Title: Macaque V1 representations in natural and reduced visual contexts: Spatial and temporal properties and the influence of saccadic eye movements

Authors and affiliations:

Octavio Ruiz
Department of Neuroscience, Brown University
Providence, RI. 02912

Michael A. Paradiso
Department of Neuroscience, Brown University
Providence, RI. 02912

Running Head: V1 responses in natural and reduced contexts

Corresponding author:

Michael A. Paradiso
Department of Neuroscience, Brown University
Providence, RI. 02912
michael_paradiso@brown.edu
Ph. 401 863 1159
Fax: 401 863 1074
Abstract

Vision in natural situations is different from the paradigms generally used to study vision in the laboratory. In natural vision, stimuli usually appear in a receptive field as the result of saccadic eye movements rather than suddenly flashing into view. The stimuli themselves are rich with meaningful and recognizable objects rather than simple abstract patterns. In this study we examined the sensitivity of neurons in macaque area V1 to saccades and to complex background contexts. Using a variety of visual conditions we find that natural visual response patterns are unique. Compared to standard laboratory situations, in more natural vision, V1 responses have a longer latency, slower time course, delayed orientation selectivity, higher peak selectivity, and lower amplitude. Furthermore, the influences of saccades and background type (complex picture vs. uniform gray) interact nonlinearly to give a distinctive, and presumably more natural, response pattern. While in most of the experiments, natural images were used as background, we find that similar synthetic unnatural background stimuli produce nearly identical responses (i.e. complexity matters more than “naturalness”). These findings have important implications for our understanding of vision in more natural situations. They suggest that with the saccades used to explore complex images, visual context (“surround effects”) would have a far greater effect on perception than in standard experiments with stimuli flashed on a uniform background. Perceptual thresholds for contrast and orientation should also be significantly different in more natural situations.

Keywords: Primary visual cortex, Natural vision, Orientation selectivity, Temporal aspects of visual processing, Visual Context and Saccades.
Introduction

Our visual system faces the daunting challenge of recognizing objects quickly and precisely as we explore complex natural scenes. Yet most of what is known about the primate visual system is derived from experiments that greatly reduce the complexity of visual input and behavior (e.g. showing spots or bars on uniform backgrounds rather than complex scenes; enforced fixation rather than exploration by saccades). The differences between these experimental simplifications and the complexity of natural scenes and behavior raise the question: does “naturalness” matter?

A number of labs have addressed this issue by independently exploring two questions: Does a natural stimulus produce different responses than a similar unnatural stimulus? And does a stimulus swept into view by a saccade give a different response than a flashed stimulus? For example, Gawne and Martin (2002) and Wurtz (1969) reported that the responses of primary visual cortex (V1) neurons are the same when a stimulus is flashed on the receptive field (RF) and when the stimulus enters the RF via saccade. However, recordings made from freely viewing animals suggest that there is something different about V1 responses when eye movements are allowed (Gallant et al. 1998; Livingstone et al. 1996). Furthermore, experiments in the LGN (Reppas et al. 2002), and in extrastriate visual cortex and parietal cortex (Duhamel et al. 1992; Kusunoki et al. 2000; Nakamura and Colby 2002; Tolias et al. 2001) show that eye movements can alter basic receptive field properties. Concerning natural images, several studies suggest that the statistics of natural images influence cortical responses (David et al. 2004; Kayser et al. 2004; Kayser et al. 2003; Vinje and Gallant 2000), though characterization of what is “natural” is a complex issue.
The present study is motivated by previous research in our lab that suggests that scene complexity and eye movements affect V1 neuron responses, and that their effects interact nonlinearily (Huang and Paradiso 2005; MacEvoy et al. 2008). As a first approach to study the effects of saccades on complex backgrounds, Huang et al. (2005) introduced changes in background luminance or texture simultaneously with a bar flashed on the RF (meant to mimic in a fixation paradigm the changes in luminance and texture around a cell’s RF that occur with a saccade). This stimulation paradigm significantly delayed orientation selectivity and contrast sensitivity. We ask in the present study whether the timing of feature selectivity is also altered in more natural visual situations. MacEvoy et al. (2008) manipulated both stimulus-presentation mode and complexity and found an interaction between the two factors; this may account for why some studies find significant effects of saccades on cortical responses and other studies do not.

The approach we take here is to contrast the influence of stimulus complexity and saccades separately and together to improve our understanding of cortical coding in more natural visual situations. Motivated by sometimes conflicting previous work in the field, we set out to address a number of key questions: First, how do complex images and saccadic eye movements combine to sculpt V1 responses? Second, is feature selectivity delayed in natural vision as Huang et al. (2005) suggest? Third, can we reconcile our findings of different flash and saccade responses with previous reports of similar responses? Fourth, are the effects of saccades a result of image translation on the retina or of a corollary discharge signal associated with the eye movement? Finally, if saccades modulate V1 responses more on a natural background than on a uniform gray
background, is this response difference particular to the background being a natural scene, or do similar unnatural backgrounds produce the same modulation?

**Materials and Methods**

All procedures used in these experiments conformed to National Institutes of Health guidelines and were approved by the Brown University Institutional Animal Care and Use Committee. Further details on animal procedures are available in Huang and Paradiso (2005) and MacEvoy et al. (2008).

*Animals and recordings*

Three monkeys (8.5 – 11 kg) were used. In an aseptic procedure, each animal was implanted with a titanium headpost. Two animals were implanted with a 19 mm diameter cylinder over V1 for recording with conventional microelectrodes. We used epoxy-coated tungsten or glass-coated platinum/iridium electrodes (FHC, Bowdoinham, ME, USA), or glass-coated tungsten electrodes (Alpha Omega Engineering, Israel). The other animal received a “Utah” array consisting of a 10×10 grid of 1 mm long electrodes (Blackrock Microsystems, USA). Potentials from conventional electrodes were amplified and recorded using a Power1401 converter and Spike2 software (Cambridge Electronic Design, England). Electrode-array potentials were amplified and recorded using a Cerebus system (Blackrock Microsystems, USA).

Animals were accustomed to sit comfortably in a primate chair (Crist Instruments, Hagerstown, MD) with their heads fixed, and trained to saccade to and fixate a small red spot presented on a
computer display for juice reward. Eye position was monitored by a 1-kHz infrared pupil-
tracking system (EyeLink 1000; SR Research).

