Exp Brain Res (2003) 150:142–153
DOI 10.1007/s00221-003-1417-5

RESEARCH ARTICLE

Bruno B. Averbeck · David A. Crowe · Matthew V. Chafee · Apostolos P. Georgopoulos

Neural activity in prefrontal cortex during copying geometrical shapes

II. Decoding shape segments from neural ensembles

Received: 3 July 2002 / Accepted: 7 January 2003 / Published online: 1 April 2003

Abstract We trained two monkeys to draw copies of geometrical shapes (e.g. squares, triangles) using a joystick, and found that several variables describing the arm trajectories were encoded in the activity of individual prefrontal neurons (Averbeck et al. 2003). Copy trajectories were drawn as sequences of segments, identified by the serial order in which they were drawn and the shape that they together produced. Here we use linear discriminant analysis to test how well the segments of copied shapes could be decoded from the neural activity patterns of small ensembles (3–22 neurons) of simultaneously recorded cells in prefrontal cortex. Using this analysis, the proper segment (drawn by the monkey) was correctly decoded from the ensemble activity pattern during the drawing of that segment in 60–80% of the cases when the largest ensembles were considered. The information transmitted by these ensembles, as well as by single neurons, was also calculated. We found that the information transmitted by the ensembles increased on average with the number of neurons they contained. Each neuron conveyed information about multiple segments within the drawing trajectory, suggesting that neurons were 'broadly tuned' across segments and that the neural code of segment was distributed.

Keywords Neural decoding · Discriminant analysis · Information theory · Prefrontal cortex · Neural ensemble

B. B. Averbeck · D. A. Crowe · M. V. Chafee · A. P. Georgopoulos (✉)
Brain Sciences Center (11B),
Veterans Affairs Medical Center,
One Veterans Drive, Minneapolis, MN 55417, USA
e-mail: omega@umn.edu
Tel.: +1-612-7252282
Fax: +1-612-7252291

B. B. Averbeck · D. A. Crowe · A. P. Georgopoulos
Graduate Program in Neuroscience,
University of Minnesota,
Minneapolis, MN 55455, USA

M. V. Chafee · A. P. Georgopoulos
Department of Neuroscience,
University of Minnesota Medical School,
Minneapolis, MN 55455, USA

A. P. Georgopoulos
Department of Neurology,
University of Minnesota Medical School,
Minneapolis, MN 55455, USA

Introduction

In the companion report to this article, we present evidence that multiple variables relating to movements of the arm were encoded in the activity of prefrontal neurons as monkeys drew copies of geometrical shapes (Averbeck et al. 2003). The copy operation was characterized as a serial-order process in which the proper segments must be drawn in the proper order to produce the correct shape. Using a stepwise multiple linear regression analysis we were able to determine that individual prefrontal neurons encoded the position of the hand, the direction of movement, movement speed, the serial order of a movement in a sequence, and the shape of the template object (Averbeck et al. 2003). Here we take the ensemble activity pattern in prefrontal cortex as a point of departure, and use it to decode the behavioral variables relating to copy performance. Specifically, we predict the identity of the drawn segment from the ensemble activity pattern at the time the segment was drawn. This prediction is compared with the segment the monkey drew to determine how accurately segment identity can be extracted from the neural data. Using this technique, we sought to quantify the degree to which neural representations in prefrontal cortex are local or distributed (Barlow 1972; Hinton et al. 1986; Foldiak and Young 1995; Thorpe 1995). A local representation is characterized by activation of a few, narrowly tuned neurons. A distributed representation, in contrast, is characterized by the activation of a large number of broadly tuned neurons. To investigate this question in relation to the neural representation of shape segments, we used an information theoretic analysis (Shannon 1948)

based on the performance of the segment classification procedure. Information theory has been used in the visual system (Optican and Richmond 1987; McClurkin et al. 1991), as well as in the motor system (Georgopoulos and Massey 1988), to provide a measure of how many levels of a behavioral factor (i.e. either a set of visual stimuli or motor responses) can be discriminated by the activity of single neurons or small ensembles. In this case, we quantified how many bits of information about shape segments were transmitted by single neurons and ensembles. This enabled us to characterize segment tuning functions in information space.

If the segment representation proved to be distributed in the prefrontal cortex, for example, tuning functions would be broad, each prefrontal neuron would convey some information about multiple segments, and each segment would be represented by a large number of neurons.

Materials and methods

The companion paper contains a more thorough description of the experimental procedures (Averbeck et al. 2003). Here we describe mainly the analyses that are specific to this paper.

Animals

Two male rhesus macaques, *Macaca mulatta* (M157 and M555, 8–10 kg body weight), were used in the experiments. Care and treatment of the animals during all stages of the experiments conformed to the Principles of Laboratory Animal Care (NIH publication no. 86–23, revised 1995). All experimental protocols were approved by the appropriate institutional review boards.

Task and data preprocessing

The monkeys were trained to use a joystick-controlled cursor to copy shapes on a projection screen. Monkeys moved the cursor into a start dot appearing on the left half of the screen to start the trial. After a waiting period (1–2 s), a template shape appeared on the right half of the screen. This was the 'go' signal. Monkeys were required to draw an accurate copy of the shape in order to be rewarded. For each set of simultaneously recorded cells, monkey 157 (M157) made 30 copies each of four shapes: a triangle, a square, a trapezoid and an inverted triangle. Monkey 555 (M555) attempted 60 trajectories of each shape for each ensemble of simultaneously recorded cells. For monkey M555, correct trials were assessed offline.

Each shape was drawn as a sequence of discrete segments, each segment being characterized by a bell-shaped velocity profile, and corresponding to one of the sides of the template. We found the segments within the trajectories of each shape by determining the points where the acceleration went from negative to positive. These acceleration zero-crossings correspond to minima in the velocity, where drawing slowed near the corners of each shape. For each set of simultaneously recorded neurons in prefrontal cortex, we determined the ensemble activity pattern during the drawing of each segment of each shape. The ensemble activity pattern was defined as a vector of firing rates of all the cells in the ensemble, averaged across the period during which each segment was drawn.

Multivariate analysis of variance across segments

We first sought to test whether ensemble activity patterns varied significantly across segments. To determine whether this was the case, we applied a multivariate analysis of variance (MANOVA) (Johnson and Wichern 1998) to the ensemble activity patterns. The MANOVA was implemented by the IMSL statistical libraries (Visual Numerics, Houston, TX, USA). A separate MANOVA was performed for each neural ensemble and each shape. The dependent variables in each analysis were the firing rates of the neurons in the ensemble. The single independent variable was the segment within that shape, defined here by its serial order. We used Wilk's lambda and its associated probability value to determine the statistical significance for rejecting the null hypothesis that ensemble activity did not vary across segments.

Classification of segment using discriminant analyses

Discriminant analyses were carried out using the DDSCRM routine of the IMSL statistical libraries. This routine calculates posterior probabilities of class (i.e. category) membership using standard statistical discriminant analysis techniques (Johnson and Wichern 1998; Duda et al. 2001). Discriminant analysis assigns a probability ('posterior probability') to each of the segments in the set (across all shapes) given the pattern of activity in the neural ensemble. The segment with the highest posterior probability is taken as the best estimate of the motor output on the basis of the neural activity. It can then be ascertained how well this decoding algorithm operates by comparing the predicted segment with the one the monkey actually drew. Thus, an error might predict a segment in the wrong serial order or in the wrong shape. Conversely, correct classification indicates that the segment in the correct serial order and shape was identified. The set of possible segments for this classification was taken to be any of the segments the monkey produced in any of the shapes it copied (15 different segments across four shapes for M157 and six segments across two shapes for M555).

