Current Biology 22, 510–514, March 20, 2012 ©2012 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2012.01.050

Report

Temporal Encoding of Spatial Information during Active Visual Fixation

Xutao Kuang,[1] Martina Poletti,[1] Jonathan D. Victor,[3] and Michele Rucci[1,2,*]
[1]Department of Psychology
[2]Program in Neuroscience
Boston University, Boston, MA 02215, USA
[3]Department of Neurology and Neuroscience, Weill Cornell Medical College, New York, NY 10065, USA

Summary

Humans and other species continually perform microscopic eye movements, even when attending to a single point [1–3]. These movements, which include drifts and microsaccades, are under oculomotor control [2, 4, 5], elicit strong neural responses [6–11], and have been thought to serve important functions [12–16]. The influence of these fixational eye movements on the acquisition and neural processing of visual information remains unclear. Here, we show that during viewing of natural scenes, microscopic eye movements carry out a crucial information-processing step: they remove predictable correlations in natural scenes by equalizing the spatial power of the retinal image within the frequency range of ganglion cells' peak sensitivity. This transformation, which had been attributed to center-surround receptive field organization [17–19], occurs prior to any neural processing and reveals a form of matching between the statistics of natural images and those of normal eye movements. We further show that the combined effect of microscopic eye movements and retinal receptive field organization is to convert spatial luminance discontinuities into synchronous firing events, beginning the process of edge detection. Thus, microscopic eye movements are fundamental to two goals of early visual processing: redundancy reduction [20, 21] and feature extraction.

Results and Discussion

To determine the influence of fixational eye movements (FEM), we examined the visual input to the retina during periods of visual fixation. The eye movements of human observers were recorded during free viewing of pictures of natural scenes, a condition in which saccades occurred frequently (average intersaccadic duration: 248 ± 54 ms) and possessed highly variable amplitude (Figure 1C). As expected from previous studies [22], microsaccades were rare in this condition, and their frequency varied substantially across subjects, ranging from approximately one microsaccade every 2 s to one every 20 s (average rate: 0.17 ± 0.15 microsaccades/s). Nevertheless, the eye was continually in motion, as it drifted with apparently random trajectories in between saccades (Figures 1D and 1E). Across all subjects, the probability distribution of the center of gaze during the intersaccadic periods covered an area of 363 ± 151 arcmin^2. Thus, the projection of the scene moved significantly on the retina during fixation.

To examine the impact of fixational instability on visual input signals, we reconstructed the spatiotemporal stimulus on the retina, i.e., the visual input resulting from scanning the image according to the subject's eye movements. We then quantified the characteristics of this signal via spectral analysis (Figure 2).

Figure 2A shows the results obtained during the intersaccadic periods of ocular drift. With spatial frequency held constant, power declined as a function of temporal frequency (Figure 2B). However, with temporal frequency held constant at a nonzero value, power was equalized over a wide range of spatial frequencies (Figure 2C). This was strikingly different from the spatial characteristics of the images displayed on the monitor: as is typical for pictures of natural scenes [23], the spectral density of the images in our database declined steeply with spatial frequency ($k^{-1.9}$; k denotes spatial frequency). In sum, the effect of fixational eye movements was to redistribute the spatial power of the scene across temporal frequencies so that the power available in the form of temporal modulations (the sum of all power at resolvable temporal frequencies) was constant up to approximately 10 cycles/deg (Figure 2D).

These results were robust with respect to the method used for spectral analysis. Figure 2C shows that equalization of spatial power also occurred when eye movements were modeled by the best-fitting Brownian motion process, a case in which the influence of eye movements can be expressed in closed form (see Supplemental Experimental Procedures available online). Furthermore, very similar results were obtained by including in the analysis the periods of fixation that also contained microsaccades in addition to ocular drift (Figure 2D). Although the fixations recorded in these experiments were on average too short to allow spectral estimation below 4 Hz, power equalization could be assessed down to 1 Hz when observers were instructed to maintain prolonged fixation on a single point in the image. In all cases, fixational eye movements attenuated the low spatial frequencies, which predominate in natural images, yielding a retinal stimulus with "whitened" spatial distributions in the temporal domain. This transformation is significant for neural coding, because it occurs within the range of peak spatiotemporal sensitivity of neurons in the retina and lateral geniculate nucleus [24–26] (i.e., 0.1–10 cycles/deg and above 2 Hz).