During an experiment, the animal sat with its head fixed facing a computer monitor in a diffusely
lit room. Extracellular recordings were made of single unit and multiunit activity from area V1 in
one hemisphere of each animal. Neural activity was collected and included in our data set if it
responded to a stimulus bar regardless of its preference for stimulus size, contrast, orientation, or
movement. We recorded from 120 V1 units or multiunit sites from the three monkeys.

Recordings were categorized as single-unit or multiunit based on (i) compactness and
separability of the spike shapes examined off-line in a PCA space, and (ii) compatibility or
incompatibility of the recording inter-spike-interval histograms (ISIH) with the refractory period
of cortical neurons. ISIHs of most of our recordings could be classified as (1) ISIHs that
increased monotonically or remained constant with decreasing ISI were considered to come from
multiunit recordings, and (2) histograms that peaked somewhere between 2 to ~15 ms, and had
sparsely occupied or empty bins for shorter intervals were categorized as single units. Not all the
neurons recorded were used in this study: Data Analysis, below, describes how experiments were
screened, off line, for recording stability and accuracy of the subjects’ behavior. Our study
includes results obtained from 74 units or multi-unit sites, recorded from 72 experiments. Of
those 74 recordings, we classified 35 as single units, 29 as multiunits, and 10 unclassified. As
differences between single and multiunit responses were not evident, we grouped them together
in the analysis. For brevity, we will use the terms “neuron” and “cell” to refer to single units and
multi-units. The units/multiunits included in our study come from our three monkeys as follows:
46 from monkey P (single electrodes), 23 from monkey S (electrode array), and 5 from monkey
Z (single electrodes). Their RF eccentricities ranged from 2.1 to 7.6 degrees of visual field, with mean ± std.dev. = 4.5 ± 1.2 degs.

Experimental procedures, conditions, and stimuli

Experiments started by hand-mapping the receptive field of an identified neuron. With the monkey fixating the center of the screen, we defined the RF of a neuron by the location, size, width, orientation, and color (black or white), of a bar that produced the maximal response. The main procedure (described below) was then run using a stimulus with those optimal parameters. Each experiment lasted 2 – 4 hours and collected data on 20 – 40 trials of each condition; different conditions and bar orientations were randomly interleaved. Visual stimuli were generated using the Psychophysics Toolbox (Brainard 1997; Pelli 1997) in Matlab (The Mathworks, USA). The visual display was an Iyama CRT (Master Pro514, model HM204DTA, Japan), running at 150 frames/s, with 1024 by 768 pixel resolution, positioned 64 cm from the eyes of the subject and subtending 34 by 25 degrees of visual angle.

A set of six stimulus conditions was used to assess the contribution of saccades and background type to the response (Fig. 1; the yellow circle represents the RF; it was not on the screen during the experiments). The Michelson contrast of the bar stimulus in the RF was set to 48% above or below the gray background, whichever gave the greater response. As an efficient way to estimate the orientation sensitivity of responses, recordings were made with the bar stimulus at the optimal orientation and at 60 deg counterclockwise from optimal. Orientation selectivity was
not based on a comparison of responses to optimal and orthogonal stimuli because the orthogonal bars often gave no response.

--- Fig. 1 near here ---

In the Flash-on-Gray condition (Fig. 1A), a trial began with a uniform gray visual display and a red fixation spot 5° to the left or right of center (not illustrated in Fig 1). If the monkey acquired this peripheral fixation spot and maintained fixation for 300 ms, the red spot was extinguished and a new fixation spot turned on at the center of the screen (left frame in Fig. 1A). If the monkey moved its gaze to the central fixation spot within 300 ms, the stimulus bar was flashed in the RF (right frame of Fig 1A) and remained on the screen for 0.3 s. If the monkey fixated the central fixation for a total of 0.5 s, a liquid reward was given.

The Flash-on-Pic condition (Fig. 1C) followed the same sequence of events as the Flash-on-Gray condition except that the background was a complex picture taken from van Hateren and van der Schaaf (1998). The mean luminance of the picture was matched to the luminance of the gray background in the Flash-on-Gray condition (16 cd/m²) as well as the mean luminance of the walls adjacent to the visual display in the testing room. To ensure that stimulation of the RF was identical in the Gray and the Picture backgrounds, the stimulus bar was shown on a circular gray patch that we call a buffer. The buffer had the same luminance as the picture and gray backgrounds and a radius 1.25 - 1.3 times the length of the bar stimulus (Fig. 1C. For clarity, the size of the bar and the buffer are exaggerated in the figure).

The flash conditions were complemented by two saccade conditions. The important difference was that in the saccade conditions the RF bar stimulus was present before the animal made a 5-
deg saccade to the center of the display. Thus the saccade brought the stimulus bar into the RF rather than a flash. In the Sac-on-Gray condition (Fig. 1B), the screen showed the homogeneous gray background, a peripheral fixation spot, and the stimulus bar (left frame of Fig. 1B). If the monkey acquired the peripheral fixation spot and maintained fixation for 300 ms, the red spot was extinguished and the central fixation spot was turned on (right frame in Fig. 1A). The bar remained present all this time. When the monkey moved its gaze to the central fixation spot (small red arrow), the RF moved onto the bar (large red arrow). If the animal fixated the central spot for 0.5 s, a liquid reward was given. The Sac-on-Pic condition (Fig. 1D) was identical to the Sac-on-Gray condition except that the background was a complex picture. As in the Flash-on-Pic condition, a buffer surrounded the bar stimulus and filled the RF.

In addition to the four primary conditions described above, several controls were run to address finer points. To determine the extent to which neural activity is influenced by visual stimulation during saccades, we included two no-bar control conditions; they were identical to the Sac-on-Gray and Sac-on-Pic conditions except that the RF bar was never shown on the screen. These conditions are referred to as “Sac-on-Gray-NoBar” (not illustrated), and “Sac-on-Pic-NoBar” (Fig. 1E).

