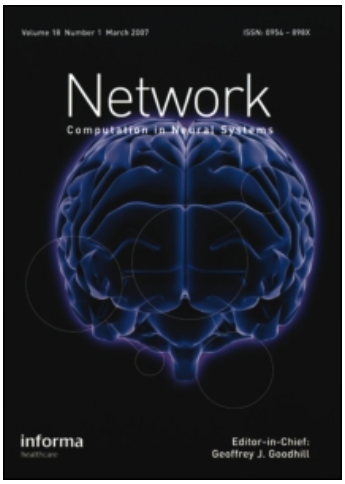
This article was downloaded by: *[Boston University]*
On: *21 November 2008*
Access details: *Access Details: [subscription number 790384933]*
Publisher *Informa Healthcare*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Network: Computation in Neural Systems
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713663148

Fixational eye movements, natural image statistics, and fine spatial vision

Michele Rucci [a]
[a] Departments of Psychology and Biomedical Engineering and Program in Neuroscience, Boston University, Boston, MA 02215, USA

Online Publication Date: 01 January 2008

To cite this Article Rucci, Michele(2008)'Fixational eye movements, natural image statistics, and fine spatial vision',Network: Computation in Neural Systems,19:4,253 — 285
To link to this Article: DOI: 10.1080/09548980802520992
URL: http://dx.doi.org/10.1080/09548980802520992

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Network: Computation in Neural Systems
December 2008; 19(4): 253–285

informa
healthcare

VIEW POINT

Fixational eye movements, natural image statistics, and fine spatial vision

MICHELE RUCCI

Departments of Psychology and Biomedical Engineering and Program in Neuroscience, Boston University, Boston, MA 02215, USA

(Received 21 August 2008; revised 26 September 2008; accepted 1 October 2008)

Downloaded By: [Boston University] At: 23:24 21 November 2008

Abstract
Perception and motor control are often regarded as two separate branches of neuroscience. Like most species, however, humans are not passively exposed to the incoming flow of sensory data, but actively seek useful information. By shaping input signals in ways that simplify perceptual tasks, behavior might play an important role in establishing efficient sensory representations in the brain. Under natural viewing conditions, the main source of motion of the stimulus on the retina is not the scene but our own behavior. The retinal image is never still, even during visual fixation, when small eye movements combine with movements of the head and body to continually perturb the location of gaze. This article examines the impact of the fixational motion of the retinal image on the statistics of visual input and the neural encoding of visual information. Building upon recent theoretical and experimental results, it is argued that an unstable fixation constitutes an efficient strategy for acquiring information from natural scenes. According to this theory, the fluctuations of luminance caused by the incessant motion of the eye equalize the power present at different spatial frequencies in the spatiotemporal stimulus on the retina. This phenomenon yields compact neural representations, emphasizes fine spatial detail, and might enable a temporal multiplexing of visual information from the retina to the cortex. This theory posits motor contributions to early visual representations and suggests that perception and behavior are more intimately tied than commonly thought.

Keywords: *Eye movements, microsaccade, drift, retina, lateral geniculate nucleus, pairwise correlation, computational model*

Introduction

Visual perception studies often attempt to establish a relationship between selected characteristics of the stimulus (e.g. contrast, frequency, orientation) and the responses these features evoke in the observer (e.g. performance, neural activity).

Correspondence: M. Rucci, Department of Phychology and Biomedical Engineering and Program in Neuroscience, Boston University, Boston, MA 02215, USA. Tel: 617–353–7671. Fax: 617–358–1736. E-mail: mrucci@bu.edu

ISSN 0954-898X print/ISSN 1361-6536 online/08/040253–285 © 2008 Informa UK Ltd.
DOI: 10.1080/09548980802520992

A presumption of this approach is that the stimulus presented by the experimenter is the actual input signal to the visual system. Yet, under natural viewing conditions, the signals entering the eyes depend not only on the external scene, but also on the observer's behavior during the acquisition of visual information. The spatiotemporal stimulus on the retina differs in important ways from the stimulus displayed on the monitor.

Eye movements are an important source of input modulations, as the eyes are never at rest. Even during visual fixation, small eye movements keep the projection of the visual scene on the retina continually in motion (Figure 1). Under conditions of sustained fixation, humans exhibit several types of fixational eye movements including occasional saccades with amplitudes less than half a degree–known as fixational saccades or microsaccades –, slow drifts, and physiological nystagmus, a high-frequency tremor with amplitude smaller than 1 (Ratliff and Riggs 1950; Ditchburn 1955; Steinman et al. 1973).[1] Under more natural viewing conditions, other small eye movements, including corrective saccades and post-saccadic drifts, together with movements of different parts of the body, significantly increase the instability of visual fixation. In this article, we will use the term "fixational instability" to denote the global motion of the retinal image that results from the combination of multiple eye and body movements during natural visual fixation. Fixational instability has a profound influence on the spatiotemporal stimulus on the retina.

The possible visual functions of the fixational motion of the retinal image have been studied for almost a century (Martinez-Conde et al. 2004). Initial interest was triggered by dynamic theories of visual acuity (Averill and Weymouth 1925;

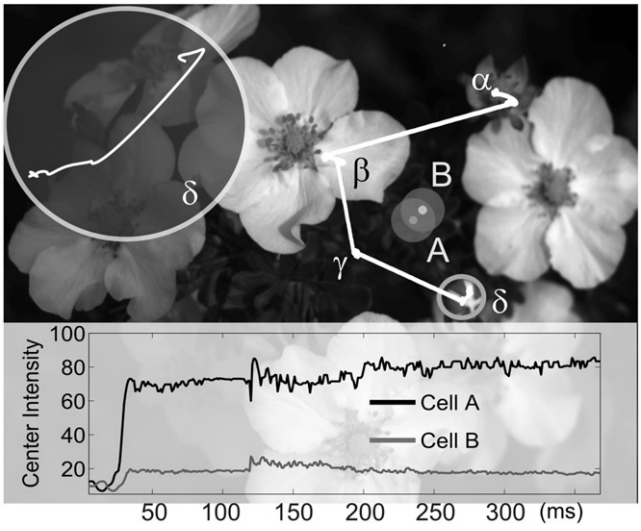


Figure 1. An example of fixational instability. A trace of eye movements recorded by a high-resolution eye-tracker (yellow trajectory) is shown superimposed on the original image. The trace is composed of four fixation periods, α, β, γ, and δ, separated by saccades. The insert on the top left shows a zoomed portion of the trace (red circle) in which small fixational eye movements occur. The bottom graph shows the spatiotemporal input to two hypothetical neurons, A and B, as the gaze location moves from γ to δ. The receptive fields of these two neurons are shown at their respective locations at time $t = 0$.

Downloaded By: [Boston University] At: 23:24 21 November 2008

Marshall and Talbot 1942), which argued for contributions of fixational eye movements in the process underlying the exquisite acuity of the human visual system. Interest was then vigorously renewed over 40 years ago by the striking discovery that stabilized images, i.e., images that move with the eyes, tend to lose contrast with time and possibly fade altogether over a period of several seconds or minutes (Ditchburn and Ginsborg 1952; Riggs and Ratliff 1952; Pritchard 1961; Yarbus 1967). Prevention of image fading has been commonly held as an explanation for the existence of fixational instability. According to this view, the physiological motion of the retinal image is necessary in order to refresh neuronal responses and prevent the disappearance of a stationary scene. Much effort in this area has been dedicated to investigating the link between microsaccades and image fading (Ditchburn et al. 1959; Martinez-Conde et al. 2006), a relationship that has been the subject of fierce controversy (Ditchburn 1980; Kowler and Steinman 1980).

Several other functions of fixational eye movements have also been hypothesized besides the prevention of image fading (Hering 1899; Averill and Weymouth 1925; Marshall and Talbot 1942; Arend 1973; Rucci et al. 2000; Ahissar and Arieli 2001; Greschner et al. 2002; Olveczky et al. 2003; Rucci and Casile 2005; Pitkow et al. 2007). While these latter theories differ in the specific functions and mechanisms they propose, they share the common hypothesis that fixational instability plays a central role in the *structuring*, rather than the simple *refreshing*, of neural activity. That is, according to these proposals, fixational eye movements are necessary to modulate input signals in ways that facilitate the neural encoding of visual information.

A significant influence of fixational instability on neural representations is also supported by recent neurophysiological investigations in the monkey. Excitatory and inhibitory influences have been reported in different cortical areas following fixational saccades (Gur and Snodderly 1987; Bair & O'Keefe 1998; Leopold and Logothetis 1998; Martinez-Conde et al. 2000; Snodderly et al. 2001). These modulations appear to depend on the direction of the saccade and on the stimulus present in the receptive field. Furthermore, distinct populations of V1 neurons have been identified on the basis of their response to the two main components of fixational eye movements, saccades and drifts. Some cells tend to be selectively activated during fixational saccades while others respond during the periods of drifts (Snodderly et al. 2001). These neurophysiological studies have shown that fixational eye movements affect neuronal activity in complex ways. These neural modulations are the result of the interactions between the spatiotemporal stimulus on the retina and the response characteristics of individual neurons.

While these recent neurophysiological findings suggest important oculomotor influences on visual input signals, little effort has actually been dedicated to studying the spatiotemporal stimulus on the retina during natural fixation. This limited characterization of the statistics of the retinal image is surprising given that this stimulus provides the real input to the visual system. In fact, analysis of the statistics of retinal input offers a powerful approach toward elucidating the possible visual functions of fixational instability, as it enables the establishment of a direct link between the statistical properties of the visual world and the response selectivity of neurons in the early stages of the visual system. This approach can give insights into the question of whether evolutionary pressures

Downloaded By: [Boston University] At: 23:24 21 November 2008

exist for acquiring information by means of jittering eyes instead of evolving more stable fixation procedures.

To start filling this gap in the literature, this article describes a model of the retinal input during natural fixation and examines the way this input signal might interact with the response characteristics of ganglion cells in the primate retina. The material presented in the following sections builds upon a series of recent theoretical and experimental studies, which have examined the impact of fixational eye movements on visual performance (Rucci and Desbordes 2003; Rucci and Beck 2005; Rucci et al. 2007), neural encoding (Rucci and Casile 2005; Desbordes and Rucci 2007; Poletti and Rucci 2008), and on the long-term developmental consequences of chronic exposure to a constantly moving retinal image (Rucci et al. 2000; Rucci and Casile 2004; Casile and Rucci 2006). These studies have suggested that an unstable fixation is an effective strategy for analyzing natural scenes.

Downloaded By: [Boston University] At: 23:24 21 November 2008

Retinal input during visual fixation

This section analyzes the influence of fixational instability on the retinal stimulus. For simplicity, we focus on the case of a fixation of infinite duration on a static, one-dimensional visual scene, $L(x)$, as illustrated in Figure 2. The model has, however, general validity and can be directly extended to the case of a two-dimensional stimulus. The extension to the case of fixation with finite duration is discussed in section "A visuomotor neural code?".

An immobile retinal image

In order to establish a reference point with respect to which evaluating the influence of eye movements, we first consider the visual input occurring in the absence of retinal image motion, that is, the signals entering the eye under the assumption that the direction of gaze is held perfectly steady. This condition cannot be achieved in real life and can only be approximated in the laboratory by experiments of retinal stabilization.

As schematically represented in Figure 2(a), the input to the retina does not change during a perfectly still visual fixation, and the luminance experienced by retinal receptors remains constant in time. In this condition, the visual input signal I_{Still} can be analytically expressed as a function of two variables, space x and time t, in the following way:

$$I_{\text{Still}}(x, t) = L(x)v(t) \tag{1}$$

where $L(x)$ is the spatial luminance profile of the stimulus, and $v(t) = 1$ at all times. The two-dimensional function described in Equation 1 is space–time separable. That is, it can be represented as the product of two separate functions of only one variable, either time or space.