To determine the factors responsible for the spatial whitening, we investigated the dependence of input signals on the probability distribution of eye movement. Under the assumption that fixational eye movements are independent of the observed image I, the spectrum of the retinal stimulus is given by:

$$S(k,\omega) = I(k)Q(k,\omega) \quad (1)$$

where $I(k)$ represents the spectral density of the image, and $Q(k,\omega)$ is the Fourier transform of the probability, $q(x,t)$, that the eye moves by x in an interval t (Supplemental Experimental Procedures). This equation provided excellent approximation of the input power spectrum estimated in the experiments (Figure 2D). It shows that the spectral distribution of the retinal stimulus depends both on the image and on the way the eye moves.

*Correspondence: mrucci@bu.edu

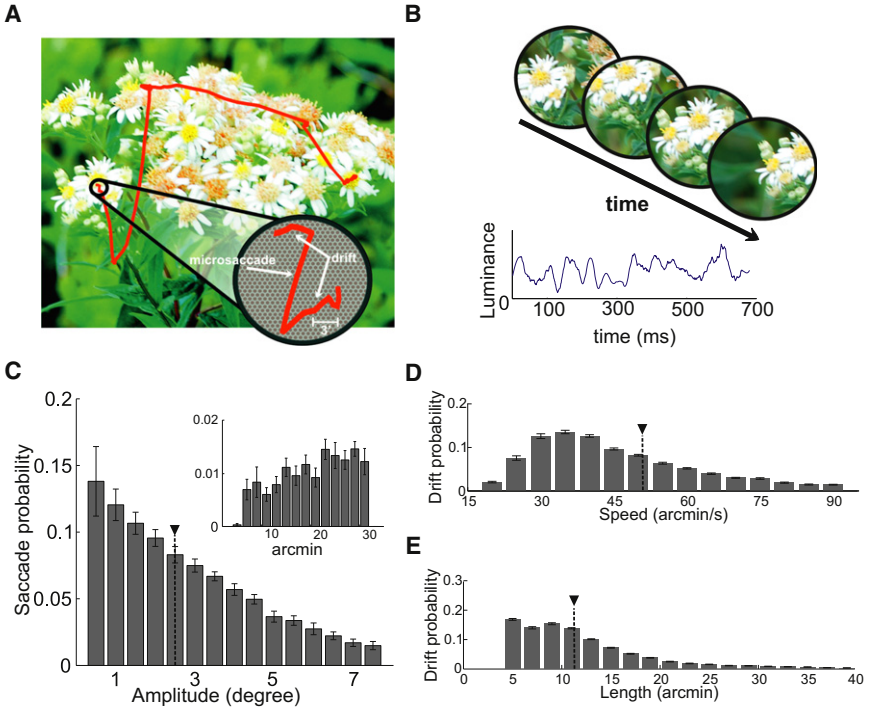

Figure 1. Microscopic Eye Movements Modulate Visual Input Signals during Normal Fixation

(A) An example of recorded eye movements superimposed on the observed image. The enlarged region shows the eye movements occurring during a period of fixation in relation to the size of photoreceptors.

(B) The resulting spatiotemporal stimulus impinging on a region of the retina (top) and on a single photoreceptor (bottom). Note the luminance fluctuations in the input signal.

(C–E) Characteristics of eye movements during free viewing of natural scenes. Average distributions of saccade amplitudes (C), ocular drift speeds (D), and lengths (E) over $N = 14$ observers. The insert in (C) shows the range of microsaccades. Data in (E) represent the arc lengths of the trajectories followed by the eye during the intersaccadic periods. Black triangles indicate means. Error bars represent one SD.

The probability density function of ocular drift and its frequency distribution are reported in Figures 3A and 3B, respectively. For a wide range of nonzero temporal frequencies, $Q(k, \omega)$ increased proportionally to the square of the spatial frequency. This dependence counterbalanced the spectral distribution of natural scenes [$I(k)$ in Equation 1], yielding a retinal stimulus whose power did not depend on spatial frequency. Thus, whitening of the retinal stimulus originated from the interaction between the statistics of natural images and the spatiotemporal structure of fixational eye movements.