To study the influence of visual input and internal signals during saccades, we included a “simulated saccade” condition (“Sim-Sac”, Fig. 1F). The goal was to duplicate, in a fixating animal, the pattern of image movement across the retina that resulted from an actual saccade. To do this, we sampled the eye position during saccades at intervals equivalent to the refresh rate of the video monitor (150-frames/s). The average samples across many saccades were used to create 5-frame movies that moved the scene with the time course of real saccades. In this
condition, a trial began with a fixation spot at the center of the screen and the picture, buffer, and bar shown displaced 5° to the left or to the right (left frame in Fig. 1F). If the fixation point was acquired and held for 300 ms, the display contents (except for the fixation point) were shifted leftward or rightward by the 5-frame movie (red arrows), ending with the same position used in Flash-on-Pic and Sac-on-Pic trials (right frame in Fig. 1F).

To test if natural images have a unique effect on V1 activity, we compared responses in all Picture conditions shown in Figure 1 when the background was either a natural or an “unnatural” scene. The unnatural background scenes were derived from van Hateren and van der Schaaf (1998) images using the Portilla and Simoncelli texture-synthesis algorithm (Portilla and Simoncelli 2000). Examples of unnatural images synthesized from the natural image in Fig. 1C are shown in Fig. 2A-C. The synthesized unnatural images had approximately the same contrast distribution, spatial-frequency power spectrum, and anisotropies as the original image.

--- Fig. 2 near here ---

Data analysis

Recordings were screened offline for spike stability and eye-fixation accuracy. Experiments were discarded if spike waveforms were not stable enough to obtain at least 20 trials of each stimulus condition. Individual trials were discarded if a corrective saccade was needed to foveate the post-saccade fixation point or if eye position was not stable within a 0.15 deg-radius window for all the conditions. The results reported here were obtained from 74 units or multi-unit sites, recorded from 72 experiments.
We calculated peri-stimulus spike histograms of the response (PSTH) using 5 ms width bins. To compute average PSTHs from all the cells in our sample, we first normalized the PSTHs from all stimulus conditions in an experiment to the response peak elicited by the optimal-orientation bar in the Flash-on-Gray condition. Orientation selectivity was calculated as

$$\text{Orient.Selectivity}(t_j) = \frac{V_0(t_j) - V_{60}(t_j)}{V_0(t_j) + V_{60}(t_j)}$$

Eq. (1)

where $t_j$ is the time, of the $j^{th}$ 5 ms bin, relative to stimulus onset (in the Flash conditions) or to the end of a saccade (in the Saccade conditions); $V_0(t_j)$ is the spike rate, in the $j^{th}$ bin, in response to an optimally-oriented bar, and $V_{60}(t_j)$ is the spike rate in response to a bar tilted 60° counterclockwise from the optimal orientation. We used a 60° instead of a 90° shift, because, in preliminary experiments, many V1 neurons did not respond at all to orthogonal orientations.

Quantifying response timing and amplitude

To quantify the response timecourse, we defined three epochs in each neuron’s response (see exemplary PSTH in Fig. 3): pre-peak (red bar), peak (green) and late phase (blue). Peak activity was measured counting the spikes in the peak epoch (green), early and/or pre-peak activity were measured in a 16-ms window preceding the Flash-on-Gray response peak time (red bar), and late activity was measured in the interval $t = 100$ to 200 ms (blue bar).

The peak time, width, and the latency of the response were determined with an automated procedure that fitted a function to the spikes occurring after a flash or saccade brought the bar stimulus into the RF. Various strategies have been used previously to estimate response latency but there is no accepted standard (Friedman and Priebe 1998). Because of our interest in
following the temporal response dynamics in detail, we sought a single approach to measure the latency and the peak time of the response’s initial transient. We did so by fitting a simple model to the initial response, and using the parameters of the fitted model as indicators of the latency $c$, and time to peak, $b$, of the initial-response transient (fitted curve in Fig. 3). The function used in the fit was the Rayleigh probability density function (Papoulis and Pillai 2002), in the following model:

$$f(t; c, b) = \begin{cases} p_0; & t \leq c \\ p_0 + R(t-c, b); & t > c \end{cases} \quad \text{Eq. (1)}$$

Where $p_0$ represents the basal spike rate; $R(t-c, b)$ is the Rayleigh probability density function; $b$ determines the shape of the Rayleigh function and specifies the difference between the beginning and the maximum of the function (in our case: the time to peak); $c$, represents the response latency, and $t$ is the time relative to the stimulus-onset time (in the Flash conditions) or to the end of a saccade (in the Saccade conditions). To avoid artifacts and losses of resolution introduced by the width and position of bins in a spike histogram, the model was fit directly to the spike times using the Maximum-Likelihood Estimator (MLE; Eliason (1993)). Only the initial peak of the response was fitted (in the Fig. 3 example, the domain was limited to $0 \leq t \leq 65$ ms). Initial solutions were set by visually inspecting the PSTHs. Although our method can theoretically find latency and peak response times with no time-binning, its resolution is limited by the sampling rate of the signal acquisition system (0.06 ms of CED/Spike2, and 0.03 ms of Cerebus).

Our approach can be considered an extension of Friedman and Priebe (1998), with the following differences: (i) our method estimates also the peak time of the initial transient, (ii) their method estimates the latency as the transition between two stationary stochastic processes (basal vs.
response), but our model assumes that only the basal activity is stationary and that the response is continuously changing; and (iii) their method takes PST histograms as inputs, which involve arbitrary binning, while ours operates on the raw spike times without binning.

Although we show PSTHs to illustrate our results, all quantitative analyses were done using the fitted-model numerical values of latency and amplitude measured from non-binned spike times. All the statistics were performed on these bin-free numerical values. Data analysis was performed using Spike2, Matlab, and Stata (StataCorp, USA).