To fully evaluate the impact of the fixational motion of the retinal image, it is useful to also examine visual input signals in the frequency domain in addition to the spatiotemporal domain. Calculation of the power spectrum of $I_{\text{Still}}(x, t)$, P_{Still}, is straightforward. As in the spatiotemporal domain, in the frequency domain $I_{\text{Still}}(x, t)$

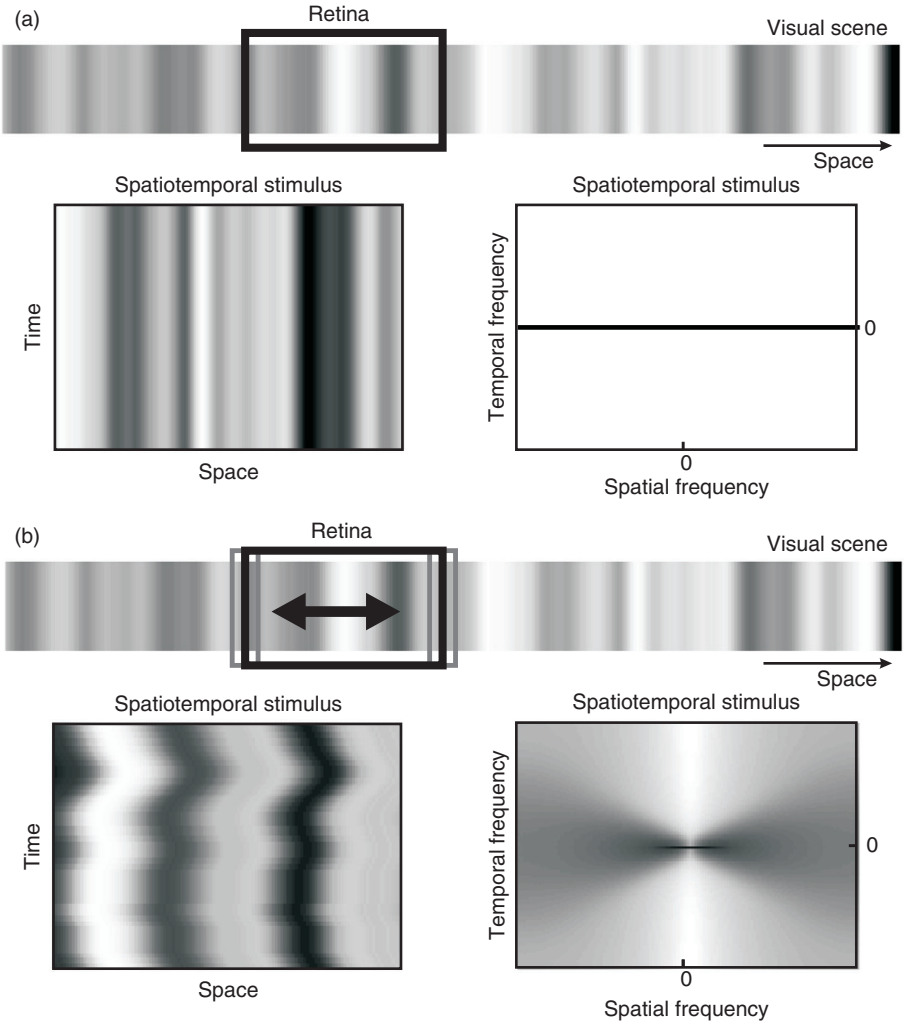


Figure 2. Spatiotemporal input to the retina during two different visual fixations: (a) a perfectly steady fixation in which the retinal image is immobile; and (b) an unstable fixation in which the retina moved as a random walk. For simplicity, a visual scene with a single spatial dimension is considered. In both (a) and (*b*), the top panel illustrates the one-dimensional scene and the retina. The two bottom panels show the visual input signals in both the space–time domain (*left*) and the frequency domain (*right*).

will also be represented by a signal that is a function of two variables, the spatial frequency u and the temporal frequency ω. Furthermore, since $I_{\text{Still}}(x, t)$ is separable into spatial and temporal components, its power spectrum will also be decomposable into separate functions of spatial and temporal frequency. From Equation 1, we obtain

$$P_{\text{Still}}(u, \omega) = P_L(u)\delta(\omega) \tag{2}$$

where $P_L(u)$ is the spatial power spectrum of $L(x)$ and $\delta(\omega)$ is the unit impulse function. Note that since I_{Still} does not change with time, all power of P_{Still}

Downloaded By: [Boston University] At: 23:24 21 November 2008

is concentrated on the spatial frequency axis at zero temporal frequency ($\omega = 0$). As illustrated in Figure 2(a), there is no power for temporal frequencies different from zero.

Normal fixation

Unlike the previous unrealistic condition of an immobile retinal stimulus, in real life, retinal receptors are continually exposed to the fluctuations of luminance resulting from the impossibility of holding the direction of gaze perfectly steady. As a consequence of these modulations, the retinal image no longer remains equal to $L(x)$ for the entire duration of fixation, as in the previous example, but shifts following the path $\xi(t)$ of the direction of gaze. Thus, the retinal stimulus can be written as:

$$I(x, t) = L[x + \xi(t)]. \tag{3}$$

In this case, $I(x, t)$ is not space-time separable as in Equation 1, but will instead change with time in a complex manner. The spatiotemporal characteristics of this signal will depend on the specific combination of the spatial statistics of $L(x)$ and the temporal statistics of $\xi(t)$.

Figure 2(b) shows an example of fixational modulations of luminance in a simulation in which the trajectory $\xi(t)$ was modeled as a discrete random walk with time step of 1 ms. As illustrated in the left panel of Figure 2(b), because of the motion of the retinal image, the visual input at each position x on the retina appears to fluctuate around its mean. The panel to the right of Figure 2(b) shows the power spectrum of $I(x, t)$, $P_I(u, \omega)$, estimated by means of the Welch method. This figure illustrates two important consequences of the fixational motion of the retinal image. First, given that the retinal stimulus changes with time, its spectral density is no longer concentrated on the temporal DC axis as in the previous example. That is, unlike Figure 2(a), power at nonzero temporal frequencies becomes available at all the spatial frequencies contained in the stimulus. Second, the bandwidth of the temporal power resulting from retinal image motion increases with spatial frequency. This effect is a consequence of the fact that the average change of luminance caused by a small translation of the retinal image increases with the spatial frequency of the stimulus. To clarify this point, consider the luminance change, DI, experienced by a retinal receptor during viewing of a grating $L(x) = A\sin(ux)$. For a small movement ξ, DI is approximately equal to the first derivative of the stimulus multiplied by the amplitude of the translation: $\xi \frac{\mathrm{d}}{\mathrm{d}x} L(x) = \xi A u \cos(ux)$. DI is proportional to the spatial frequency. Therefore, during fixational instability, the visual input at each location of the retina will change more rapidly (i.e., will possess higher temporal frequencies) for high spatial frequency stimuli than for low spatial frequency stimuli.

It is important to observe that the fixational motion of the retinal stimulus does not generate new power, but merely redistributes the power of the stimulus. In fact, since the external stimulus is the same, whether or not it is observed through jittering eyes, the total powers of the visual input signals in the presence and in the absence of fixational instability, $I(x, t)$ and $I_{\mathrm{Still}}(x, t)$, must be identical. More specifically, I and I_{Still} possess the same total amount of power at each spatial frequency. However, the temporal distribution of power is different between these

Downloaded By: [Boston University] At: 23:24 21 November 2008

two signals and varies with the spatial frequency. As explained below, this differential impact of fixational instability at different spatial frequencies might have important implications regarding the encoding of visual information in the brain.

An estimate of the input spectrum during fixational instability

To better understand the impact of fixational eye movements on the visual input, it is instructive to analytically estimate the power spectrum of the spatiotemporal stimulus on the retina. This task is not simple for an arbitrary gaze trajectory $\xi(t)$. However, an approximation of the power spectrum can be derived under a few simplifying assumptions.

We start by obtaining an approximation of the visual input itself, $I(x,t)$. Equation 3 makes clear that, given that the retinal image translates during fixational instability, the input luminance impinging on retinal location $x+\xi(t)$ at time t will be identical to the input at location x at time $t=0$. Under the plausible assumption of motion with small amplitude, the two locations $x+\xi(t)$ and x will be in proximity to each other. We can, thus, use a Taylor approximation to estimate the luminance at one location on the basis of the luminance level at the other location:

$$I(x,t) = L[x+\xi(t)] \approx L(x) + \frac{\mathrm{d}L(x)}{\mathrm{d}x}\xi(t) + \frac{1}{2}\frac{\mathrm{d}^2L(x)}{\mathrm{d}x^2}\xi^2(t) \tag{4}$$

where the terms with a higher order than the second were assumed to be negligible because of the small amplitude of the fixational motion of the retinal image.

In the Fourier domain, this equation becomes

$$I_{W,T}(u,\omega) \approx L_W(u)\delta_T(\omega) + juL_W(u)\xi_T(\omega) - \frac{1}{2}u^2L_W(u)\Gamma_T(\omega) \tag{5}$$

where the Fourier Transforms were evaluated over finite spatial and temporal windows of observation with length W and duration T, respectively: $I_{W,T}(u,\omega) = \mathcal{F}_{W,T}\{I(x,t)\}$, $\xi_T(\omega) = \mathcal{F}_T\{\xi(t)\}$, $\Gamma_T(\omega) = \mathcal{F}_T\{\xi^2(t)\}$, and $\delta_T(\omega) = (2\sin(\omega T/2)/\omega)$.

Equation 5 enables an estimation of the power spectrum of the input signal $I(x,t)$. Since

$$P_I(u,\omega) = \lim_{W\to\infty}\lim_{T\to\infty}\frac{1}{W}\frac{1}{T}\langle|I_{W,T}(u,\omega)|^2\rangle,$$

we obtain

$$P_I(u,\omega) \approx P_L(u)\delta(\omega)\{1-u^2\sigma_\xi^2\} + u^2P_L(u)P_\xi(\omega) \tag{6}$$

where σ_ξ^2 represents the variance of the motion of the eye, $\delta(\omega)$ is the unit impulse function, and we have assumed the trajectory of the retinal image to possess zero mean.

Equation 6 shows that the power spectrum of the visual input during fixational instability can be approximated by the sum of two terms:

$$P_I(u,\omega) \approx P_I^S(u,\omega) + P_I^D(u,\omega) \tag{7}$$

Downloaded By: [Boston University] At: 23:24 21 November 2008

with

$$\begin{cases} P_I^S(u,\omega) = P_L(u)\delta(\omega)\{1 - u^2\sigma_\xi^2\} \\ P_I^D(u,\omega) = u^2 P_L(u) P_\xi(\omega). \end{cases}$$

This equation explains the way the spectral density of the stimulus is reorganized by the fixational motion of the retinal image. The power of the first term, P_I^S, is entirely concentrated on the temporal DC axis ($P_I^S(\omega) = 0$ for $\omega \neq 0$). The second term, P_I^D, is instead nonzero at $\omega \neq 0$ and contains the power of I_{Still}, which, as a consequence of the motion of the eye, spreads to the temporal domain. Unlike P_I, P_I^D is space-time separable. As shown by Equation 7, on the spatial frequency axis, the spectral density of the stimulus $P_L(u)$ is multiplied by the square of spatial frequency. At each temporal frequency ω, the term $P_L(u)u^2$ is multiplied by a gain factor $P_\xi(\omega)$ that depends on the characteristics of fixational instability.

The approximation of P_I given by Equation 7 obeys the conservation of power. At each spatial frequency u, the total power available across all temporal frequencies is identical to the amount of power present in the case in which the stimulus is viewed with no fixational motion:

$$\int_{-\infty}^{\infty} P_I(u,\omega)\mathrm{d}\omega = \int_{-\infty}^{\infty} P_{\text{Still}}(u,\omega)\mathrm{d}\omega = P_L(u) \tag{8}$$

An estimate of the total *dynamic* power (the power at temporal frequency different from zero), can be calculated from Equation 8 for every spatial frequency. Since the total input power does not change with or without retinal image motion, $P_I^D(u)$ can be approximated by subtracting $P_I(u,0)$ from the power of stimulus:

$$\widehat{P}_I^D(u) = \int_{-\infty}^{\infty} P_I^D(u,\omega)\mathrm{d}\omega = P_L(u) - P_I(u,0) \approx u^2 P_L(u)\sigma_\xi^2. \tag{9}$$

$P_I^D(u)$ is the amount of power which becomes available in the form of temporal modulations of luminance because of fixational instability.

Equation 9 shows that the amount of power that spreads into the temporal domain tends to increase with spatial frequency as u^2. This effect is schematically illustrated in Figure 3. Fixational instability can be regarded as an operator, which transforms spatial power into temporal power. Given a static stimulus with spectral density $P_L(u)$, the temporal power available at each temporal frequency will tend to be proportional to $P_L(u)u^2$, and the total temporal power (integrated over all temporal frequencies) will be approximately equal to $u^2 P_L(u)\sigma_\xi^2$. As shown in Figure 3(a), multiplication by u^2 attenuates low spatial frequencies and enhances high spatial frequencies, an effect that can be observed in the simulation of Figure 2(b).

Figure 3(b) shows the impact of fixational instability during viewing of natural images. Unlike other categories of images, images of natural scenes are characterized by a very specific spectral distribution; their power spectrum declines proportionally to the square of the spatial frequency (Field 1987). The spatiotemporal transformation resulting from fixational instability implies that the total amount of temporal power available in the form of fixational modulations of luminance is constant during viewing of natural images. That is, fixational instability equalizes the amount of power at all spatial frequencies.

Downloaded By: [Boston University] At: 23:24 21 November 2008

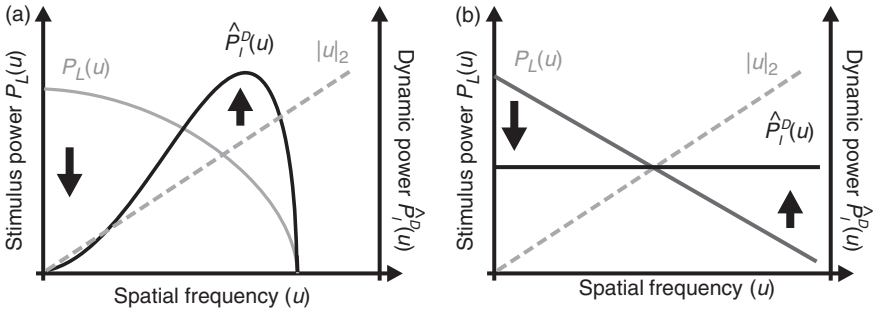


Figure 3. Redistribution of spectral density caused by fixational instability. The fixational motion of the retinal image transforms the spatial power of the stimulus into temporal power. (a) Given a static stimulus with power spectrum $P_L(u)$, the total temporal power resulting from fixational instability at spatial frequency u is approximately equal to $P_L(u)|u|^2$. (b) During viewing of natural images, the temporal power resulting from fixational instability is constant across spatial frequency.