Equation 1 predicts that whitening of the retinal stimulus by normal fixational instability will only occur during viewing of images with approximately scale-invariant spectral density ($I(k) \propto k^{-2}$), like natural images. Figure 3C tests this prediction. It shows the input temporal power measured when the same eye movement traces scanned images whose power spectra declined according to other power-law functions ($I(k) \propto k^{-\alpha}$, for α = 3,4,5). No power equalization occurred with these images. Whitening was also greatly attenuated when images of natural scenes were scanned by abnormally large eye

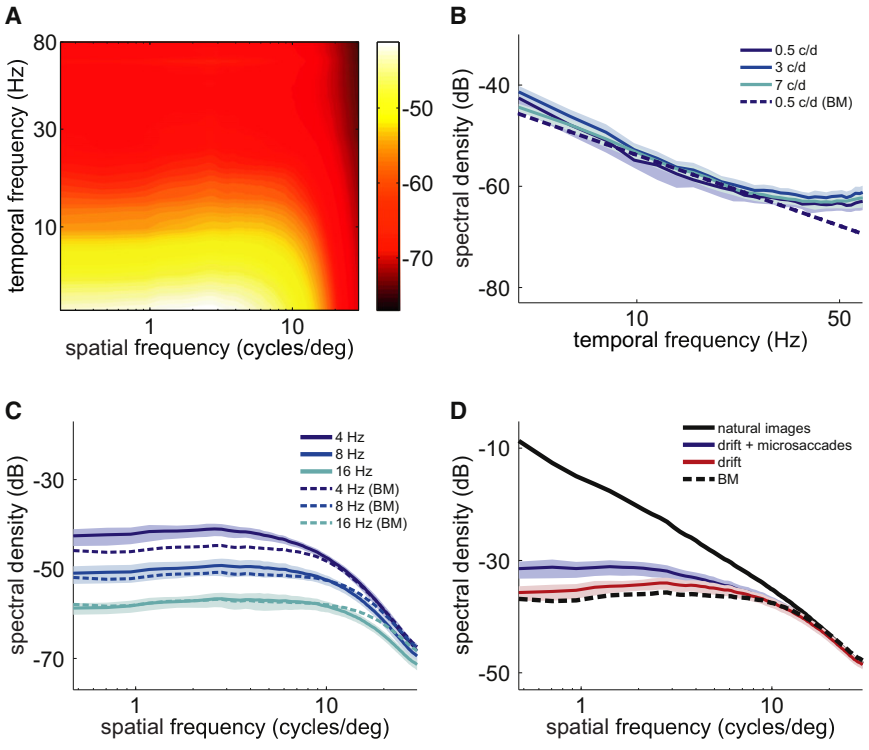

Figure 2. Fixational Modulations of Luminance Equalize Spatial Power

(A) Spatiotemporal frequency content of the retinal stimulus during intersaccadic fixation.

(B and C) Sections at several temporal and spatial frequencies. At nonzero temporal frequencies, the power of the retinal stimulus was constant over a wide range of spatial frequencies. Theoretical predictions from a Brownian motion model of eye movement (BM) are also shown.

(D) Comparison between the power spectrum of the images observed by subjects (black line) and the power available in the form of temporal modulations in the retinal stimulus (the sum over all nonzero temporal frequencies) during drift (red) and during periods of fixation that included both drifts and microsaccades (blue). The dotted line represents the prediction of the Brownian motion model. Shaded areas in (B)–(D) represent one SD.

