparse representation (Barlow 1972). Instead, the data presented in this paper indicate that they are very selective, in that they respond rather selectively to stimuli within the class faces, and provide little information about stimuli that are not faces. However, within the class for which they encode information (faces), the representation is very distributed, implying great discriminative capacity, including the representation of small differences between faces presented simultaneously. It will be of interest in future studies to determine the sparseness of the representation of nonface stimuli provided by other neurons in the temporal cortical visual areas.

The broad tuning, or distributed representation, of the neurons described here would only implement an efficient system for implementing pattern recognition and discrimination if each neuron were tuned differently to the set of stimuli. This does appear to be the case, as shown by the different profiles of different neurons to the same set of stimuli (see, e.g., Figs. 2–6) (and see Baylis et al. 1985) and by the

demonstration in the multidimensional scaling and cluster analysis (Figs. 12 and 13) that the different face stimuli are well separated by the *population* of neurons. In information theory terms, if the neurons were all tuned in the same way to the set of stimuli, then increasing the number of neurons over which the information is calculated would add little to the total information available from the population about which stimulus had been seen. In fact, our current analyses (E. T. Rolls, M. J. Tovee, and A. Treves, unpublished observations) are showing that as neurons are added to the sample, then the information available about the stimuli increases essentially in proportion to the number of neurons in the sample (up to the sample sizes of 10 neurons for which the computation has been possible to date). This is quantitative evidence that an efficient distributed representation is provided by these neurons.

The distributed representation provided by these neurons would, as noted above, be very appropriate for fine discrimination, because the responses of a large number of cells would alter by graded amounts when the details of the input stimulus changed. Such a distributed representation would also have the useful property of graceful degradation, in that if individual neurons in the representation were lost, the representation provided by the ensemble would not be greatly impaired.

The approach described in this paper provides objective methods for specifying what is represented in a brain region. One method is to specify the aspects of the input stimuli about which information (in information theoretic terms) is provided. In the present case, information about which face was seen, but not about which nonface was seen, was provided by the population of neurons. This is reminiscent of the "information bearing parameter" approach of Suga (1989), with the information being specified quantitatively. Another application of this method had been to the demonstration that the responses of the neurons described here are translation invariant, in that in their activity information about which face was seen, but not about where the face was in the visual field, is provided (Rolls et al. 1994). An important concept is what is made explicit in the neuronal responses. In the case of the neurons described here, information about which face was seen, but not about which nonface was seen, is made explicit in the neuronal responses. Other methods used here are multidimensional scaling and cluster analysis, which provide an indication about how the information available in the responses of the neurons separates the stimuli and enables discrimination between them (Figs. 12 and 13).

We are very grateful for helpful comments and assistance provided by Drs. A. Treves and R. Baddeley.

This research was supported by Medical Research Council Grant PG8513790 to E. T. Rolls.

Present address of M. J. Tovee, Dept. of Psychology, University of Newcastle, Ridley Bldg., Newcastle upon Tyne, NE1 7RU, UK.

Address reprint requests to E. T. Rolls.

Received 10 March 1994; accepted in final form 14 September 1994.

REFERENCES

BAIZER, J. S., UNGERLEIDER, L. G., AND DESIMONE, R. Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *J. Neurosci*. 11: 168–190, 1991.

BARLOW, H. B. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1: 371–394, 1972.

BARNES, C. A., MCNAUGHTON, B. L., MIZUMORI, S. J., AND LIM, L. H. Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. *Prog. Brain Res*. 83: 287–300, 1990.

BAYLIS, G. C., ROLLS, E. T., AND LEONARD, C. M. Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Res*. 342: 91–102, 1985.

BAYLIS, G. C., ROLLS, E. T., AND LEONARD, C. M. Functional subdivisions of temporal lobe neocortex. *J. Neurosci*. 7: 330–342, 1987.

BRUCE, C., DESIMONE, R., AND GROSS, C. G. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol*. 46: 369–384, 1981.

CAHUSAC, P. M. B., MIYASHITA, Y., AND ROLLS, E. T. Responses of hippocampal formation neurons in the monkey related to delayed spatial response and object-place memory tasks. *Behav. Brain Res*. 33: 229–240, 1989.

CHURCHLAND, P. S. AND SEJNOWSKI, T. J. *The Computational Brain*. Cambridge, MA: MIT, 1992.

DESIMONE, R. Face-selective cells in the temporal cortex of monkeys. *J. Cognit. Neurosci*. 3: 1–8, 1991.

DESIMONE, R., ALBRIGHT, T. D., GROSS, C. G., AND BRUCE, C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci*. 4: 2051–2062, 1984.