Downloaded By: [Boston University] At: 23:24 21 November 2008

The estimates of spectral densities obtained in this section apply to any type of eye movements yielding microscopic motion of the retinal image. These estimates should, however, be used with caution. Since in the Taylor series of Equation 4 we have considered only terms up to the second order, this approximation holds only for small amplitudes of motion. In the frequency domain, our estimates are valid only at low spatial frequencies. However, errors at high spatial frequencies may be tolerated when these estimates are used to study neural activity, as the contrast sensitivity functions of neurons in the early visual system tend to be low-pass. The analysis of section "Influence of fixational instability on neural activity" shows that these spectral estimates provide an intuitive understanding of the results obtained in simulations of retinal activity.

A decomposition of the visual input

The approximation of the input power spectrum given by Equation 7 has a useful correspondence in the spatiotemporal domain. Because of the small amplitude of the fixational motion of the retinal image, it comes natural to think of the visual input received by a retinal receptor in terms of temporal fluctuations around a mean luminance value. That is, the input signal experienced at time t by a receptor at location x on the retina, $I(x, t)$, can be written as

$$I(x, t) = \bar{I}(x) + \tilde{I}(x, t) \tag{10}$$

where $\bar{I}(x)$ is the average luminance impinging on retinal position x, $\bar{I}(x) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} I(x, t)\mathrm{d}t$, and $\tilde{I}(x, t)$ is the zero-mean signal given by the deviation at time t from this average.

This input decomposition is shown in Figure 4 for the case of a fixation on a natural image. In 2D space, the visual input entering the eye can be regarded as a movie, $I(\mathbf{x}, t)$. Each frame of this movie corresponds to a snapshot at a particular moment in time, that is, the input image resulting from viewing the scene from the

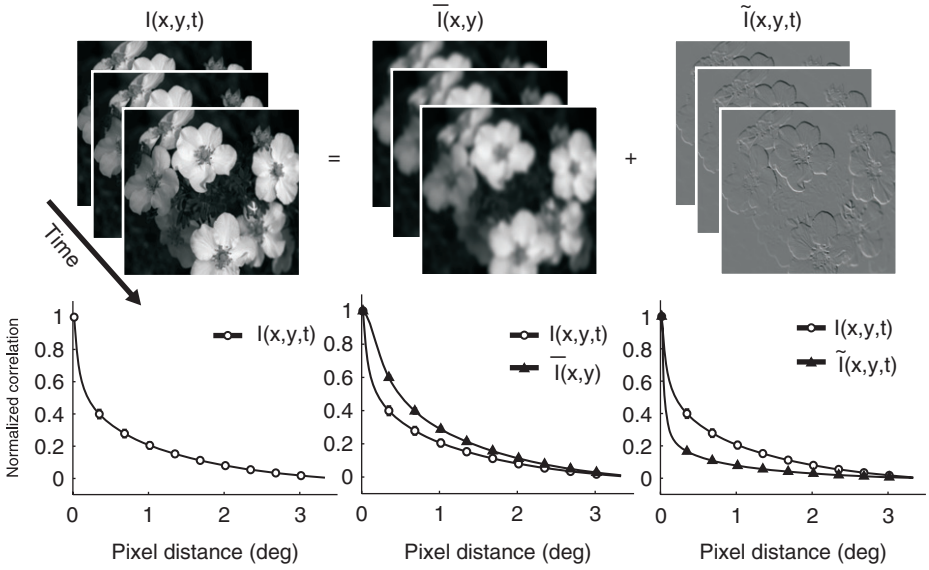


Figure 4. Deconstructing the retinal stimulus during fixational instability. (*Top*) At each instant in time t, the input signal $I(\mathbf{x}, t)$ impinging on the retina can be split into the sum of two images: a static image $\bar{I}(\mathbf{x})$, in which every pixel consists of the temporal average of the luminance value experienced by retinal location $\mathbf{x}$ over the period of fixation; and a time-varying image $\tilde{I}(\mathbf{x}, t)$, in which each pixel represents the instantaneous deviation with respect to $\bar{I}(\mathbf{x})$. (*Bottom*) Spatial correlation in pixel intensity in the frames of the three movies. $\bar{I}(\mathbf{x})$ and $\tilde{I}(\mathbf{x}, t)$ present, respectively, broader and narrower spatial correlations than those in the original stimulus.

Downloaded By: [Boston University] At: 23:24 21 November 2008

position assumed by the retina at time t. Each frame of $I(\mathbf{x}, t)$ can be reconstructed from the sum of two separate images: (1) the image $\bar{I}(\mathbf{x})$ given by the average of all the frames of $I(\mathbf{x}, t)$, and (2) the image $\tilde{I}(\mathbf{x}, t)$ given by the difference between $I(\mathbf{x}, t)$ and $\bar{I}(\mathbf{x})$. Thus, the visual input $I(\mathbf{x}, t)$ can be separated into the sum of two signals, $\bar{I}(\mathbf{x}, t)$ and $\tilde{I}(\mathbf{x}, t)$. Note that only the latter signal varies with time. The first signal is a movie which contains the same image at every time step: $\bar{I}(\mathbf{x}, t) = \bar{I}(\mathbf{x})$.

The input representation given by Equation 10 is the equivalent in the spatiotemporal domain of the power spectrum approximation of Equation 7. This equivalence can be shown by using Equation 10 to calculate the cross-correlation function of I, $R_{II}(x, t)$. The cross-correlation of the visual input can also be written as the sum of two terms:

$$R_{II}(x, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(\lambda, \gamma) I(x + \lambda, t + \gamma) \mathrm{d}\lambda \mathrm{d}\gamma = R_{\bar{I}\bar{I}}(x, t) + R_{\tilde{I}\tilde{I}}(x, t)$$

where $R_{\bar{I}\bar{I}}$ and $R_{\tilde{I}\tilde{I}}$ are the cross-correlations of $\bar{I}$ and $\tilde{I}$, respectively. The two cross-terms are both zero because $\bar{I}$ does not change with time and $\tilde{I}$ has zero mean.

Since the power spectrum of the visual input is by definition the Fourier transform of $R_{II}(x, t)$, $P_I(u, \omega)$ can be written as the sum of the power spectra of $\bar{I}$ and $\tilde{I}$:

$$P_I(u, \omega) = \mathcal{F}\{R_{\bar{I}\bar{I}}(x, t)\} + \mathcal{F}\{R_{\tilde{I}\tilde{I}}(x, t)\} \tag{11}$$

These two terms can be approximated by the two power spectra in Equation 7, as it can be shown by explicitly calculating one of them. For example, under the assumption of retinal image motion with zero mean, we obtain

$$\bar{I}(x) = I(x, 0) + \frac{\partial^2 I}{\partial x^2}\sigma_\xi^2$$

which, using an approach similar to the one of section "An estimate of the input spectrum during fixational instability", gives

$$P_{\bar{I}\bar{I}}(u, \omega) \approx P_L(u)\delta(\omega)(1 - u^2\sigma_\xi^2) \tag{12}$$

Comparison of Equations 12 and 14 shows that the two decompositions are equivalent: $P_{\bar{I}\bar{I}} \approx P_I^S$ and $P_{\tilde{I}\tilde{I}} \approx P_I^D$.

The three curves in the bottom row of Figure 4 show the average correlations in pixel intensity exhibited by the frames of the three movies. In $I(\mathbf{x}, t)$, levels of correlation tend to decline slowly with pixel distance. This broad correlation is to be expected from the scale-invariant structure of the power spectrum of natural images. That is, intensity values tend to vary smoothly in natural images, as most power occurs at low spatial frequencies. $\bar{I}$ exhibits even broader correlations than those in $I(\mathbf{x}, t)$ because of the spatial smoothing operated by fixational instability. In fact, the intensity of each pixel in $\bar{I}$ contains an average evaluated over a neighborhood of pixels in $I(\mathbf{x}, t)$. In contrast, pixel pairs in $\tilde{I}(\mathbf{x}, t)$ tend to possess correlated intensities only if they are located in close proximity to each other. This reduction in the extent of spatial correlations can be understood by remembering that at each temporal frequency the power spectrum of this movie is approximately proportional to $P_L(u)|u|^2$. For a natural image, these two terms counterbalance each other, yielding an almost flat spectral distribution. A broad power spectrum corresponds to a narrow spatial correlation.

Downloaded By: [Boston University] At: 23:24 21 November 2008

Influence of fixational instability on neural activity

The analysis of visual input described in section "Retinal input during visual fixation" provides a tool for investigating the possible influences of the fixational motion of the retinal image on neural activity. Many neurons in the early stages of the visual system respond to visual stimuli in a way that can be approximated by linear or quasi-linear models (Carandini et al. 2005). Since a linear system verifies the superposition property, the visually-evoked response of a linear neuron during fixational instability can be evaluated on the basis of Equation 10, by summing up individual responses to each of the two input components, $\bar{I}(\mathbf{x})$ and $\tilde{I}(\mathbf{x}, t)$.

Let us consider an ideal linear neuron η with mean instantaneous firing rate given by $\eta(t) = T_\eta\{I(\mathbf{x}, t)\}$, where T_η represents the linear transformation operated by the neuron. From Equation 10, we obtain

$$\eta(t) = T_\eta\{I(\mathbf{x}, t)\} = T_\eta\{\bar{I}(\mathbf{x})\} + T_\eta\{\widetilde{I}(\mathbf{x}, t)\}. \tag{13}$$

That is, the response of the neuron during fixational instability will depend on its sensitivity to two input signals with very different spatiotemporal characteristics.

The analysis of the previous section allows estimation of the structure of correlated activity in a population of linear, time-invariant neuronal elements with transfer function $H_\eta(u,\omega)$. From Equation 7, we obtain the power spectrum of neural activity $P_\eta(u,\omega)$:

$$P_\eta(u,\omega) \approx |H_\eta(u,\omega)|^2 P_I^S(u,\omega) + |H_\eta(u,\omega)|^2 P_I^D(u,\omega). \tag{14}$$

The inverse Fourier Transform of $P_\eta(u,\omega)$ gives the spatiotemporal correlation $r_\eta(x,t)$, i.e., the correlation at time delay t in the responses of two neurons with receptive fields center at distance x:

$$r_\eta(x,t) = r_\eta^S(x,t) + r_\eta^D(x,t) \tag{15}$$

where

$$\begin{cases} r_\eta^S(x,t) = \mathcal{F}^{-1}\{|H_\eta(u,\omega)|^2 P_L(u)\delta(\omega)[1-u^2\sigma_\xi^2]\} \\ r_\eta^D(x,t) = \mathcal{F}^{-1}\{|H_\eta(u,\omega)|^2 u^2 P_L(u) P_\xi(\omega)\}. \end{cases}$$

Equation 15 shows that the degree of correlation in the responses of two cells depends on the correlation given by the static stimulus, $\bar{I}(x)$, and on the correlation given by the fixational modulations of luminance, $\tilde{I}(x,t)$. The influence of each of these two terms depends on neuronal selectivity. In general, the responses of neurons sensitive to time-varying stimuli will be influenced by the fluctuations of luminance caused by fixational eye movements. Even if fixational fluctuations tend to possess small amplitudes, these modulations may elicit strong responses in linear filters that are tuned to them. Neurons that are instead sensitive to stationary patterns will be strongly stimulated by $\bar{I}(\mathbf{x})$. Examples of neurons with different characteristics are given in the following sections.

Downloaded By: [Boston University] At: 23:24 21 November 2008

Two biologically-implausible examples

To illustrate the usefulness of Equation 15, this section describes two model "neurons" for which levels of correlation can be expressed in closed analytical forms. The response characteristics of these two neuronal elements are unrealistic, but the formulas obtained in this section provide a good reference for understanding the correlated activity of more biologically-plausible models.

