neuroReport

NeuroReport 9, 2153–2158 (1998)

Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing

Jack L. Gallant,[1,CA] Charles E. Connor,[2] and David C. Van Essen

Department of Anatomy and Neurobiology, Washington University Medical School, Box 8108, 660 S. Euclid Ave., St. Louis, MO 63110; [1]Department of Psychology, #1650, University of California at Berkeley, Berkeley, CA 94720-1650; [2]Krieger Mind-Brain Institute, 338 Krieger Hall, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218–2685, USA

[CA]Corresponding Author

UNDER natural viewing conditions primates make frequent exploratory eye movements across complex scenes. We recorded neural activity of 62 cells in visual areas V1, V2 and V4 in an awake behaving monkey that freely viewed natural images. About half of the cells studied showed a modulation in firing rate following some of the eye movements made during free viewing, though the proportions showing a discernible modulation varied across areas. These cells were also examined under controlled viewing conditions in which gratings or natural image patches were flashed in and around the classical receptive field while the animal performed a fixation task. Activity rates were generally highest with flashed gratings and lowest during free viewing. Flashed natural image patches evoked responses between these two extremes, and the responses were higher when the patches were confined to the classical receptive field than when they extended into the non-classical surround. Thus the reduction of activity during free viewing relative to that obtained with flashed gratings is partly attributable to natural images being less effective stimuli and partly to suppressive spatio-temporal neural mechanisms that are important during natural vision.

Key words: Free viewing; Natural image; Visual cortex; V1; V2; V4

Introduction

Natural vision differs from visual stimulation during conventional neurophysiological experiments in two important respects. First, natural images are quite different from the stimuli typically used in controlled experiments. Natural images have broadband Fourier and color spectra,[1–5] and luminance contrast varies widely within and between images.[6] In addition, natural images are present continuously and extend across most of the visual field. Second, during natural vision primates make exploratory eye movements across the visual scene.[7,8] These eye movements typically consist of 100–400 ms fixations separated by 20–40 ms saccadic eye movements from one point to another. The fixations themselves include small eye movements consisting of microsaccades and irregular drift.

These differences suggest that neural responses evoked during natural vision will be affected by complex linear and nonlinear processes such as contrast normalization and non-classical surround interactions, and by mechanisms of para-saccadic masking.[9–13] In fact previous studies in anesthetized animals using natural images[14] or other broadband stimuli[15,16] have already revealed complex response properties not observed with simpler stimuli.

As a next step in understanding natural vision we have recorded from visual cortical cells in a monkey freely viewing natural scenes. We compare the free viewing activity to that obtained under controlled viewing conditions with flashed gratings and natural image patches. Preliminary versions of these results have been reported elsewhere.[17,18]

Methods

All surgical, training and recording procedures complied with USDA and NIH guidelines and were conducted according to a University-approved animal protocol. Methods were generally similar to those described previously,[19] except as noted below.

Equipment: Stimuli were presented on a Silicon Graphics Indigo workstation RGB monitor (1280 × 1024 pixels, 29.5 × 23.6 cm, 32.9 × 25.3° at 50 cm viewing distance). For experiments requiring controlled fixation, and for early free viewing experiments (all 20 V4 cells and 12 of 17 V2 cells), eye position was digitized to 8 bits at 50 Hz (Z-80-based A/D controller, Caltech Biology Shop). In later free viewing experiments (25 V1 cells and 5 of 17 V2 cells) eye position was digitized to 12 bits at 1 kHz (Keithly Metrabyte A/D board, Brainstorms Monitoring Corp.). There were no significant differences between these systems for the analyses presented here, so they are presented together.

Eye calibration: Eye position was monitored using the scleral search coil method.[20] Eye coil voltages were mapped to the screen via a calibration procedure run each day. The animal fixated points on a grid of 9 (in early experiments) or 35 (later experiments) locations distributed across the monitor. The animal ran 4–6 5 s fixation trials at each point before proceeding to the next location. Calibration data were fit with a plane or third order polynomial. The calibration used for V1 recording was typically accurate to 3 pixels (0.07°) at the center of the screen, and 10 pixels (0.2°) at the edge of the screen. Thus in the worst case the error might be a substantial fraction of a V1 CRF (though much less than a typical V2 or V4 CRF). The analyses presented in this paper are not critically dependent on exact eye position estimates, and the conclusions would not be affected by modest inaccuracies in calibration.