DESIMONE, R. AND GROSS, C. G. Visual areas in the temporal lobe of the macaque. *Brain Res*. 178: 363–380, 1979.

FUKUNAGA, K. *Introduction to Statistical Pattern Recognition*. New York: Academic, 1972.

GROSS, C. G. Representation of stimuli in inferior temporal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci*. 335: 3–10, 1992.

GROSS, C. G., DESIMONE, R., ALBRIGHT, T. D., AND SCHWARTZ, E. L. Inferior temporal cortex and pattern recognition. *Exp. Brain Res. Suppl*. 11: 179–201, 1985.

HASSELMO, M. E., ROLLS, E. T., AND BAYLIS, G. C. The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav. Brain Res*. 32: 203–218, 1989.

HINTON, G. E., MCCLELLAND, J. L., AND RUMELHART, D. E. Distributed representations. In: *Parallel Distributed Processing,* edited by D. E. Rumelhart and J. L. McClelland. Cambridge, MA: MIT Press, 1986, chapt. 3, p. 77–109.

HORNAK, J., ROLLS, E. T., AND WADE, D. Face and voice expression identification and their association with emotional and behavioural changes in patients with frontal lobe damage. 1995.

JUDGE, S. J., RICHMOND, B. J., AND CHU, F. C. Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res*. 20: 535–538, 1980.

JUNG, M. W. AND MCNAUGHTON, B. L. Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus* 3: 165–182, 1993.

LEONARD, B. W. AND MCNAUGHTON, B. L. Spatial representation in the rat: conceptual, behavioral and neurophysiological perspectives. In: *Neurobiology of Comparative Cognition,* edited by R. P. Kesner and D. S. Olton. Hillsdale, NJ: Erlbaum, 1990, p. 363–422.

MAUNSELL, J. H. R. AND NEWSOME, W. T. Visual processing in monkey extrastriate cortex. *Annu. Rev. Neurosci*. 10: 363–401, 1987.

MACRAE, A. W. On calculating unbiased information measures. *Psychol. Bull*. 75: 270–277, 1971.

MERRILL, E. G. AND AINSWORTH, A. Glass-coated platinum-plated tungsten microelectrodes. *Med. Biol. Eng*. 10: 662–672, 1972.

O'MARA, S. M., ROLLS, E. T., BERTHOZ, A., AND KESNER, R. P. Neurons responding to whole-body motion in the primate hippocampus. *J. Neurosci*. In press.

OPTICAN, L. AND RICHMOND, B. J. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J. Neurophysiol*. 57: 132–146, 1987.

OPTICAN, L. M., GAWNE, T. J., RICHMOND, B. J., AND JOSEPH, P. J. Unbiased measures of transmitted information and channel capacity from multivariate neuronal data. *Biol. Cybern*. 65: 305–310, 1991.

PERRETT, D. I., ROLLS, E. T., AND CAAN, W. Visual neurons responsive to faces in the monkey temporal cortex. *Exp. Brain Res*. 47: 329–342, 1982.

RICHMOND, B. J. AND OPTICAN, L. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. *J. Neurophysiol*. 57: 147–161, 1987.

RICHMOND, B. J. AND OPTICAN, L. M. Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. II. Information transmission. *J. Neurophysiol*. 64: 370–380, 1990.

RICHMOND, B. J., OPTICAN, L. M., PODELL, M., AND SPITZER, H. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *J. Neurophysiol*. 57: 132–146, 1987.

ROLLS, E. T. Processing beyond the inferior temporal visual cortex related to feeding, learning, and striatal function. In: *Brain Mechanisms of Sensation,* edited by Y. Katsuki et al. New York: Wiley, 1981a, chapt. 16, p. 241–269.

ROLLS, E. T. Responses of amygdaloid neurons in the primate. In: *The Amygdaloid Complex,* edited by Y. Ben-Ari. Amsterdam: Elsevier, 1981b, p. 383–393.

ROLLS, E. T. Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Hum. Neurobiol. 3*: 209–222, 1984.

ROLLS, E. T. Functions of neuronal networks in the hippocampus and neocortex in memory. In: *Neural Models of Plasticity: Experimental and Theoretical Approaches,* edited by J. H. Byrne and W. O. Berry. San Diego, CA: Academic, 1989, chapt. 13, p. 240–265.

ROLLS, E. T. Neural organisation of higher visual functions. *Curr. Opin. Neurobiol*. 1: 274–278, 1991.

ROLLS, E. T. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philos. Trans. R. Soc. Lond. B Biol. Sci*. 335: 11–21, 1992a.