In the first example, we consider a neuronal element that responds equally to all nonzero temporal frequencies, but does not respond to temporal DC:

$$H_\eta(u,\omega) = \begin{cases} 1 & \text{if } \omega \neq 0 \\ 0 & \text{if } \omega = 0. \end{cases} \tag{16}$$

From Equation 15, we obtain that the spatiotemporal correlation function, $r_\eta(x,t)$, in the responses of a population of neurons with these characteristics will depend only on $r_\eta^D(x,t)$, as $r_\eta^S(x,t) = 0$. To calculate the instantaneous correlation in the responses of two units (the correlation at delay $t=0, r_\eta(x,0)$), we need to integrate over temporal frequency:

$$r_\eta(x,0) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_I^D(u,\omega)e^{-jux}\,\mathrm{d}\omega\,\mathrm{d}u = \frac{1}{2\pi}\int_{-\infty}^{\infty} u^2 P_L(u)\sigma_\xi^2\,\mathrm{d}u \tag{17}$$

where the term within the integral is the total dynamic power resulting from fixational instability. As shown in Figure 3, $\widehat{P}_I^D(u)$ is constant at all spatial frequencies during viewing of natural images, as the term u^2 counterbalances the scale-invariant spectrum of natural images. Thus, from Equation 17, we obtain

$$r_\eta(x, 0) = \delta(x)\sigma_\xi^2 \tag{18}$$

which shows that the responses of nonoverlapping units are uncorrelated. This result occurs because, with presentation of natural images, an ideal neuron with the characteristics of Equation 16 is effectively driven by a component of the visual input with flat spatial spectrum, i.e., a signal that is spatially uncorrelated.

The second example of ideal "neuron" is given by a filter that only responds to input modulations at a specific temporal frequency independent of the spatial frequency:

$$H_\eta(u, \omega) = \frac{1}{2}[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)] \tag{19}$$

As in the previous example, the spatiotemporal correlation function $r_\eta(x, t)$ will only depend on $r_\eta^D(x, t)$. The term $r_\eta^S(x, t)$ in Equation 15 is zero, because this neuron is not sensitive to input stimuli that do not change in time ($H_\eta(u, 0) = 0$). In this case, the spatial correlation at zero time delay during viewing of natural images is given by

$$r_\eta(x, 0) = k\delta(x) \tag{20}$$

where the constant of proportionality depends on the power of fixational instability at frequency ω_0. Again, the result of Equation 20 occurs because this ideal filter is effectively driven by a spatially uncorrelated component of the visual input.

The impulse responses of the models in Equations 16 and 19 have infinite durations and lack the property of causality. Neurons with these characteristics do not exist in the brain. Next section examines the patterns of correlated activity established by more realistic models.

Retinal ganglion cells

Retinal ganglion cells differ in important ways from the ideal neurons described in the previous section. While the visual responses of many of these cells can be well approximated by rectified linear filters, unlike the example of section "Two biologically-implausible examples", the transfer functions of these filters only differ from zero in limited ranges of spatial and temporal frequencies. Thus, the structure of correlated activity in populations of retinal ganglion cells will depend on the precise interaction between the retinal stimulus and neuronal characteristics, as specified by Equation 15. More specifically, the impact of each of the two terms r_η^S and r_η^D in Equation 15 will depend on the relative sensitivity of retinal neurons to static and time-varying stimuli: r_η^S will prevail in cells that respond strongly to unchanging stimuli in their receptive fields; r_η^D will be influential in neurons that prefer time-varying input signals. Neuronal sensitivity to these two types of stimuli is determined by the characteristics of the neuron's transfer function on, and outside of, the DC axis: $H_\eta(u, 0)$ and $H_\eta(u, \omega \neq 0)$, respectively.

Downloaded By: [Boston University] At: 23:24 21 November 2008

Figures 5 and 6 show the spatial structure of correlated activity measured in simulations of ganglion cells in the retina of the macaque. Neuronal models in these simulations consisted of rectified linear filters, which were separable in their spatial and temporal kernels, $F(\mathbf{x})$ and $G(t)$: $h(\mathbf{x}, t) = F(\mathbf{x})G(t)$. Kernels were designed on

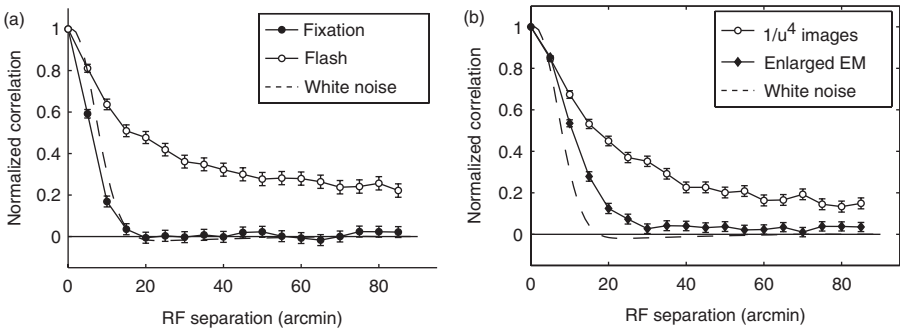


Figure 5. Spatial structure of correlated activity in populations of magnocellular ganglion cells in the macaque's retina. Data represent mean levels of pairwise correlation between neurons with receptive fields at various separations (x-axis) measured under various conditions. (a) Spatial correlations measured during normal fixation on natural images (Fixation) and when the same images were flashed without eye movements (Flash). (b) Spatial correlations measured during fixation on images with u^{-4} spectral density and during fixation on natural images with enlarged fixational eye movements. The dotted lines in both panels (white noise) represent the correlation predicted by the spatial structure of the cell kernel during stimulation with static uncorrelated input. Real traces of eye movements recorded from human observers were used in these simulations. Error bars represent 95% confidence intervals based on the SE of the mean.

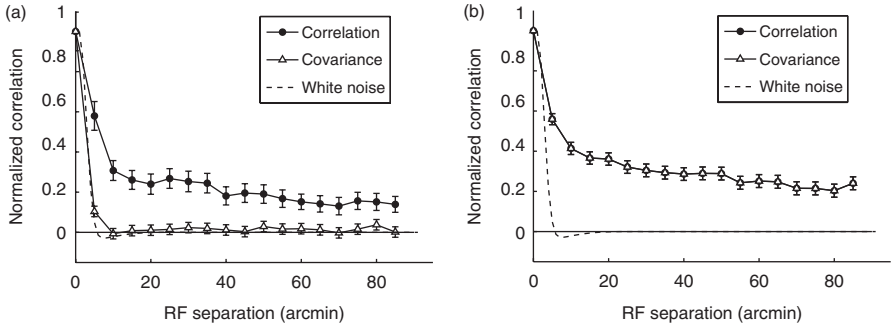


Figure 6. Spatial structure of correlated activity in populations of parvocellular ganglion cells in the macaque's retina. Data represent mean levels of pairwise correlation and covariance between neurons with receptive fields at various separations (x-axis) measured during (a) normal fixation on natural images and (b) flashing of the same set of images without eye movements. The dotted lines in both panels (white noise) represent the correlation predicted by the spatial structure of the cell kernel during stimulation with static uncorrelated input. Real traces of eye movements recorded from human observers were used in these simulations. Error bars represent 95% confidence intervals based on the SE of the mean.

Downloaded By: [Boston University] At: 23:24 21 November 2008

the basis of published physiological data from ON-center parvocellular (P) and magnocellular (M) ganglion cells (Croner and Kaplan 1995; Benardete and Kaplan 1997, 1999). Neuronal responses were modeled while receptive fields scanned natural images following traces of eye movements previously recorded from human observers during viewing of the same images presented to the model. Levels of correlation in the responses of pairs of cells were averaged over a large number of fixations on many images and also averaged over all pairs of cells with receptive fields at the same distance, independent of the orientation of the axis on which the two receptive fields were aligned. In this way, we obtained a function of only one spatial dimension: the distance between receptive field centers.

The left panel in Figure 5 shows that fixational eye movements strongly influenced correlated activity in populations of M cells. The curve marked by dark circles in Figure 5(a) represents the instantaneous correlation (the correlation function at zero delay, $r_\eta(x, 0)$) in the responses of pairs of neurons with receptive fields at separation x measured during fixation on natural images. As illustrated by these data, the responses of neurons with nonoverlapping receptive fields were completely uncorrelated during fixation on a natural image. Like the examples of section "Two biologically-implausible examples", magnocellular neurons tend to respond poorly to stationary stimuli and are instead strongly sensitive to time-varying input. This temporal sensitivity discards the influence of the broad correlations of natural images and is responsible for the shape of the correlation in Figure 5(a).

It is important to note that the narrow spatial correlation profile exhibited by the responses of magnocellular neurons originates from the spatiotemporal modulations of luminance caused by fixational eye movements during viewing of natural images. Spatial correlations were much broader when the same set of natural images was observed in the absence of eye movements or during presentation of stimuli with spectral density distribution different from that of natural images. The line marked by white circles in Figure 5(b) represents the spatial correlation function measured when the same natural images used in the previous simulation were flashed while keeping receptive fields stationary. In this case, the sudden change in visual input caused by the flashing of the stimulus gave rise to wide pools of synchronously active cells. As illustrated by these data, in the absence of retinal image motion, model neurons exhibited correlated responses even when their receptive fields were at considerable distance. This pattern of activity was a consequence of the broad intensity correlations of natural images.

The different structure of correlated activity measured during fixational instability and during flashing of the stimulus can be explained on the basis of Equation 15. Since M cells respond poorly to unchanging stimuli, $H_\eta(u, 0) \approx 0$ and $r_\eta^D(x, t) \gg r_\eta^S(x, t)$. Therefore, during fixational instability, we obtain

$$r_\eta(x, 0) \approx \mathcal{F}^{-1}\{|H_\eta(u, \omega)|^2 P_\xi(\omega)\}_{t=0}$$

$$r_\eta(x, 0) \approx \frac{1}{4\pi^2}\int_{-\infty}^{\infty} |G_\eta(\omega)|^2 P_\xi(\omega)\mathrm{d}\omega \int_{-\infty}^{\infty} |F_\eta(u)|^2 e^{-jux}\,\mathrm{d}u \propto \int_{-\infty}^{\infty} |F_\eta(u)|^2 e^{-jux}\,\mathrm{d}u \tag{21}$$

where $F(u)$ and $G(\omega)$ are the Fourier Transforms of $F(x)$ and $G(t)$, and we used the fact that $u^2 P_L(u)$ is approximately constant for natural images. Equation 21 explains

Downloaded By: [Boston University] At: 23:24 21 November 2008

that the only correlation exhibited by M cells should result from the overlap between receptive fields, as during the static presentation of spatial white noise. The dotted line in Figure 5(a) shows the spatial correlation profile predicted by Equation 21. In agreement with this prediction, levels of correlation measured in our simulations during fixation on natural images could be fully accounted for by the overlap in receptive fields, as during presentation of uncorrelated input.

The pattern of correlation given by Equation 21 should be compared to the spatial correlation predicted by flashing the stimulus in the absence of retinal image motion. In this case, levels of correlation are given by

$$r_\eta(x, 0) \propto \int_{-\infty}^{\infty} |F_\eta(u)|^2 P_L(u) e^{-jux}\, \mathrm{d}u \tag{22}$$

where the constant of proportionality depends on the sensitivity of M cells to the temporal arrangement of the flashes. The only difference with respect to Equation 21 is the presence of the power spectrum of the stimulus, $P_L(u)$. This equation makes clear that the broad correlations measured in our simulations during flashing of the stimulus originated from the correlated structure of natural images.

The structure of correlated activity was significantly different in simulations of P cells, as illustrated by the data in Figure 6. Wide ensembles of coactive P units emerged during fixation and extended over several degrees of visual field. Even the responses of pairs of cells with no overlap in their receptive fields were strongly correlated during viewing of natural images. The reason for this difference with respect to M cells lies in the temporal sensitivity of P cells. Unlike M cells, P neurons tend to exhibit sustained responses to stationary stimuli. Thus, $H_\eta(u, 0)$ is different from zero, and the contribution of $r_\eta^S(x, t)$ is not negligible, as in the case of M cells. Because of this temporal sensitivity, P cells were strongly influenced by the broad correlations of natural images. Therefore, as predicted by Equation 15, the patterns of correlated activity measured in P cells during fixational instability and during flashing of the stimulus were more similar to each other than in M cells (Figure 6(b)).

Even though levels of correlations in P cells were less influenced by fixational eye movements, image motion maintained a critical impact on the second-order statistics of neural activity in this population. The curves marked by triangles in Figure 6 show the patterns of covariance in neural responses. These curves were obtained by subtracting from the instantaneous firing rates the mean responses of modeled cells during fixation. They reveal that P cell responses fluctuated around their mean values were spatially uncorrelated during fixation on natural images. This effect is important, as neurons downstream in the visual pathway might be sensitive to changes in the incoming firing patterns (Abbott et al. 1997). Therefore, cortical neurons could be effectively driven by an uncorrelated parvocellular input, even if the mean firing rates of P cells are correlated. In the patterns of correlation, this effect is hidden by responses to the unchanging component of the retinal stimulus. Levels of correlation and of covariance were almost indistinguishable in magnocellular populations because M cells are not strongly sensitive to stationary stimuli.

To summarize, fixational eye movements effectively decrease the spatial extent of correlated activity in simulations of retinal activity. Pairs of neurons with

Downloaded By: [Boston University] At: 23:24 21 November 2008

nonoverlapping receptive fields exhibit no covariance (M and P neurons) and no correlation (M neurons) in their firing rates. This effect is a direct consequence of the linear interaction between the temporal characteristics of cells and the statistics of visual inputs. It originates from an input component that lacks of spatial correlations (the term $\tilde{I}$ in Equation 10). In the following of this article, we collectively refer to these changes in the statistics of neuronal firing as "neural decorrelation" to indicate that the pattern of coactivation during fixation is identical to that resulting from the static presentation of a white noise stimulus. Neural decorrelation occurs only during unstable fixation on images with a spectral density like that of natural images. In simulations, this decorrelation of retinal activity was extremely robust. It was little affected by the spatial parameters of simulated receptive fields, their eccentricity, their degree of space–time inseparability, or the presence of nonlinear mechanisms – such as rectification and contrast gain control – in the generation of neural responses (Desbordes and Rucci 2007).