In our analyses, we assumed equal prior probabilities for all classes. Therefore, the posterior probabilities were calculated by modeling the relationship between the neural activity and each segment using a multivariate normal probability distribution. The variables of the multivariate normal distribution were the activities of the simultaneously recorded cells. To carry out the analysis, separate density functions were estimated for each category to be classified, which in our case were the set of segments across all the shapes. Each density function represented the estimated mean and standard deviation of the neural activity, while the monkey was drawing a particular segment. The density functions were the following:

$$p_i(\mathbf{x}) = \frac{1}{2\pi^{d/2}|\Sigma_i|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu_i)^t \Sigma_i^{-1}(\mathbf{x}-\mu_i)\right] \quad (1)$$

where p_i is the probability that the vector of cell firing rates $\mathbf{x}$ came from category (i.e. segment) i, Σ is the covariance matrix of cell firing rates, the $|\,|$ symbol indicates the determinant, μ_i is the mean firing rate vector during the drawing of segment i, and d is the number of neurons used in the classification procedure, equivalent to the dimensionality of the gaussian function. Taking the natural logarithm of Eq. 1 and eliminating terms common to all categories, we derived the following equation for the quadratic discriminant function:

$$d_i(x) = -\frac{1}{2}(\mathbf{x}-\mu_i)^t \Sigma_i^{-1}(\mathbf{x}-\mu_i) - \frac{1}{2}\ln|\Sigma_i| \quad (2)$$

Since this is no longer a probability, the $p(x)$ is replaced by $d(x)$ representing the relative score of the vector $\mathbf{x}$ for this category, often called the discriminant function. This is the most complete model that, in contrast to linear discriminant analysis, does not assume equality of the covariance matrices for the categories. If equal covariance matrices are assumed then they can be pooled, yielding the following linear discriminant function:

$$d_i(x) = -\frac{1}{2}(\mathbf{x} - \mu_i)^t \Sigma_p^{-1}(\mathbf{x} - \mu_i) \quad (3)$$

where the p subscript on the covariance matrix (Σ) indicates the use of the pooled covariance matrix. The squared terms in Eq. 3 can also be eliminated, as the covariance matrices are assumed to be equal, and thus only terms involving the means for each category remain. The classification procedure calculates d for each trial for each category, and it classifies a given observation as belonging to the category for which it has the highest discriminant score. This yields a classification matrix with rows representing the known category membership (i.e. the segment being classified) and columns representing the category (i.e. the segment) to which each trial was classified. Therefore, correctly classified trials contribute to the diagonal of the classification matrix, and incorrectly classified trials contribute to the off-diagonal cells of the matrix.

Within the discriminant analysis, the definition of the means and the covariance matrices for each group were performed using two methods: reclassification, and leave-one-out (or jackknife). In the reclassification method, all trials are used to calculate the means and covariance matrices, and each trial is classified based on the complete model. In the leave-one-out method, the means and covariance matrices are calculated without the trial to be tested, and then the trial is classified based on the model *estimated without it*. This method assesses the ability of the model to generalize. We compared results obtained using each of these analyses in this study.

Information transmitted

The information about a segment carried by an ensemble was calculated using standard techniques (Shannon 1948; Abramson 1963; Georgopoulos and Massey 1988). In brief, the classification matrix was considered as a stimulus–response (SR) matrix, where the 'stimulus' was the known category membership of the segment, and the 'response' was the segment to which a given trial was classified. The elements of the matrix were counts of the times for which a given segment (stimulus) was classified in a particular category (response). These simple counts were then used to calculate information using the following equations:

$$T = H(S) + H(R) - H(S,R) \quad (4)$$

$$H(S) = \log_2 N - \frac{1}{N}\sum_s n_s \log_2 n_s \quad (5)$$

$$H(R) = \log_2 N - \frac{1}{N}\sum_r n_r \log_2 n_r \quad (6)$$

$$H(S,R) = \log_2 N - \frac{1}{N}\sum_{s,r} n_{s,r} \log_2 n_{s,r} \quad (7)$$

where T is the information transmitted, $H(S)$ is the uncertainty in the stimulus, $H(R)$ is the uncertainty in the neural response, $H(S,R)$ is the joint stimulus–response uncertainty, N is the total number of trials, and n is the number of trials in a particular bin. These computational formulas were derived from the following standard formulas for calculating information:

$$T = \sum_{s,r} P(s,r) \log_2 \frac{P(s,r)}{P(s)P(r)} \quad (8)$$

$$H = p_i \log_2 p_i \quad (9)$$

$$p_i = \frac{n_i}{N} \quad (10)$$

where p denotes probability. Substitution of Eq. 10 into Eq. 9 leads to Eqs. 5, 6 and 7 above, which can be algebraically transformed into Eq. 8. The redundancy was calculated using the following formula (Gochin et al. 1994):

$$R = \frac{T_{is} - T_{ie}}{T_{is}} \quad (11)$$

where T_{is} is the sum of the information carried by single cells considered in isolation, and T_{ie} is the information carried by the single cells considered as an ensemble.

We also calculated the partial information transmitted by a cell about each segment. The total information calculated above is the average of the partial information carried about each segment. Thus the partial information about each stimulus is:

$$T(s) = \sum_r P(r|s) \log_2 \frac{P(s,r)}{P(s)P(r)} \quad (12)$$

When calculating information transmission with the data sets commonly obtained in behaving monkey experiments, the small-sample bias in the information calculation must be corrected (Optican et al. 1991; Treves and Panzeri 1995; Panzeri and Treves 1996; Golomb et al. 1997). The small-sample bias results in an overestimation of the information when an insufficient number of samples are used to generate the classification matrix. We calculated the Bayesian estimate of the bias term according to Panzeri and Treves (1996), and subtracted it from all information measures. This approach was found to perform well in a Monte Carlo simulation using the same size SR matrix and the same number of trials.

Another approach that can be used to calculate mutual information is to use untransformed activity rates. These rates would be binned by dividing the range of responses, or some function of the responses (e.g. principal components) into a number of bins. The SR matrix is then calculated as the number of times a particular response occurs when a specific segment is being drawn. However, these approaches do not adapt well to the analysis of ensembles because the number of bins grows exponentially with the number of cells to be considered. This leads to a dramatic increase in the small-sample bias problem. Furthermore, the number of bins considered has to be chosen arbitrarily. The approach we used to estimate the information was to first use a parametric method to decode the neural response rates, and then use these decoded estimates of the segment being drawn as the values in the SR matrix.

Since we did not calculate directly the information in the activity rates, we necessarily underestimated the actual mutual information between the neuron and the segments. In order to directly estimate the information contained in the firing rates, large numbers of trials would be necessary (Reinagel and Reid 2000), and these would be difficult to obtain in behaving monkey experiments. The information in the raw firing rates is greater than or equal to the information in the decoded signal, and equality only holds when the decoding technique attains the limit defined by the Fisher information (Brunel and Nadal 1998).