--- Fig. 3 near here ---

Defining the stimulus onset time in saccade conditions

Strictly speaking it is impossible to equate stimulus onset time in flash and saccade conditions. In the flash condition the stimulus appears instantaneously within a static RF. In the saccade condition the stimulus enters the RF gradually before stopping. Furthermore, the movement of the stimulus on the retina is not only a function of the movement of the eye ball, but also of the lens within the eye, which lags and then overshoots the movement of the ocular globe (Deubel and Bridgeman 1995a; b). Therefore, following Deubel and Bridgeman (1995a) we defined the end of a saccade as the peak of the first overshoot in the eye-position trace provided by our optical eye tracker, and we corrected this time by the 3-ms delay resulting from the eye-tracker standard filter setting.
Results

Effects of visual context and saccades on response timing and amplitude

Across the four primary experimental conditions, V1 neurons show a diversity of latencies, magnitudes, and time courses (See examples in Fig. 4A and B). In most cells, responses were of progressively lower amplitude and longer latency when moving from the Flash-on-Gray (black), to the Flash-on-Pic (blue), Sac-on-Gray (magenta), and Sac-on-Pic (green) conditions. In some cells, these changes were quite pronounced (Fig 4B). Average responses across all neurons in our sample are shown in Fig 4C and mean values of different response parameters are given in Table 1. The thin traces in Fig. 4C show the responses to bar stimuli tilted 60 deg counterclockwise from optimal.

As shown in Fig. 4C (thick traces) and Table 1, response latencies to a stimulus flashed on a homogeneous gray (black trace) or a complex background (blue) are similar and significantly shorter than latencies in saccade conditions (magenta, green) (two-way ANOVA, $p < 0.00005$; $p($flash vs. saccade$) < 0.00005$). Saccades not only delay the beginning of the response to an object, but also reduce the speed of the response (rise time depends on the condition: two-way ANOVA, $p = 0.002$; $p($background$) = 0.01$). Overall, the response is most delayed and sluggish after a saccade on a complex background (green trace in Fig. 4C) and this translates into a
response peak that is delayed relative to all the other conditions. It is important to note that the
delay in the saccade condition responses is not simply the time it takes the bar stimulus to sweep
across the RF; from the time the bar first enters the RF to the time the eyes reach the final
fixation point there are only a few milliseconds, whereas the response latency and peak are
delayed 7 - 14 ms. Even more interesting, the response is much more delayed when the
background is complex rather than homogeneous, even though the eye dynamics are comparable.
The peak response amplitude is affected by both saccades ($p$(Flash vs. Saccade) = 0.001) and
background type ($p$(Gray vs. Picture) < 0.00005). As summarized in Table 1: compared to the
Flash-on-Gray, condition, the peak response is 92% in the Flash-on-Pic condition, 85% in the
Sac-on-Gray condition, and 59% in the Sac-on-Pic condition. A reduced response with a
complex background compared to a uniform background is expected from known modulatory
effects of the extra-classical receptive field (Albright and Stoner 2002; Allman et al. 1985). More
surprising is the reduced response that results when a saccade brings the stimulus into the RF
with either type of background. Overall, the most natural condition, Sac-on-Pic (green),
produced the smallest response (59% of the Flash-on-Gray condition). The reduced response in
the most natural condition is significantly smaller than a simple combination of the reductions
produced independently by the background and saccade factors ($0.92 \times 0.85 = 0.78 > 0.59$).
Evidently there is a nonlinear response depression from the combined factors of presentation and
background types in the Sac-on-Pic condition. Consistent with the free-viewing study by Gallant
et al. (1998), surround suppression is not sufficient to explain the reduced response we observed
with saccades on natural images.
Comparing the flash conditions with the saccade conditions makes an important point about the effectiveness of surround modulation in the two situations. In figure 4 the traces for the two flash conditions (black and blue) are much more similar than the traces for the two saccade conditions (magenta and green). When the stimulus is flashed in the receptive field, the peak response with a complex picture background (blue) is 8% less than with a gray background (black). On the other hand, when the stimulus comes into the receptive field with a saccade, the picture background response (green) is 26% lower than on a gray background (magenta). Thus, in the more natural conditions with saccades, the influence of visual context is much greater than in the flash conditions commonly used in experiments. The implication is that the many studies conducted with fixation paradigms may significantly underestimate the influence of visual context in natural vision.

Early activity in the saccade-on-picture condition

We found that a few cells in our sample (7 cells = 10% of the recorded units) showed an increase of activity in the Sac-on-Pic condition that preceded the responses to the other conditions (example in Fig. 5A; 7 cell average in Fig. 5B). This early activity is manifested as a small bump, around $t = 27$ ms, in the average Sac-on-Pic response (Fig. 4C, green trace), that disappears from the average if the 7 early-response units were removed (not shown). The origin of the Sac-on-Pic early response component could be (i) an internal, non visual, signal present in V1 during a saccade (i.e. corollary discharge), (ii) an anticipatory response to the stimulus that would enter the RF at the end of the saccade (i.e. RF remapping), or (iii) a response of the neuron to the background image sweeping across the RF during the saccade. To discriminate between these alternatives, we used Sac-on-Gray-NoBar and Sac-on-Pic-NoBar control conditions in
some experiments (Fig. 1). Responses in the Sac-on-Gray-NoBar condition (gray trace in fig. 5B) did not show any early activity, suggesting that the early response is not simply a manifestation of a corollary discharge. The green bold and thin traces in Fig. 5B show responses in the Sac-on-Pic condition to the optimal and 60 deg CCW bars in the RF, respectively. These two responses peak at 27 ms and appear identical until about 40 ms, where they diverge. This indicates that the early response is insensitive to the orientation of the bar that will enter the RF at the end of the saccade. This observation does not favor the hypothesis of an anticipatory response (remapping); neither does the fact that we never saw an early peak during saccades on the gray background (magenta).

The Sac-on-Pic-NoBar condition (cyan trace) produced little response at times when the presence of the bar at the end of a saccade would make the cells respond ($t > 40$ ms). However, the early component ($t \approx 27$ ms) was indistinguishable from the Sac-on-Pic condition (green). This suggests that the early component is a response to the relative movement of the background scene during saccades.

--- Fig. 5 near here ---

Effects of visual context and saccades on orientation selectivity

In addition to the effects on response amplitude and timing, saccades and background type also influence the latency, time course and magnitude of orientation selectivity (Fig. 4D). As expected from the delayed response, orientation selectivity is slowest to develop in the Sac-on-
Pic condition (green). More surprising is the greater peak selectivity in the most natural condition (0.47 compared with 0.35 in the Flash-on-Gray condition, at $t \approx 65$ ms). Closer inspection of Fig. 4C (thick green line for optimally-oriented stimulus, thin line for stimulus 60 deg from optimal) shows that the responses to the preferred and non-preferred orientation stimuli are not simply scaled versions of each other. Rather, the response to the non-preferred stimulus (thin green) rises more slowly and peaks later than the preferred-orientation response (thick green). By comparison, in the other three conditions the preferred- and non-preferred responses are more similar with a change in overall gain (thick and thin black, blue and magenta). At this point we don’t know why the timecourse of the Sac-on-Pic response changes so much with orientation. Perhaps the combined effects of saccades and background push the response below a linear operating region in the already weaker Sac-on-Pic response. For unknown reasons, of the four conditions tested, the Flash-on-Pic response showed the largest steady-state selectivity (Fig. 4D, blue trace, $t > 100$ ms; two-way ANOVA, $p = 0.004$).