ROLLS, E. T. The processing of face information in the primate temporal lobe. In: *Processing Images of Faces,* edited by V. Bruce and M. Burton. Norwood, NJ: Ablex, 1992b, chapt. 3, p. 41–68.

ROLLS, E. T. Neurophysiology and functions of the primate amygdala. In: *The Amygdala,* edited by J. P. Aggleton. New York: Wiley-Liss, 1992c, chapt. 5, p. 143–165.

ROLLS, E. T. Brain mechanisms for invariant visual recognition and learning. *Behav. Processes*. In press.

ROLLS, E. T. Learning mechanisms in the temporal lobe visual cortex. *Behav. Brain Res*. In press.

ROLLS, E. T. AND BAYLIS, G. C. Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Exp. Brain Res*. 65: 38–48, 1986.

ROLLS, E. T., BURTON, M. J., AND MORA, F. Hypothalamic neuronal responses associated with the sight of food. *Brain Res*. 111: 53–66, 1976.

ROLLS, E. T., MIYASHITA, Y., CAHUSAC, P. M. B., KESNER, R. P., NIKI, H., FEIGENBAUM, J., AND BACH, L. Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. *J. Neurosci*. 9: 1835–1845, 1989.

ROLLS, E. T. AND O'MARA, S. Neurophysiological and theoretical analysis of how the hippocampus functions in memory. In: *Brain Mechanisms of Perception and Memory: From Neuron to Behavior,* edited by T. Ono, L. R. Squire, M. E. Raichle, D. I. Perrett, and M. Fukuda. New York: Oxford Univ. Press, 1993, chapt. 17, p. 276–300.

ROLLS, E. T., SANGHERA, M. K., AND ROPER-HALL, A. The latency of activation of neurons in the lateral hypothalamus and substantia innominata during feeding in the monkey. *Brain Res*. 164: 121–135, 1979.

ROLLS, E. T. AND TOVEE, M. J. Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc. R. Soc. Lond. B Biol. Sci*. 257: 9–15, 1994a.

ROLLS, E. T. AND TOVEE, M. J. The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the visual field. *Exp. Brain Res*. In press.

ROLLS, E. T., TOVEE, M. J., PURCELL, D. G., STEWART, A. L., AND AZZOPARDI, P. The responses of neurons in the temporal cortex of primates, and face identification and detection. *Exp. Brain Res*. 101: 474–484, 1994.

ROLLS, E. T. AND TREVES, A. The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network* 1: 407–421, 1990.

ROLLS, E. T., YAXLEY, S., AND SIENKIEWICZ, Z. J. Gustatory responses of single neurons in the orbitofrontal cortex of the macaque monkey. *J. Neurophysiol*. 64: 1055–1066, 1990.

SELTZER, B. AND PANDYA, D. N. Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res*. 149: 1–24, 1978.

SMITH, D. V. AND TRAVERS, J. B. A metric for the breadth of tuning of gustatory neurons. *Chem. Senses* 4: 215–229, 1979.

SUGA, N. Principles of auditory information-processing derived from neuroethology. *J. Exp. Biol*. 146: 277–286, 1989.

TANAKA, K., SAITO, H.-A., FUKADA, Y., AND MORIYA, M. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol*. 66: 170–189, 1991.

TOVEE, M. J. AND ROLLS, E. T. Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. *Visual Cognit*. In press.

TOVEE, M. J., ROLLS, E. T., AND AZZOPARDI, P. Translation invariance and the responses of neurons in the temporal visual cortical areas of primates. *J. Neurophysiol*. 72: 1049–1060, 1994.

TOVEE, M. J., ROLLS, E. T., TREVES, A., AND BELLIS, R. P. Information encoding and the responses of single neurons in the primate temporal visual cortex. *J. Neurophysiol*. 70: 640–654, 1993.

TREVES, A. Graded-response neurons and information encodings in autoassociative memories. *Phys. Rev. A* 42: 2418–2430, 1990.

TREVES, A. AND PANZERI, S. The upward bias in measures of information derived from limited data samples. *Neural Computation*, 1995.

TREVES, A. AND ROLLS, E. T. What determines the capacity of autoassociative memories in the brain? *Network* 2: 371–397, 1991.

TREVES, A. AND ROLLS, E. T. Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* 2: 189–199, 1992.

TREVES, A. AND ROLLS, E. T. A computational analysis of the role of the hippocampus in memory. *Hippocampus* 4: 374–391, 1994.

WEISKRANTZ, L. AND SAUNDERS, R. C. Impairments of visual object transforms in monkeys. *Brain* 107: 1033–1072, 1984.

YOUNG, M. P. AND YAMANE, S. Sparse population coding of faces in inferotemporal cortex. *Science Wash. DC* 256: 1327–1330, 1992.