

# Goal-Related Activity in V4 during Free Viewing Visual Search: Evidence for a Ventral Stream Visual Saliency Map

James A. Mazer\* and Jack L. Gallant

Department of Psychology  
University of California  
Berkeley, California 94720

## Summary

Natural exploration of complex visual scenes depends on saccadic eye movements toward important locations. Saccade targeting is thought to be mediated by a retinotopic map that represents the locations of salient features. In this report, we demonstrate that extrastriate ventral area V4 contains a retinotopic saliency map that guides exploratory eye movements during a naturalistic free viewing visual search task. In more than half of recorded cells, visually driven activity is enhanced prior to saccades that move the fovea toward the location previously occupied by a neuron's spatial receptive field. This correlation suggests that bottom-up processing in V4 influences the oculomotor planning process. Half of the neurons also exhibit top-down modulation of visual responses that depends on search target identity but not visual stimulation. Convergence of bottom-up and top-down processing streams in area V4 results in an adaptive, dynamic map of saliency that guides oculomotor planning during natural vision.

## Introduction

During visual exploration of complex natural scenes, both humans and nonhuman primates make highly stereotyped eye movements. These movements consist of periods of stable fixation lasting 200 ms or longer, interspersed with rapid eye movements (saccades) that shift gaze from one point to another in the scene (Yarbus, 1967). During natural vision, saccades bring salient visual features onto the fovea, where they can be processed with maximal spatial resolution (Motter and Belky, 1998). Natural scenes often contain many salient features, but the fovea can be directed toward only one at a time. Efficient allocation of limited foveal resources requires careful selection of foveation targets during natural vision. Many factors influence saccade targeting: behavioral goals, motivational state, and both the local and global spatial properties of the visual scene. This report explores the relationship between visual activity in an extrastriate form processing area, V4, and oculomotor planning during free viewing visual search.

The influence of spatial scene properties on saccade targeting (Reinagel and Zador, 1999; Yarbus, 1967) suggests that visual form processing should play a key role in guiding eye movements during natural vision. In 1985, Koch and Ullman (Koch and Ullman, 1985; also see Itti and Koch, 2000, and Niebur and Koch, 1996) proposed an influential model describing how salient features could be identified in natural scenes. The model post-

ulates a hierarchical arrangement of retinotopically organized filters that represent different spatial scales, orientations, and positions, analogous to the known properties of neurons found in early and intermediate visual areas. Nonlinear winner-take-all (WTA) interactions between filters produce a retinotopic saliency map. Oculomotor plans based on such a saliency map can be used to facilitate efficient use of limited foveal resources. If a saliency map is indeed used to guide eye movements, then the model suggests that visual selectivity and form processing should influence the oculomotor planning process.

From a theoretical perspective, a neural instantiation of the saliency map model should satisfy several constraints. The map must be retinotopically organized, it must receive input from visual areas capable of representing fine spatial details, and it must project to the oculomotor system. Single-neuron recording studies have demonstrated that the superior colliculus (SC), the frontal eye fields (FEF), and lateral intraparietal cortex (LIP) satisfy some, but not all of, these requirements. These areas are all active prior to saccades directed toward salient or behaviorally relevant visual features, and each connects, directly or indirectly, to the oculomotor system (Bichot et al., 2001a; Constantinidis and Steinmetz, 2001; McPeck and Keller, 2002). However, neurons in these areas are generally insensitive to the structural properties of visual stimuli such as orientation, spatial frequency, and color (Bichot et al., 1996). For this reason, it is unlikely that these areas alone can support the analysis of fine spatial detail required to facilitate saccade targeting during natural vision. In this report, we seek to identify the source of visual input to the saliency system.

One possibility is that saliency is an emergent property arising from a network of interconnected brain regions, each of which mediates a different aspect of saliency computation. Some areas, like the FEF, might be responsible for generating eye movements (Bruce and Goldberg, 1985). Parietal areas, like LIP, might mediate multimodal saliency extraction (Cohen and Andersen, 2000). Ventral areas, like V4, might be critically involved in computing saliency based on the spatiotemporal properties of the scene. This alternative framework for saliency computation represents a modification of the Koch and Ullman model because early spatial filtering stages are anatomically separated from the WTA stage.

Extrastriate area V4 is an important intermediate stage of visual form processing; virtually all visual signals in the ventral pathway pass through V4 on the way to inferotemporal (IT) areas (Ungerleider and Mishkin, 1982). In contrast to FEF and parietal areas previously implicated in saliency computation (Bichot et al., 2001b; Constantinidis and Steinmetz, 2001), V4 neurons are highly selective for complex visual attributes like shape and color (Gallant et al., 1993; Gattass et al., 1988; Pasupathy and Connor, 1999; Schein and Desimone, 1990). Although V4 has no direct anatomical connection to the oculomotor system, Fischer and Boch (1981a, 1981b) described presaccadic enhancement of activity in V4 neurons when stimuli in the receptive field (RF) were

\*Correspondence: mazer@socrates.berkeley.edu

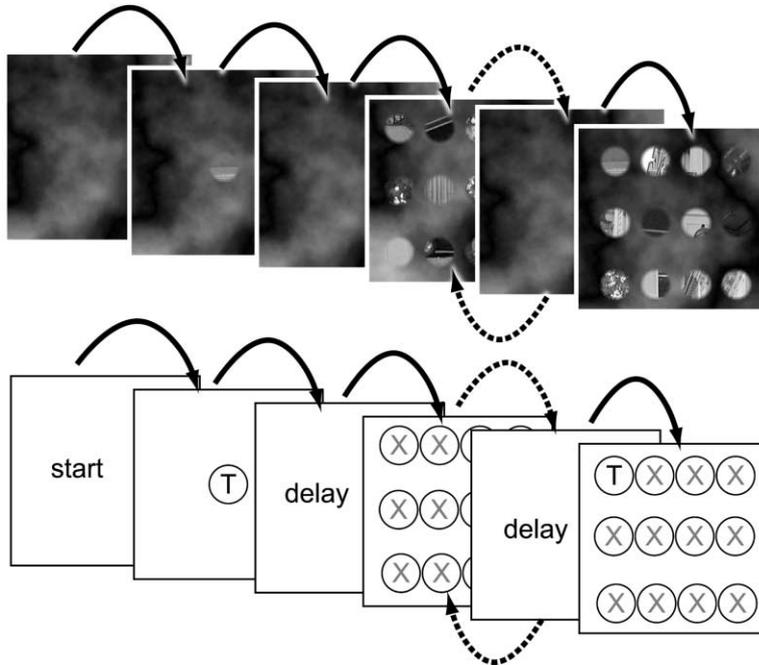


Figure 1. Free Viewing Visual Search Task

Top panel shows typical stimuli used during free viewing visual search task (FVVS); bottom panel shows schematic trial structure. The textured noise pattern cued onset of a search trial. Animals responded to the start cue by grabbing a touch bar. The search target (T) appeared at the center of the screen and could be inspected for 3–5 s. The search target was extinguished for a delay period (2–4 s), and an array of possible matches was presented (2–5 s). If any array contained the search target, the bar had to be released within 500 ms of array offset. If only distracters (X) were present in the array, no bar release was permitted. The delay-array cycle could repeat up to seven times on each trial. During each experiment, the search target and distracters were circular image patches of the same size selected randomly from a single photograph.

also targets for upcoming saccades. More recently, Tolias and colleagues (Tolias et al., 2001) used probe stimuli to map spatial RFs immediately before saccades toward highly salient targets. They described small changes in RF size and location immediately before saccade execution. The observation that activity in V4 is correlated with eye movements raises the possibility that V4 could be one source of visually selective input to the salience network.