Factors responsible for the unique response in the most natural condition

We considered three factors that might be involved in the altered responses in the most natural condition. One that we have largely ruled out is adaptation. In the Sac-on-Pic condition, the RF is stimulated on the penultimate fixation, something that does not occur in the flash conditions. If adaptation during one fixation reduces the response on the next, we would expect to see a negative correlation between activity on a fixation and activity on the previous fixation. However, instead of a negative correlation, we found a small positive correlation between the response on each fixation and the response on the previous-fixation ($R^2 < 0.11$, $p = 0.004$).
Furthermore, no correlation was found between the fixation response and activity just prior to the response (possibly resulting from visual stimulation during saccades). Therefore adaptation across saccades does not appear to be a major factor.

The neural signals responsible for reduced responses in more natural conditions might come from visual stimulation of the entire visual field during saccades or from a corollary discharge associated with the eye movement. We found that both these factors are important. To test the role played by global image motion during the saccade, we added simulated saccade controls in which natural images were moved across the retina by computer animation with the dynamics of actual saccades (condition described in Fig 1F, results in Fig. 6). The Sim-Sac responses (red lines in Fig. 6A) were, on average, similar to the Sac-on-Pic response (green lines). In both cases the responses are slower and their peak-amplitude lower than the other three conditions.

The movement of the background image during a saccade accounts for much, but not all, of the effects observed with saccades on a complex background. The response latency in simulated saccades falls short of the actual-saccade response \((p = 0.0001, \text{ paired t-test})\) and is similar to the Sac-on-Gray condition (magenta; see Table 1). In the Sim-Sac condition (red) the response is about 60\% of the Flash-on-Gray (black) response, but actual saccades reduce this amplitude a further 10 \% (Table 1). The early response (that we attributed to visual input during a saccade) is also present in the simulated saccade activity \((t \approx 20 \text{ ms and on})\), but this early activity is smaller
in the actual saccade condition ($p < 0.00005$, Wilcoxon paired test). Orientation selectivity develops in similar ways in actual and simulated saccades (Fig. 6B; green and red traces, respectively) and attains similar values (peak: $p = 0.99$; late phase: $p = 0.06$; Wilcoxon paired tests).

The suppression of early activity observed in the real-saccade condition is consistent with the idea of a corollary-discharge gating of input information to or in V1 during a saccade. However, unexpectedly, our data suggests that this corollary discharge, or some other signal related to actual saccades, may extend beyond the time-window of the saccadic visual input into the response itself, reducing the response.

"Naturalness" is not critical for the effects of a complex background

The discussion to this point has addressed the importance of saccadic eye movements and complex stimulus surrounds in natural visual responses. Not clear from these findings is whether the effects of the complex backgrounds were a consequence of them being natural images or instead simply images with complex arrangements of light and contours. To investigate this point we compared responses based on natural image surrounds with responses involving unnatural scenes synthesized using the Portilla and Simoncelli texture algorithm (see Methods). Fig. 7A shows responses in the Flash-on-Pic (blue), Sac-on-Pic (green) and Sim-Sac (red) conditions. There are two traces of each color that represent the responses when the background was either the natural scene (shown in Fig. 1C-F) or an unnatural version of that scene (Fig. 2A). For each condition, the responses superpose so well that is difficult to see any difference. The
orientation selectivity traces are noisier (Fig. 7B) but responses in each condition are similar regardless of the particular scene shown in the background. Similar results were observed when responses with three different natural scene backgrounds were compared (Fig. 1, Fig. 2B and C) (results not shown). Therefore, neither the “naturalness” of the scene, nor its particular elements, are critical for the contextual effects seen in V1 responses as long as the statistics of the scene remain globally similar.

--- Fig. 7 near here ---

Discussion

The data presented here show that in more natural situations V1 responses are different from responses in more reduced paradigms in the following ways: lower amplitude, longer latency, slower timecourse, delayed orientation selectivity, and higher peak orientation selectivity. In a previous study (MacEvoy et al. 2008) we examined the first of these properties (amplitude); more on that below. The aspects of natural vision that we have studied are the complexity of the background beyond the receptive field (complex picture vs. uniform gray) and saccades that bring stimuli into RFs rather than flashing the stimuli to a fixating animal. The results suggest that the many studies using flashed stimuli in reduced contexts overestimate the speed and especially the amplitude of V1 responses when extrapolating their measurements to natural vision. Response speed and amplitude are important because, in natural conditions, the typical
duration of fixations in humans is short (300 ms (Otero-Millan et al. 2008)), and because the
processing time of the primate visual system is very fast, on the order of 100 ms or less (Crouzet
et al. 2010; Fabre-Thorpe et al. 1998; Keysers et al. 2001; Stanford et al. 2010; Thorpe et al.
1996). This means that the actual spikes taken into account to perform a computation are few,
plausibly coming from the initial response in each fixation, and that they must be processed in a
brief time window. This makes relevant the delay, response reduction, and early orientation
selectivity differences found in the present study. Whether or not they are advantageous, natural-
vision response delays may represent the additional processing necessary to cope with the
enormous demands of analyzing images in non-constrained natural settings.

