

Information, movement, and adaptation in human vision

Alexander J. H. Houston^{1,*}, David H. Brainard², Hannah E. Smithson³, and Daniel J. Read^{4,†}¹*School of Mathematics and Statistics, University of Glasgow, University Place, Glasgow G12 8QQ, United Kingdom*²*Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6241, USA*³*Department of Experimental Psychology, University of Oxford, Oxford OX2 6GG, United Kingdom*⁴*School of Mathematics, University of Leeds, Leeds LS2 9JT, United Kingdom*

(Received 3 September 2025; accepted 18 February 2026; published 28 April 2026)

Our eyes are never still. Even when fixating, they exhibit small, jittery motions. While it has long been argued that these fixational eye movements (FEMs) aid the acquisition of visual information, a complete theoretical description of their impact on the information available in the early visual system has been lacking. Here, we consider the drift component of FEMs. We describe these FEMs as a diffusive process and build a minimal theoretical model of the early visual response, including the critical process of adaptation (a fading response to a fixed image). We establish the effect of FEMs on the mutual information between a visual stimulus and this response. Our approach identifies the key dimensionless parameters that characterize the effect of fixational eye movements and reveals the regimes in which this effect can be beneficial, detrimental, or negligible. Taking parameter values appropriate for human vision, we show that the spatiotemporal couplings induced by fixational eye movements can explain the qualitative features of the human contrast sensitivity function as well as classic experiments on temporal integration. To our knowledge, the consideration of fixational eye movements in this latter context is unique, and our results suggest the need for future experiments to determine the mechanisms by which spatial and temporal responses are coupled in the human visual system.

DOI: [10.1103/l4qd-5s7d](https://doi.org/10.1103/l4qd-5s7d)

I. INTRODUCTION

Human vision is a dynamic process, and our eyes are in constant motion [1]. At the largest scales, as we look around a scene, jumplike movements of the eye called saccades locate the images of targets of interest on the high-resolution fovea in the center of the retina [2]. In between saccades, we fixate on a target, and yet our eyes still move. These fixational eye movements (FEMs) comprise occasional small jumps of order 10 arcmin known as microsaccades, but for the majority of the time the eye undergoes a seemingly random, diffusionlike motion referred to as drift [1,3,4]. Superimposed on drift is a small oscillatory motion known as tremor that can be difficult to measure [5]. In this paper, we focus on the drift that occurs in approximately 0.5–1 s intervals [3,6,7] between microsaccades. We neglect tremor as recent measurements made using high-resolution retinal tracking indicate that it is small relative to the size of cone entrance apertures [8]. In the remainder of this paper, we will use the term fixational eye movements to refer to the drift component of fixational eye movements.

That the eye performs fixational movements has been known for over 250 years, thanks to the observations of Jurin [9], Darwin and Darwin [10], and Helmholtz [11], who made reference to the “wandering of the gaze.” As to why these movements occur, a pessimistic view is that FEMs are merely involuntary motions resulting from imperfect motor control of the eye that simply blur or degrade an image [12]. However, this view seems unlikely to provide a full explanation, given the many aspects of vision that have evolved to achieve optimal performance given some physical constraint [13,14], such as single photon detection [15–17], the diffraction-limited construction of insect eyes [18], discussed by Feynman *et al.* [19] and McLeish [20], and the use of Cartesian ovals in Trilobite eyes to correct aberration [21]. Indeed, the idea that FEMs may be beneficial dates back to 1899 [22] and there has been growing empirical evidence that they can improve visual acuity [23–25]. Further, it has been suggested that the magnitude of drift can be tuned [26–28], such that the view of FEMs has developed from that of an aimless “wandering eye” to a more purposeful, controlled “active vision” [29,30].

How, then, might FEMs be beneficial, how should this benefit be optimized, and in what fashion does it depend on the features of the visual stimulus? The visual system is designed to detect changes. So much so that, were the environment completely static and our eyes completely still, we would not see anything at all. This visual adaptation, which functions as a bandpass temporal response, has been well established [31,32] and opens the door for FEMs to beneficially refresh the image by moving the stimulus over the retina. However, the effect of FEMs is not uniform for all spatial frequencies

*Contact author: Alexander.Houston@glasgow.ac.uk

†Contact author: D.J.Read@leeds.ac.uk

with an image. Rather, they cause the spatial structure of a static image to be converted into spatiotemporal structure in a manner that depends on the spatial frequency [1,33]. Accordingly, it has been hypothesized that FEMs serve not simply to overcome adaptation by refreshing the image, but to encode visual information within a spatiotemporal structure to which subsequent processing stages are most sensitive [22,34–38]. Some early psychophysical experiments [39] did not support this view, however, and there remains debate as to the degree to which human visual performance can be understood with or without considering FEMs [30,40].

To address the issue of how FEMs might be expected to affect visual performance, here we develop an analytically tractable model of the effect of drift on the information about a stimulus available in the early visual system. The model exposes basic principles of how stimulus properties (spatial frequency, temporal frequency, duration), spatial and temporal filtering in the visual pathways, and properties of fixational eye movements (rate of drift) interact to shape the limits of visual performance. Related to this, the model derives key dimensionless quantities that capture aspects of these interactions and allows quantitative prediction of how visual performance depends on both the stimulus and key components of the neural processing chain. One key implication of the model is that across a range of stimulus parameters (e.g., duration and spatial frequency), it is possible to find regimes in which FEMs enable increased performance and other regimes in which they are deleterious. The existence of both types of regimes illustrates the richness of the interplay between FEMs and the visual information available for perceptual decisions, and clarifies why effects of both signs can potentially be observed in experimental data.

The thread that runs through this paper is that spatial variation in a stimulus combined with retinal motion produces a time-dependent visual signal. Although simple, we find this principle to have great expository power. We make it quantitative in Sec. II by constructing a minimal model for the early stages of the visual system, allowing theoretical investigation of the influence of ocular drift on information acquired about an external stimulus.