In this report, we describe experiments designed to test the hypothesis that area V4 provides visual input to a salience network that could guide eye movements during natural vision. Two macaques were trained to perform a free viewing visual search (FVVS) task that requires both accurate oculomotor planning and fine visual discrimination (see Figure 1 and Experimental Procedures). This task was specifically designed to dissociate the spatial properties of visual stimuli from their perceptual salience. We expected that if V4 is involved in salience computation, then during FVVS, eye movements should be preferentially directed toward retinotopic locations with elevated V4 activity. If this were the case, then the visual responses of single V4 neurons should be correlated with the direction of subsequent eye movements during FVVS.

While the original salience map model was entirely stimulus driven, there is substantial evidence that extraretinal factors can influence salience computations. Previous studies have shown that attention can dramatically affect visual behavior (Treisman and Gelade, 1980). Neurophysiological studies in V4 have revealed two classes of attentional modulation. First, attention to particular spatial locations (spatial or focal attention) can affect both the spatial sensitivity profiles (Connor et al., 1997; Moran and Desimone, 1985) and tuning properties (McAdams and Maunsell, 1999) of V4 neurons. Second, attention to particular stimulus features or attributes (feature attention) can suppress or facilitate responses to stimuli falling in the RF (Haenny et al., 1987; Motter, 1994a).

Top-down modulation of visual inputs to the salience network might facilitate visual search. Modulation of specific classes of visual input based on the spatial properties of the search target could serve to make the entire network more sensitive to particular features and thereby increase the likelihood of target detection. Therefore, we also sought evidence of dynamic changes in V4 responses that were correlated with search target identity and uncorrelated with either visual stimulation or eye movements.

## Results

We recorded from 104 well-isolated V4 neurons in two adult male macaques while they performed the FVVS task (see Figures 1 and 2). Performance was typically 80%–95% correct on the search task. In this report we discuss only the data from trials in which the search target was correctly detected. To determine the relationship between the neuronal firing rate during each fixation and the direction of the subsequent saccade, we first analyzed the recorded eye position signal to identify fixations during FVVS (see Experimental Procedures). Spikes associated with each fixation were extracted and aligned to fixation onset. These data were used to compile fixation-aligned response rasters. Rasters were sorted by the direction of the subsequent saccade and binned ( $15^\circ \times 15^\circ \times 25$  ms bins) to form histograms. Figure 3B shows the activity of a single V4 neuron, plotted as a two-dimensional function of subsequent saccade direction. This neuron shows an enhanced visual response during fixations immediately prior to saccades directed toward the RF and a relative reduction in firing prior to saccades directed away from the RF. The time course of enhancement is correlated with the time course of the visual response itself; maximal enhancement occurs at the peak latency of the neuron.

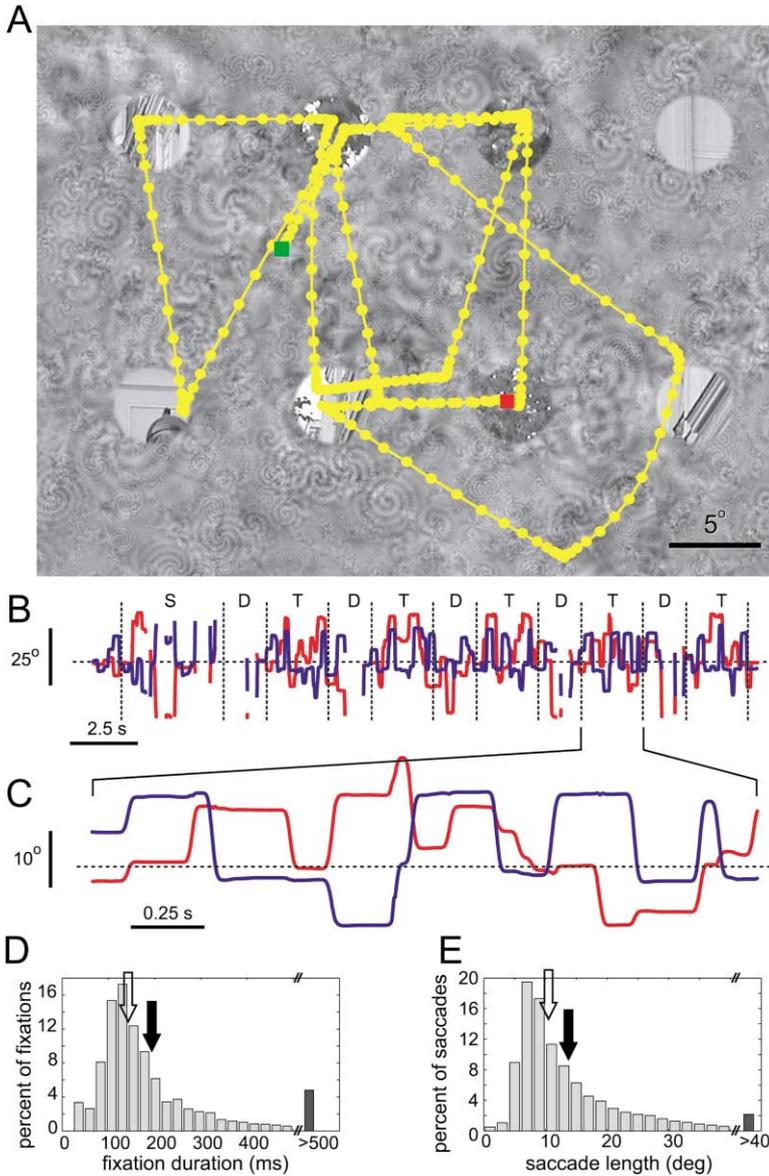


Figure 2. Typical Eye Movements during FVVS

(A) One stimulus frame from a visual search trial is shown, overlaid with the animal's voluntary eye movements (yellow trace). The background pattern is a textured noise stimulus with  $1/f^2$  power spectrum. A  $4 \times 2$  array of image patches is blended into the background pattern. The green and red squares indicate the beginning and end of the eye trace, respectively.

(B) Eye position for the entire trial (~25 s), plotted as a function of time. Red and blue curves indicate horizontal and vertical eye position, respectively. Vertical dashed lines demarcate the different phases of the trial (S, sample presentation; D, delay period; and T, presentation of the test array). Gaps in the eye traces are caused by blinks and short periods of time when the monkey looked outside the calibration range (away from the CRT). Note that gaps occur predominantly during the delay period and at the end of the sample presentation period. Dashed horizontal line indicates  $0^\circ$  (dead ahead).

(C) The indicated portion of (B) is shown in an expanded time scale. The time period exactly corresponds to the frame shown in (A).

(D and E) The distribution of fixation durations and saccade lengths across the entire data set (82 neurons; 224,529 fixations and saccades). Solid and open arrows indicate mean and median values, respectively. Average fixation duration was  $195 \pm 168$  ms (mean  $\pm$  SD; median, 144 ms). Average saccade length was  $13.7^\circ \pm 11.9^\circ$  (median  $13.7^\circ$ ).

