Does the brain de-jitter retinal images?

Bruno A. Olshausen[a,1] and Charles H. Anderson[b]

[a]Helen Wills Neuroscience Institute and School of Optometry, University of California, Berkeley, CA 94720; and [b]Department of Anatomy and Neurobiology, Washington University, St. Louis, MO 63110

As anyone who has taken a photograph with shaky hands can attest, camera movement creates image blur. Because photosensors must integrate over a finite interval to acquire a noise-free image, any motion that occurs during that interval will blur over the spatial details of an object by the amount of space traversed (Fig. 1). This simple fact creates headaches for engineers designing scientific imaging systems and commercial digital cameras. It is also a problem that the nervous system of any animal with moving eyes—from jumping spiders to humans—must deal with in processing images, but little is currently known about how they do it. In PNAS, Burak et al. (1) propose a potential neural mechanism for solving this problem in the human visual system.

During normal vision the eyes are constantly in motion. Typically the eyes shift their gaze from one point to another within a scene, or they may "lock on" a moving object (or during head or body movement) in a smooth tracking motion to stabilize its image on the retina. However, even when one's gaze is fixed upon a stationary object with the head still, the eyes continue to move involuntarily due to small drifts and corrective saccades. In recent years, these small fixational eye movements have been increasingly studied by vision scientists (2–4). What is especially striking about these movements is that, despite the fact that they can sweep the image of an object over many photoreceptors at speeds well within the range of perceptible motion, a fixated object nevertheless appears stable, and fine details at the same spatial scale as the photoreceptor spacing are preserved. A possible neural mechanism for canceling motion signals due to eye movement has been discovered in the retina (5), which may account for why objects appear stable. However, the fact that detail remains sharp is still baffling, not only because of the integration time of photoreceptors but also because downstream computational processes must integrate information over time as well to make perceptual decisions. A naive integration process uninformed about changes in eye position would lose information about fine details through averaging. This begs the question: How do we resolve and make discriminations about detailed structure despite the temporal integration processes that exist at various stages of neural processing?
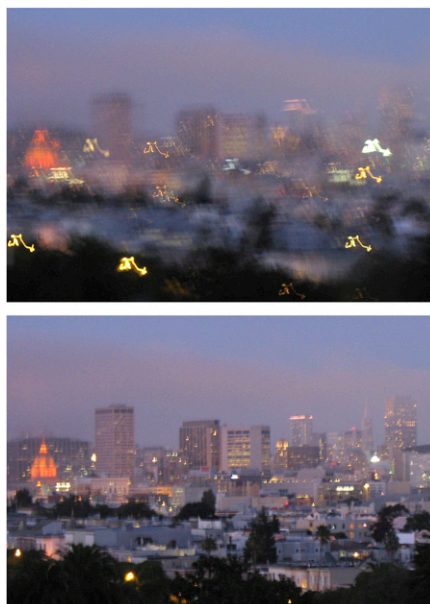

Fig. 1. (*Upper*) Camera shake during exposure creates image blur. (*Lower*) The same scene captured with the same exposure duration and a similar degree of camera shake but with an image stabilization mechanism (Canon PowerShot SD700 IS) that compensates for camera movement. How the visual system compensates for retinal image motion so as to prevent detail from being similarly corrupted by temporal integration processes in the brain is currently unknown.

Answering this question will ultimately require detailed probing of neural mechanisms at each stage of processing during eye movements. In the meantime, computational models that demonstrate neural mechanisms that could solve the problem in principle may provide useful insights and guidance for what to look for in experiments. The model recently proposed by Burak et al. (1) proposes that neural circuitry in the visual cortex internally compensates for small eye movements so as to form a stable image representation. Their model shares similarities with previous proposals based on shifter circuits (6), dynamic routing (7), and map-seeking circuits (8). The basic premise behind all of these models is that there exist two distinct classes of neurons—those that carry visual information, and those that control the flow of visual information by dynamically gating connections (and hence signal flow) from one neuron to another. By dynamically routing information flow from one array of neurons to the next so as to compensate for image motion, vari-

ability in the input array due to changes in position can be removed to form an invariant representation of objects on the output array. Computational processes required for discrimination can then accumulate information in the stabilized representation.

Burak et al. (1) take this basic idea a step further by incorporating relevant neurobiological details such as the spatial extent and dynamics of fixational eye movement, neural integration times, and most importantly the spiking properties of visual neurons. By casting the problem in a Bayesian framework, and by considering the information conveyed by each spike coming from an array of retinal ganglion cells, they derive an optimal computation for cortical neurons to perform to simultaneously recover a stabilized image and an estimate of retinal position. They evaluate the model's performance on a number of pattern discrimination tasks, demonstrating that it is capable of stabilizing retinal images in a manner that is consistent with realistic assumptions of spike rates and eye movement dynamics. They also propose a number of experimental tests of the model—for example, the ability to stabilize images should depend on the degree of spatial context provided.

Whether or not cortical neurons provide an explicitly stabilized image representation is subject to debate. In fact, one may wonder whether it is even necessary in the first place. It has been argued, for example, that what really matters are sensorimotor contingencies (9), so as long as the system has a representation of eye movement then all of the information needed to make proper inferences about the world is available. Insisting that the cortex create a stable image representation, according to this argument, may be a bit like insisting that the inverted image projected on the retina be reinverted in the cortex so that objects appear "upright." It makes a nice picture, but who is watching? However, the central issue for us is not about making a nice picture, but rather about properly integrating image data to support

www.manaraa.co

inferential computations. Such computations require that image data be routed and combined with past image data to be in register. There may be other ways of doing this that do not involve explicitly stable representations, but visual information must still be somehow properly routed in response to changes in eye position.