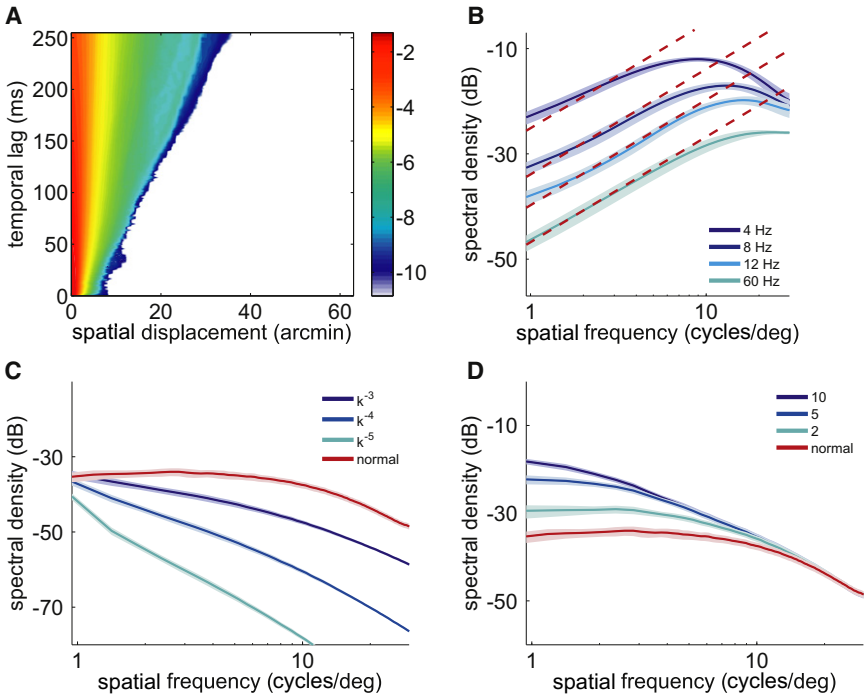

Figure 3. Whitening of the Retinal Stimulus Depends on the Characteristics of Natural Images and Fixational Eye Movements

(A) Probability distribution $q(x,t)$ of eye movement during intersaccadic fixation. The color of a pixel at coordinates (x,t) represents the probability (in log scale) that the eye moved by x in an interval t.

(B) Frequency content of $q(x,t)$. For a wide frequency range, energy increased proportionally to the square of spatial frequency (dashed lines), a dependence that counterbalanced the power spectrum of natural images.

(C and D) Whitening was lost with presentation of images that lacked scale invariance and was greatly attenuated when eye movements were artificially enlarged. In (C), natural images were filtered to yield various spectral densities: the red curve is the unfiltered case and corresponds to the scale-invariant spectrum. In (D), eye movements were scaled by different factors: the red curve, a scale factor of 1, corresponds to normal eye movements. Each curve represents the sum over all nonzero temporal frequencies. Shaded areas represent one SD.

movements (Figure 3D). This happened because normal eye movements only counterbalanced the spectral density of natural scenes up to approximately 10 cycles/deg (Figure 3B). Artificial amplification of eye movements shifted this whitening range to lower spatial frequencies, away from the region in which ganglion cells have their greatest sensitivity. Thus, fixational eye movements match the characteristics of natural scenes to the sensitivity of retinal neurons.

The data in Figures 2 and 3 show that fixational eye movements carry out a specific conversion of spatial patterns into temporal modulations. What are the implications of this reformatting of the retinal input for the neural encoding of visual information?

To gain some intuition, Figure 4A decomposes the stimulus on the retina into a sum of two separate signals: a static image, in which the intensity of each pixel represents the average luminance experienced by a retinal receptor over the fixation period; and a dynamic signal containing the changes in luminance caused by eye movements. These two signals have very different correlation properties. The static signal corresponds to the power in the retinal stimulus at zero temporal frequency. Most of this power is at low spatial frequencies: correspondingly, pairs of pixel have extensive luminance correlations. In contrast, the dynamic signal corresponds to the power that eye movements shifted to nonzero temporal frequencies and equalized over a broad range of spatial frequencies. Because equalization of spectral power is equivalent to elimination of second-order correlations, this signal no longer contains the long-range correlations attributable to the average second-order statistics of natural images. The surviving correlations are specific to the scene being observed and emphasize the luminance discontinuities that occur at edges and object boundaries.

Both parvocellular and magnocellular retinal ganglion cells possess peak sensitivity at nonzero temporal frequencies and will therefore be strongly influenced by the dynamic component of the visual input. Neurons with transient responses—which are not sensitive to zero temporal frequency—will effectively see a spatially whitened input. Their responses will therefore be less influenced by low spatial frequencies and enhance high spatial frequencies significantly more than expected from their contrast sensitivity functions measured in neurophysiological recordings with immobile retinas (Figures 4B and 4C; see Supplemental Experimental Procedures). More sustained neurons will be driven by both the static and dynamic signals. Their tonic responses will depend on the local mean luminance conveyed by the static signal. Superimposed on this, phasic modulations in their responses driven by the dynamic, spatially-whitened component of the visual input will emphasize high spatial frequencies. These synchronous modulations are likely to elicit strong responses in cortical neurons [27, 28]. Interestingly, in physiological recordings, these modulations would be mistaken for correlations arising from neural noise, unless careful experimental measurement of microscopic eye movements was undertaken [10]. They are instead generated by the jittering retinal stimulus itself.