### Saccades Are Directed toward Retinotopic Locations of High V4 Activity

To identify all neurons in which visual response magnitude predicts the direction of subsequent saccades, we first collapsed each two-dimensional activity map (e.g., Figure 3B) along the radial direction, preserving the angular component ( $30^\circ$  bins). The resulting angular activity function represents the average visual response magnitude as a function of saccade angle, independent of saccade length. A permuted Rayleigh test was used to identify neurons with significantly nonuniform angular activity functions ( $p < 0.05$ , corrected for multiple comparisons across time bins). 58% (42/73) of the V4 neurons in our sample with sufficient data for this test exhibit statistically significant, nonuniform angular activity functions. During FVVS, the activity of these cells predicts the direction of the subsequent saccade. Note that the Rayleigh test is statistically conservative; it is most sensitive to distributions with a clear, unimodal peak and is relatively insensitive to multi-peaked distributions. For

this reason, this analysis probably underestimates the true proportion of V4 neurons whose activity predicts the direction of the subsequent saccade.

For fixations falling along the lower left boundary of each search array, subsequent saccades were almost always directed away from the RF (all neurons studied had RFs located in the lower left quadrant). On these fixations, only the background pattern could fall in the RF. This could lead to a biased assessment of saccade-related activity. For instance, if image patches consistently drove a neuron more strongly than the background pattern, regardless of salience, the resulting correlation between low firing rates and saccades away from the RF (biased by lower-left fixations) could spuriously appear to be salience-related activity. To address this, we repeated the Rayleigh test after excluding all fixations in which the RF fell outside the bounds of the search array. After exclusion, we found 59% (43/73) of the neurons tested have nonuniform angular activity functions, consistent with visual responses predictive

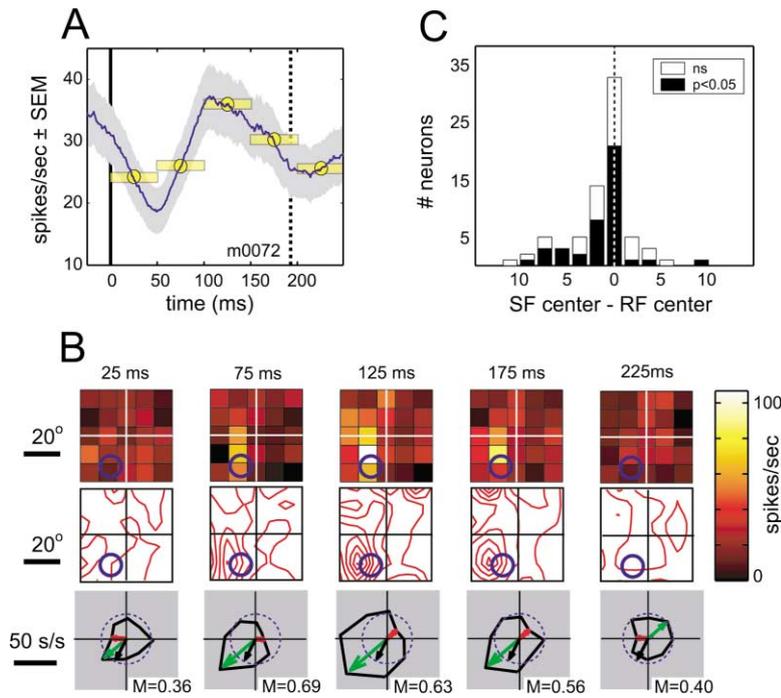


Figure 3. Visual Responses in V4 Predict the Direction of the Subsequent Saccade

(A) The fixation-aligned response histogram (1 ms bins) for a single neuron indicates the average response of the neuron during fixation, independent of the direction of the previous and subsequent saccades. The solid vertical line (0 ms) indicates fixation onset; the dotted vertical line represents the mean latency of the subsequent saccade. (B) To visualize the relationship between neural activity and the direction of the subsequent saccade, responses are sorted by direction of the subsequent saccade and plotted on two-dimensional maps reflecting activity at Hme points indicated by yellow bars in (A). Maps are smoothed for illustration ( $\sigma = 1$  bin). Blue circles indicate the spatial RF. In this neuron, visual responses are enhanced just before saccades that drive the fovea toward the RF (87 and 137 ms). The region of increased visual response is the salience field (SF). Bottom panel shows one-dimensional polar plots of activity as a function of subsequent saccade direction, independent of saccade magnitude (see Experimental Procedures). Black arrows indicate direction toward RF center; green and yellow arrows indicate directions preceded by maximum and minimum activity levels, respectively.

(C) Summary histogram of normalized differences between the locations of SFs and RFs across the sample of V4 neurons ( $n = 59$ ; see text for details). The nonuniform shape of the histogram indicates significant correlation between SF and RF.

of subsequent saccade direction. These data indicate that correlations between visual responses and saccade direction in V4 are not attributable to the specific sequence of eye movements the monkey uses to perform the FVVS task.

We defined the salience field (SF) of each neuron as the spatial region of enhanced presaccadic activity (i.e., the bright yellow regions in Figure 3B). We quantified the correspondence between SF and RF in each neuron by calculating the angular difference between SF and RF centers (mean of the best fit Gaussians). To account for SF and RF size differences, the angular difference was normalized by the average of the SF and RF sizes (see Experimental Procedures for complete details). Figure 3C shows the distribution of normalized angular differences across the sample. The correspondence between the spatial locations of the SF and RF is highly significant ( $\chi^2 = 36.1$ ,  $df = 9$ ,  $n = 59$ ,  $p < 0.0001$ ), indicating that V4 visual responses are typically enhanced prior to saccades that drive the fovea toward a neuron's RF.

#### V4 Predictive Activity Does Not Depend on Stimulus Luminance or Contrast

The image patches in this experiment can vary substantially in brightness and contrast, and it is theoretically possible that this variability could have influenced our results. Consider, for example, a situation in which high-energy (i.e., luminance or contrast) targets falling within the RF tend to attract saccades, while low-energy targets do not. If high-energy targets also tend to elicit stronger responses in V4 than do low-energy targets, then for those neurons whose responses predict the direction of subsequent saccades, the distribution of

correlations between target energy and saccade direction will be biased: stimuli that fall in the RF and are the target for the subsequent saccade will tend to have higher energy than RF stimuli that are not the target for a subsequent saccade.

We investigated this possibility by plotting stimulus energy (average luminance, RMS contrast, and spectral content of pixels falling in the RF) as a function of subsequent saccade direction. Figure 4A shows a polar plot of RF luminance as a function of subsequent saccade direction. It is clear in this case that there is no systematic bias in the relationship between luminance energy and saccade direction. Figure 4B summarizes this relationship for the 73 neurons with sufficient data to perform this analysis. Each point represents the difference between RF angle and the mean vector of the polar luminance-saccade plot (Figure 4A). If the predictive activity we report here reflects saccades directed toward high luminance stimuli falling within the RF, then all points should lie on the line of unit slope ( $y = x$ ). Instead, the points are completely scattered, demonstrating that there is no systematic relationship between luminance, saccade direction, and RF position. We carried out additional tests to determine whether stimulus RMS contrast (Figure 4C) or spectral energy (Figure 4D; see Experimental Procedures) could account for our results. These tests show that contrast and spectral energy are also unrelated to saccade direction and RF position.