In terms of neurophysiological evidence, the studies of Motter and Poggio (10) show that the receptive fields of V1 neurons shift in the opposite direction of eye movement so as to remain stable in world coordinates, thus supporting the idea of image stabilization. However, contrary to these findings, Gur and Snodderly (11) report that V1 receptive fields are locked to retinotopic coordinates. It is difficult to account for these discrepancies because errors in eye position measurement or other artifacts would not tend to err systematically in favor of one scheme or the other. Thus, additional experiments will likely be needed to resolve this issue. It is important to bear in mind though that Motter and Poggio's study (10) was originally motivated by previous findings from Poggio's lab showing that binocular neurons in the fovea can relay differences in disparity as small as 3 min of arc despite the fact that the eyes drift independently over considerably large distances. As they observe, "The spatial domain of a stereoscopic neuron appears to be dynamically maintained over a spatial extent of the external world which is considerably narrower than the range of positions taken by the eye during fixation on a small target" (10). Indeed, it is difficult to imagine how a purely retino-

topic representation could support the dynamic computations necessary for stereopsis in the face of differential binocular eye movements, and so it may prove fruitful to direct future experiments toward this issue.

A question that naturally arises when considering fixational eye movements is

Burak et al. have proposed a model that is mathematically and computationally sound and neurobiologically plausible.

whether they are a bug or a feature. It is often pointed out, for example, that without motion the retinal image will fade, and thus eye movements are necessary to prevent fading. But we find this reasoning a bit circular: Presumably, retinal images fade because there is no use for neurons to support the encoding of constant signals when the eye is constantly moving. We speculate that there are good reasons for introducing image motion based on first principles. For example, many super-resolution schemes rely upon accumulating information from a moving image array to achieve higher-resolution images. Scanning images over the retinal cone array may also be advantageous for averaging over the idiosyncrasies of particular cones, such as differences in gain or

wavelength selectivity. However, such schemes, although improving image quality, increase the burden on postreceptor processing because signals must be properly rerouted downstream so as to properly accumulate information.

The model of Burak et al. (1) brings us a step toward demystifying what was otherwise a puzzling conundrum about how cortical neurons could properly integrate the spiking activity coming from a moving retinal image array. Here, we have a plausible model for how it *could* be done, at least for simple binary images. An important extension will be to show that the model can also work for continuous, natural images that contain high spatial correlations and thus introduce extra challenges for cross-correlation-based stabilization schemes such as that proposed by the authors. It would also be useful to explore more realistic models of eye drift beyond the random walk model considered here. How such a detailed circuit could be wired up developmentally, or at least partially self-organized from visual experience, will also need to be explained. Nevertheless, this is a good start. Burak et al. (1) have proposed a model that is mathematically and computationally sound and neurobiologically plausible. Now it needs to be tested. Recent advances in adaptive optics combined with real-time tracking for photoreceptor-precision stimulus delivery (12) and neurophysiological recording (13) provide a promising avenue for determining how or whether the brain de-jitters retinal images.

1. Burak Y, Rokni U, Meister M, Sompolinsky H (2010) Bayesian model of dynamic image stabilization in the visual system. *Proc Natl Acad Sci USA* 107:19525–19530.
2. Martinez-Conde S, et al. (2008) Eye movements and the perception of a clear and stable visual world. *J Vis* 8:1.
3. Rucci M, Iovin R, Poletti M, Santini F (2007) Miniature eye movements enhance fine spatial detail. *Nature* 447:851–854.
4. Poletti M, Listorti C, Rucci M (2010) Stability of the visual world during eye drift. *J Neurosci* 30:11143–11150.
5. Olveczky BP, Baccus SA, Meister M (2003) Segregation of object and background motion in the retina. *Nature* 423:401–408.

6. Anderson CH, Van Essen DC (1987) Shifter circuits: A computational strategy for dynamic aspects of visual processing. *Proc Natl Acad Sci USA* 84:6297–6301.
7. Olshausen BA, Anderson CH, Van Essen DC (1993) A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J Neurosci* 13:4700–4719.
8. Arathorn DW (2002) *Map-Seeking Circuits in Visual Cognition: A Computational Mechanism for Biological and Machine Vision* (Stanford Univ Press, Stanford, CA).
9. O'Regan JK, Noë A (2001) A sensorimotor account of vision and visual consciousness. *Behav Brain Sci* 24: 939–973, discussion 973–1031.

10. Motter BC, Poggio GF (1990) Dynamic stabilization of receptive fields of cortical neurons (VI) during fixation of gaze in the macaque. *Exp Brain Res* 83: 37–43.
11. Gur M, Snodderly DM (1997) Visual receptive fields of neurons in primary visual cortex (V1) move in space with the eye movements of fixation. *Vision Res* 37: 257–265.
12. Arathorn DW, et al. (2007) Retinally stabilized cone-targeted stimulus delivery. *Opt Express* 15: 13731–13744.
13. Sinich LC, Zhang Y, Tiruveedhula P, Horton JC, Roorda A (2009) Resolving single cone inputs to visual receptive fields. *Nat Neurosci* 12:967–969.