Downloaded By: [Boston University] At: 23:24 21 November 2008

Fixational eye movements and fine spatial vision

An interesting implication of the decorrelation of neural activity described in section "Influence of fixational instability on neural activity" is an involvement of fixational eye movements in fine-scale spatial vision. This hypothesis stems directly from the redistribution of spectral density that occurs when a scene is viewed through jittering eyes. As explained in section "Retinal input during visual fixation", the extent by which the fixational motion of the retinal image spreads the spatial power of the stimulus across temporal frequencies is not constant throughout the spatial frequency plane, but increases with the spatial frequency. In other words, during the normal instability of visual fixation, neurons in the early visual system are exposed to changes in luminance (the term $\tilde{I}(x, t)$ in Equation 10), which enhance the high spatial frequency components of the stimulus. Therefore, visual processes that exploit these input changes should benefit from the presence of fixational eye movements in tasks in which high-frequency information is needed. An example of the enhancement of high spatial frequencies in the fixational modulations of luminance is given in Figure 4, where $\tilde{I}(x, t)$ emphasizes discontinuities in the scene such as edges and object boundaries (see right panel in the figure).

To clarify this point, consider the ideal neuronal element η described in section "Two biologically-implausible examples", which responds to modulations at a single temporal frequency (Equation 19). This neuron will not be active in the absence of temporal changes in the input. During fixation on a static stimulus, the input effective in driving η will depend on both the characteristics of fixational instability and the spatial frequency content of the stimulus. With presentation of a white noise pattern, which contains equal power at all spatial frequencies, η will respond only to spatial frequencies higher than a threshold. The value of this threshold is determined by the motion of the retinal image and decreases with its velocity, so that η responds to lower frequencies with higher velocities of retinal image motion. The contrast sensitivity functions of retinal ganglion cells clearly differ from the transfer function of η. Yet many ganglion cells, in both magnocellular and parvocellular streams, prefer nonzero temporal frequencies. In these neurons, the spectral redistribution caused by fixational instability transfers part of the power

which would otherwise be at zero temporal frequency to a range of nonzero temporal frequencies to which neurons are more responsive. This redistribution has two consequences: (a) an enhancement of responses to high spatial frequencies; and (b) a suppression of responses to low spatial frequencies relative to responses evoked by modulation of a stationary stimulus.

The analysis of the previous sections enables the formulation of specific predictions regarding the possible role of fixational eye movements in the perception of spatial detail. Two of these predictions are schematically shown in Figure 7. The stimulus in Figure 7(a) consists of a grating at frequency f_G embedded in a noise field with maximum frequency f_N. A critical feature of this stimulus is that the maximum frequency of the pattern of noise is lower than the frequency of the grating, $f_N < f_G$. We can use the method of Figure 3 to estimate the temporal power of the luminance modulations resulting from viewing this stimulus during the normal instability of visual fixation. Obviously real neurons cannot integrate over all temporal frequencies as in Figure 3. However, since the spectral profiles of Figure 3 also occur at each individual temporal frequency, this method gives insights into the way retinal ganglion cells would respond. As shown in Figure 7(a), because of the power-law rule governing the transformation of spatial power into temporal power, the changes of luminance resulting from fixational eye movements will emphasize the grating relative to the noise. Neurons sensitive to these temporal changes would be driven primarily by the grating rather than the pattern of noise. In fact, selectivity to fixational modulations of luminance is a powerful mechanism for rejecting noise at low spatial frequencies. To provide an example of the efficiency of this mechanism, the image on the right of Figure 7(a) shows the effect of filtering the stimulus by means of a spatial filter with transfer function proportional to the frequency. This operation significantly attenuates the pattern of noise. In practice,

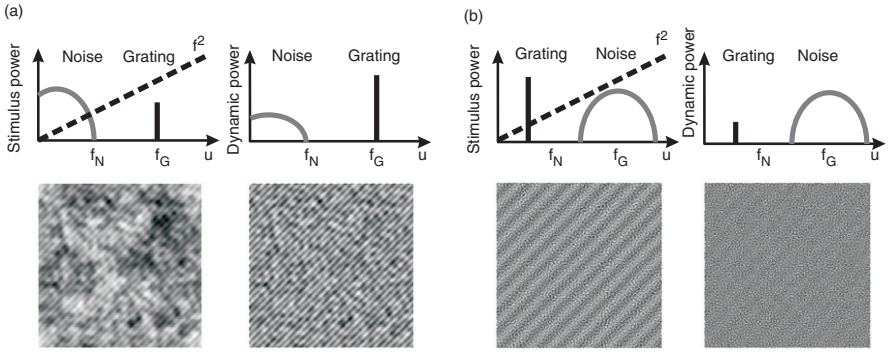


Figure 7. Two predictions of the redistribution of power caused by the fixational motion of the retinal image. (a) The changes in the visual input occurring during fixational eye movements should facilitate detection/discrimination of gratings at frequencies higher than the frequency range of a noise field in which they are embedded. (b) Fixational eye movements should instead have no beneficial effect with gratings at frequencies lower than the pattern of noise. In both panels, the top row shows the spectral density of the stimulus (*left*) and the total temporal power of fixational modulations predicted by the theory (*right*). The bottom row shows the stimulus (*left*) and the effect of altering its frequency content by means of a u^2 filter (*right*), a spatial effect similar to that attributed to fixational eye movements in the temporal domain (see text for details).

Downloaded By: [Boston University] At: 23:24 21 November 2008

the high-frequency enhancement resulting from retinal image motion is not static, as in the example given by this image, but occurs in the time-varying stimulus $\tilde{I}(x, t)$ at temporal frequencies different from zero.

The prediction of Figure 7(a) relies on the fact that the frequency of the grating is higher than the frequency range of the pattern of noise in which the grating is embedded. When the frequency order of these two components of the stimulus is reversed, so that the frequency of the grating is lower than the minimal frequency of the noise field ($f_G < f_N$), a diametrically different prediction emerges, as shown in Figure 7(b). In this case, the temporal modulations resulting from fixational eye movements would emphasize the pattern of noise relative to the grating, and neurons sensitive to these input changes would be primarily driven by noise. Therefore, fixational instability should improve detection and discrimination of a grating masked by low-frequency noise as in Figure 7(a) and should have no effect with a low-frequency grating masked by noise at higher frequencies as in Figure 7(b).[2]

The hypothesis that fixational eye movements contribute to the perception of fine spatial detail has a long history. This idea was originally speculated by Hering (1899) and later refined into the so-called dynamic theories of visual acuity (Averill and Weymouth 1925; Marshall and Talbot 1942; Arend 1973). Interest in these proposals declined after classical experiments on retinal image stabilization – a procedure in which retinal image motion is eliminated – reported little or no change in visual acuity (Riggs et al. 1953; Tulunay-Keesey and Jones 1976). Studies with long stimulus presentations found only a global reduction in contrast sensitivity, which appeared more pronounced at low–rather than high–spatial frequencies (Koenderink 1972; Kelly 1979; Tulunay-Keesey 1982). In addition, studies with brief stimulus durations found little or no effect of image stabilization on visual acuity and contrast sensitivity (Riggs et al. 1953; Tulunay-Keesey and Jones 1976). In spite of these early results, the proposal of an involvement of fixational eye movements in the perception of spatial detail has recently found new support from neurophysiological investigations (Snodderly et al. 2001; Greschner et al. 2002; Olveczky et al. 2003), psychophysical results (Rucci and Desbordes 2003; Rucci and Beck 2005), and theoretical analysis of the influence of a moving retinal image on the responses of neurons in the early stages of the visual system (Ahissar and Arieli 2001; Henning and Worgotter 2004; Rucci and Casile 2005; see also Pitkow et al. 2007). While these studies propose different functions and mechanisms, they all argue for a contribution of fixational eye movements to fine-scale vision.

Early studies of retinal stabilization might have failed to detect visual contributions from fixational eye movements because of both technological and methodological problems. Traditional methods for eliminating retinal image motion suffered from important technological limitations (see Steinman and Levinson 1990 for a comprehensive review). First, the devices used to stabilize images on the retina required careful calibration procedures, which did not allow fast and flexible switching between the conditions of presence and absence of retinal image motion. It was not possible to selectively stabilize the image during periods of visual fixation between macroscopic saccades, as would have been desirable in order to study fixational eye movements in their natural context (Steinman et al. 1967; Steinman and Collewijn 1980; Kapoula et al. 1986). Instead, all trials with stabilized vision had to be run in uninterrupted blocks while the subject maintained fixation, a highly unnatural condition that unavoidably led to visual fatigue and fading.

Downloaded By: [Boston University] At: 23:24 21 November 2008

Moreover, these studies lacked a method for objectively assessing the quality of stabilization. The experimenter had to rely on the subject to judge when retinal image motion had been eliminated. As a result, only a small number of very experienced subjects, in most cases the experimenters themselves, participated in these experiments.

In addition to these technical limitations, classical studies on retinal stabilization also suffered from methodological problems. First, these studies compared results obtained in the absence of retinal image motion to levels of performance measured while subjects maintained steady and accurate fixation for many seconds or minutes. Prolonged sustained fixation is an unnatural viewing condition, which is known to reduce the amount of fixational instability (Steinman et al. 1967; Skavenski et al. 1979; Kapoula et al. 1986). It is possible that the reduction in retinal image motion caused by this viewing condition contributed to the similarity in the performance levels with and without retinal motion measured by these studies. Second, early stabilization studies might not have used the best stimuli to emphasize the visual impact of fixational eye movements. Unlike the stimuli of Figure 7, the stimuli used by previous studies were either plain gratings or binary images which did not contain noise. The presence of the patterns of noise plays an important role in the predictions of Figure 7, as it enhances the putative influence of fixational eye movements. These considerations suggest that the hypothesis of an involvement of fixational eye movement in fine spatial vision might have been discarded prematurely.

Building on the predictions of Figure 7, we have recently examined the influence of fixational eye movements on the discrimination of gratings at different spatial frequencies (Figure 8). In a forced-choice discrimination task, subjects reported

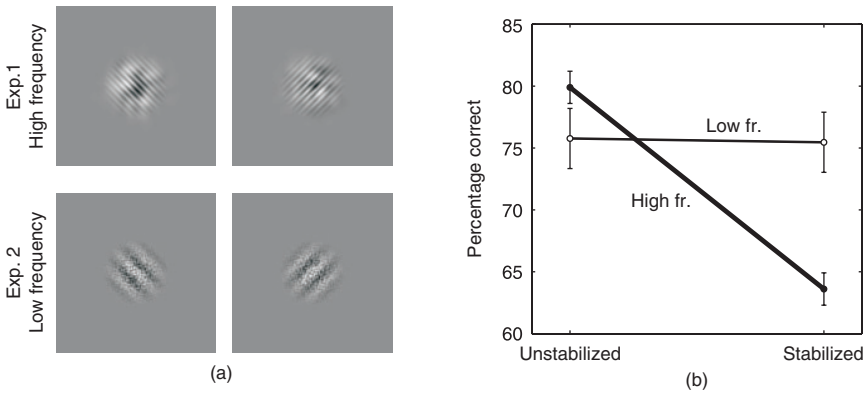


Figure 8. Results of experiments in which subjects reported the orientation ($\pm 45°$) of a noisy grating. (a) Examples of stimuli used in the experiments. In Experiment 1, an 11 cycles deg^{-1} grating was perturbed by low spatial frequency noise (low-pass cut-off frequency $f_N =$ 5 cycles deg^{-1}). In Experiment 2, the stimulus was a 4 cycles deg^{-1} grating overlapped by high spatial frequency noise (high-pass $f_N = 10$ cycles deg^{-1}). (b) Mean subject performance in the presence (unstabilized) and absence of retinal image motion (stabilized). Retinal stabilization was selectively applied to the period of fixation following a saccade. In both experiments, the contrast of the gratings was individually adjusted for each subject so that performance levels in the presence of normal retinal motion were ~75%. Error bars represent 95% confidence intervals. Modified from Rucci et al. (2007).

Downloaded By: [Boston University] At: 23:24 21 November 2008

whether a noisy grating was tilted by 45° clockwise or counter-clockwise. Two separate experiments investigated the discrimination of the stimuli shown in Figure 8(a). In Experiment 1, the stimulus was a grating at $f_G = 11$ cycles deg^{-1} perturbed by noise at low spatial frequencies ($f_N = 5$ cycles deg^{-1}). In Experiment 2, the frequency order of the grating and the noise was reversed, that is, the grating was at lower spatial frequency than the pattern of noise ($f_G = 4$ cycles deg^{-1}, $f_N = 10$ cycles deg^{-1}). In both cases, the power of the noise decreased proportionally to the square of the spatial frequency, as occurs in the power spectrum of natural images.