Multidimensional scaling (MDS)

MDS analysis was performed using SPSS (release 10, SPSS Inc., Chicago, IL, USA), and was carried out on the mean classification matrix of the population derived from the discriminant analysis. Thus, the rows of the matrix corresponded to the known category membership of each trial, and the columns corresponded to the category to which the trial was classified. Each cell of the matrix was then converted to a difference measure by re-expressing it as a probability of classification and then subtracting it from 1. This matrix was subsequently analyzed by the MDS algorithm of SPSS using an interval-metric model.

Results

Discriminant analysis was used to decode the identity of drawn segments from the activity of neuronal ensembles in prefrontal cortex. Information theory was also applied to calculate the information transmitted by single neurons and small neuronal ensembles about shape segments. The average ensemble size was eight cells (range 3–22, *N*=68 ensembles), and was distributed in this dataset as shown in Fig. 1. Ensembles were recorded while monkeys copied triangles, squares, trapezoids and inverted triangles.

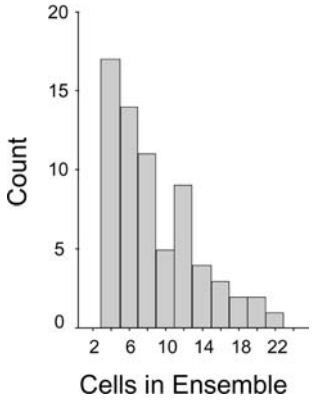

Fig. 1 Distribution of size of ensembles recorded in the prefrontal cortex of two male rhesus macaques. The largest ensemble of simultaneously recorded cells was 22; 68 ensembles were recorded, 34 in monkey M555 and 34 in monkey M157

Analysis of ensemble activity across shape segment

We were interested in investigating the relationship between ensemble activity and the execution of segments of shapes. The first step was to divide the shape trajectories into segments (see Methods section above, and Averbeck et al. 2003). This allowed us to calculate discharge rates for every cell during the drawing of each segment of a shape. We began the analysis of the neural data by determining whether ensembles altered their firing rates during the drawing of different segments within a shape. For that purpose we performed a one-way MANOVA for each ensemble and shape, with the firing rates of the cells in the ensemble as the dependent variable. Overall, the activity of every ensemble in the sample varied significantly across the segments of at least one shape. The vast majority varied across segments of all shapes (Table 1). This indicated that the ensembles were encoding the segments of the shapes in the copy task.

Table 1 Percentages of ensembles related to segments of the shapes in the copy task. Results for trapezoid and inverted triangle reflect only monkey M157, whereas those for the remaining shapes reflect both monkeys

Shape	Significant ensembles
Triangle	66/68 (97%)
Square	66/68 (97%)
Trapezoid	34/34 (100%)
Inverted triangle	34/34 (100%)
Any shape	68/68 (100%)

Decoding shape segment

The next step was to determine whether we could use the neuronal signal to predict which segment was being drawn; specifically, we wanted to decode the signals carried by the ensembles regarding the segments of the shapes. Qualitatively, if a categorical variable is encoded by a neural ensemble (as assessed in the MANOVA), the pattern of firing rates in the ensemble will differ significantly among the different values of the variable. We took advantage of this fact in order to decode the signal in the ensemble, applying discriminant analysis to the ensemble firing rate patterns to predict which segment was being drawn. For example, if the encoding analysis indicates that the patterns of firing rate differ for the different segments of the shapes, the discriminant analysis can determine which segment is being drawn from the pattern of activity in the neuronal ensemble. This is done by comparing the current ensemble activity pattern with the mean ensemble activity pattern for each segment: the analysis predicts the current segment being drawn on the basis of its statistical proximity to a segment's mean activity pattern.

Discriminant analysis can include only linear components of the cell firing rate vectors (linear discriminant analysis, LDA), or it can include both linear and quadratic components (quadratic discriminant analysis, QDA). We wanted to assess the relative effectiveness of the two methods on our data, given that each method possesses different advantages and drawbacks. Specifically, LDA operates on the assumption that the covariance matrices of the cells in the ensemble do not differ significantly across categories (i.e. segments). However, LDA is relatively robust (meaning that discriminant functions generalize well to new data), and may still perform well in cases for which there may not be enough trials to accurately estimate the covariance matrices for each individual category, a case of the bias–variance trade-off (Bishop 1995). In this case, the pooled covariance matrix could be a better estimate of the population covariance matrix than covariance matrices calculated separately for each segment. On the other hand, QDA is generally less robust and, in addition, it could not be performed on all our ensembles because the individual segment covariance matrices were occasionally non-invertible, due to the lack of neuronal firing during the drawing of some segments of certain shapes. In general, QDA performed better than LDA in reclassification, but it performed worse than LDA in leaving-one-out analyses. The improved performance of QDA in reclassification was probably due to the extra parameters in the quadratic model. The decrease in QDA performance when the leaving-one-out method was used indicates that it is less robust. This could be due to the

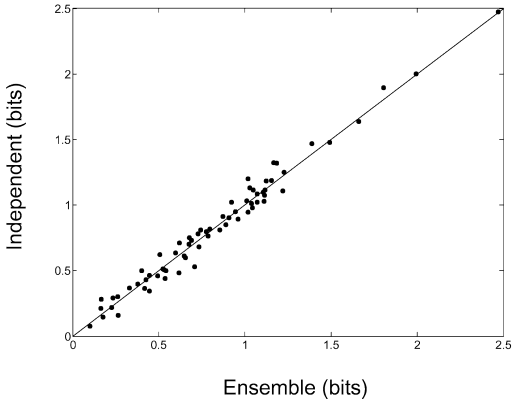

Fig. 2 Comparison of bits of information for surrogate (independent) and ensemble datasets. Surrogate datasets were generated by shuffling neural response for single cells across trials. Correlation in the trial-to-trial variability was thus eliminated in the surrogate dataset

limited number of trials available (30 per shape × number of segments) versus the number of cells simultaneously recorded (up to 22). In the remainder of the analyses we used the more robust LDA and computed all results using the leaving-one-out method.

The ensemble analyses that follow were all done on simultaneously recorded cells. An interesting question is what effect this simultaneous recording had on the performance of the decoding model. We tested this by permuting the neural responses across trials within datasets of simultaneously recorded neurons. The responses in the resulting surrogate dataset were independent. Figure 2 shows the performance of the decoding model on the original (ensemble) and surrogate dataset (independent). It can be seen that there is very little change in performance between the original and surrogate datasets.

Performance of decoding algorithm

The classification performance of LDA applied to neuronal ensembles is shown in Fig. 3A. The neural data from the two monkeys in this experiment are segregated (ensembles from M555 in *red*, from M157 in *blue* in Fig. 3). One monkey (M157) mastered a larger repertoire of shapes, and consequently produced a larger number of different shape segments. One consequence of this was that the LDA classification employed a larger number of possible categories (segments) in this case. We found that the overall performance of the classification was related to the number of categories included (see below). For both animals, however, it is evident that classification performance of ensembles increased with the number of neurons they contained. In the case of M157, who drew 15 different segments across four different shapes, classification performance approached 60% correct for some of the larger ensembles. This performance is considerably in excess of the rate predicted from random classification (7% in the case of 15 categories). In the case of M555, who drew six different segments, correct classification rates approached 80%, where random classification would have been correct in 17% of cases (with six categories).