As an additional control, for 14 neurons we normalized all the image patches so they all had the same RMS contrast and mean luminance as the background. In these experiments it was impossible to perform the task using simple contrast or luminance cues. In 6/14 (43%) of these neurons, visual responses significantly pre-

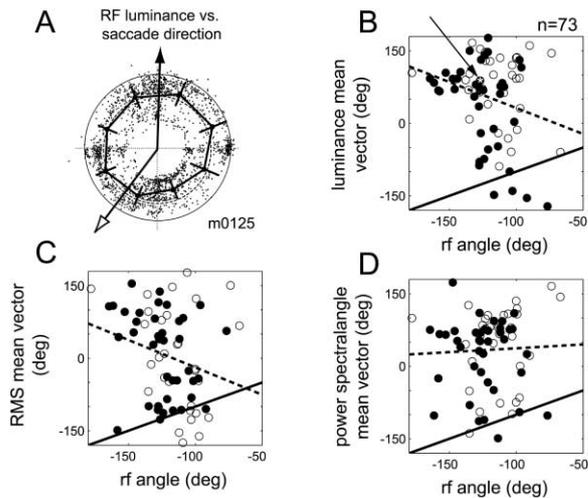


Figure 4. Simple Statistical Properties of Stimuli in the RF Do Not Predict Subsequent Saccade Direction

(A) Polar plot of the normalized luminance (0%–100%) of the pixels falling in the spatial RF as a function of subsequent saccade direction, for one neuron. Each point represents a single fixation. The solid black line (with error bars) indicates average RF luminance ( $\pm$ SD, 45° bins); the dashed circle indicates the mean luminance across all fixations; and the solid outer circle indicates 40% luminance. Open arrow and filled arrow indicate the direction toward the spatial RF and the mean vector of the polar luminance distribution, respectively.

(B) The angular location of the spatial RF (open arrow in A) is plotted on the abscissa and the mean vector of the luminance-RF profile (filled arrow in A) on the ordinate. Each point represents a single neuron; filled points denote significantly modulated neurons. The arrow and star denote the cell shown in (A). The dashed line indicates the best linear fit to the complete data set ( $n = 73$ ); solid line indicates unity slope ( $y = x$ ). The substantial scatter in these points indicates that there is no relationship between luminance of the stimulus within the receptive field and the direction of the subsequent saccade.

(C) Same analysis as in (B), but for rms contrast.

(D) Same analysis as in (C), but for spectral energy.

dicted the direction of the subsequent saccade. These data demonstrate that V4 neurons can still predict saccade direction even when luminance and contrast cues are not available.

#### V4 Activity Predicts Saccade Direction, not Gaze Angle

Previous studies reported that the activity of a subpopulation of V4 neurons can be modulated by changes in gaze direction (Rosenbluth and Allman, 2002). However, the effects reported here are not related to gaze direction, but rather to saccade direction. Gaze direction describes the angular displacement of the fovea relative to a fixed reference position (typically straight ahead). In contrast, saccade angle refers to the change in gaze angle as the animal redirects its eyes from one location to another. In this study, all saccade vectors were measured relative to the position of the eye at the start of each saccade and are therefore independent of gaze angle. The relationship between gaze angle and saccade vectors is illustrated in Figure 5. The circular shape of the saccade vector distribution (Figure 5C) indicates that saccade directions were sampled uniformly in our

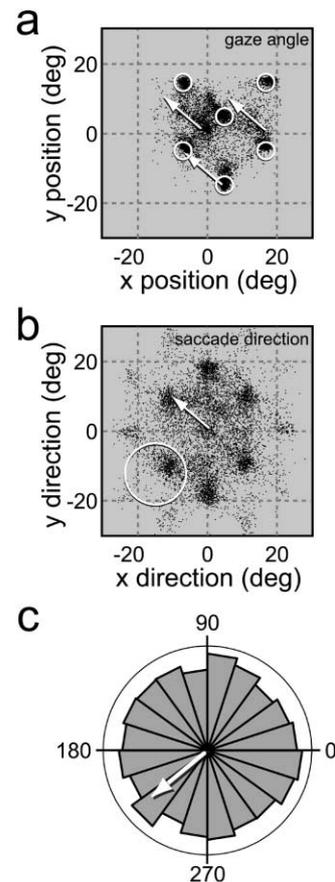


Figure 5. Relationship between Gaze Angle and Saccade Direction

(A) Summary of gaze positions in a typical FVVS data set (one neuron). Each dot indicates the gaze angle for a single fixation recorded during 2384 s of search (8358 fixations). Gaze is directed toward positions occupied by the search targets (white circles). Arrows indicate three equivalent saccades initiated from three different gaze positions (see text for details).

(B) Saccade vectors for the same data set. In our analyses, each saccade is a vector originating at the origin; for clarity only the vector endpoints are plotted. Circle indicates RF location. The single arrow represents all three saccades plotted in (A).

(C) Frequency histogram for saccade directions (9°/bin; circle = 500 saccades/bin). Saccades in the direction of the arrow are saccades directed toward the RF. Note that this distribution is circular, indicating that all saccade directions were sampled densely.

experiment, even though the search targets were presented on a regular grid.

#### V4 Predictive Activity Is Visual, not Motor

The data presented thus far suggest that the responses of each V4 neuron reflect, in part, the salience of the features in the spatial RF. During FVVS, the spatial distribution of activity across the surface of V4 feeds forward to higher stages of processing, where it is used to guide oculomotor planning. However, previous studies have suggested that saccade-related activity in V4 could reflect oculomotor command signals, as opposed to visual inputs used to make oculomotor decisions (Fischer and Boch, 1981a, 1981b; Tolia et al., 2001). Therefore, we examined saccade-related activity more closely to determine whether correlations between V4 activity and subsequent saccade direction reflect an oculomotor

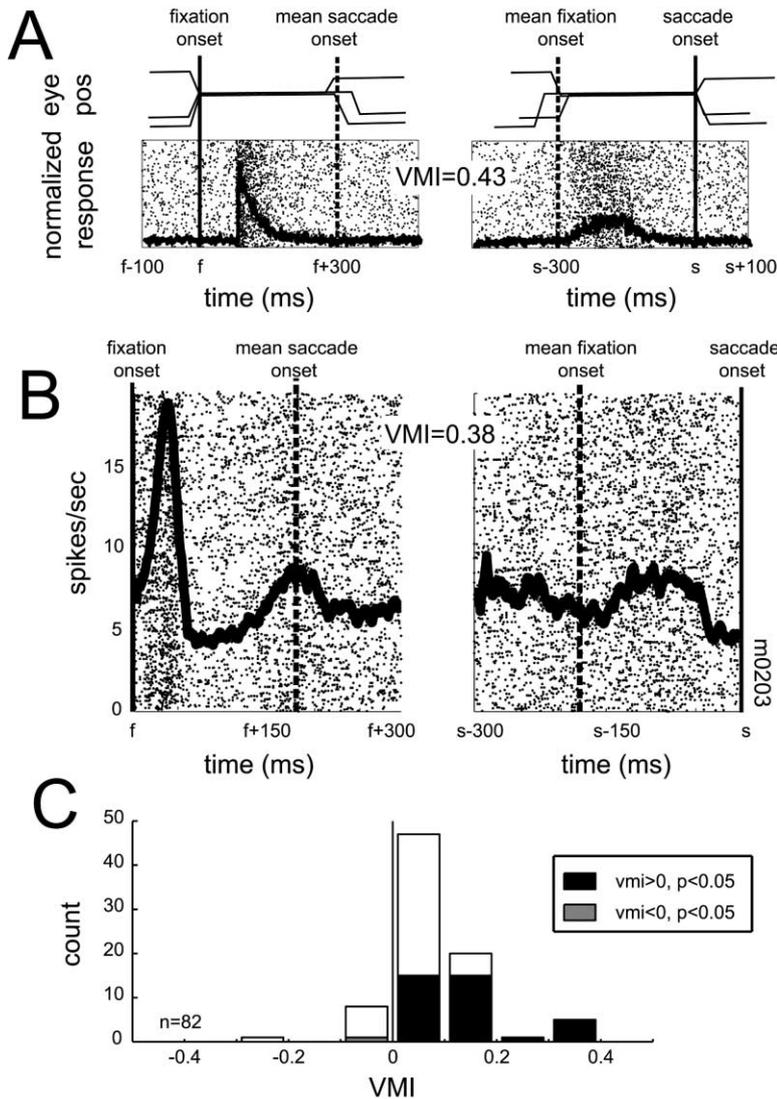


Figure 6. V4 Activity during FVVS Reflects Visual Responses rather than Premotor Activity