In Sec. III A, we use the model to determine the optimal form of diffusive eye motion to gain information from a given spatial wave vector. Taking this further, we establish the combinations of stimulus presentation time and spatial frequency for which ocular drift is beneficial, detrimental, or has negligible effect on sensitivity to individual spatial wave vectors. Although we do not attempt a detailed account of specific experiments on human contrast spatial sensitivity, by employing reasonable numerical estimates of properties of the human visual system we find that measured rates of fixational ocular drift are tuned so as to provide maximum benefit given the presence of optical blur—an indication that once again a biological process is operating near a limit imposed by physical principles.

In Sec. III B, we incorporate temporal modulation of the stimulus. Here, we find that the fact that eye movements convert spatial variations into temporal ones, in conjunction with temporal adaptation and blur, is able to account qualitatively for the features of the spatiotemporal contrast sensitivity function without the need to invoke a bandpass spatial filtering

stage. This is not to imply that such stages do not exist nor that a quantitative account can be achieved without incorporating them, but rather to again illustrate the range of effects that can be understood when the interactions between FEMs and visual system filtering are jointly considered in a systematic framework.

We consider a second example in Sec. IV, the extent to which the experiments of Barlow [41] support the notion that the temporal adaptation of the human visual system is dependent on the scale of the stimulus, as has been understood previously. By using a theoretical model of detection [42], we again find that the modulation of information caused by ocular drift is sufficient to explain the observations, without incorporating additional layers of postreceptoral processing. A given form of eye motion will induce temporal signal variations at a frequency dependent on the spatial scale of the stimulus, and this results in an effective scale-dependent adaptation that allows the features of the experimental observations to be reproduced. As with our analysis of spatiotemporal contrast sensitivity, our point is not that the inferred effects are necessarily absent, but rather that caution is required before the inferences are secure. Our analysis in this case suggests the structure of future experiments that might be decisive.

We close with a brief Discussion.

II. INFORMATION IN THE EARLY VISUAL SYSTEM

A. Minimal model of the early visual response

Our aim in this paper is to understand the role that fixational eye movements play in determining the visual information available at the retina. Our approach is to employ a minimal model of the early visual system. We take the position that much of the impact of FEMs stems from their coupling of spatial and temporal properties, specifically the spatial structure of the stimulus and a bandpass temporal response. Therefore, while we incorporate a temporal adaptation filter, we neglect any additional aspects of postretinal signal processing, although these could be incorporated into our model, as discussed in Sec. V. To further simplify the theoretical description, we treat the photoreceptor array as a homogeneous continuum of constant photoreceptor density, neglecting that photoreceptors are discrete objects and also the preferential response of L, M, and S cones to long, medium, or short wavelengths. We consider only contrast stimuli, to which we assume all photoreceptors have an identical response. While the bandpass temporal response does not occur in the photoreceptors, in the interest of concision, we shall refer to our minimal model of the response of the early visual system as the “photoreceptor response.”

The key components of our model of photoreceptor response are laid out in Fig. 1. There are three ways in which time dependence of the photoreceptor signal, y , can be induced: temporal modulation of the external stimulus, x , retinal motion, and adaptation. The first of these is described by an amplitude $A(t')$, with the example shown in Fig. 1(b) corresponding to a stationary external stimulus presented with constant amplitude for a time T . Note that we take this modulation to be uniform across the stimulus, a common experimental protocol, such that the spatial and temporal vari-