Information transmitted by neuronal ensembles and redundancy

We calculated the bits of information about the segments of the shapes carried by ensembles of various sizes. Figure 3B shows the information carried by each ensemble for the segments across all shapes. The most information carried by a single ensemble was 1.39 bits for M555 and 2.47 bits for M157. For M555, the maximum possible number of bits was $\log_2(6)=2.58$, and for M157 the maximum possible number of bits was $\log_2(15)=3.9$. It can be seen in both cases that the amount of information carried was less than the maximum possible. Figure 3C demonstrates the linear relationship between the percentage of segments correctly classified in the discriminant

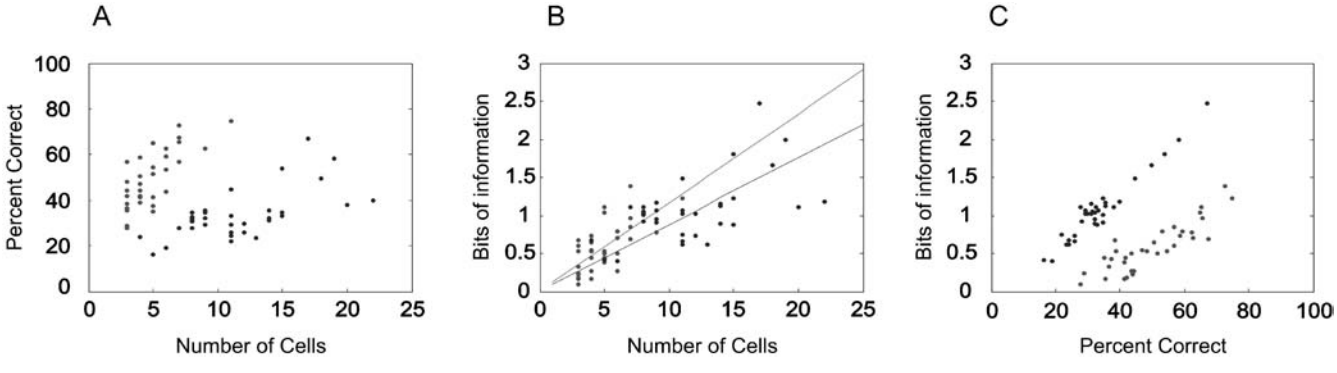

Fig. 3A–C Classification rates and bits of information versus the ensemble size. **A** Percentage correct classification rate for each shape versus the number of cells in each ensemble. **B** Bits of information versus the number of cells in each ensemble. **C** Relationship between the percentage correct and the bits of information for each ensemble

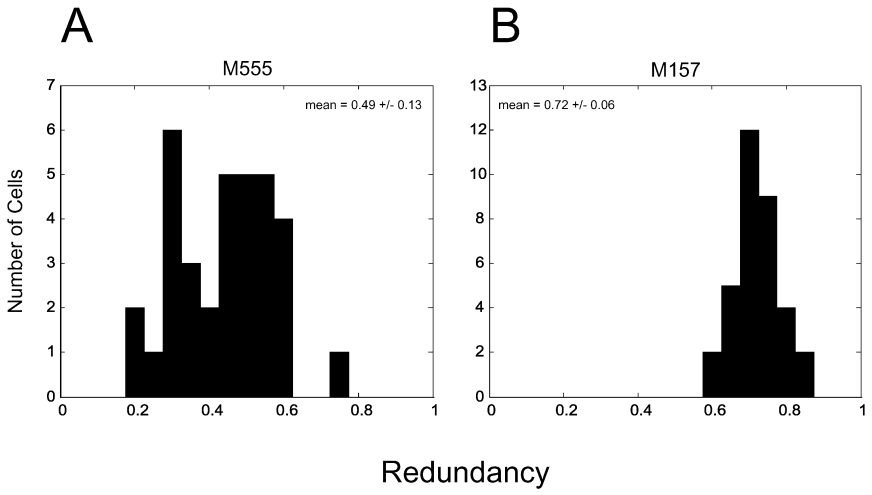

Fig. 4A,B Redundancy in neuronal ensembles of monkeys M555 (**A**) and M157 (**B**). Redundancy is a measure of the amount of information that is shared among cells in an ensemble. When cells which are carrying redundant information are combined, the total information of the ensemble is less than the sum of the information of the independent neurons

analysis and the information transmitted. This is not an obligatory relationship. The information transmitted is calculated from the entire confusion matrix, whereas the diagonal of this matrix gives the percentage correct performance.

A previous study has reported a ceiling effect on the amount of information transmitted by an ensemble, evidenced by a deviation from linearity in the plot of information transmitted versus number of cells in an ensemble (Rolls et al. 1997). The ceiling effect occurs as the information encoded by an ensemble approaches the theoretical maximum. We saw little evidence for this in our analyses — our ensembles did not reach the maximum amount of information that could be transmitted about the segments being drawn. Thus, the data were restricted to the linear portion of the relationship between number of cells and bits of information. We performed a linear regression analysis to quantify the relationship between the information carried by an ensemble (dependent variable) and the number of cells in the ensemble (independent variable). This analysis was performed across all ensembles. The slopes of the lines in these regressions indicated how much information each cell transmitted on average as part of an ensemble. The mean amount of information per neuron carried about all the segments across shapes was, on average, 0.12 bits for M555 and 0.09 bits for M157.

We also calculated the information carried by single cells when considered separately from the ensembles. The sum of the information carried by single cells is expected to exceed that carried by the ensemble due to the redundancy of coding within populations of neurons. The average amount of information carried by single cells considered independently was 0.21 bits for M555 and 0.34 bits for M157. We calculated redundancy (see Methods section) separately for each ensemble of neurons recorded. Redundancy values range from 0 (no redundancy) to 1 (full redundancy). A value of 0 indicates a state in which cells carry completely independent (i.e. uncorrelated) information. A value of 1 indicates a state in which all cells convey exactly the same (perfectly correlated) information. The distribution of redundancies is shown in Fig. 4. The mean redundancy for M555 was 0.44±0.13 and for M157 it was 0.72±0.06 (means ±SD). This indicated that there was considerable redundancy among our population of cells, particularly in the case of M157. In sum, there is an increase in the information carried by ensembles with the addition of more neurons. However, it is not the case that the total information carried by an ensemble is equal to the sum of the information transmitted by its constituent neurons. The total is less than this sum due to redundancy.

Partial information transmitted about individual segments

We assessed whether single cells carried information about more than one segment by calculating the partial information, i.e. how much information was transmitted about each of the segments in the set. After calculating the partial information about all segments for each cell, we sorted the segments for each cell in descending order based upon the amount of information transmitted about the segment. This yielded a tuning function in information space for each cell. Figure 5 plots the population information tuning function averaged across cells and segments. This function indicates the amount of information transmitted by a prefrontal neuron, on average, about segments within a given 'information rank' (i.e. from the segment about which each cell transmitted the most information to the segment about which each cell transmitted the least information). In M157, cells transmitted an average of 0.94 bits about the 'preferred' segment in information space, i.e. the segment for which they transmit the most information, and in M555 they transmitted about 0.58 bits of information. It can be seen that ensembles transmit information about a number of different shape segments (Fig. 5).