(A) Computer simulation illustrating how the visuomotor index (VMI, see text for details) can distinguish between visual and motor responses. Data were obtained from a model V4 neuron (latency 70 ms) responding to 1000 fixations. The left column shows simulated eye movements (top) and spike rasters (bottom) aligned to fixation onset. The right column shows the same data aligned to onset of subsequent saccade. *x* axes indicate time relative to fixation (*f*) or saccade (*s*) onset. On the left, the solid vertical line indicates fixation onset and the dashed line indicates the mean latency of the subsequent saccade; on the right, the dashed line indicates saccade onset and the solid line indicates the mean onset of the previous fixation. Solid traces are response histograms for the two alignment conditions, plotted on the same scale. The histogram peak is higher in the fixation-aligned condition (VMI > 0), confirming that the underlying simulated process is visual and not premotor.

(B) Fixation- and saccade-aligned responses from one V4 neuron, along with average response histograms. Like the model cell, this neuron has a strong visual response and a positive VMI.

(C) Summary histogram of VMI frequencies in the population of 82 neurons tested. The average VMI ( $0.08 \pm 0.01$ , mean  $\pm$  SEM) is significantly greater than zero (permuted *t* test,  $p < 0.0001$ ). Individual significance testing reveals that in 36/82 (44%) of the cells, the VMI is significantly greater than zero (permuted *t* test,  $p < 0.05$ , indicated by the black bars); one neuron (gray bar) has a significantly negative VMI.

plan, either premotor activity originating in V4 itself or a corollary discharge signal originating in oculomotor or parietal areas.

To distinguish between visual activity and oculomotor commands, we calculated two response histograms for each neuron, one aligned to fixation onset and one to saccade onset. The ratio of peak magnitudes from the fixation- and saccade-aligned response histograms indicates whether neuronal activity is influenced predominantly by visual or oculomotor factors (see Figure 6). For each neuron, we computed a visuomotor index,  $VMI = (f_{max} - s_{max}) / (f_{max} + s_{max})$ , where  $f_{max}$  and  $s_{max}$  are the maximum firing rates in the fixation- and saccade-aligned response histograms, respectively. Positive VMIs indicate that neuronal modulation is coupled to the visual response; negative VMIs suggest that activity reflects a premotor response. 89% (73/82) of V4 neurons in our sample have positive VMIs (see Figure 6C), and of these, 36 VMIs are significantly greater than zero ( $p < 0.05$ ). Only one cell has a significantly negative VMI. This analysis demonstrates that V4 activity during FVVS is predominantly visual: visual responses in V4 influence oculomotor planning, and not vice versa.

#### V4 Responses Are Modulated by Task Demands

We hypothesized that if top-down feature attention serves to facilitate visual search, then changes in visually driven activity in V4 should be correlated with search target identity. During FVVS, the cued sample at the start of each trial (see Figure 1) provides complete information about the spectral properties of the target, i.e., orientations, spatial frequencies, contrast, etc., but no information about the spatial location of the match. Feature attention could serve to modulate V4 selectivity based on the spectral properties of the sample. (Based on the uniform distribution of saccade directions shown in Figure 5C, we assumed that spatial attention, being tightly coupled to eye movement generation [Bisley and Goldberg, 2003], was uniformly distributed.)

In our experiments, we typically selected four different search targets (and 90–100 distracters) at the start of each run. For each neuron in our sample, we compiled fixation-aligned response histograms conditioned on the search target (see Figure 7). To assess the statistical significance of modulations caused by feature attention, each target-contingent, fixation-aligned response histogram was compared to a grand mean fixation-aligned

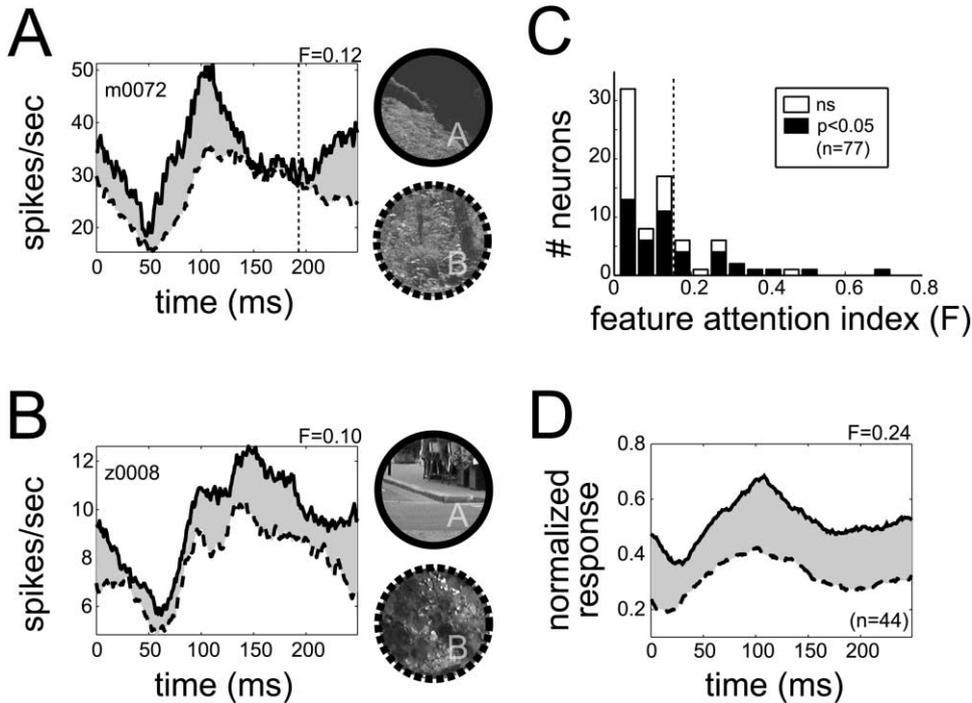


Figure 7. V4 Neurons Are Modulated by Feature Attention during FVS

Fixation-aligned response histograms (25 ms bins) conditioned by search target are shown for two different neurons (see text for details). Solid and dashed lines indicate mean firing rates; filled regions denote  $\pm$ SEM. Inset image patches are the actual search targets used in each condition. Vertical solid and dashed lines indicate fixation onset and mean saccade onset times, respectively.

(A) One V4 neuron whose activity increases during search for the upper patch, relative to search for the lower patch. Note that this change in activity is completely independent both of the specific visual stimuli falling in the receptive field and of subsequent saccade direction. Maximum modulation occurs near the peak of the visual response.

(B) A second V4 neuron that shows search target-dependent modulation. For this neuron, modulation occurs throughout the fixation response.