To overcome the limitations of previous experiments and selectively isolate the motion of the retinal image that occurs during natural fixation, we developed a new method of retinal stabilization. This method relies on EyeRIS (Eye-Movement Real-Time Integrated System), a general-purpose system for flexible gaze-contingent display control, which is based on a digital signal processor coupled with analog and digital interfaces (Santini et al. 2007). EyeRIS is an integrated hardware and software system specifically designed to process eye-movement signals in real time. The system links oculomotor events to changes in the stimulus so that the image on the monitor can be modified in real-time according to the user's specification. EyeRIS guarantees accurate retinal stabilization while offering a degree of experimental flexibility far superior to that of all other methods of retinal stabilization. This technique allows on and off switching of retinal stabilization with video-frame resolution up to 200 Hz. We used EyeRIS to display the stimulus at the onset of fixation after the subject performed a saccade toward a randomly cued location. Stimuli were either maintained at a fixed location on the screen (unstabilized condition) or were moved with the eye so as to cancel the retinal motion resulting from fixational eye movements (stabilized condition).

Figure 8(b) shows mean percentages of correct discrimination across six subjects. Performance dropped drastically with high-frequency gratings (Experiment 1), when stimuli were stabilized on the retina. On the contrary, the retinal image motion produced by fixational eye movements did not improve performance with the low spatial frequency gratings of Experiment 2. Thus, in keeping with the predictions of Figure 7, fixational eye movements improved discrimination of the orientation of a high-frequency grating masked by low-frequency noise but did not help with a low spatial frequency grating masked by high-frequency noise. This result contradicts traditional views of the influence of fixational eye movements on vision. In contrast with our findings, the pronounced reduction in contrast sensitivity at low spatial frequencies measured by previous experiments with prolonged retinal stabilization (Koenderink 1972; Kelly 1979; Tulunay-Keesey 1982) would have predicted a more significant drop in performance with low-frequency rather than high-frequency gratings. As with the stimuli of Experiment 1, the power spectra of natural visual environments are dominated by low spatial frequencies. Thus, the results of Figure 8 suggest that sampling visual information by means of a jittering fixation is an effective strategy for analyzing natural scenes, as it facilitates the processing of spatial detail in the face of otherwise overwhelming low-frequency power.

A visuomotor neural code?

The theory described in this article has important implications regarding the neural encoding of visual information. It suggests that the physiological instability of

Downloaded By: [Boston University] At: 23:24 21 November 2008

fixation is an important contributor to neural representations in the early stages of the visual system, and that a full understanding of these representations needs to take into account not only the stimulus but also the motor activity of the observer. Three important implications of this theory are described in this section.

Discarding redundant correlations

The first implication of the results shown in the previous sections is a role of the fixational motion of the retinal image in the establishment of compact neural representations during viewing of natural stimuli.

The fact that most power in images of natural scenes is at low spatial frequencies implies that these images tend to vary smoothly in space. That is, it is often possible to predict the intensity of a pixel at a certain location on the basis of the intensity values of adjacent pixels. This input redundancy has long been regarded as a challenge faced by the visual system, as it implies that many neurons, even with receptive fields far apart, would tend to represent similar information. While a moderate degree of redundancy in neural firing might help establish robust representations, a scheme of neural encoding in which massive ensembles of coactive neurons signal the same visual features does not appear to be an efficient way to allocate resources. Furthermore, the neural pathway from the retina to the cortex constitutes a bottleneck in the flow of visual information. It would make sense to first compress the input signals in order to transmit them more efficiently from the retina to the cortex. Similar challenges confront communication engineers when transmitting data. It has been proposed that an important function of early visual processing is the removal of part of the redundancy that characterizes natural visual input, much in the same way that engineers preprocess data before sending it over a communication channel (Attneave 1954; Barlow 1961).

Input redundancies can be attenuated in several ways (Field 1994). It has long been postulated that lateral inhibition in the retina and LGN might be involved in this process (Srinivasan et al. 1982; van Hateren 1992), a hypothesis based on the observation that inhibition makes the responses of neurons in these structures less predictable. In an extreme version of this idea, the antagonistic center-surround organization of retinal ganglion cells has been held responsible for preventing the broad correlations of natural images from affecting neural responses (Atick and Redlich 1990, 1992; Atick 1992). This proposal requires the contrast sensitivity functions of ganglion neurons to increases proportionally with spatial frequency as shown in the insert of Figure 9. This frequency behavior counterbalances the spectral density of natural images and results in a spatially-decorrelated output.

Atick and Redlich's hypothesis was based on psychophysical measurements of human contrast sensitivity, which were taken to represent a cumulative envelope of the frequency responses of ganglion cells. However, data from neurophysiological recordings have shown that the sensitivities to low spatial frequencies of retinal and geniculate neurons deviate from the frequency responses of ideal decorrelating spatial filters in a way that is not consistent with a decorrelation of neural responses (Figure 9). While deviations from the ideal frequency behavior are to be expected and can be tolerated at high spatial frequencies, a strict proportionality with frequency is required in the low-frequency range, where most of the power of natural images is concentrated. The deviations shown in Figure 9 occur because,

Downloaded By: [Boston University] At: 23:24 21 November 2008

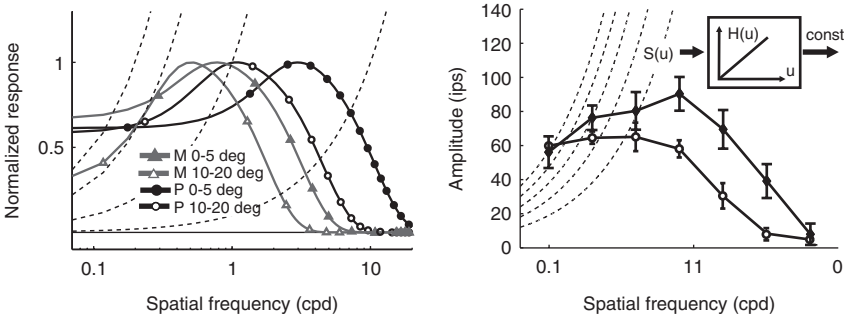


Figure 9. The response characteristics of neurons in the retina and lateral geniculate nucleus deviate from those of ideal decorrelating filters. (*Left*) Contrast sensitivity of parvocellular (P) and magnocellular (M) ganglion cells in the macaque retina. Data represent median values recorded by Croner and Kaplan (1995) in the central visual field (0–5°) and in a more eccentric area (10–20°). (*Right*) Contrast sensitivity functions of two geniculate neurons recorded in the alert macaque (unpublished data, courtesy of Alan Saul). In both panels, the dotted lines represent the transfer functions of ideal decorrelating filters, which counterbalance the spectral distribution of natural images ($H(u) = Au$, see insert panel). These functions have curvilinear shapes because the x-axis is logarithmically scaled. The frequency characteristics of the neurons shown in this figure predict neuronal responses to be strongly influenced by the broad correlations of natural images.

Downloaded By: [Boston University] At: 23:24 21 November 2008

in the macaque, inhibitory cell surrounds tend to be too weak to cancel out the influence of receptive-field centers. As shown in Figures 5 and 6, the responses of model neurons with these contrast sensitivity functions are broadly correlated when natural images are viewed without fixational eye movements.

The spatial decorrelation speculated by Atick and Redlich (1992) implicitly assumes a direct projection of the visual scene onto the retina without considering that fixational instability continually modulates input signals. The results of Figures 5 and 6 suggest a different scenario. In these simulations, cell responses were uncorrelated during the normal motion of the retinal image even if the spatial characteristics of neurons were far from those of ideal decorrelating filters. Unlike the static decorrelation mechanism of Figure 9, the neural decorrelation shown in Figures 5 and 6 originated from the interaction between fixational modulations of luminance and the temporal sensitivity of ganglion cells, not from their spatial characteristics. Three main components cooperated to this result: (1) the sensitivity of simulated neurons to fixational modulations, (2) the small spatial region covered by eye movements, and (3) the second-order statistics of natural images. Fixational instability transforms spatial changes within a neighborhood of a cell receptive field into temporal fluctuations. In natural images, pixel intensities tend to be correlated over large distances, but local changes of luminance are spatially uncorrelated (Rucci and Casile 2005).[3] Thus, neurons sensitive to fixational modulations of luminance exhibit uncorrelated responses during viewing of natural images.

It should be noted that the spatial decorrelation of neuronal responses caused by fixational instability does not come at the expense of an increase in the temporal correlation. This point might appear counterintuitive, as it may seem that retinal image motion transforms spatial correlations into temporal correlations. However, since the actual pattern of fixational eye movements changes at every fixation, there

is no systematic relationship between the firing of neurons with receptive fields at any given relative position, besides the correlation caused by receptive field overlap. On the contrary, also levels of correlation at nonzero time delays tend to decrease during fixational instability. This result can be explained on the basis of Equation 6, our approximation of the power spectrum of visual input to the retina. This equation shows that the temporal spectral density of static images scanned by fixational eye movements is approximately equal to the power spectrum of fixational eye movements. This spectrum is known to decline with temporal frequency as ω^{-2} (Eizenman et al. 1985), an input regime that appears to be well counterbalanced by the temporal sensitivity of many ganglion cells. A similar mechanism has already been proposed to explain the temporal decorrelation of geniculate responses observed during static presentation of time-varying natural scenes (Dan et al. 1996).

Downloaded By: [Boston University] At: 23:24 21 November 2008

Emphasizing useful correlations

An early compression of information does not only save metabolic resources and facilitate the transfer of information to the cortex; it is also beneficial for emphasizing the most interesting elements of the visual stimulus, those that cannot be predicted from the mere knowledge of the statistical properties of the environment. That is, compact early representations might simplify the computational tasks of later processing stages, where these elements are extracted and used.

Indeed, the changes in the second-order statistics of neural activity shown in the previous sections acquire importance in the light of theories that emphasize a role for the precise spatiotemporal organization of neuronal firing in visual perception. The temporal structure of cortical activity in the processes of image segmentation and feature binding has been one of the most debated issues of current neuroscience (for reviews see Singer and Gray 1995; Gray 1999; Singer 1999; Shadlen and Movshon 1999). Correlated cell responses, both in the form of coactivity of instantaneous firing rates (Roelfsema et al. 2004) and as synchronous spikes (Singer and Gray 1995), might signal the presence of important features in the visual scene such as an edge or an object. Furthermore, in the retina and thalamus, simultaneously active neurons are more likely than individual cells to evoke responses at later stages (Alonso et al. 1996; Usrey and Reid 1999; Usrey et al. 2000). It is surprising that relatively little attention has been paid in this context to the modulations exerted by eye movements. Eye movements are a powerful source of correlation in neural activity, as they induce synchronous changes in the stimuli within receptive fields. A synchronization in the responses of ganglion cells during motion of the stimulus that replicated the effect of fixational eye movements has already been observed in the retina (Greschner et al. 2002).

The average decorrelation of neural activity shown in Figures 5 and 6 should not be taken to imply that fixational instability does not induce useful correlations. This decorrelation is a statistical phenomenon that occurs over multiple fixations. That is, because of the characteristics of natural images, it is not possible to systematically predict the influence of fixational instability on the response of a ganglion cell from the activity of another neuron with nonoverlapping receptive field. During each individual fixation, however, a specific pattern of correlated activity exists, which is determined by both the stimulus and the motion of the

retinal image. This pattern differs from one fixation to the next. The average decorrelation caused by fixational eye movements guarantees that the short-lived correlations in cell responses established during individual fixations do not depend on the "uninteresting" broad correlations of natural scenes, but emphasize instead the unpredictable elements of the stimulus. Thus, fixational instability might also emphasize useful information during viewing of naturalistic stimuli.

To clarify this point, Figure 10 summarizes the results of a study, in which we modeled retinal activity during the psychophysical experiments of section "Fixational eye movements and fine spatial vision" (Poletti and Rucci 2008). In these experiments, subjects discriminated the orientation ($\pm 45^\circ$) of a noisy grating as explained in Figure 8. Each trial of the simulations modeled a trial in the experiments. That is, in each simulated trial, model neurons received as input a movie, which reconstructed the retinal stimulus experienced by the subject during the corresponding experimental trial. This movie was generated by translating the stimulus that had been displayed in the considered trial following the eye movements performed by the subject. As explained in Figure 10(a), we simulated the responses of two line-arrays of retinal ganglion cells with receptive fields aligned along the two possible orientation axes of gratings. The output signals of these two cell arrays eventually converge on different orientation-selective neurons in the primary visual cortex. Thus, these signals provide an estimate of the input received by V1 neurons with preferred orientation parallel and orthogonal to the grating.