Neuronal ensembles recorded in monkey M157 transmitted more information than neuronal ensembles in M555. To test the possibility that this difference was due to the different number of segments drawn by the two

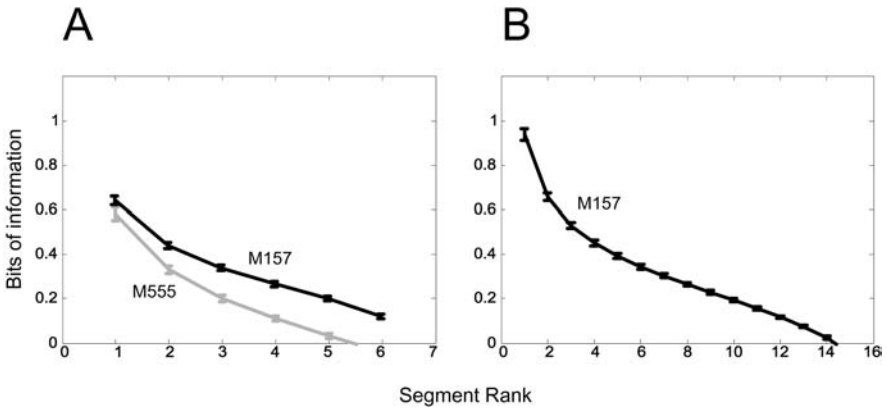

Fig. 5A,B Mean information tuning curves for single cells. **A** Information tuning curve for monkey M555 and partial information about the best six segments for monkey M157. **B** Information tuning curve for original dataset for M157. Data points show means ±SEM (n=34 for each of M555 and M157). Information transmitted about each segment was calculated for each cell. The segments were then rank-ordered by the amount of information transmitted. Segment 1 corresponds to the segment on which each cell transmitted the most information. The last segment corresponds to the segment on which the cell transmitted the least amount of information

monkeys we calculated the partial information about the best six segments for M157 (equating the number of segments across the two monkeys), for each cell. Most of the difference between the amounts of partial information evident in the original analysis (Fig. 5) could be ascribed to the larger number of categories in the analysis of data from M157. When a smaller subset was considered, the mean partial transmitted information for the two monkeys was quite similar (Fig. 5A).

Partial information transmitted as a function of number of segments considered

In order to obtain a more general relation between transmitted information and the total number of segments in the analysis we calculated a family of information tuning curves for each neuron by subsampling the original 15 segments for M157 and six segments for M555. The procedure was as follows. For each cell the information about all segments was calculated. Next, the segments were rank-ordered by their partial information, as above. The segment about which a cell transmitted the least information was then discarded, and the analysis repeated. The subsampling produced a family of information tuning curves for each cell. These were then averaged across cells to obtain the population information tuning curves shown in Fig. 6. Each curve characterizes the information transmitted by a prefrontal neuron across segments when a different total number of segments were included in the analysis. The number of segments represented in the curves ranges from 3 (leftmost curve) to 15 (rightmost curve). The curves with six or fewer categories represent combined data from M555 and M157, those with greater than six categories reflect the data of M157 only. The curves were similar in shape but shifted with respect to one another. Figure 7 illustrates the maximum information as a function of the number of segments considered,

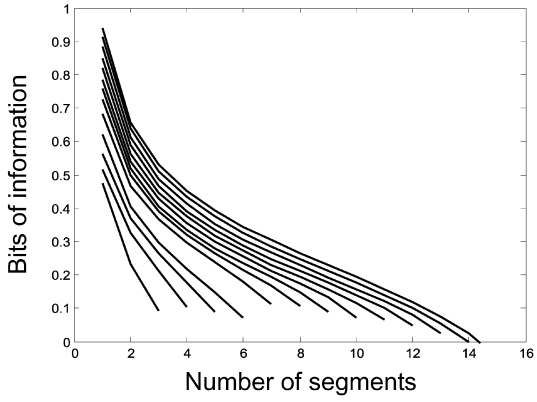

Fig. 6 Family of information tuning curves derived for different numbers of categories (segments)

as well as the mean slope of the information tuning function. Figure 7A shows that the maximum information increased roughly linearly as a function of the number of segments being considered. Correspondingly, the mean slope plotted in Fig. 7B (i.e. the mean of the change in information with the segment rank) decreases as a function of the number of segments considered, but asymptotes for set sizes of 10 or 12 segments. This suggests that the slope remains relatively constant for large set sizes.

Factors contributing to modulation of ensemble activity

The MANOVA and decoding techniques indicated that ensemble activity patterns were systematically related to shape segments. These analyses did not themselves define which attributes of those segments were important factors influencing firing rate. Each segment was characterized by a combination of variables, including shape,

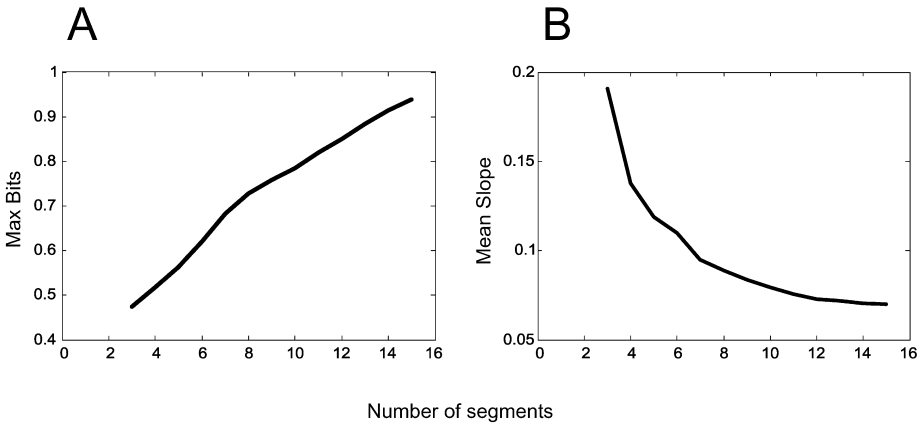

Fig. 7A,B Plots of the increase in the maximum information transmitted (**A**) and the mean slope (i.e. change in information with increasing rank) (**B**). Both functions are monotonically related to the number of categories (segments) being represented. The maximum bits of information increases with an increasing number of segments, and the slope decreases with the number of segments