(C) Distribution of feature attention modulation indices. Filled black bars indicate neurons showing statistically significant attentional modulation ( $F > 0$ ;  $p < 0.05$ ). Dashed line is the average feature attention index across all significantly modulated neurons ( $0.16 \pm 0.14$ , mean  $\pm$  SD;  $n = 44$ ).

(D) Population effect of feature attention on the fixation-aligned response histogram. Solid and dashed lines show the population level fixation-aligned histograms obtained during search for the two different search targets, as described above for individual neurons. For each of the 44 significantly modulated neurons, fixation-aligned response histograms were normalized (0-1) and then averaged to compute the population effect.

histogram obtained by pooling across all target conditions. Each target-contingent histogram was compared to the grand mean histogram using a permuted t test. If any target-contingent histogram showed significant differences from the mean ( $p < 0.05$ , corrected for multiple comparisons), the neuron was considered to be modulated by feature attention. Over 25% (28/104) of V4 neurons show significant differences in fixation-aligned responses across search targets. We typically obtained several thousand fixations in each target condition. Each fixation brought a unique visual stimulus into the RF, and each was followed by a unique saccade vector. Sample-dependent response histogram modulations are therefore not attributable to variations in either the visual stimuli or the pattern of saccades generated by the animals in each experimental condition.

### Discussion

Our results reveal several novel findings about V4 activity during visual search. First, visually driven activity in single V4 neurons predicts the direction of subsequent saccades. More specifically, the correspondence between SF and RF indicates that enhanced visual re-

sponses are most likely to be followed by saccades that move the fovea toward the location currently occupied by an active neuron's spatial RF. Second, presaccadic enhancement in V4 reflects the salience of features in the RF and is not merely an oculomotor command signal originating elsewhere. The strong correlation between oculomotor behavior and visual activity in V4 suggests that V4 contributes to the oculomotor planning process; however, it does not appear to actually encode the motor plan itself. This is consistent with the critical role of area V4 in the visual form processing stream (Gallant et al., 2000). Third, salience signals in V4 are influenced by feature attention. Many previous studies have shown that V4 activity is influenced by both feature and spatial attention (McAdams and Maunsell, 1999; Moran and Desimone, 1985; Motter, 1994a, 1994b). Based on the results described here, we suggest that during natural vision, feature attention serves to modulate visual selectivity to facilitate target detection. During both free viewing visual search and visual exploration of natural environments, feature attention can act to facilitate accurate oculomotor planning.

The original salience map hypothesis (Koch and Ullman, 1985) postulated that target detection in complex

scenes could be supported by a single retinotopic, stimulus-driven salience map requiring no top-down modulation. We suggest that salience computation is actually implemented by a network of cortical and subcortical areas working in concert. Some areas, like V4, provide visual input to the salience network. Others, like the FEF, reflect the salience network's output and tie the network into the oculomotor system. Extraretinal modulation in V4 reflects the influence of behavioral goals and task demands. These signals may arise in IT and prefrontal cortex (Miller et al., 1993; Wilson et al., 1993). Although the computational principle of this network are similar in many respects to the original conception of the salience map (Koch and Ullman, 1985), in this formulation no single brain structure is solely responsible for salience computation.

It is interesting to compare the V4 activity observed here with an earlier study of IT activity during FVVS (Sheinberg and Logothetis, 2001). Our results suggest that the spatial distribution of activity in V4 encodes the retinotopic locations of salient features throughout the visual field. Our analysis excluded periods during which the search target was visible (see Experimental Procedures). Therefore, we can be sure that the salience representation in V4 depends on neither successful target detection nor the decision to execute a motor act in response to that information. In contrast, neurons in anterior IT cortex appear to encode target detection and not target salience. Sheinberg and Logothetis (2001) trained animals to perform a search for small objects embedded in natural scenes, using natural eye movements. Animals were required to execute a saccade to the detected target when the match was detected. Under these conditions, a subset of neurons in anterior IT is active only when the target appears in the RF and the subsequent saccade is directed toward the target. When the target falls in the RF but is not detected, these neurons are unresponsive. Taken together, our findings and those of Sheinberg and Logothetis suggest that the WTA stage of salience computation is implemented somewhere along the ventral stream between V4 and IT.

V4 does not make a direct projection to oculomotor areas but does receive feedback from the FEF (Stanton et al., 1993). Previous reports of presaccadic enhancement in V4 (Fischer and Boch, 1981a, 1981b; Tolias et al., 2001) suggested that enhancement was likely due to feedback from an oculomotor area, possibly the FEF. Under free viewing conditions, we find that the time course of subsequent saccade prediction closely follows the time course of the normal V4 visual response (see Figure 3). This indicates that presaccadic enhancement in V4 reflects an enhanced visual response due to behavioral context or salience. Although this finding is consistent with previously reported data, it suggests an alternative interpretation of those results. In the previous studies, animals were trained to perform a traditional saccade-to-target task in which they fixated for 300–1000 ms while a saccade target was presented at a peripheral location. At the end of the fixation period, saccades were executed toward the location cued by the saccade target (Fischer and Boch, 1981a, 1981b). The use of extended fixations facilitated dissociation of visual and motor responses by temporally separating the visual response from the motor act. Natural fixations

observed during FVVS are considerably shorter than those required in the earlier studies (mean 195 ms compared to 300–3000 ms; see Figure 2D), and we observe significant parasaccadic suppression in virtually all V4 neurons (Gallant et al., 1998). Our results suggest that previously reported presaccadic enhancement effects in V4 are either suppressed or obscured during FVVS. FVVS closely mimics natural visual exploration, both in the spectral properties of the stimuli and the pattern of eye movements. We must therefore consider the possibility that the presaccadic enhancement described by Fischer and Boch is not a significant factor in V4 during natural vision.

Taken together, our results demonstrate that area V4 is a key stage of the salience network during free viewing search. The spatial distribution of activity within V4 encodes information that downstream areas could use to guide subsequent saccadic eye movements toward interesting or behaviorally relevant points in the visual scene. Natural visual search requires coordination between salience computation, target selection, and oculomotor planning. Salience computation under natural viewing conditions appears to involve several areas, both visual and oculomotor, acting in concert. Our results suggest that V4 may be a key component of this network. FEF and LIP are clearly involved in target selection and oculomotor planning (Bichot and Schall, 1999; Biseley and Goldberg, 2003; Gottlieb et al., 1998), but they do contain the visually selective neurons required to support target selection during natural vision. V4 neurons provide visual input to the salience network that can be used to facilitate identification and localization of behaviorally relevant scene features. This information can be used to efficiently place potential matches at the fovea, where they can be scrutinized with high spatial resolution.

#### Experimental Procedures

Data were collected from two adult male monkeys (*Macaca mulatta*), 8 and 10 kg. All procedures were in accordance with the NIH Guide for the Care and Use of Laboratory Animals and approved by the Animal Care and Use Committee at the University of California, Berkeley. A metal head post, scleral search coil (in one animal), and an acrylic recording platform overlying area V4 were implanted using sterile surgical techniques under isoflurane anesthesia. During recording, V4 neurons were identified on the basis of both stereotaxic coordinates and physiological properties.