Although there are previous studies that explored the effects of saccades or background type,
most of the results presented here are novel. We confirmed the well-established point that
stimulus contrast outside the receptive field suppresses the response to an RF stimulus (Albright
and Stoner 2002; Allman et al. 1985). More unexpected however, is the large response reduction
that results from a saccade, even when the stimulus within and outside the RF is the same as in a
fixation/flash condition. This result is particularly surprising because several well-conducted
studies reported that V1 responses were unaffected by eye movements
(Gawne and Martin 2002; Wurtz 1969). Experimental details probably account for the
differences found in our study. For example (Wurtz 1969) recorded V1 responses when stimuli
moved rapidly across receptive fields due to either image motion or a saccade. Thus the results
apply to V1 responses to visual input during saccades rather than during fixations, as in our
study. Moreover, the Wurtz experiments were conducted with a blank background most similar
to our Flash-on-Gray and Sac-on-Gray conditions; they did not include a Sac-on-Pic condition,
the most natural condition in our study, and the one that produced the most unique visual responses. Gawne and Martin (2002) reported that most neurons gave similar responses to checkered patterns in flash and saccade conditions, but a minority showed amplitude or latency differences in their responses. Again, the background used in the study by Gawne and Martin was always gray; in this situation we find the saccade effects are smaller than with a complex background. Overall our findings are not at odds with previous studies that used saccades to bring stimuli into receptive fields; rather, they extend the studies to the more natural situation in which saccades are made across complex backgrounds.

In addition to saccades, the other aspect of natural vision we have examined is the “naturalness” of visual context – controlled bar stimuli were presented on photographs of outdoor scenes. Natural scenes have unique statistical properties and there is evidence that cortical processing may be adapted to them (Olshausen and Field 1996). Kayser et al. (2003) found that natural image responses in cat area 17 are qualitatively different than responses to gratings and to a number of other less natural stimuli. David et al. (2004) found that, in general, V1 responses are better predicted by responses to natural images than to more reduced stimuli. A critical distinction between our work and these earlier contributions is that we considered natural images only as background stimuli, not as the stimulus driving the receptive field. In our case, we wanted to assure that across conditions the RF was stimulated in as identical a manner as possible. Thus it was impractical to have the complex features of natural images entering the RFs. We find that responses to bar stimuli in the receptive field are strongly modulated by the surround but that modulation by natural image backgrounds is indistinguishable from modulation by similar unnatural image backgrounds (i.e. unnatural images synthesized with the Portilla and
Simoncelli texture algorithm). In our hands, background stimulus complexity mattered but naturalness did not.

The effects of saccades and background type on the amplitude of V1 responses were studied previously by MacEvoy et al. (2008). In that study we found that V1 responses were much less different and that responses in natural conditions were slightly larger than in reduced conditions. We attribute the differences between that study and the present one to the more refined experimental and analysis techniques used in the present work. In the MacEvoy et al. study, there was uncertainty in the start times of saccades because a 60Hz eye tracker was used compared to the 1kHz tracker used in the present study. Also, firing rates were analyzed in long epochs (500 msec). The present experiments used faster eye tracking (and consequent better estimates of saccade and fixation onsets) and shorter analysis windows, and provide a more accurate measure of V1 responses just after the start of new fixations.

The fact that one of our stimulus conditions included both saccades and complex-image backgrounds appears to be very important. There have been numerous studies of saccades and natural images separately, but few combined these two factors. Our results suggest that natural vision responses are unique because saccade and background factors occur in combination, and interact nonlinearly. This effect was most clear in the response amplitude. A complex background gives a lower response than a gray background and a saccade gives a lower response than flashing a stimulus. However, the combined effect of a stimulus introduced by a saccade on a complex background is much greater (much more reduced response amplitude) than would be predicted from the separate effects. This point is probably related to the report by Gallant et al.
(1998) that surround suppression is stronger in free viewing. Our stimulus paradigms were not as natural as free viewing, but they did allow a level of control making it possible to directly compare flash and saccade conditions.

As to the source of the response changes we see in the more natural conditions, we can only speculate. The reduced response after a saccade on a complex background could be a result of adaptation. It has been reported that, after exposure to natural images, the human contrast sensitivity function shows a selective loss in low spatial frequencies (Bex et al. 2009; Webster and Miyahara 1997). Further experiments will be required to assess the role, in our results, of adaptation to natural or natural-like complex images. In any case, note that the normal-day environment of a sighted primate is visually complex. Our results, together with those by Bex et al. and Webster and Miyahara, suggest that, in visual experiments, gray backgrounds between trials and isolated stimuli are actually elevating visual sensitivity at low spatial frequencies. This calls into question the generality of the standard contrast sensitivity function determined in most experiments.

Another contribution to the response changes we see in the more natural conditions, might come from the inhibitory circuitry of the LGN and V1. It has been shown, in cats, that electric stimulation of the optic tract produces fast and transient excitatory synaptic potentials in pyramidal neurons, quickly followed by a strong and long GABA-mediated hyperpolarization (Douglas and Martin 1991). Also in the cat, similar excitation/inhibition patterns can be produced by electrical stimulation of neighboring regions of V1 (Chung and Ferster 1998; Hirsch and Gilbert 1991). The response reduction, and the striking resemblance between our responses in real and simulated saccades on complex backgrounds, suggest that feed-forward and lateral
inhibition may be involved in the suppression of the response. When a saccade on a complex background occurs, the visual image redistributes abruptly on the retina. This would translate into a massive change in neural activity in the optic nerve, LGN, and V1. We hypothesize that the firing of many newly activated axons will activate feed-forward inhibitory interneurons in the LGN and V1, and that this transient of inhibition might account for much of the suppression that we observed after saccades on complex scenes.

Many of the other properties we studied, such as the timing of V1 responses and feature selectivity have not been previously examined in more natural paradigms, though there are studies (e.g. Guo et al. (2007)) showing that responses to bar stimuli are influenced by previous stimuli appearing beyond the receptive field. Combined with the decreased response amplitude noted already, our results show that V1 responses in more natural situations exhibit a spectrum of differences from responses in more reduced paradigms. Our working hypothesis is that the response differences in natural situations result from altered activation of the cortical network. By using simulated saccades we found that motion of the scene during saccades was responsible for much of the response difference in flash and saccade conditions. The effect of simulated saccades also distinguishes the response reduction we observed from saccadic suppression mediated solely by corollary-discharge. Corollary discharge signals seem to play a role, nevertheless, as real saccades decrease response amplitude somewhat more than simulated saccades.