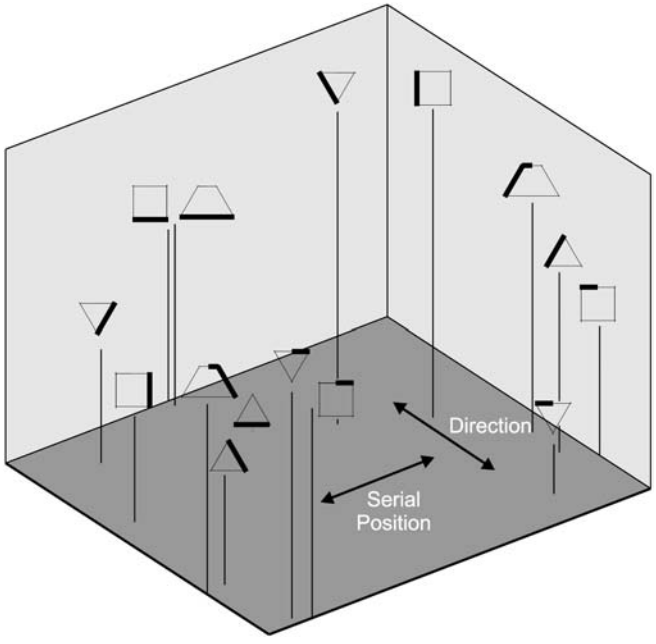

Fig. 8 Multidimensional scaling (MDS) analysis of confusion matrix for segments of all shapes. Each segment across the shapes is represented in the MDS plot. The segment is indicated by the *bold line* on the shape. Thus the segment being represented at the farthest right side of the plot is segment one of the square, and the segment being represented on the farthest left side of the plot is segment 3 of the inverted triangle

serial order, position, and direction, among others. We performed a multi-dimensional scaling analysis to determine which of these segment parameters influenced the neural activity. MDS reduces the dimensionality of this segment space while preserving the relative distances between shape segments. An examination of how segments are grouped in this space allows inferences regarding which attributes of shape segments are heavily weighted in the neural representation. The basis for the MDS analysis was the classification of segments by linear discriminant analysis. The LDA provided the number of times a segment was correctly classified (as itself), or mistakenly classified as each of the other possible segments (yielding the SR matrix). The values in each cell of the mean population SR matrix averaged across ensembles were converted into distance measures in the probability of classification (Methods section). MDS was applied to this matrix of pairwise distances between segments. The MDS analysis was configured to reduce the number of dimensions from 15 to 3. Figure 8 illustrates the results of this analysis for M157, plotting the distances between segments in the 3-dimensional space. The distribution of shape segments within this space suggests that serial position and direction are the factors that primarily modulate the activity of the ensembles (and consequently of the single cells). Thus, segment serial position increases in an orderly manner moving from the right to the left side of the MDS space (along the 'serial position' axis, Fig. 8). Segment direction changes in an orderly manner moving along the 'direction' axis (Fig. 8). Both of these factors were frequently observed to influence firing rate in the stepwise multiple linear regression analysis described in the companion report (Averbeck et al. 2003).

Discussion

Drawing copies of model objects requires a visuomotor process, which is disrupted by damage to several brain regions, including prefrontal cortex (Luria and Tsvetkova 1964; Benton 1968; Gainotti 1985; Koski et al. 2002). Our interest in this series of experiments (Averbeck et al. 2003) has been to investigate neural representations in prefrontal cortex that are related to copy performance and whose loss might produce the spatial deficits that characterize drawing deficits in constructional apraxia. In the companion report, we show that several parameters describing the motor output during the copy task are simultaneously encoded in the activity of individual prefrontal neurons (Averbeck et al. 2003). Here we explore the coding of the segment being drawn in this task by applying linear discriminant analysis in an information theoretic framework to decode segment information from small ensembles of simultaneously recorded neurons. LDA was employed as an initial step to decode the

activity of small ensembles and single cells related to the segment being drawn. The decoded representation was used to construct SR matrices, from which we were able to calculate the information transmitted or mutual information, as well as the partial information transmitted.

Decoding shape segment

The stepwise multiple linear regression analysis employed in the companion paper (Averbeck et al. 2003) demonstrated that metric, sequence and shape variables were encoded in the pattern of activity in neuronal ensembles in prefrontal cortex. In this report, discriminant analysis was used to decode the identity of the trajectory segment being drawn from this ensemble activity. Behavioral variables were in this way extracted from the neural data. In general, ensembles that contained more neurons classified segments correctly more often than did smaller ensembles (Fig. 3). Classification rates were in the range of 60–80% correct when considering the larger ensembles in each monkey. It is noteworthy that classification of this accuracy could be achieved on the basis of the activity of ensembles of between 10 and 20 neurons randomly encountered in a small area of the periprincipalis region of prefrontal cortex. Because a small region of prefrontal cortex was sampled, we do not know from our data whether comparably high classification rates would be obtained from recordings at other prefrontal sites. Thus, neurons whose activity is related to the task are sufficiently dense in prefrontal cortex that by examining the activity of this small number of prefrontal neurons at random enough information can be recovered to give a high degree of confidence regarding the identity of the drawn segment.

Transmitted information in small ensembles and redundancy

The decoding approach we have used provides a lower limit on the mutual information between the neural activity and the segment being drawn. The decoding strategy can only lose information that is present in the original signal. However, it is necessary since brute-force information calculations require large numbers of trials (Reinagel and Reid 2000). We found that single cells considered as part of an ensemble transmitted 0.12 bits for M555 and 0.09 bits of information for M157. When considered independently, single cells carried 0.21 bits for M555 and 0.34 bits for M157. As in the LDA analysis above, these numbers reflect the amount of information transmitted by a randomly encountered cell in prefrontal cortex. We also calculated the redundancy separately for each ensemble. The mean redundancy for M555 was 0.49, and for M157 it was 0.72. This difference perhaps reflects the larger size of the ensembles for M157.

Redundancy arises from similar (i.e. correlated) patterns of activity among cells. In the absence of redundancy among cells, the information carried by the ensemble is the sum of the information carried by single cells (Gawne and Richmond 1993); otherwise, the ensemble information is less than this sum. Thus, two cells responding similarly to all segments will provide redundant information. When considered as part of an ensemble, they will transmit no more information than either of the cells considered in isolation. It should be noted that although we recorded neurons randomly, they were sampled using a spatially limited electrode array (either a 2.5-mm diameter circular electrode array with an inter-electrode distance of ~500 μm or a 5-mm linear electrode array with an inter-electrode distance of ~300 μm). There is evidence that neurons with common tuning properties tend to cluster (Rao et al. 1999; O Scalaidhe et al. 1999; Constantinidis et al. 2001). It may be possible that neurons sampled at greater distances would show more dissimilar tuning properties, and would transmit less redundant information. Given a large enough variation in inter-cell distances within ensembles, it would be of interest to map redundancy as a function of this distance.

Although it appears that the information-carrying capacity of a system is reduced by redundancy, it can nevertheless be useful in a noisy system if the noise is uncorrelated across neurons (Zohary et al. 1994; Panzeri et al. 1999). In this case, the averaging of cells with similar tuning curves can lead to a better estimate of the underlying variable. Furthermore, a balance between redundancy and independence may optimize the amount of information being transmitted by a system with a given amount of noise (Atick and Redlich 1990; Atick et al. 1992). Previous analyses of redundancy reported values between adjacent visual cortical neurons of about 20% (Gawne and Richmond 1993; Gawne et al. 1996), which, though smaller, are in line with the values we report here.