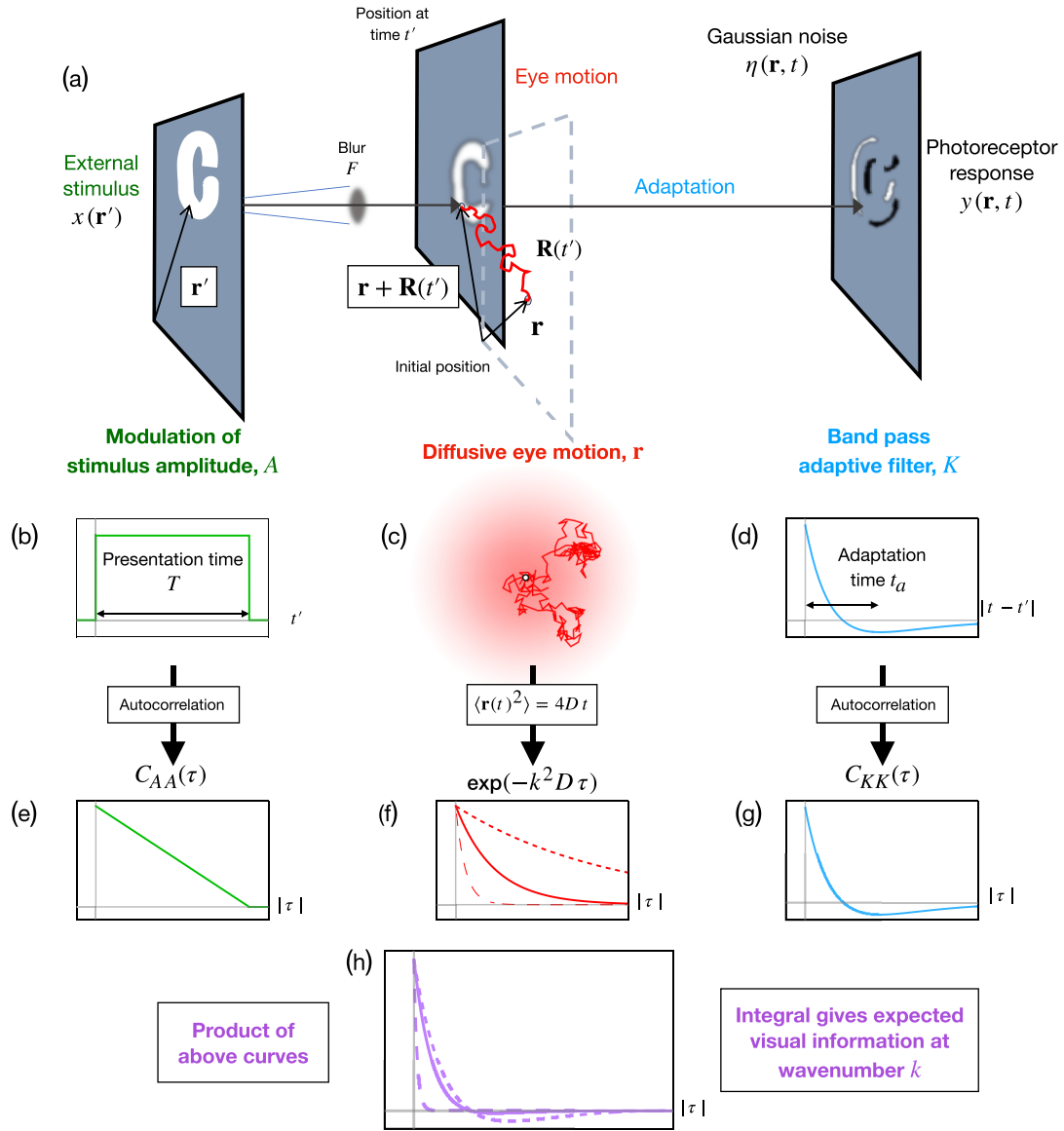


FIG. 1. A schematic of our minimal model for the visual information acquired from a stimulus. (a) An external stimulus is related to a photoreceptor response through optical blur, eye motion, adaptation, and Gaussian noise. Setting aside blur for the moment, in our minimal model the central elements of the stimulus, eye motion, and retinal physiology are captured by the stimulus amplitude, diffusive trajectories, and a linear adaptive response, which are shown in panels (b), (c), and (d), respectively. (Although we focus on diffusive eye motion in the paper, our model allows more general motion to be incorporated.) In particular, it follows from a calculation presented in the Supplemental Material [49] that the aspects of the stimulus, eye motion, and adaptive response that are relevant for the acquisition of visual information are the functions shown in panels (e), (f), and (g), where C_{AA} and C_{KK} denote the autocorrelations of the stimulus and adaptive filter and $4Dt$ is the mean squared displacement of the retina, with D the diffusion constant. The solid, small dashed and large dashed curves correspond to a medium, small, and large value of $k^2 D$, where k is a spatial wave number of the stimulus. The product of the functions shown in panels (e)–(g) produces the curves shown in panel (h), the integral of which gives the visual information acquired from the stimulus at wave number k , up to a given time, modulo additional effects due to blur as discussed in the main text. The different curves in panel (h) result from the corresponding curves in panel (f).

ation separate and hence appear as the product $x(\mathbf{r}')A(t')$ in Eq. (1). As a convention, we take $|A(t')| \leq 1$, with the overall stimulus amplitude incorporated into $x(\mathbf{r}')$.

Second, retinal motion will convert spatial variation of the stimulus into temporal variation of the photoreceptor response. A photoreceptor initially located at $\mathbf{r}$ will, at some later time t' , have moved to a new position $\mathbf{r} + \mathbf{R}(t')$ via

retinal motion encoded by $\mathbf{R}(t')$. Due to the optical blur of the eye's lens, and because of its finite size, a photoreceptor will collect photons originating from a distribution of positions within the stimulus. We represent this distribution by a blurring function F centered on the current photoreceptor position, leading to the convolution between the stimulus and blurring function shown in Eq. (1). In what follows, we

assume that the path $\mathbf{R}(t')$ taken by the eye is “known” to the neural processors, either via mechanical or electrical signals or by inference from a wider set of visual cues, and hence this motion is not itself a source of noise. The inference of motion is itself an interesting area of study [43,44]. Here, we simply assume that there are sufficient data available for this inference to be accurately made.

Finally, the temporal response of photoreceptors (and early-stage signal processing in the retina) is not instantaneous. These early components of the visual signaling pathway integrate an incoming signal over some past time to accumulate a response [42] and adapt to steady stimuli such that a constant stimulus will eventually produce a zero, or near zero response. This can be modeled by a linear temporal response function, or memory kernel, $K(t - t')$ [42]. The present photoreceptor response at time t corresponds to integration over the past signal from earlier times t' , weighted by the memory kernel, as can be seen below in Eq. (1). Typically, the memory kernel is considered to be positive for small $t - t'$, to model the accumulation of input signal at short times, before becoming negative for larger $t - t'$ to model the adaptation process. For perfect adaptation, that is, zero response to a constant signal in the long time limit, the integral of K should be zero. These features are apparent in the example shown in Fig. 1(d), which takes K to be biexponential. We shall use this form of memory kernel throughout this work and will describe it more fully shortly.

The preceding elements, with the inclusion of additive noise, combine to give the following minimal model for the photoreceptor array response:

$$y(\mathbf{r}, t) = \int_0^t K(t - t') \int F(\mathbf{r} + \mathbf{R}(t') - \mathbf{r}') x(\mathbf{r}') \times A(t') d^2 \mathbf{r}' dt' + \eta(\mathbf{r}, t), \quad (1)$$

as has been used previously in the context of motion estimation [43]. We reiterate that Eq. (1) provides the response at time t of a photoreceptor initially located at $\mathbf{r}$. We take the noise $\eta(\mathbf{r}, t)$ to be Gaussian white noise, i.e., uncorrelated between different points in space or time. One could also include noise on the stimulus $x(\mathbf{r}')$, for example, due to photon shot noise, which becomes a significant factor at low light levels [45–48], or noise in the early stages of photoreceptor excitations. Such noise would be passed through some of the temporal and spatial filtering represented by the functions K and F , resulting in colored noise at the final signal. We neglect this for simplicity.