Free viewing: During free viewing the animal viewed a sequence of 4–8 natural images, each presented for 5–10 s, and separated by intervals of 30 s to several minutes. No behavioral response was required and no liquid reward was provided, leaving the animal free to look at any part of the image or not to look at it at all. Free viewing was performed with the room lights off, but during the inter-image interval the room light was often turned on to prevent dark adaptation and keep the animal awake and alert. A liquid reward was occasionally given during the inter-image interval to enhance alertness. A wide range of natural images were selected from photo-CDs (Corel Corporation), including pictures of animals, fruits and vegetables, forest scenes, people and residential interiors. For display, photo-CD images were converted to RGB using a gamma value of 0.7, and only the central 1280 × 1024 region of each 1536 × 1280 pixel image was shown. Artificial images such as noise patterns and checkerboards were also used occasionally.

Grating test: The grating experiments used 30 Cartesian, 60 polar and 20 hyperbolic sinusoidal gratings, presented at either 2 or 3 phases.[21] (For V1 cells, only Cartesian gratings were used.) Gratings were shown at maximum contrast in the optimal color for the cell under study. The outer 10% of each circular grating patch was blended into the gray background by ramped dithering. For V1 cells each trial used eight randomly chosen gratings, flashed in the receptive field for 250 ms each and separated by a 250 ms blank interval. For V2 and V4 cells each trial used four randomly chosen gratings flashed in the receptive field for 500 ms each and separated by a 700 ms blank interval. During a trial the animal was required to maintain fixation within a 0.25–0.5° diameter window.

Image patch test: The image patches that fell within the CRF of a cell during free viewing were extracted via an automated procedure. The procedure registered a fixation when the eye remained within an 0.3 CRF-diameter window for at least 70 ms, and a change of fixation was registered when the eye moved more than 0.3 receptive field diameters from its original location, irrespective of whether the change was caused by a saccade, microsaccade, or drift. Each image patch test included 10–25 such patches. The patches were presented in and around the cell's CRF while the animal performed a fixation task. Two image patch conditions were interleaved in most experiments. In the first condition the patches were confined to the CRF of the cell under study. In the second the patches were three times the size of the CRF, and included a larger portion of the original image. On each trial four random image patches were flashed for 500 ms each and were separated by a 700 ms blank interval. Fixation requirements were the same as for the grating test.

Contrast: Image contrast was expressed as rms contrast, the average rms deviation of each pixel within a region relative to the mean luminance within that region.[6] Gratings had full Michelson contrast and an rms contrast of about 0.7. Patches randomly selected from the natural images ranged from < 0.001 to > 2.0 rms contrast. For a few cells the average rms contrast across the sequence of all image patches falling in the CRF during free viewing was computed; it ranged from about 0.3 to 0.6. Thus the rms contrast of the portions of natural images that entered the CRF during free viewing was somewhat lower than that of the gratings, but should have been suprathreshold for most V1 neurons.

Stimulus averaged rate: In order to compare activity observed during free viewing with the activity evoked in conventional experiments we used the stimulus averaged rate (SAR). To calculate the SAR a window of specified duration (e.g. 50 ms) was moved in 10 ms increments, beginning at the onset of each fixation or stimulus flash. The mean firing rate across all fixations or stimuli was calculated for each window position, and the maximum average rate (across all window positions) was defined as the SAR. Stimulus averaging of this sort is necessary because the reliability of each unique free viewing fixation is likely to be low;[22] to obtain comparable rate estimates this method was also used for the grating and image patch tests.

SAR analyses were performed for window lengths of 50, 100 and 200 ms. The three window sizes produced similar results; to reduce noise the SARs obtained with the three window lengths were aver-

aged together. (Averaging over different windows can increase signal strength by reducing the effects of uncorrelated noise.) SARs could not be calculated for large windows when the peak of activity was too near the beginning or end of the stimulation period, and so some cells were excluded from some analyses.