Werner and Mountcastle (1965) were the first to use information theory to analyze neural activity. In their study, analyses were made of the activity of primary afferents in the saphenous nerve. Single primary afferents were found to carry a maximum of about 2.7 bits of information, an amount similar to that human subjects are capable of discerning across many different sensory modalities (Miller 1956). This is considerably higher than the amount of information our cells were carrying. However, other authors have carried out similar information transfer analyses on inferotemporal (IT) neurons (Gawne and Richmond 1993; Gochin et al. 1994). Gawne and Richmond found that their cells carried about 0.23 bits of information when several principal components were used, and 0.14 bits of information when the first principal component (which is similar to firing rate) was used. They used a stimulus set composed of the 32 Walsh functions that describe a 4×4 grid of visual space. Thus the maximum information they could have obtained, assuming equal prior probabilities, was $\log_2(32)=5$ bits. Their results using the first principal component were similar to those reported in this study. Their estimates, however, were computed using pairs of simultaneously

recorded neurons, as opposed to the small ensembles of up to 22 neurons we used in our analysis. Gochin et al. (1994) found that individual cells, assessed as ensembles, carried about 0.26 bits of information on their stimulus set. They used a relatively small stimulus set of only five objects. Their cells were also preselected on the basis of whether or not they responded in the task, whereas we analyzed all cells recorded, without preselection. These examples, using data recorded from IT cortex, suggest a distributed code. IT cortex has often been thought to be the home of the putative grandmother cell (Barlow 1972; Gross et al. 1972). The analyses on IT neural activity reported in these papers, however, suggest that the code for objects in IT cortex is relatively distributed, and, specifically, that single cells carry small amounts of information about a variety of objects. Thus, ensembles of neurons are necessary for accurate stimulus identification. This same conclusion can be drawn from our data, with respect to the coding of the segment being drawn.

Partial information transmitted about individual segments

In order to explicitly explore how much information single cells were transmitting about individual segments, we calculated the partial information for each cell for each segment. Then, we ranked the segments from the most information to the least information within a cell, and calculated the mean across the population of cells for each rank. This determined the mean tuning function in terms of partial information transmitted. The plots of partial information show that single cells carry information about multiple segments (Fig. 5). The difference between the preferred segment about which cells transmit the most information and the second segment in the rank is about 0.3 bits for M157 and 0.15 bits for M555. The negative slope, however, decreases as the relative rank of the segment increases. Thus, the amount of information a cell transmits about segments decreases quickly at first and then begins to asymptote. This curve explicitly represents the way in which information is distributed across single cells.

This representation differs somewhat from the tuning functions usually given for neurophysiological data. Most tuning functions represent the firing rate as a function of the variable of interest. When the data are analyzed in this way, a neuron is considered broadly tuned when it has an elevated firing rate for a large portion of the space covered by the variable of interest, as for instance cosine tuning in the motor cortex (Amirikian and Georgopoulos 2000). However, tuning functions in terms of firing rates do not indicate directly how well a neuron can discriminate different values of a variable. It is the slope of the tuning function that determines how much information a cell provides about a given value of a variable (Pouget et al. 1999), not the magnitude of the firing rate (although magnitude and information can be correlated; McAdams and Maunsell 1999). Recent experiments suggest this perspective is not only statistically valid but that this information is actually utilized by the brain (Schoups et al. 2001). Furthermore, it is important to point out that cells transmit less information about a variable at the peak of a gaussian or sinusoidal tuning function because the slope is small at these points, and therefore the cell changes its firing rate very little in the neighborhood of the peak. This is different from the classical view of neural tuning (called in the extreme the 'labeled line'), which assumes that a cell is most interested in the value of a variable at the peak of its tuning function. It should be noted that the classification procedure and information theoretic analysis applied in the current study assumed that neurons had stable tuning with respect to shape segments over the time that the activity of a given ensemble was recorded. There is evidence that the responsiveness and tuning of prefrontal neurons is variable and under the control of neuromodulatory systems (Williams and Goldman-Rakic 1995; Williams et al. 2002). Fluctuation in neuronal properties of this type would degrade the classification performance and lower transmitted information in the current analysis. For this reason, we may underestimate these parameters.

Discussion about whether single cells carry information about multiple items or only a few items is usually framed in terms of a local versus a distributed representation (Foldiak and Young 1995; Thorpe 1995). A network is said to use a local representation if the activity of a single unit in the system provides a great deal of information about a narrow range of values of a variable that is being represented (Barlow 1972), and very little information outside of this narrow range (or, accordingly, about only a few stimuli in a set if a discrete variable is being considered). In contrast, units in a network using a distributed representation provide an intermediate amount of information about a large range of values of a variable being represented. Such units are characterized by broad tuning functions (Hinton et al. 1986). A network made up of broadly-tuned units results in a larger fraction of the units being active for a given stimulus. In any case, it is necessary to assess activity across the population to estimate the value of a variable in both a distributed and a local representation.

Given that information transfer and level of activation are dissociable quantities, it seems relevant that a network of local units be conceived of as the number of cells *transmitting information* about a given value of a variable. Thus a local network would be one in which only a few cells would be transmitting a large amount of information about a given value of a variable, as opposed to a distributed network in which a large number of cells would be transmitting less information each. The total amount of information transferred could remain the same with both strategies, but the distribution of the information would be different.

Although it may be possible to explicitly discuss localized versus distributed codes in artificial situations (Foldiak and Young 1995; Thorpe 1995), the application of these concepts to real neural coding is difficult. For example, the width of the tuning function seems to be a

good criterion for whether a localized or distributed code is being used. However, the width of a tuning function varies over a constant range, and the selection of a specific point at which a tuning function is characteristic of a local or a distributed signal is somewhat arbitrary. One approach is to assess whether a system becomes more or less distributed with changes in task conditions (Vinje and Gallant 2000). Another approach is to examine the coding from multiple perspectives (Rolls and Tovee 1995; Rolls et al. 1997), and to demonstrate that strict local coding does not hold.

Distributed representations have several interesting computational properties. For example, systems using distributed coding are more resistant to damage. Their representations degrade gradually as units are removed, since the value of a variable is represented across many units. Local systems degrade catastrophically, since the loss of a unit, which represents a given value of a variable, results in the inability of the system to represent that variable, although this consideration ignores the possibility of redundancy in the local system. Another advantage of a distributed system is the ability to represent many more items with a given number of units than a local system (Hinton et al. 1986). A local system is inefficient with respect to the number of units necessary to represent a given number of items, making redundancy even more costly. Even Barlow's original work (Barlow 1972), maintained that the representation of a given concept would be by several 'cardinal' cells, and not by a single 'pontifical' cell.

Partial information transmitted as a function of the number of segments considered

By subsampling the number of categories, we were able to develop segment-tuning curves for various numbers of categories. We found that the segment-tuning curves for the different numbers of categories had a similar shape. The maximum bits of information transmitted was an increasing function of the number of categories, whereas the mean slope was a decreasing function. The exact shape of these curves depends to a certain extent on the shape of the original partial information curves. Thus, if a cell transmits an equal amount of information about all the members of a set of stimuli, adding stimuli to which the cell does not respond will simply result in an increase in the information transmitted about each of the other stimuli because there will be more information to transmit about these stimuli. These curves provide an empirical description of the amount of information about different-sized sets of segments transmitted by the population of neurons from which we recorded.

Potential factors modulating ensemble activity

In order to relate our ensemble analysis to our single cell analysis we also explored the specific variables being represented by the population of prefrontal cells recorded at the ensemble level. Each segment is defined by several variables including speed, direction, position, and segment number. We used MDS to explore the dimensions that accounted for the largest quantity of variance in our prefrontal ensembles. This analysis suggested that the relative segment number and the direction were most strongly affecting the responses of the neurons. This is consistent with the single cell analysis (Averbeck et al. 2003) that found direction and segment number were significant factors in 36% and 46% of the prefrontal cortex cells, respectively. However, the MDS analysis considers not only a cell's response to a given variable, but also the degree to which that variable affects the variance of the firing.