B. Quantifying visual information

We are now in a position to consider the rate at which visual information can be gained about the external stimulus x from the photoreceptor response y . This process is hindered by noise, adaptation, and optical blur. A central question, which our model allows us to address, is to what extent eye movements mitigate these factors. It is convenient to take the spatial Fourier transform of Eq. (1) and to consider the photoreceptor array response at wave vector $\mathbf{k}$ (i.e., with a spatial frequency $|\mathbf{k}|/2\pi$). Denoting the Fourier transform by $\hat{\bullet}$, this gives $\hat{y}_\mathbf{k}(t) = \hat{F}_\mathbf{k} \hat{x}_\mathbf{k} b_\mathbf{k}(t) + \hat{\eta}_\mathbf{k}(t)$, where

$b_\mathbf{k}(t) = \int_0^t dt' K(t - t') e^{i\mathbf{k} \cdot \mathbf{R}(t')} A(t')$. Maximizing the information transmitted at wave vector $\mathbf{k}$ corresponds to making $b_\mathbf{k}(t)$, in some sense, as large as possible. We have found that the appropriate measure of the magnitude of $b_\mathbf{k}(t)$ is the quantity

$$\psi_\mathbf{k} = \int_0^\infty |b_\mathbf{k}(t)|^2 dt. \quad (2)$$

Making $\psi_\mathbf{k}$ large corresponds to making the signal at wave vector $\mathbf{k}$ large, in a time integrated sense. Once the other components of the visual system have been specified, the only remaining optimization problem is how to maximize this quantity via eye movements. In Sec. I of the Supplemental Material [49], we analyze multiple different ways of quantifying the process of learning about external stimulus x from the photoreceptor response y , and in each case demonstrate that $\psi_\mathbf{k}$ emerges as the correct quantity to consider. We show that maximizing $\psi_\mathbf{k}$ minimizes the mean square error when estimating the stimulus $\hat{x}_\mathbf{k}$ from the response $\hat{y}_\mathbf{k}(t)$. A robust and often-used measure is the mutual information: Assuming a Gaussian prior distribution of the stimulus $\hat{x}_\mathbf{k}$, we show that maximizing $\psi_\mathbf{k}$ gives maximal mutual information between stimulus x and photoreceptor response y at wave vector $\mathbf{k}$. We finally show that this same quantity $\psi_\mathbf{k}$ emerges from an existing model of signal detection [42], which we will employ in Sec. IV. Note that in all cases that the information transmitted depends on the product $\psi_\mathbf{k} |\hat{F}_\mathbf{k}|^2$, and in this way depends also on the optical blur. This simple dependence on the blur function arises because we assume F does not have a temporal component (i.e., there is a factorization into separate functions governing spatial (F) and temporal (K) response).

The value of $\psi_\mathbf{k}$ depends on the specific path $\mathbf{R}(t')$ of the eye motion; i.e., it takes a different value for different paths. In the present work, we focus on its average $\langle \psi_\mathbf{k} \rangle$ across multiple paths, reasoning that forms of eye motion that maximize the average are likely to be beneficial. Here and throughout, we use $\langle \bullet \rangle$ to denote an average over the ensemble of possible paths taken by the eye. Under the assumption that the eye movements are a statistically stationary process, meaning that there is no special reference time and so the statistical properties of the motion do not vary over time, we show in Sec. II A of the Supplemental Material [49] that the average value of $\psi_\mathbf{k}$ can be conveniently written as

$$\chi_\mathbf{k} = \frac{\langle \psi_\mathbf{k} \rangle}{t_a} = \frac{2}{t_a} \int_0^\infty C_{AA}(\tau) C_{KK}(\tau) S_\mathbf{k}(\tau) d\tau, \quad (3)$$

where C_{ff} denotes the autocorrelation of the function f ,

$$C_{ff}(\tau) = \int_{-\infty}^\infty f(t + \tau) f(t) dt, \quad (4)$$

$S_\mathbf{k}(\tau) = \langle \exp[i\mathbf{k} \cdot (\mathbf{R}(t + \tau) - \mathbf{R}(t))] \rangle$ and t_a is an adaptation time that we define in the following paragraph. The three terms under the integral for $\chi_\mathbf{k}$ in Eq. (3) capture the effect on information gain of the three sources of time dependence in the photoreceptor response, these being temporal modulation of the stimulus, adaptation, and eye movements, as discussed previously. Typical forms of these three terms are shown in Fig. 1, and we now describe each in turn.

We begin with the contribution to the retinal information due to the memory kernel. Since, in our minimal model, the

memory kernel captures the retinal physiology, it provides the relevant timescale by which to normalize all others. As mentioned in Sec. II A, throughout this work we take the memory kernel to be biexponential, as shown in Fig. 1(d). It is composed of an excitatory response on a short timescale t_s and an inhibitory one on a long timescale t_l . From these, we define two auxiliary parameters: an “adaptation time” $t_a = \sqrt{t_s t_l}$ and a “shape parameter” $\sigma = \sqrt{t_l/t_s}$. The adaptation time sets the characteristic timescale for the memory kernel and it is useful to use it to nondimensionalize other times in the system, which we shall denote with a tilde, for example, $\tilde{t} = t/t_a$. The shape parameter encodes the separation between short and long time responses of the temporal filter (so, $t_s = t_a/\sigma$ and $t_l = t_a\sigma$). With this choice of parameters, the memory kernel is rendered pleasingly symmetric

$$K(\tilde{t}) = \frac{1}{t_a} \left[\sigma e^{-\sigma \tilde{t}} - \frac{1}{\sigma} e^{-\tilde{t}/\sigma} \right], \quad (5)$$

for $\tilde{t} \geq 0$, with $K(\tilde{t}) = 0$ for $\tilde{t} < 0$. As presented in Eq. (5), the memory function exhibits perfect adaptation; the excitatory and inhibitory components are balanced such that at long times the response to a constant stimulus will vanish. Imperfect adaptation can be incorporated by giving the two components different weightings. We discuss the consequences of this in Sec. IV of the Supplemental Material [49]. For a biexponential memory kernel, the autocorrelation, as calculated via Eq. (4), produces a scaled version of the original function

$$C_{KK}(\tilde{\tau}) = \frac{1}{2} \frac{\sigma^2 - 1}{\sigma^2 + 1} K(|\tilde{\tau}|), \quad (6)$$

as illustrated in Fig. 1(g). The prefactor of C_{KK} reflects that when $\sigma = 1$ the two timescales are equal and the memory kernel vanishes.

Some previous studies have used a higher-order memory kernel as results from successive application of exponential filters [42,50]. This can result in a markedly different memory kernel for which $K(0) = 0$. We note, however, that this difference is suppressed in the autocorrelation function, which remains qualitatively similar in profile to the biexponential in Eq. (6), as illustrated in Sec. III of the Supplemental Material [49] for a memory kernel of the form given by Watson [42]. In particular, the autocorrelation function in both cases is maximal at the origin and has a single minimum. Consequently, none of the results presented in this paper are qualitatively changed by using a higher-order memory kernel.