These surprising effects were demonstrated by computer simulations of populations of transient parvocellular neurons. Whereas a static (no eye movements) flash of a natural image led to spatially extensive pools of coactive neurons, neural responses during viewing of the same image with normal fixational eye movements emphasized object contours (see Movie S2). This pattern of activity occurred even though the receptive fields of model neurons were circularly symmetric and did not possess a preference for oriented stimuli. Thus, fixational eye movements facilitated the extraction of important features in the scene.

The observed whitening of the retinal input may provide an explanation for the lack of perceived blurring during fixational instability [29]. With normal eye movements, a predominance of low spatial frequency power only occurs at zero temporal frequency (the static component of the visual input in Figure 4A), a range in which retinal ganglion cells are only

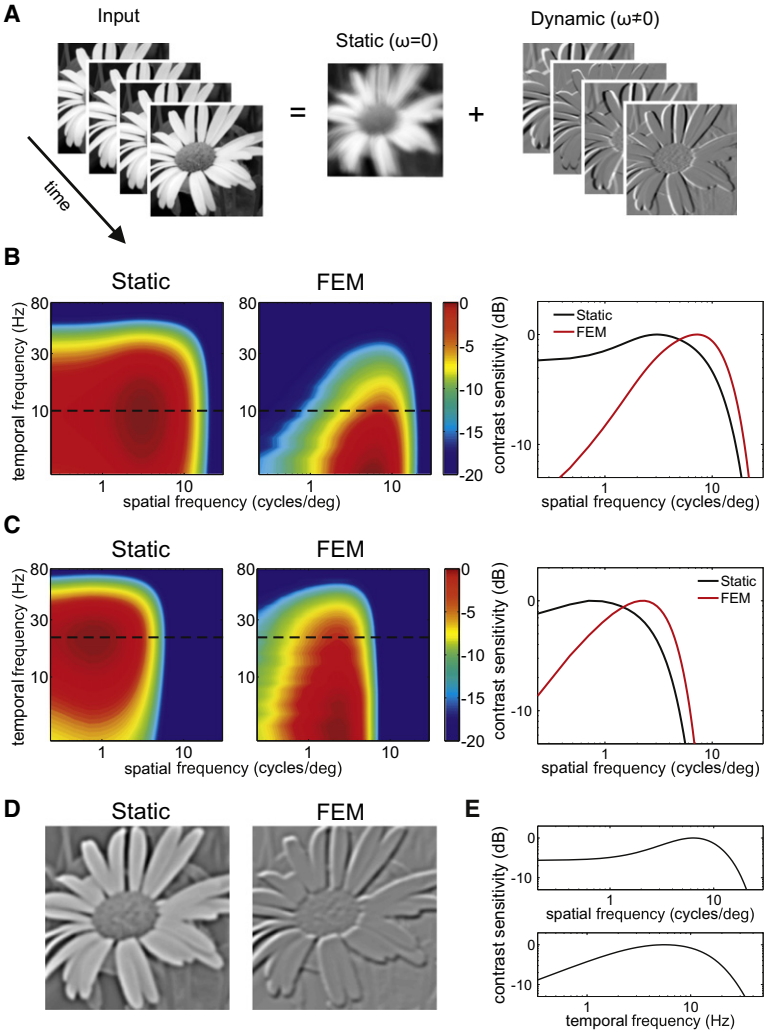

Figure 4. Implications of Spatial Whitening for Neural Encoding

(A) The retinal stimulus during fixation can be decomposed into a blurred image (the power at temporal DC) and the spatially decorrelated modulations caused by eye movements (see Movie S1).

(B and C) Influences of eye movements on the response characteristics of parvocellular (B) and magnocellular (C) retinal ganglion cells. Contrast sensitivity functions measured with immobile retinas [24, 31, 32] (static) are compared to equivalent functions, which incorporate the effect of the redistribution of input power caused by eye movements (FEM). The two panels on the right show spatial sections at the original peak temporal frequencies (dashed lines).

(D) Snapshots of activity in an array of computer-simulated transient neurons during a recorded trace of fixational eye movements (FEM) and when the same scene was displayed without retinal image motion (static). The intensity of each pixel represents the mean instantaneous firing rate of a neuron with receptive field centered at the pixel coordinates. Notice the enhancement of important edges in the scene during fixational instability. See Movie S2 for full simulations of neural activity.

(E) Neurons were modeled as linear filters replicating the spatial and temporal contrast sensitivity of an ON-center parvocellular cell recorded in the macaque's lateral geniculate nucleus [26].

marginally responsive. At nonzero temporal frequencies, fixational eye movements enhance high spatial frequency power, an effect that also supports a contribution of fixational eye movements to the perception of fine spatial detail [12, 15] and explains the perceptual impairments reported in the absence of retinal image motion [16].