Comparisons of the SARs obtained in different test conditions were made using two methods: a non-parametric binomial sign test and a randomized

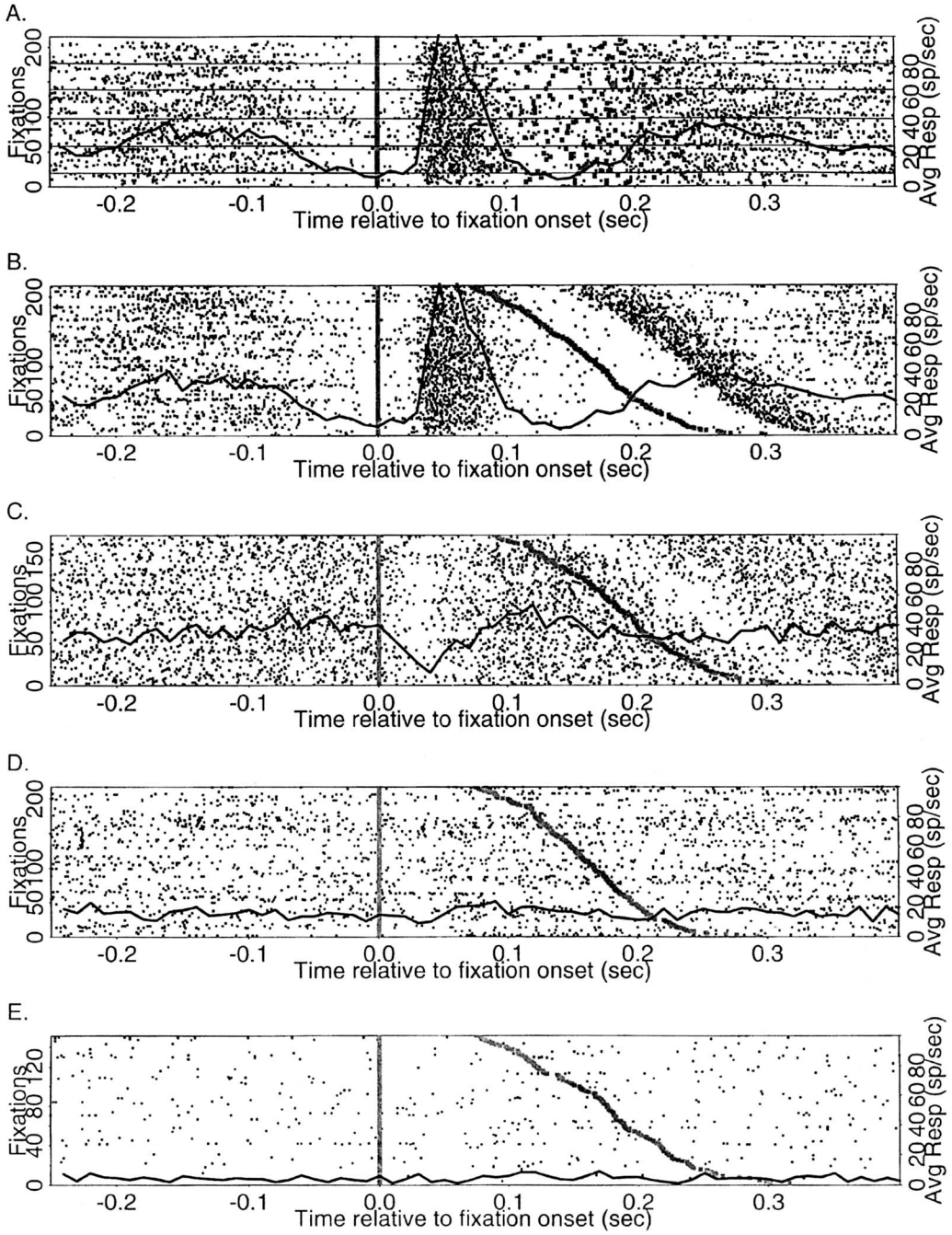

FIG. 1. **(A)** Activity of one V1 cell during free viewing of six natural images. Plots only include data from fixations in which both the fovea and the receptive field were within the boundaries of the image. Each row represents the epoch around one fixation, and fixations are listed in the order they occurred during free viewing, from bottom to top. The thin horizontal lines separate fixations made to different images. All fixations are aligned at onset, given by the thick vertical line. Fixation offsets are given by the large dots scattered to the right of the plot; fixation duration was variable because it was under the animal's control. Small dots represent individual spikes. The activity rate in each 10 ms bin around fixation onset is overlaid on the plot (PSTH), and corresponds to the scale at right. This cell gives a phasic discharge after each eye movement, when a new stimulus enters the receptive field. **(B)** The same data shown in (A), except with fixations sorted according to duration. **(C)** Sorted free viewing activity profiles for another V1 cell showing phasic suppression following an eye movement. **(D)** Sorted profiles for a V2 cell that shows little evidence of modulation after eye movements made during free viewing. **(E)** Sorted profiles for another V2 cell showing no modulation during free viewing.