The contribution of the stimulus temporal modulation to the retinal information is also captured via its autocorrelation function, again calculated via Eq. (4). For a constant amplitude stimulus presented for time T , we take $A(t)$ to be 1 when $0 \leq t \leq T$ and zero otherwise. For nondimensional time $\tilde{t} = t/t_a$ and presentation time $\tilde{T} = T/t_a$, this gives $A(\tilde{t}) = 1$ when $0 \leq \tilde{t} \leq \tilde{T}$ and zero otherwise, as shown in Fig. 1(b). The resulting autocorrelation function is

$$C_{AA}(\tilde{\tau}) = \begin{cases} t_a(\tilde{T} - |\tilde{\tau}|), & 0 \leq |\tilde{\tau}| \leq \tilde{T}, \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

which has the decreasing triangular form shown in Fig. 1(e). The autocorrelation function for sinusoidal stimulus modulation is shown in Sec. V of the Supplemental Material [49].

We provide motivation for these particular forms of temporal modulation at the start of Sec. III.

The impact of retinal motion on gaining visual information is captured by $S_k(\tau) = \langle e^{ik \cdot (\mathbf{R}(t+\tau) - \mathbf{R}(t))} \rangle$. This path “structure factor” involves summing the signal amplitudes due to different retinal positions, averaging over all paths that might occur and averaging over time, t . We emphasize that this expression for S_k is valid for all forms of eye movement that are statistically stationary (i.e., whose statistical properties are independent of time). In the present work, following earlier studies [1], we treat the eye’s motion as diffusive [Fig. 1(c)]. In reality, FEMs are probably not purely diffusive: Other studies have revealed a tendency for the eye to drift toward a preferred retinal location [51,52] and have detected a tendency toward short time persistence and longer time antipersistence [53,54]. Nevertheless, we consider diffusive motion to be an appropriate minimal model that allows us to elucidate key features of the interaction between FEMs and other components of visual processing, while retaining analytical simplicity.

The assumption of diffusive motion means that the probability distribution of motion vectors is zero-mean Gaussian, so that the path structure factor is specified wholly by the mean squared displacement (MSD) as $S_k = e^{-\frac{k^2}{4} \text{MSD}(\tau)}$. Diffusive motion is additionally constrained so that MSD is proportional to time, $\text{MSD}(\tau) = 4D\tau$, with D the diffusion constant. Hence, the expression used throughout this paper is

$$S_k = e^{-\Gamma |\tilde{\tau}|}, \quad (8)$$

where we define the dimensionless parameter $\Gamma = k^2 \text{MSD}(t_a)/4 = k^2 D t_a$, the square of the spatial frequency, made dimensionless by the diffusion constant and adaptation time. At fixed diffusion constant D , increasing Γ corresponds to increasing the square of the spatial frequency k ; at fixed spatial frequency, increasing Γ corresponds to increasing the diffusion constant.

The logic behind the exponential decay, shown in Fig. 1(f), is then clear. The dimensionless number Γ quantifies the typical magnitude of diffusive motion relative to the wave number k during the adaptation time t_a , and hence determines the rapidity of the temporal variation of the photoreceptor signal induced by diffusive eye motion. The faster this variation, the smaller the amplitude that is produced upon averaging over time $\tilde{\tau}$, giving the exponential decay. We illustrate this in Fig. 1(f) by showing $S_k(\tau)$ for small Γ (small dash), intermediate Γ (solid line), and large Γ (large dash).

We have now introduced all components of our minimal model. For ease of reference, in Table I we summarize the key elements of our notation.

As stated earlier, the three aspects of adaptation, modulation, and motion combine to produce the gained information. The product of the curves in Figs. 1(e)–1(g) gives the three curves in Fig. 1(h). Following Eq. (3), the integral of the curves in Fig. 1(h) gives the information gain as encoded in χ_k . Different aspects of the system dominate for different parameter choices. For example, when Γ is large the exponential decay of S_k in Fig. 1(f) is much faster than that the memory kernel autocorrelation in Fig. 1(g), and so S_k is almost perfectly reproduced in the large Γ (large dash) curve in Fig. 1(h). Conversely, when Γ is small, the slow

TABLE I. A summary of the key notation used in this paper.

Notation	Meaning
t_a	<i>Adaptation time:</i> $\sqrt{t_s t_l}$, with t_s and t_l the short and long timescales of the memory kernel
σ	<i>Shape parameter:</i> $\sqrt{t_l/t_s}$, with t_s and t_l the short and long timescales of the memory kernel
$\tilde{T}$	<i>Dimensionless presentation time:</i> T/t_a , with T the presentation time
Γ	$k^2 D t_a$, with $2\pi/k$ a signal wavelength and D the diffusion constant of the retinal motion
χ	<i>Measure of retinal information:</i> defined in Eq. (3) and surrounding text

decay of S_k in Fig. 1(f) means that the corresponding small Γ (small dash) curve in Fig. 1(h) bears the imprint of the memory autocorrelation function. For all regimes, the finite presentation time, as encoded by the autocorrelation function in Fig. 1(e), provides a cutoff for the curves in Fig. 1(h). A particular case worth highlighting is that of very short presentation times (i.e., shorter than both the excitatory response of the memory kernel and the exponential decay of S_k). Then, for a constant amplitude stimulus, we can approximate the gained information by

$$\chi_k \approx \frac{1}{t_a} C_{KK}(0) \tilde{T}^2. \quad (9)$$

This quadratic dependence on presentation time will be relevant to our discussion of detection thresholds in Sec. IV.