#### Visual Search Task

Animals were trained to perform the FVVS task (see Figure 1) prior to the onset of recording. Search trials were cued by onset of a textured noise pattern. Animals initiated each trial by grasping a touch bar following cue presentation. A search target was presented at the center of a 21 inch CRT (Viewsonic PS790; 37 or 45 cm viewing distance) blended smoothly into the background pattern (filtered white noise and textured patterns with  $1/f^2$  power spectra). Animals were given 2–4 s to inspect the search target using voluntary eye movements. After a 2–4 s delay (during which only the textured background pattern was visible) an array of 4–25 potential match stimuli was presented (also blended into the background pattern). Each array remained on the monitor for 2–5 s. If the search target appeared anywhere in the array, the touch bar had to be released no later than 500 ms after array offset. Following nonmatch arrays, there was another 2–3 s delay period. The delay-array sequence repeated 1–7 times (uniform probability distribution); the final array on each trial contained the search target. If the target was correctly

detected, as indicated by the bar release, monkeys received a liquid reward. Failures to detect the target were indicated by an error cue followed by a brief timeout period. The monkeys were not required to indicate the position of the match. Both the spatial location of the target and the time at which it appeared in the trial (frame number) were selected randomly at the beginning of each trial. All frame durations and trial lengths were fully randomized to eliminate anticipatory effects.

Both search targets and distracters were circular image patches extracted from black and white photographs. For most neurons studied, patches were cropped using a circular mask the size of the spatial RF. In a few eccentric neurons with large RFs, patches were limited to a maximum radius of 5°. The outer 10% (by radius) of each patch was  $\alpha$ -blended smoothly into the background pattern. During each recording session, image patches were selected randomly from a single high-quality digital black and white photograph. The grid spacing and geometry of each search array were individually adjusted for each neuron studied such that fixation of one patch usually placed a different patch at the center of each neuron's RF. For parafoveal neurons (<2° eccentricity), patch size was adjusted to encompass both the RF and fovea, and array spacing was adjusted to prevent any patch overlap. RFs ranged from 0.5° to 16° in radius (mean 4.6°).

#### Fixation Task

To facilitate RF mapping and eye tracker calibration, animals were also trained to acquire and maintain fixation on a small (2–3 arcmin) high-contrast target for 3–6 s. During fixation trials, eye movements (>0.5°) ended the trial. For RF mapping, small high-contrast probe stimuli (circular spots or oriented bars) were presented at random locations (5–10 Hz). Spot locations were determined by sampling the hand-mapped RF with a 10 × 10 (or denser) grid. Conventional reverse correlation methods were used to estimate RF location and size (Mazer et al., 2002). Quantitative estimates of RF size and location were obtained by fitting spot map data with circularly symmetric two-dimensional Gaussians; mean and standard deviation of the best fit Gaussian were taken as the RF center and radius, respectively. During eye tracker calibration, animals fixated on the same small targets placed at 20–40 different locations on the CRT in random order, one location per trial. Eye positions measured during FVVS were converted to CRT coordinates by interpolation based on the data obtained during tracker calibration runs. Interpolation was performed offline using a two-dimensional cubic spline function (MATLAB, MathWorks, Natick, MA). Eye tracker calibration was performed for every neuron studied, typically at the end of each FVVS run. This calibration procedure was sufficient to recover gaze direction during FVVS with a resolution of ~0.25° for both the scleral coil and the infrared tracking system. Blinks and periods when gaze was directed outside the calibration range (i.e., off-screen) were excluded from analysis.

#### Data Collection

Behavioral control, stimulus presentation, and data collection were performed on a Linux microcomputer using custom software. Eye movements were recorded either with a scleral search coil (1000 Hz) (Judge et al., 1980) or with an infrared eye tracker (120 Hz: RK-801, ISCAN, Burlington, MA; or 500 Hz: Eyelink II, SR Research, Toronto, Canada). Latencies associated with the video-based trackers were compensated during offline analysis (Gawne and Martin, 2000). Single neuron responses were recorded using high impedance (nominally 10–25 M $\Omega$ ) epoxy-coated tungsten microelectrodes (125  $\mu$ m diameter, 20°–25° taper; Frederick Haer Co., New Brunswick, ME). A microdrive system (MM-3BF, National Aperture, Nashua, NH) was used to advance electrodes through the intact dura perpendicular to the cortical surface. Signals were amplified (Model 1800, AM-Systems, Seattle, WA) and band pass filtered (0.1–10 kHz; custom filter), and spikes were isolated using a conventional window discriminator. In later experiments, neural signals were recorded with a dedicated multichannel recording system (amplification, filtering, and spike detection in a single unit; MAP, Plexon Inc, Dallas, TX). Spike times were recorded with 1 ms resolution by the same computer system controlling the behavioral task and recording eye movements.

#### Response Histograms

To construct response histograms, the continuous data record obtained during each FVVS trial was segmented into fixations. First, saccades were identified from the calibrated eye position signal by thresholding the eye velocity signal (>12°/s). Continuous periods of stable fixation (>25 ms) were isolated along with corresponding spike rasters (beginning 500 ms before fixation onset and ending at the onset of the following saccade). Rasters were aligned to either fixation onset or the start of the following saccade, depending upon the analysis. Rasters were collected into response histograms using either 25 ms or 1 ms bins, depending on the analysis. Histograms binned at 1 ms resolution were convolved with a Gaussian ( $\sigma = 3$  ms). Only data from correct trials were analyzed in this report (80%–95% of all trials). The first and last frames of each trial (when the search target was present on screen) were also excluded from analysis.

To distinguish between visual and motor activity, we calculated a visuomotor selectivity index (VMI) from the 1 ms resolution response histograms. This index,  $VMI = (f_{max} - s_{max}) / (f_{max} + s_{max})$ , where  $f_{max}$  and  $s_{max}$  are the maximum rates in the fixation- and saccade-aligned response histograms, respectively, varies between  $-1$ , for a pure motor response, and  $+1$ , for a pure visual response.

Fixation-aligned response histograms were used to estimate the ability of each neuron to predict subsequent saccade vectors. For each neuron, fixation-aligned rasters were sorted by direction of the subsequent saccade, and response histograms (25 ms bins) were computed for each saccade vector (15° × 15° bins). For each histogram time bin, response rate was plotted as a two-dimensional function of saccade vector (see Figure 3B). To identify neurons in which visual activity was correlated with the direction of subsequent saccade, angular response functions were computed by plotting firing rates as a function of saccade direction (independent of amplitude; 30°/bin) for each time bin and tested for nonuniform distributions using a permuted Rayleigh test ( $p < 0.05$ , corrected for multiple comparisons, i.e., the number of time bins analyzed). For each neuron with significantly correlated activity, the angular center and width of the salience field (SF) was taken as the mean and standard deviation of the best-fit one-dimensional Gaussian to the angular response function with a maximum response rate. Differences between RF and SF centers were quantified by computing the normalized angular difference between the two centers:  $d = (\mu_{SF} - \mu_{RF}) / [(\sigma_{SF} + \sigma_{RF})/2]$ , where  $\mu_{SF}$  and  $\mu_{RF}$  are the angular SF and RF positions and  $\sigma_{SF}$  and  $\sigma_{RF}$  are their widths. If SF and RF are uncorrelated, then the distribution of  $d$  should be uniform. The measured distribution was tested for uniformity using a standard  $\chi^2$  test ( $df = 9$ ,  $p < 0.001$ ).

Fixation-aligned response histograms were also used to identify neurons with significant search target-dependent activity. For each neuron, fixation-aligned rasters were sorted by search target and averaged to compute a response histogram (25 ms) for each search target tested (typically four). Histograms were compared to a grand mean response histogram (computed from all fixations, independent of search target) as follows. For each time bin, each conditional histogram was compared to the grand mean histogram using a permuted t test. Neurons with significant deviations from the grand mean histogram, in any condition ( $p < 0.05$ , corrected for multiple comparisons), were considered to be modulated by the search target.