nonparametric analog of the pairwise t-test.[21] In the latter test the distribution of the test statistic under the null hypothesis of no difference in the SARs was obtained by pairwise randomization of data from the conditions being compared.

Results

Free viewing: The activity patterns obtained during free viewing from different visual cortical cells were highly variable, both within and between areas. Some cells gave a robust phasic discharge after almost every eye movement (Fig. 1A,B), while a few cells showed phasic suppression (Fig. 1C). Many other cells, particularly in areas V2 and V4, showed no clear modulation with respect to eye movements (Fig. 1D,E).

In order to objectively identify cells whose activity was modulated following eye movements made during free viewing, we used a combination of two criteria. A cell was categorized as modulated if either the PSTH showed a change of 20 sp/s in any 15 ms bin from 50 ms before to 200 ms after fixation, or the PSTH showed a change of 10 sp/s in any 15 ms bin from 50 ms before to 200 ms after fixation, and that change occurred on the maximum slope of an S-shaped (ogive) section of the PSTH waveform profile. These two criteria together identified cells with either fast or slow modulation patterns, and most importantly they produced a classification that was consistent with visual assessment.

According to this quantitative classification scheme the cells shown in Fig. 1A–C were modulated, while the cells in Fig. 1D,E were unmodulated. Across the entire sample of 62 cells, 26 (42%) showed modulation related to eye movements made during free viewing. In area V1 14 out of 24 cells (58%) were modulated, while three of 17 V2 cells (18%) and 9 of 20 V4 cells (45%) were modulated. Of the 14 modulated V1 cells, 10 showed phasic enhancement and four showed phasic suppression, while all of the modulated V2 and V4 cells showed phasic enhancement. Of the cells showing modulation, most showed substantial variability in activity after individual eye movements. This variability presumably depended on the structure of the stimuli in the CRF during each fixation, and on other factors such as the sequence of fixations and selective attention.

Our ability to detect modulation in the free viewing data was limited by the size of the data set obtained for each cell (generally < 1 min in total duration). One way to increase the signal-to-noise ratio in noisy data is burst filtering, in which individual spikes are accepted only if they form part of a brief burst of spikes.[23] To test whether burst filtering might reveal modulation in any of the cells that were unmodulated according to our PSTH analysis we reanalysed the unmodulated cells after recording bursts as single neural events and eliminating non-burst spikes. Two different burst filters were used, four spikes occurring within 50 ms and four spikes in 25 ms. Burst filtering did not reveal modulation in any of the cells designated as unmodulated by the previous classification.

Gratings: Figure 2A shows the SARs obtained with gratings *vs* those obtained during free viewing. For the free viewing data the SAR quantifies the rate of activity obtained after each fixation, averaged over all fixations. For the grating test the SAR quantifies the rate of activity obtained after the onset of each grating, averaged over all gratings used in the test. The ratio between the free viewing and grating SARs is 0.32; across the sample free viewing SARs were about one-third as large as those obtained with flashed gratings. The SARs obtained in the two conditions are significantly different according to a binomial sign test ($p = 0.02$), but not according to a randomized pairwise t-test ($p = 0.13$). The failure of the t-test to reach significance is probably due to the large variation in the SAR ratio across cells. In particular a few V1 cells produced markedly higher SARs with flashed gratings than during free viewing of natural images (two square symbols at bottom right of Fig. 2A), while some V4 cells produced markedly higher SARs during free viewing of natural images than with flashed gratings (two diamond symbols at upper left of Fig. 2A). Thus although free viewing SARs are on average about one-third the size of grating SARs, the SAR ratio obtained for a specific cell may deviate substantially from this relationship.