In the remainder of this paper, we first use the above model to consider contrast sensitivity at a specified wave vector for both constant and modulated stimuli. Then, in Sec. IV we provide a more holistic view of detection thresholds, incorporating the sum over wave vectors present in most realistic stimuli. At certain points, we present the consequences of our model with parameters appropriate for human vision. For the adaptation time and shape parameter, we take $t_a = 40$ ms and $\sigma = \sqrt{2}$, based on achieving approximate agreement between the autocorrelation of our memory kernel and that of a more involved one that has been shown to accord with the human visual system [42]; see Sec. III of the Supplemental Material [49] for details. There is variability in the values given in the literature for the diffusion constant, and care is needed to ensure that microsaccades and head movements are not incorporated, but values of 0.01–0.04 arcmin²/ms have been reported [25,55]. Based upon this, we take $D = 0.02$ arcmin²/ms unless otherwise stated. Finally, we take blurring function F to have the form $|\hat{F}_k| \sim e^{-\frac{1}{2}\lambda^2 k^2}$, with $\lambda \approx 1$ arcmin [56,57].

III. EFFECTS OF RETINAL MOTION ON CONTRAST SENSITIVITY

With all the components in place, we now calculate explicitly the information gained from a stimulus presented at a particular wave vector, in terms of χ_k . The results of this section can also be viewed in terms of the contrast sensitivity as a function of the spatial frequency, $|\mathbf{k}|/2\pi$, of a stimulus. The contrast sensitivity is proportional to $|\hat{F}_k|/\sqrt{\chi_k}$, with $\hat{F}_k$ the Fourier transform of the blurring function. Therefore, $\sqrt{\chi_k}$ gives the contributions to contrast sensitivity from eye movement and signal modulation when passed through the temporal response. The contrast sensitivity function is a quintessential measure of the system-level performance of the

visual system [58–62]. Although the classic experiments of Kelly [61] attempted to stabilize stimuli on the retina, it is now widely accepted that FEMs contribute to the contrast sensitivity function, and there have been a number of computational works investigating the role they play in this regard [1,33,63].

A widely used experimental approach to characterizing vision has been to probe it with simple, well-controlled stimuli, for which key parameters can be varied to provide information about how fundamental properties such as the spatial and temporal scale of the stimuli affect performance. Chief among such stimuli are those with constant amplitude and those with sinusoidal amplitude modulation. Considering these forms of stimuli is especially natural in the context of our theory since, having modeled the visual response as stationary and linear, the sinusoids form an eigenbasis for the system. Doing so amounts to taking the Fourier transform of the system response, as is employed near-ubiquitously in signal processing. We therefore consider first stimuli with constant amplitude, before moving onto sinusoidally modulated stimuli, for which the former is of course a limiting case.

A. Constant amplitude

Taking our generic expression for χ_k (3), substituting the particular forms of C_{KK} , C_{AA} , and S_k appropriate to a biexponential memory kernel, constant amplitude, and diffusive retinal motion, and nondimensionalizing yields

$$\begin{aligned} \chi_k &= \frac{\sigma^2 - 1}{\sigma^2 + 1} \int_0^{\tilde{T}} (\tilde{T} - \tilde{\tau}) \left[\sigma e^{-\sigma \tilde{\tau}} - \frac{1}{\sigma} e^{-\tilde{\tau}/\sigma} \right] e^{-\Gamma \tilde{\tau}} d\tilde{\tau} \\ &= \frac{\sigma^2 - 1}{\sigma^2 + 1} \left[\left(\frac{\sigma}{\Gamma + \sigma} - \frac{1}{1 + \Gamma\sigma} \right) \tilde{T} \right. \\ &\quad \left. + \frac{\sigma}{(\Gamma + \sigma)^2} (e^{-(\Gamma + \sigma)\tilde{T}} - 1) \right. \\ &\quad \left. - \frac{\sigma}{(1 + \Gamma\sigma)^2} (e^{-(\Gamma + \frac{1}{\sigma})\tilde{T}} - 1) \right]. \end{aligned} \quad (10)$$

This is plotted as a function of Γ in Fig. 2(a) for $\sigma = \sqrt{2}$ and a range of stimulus presentation times $\tilde{T}$, while in Fig. 2(b) we show χ_k as a function of $\tilde{T}$ for several choices of Γ . There are several regimes of behavior, which we now delineate. Figure 3(a) is a phase diagram indicating the values of Γ and $\tilde{T}$ where each regime of behavior applies. In what follows, it helps to define two constants that depend only upon the shape of the temporal response, parametrized via σ :