As shown by the top panel of Figure 10(b), the structure of correlated activity was minimally influenced by retinal image motion when the grating was at lower frequency than the pattern of noise. Data points in Figure 10(b) represent the mean

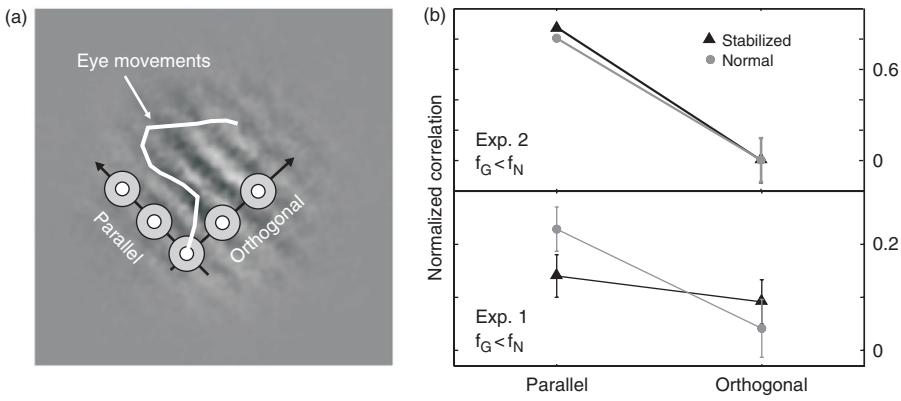


Figure 10. Structure of retinal activity in simulations of the experiments of Figure 8. (a) Models of parvocellular ganglion cells were exposed to input signals that replicated the retinal stimuli experienced by subjects. The model consisted of two arrays of cells with receptive fields aligned along the two possible orientations that a grating could assume in a trial. In unstabilized trials, receptive fields moved following recorded traces of eye movements. No retinal image motion occurred in stabilized trials. (b) Mean correlation coefficients in the responses of neurons on the two axes. Both the cases of a grating at lower (Experiment 2) and higher frequency than the band of noise (Experiment 1) are shown. In both experiments, the contrast sensitivity functions of model neurons peaked at the grating's frequency. These functions were taken from published physiological data in the macaque (Derrington and Lennie 1984). Modified from Poletti and Rucci (2008).

Downloaded By: [Boston University] At: 23:24 21 November 2008

correlation coefficients averaged over all pairs of cells on each axis. Neurons with receptive fields aligned parallel to the grating were on average more coactive than neurons on the orthogonal axis, indicating that a discrimination mechanism based on correlated activity was on average capable of detecting the orientation of the grating, as were the subjects in our experiments. However, correlation coefficients measured during the normal fixational motion of the retinal image were very similar to those measured in simulations of retinal stabilization. That is, the accuracy of the model in discriminating the orientation of the grating was virtually identical in normal and stabilized vision. This behavior mirrored the similar percentages of correct discrimination exhibited by subjects in the two viewing conditions of Experiment 2.

The bottom panel of Figure 10(b) shows the structure of correlated activity in the simulations of Experiment 1, when the frequency of the grating was higher than the frequency band of noise. As in the simulations of Experiment 2, levels of covariance in neuronal responses were higher on the parallel axis than on the orthogonal axis, both in the presence and in the absence of retinal image motion. Thus, on average, a discrimination mechanism based on correlated activity correctly identified the orientation of the grating. However, unlike the previous simulations, the two viewing conditions yielded different results in Experiment 1. Fixational eye movements enhanced the influence of the grating on neural activity. They increased the degree of synchronization in the responses of arrays of neurons with receptive fields aligned parallel to the grating and decreased synchronization on the orthogonal axis. Thus, consistent with our experimental results, the model was more accurate in discriminating the orientation of a grating in the presence of retinal image motion than under retinal stabilization.

The data in Figure 10 show that the degree of synchronization in the responses of ganglion cells predicts the performance of human observers in the task of discriminating between two orthogonally-oriented gratings. In keeping with psychophysical results, fixational eye movements synchronously modulated the responses of modeled ganglion cells during viewing of high-frequency gratings masked by low-frequency noise, but not during viewing of low-frequency gratings masked by high-frequency noise. These results suggest that the synchronization of retinal activity caused by fixational eye movements is an important component of the neural substrate for fine-grain spatial vision.

As mentioned in section ''Fixational eye movements and fine spatial vision'', the idea that the spatiotemporal structure of neural activity could be part of the strategy by which the visual system encodes fine spatial detail goes back to the ''dynamic theories'' of visual acuity (Averill and Weymouth 1925; Marshall and Talbot 1942; Arend 1973). These theories argued that the limiting retinal factor in visual acuity was ''the relation of receptor width to the highest optical gradient in a moving pattern rather than the static differential illumination in one cone, compared with its neighbors.'' (Marshall and Talbot 1942). More recently, this line of thought has found renewed interest in the proposal formulated by Ahissar and Arieli (2001), who examined the significance of a changing retinal image in the light of current knowledge on the response properties of neurons in the early visual system. According to this proposal, the precise timing in neuronal responses during the fixational motion of the retinal image provides a substrate for encoding the fine characteristics of the stimulus. Our experimental and theoretical results provide

Downloaded By: [Boston University] At: 23:24 21 November 2008

support for this proposal. In this regard, it should be noted that the synchronization of neuronal responses shown in Figure 10 does not exclude that the visual system might also take advantage of other aspects of the correlated structure of retinal activity besides zero-delay covariance, as hypothesized by Ahissar and Arieli (2001).

Selecting what is relevant

So far, we have worked under the assumption of prolonged visual fixation. Modeling results have referred to steady-state conditions of neuronal responses, and the frequency analysis of section "Retinal input during visual fixation" implicitly assumed a fixation of infinite duration. However, under natural viewing conditions, saccades separate brief periods of fixational instability of the order of a few hundred milliseconds. How do the results presented in this article extend to intersaccadic fixation?

In a recent modeling study, we have examined the impact of the nonstationary visual input that characterize natural viewing on the statistics of retinal activity (Desbordes and Rucci 2007). The results of this study are summarized in Figure 11. The normal alternation between large saccades and periods of minute eye movements profoundly influenced the structure of correlated activity in models of retinal ganglion cells. Data points in Figure 11 represent mean levels of correlation in the responses of pairs of M cells during the period of fixation following a saccade. As in the analysis of section "Retinal ganglion cells", levels of correlation were averaged over a large number of fixations on different images and also averaged over all pairs of simulated cells with receptive fields at a given separation. In this case, however, levels of correlation were also evaluated at several different times after the end of the preceding saccade. Therefore, the resulting pattern of correlation was

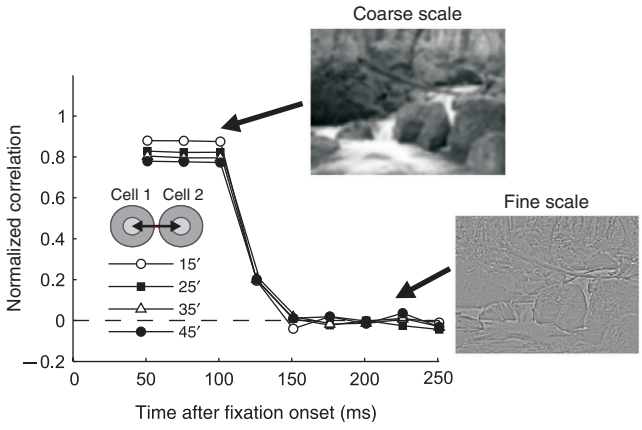


Figure 11. The normal alternation between macroscopic and microscopic eye movements might enable a temporal multiplexing of visual information, in which the same neuronal populations encode features at different spatial scales during the course of fixation. Data represent the dynamics of correlation in the responses of pairs of M cells during post-saccadic fixation. Data from pairs of cells with receptive fields at various separations are shown. Cell responses are uncorrelated 150 ms after fixation onset. Modified from Desbordes and Rucci (2007).

Downloaded By: [Boston University] At: 23:24 21 November 2008

a function of two variables: the distance between receptive field centers and the time relative to fixation onset at which correlation was evaluated. Data for pairs of cells with receptive fields at several distances are shown in Figure 11.

These data show that patterns of correlation change in a stereotypical manner during the course of visual fixation. Wide ensembles of coactive units emerge immediately after the onset of fixation and extend over several degrees of visual field, a region much larger than the size of simulated receptive fields. These neuronal ensembles shrink in size during the course of fixation, when eye movements modulate the visual input. Results similar to the patterns of correlation shown in Figure 11 were also obtained for the patterns of covariance of both P and M cells. That is, following the initial pattern of activity, the covariance between the responses of pairs of P and M cells and the correlation between the responses of pairs of M cells dropped drastically during the course of fixation. Cell responses were completely uncorrelated by the end of a typical 300 ms fixation. A highly similar dynamics occurred in simulations of neurons with receptive fields in the central region of the visual field and in the visual periphery (Desbordes and Rucci 2007).

The dynamics of correlated activity shown in Figure 11 can be explained on the basis of the results of section "Retinal ganglion cells". Because of the high speed of eye velocity during a saccade, the sudden appearance of a new stimulus at the onset of fixation has an effect similar to that resulting from flashing the stimulus without retinal image motion. Flashing of the stimulus spreads the spatial power of the image uniformly across temporal frequencies. Therefore, given that the spatial contrast sensitivities of ganglion cells do not counterbalance the spectral distribution of natural scenes (Figure 9(a)), even neurons with distant receptive fields exhibit correlated responses immediately after a saccade. Following this initial phase, the steady-state results of section "Retinal ganglion cells" apply, and neural activity tends to be more sparse.

An interesting consequence of this dynamics of neural activity is the possibility of a "temporal multiplexing" of visual information, in which the same neuronal populations encode image features at different spatial scales at different times during the course of fixation. This scheme of neural encoding is illustrated in Figure 11. In the period immediately following the onset of fixation, neurons are sensitive to the differences between the new stimulus brought in by the saccade and the stimulus of the previous fixation. For example, the correlated firing of two neurons at this stage might signal that both receptive fields moved from a dark region of the image to a brighter one. These neuronal responses are strongly driven by the broad correlations of natural images. Thus, the wide pools of coactive units present in this period convey information about the scene at a large scale. Conversely, responses at a later interval are determined by the spatiotemporal input received from the narrow region scanned during the period of fixation. At this time, the modulations of neural activity caused by fixational eye movements signal spatial changes within neighborhoods of cell receptive fields. These responses are significant when receptive fields are located in proximity to abrupt changes in contrast, such as an edge or a contour.

Thus, by bringing information from different spatial scales within the cell's temporal window of integration, eye movements enable a hierarchical representation in which the global structure of the scene is encoded before its fine details.

Downloaded By: [Boston University] At: 23:24 21 November 2008

Consistent with this observation, psychophysical evidence exists in support of both (a) a tendency to process global structures before local ones, i.e., the forest before the trees – a phenomenon known as the "global precedence effect" (Navon 1977); and (b) a coarse-to-fine integration of visual information, in which low spatial frequencies are processed before high frequencies (Parker et al. 1992; Schyns and Oliva 1994; Parker and Costen 1999). Psychophysical studies of the time-course of visual perception have not examined the fine eye movements performed by subjects. The results of our modeling studies suggest a contribution of fixational eye movements to the time-course of visual perception.

A temporal multiplexing of visual information has two important consequences. First, it implies that the flow of sensory data reaching the cortex needs to be interpreted according to the time elapsed from fixation onset. In this view, the occurrence of a saccade acts as a clock for decoding thalamic inputs to V1. This clock is necessary because coactive neurons in the retina and LGN carry different information at different times. There are several biologically-plausible mechanisms which could serve a clock function in the brain, ranging from motor efferent copies to the release of neuromodulators timed with the occurrence of saccades (Steinfels et al. 1983; Aston-Jones et al. 1991).

A second important implication of a neural multiplexing is the possibility of actively controlling the scale of visual processing. As illustrated in Figure 11, the duration of visual fixation gives control on the impact of high spatial frequencies on neural activity. In everyday life, many visual operations can be conducted efficiently at low spatial frequencies without the need for fine detail. The duration of fixation can be maintained short in these tasks, so that saccades can quickly relocate the fovea from one point of the scene to the next (Andrews and Coppola 1999). When fine spatial detail is needed, fixations can be prolonged so to exploit the uncorrelated fluctuations brought in by fixational instability. In this view, eye movements enable not only selection of the region of the scene to process, but also of the scale at which processing occurs. Consistent with this idea, longer fixations have been associated with tasks that require fine spatial vision (Pollatsek et al. 1986; Hooge and Erkelens 1998).

The early visual system is often regarded as a purely sensory processing stage. Neuronal responses are commonly linked to stimulus features without considering the observer's behavior. Yet, during natural viewing, egomotion is the most important contributor to the temporal structure of retinal activity. The body of work summarized in this article suggests that the fixational motion of the retinal image is an important component of early visual representations. Many of the predictions of the modeling studies described in these pages now await experimental validation.

Downloaded By: [Boston University] At: 23:24 21 November 2008

Acknowledgments

This work was supported by the National Institute of Health Grants EY18363 and EY015732 and by the National Science Foundation Grants BCS-0719849 and CCF-0720691. Thanks to Antonino Casile and Martina Poletti for many insightful discussions. A special thank to Francesco Saverio Rucci for countless discussions, teachings, and a lifetime of help and support.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Notes

[1] We are normally not aware of our own fixational eye movements. Several methods exist to reveal the motion of the retinal image during fixation. A striking demonstration is available on the world wide web at http://www.brl.ntt.co.jp/people/ikuya/demo/visualjitter/VisualJitter.html (Murakami and Cavanagh 1998).