#### Stimulus Statistics

Additional analyses were used to assess the influence of stimulus statistics on neuronal activity and saccade direction (see Figure 4). The pattern falling within the RF on each fixation was first extracted and several statistical indices were computed: average luminance (normalized from 0%–100%), RMS contrast, and spectral content. Spectral content was characterized in terms of the power spectrum, expressed as a 256 element vector. Spectral differences between different stimuli falling within the RF were quantified by computing the angular difference between each measured power spectrum vector and a standard reference vector (all ones). Next, each index was plotted as a function of subsequent saccade direction (see Figure 4A), and the direction of maximum correlation was expressed as mean vector angle (see Figures 4B–4D).

## Acknowledgments

We thank K. Gustavsen, B. Willmore, K. Bradley, and T. Annau for critical comments on the manuscript. This research was funded by a grant from the National Eye Institute.

Received: May 20, 2003

Revised: October 10, 2003

Accepted: October 25, 2003

Published: December 17, 2003

## References

- Bichot, N.P., and Schall, J.D. (1999). Effects of similarity and history on neural mechanisms of visual selection. *Nat. Neurosci.* **2**, 549–554.
- Bichot, N.P., Schall, J.D., and Thompson, K.G. (1996). Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature* **381**, 697–699.
- Bichot, N.P., Rao, S.C., and Schall, J.D. (2001a). Continuous processing in macaque frontal cortex during visual search. *Neuropsychologia* **39**, 972–982.
- Bichot, N.P., Thompson, K.G., Rao, S.C., and Schall, J.D. (2001b). Reliability of macaque frontal eye field neurons signalling saccade targets during visual search. *J. Neurosci.* **21**, 713–725.
- Bisley, J.W., and Goldberg, M.E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science* **299**, 81–86.
- Bruce, C.J., and Goldberg, M.E. (1985). Primate frontal eye fields. I. Single neurons discharging before saccades. *J. Neurophysiol.* **53**, 603–635.
- Cohen, Y.E., and Andersen, R.A. (2000). Reaches to sounds encoded in an eye-centered reference frame. *Neuron* **27**, 647–652.
- Connor, C.E., Preddie, D.C., Gallant, J.L., and Van Essen, D.C. (1997). Spatial attention effects in macaque area V4. *J. Neurosci.* **17**, 3201–3214.
- Constantinidis, C., and Steinmetz, M.A. (2001). Neuronal responses in area 7a to multiple stimulus displays. I. Neurons encode the location of the salient stimulus. *Cereb. Cortex* **11**, 581–591.
- Fischer, B., and Boch, R. (1981a). Enhanced activation of neurons in prelunate cortex before visually guided saccades of trained rhesus monkeys. *Exp. Brain Res.* **44**, 129–137.
- Fischer, B., and Boch, R. (1981b). Selection of visual targets activates prelunate cortical cells in trained rhesus monkey. *Exp. Brain Res.* **41**, 431–433.
- Gallant, J.L., Braun, J., and Van Essen, D.C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* **259**, 100–103.
- Gallant, J.L., Connor, C.E., and Van Essen, D.C. (1998). Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing. *Neuroreport* **9**, 2153–2158.
- Gallant, J.L., Shoup, R.E., and Mazer, J.A. (2000). A human extrastriate cortical area functionally homologous to macaque V4. *Neuron* **27**, 227–235.
- Gattass, R., Sousa, A.P.B., and Gross, C.G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *J. Neurosci.* **8**, 1831–1844.
- Gawne, T.J., and Martin, J.M. (2000). Activity of primate V1 cortical neurons during blinks. *J. Neurophysiol.* **84**, 2691–2694.
- Gottlieb, J.P., Kusunoki, M., and Goldberg, M.E. (1998). The representation of visual salience in monkey parietal cortex. *Nature* **391**, 481–484.
- Haenny, P.E., Maunsell, J.H.R., and Schiller, P.H. (1987). State-dependent activity in monkey visual cortex. II. Retinal and extraretinal factors in V4. *Exp. Brain Res.* **69**, 245–259.
- Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* **40**, 1489–1506.
- Judge, S.J., Wurtz, R.H., and Richmond, B.J. (1980). Vision during saccadic eye movements. I. Visual interactions in striate cortex. *J. Neurophysiol.* **43**, 1133–1155.
- Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* **4**, 219–227.
- Mazer, J.A., Vinje, W.E., McDermott, J., Schiller, P.H., and Gallant, J.L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proc. Natl. Acad. Sci. USA* **99**, 1645–1650.
- McAdams, C.J., and Maunsell, J.H.R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* **19**, 431–441.
- McPeck, R.M., and Keller, E.L. (2002). Saccade target selection in the superior colliculus during a visual search task. *J. Neurophysiol.* **88**, 2019–2034.
- Miller, E.K., Gochin, P.M., and Gross, C.G. (1993). Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Res.* **616**, 25–29.
- Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science* **229**, 782–784.
- Motter, B.C. (1994a). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J. Neurosci.* **14**, 2178–2189.
- Motter, B.C. (1994b). Neural correlates of feature selective memory and pop-out in extrastriate area V4. *J. Neurosci.* **14**, 2190–2199.
- Motter, B.C., and Belky, E.J. (1998). The guidance of eye movements during active visual search. *Vision Res.* **38**, 1805–1815.
- Niebur, E., and Koch, C. (1996). Control of selective visual attention: modeling the “where” pathway. In *Advances in Neural Information Processing Systems*, D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, eds. (Cambridge, MA: MIT Press), pp. 802–808.
- Pasupathy, A., and Connor, C.E. (1999). Responses to contour features in macaque area V4. *J. Neurophysiol.* **82**, 2490–2502.
- Reinagel, P., and Zador, A.M. (1999). Natural scene statistics at the center of gaze. *Network* **10**, 341–350.
- Rosenbluth, D., and Allman, J.M. (2002). The effect of gaze angle and fixation distance on the responses of neurons in V1, V2 and V4. *Neuron* **33**, 143–149.
- Schein, S.J., and Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. *J. Neurosci.* **10**, 3369–3389.
- Sheinberg, D.L., and Logothetis, N.K. (2001). Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J. Neurosci.* **21**, 1340–1350.
- Stanton, G.B., Bruce, C.J., and Goldberg, M.E. (1993). Topography of projections to the frontal lobe from the macaque frontal eye fields. *J. Comp. Neurol.* **330**, 286–301.
- Tolias, A.S., Moor, T., Smirnikas, S.M., Tehovnik, E.J., Siapas, A.G., and Schiller, P.H. (2001). Eye movements modulate visual receptive fields of V4 neurons. *Neuron* **29**, 757–767.
- Treisman, A.M., and Gelade, G. (1980). A feature-integration theory of attention. *Cognit. Psychol.* **12**, 97–136.
- Ungerleider, L.G., and Mishkin, M. (1982). Two cortical visual systems. In *Analysis of Visual Behavior*, D.G. Ingle, M.A. Goodale, and R.J.Q. Mansfield, eds. (Cambridge, MA: MIT Press), pp. 549–586.
- Wilson, F.A.W., O’Scalaidhe, S.P., and Goldman-Rakic, P.S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* **260**, 1955–1958.
- Yarbus, A.L. (1967). *Eye Movements and Vision* (New York: Plenum).