$$S_1 = \frac{(\sigma^2 - 1)^2}{\sigma(\sigma^2 + 1)}, \quad (11)$$

$$S_2 = \frac{\sigma^2 - 1}{\sigma^2 + 1}. \quad (12)$$

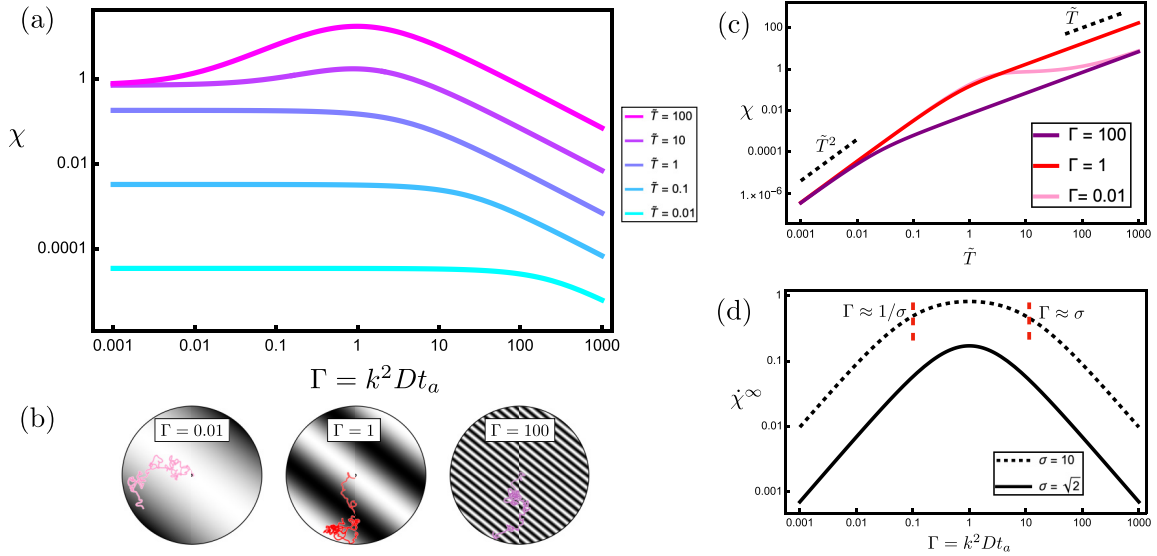


FIG. 2. The retinal information gained from a constant amplitude signal with diffusive ocular drift. (a) The information gained is captured by χ , given in general in Eq. (3) and in the particular case of a constant amplitude stimulus in Eq. (10). This is shown as a function of Γ , that is, how far the eye diffuses in the adaptation time, measured in the wavelength of the signal ($2\pi/k$). The stimulus is presented for a time $T = \tilde{T}t_a$, with $\tilde{T}$ varying from 0.01 (cyan) to 100 (magenta) in powers of 10. The shape parameter, defined in terms of the long and short timescales of the memory function (t_l and t_s) as $\sqrt{t_l/t_s}$, is $\sigma = \sqrt{2}$. (b) Schematic illustrations of different values of Γ , showing a gray scale stimulus with the path taken by the retina within the adaptation time. We emphasize that there is no absolute scale associated with these plots, and an increase in $\Gamma = k^2 D t_a$ may result from either increasing the spatial frequency in the stimulus, k , or increasing the diffusion constant, D . (c) The information gained as a function of $\tilde{T}$ for three choices of Γ , highlighting the different power laws that arise. (d) The asymptotic rate of information gain, $\dot{\chi}^\infty$. The solid curve corresponds to the same adaptive memory kernel as in panel (a). The dashed curve illustrates the effect of a larger shape parameter $\sigma = \sqrt{t_l/t_s}$, corresponding to a greater separation between the timescales of excitation and inhibition.

We note that for a stimulus at wave vector $\mathbf{k}$, diffusive eye motion produces a temporal signal variation on timescale $(k^2 D)^{-1}$. Comparing this with the presentation time T , we distinguish long presentation times ($T > (k^2 D)^{-1}$, so that $\tilde{T} > 1/\Gamma$) in which there is significant retinal motion over the spatial scale of the signal during stimulus presentation, and short presentation times ($\tilde{T} < 1/\Gamma$) when retinal motion is small on the spatial scale of the signal during stimulus presentation.

For long presentation times, $\tilde{T} > 1/\Gamma$, the first term of Eq. (10) dominates and the information gain grows linearly with time, which is the behavior seen in the upper right part of Fig. 2(a) for large $\tilde{T}$. Schematic illustrations of different Γ values are shown in Fig. 2(b), and the corresponding time dependence of $\chi_{\mathbf{k}}$ is shown in Fig. 2(c), with all curves exhibiting linear growth at large times. The resultant rate of information gain $\dot{\chi}^\infty = \lim_{\tilde{T} \rightarrow \infty} (\chi_{\mathbf{k}}/\tilde{T})$ is shown in Fig. 2(d), giving the shape of the peak in $\chi_{\mathbf{k}}$ that emerges at large presentation times. This depends on the extent to which the eye motion can drive information through the adaptive temporal response function. To assess this, we compare the signal variation timescale $(k^2 D)^{-1}$ with the two timescales in the memory kernel for retinal response, t_l and t_s . This gives rise to three regimes of behavior: (1) The temporal signal variation is slower than the inhibitory component of retinal response ($k^2 D t_l \ll 1$ so that $\Gamma \ll 1/\sigma$), (2) temporal variation is slower than the excitatory response but faster than the inhibitory component ($k^2 D t_l \gg 1$ but $k^2 D t_s \ll 1$ so that $1/\sigma \ll \Gamma \ll \sigma$), or (3) temporal variation is faster than both excitatory and

inhibitory responses ($k^2 D t_s \ll 1$ so that $\Gamma \gg \sigma$). We briefly describe the rate of information gain in each of these regimes.

When $\Gamma \ll 1/\sigma$, the signal variation at a receptor due to eye motion is slower than any component of the temporal response function. In the absence of any eye movement, when $\Gamma = 0$, adaptation is perfect and there is no component of $\chi_{\mathbf{k}}$ proportional to $\tilde{T}$, meaning increasing the presentation time has no further effect on the transmitted information. For small nonzero Γ , the eye movement marginally offsets the adaptation, so that some signal is transmitted through the temporal filter. The extent to which this occurs is proportional to Γ , so that

$$\dot{\chi}^\infty \approx \frac{\chi_{\mathbf{k}}}{\tilde{T}} \approx S_1 \Gamma. \quad (13)$$

This appears in Fig. 2(d) as a gradient of 1.

For $1/\sigma \ll \Gamma \ll \sigma$, the short-time initial exponential decay of the temporal response dominates, yet the eye motion decorrelates the signal before the negative, adaptive part of the temporal response can take effect; hence, the eye movement overcomes adaptation. Within this regime, the rate becomes independent of Γ , with

$$\dot{\chi}^\infty \approx \frac{\chi_{\mathbf{k}}}{\tilde{T}} \approx S_2. \quad (14)$$

For this regime to be well defined, we must have $\sigma \gg 1$; that is, the short and long timescales of the temporal filter, t_s and t_l , must be well separated. We illustrate this in Fig. 2(d) by showing the rate of information gain for $\sigma = 10$, so that we can see an intermediate plateau beginning to open up between