Our results make predictions about how visual perception in more natural paradigms should differ from perception in simplified situations. For example, when saccades are made across
complex backgrounds, response amplitude is significantly lower than with flashed stimuli
presented during fixation. In an experiment such as contrast detection, it may take more contrast
to reach perceptual threshold with a saccade across a picture than with a flashed stimulus
presented in isolation. Our findings also make predictions about the effects on perception of
visual context. Context is hugely important in virtually every aspect of visual perception; there
are strong interactions between the lightness, color, orientation, and motion at different locations
in the visual field. For this reason it seems highly significant that we found that surround
interactions are much stronger in natural visual paradigms than in reduced situations. Virtually
everything that is known about visual context and the influence of surround stimuli is based on
fixation experiments with flashed stimuli. Our results suggest that surround interactions would
be significantly stronger, and more characteristic of natural vision, if experiments were
conducted with stimulus presentation via saccades. Finally, the data presented here suggest that
a careful perceptual study would reveal that the sensitivity to stimulus orientation, and perhaps
other attributes, is highest when new stimuli are viewed at the end of saccades. Moreover,
perceptual sensitivity should follow a timecourse that is unique to more natural vision.
Acknowledgements

We thank Aaron Gregoire for his skilled assistance with animal training and care, Dr. Moses Goddard for advice in surgical procedures, and Drs. Stuart Geman and Hernando Ombao for advice with statistical analyses. We thank the reviewers of this paper for their insightful comments and suggestions. This research was supported by grant EY09050 from the National Eye Institute.
References


Figure Legends

**Fig. 1.** Visual stimulation and behavioral paradigms. Monkeys were trained to fixate and follow a small red spot on a computer display while a bar of light probed V1 neuron responses. Each frame in the figure depicts the image shown on the screen, the fixation spot, and the position of the RF on the screen (yellow circle; not displayed in the actual experiment). Each panel represents the two main epochs within a trial: before (left frame) and after (right frame) the stimulus onset or a saccade. **A**, Flash-on-Gray condition: monkey fixates a red spot in the center of the display (left frame) and a bar of light is then flashed in the RF (right frame). **B**, Saccade-on-Gray: monkey fixates a spot at a peripheral location while the bar is already present (left frame); the peripheral fixation spot disappears and a new one appears at screen center (right frame); the monkey makes a saccade to the screen center (small red arrow) bringing the RF onto the bar of light (big red arrow). **C**, Flash-on-Picture: a natural scene fills the screen, except for a circular buffer patch that encircles the RF; the animal fixates centrally; the stimulus bar is flashed on the buffer. **D**, Saccade-on-Picture: a natural scene fills the screen except for a circular buffer patch that will encircle the RF after the saccade; the bar is present in the buffer; the animal fixates peripherally (left frame); the peripheral fixation spot disappears and a new one appears at screen center (right frame); the monkey makes a saccade to screen center (small red arrow).
bringing the RF onto the bar of light (big red arrow). The size of the buffer and the bar have been exaggerated for illustration purposes. In practice, the size of the bar was comparable to a small tree branch in the figure. The buffer radius was about 1.3 times the length of the bar. E, Saccade-on-Picture-NoBar: control condition to test the response produced by the saccade itself; similar to D, but a bar of light was never presented in the buffer. F, Simulated-Saccade: monkey fixates the center of the display during the entire trial; the scene, buffer and bar move, duplicating the retinal translation that occurs during a real saccade. All figures in the paper use the color code established in the panel titles of this figure.

Fig. 2. Background images used to assess the importance of naturalness in response modulation. A, Natural image scene taken from van Hateren and van der Schaaf (1998). B, Luminance histogram of the image in A: arrows show the contrasts of the stimulus bars used to probe the response of the RF, either 50% above or below the mean luminance level of the image. C, Normalized power spectrum of the original image; brighter indicates greater energy. Horizontal axis corresponds to horizontal spatial frequencies increasing, from the center, to the left and right borders. Vertical axis correspond to vertical frequencies. Center corresponds to zero spatial frequency. D, Synthetic “unnatural” image, derived from the image in A using the texture-synthesis algorithm by Portilla and Simoncelli (2000). E, Luminance histogram of the synthetic image in D. F, Power spectrum of the synthetic image. G, H, Two other “unnatural” images synthesized from A, using different initial seeds. The luminance histograms of the images in G and H are identical to E, their power spectra is identical to F.
Fig. 3. Temporal quantification of the neural response. The peri-stimulus time histogram (PSTH; vertical bars) shows a typical response of a V1 neuron to a bar flashed in its RF for 300 ms. The horizontal bars above the PSTH represent the time epochs used for quantitative analysis: pre-peak (red), peak (green) and late (blue). The pre-peak epoch (red) was defined for each neuron as a 16 ms window beginning at the minimum response latency across conditions. The peak epoch (green) was determined independently for each neuron and each condition using an automatic procedure to estimate the peak time, its width, and the latency of the response (see Methods). The late epoch was defined as a fixed interval ($t = 100$ to $200$ ms; blue bar) based on the end of the initial response transient of the grand average of all cells in our sample.

Fig. 4. The dependence of V1 responses on the mode of stimulus presentation and the type of background image. A, Response of one neuron to an optimally-oriented bar in four different conditions: Flash-on Gray (black), Flash-on-Pic (blue), Sac-on-Gray (magenta), and Sac-on-Pic (green). In Flash conditions, $t = 0$ indicates stimulus onset; in Saccade conditions, $t = 0$ is the end of the saccade that brought the stimulus into the RF. Note the similarity of the Flash-on-Gray and Flash-on-Pic responses (black and blue), and the larger latency of the Sac-on-Gray (magenta) and especially the Sac-on-Pic (green) response. Responses were normalized to the peak response in the Flash-on-Gray condition (black). B, A neuron showing greater response amplitude reduction in the saccade conditions (magenta and green) compared to the cell in A. C, Average responses from 74 V1 recording sites. Thick lines indicate responses to optimally-oriented bar stimuli and thin lines show responses to bars tilted $60^\circ$ CCW from optimal. Response latencies in Flash-on-Gray (black trace) and Flash-on-Picture (blue) conditions are similar; Sac-on-Gray
and Sac-on-Pic (green) conditions have progressively longer latencies. Response amplitude also decreases across conditions. Before averaging, all the responses from a particular cell were normalized to the Flash-on-Gray peak. The grand Flash-on-Gray average (black) is less than 1.0 because different cells have different latencies and time courses. The quantitative analysis summarized in Table 1, shows a mean peak value = 1 for the Flash-on-Gray condition because it used the peak values measured independently for each cell and each condition with the procedure described in Fig 3 and Methods. The Sac-on-Pic response (green) exhibits a small, early increase of activity around 25 ms, found to come from a subset of neurons that fired to the movement of the background scene during saccades (see Results and Fig. 5). D, Mean orientation selectivity of the cells in our sample. For each condition, orientation selectivity rises following the average responses, peaks and then exhibits a plateau for the rest of the response. The latency of the selectivity in different conditions follows the latencies of the average responses per condition. Three epochs of the orientation-selectivity timecourse are of notice: the different latencies in each condition, the initial peak of the Sac-on-Pic response (green trace, around $t = 60$ ms; larger than the other three), and the plateau corresponding to the late Flash-on-Picture response (blue trace, $t > 100$ ms; larger than the other conditions). These differences are discussed in Results.