Furthermore, these results radically alter standard views about the goals of early neural processing in the visual system. Whereas center-surround interactions [17–19] and nonlinearities [30] in retinal neurons might contribute to spatial decorrelation in ganglion cells with sustained responses, transient neurons are driven by signals that are already spatially whitened. Thus, the band-pass spatial sensitivity of transient neurons actually introduces correlations in their responses, which further emphasize edges in the image.

In sum, our findings show that the incessant fixational motion of the eye reshapes the spatiotemporal stimulus on the retina into a format that facilitates subsequent neuronal processing. This transformation takes advantage of the temporal tuning of visual neurons by (1) spreading the energy of the image into the temporal domain and (2) equalizing the resulting spatial distribution within the range of peak neuronal sensitivity. Much more than just a means for refreshing neural activity, fixational eye movements are a crucial mechanism for encoding a spatial sensory domain in time before any neuronal processing occurs.

Experimental Procedures

Fourteen naïve subjects with normal vision participated in the experiments. All subjects gave written informed consent according to the procedure of the Charles River Campus Institutional Review Board at Boston University. Subjects were instructed to memorize 60 grayscale pictures of natural scenes. Images were displayed sequentially, each for 10 s, on an Iyama HM204DT CRT monitor at 1,024 × 768 pixel resolution and 100 Hz refresh rate. Each pixel subtended 1 arcmin, an angle similar to that covered when the image was originally acquired. Stimuli were observed monocularly, with the left eye patched, in a dimly illuminated room.

Eye movements were recorded by means of a Dual Purkinje Image eyetracker (Fourward Technology) and sampled at 1 Khz. A head rest and a custom dental-imprint bite bar prevented head movements. Recorded traces were segmented into periods of fixation and saccades based on eye velocity. Movements with speeds larger than 3°/s and amplitudes exceeding 3 arcmin were classified as saccades or microsaccades depending on whether their amplitudes were larger or smaller than 30 arcmin. Saccade amplitude was defined as the length of the vector connecting the two locations at which velocity became greater (onset) and lower (offset) than 2°/s.

For each period of fixation, we created a movie of the retinal stimulus by translating the image following the recorded eye trajectory, so that each frame was centered at the current gaze location. Power spectra of these movies were evaluated by means of Welch algorithm with 50% block overlap (window lengths: 512 ms and 256 arcmin). Spectra were estimated separately for each subject and then averaged. 3D power spectra were also rotationally averaged across all directions in each spatial frequency plane (k_x, k_y) to yield 2D maps of spatial and temporal frequency. Data in Figure 2 were obtained by selecting fixations longer than 512 ms. Results were virtually identical (other than a lower temporal resolution) when fixations longer than 128 ms were selected for analysis.

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures and two movies and can be found with this article online at doi:10.1016/j.cub.2012.01.050.

Acknowledgments

This work was supported by National Institutes of Health (NIH) grants EY07977 and EY09314 to J.D.V. and NIH grant EY18363 and National Science Foundation grants BCS-0719849, BCS-1127216, and IOS-0843304 to M.R. The authors thank Antonino Casile and Eric Schwartz for helpful comments.

Received: December 20, 2011
Revised: January 25, 2012
Accepted: January 25, 2012
Published online: February 16, 2012