Image patches: Given the smaller SARs obtained during free viewing relative to the grating test, an obvious question is whether this is because natural images are less effective visual stimuli than are high contrast gratings. This question can be addressed by comparing the SARs obtained under controlled viewing conditions with flashed gratings to those obtained with natural image patches flashed within the CRF. As shown in Fig. 2B, the ratio of grating to image patch SARs is 0.73. This is not a significant difference according to the sign test ($p = 0.15$) or the paired t-test ($p = 0.61$). Thus although the SARs obtained with image patches flashed in the CRF are on average slightly lower than those obtained with flashed gratings, this difference is modest and is not statistically significant.

The SARs obtained with flashed images are compared to those obtained during free viewing in Fig. 2C. (Recall that the flashed image patches were a subset of all the natural images that had fallen in the CRF during free viewing.) The ratio of free

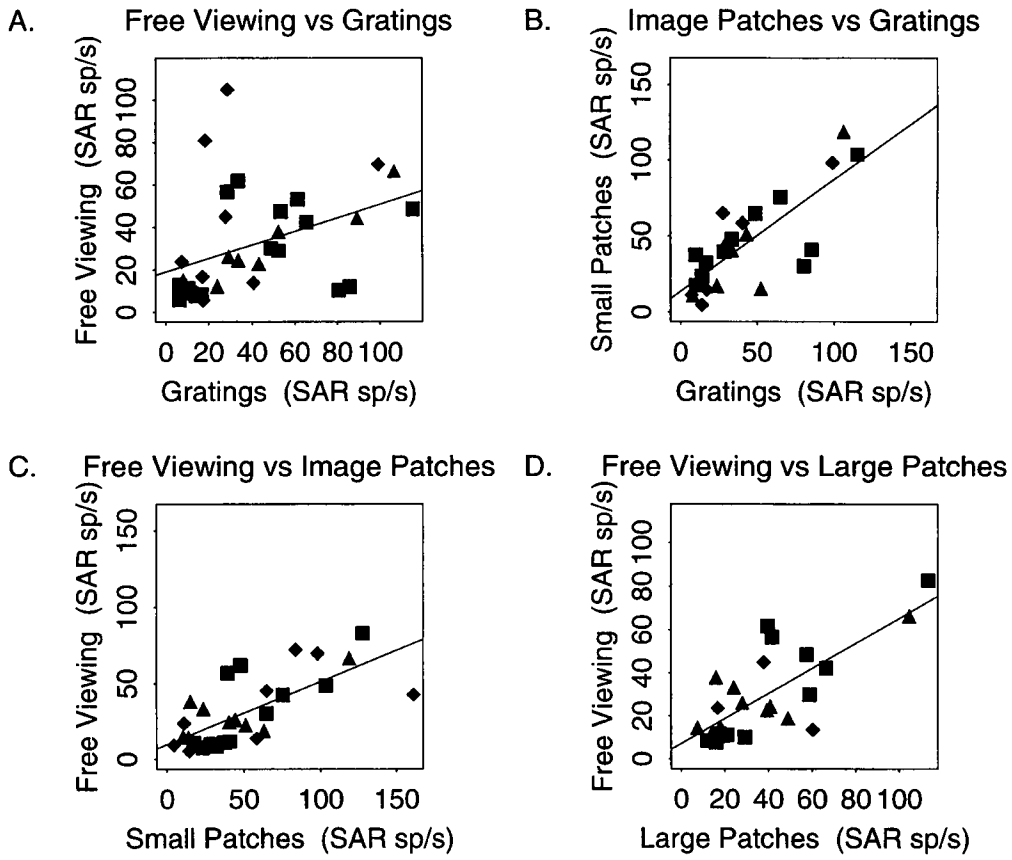

FIG. 2. Comparison of stimulus averaged rates (SARs) obtained under various test conditions. The SAR is the rate of activity obtained after each fixation or stimulus onset, averaged over all fixations or stimuli. For these analyses SARs have also been averaged across time windows of 50, 100 and 200 ms. All panels refer to the same sample of cells, but some cells have been excluded from some plots because of limitations on computing the SARs. (**A**) Comparison of SARs obtained with flashed gratings and those obtained during free viewing. This analysis included 16 V1 cells (squares), 8 V2 cells (triangles) and 10 V4 cells (diamonds). The SAR ratio is 0.32. (**B**) Comparison of SARs obtained with flashed gratings and with flashed image patches confined to the CRF, for 11 V1 cells, 7 V2 cells and 6 V4 cells (symbols same as in panel A). The SAR ratio is 0.73. (**C**) Comparison of SARs obtained with flashed image patches confined to the CRF and during free viewing, for 12 V1 cells, 11 V2 cells and 8 V4 cells. The SAR ratio is 0.41. (**D**) Comparison of SARs obtained with flashed image patches that encompassed both the CRF and the nearby surround *vs* those obtained during free viewing, for 12 V1 cells, 11 V2 cells and 3 V4 cells. The SAR ratio is 0.58.

viewing to image patch SARs is 0.41, a significant difference in SARs (sign $p = 0.01$; t-test $p = 0.001$). Thus SARs obtained during free viewing are substantially smaller than those obtained when the same natural image patches are flashed within the CRF.

The data presented thus far suggest that the large difference between SARs obtained during free viewing versus with flashed gratings is only partly attributable to the composition of the stimuli falling within the CRF. Another important factor may be the suppressive effects of the natural image surrounding the CRF.[9,12,24,25] The SAR ratio for flashed image patches 2–3 times larger than the CRF versus those confined to the CRF is 0.78. This difference just misses significance by the sign test ($p = 0.076$) and is significant according to the t-test ($p = 0.002$). Figure 2D compares SARs obtained during free viewing with those obtained with flashed image patches 2–3 times larger than the CRF. The ratio is 0.58, again a significant difference (sign $p = 0.03$; t-test $p = 0.02$).