Fig. 5. A sub-population of neurons had an unusual early response in the Sac-on-Pic condition (green). A, Single cell example. B, Average response of 7 cells that showed an early component in the Sac-on Pic condition. Color code as in Fig. 1 and 4; thick lines show optimal-orientation responses, thin lines show responses to bars tilted 60° CCW from optimal. The gray thin trace is
the response in the Sac-on-Gray-NoBar condition and the cyan thin trace is the response in the
Sac-on-Pic-NoBar condition. The Sac-on-Pic responses (green, thick and thin traces) shows an
early increase of activity between 10 and 30 ms. The early response in the Sac-on-Pic-NoBar
condition (cyan) is nearly identical in that period, suggesting that the early activity is due to
visual stimulation during the saccade rather than afterward.

Fig. 6. Responses in simulated saccade conditions are similar to responses with real saccades.
A, Average response of 42 cells to an optimally-oriented bar of light in the four principal
conditions (color code as in Figs. 1 and 4). The response in the simulated-saccade condition (red
trace) is largely similar to the response with a real saccade on a complex scene (green). However,
with real saccades the response amplitude is lower. B, Orientation selectivity in the Simulated-
Saccade condition (red) follows a pattern similar to the Sacc-on-Picture condition (green).

Fig. 7. Effects of background scene “naturalness”. A, Average responses in the Flash-on-Gray
(gray trace), Flash-on-Pic (blue traces), Sac-on-Pic (green traces), and Simulated-Saccade
conditions (red traces). Traces of the same color show responses in the same condition with
natural and “unnatural” background images. Multiple unnatural images, such as those in Figure
2, gave responses indistinguishable from responses to the natural images they were derived from.
B, Orientation selectivity of the cells shown in A. No systematic changes were observed as a
function of the “naturalness” of the background scene.
Table 1

Latency, peak magnitude and orientation selectivity of the responses to different contexts and modes in which stimulus enters the RF of V1 neurons.

<table>
<thead>
<tr>
<th>Condition (and color used in Figures)</th>
<th>Response Latency</th>
<th>Peak Time</th>
<th>Normalized amplitude</th>
<th>Orientation selectivity</th>
<th>Amplitude</th>
<th>Orientation selectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flash-on-Gray (black)</td>
<td>35 ± 8 ms</td>
<td>50 ± 8 ms</td>
<td>1</td>
<td>0.35 ± 0.3</td>
<td>0.43 ± 0.2</td>
<td>0.40 ± 0.3</td>
</tr>
<tr>
<td>Flash-on-Picture (blue)</td>
<td>36 ± 8 ms</td>
<td>52 ± 9 ms</td>
<td>0.92 ± 0.2</td>
<td>0.43 ± 0.3</td>
<td>0.35 ± 0.2</td>
<td>0.51 ± 0.3</td>
</tr>
<tr>
<td>Sac-on-Gray (magenta)</td>
<td>42 ± 10 ms</td>
<td>57 ± 10 ms</td>
<td>0.85 ± 0.2</td>
<td>0.33 ± 0.3</td>
<td>0.33 ± 0.2</td>
<td>0.35 ± 0.3</td>
</tr>
<tr>
<td>Sac-on-Picture (green)</td>
<td>46 ± 13 ms</td>
<td>64 ± 13 ms</td>
<td>0.59 ± 0.2</td>
<td>0.47 ± 0.3</td>
<td>0.30 ± 0.2</td>
<td>0.42 ± 0.3</td>
</tr>
<tr>
<td>Sim-Sac on Picture (red)</td>
<td>44 ± 12 ms</td>
<td>65 ± 14 ms</td>
<td>0.70 ± 0.2</td>
<td>0.57 ± 0.3</td>
<td>0.37 ± 0.2</td>
<td>0.55 ± 0.3</td>
</tr>
</tbody>
</table>

All values reported are mean ± standard deviation.

All amplitudes, included those of the late response phase, are expressed as relative to the Flash-on-Gray peak amplitude.

The values in the first 4 rows were obtained from our entire sample of 74 recording sites. The values reported in the last row (shaded) come from a subset of that sample, with 42 recording sites.
Example from one unit

Average from 74 units

Example from another unit

Orientation selectivity: average from 74 units

Color code:
- Flash on Gray
- Flash on Picture
- Saccade on Gray
- Saccade on Picture
Example: Unit showing an early-response in the Sac-on-Pic

A

![Graph A]

Normalized Firing Rate

Time (s)

0 0.05 0.1 0.15 0.2

B

Average of 7 "early-response" units

![Graph B]

Normalized Firing Rate

Time (s)

0 0.05 0.1 0.15 0.2

Color code:

- Flash on Gray
- Flash on Picture
- Sac-on-Pic-NoBar

- Saccade on Gray
- Saccade on Picture
- Sac-on-Gray-NoBar
Simulated-saccade responses

A

Averages

Normalized Firing Rate

Time (s)

B

Orientation selectivity

Orientation Selectivity

Time (s)

Color code:

Flash on Gray          Saccade on Gray
Flash on Picture       Saccade on Picture
Sac-on-Pic-NoBar       Sac-on-Gray-NoBar
                        Simulated Saccade on Picture
Natural vs. "unnatural" background

A
Average responses

B
Orientation selectivity

Color code:
Flash on Gray  Simulated Saccade on Picture
Flash on Picture  Saccade on Picture