References

1. Ratliff, F., and Riggs, L.A. (1950). Involuntary motions of the eye during monocular fixation. J. Exp. Psychol. *40*, 687–701.
2. Steinman, R.M., Haddad, G.M., Skavenski, A.A., and Wyman, D. (1973). Miniature eye movement. Science *181*, 810–819.
3. Martinez-Conde, S., Macknik, S.L., and Hubel, D.H. (2004). The role of fixational eye movements in visual perception. Nat. Rev. Neurosci. *5*, 229–240.
4. Hafed, Z.M., Goffart, L., and Krauzlis, R.J. (2009). A neural mechanism for microsaccade generation in the primate superior colliculus. Science *323*, 940–943.
5. Ko, H.K., Poletti, M., and Rucci, M. (2010). Microsaccades precisely relocate gaze in a high visual acuity task. Nat. Neurosci. *13*, 1549–1553.
6. Leopold, D.A., and Logothetis, N.K. (1998). Microsaccades differentially modulate neural activity in the striate and extrastriate visual cortex. Exp. Brain Res. *123*, 341–345.
7. Martinez-Conde, S., Macknik, S.L., and Hubel, D.H. (2000). Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. Nat. Neurosci. *3*, 251–258.
8. Snodderly, D.M., Kagan, I., and Gur, M. (2001). Selective activation of visual cortex neurons by fixational eye movements: implications for neural coding. Vis. Neurosci. *18*, 259–277.
9. Greschner, M., Bongard, M., Rujan, P., and Ammermüller, J. (2002). Retinal ganglion cell synchronization by fixational eye movements improves feature estimation. Nat. Neurosci. *5*, 341–347.
10. Kagan, I., Gur, M., and Snodderly, D.M. (2008). Saccades and drifts differentially modulate neuronal activity in V1: effects of retinal image motion, position, and extraretinal influences. J. Vis. *8*, 19, 1–25.
11. Herrington, T.M., Masse, N.Y., Hachmeh, K.J., Smith, J.E.T., Assad, J.A., and Cook, E.P. (2009). The effect of microsaccades on the correlation between neural activity and behavior in middle temporal, ventral intraparietal, and lateral intraparietal areas. J. Neurosci. *29*, 5793–5805.
12. Marshall, W.H., and Talbot, S.A. (1942). Recent evidence for neural mechanisms in vision leading to a general theory of sensory acuity. In Biological Symposia—Visual Mechanisms, H. Kluver, ed. (Lancaster, PA: Cattell Press), pp. 117–164.
13. Ditchburn, R.W., and Ginsborg, B.L. (1952). Vision with a stabilized retinal image. Nature *170*, 36–37.
14. Riggs, L.A., Ratliff, F., Cornsweet, J.C., and Cornsweet, T.N. (1953). The disappearance of steadily fixated visual test objects. J. Opt. Soc. Am. *43*, 495–501.
15. Ahissar, E., and Arieli, A. (2001). Figuring space by time. Neuron *32*, 185–201.
16. Rucci, M., Iovin, R., Poletti, M., and Santini, F. (2007). Miniature eye movements enhance fine spatial detail. Nature *447*, 851–854.
17. Srinivasan, M.V., Laughlin, S.B., and Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. Proc. R. Soc. Lond. B Biol. Sci. *216*, 427–459.
18. Atick, J., and Redlich, A. (1992). What does the retina know about natural scenes? Neural Comput. *4*, 196–210.
19. van Hateren, J.H. (1992). A theory of maximizing sensory information. Biol. Cybern. *68*, 23–29.
20. Attneave, F. (1954). Some informational aspects of visual perception. Psychol. Rev. *61*, 183–193.
21. Barlow, H.B. (1961). Possible principles underlying the transformations of sensory messages. In Sensory Communication, W.A. Rosenblith, ed. (Cambridge, MA: MIT Press), pp. 217–234.
22. Collewijn, H., and Kowler, E. (2008). The significance of microsaccades for vision and oculomotor control. J. Vis. *8*, 20, 1–21.
23. Field, D.J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A *4*, 2379–2394.
24. Croner, L.J., and Kaplan, E. (1995). Receptive fields of P and M ganglion cells across the primate retina. Vision Res. *35*, 7–24.
25. Kaplan, E., and Benardete, E. (2001). The dynamics of primate retinal ganglion cells. Prog. Brain Res. *134*, 17–34.
26. Derrington, A.M., and Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. J. Physiol. *357*, 219–240.
27. Dan, Y., Alonso, J.M., Usrey, W.M., and Reid, R.C. (1998). Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. Nat. Neurosci. *1*, 501–507.
28. Bruno, R.M., and Sakmann, B. (2006). Cortex is driven by weak but synchronously active thalamocortical synapses. Science *312*, 1622–1627.
29. Packer, O., and Williams, D.R. (1992). Blurring by fixational eye movements. Vision Res. *32*, 1931–1939.
30. Pitkow, X., and Meister, M. (2011). Decorrelation and efficient coding by retinal ganglion cells. Nat. Neurosci., in press.
31. Benardete, E.A., and Kaplan, E. (1997). The receptive field of the primate P retinal ganglion cell, I: Linear dynamics. Vis. Neurosci. *14*, 169–185.
32. Benardete, E.A., and Kaplan, E. (1999). The dynamics of primate M retinal ganglion cells. Vis. Neurosci. *16*, 355–368.