[2] From Figure 7(b), one could assume that fixational instability is detrimental when $f_G < f_N$. However, the analysis of Figure 7 implicitly assumes that stimuli are presented for infinitely long periods of time. In a real experiment, temporal input modulations do not originate only from retinal image motion. The flashing of the stimulus at the onset of each trial also produces temporal power. Unlike the temporal power resulting from fixational instability, the temporal spreading of spatial power caused by stimulus onset is uniform across spatial frequency and carries information about the grating. Therefore, it may be expected that with stimuli like the one in Figure 7(b) perceptual judgments will rely more heavily on stimulus onset than on fixational modulations that do not convey useful information. In this case, the fixational motion of the retinal image would not significantly affect performance.

[3] In order to understand this property of natural images, consider the difference in the intensities of nearby pixels. This difference is approximately equal to the first spatial derivative of the image. Since the operation of differentiation corresponds to multiplication by frequency in the Fourier domain, the derivative of an image with an u^{-2} spectral density has a flat power spectrum; that is, it is spatially uncorrelated.

Downloaded By: [Boston University] At: 23:24 21 November 2008

References

Abbott LF, Varela JA, Sen K, Nelson SB. 1997. Synaptic depression and cortical gain control. Science 275:220–224.

Ahissar E, Arieli A. 2001. Figuring space by time. Neuron 32:185–201.

Alonso JM, Usrey WM, Reid RC. 1996. Precisely correlated firing in cells of the lateral geniculate nucleus. Nature 383:815–819.

Andrews TJ, Coppola DM. 1999. Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. Vision Research 39:2947–2953.

Arend LE. 1973. Spatial differential and integral operations in human vision: Implications of stabilized retinal image fading. Psychology Review 80:374–395.

Aston-Jones G, Chiang C, Alexinsky T. 1991. Discharge of noradrenergic locus coeruleus neurons in behaving rats and monkeys suggests a role in vigilance. In: Barnes CD, Pompeiano O, editors. Neurobiology of the locus coeruleus. Amsterdam: Elsevier Science Publishers. pp 501–519. Chapter 35.

Atick JJ. 1992. Could information theory provide an ecological theory of sensory processing? Network: Computation Neural Systems 3:213–251.

Atick JJ, Redlich A. 1990. Towards a theory of early visual processing. Neural Computation 2:308–320.

Atick JJ, Redlich A. 1992. What does the retina know about natural scenes? Neural Computation 4:196–210.

Attneave F. 1954. Some informational aspects of visual perception. Psychology Review 61:183–193.

Averill HI, Weymouth FW. 1925. Visual perception and the retinal mosaic II. The influence of eye movements on the displacement threshold. Journal of Comparative Psychology 5:147–176.

Bair W, O'Keefe LP. 1998. The influence of fixational eye movements on the response of neurons in area MT of the macaque. Visual Neuroscience 15:779–786.

Barlow HB. 1961. Possible principles underlying the transformation of sensory messages, Cambridge, MA: MIT Press. pp 217–234. Chapter 13.

Benardete EA, Kaplan E. 1997. The receptive field of the primate P retinal ganglion cell, I: Linear dynamics. Visual Neuroscience 14:169–185.

Benardete EA, Kaplan E. 1999. The dynamics of primate M retinal ganglion cells. Visual Neuroscience 16:355–368.

Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant J, Rust NC. 2005. Do we know what the early visual system does? Journal of Neuroscience 25:10577–10597.

Casile A, Rucci M. 2006. A theoretical analysis of the influence of fixational instability on the development of thalamocortical connectivity. Neural Computation 18:569–590.

Croner L, Kaplan E. 1995. Receptive fields of P and M ganglion cells across the primate retina. Vision Research 35:7–24.

Dan Y, Atick JJ, Reid RC. 1996. Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory. Journal of Neuroscience 16:3351–3362.

Derrington AM, Lennie P. 1984. Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. The Journal of Physiology 357:219–240.

Desbordes G, Rucci M. 2007. A model of the dynamics of retinal activity during natural visual fixation. Journal of Vision 24:217–230.

Ditchburn RW. 1955. Eye movements in relation to retinal action. Optica Acta 1:171–176.

Ditchburn RW. 1980. The function of small saccades. Vision Research 20:271–272.

Ditchburn RW, Fender DH, Mayne S. 1959. Vision with controlled movements of the retinal image. The Journal of Physiology 145:98–107.

Ditchburn RW, Ginsborg BL. 1952. Vision with a stabilized retinal image. Nature 170:36–37.

Eizenman M, Hallett P, Frecker RC. 1985. Power spectra for ocular drift and tremor. Vision Research 25:1635–1640.

Field D. 1994. What is the goal of sensory coding? Neural Computation 6:559–601.

Field DJ. 1987. Relations between the statistics of natural images and the response properties of cortical cells. Journal of Optica Society of America A 4:2379–94.

Gray CM. 1999. The temporal correlation hypothesis of visual feature integration: Still alive and well. Neuron 24:31–47.

Greschner M, Bongard M, Rujan P, Ammermuller J. 2002. Retinal ganglion cell synchronization by fixational eye movements improves feature estimation. Nature Neuroscience 5:341–347.

Gur M, Snodderly DM. 1987. Studying striate cortex neurons in behaving monkeys: Benefits of image stabilization. Vision Research 27:2081–2087.

Henning MH, Worgotter F. 2004. Eye micro-movements improve stimulus detection beyond the Nyquist limit in the peripheral retina. In: Thrun S, Saul L, Schölkopf B, editors. Advances in Neural Information Processing Systems 16. Cambridge, MA: MIT Press.

Hering E. 1899. Uber die Grenzen der Sehscharfe. berichte der Koniglichen Sachsischen Gesellshaft der Wissenschaften. Mathematisch-Physische Klasse 20:16–24.

Hooge I, Erkelens C. 1998. Adjustment of fixation duration in visual search. Vision Research 38:1295–1302.

Kapoula ZA, Robinson DA, Hain TC. 1986. Motion of the eye immediately after a saccade. Experimental Brain Research 61:386–394.

Kelly DH. 1979. Motion and vision. I. Stabilized images of stationary gratings. Journal of the Optical Society of America 69:1266–1274.

Koenderink JJ. 1972. Contrast enhancement and the negative afterimage. Journal of the Optical Society of America 62:685–689.

Kowler E, Steinman RM. 1980. Small saccades serve no useful purpose. Vision Research 20:273–276.

Leopold DA, Logothetis NK. 1998. Microsaccades differentially modulate neural activity in the striate and extrastriate visual cortex. Experimental Brain Research 123:341–345.

Marshall WH, Talbot SA. 1942. Recent evidence for neural mechanisms in vision leading to a general theory of sensory acuity. In: Kluver H, editor. Biological Symposia – Visual Mechanisms. Vol. 7. Lancaster, PA: Cattel. pp 117–164.

Martinez-Conde S, Macknik SL, Hubel DH. 2000. Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. Nature Neuroscience 3:251–258.

Martinez-Conde S, Macknik SL, Hubel DH. 2004. The role of fixational eye movements in visual perception. Nature Reviews Neuroscience 5:229–240.

Martinez-Conde S, Macknik SL, Troncoso XG, Dyar TA. 2006. Microsaccades counteract fading during fixation. Neuron 49:297–305.

Downloaded By: [Boston University] At: 23:24 21 November 2008

Murakami I, Cavanagh P. 1998. A jitter after-effect reveals motion-based stabilization of vision. Nature 395:798–801.

Navon D. 1977. Forest before trees: The precedence of global features in visual perception. Cognitive Psychology 9:353–383.

Olveczky BP, Baccus SA, Meister M. 2003. Segregation of object and background motion in the retina. Nature 423:401–408.

Parker DM, Costen NP. 1999. One extreme or the other or perhaps the golden mean? Issues of spatial resolution in face processing. Current Psychology 18:118–127.

Parker DM, Lishman JR, Hughes J. 1992. Temporal integration of spatially filtered visual images. Perception 21:147–60.

Pitkow X, Sompolinsky H, Meister M. 2007. A neural computation for visual acuity in the presence of eye movements. PLoS Biology 5:2898–2911.

Poletti M, Rucci M. 2008. Oculomotor synchronization of visual responses in modeled populations of retinal ganglion cells. Journal of Vision 8(14):4, 1–15, http://journalofvision.org/8/14/doi:10.1167/8.14.4

Pollatsek A, Rayner K, Balota D. 1986. Inferences about eye movement control from the perceptual span in reading. Perception and Psychophysics 40:123–130.

Pritchard RM. 1961. Stabilized images on the retina. Scientific America 204:72–78.

Ratliff F, Riggs L. 1950. Involuntary motions of the eye during monocular fixation. Journal of Experimental Psychology 40:687–701.

Riggs L, Ratliff F, Cornsweet J, Cornsweet T. 1953. The disappearance of steadily fixated visual test objects. Journal of the Optical Society of America 43:495–501.

Riggs LA, Ratliff F. 1952. The effects of counteracting the normal movements of the eye. Journal of the Optical Society of America 42:872–873.

Roelfsema PR, Lamme VA, Spekreijse H. 2004. Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. Nature Neuroscience 7:982–991.

Rucci M, Beck J. 2005. Effects of ISI and flash duration on the identification of briefly flashed stimuli. Spatial Vision 18:259–274.

Rucci M, Casile A. 2004. Decorrelation of neural activity during fixational instability: Possible implications for the refinement of V1 receptive fields. Visual Neuroscience 21:725–738.

Rucci M, Casile A. 2005. Fixational instability and natural image statistics: Implications for early visual representations. Network: Computation Neural Systems 16:121–138.

Rucci M, Desbordes G. 2003. Contributions of fixational eye movements to the discrimination of briefly presented stimuli. Journal of Vision 3:852–864.

Rucci M, Edelman GM, Wray J. 2000. Modeling LGN responses during free-viewing: A possible role of microscopic eye movements in the refinement of cortical orientation selectivity. Journal of Neuroscience 20:4708–4720.

Rucci M, Iovin R, Poletti M, Santini F. 2007. Miniature eye movements enhance fine spatial detail. Nature 447:851–854.

Santini F, Redner G, Iovin R, Rucci M. 2007. EyeRIS: A general-purpose system for eye movement contingent display control. Behavior Research Methods 39:350–364.

Schyns PG, Oliva A. 1994. From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. Psychology Science 5:195–200.

Shadlen MN, Movshon JA. 1999. Synchrony unbound: A critical evaluation the temporal binding hypothesis. Neuron 24:67–77.

Singer W. 1999. Time as coding space? Current Opinion Neurobiology 9:189–194.

Singer W, Gray CM. 1995. Visual feature integration and the temporal correlation hypothesis. Annual Reviews Neuroscience 18:555–586.

Skavenski AA, Hansen RM, Steinman RM, Winterson BJ. 1979. Quality of retinal image stabilization during small natural and artificial body rotations in man. Vision Research 19:675–683.

Snodderly DM, Kagan I, Gur M. 2001. Selective activation of visual cortex neurons by fixational eye movements: Implications for neural coding. Vision Neuroscience 18:259–277.

Srinivasan M, Laughlin S, Dubs A. 1982. Predictive coding: A fresh view of inhibition in the retina. Proceedings of the Royal Society of London. Series B: Biological Sciences 216:427–459.

Steinfels GF, Heym J, Strecker RE, Jacobs BL. 1983. Behavioral correlates of dopaminergic unit activity in freely moving cats. Brain Research 258:217–228.

Steinman RM, Collewijn H. 1980. Binocular retinal image motion during active head rotation. Vision Research 20:415–429.

Downloaded By: [Boston University] At: 23:24 21 November 2008

Steinman RM, Cunitz RJ, Timberlake GT, Herman M. 1967. Voluntary control of microsaccades during maintained monocular fixation. Science 155:1577–1579.

Steinman RM, Haddad GM, Skavenski AA, Wyman D. 1973. Miniature eye movement. Science 181:810–819.

Steinman RM, Levinson JZ. 1990. The role of eye movements in the detection of contrast and spatial detail. In: Kowler E, editor. Eye movements and their role in visual and cognitive processes. Amsterdam: Elsevier Science Publishers BV. pp 115–212. Chapter 3.

Tulunay-Keesey U. 1982. Fading of stabilized retinal images. Journal of the Optical Society of America 72:440–447.

Tulunay-Keesey U, Jones RM. 1976. The effect of micromovements of the eye and exposure duration on contrast sensitivity. Vision Research 16:481–488.

Usrey WM, Alonso JM, Reid RC. 2000. Synaptic interactions between thalamic inputs to simple cells in cat visual cortex. Journal of Neuroscience 20:5461–5467.

Usrey WM, Reid RC. 1999. Synchronous activity in the visual system. Annual Review of Physiology 61:435–456.

van Hateren JH. 1992. A theory of maximizing sensory information. Biological Cybernetics 68:23–29.

Yarbus AL. 1967. Eye movements and vision. English Translation. New York: Plenum Press.

Downloaded By: [Boston University] At: 23:24 21 November 2008