To summarize, in this study the largest SARs were obtained with flashed gratings, but flashed natural image patches that were confined to the CRF produced responses that were on average nearly as large. Larger flashed natural image patches that encroached on the non-classical surround produced substantially smaller SARs, while the lowest SARs were obtained during free viewing. Individual cells showed substantial variation in this pattern, particularly in terms of the relationship between responses obtained with gratings versus natural image stimuli.

Discussion

During natural vision each eye movement brings a new stimulus into the receptive field, so one might expect that cells would generally show modulation following some but not all eye movements made during free viewing of natural scenes. In these experiments discernible modulation was observed in about 55% of V1 cells, 20% of V2 cells and 45% of V4

cells. We emphasize that these percentages probably underestimate the true incidence of modulation of activity during free viewing because the sample size is fairly small and the free viewing periods are fairly short. The different proportions observed in the three areas will presumably change when a larger number of cells is studied using longer periods of free viewing, or when more sensitive methods are used to detect modulation.

We compared activity obtained during free viewing of natural scenes to controlled viewing of gratings by means of the SAR measure. Free viewing SARs were on average only about one-third as large as those obtained with flashed gratings. In other words the modulation observed during free viewing of natural scenes was on average about one-third as large as the modulation evoked by a typical flashed sinusoidal grating. Note, however, that there is significant scatter in these data, and some cells had higher SARs during free viewing than when flashed gratings were used (see Fig. 2A).

There are several reasons why modulation might be relatively low during free viewing of natural scenes. First, natural images may be significantly less effective stimuli than sinusoidal gratings. An arbitrary natural image patch, chosen at random, is unlikely to contain much power within a given cell's multidimensional passband[3] and will tend to have low contrast.[6] Second, responses in area V1 can be actively suppressed by frequency and orientation components outside a cell's passband[9] and responses of striate as well as extrastriate areas are suppressed by stimuli falling in the non-classical surround.[12,24,25] Third, the stimulus changes arising from saccades across a natural scene have a slower onset than the abruptly flashed stimuli used in conventional experiments, and these onset differences may have large effects on responsiveness.[10,26]

The effects of image structure were examined by flashed presentation of a subset of the natural image patches that had entered the receptive field during free viewing. The free viewing SARs were significantly smaller than the SARs obtained with natural image patches flashed in the CRF even though these patches were a subset of those that had entered the CRF during free viewing. Enlarging the flashed patches so that they extended beyond the CRF reduced the difference between the SARs obtained during free viewing and with flashed patches, but the flashed condition still produced larger responses. This is consistent with observations of suppression from the non-classical surround.[9,12,24,25] The remainder of the difference may reflect other factors including para-saccadic suppression,[10] attention and arousal effects,[19] and stimulus onset characteristics that were not systematically examined in these experiments.[10,26]