In summary, these data quantify the amount of information about shape segment transmitted by individual neurons and also neuronal ensembles in prefrontal cortex. The amount of information transmitted by neural ensembles was an increasing function of the number of cells they contained. However, because redundant information was transmitted by some neurons, the total information transmitted by the ensemble did not equal the sum of the information transmitted by its member neurons taken individually. Individual neurons conveyed information about a number of segments in the set the monkey produced. In this sense, they were 'broadly tuned' with respect to segment. Because of this, each neuron participated in the neural representation of several segments, and each segment was represented by a considerable number of prefrontal neurons. A consequence of this later fact may been that it was possible to decode the correct segment with reasonable success (correct in 60–80% of cases) on the basis of comparatively small neural ensembles containing fewer than 20 neurons randomly encountered in prefrontal cortex.

Acknowledgements We would like to thank Stefano Panzeri for providing the algorithm that computed the Bayesian form of the information correction term. This work was supported by United States Public Health Service grant NS17413, the United States Department of Veterans Affairs, and the American Legion Brain Sciences Chair.

References

Abramson N (1963) Information theory and coding. McGraw-Hill, New York

Amirikian B, Georgopoulos AP (2000) Directional tuning profiles of motor cortical cells. Neurosci Res 36:73–79

Atick JJ, Redlich AN (1990) Towards a theory of early visual processing. Neural Comput 2:308–320

Atick JJ, Li Z, Redlich AN (1992) Understanding retinal color coding from first principles. Neural Comput 4:559–572

Averbeck BB, Chafee MV, Crowe DA, Georgopoulos AP (2003) Neural activity in prefrontal cortex during copying geometrical shapes. I. Single cells encode metric, sequence and shape parameters. Exp Brain Res (in press)

Barlow HB (1972) Single units and sensation: a neuron doctrine for perceptual psychology? Perception 1:371–394

Benton AL (1968) Differential behavioral effects in frontal lobe disease. Neuropsychologia 6:53–60

Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, Oxford
Brunel N, Nadal JP (1998) Mutual information, Fisher information, and population coding. Neural Comput 10:1731–1757
Constantinidis C, Franowicz MN, Goldman-Rakic PS (2001) Coding specificity in cortical microcircuits: a multiple-electrode analysis of primate prefrontal cortex. J Neurosci 21:3646–3655
Duda RO, Hart PE, Stork DG (2001) Pattern classification. John Wiley, New York
Foldiak P, Young MP (1995) Sparse coding in the primate cortex. In: Arbib MA (ed) The handbook of brain theory and neural networks. MIT Press, Cambridge MA, pp 895–898
Gainotti G (1985) Constructional apraxia. In: Frederiks JAM (ed) Handbook of clinical neurology, vol 1. Elsevier, Amsterdam, pp 491–506
Gawne TJ, Richmond BJ (1993) How independent are the messages carried by adjacent inferior temporal cortical neurons? J Neurosci 13:2758–2771
Gawne TJ, Kjaer TW, Hertz JA, Richmond BJ (1996) Adjacent visual cortical complex cells share about 20% of their stimulus-related information. Cereb Cortex 6:482–489
Georgopoulos AP, Massey JT (1988) Cognitive spatial-motor processes. 2. Information transmitted by the direction of two-dimensional arm movements and by neuronal populations in primate motor cortex and area 5. Exp Brain Res 69:315–326
Gochin PM, Colombo M, Dorfman GA, Gerstein GL, Gross CG (1994) Neural ensemble coding in inferior temporal cortex. J Neurophys 71:2325–2337
Golomb D, Hertz J, Panzeri S, Treves A, Richmond B (1997) How well can we estimate the information carried in neuronal responses from limited samples? Neural Comput 9:649–665
Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the Macaque. J Neurophys 35:96–111
Hinton GE, McClelland JL, Rumehlart DE (1986) Distributed representations. In: Rumelhart DE, McClelland JL (eds) Parallel distributed processing: explorations in the microstructure of cognition. Vol 1. Foundations. MIT Press, Cambridge MA
Johnson RA, Wichern DW (1998) Applied multivariate statistical analysis. Prentice Hall, Saddle River NJ
Koski L, Iacoboni M, Mazziotta JC (2002) Deconstructing apraxia: understanding disorders of intentional movement after stroke. Curr Opin Neurol 15:71–77
Luria AR, Tsvetkova LS (1964) The programming of constructive ability in local brain injuries. Neuropsychologia 2:95–107
McAdams CJ, Maunsell JH (1999) Effects of attention on the reliability of individual neurons in monkey visual cortex. Neuron 23:765–773
McClurkin JW, Optican LM, Richmond BJ, Gawne TJ (1991) Concurrent processing and complexity of temporally encoded neuronal messages in visual perception. Science 253:675–677
Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63:81–97
O Scalaidhe SP, Wilson FA, Goldman-Rakic PS (1999) Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: evidence for intrinsic specialization of neuronal coding. Cereb Cortex 9:459–475
Optican LM, Richmond BJ (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. J Neurophysiol 57:162–178
Optican LM, Gawne TJ, Richmond BJ, Joseph PJ (1991) Unbiased measures of transmitted information and channel capacity from multivariate neuronal data. Biol Cybern 65:305–310
Panzeri S, Treves A (1996) Analytical estimates of limited sampling biases in different information measures. Network 7:87–107
Panzeri S, Schultz SR, Treves A, Rolls ET (1999) Correlations and the encoding of information in the nervous system. Proc R Soc Lond B Biol Sci 266:1001–1012
Pouget A, Deneve S, Ducom JC, Latham PE (1999) Narrow versus wide tuning curves: what's best for a population code? Neural Comput 11:85–90
Rao SG, Williams GV, Goldman-Rakic PS (1999) Isodirectional tuning of adjacent interneurons and pyramidal cells during working memory: evidence for microcolumnar organization in PFC. J Neurophysiol 81:1903–1916
Reinagel P, Reid RC (2000) Temporal coding of visual information in the thalamus. J Neurosci 20:5392–5400
Rolls ET, Tovee MJ (1995) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. J Neurophysiol 73:713–726
Rolls ET, Treves A, Tovee MJ (1997) The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. Exp Brain Res 114:149–162
Schoups A, Vogels R, Qian N, Orban G (2001) Practising orientation identification improves orientation coding in V1 neurons. Nature 412:549–553
Shannon CE (1948) A mathematical theory of communication. Bell System Technical Journal 27:379–423, 623–656
Thorpe S (1995) Localized versus distributed representations. In: Arbib MA (ed) The handbook of brain theory and neural networks. MIT Press, Cambridge, MA, pp 549–552
Treves A, Panzeri S (1995) The upward bias in measures of information derived from limited data samples. Neural Comput 7:399–407
Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287:1273–1276
Werner G, Mountcastle VB (1965) Neural activity in mechanoreceptive cutaneous afferents: stimulus-response relations, Weber functions, and information transmission. J. Neurophysiology 28:359–397
Williams GV, Goldman-Rakic PS (1995) Modulation of memory fields by dopamine D_1 receptors in prefrontal cortex. Nature 376:572–575
Williams GV, Rao SG, Goldman-Rakic PS (2002) The physiological role of 5-HT_{2A} receptors in working memory. J Neurosci 22:2843–2854
Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. Nature 370:140–143