A more sensitive method for testing the efficacy of the stimuli occurring in and around the CRF during free viewing is to construct a movie representing the stimuli that fell in the CRF during free viewing, and to replay this movie in and around the CRF while the animal performs a fixation task. Preliminary experiments using this review procedure in V1 and V2 reveal highly consistent patterns of activity in cells whose activity was apparently unmodulated in the basic free viewing experiment.[18] They also reveal a high correlation between free viewing and review data, suggesting that attentional effects in the free viewing data obtained from areas V1 and V2 are relatively minor.[18]

Suppression of activity during free viewing of natural scenes reflects the operation of several linear and non-linear suppressive mechanisms whose joint effects are larger under natural viewing conditions than in typical controlled viewing situations. Although these mechanisms have been described previously, their importance in natural vision may not have been fully appreciated. They may increase visual efficiency by normalizing responses via contrast gain control[11,13] or by enhancing local texture contrast.[12] They may also improve transmission bandwidth by reducing responses to local correlations present in natural images[14], thereby contributing to sparse coding.[3]

References

1. Burton GJ and Moorhead IR. *Appl Opt* **26**, 157–170 (1987).
2. Dong DW and Atick JJ. *Network: Computation in Neural Systems* **6**, 345–358 (1995).
3. Field DJ. *J Opt Soc Am A* **4**, 2379–2394 (1987).
4. Ruderman DL and Bialek W. *Phys Rev Lett* **73**, 814–817 (1994).
5. Tolhurst DJ, Tadmor Y and Chao T. *Ophthalmol Phys Opt* **12**, 229–232 (1992).
6. Pelli E. *J Opt Soc Am A* **7**, 2032–2040 (1990).
7. Burman DD and Seagraves MA. *J Neurophysiol* **71**, 1266–1271 (1994).
8. Keating CF and Keating EG. *Perception* **11**, 211–219 (1982).
9. Bonds AB. *Vis Neurosci* **2**, 41–55 (1989).
10. Judge SJ, Wurtz RH and Richmond BJ. *J Neurophysiol* **43**, 1133–1155 (1980).
11. Heeger DJ. *Vis Neurosci* **9**, 181–197 (1992).
12. Knierim J and Van Essen DC. *J Neurophysiol* **67**, 961–980 (1992).
13. Wilson HR and Humanski R. *Vis Res* **33**, 1133–1149 (1993).
14. Dan Y, Atick JJ and Reid RC. *J Neurosci* **16**, 3351–3362 (1996).
15. Lehky SR, Sejnowski TJ and Desimone R. *J Neurosci* **9**, 3566–3581 (1992).
16. Reid RC, Victor JD and Shapley RM. *Vis Neurosci* **9**, 39–45 (1992).
17. Gallant JL, Connor CE, Rakshit S and Van Essen DC. *Soc Neurosci Abstr* **20**, 838 (1994).
18. Gallant JL. *ARVO Abstr* **37**, 674 (1996).
19. Connor CE, Preddie DC, Gallant JL and Van Essen DC. *J Neurosci* **17**, 3201–3214 (1997).
20. Robinson DA. *IEEE Trans Biomed Eng* **10**, 137–145 (1963).
21. Gallant JL, Connor CE, Rakshit S *et al. J Neurophysiol* **76**, 2718–2739 (1996).
22. Tolhurst DJ, Movshon JA and Dean AF. *Vis Res* **23**, 775–785 (1983).
23. Bair W and Koch C. *Neural Comp* **8**, 1185–1202 (1996).
24. DeAngelis GC, Freeman RD and Ohzawa I. *J Neurophysiol* **71**, 347–374 (1994).
25. Desimone R and Schein SJ. *J Neurophysiol* **57**, 835–868 (1987).
26. Egeth HE and Yantis S. *Annu Rev Psychol* **48**, 269–297 (1997).

ACKNOWLEDGEMENTS: We thank Bruno Olshausen and Thomas Coogan for help in designing these experiments, and Heather Drury for software support. This research was supported by NIH grant EY02091 and the McDonnell Center for Higher Brain Function.

Received 30 September 1997; accepted 15 October 1997