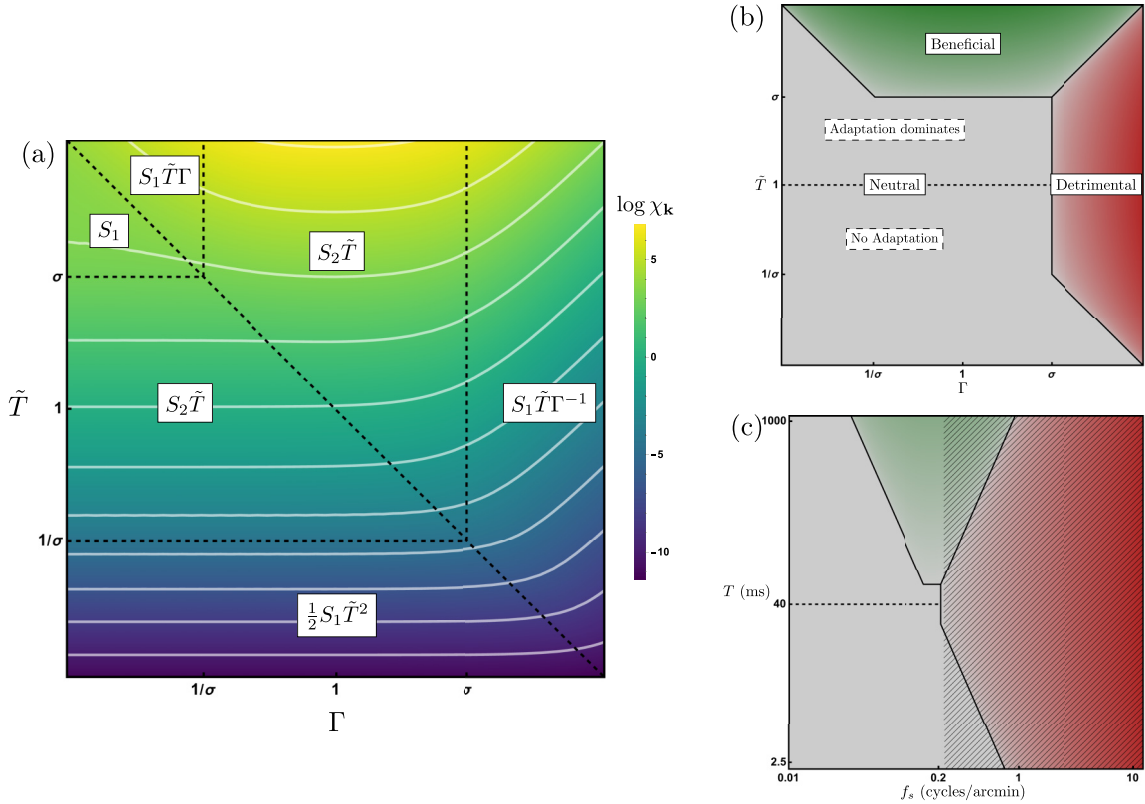


FIG. 3. The phase diagram of fixational eye motion. (a) The logarithm of χ_k , given in Eq. (10), as a function of $\Gamma = k^2 D t_a$ and the dimensionless presentation time $\tilde{T} = T/t_a$. The white lines are contours of χ_k and the labels indicate its behavior in each region, with S_1 and S_2 the constants defined in Eqs. (11) and (12). (b) A simplified version of the same phase space as in panel (a), indicating the regions in which eye motion is beneficial, detrimental, or neutral in its effect. The dashed line at $\tilde{T} = 1$ corresponds to the adaptation time. (c) The eye motion phase diagram as appropriate for human vision, shown in terms of spatial frequency, f_s , and presentation time T . In converting to dimensional units, we have taken the adaptation time $t_a = 40$ ms (again indicated by a dashed line), the diffusion constant $D = 0.02$ arcmin²/ms, and the shape parameter $\sigma = \sqrt{2}$. We have also assumed a blurring function of the form $|\hat{F}_k| \sim e^{-\frac{1}{2}\lambda^2 k^2}$ with $\lambda \approx 1$ arcmin. The hatched region indicates the spatial frequencies for which information is lost due to this blur.

$\Gamma = 1/\sigma$ and $\Gamma = \sigma$, for which $\chi_k \approx \tilde{T}$. Physiologically, it seems unlikely that there is such strong separation in the temporal filter [42,50], and so this intermediate plateau is not so well defined, as is clear when $\sigma = \sqrt{2}$. In practice, Eq. (14) then represents an upper bound of the degree to which eye motion can overcome adaptation.

Finally, when $\Gamma \gg \sigma$ the retinal motion is so rapid as to vary the incoming signal faster than the short-time positive component of the temporal response. Hence, the temporal response becomes irrelevant and the result depends only on the signal decorrelation due to eye motion, so that χ_k decreases with Γ as

$$\dot{\chi} \approx \frac{\chi_k}{\tilde{T}} \approx S_1 \Gamma^{-1}. \quad (15)$$

This appears as a gradient of -1 in Fig. 2(d).

In summary, for long presentation times $\tilde{T} > 1/\Gamma$ the rate of information gain is proportional to Γ for small Γ , proportional to $1/\Gamma$ for large Γ and, if the short and long time components of the temporal response are well separated, constant for intermediate values of Γ . This large $\tilde{T}$ limit for the contrast sensitivity, exhibiting a peak at a given spatial wave number, has been described in computational work [33,55,63]

and is consistent with experimental observation of contrast sensitivity functions [50].

We now consider how this picture is changed by shorter presentation times, $\tilde{T} < 1/\Gamma$, paying particular attention to the dependence of χ_k on $\tilde{T}$, since, as highlighted at the end of Sec. II B, this relationship will not always be linear. For short presentation times, there is no significant retinal motion on the spatial scale of the signal and the response due to switching the stimulus on and off dominates over those induced by eye motion. Hence, the retinal motion can be neglected and the information gained becomes independent of Γ , leading to the horizontal lines on the left of Fig. 2(a), the heights of which depend only on the presentation time. Focusing on this short presentation time limit, there are three subregimes, which, in analogy with those for Γ presented above, are defined by whether the presentation time is shorter than both adaptive timescales, $\tilde{T} \ll 1/\sigma$, intermediate to them, $1/\sigma \ll \tilde{T} \ll \sigma$, or longer than both, $\tilde{T} \gg \sigma$. We find that these three regimes cause the information gained to first increase quadratically with presentation time, then increase linearly, and finally stagnate to a constant value, as we now briefly explain.

When the presentation time is shorter than the short-time response of the adaptive filter, the stimulus is being integrated

such that the transmitted signal grows linearly with presentation time. Since χ_k depends on the square of the transmitted signal, it grows quadratically with presentation time

$$\chi_k \approx \frac{S_1}{2} \tilde{T}^2, \quad (16)$$

displayed by all curves in Fig. 2(c), this being the scenario highlighted at the end of Sec. II B. If the adaptive shape parameter is sufficiently large, then there will be a regime where the presentation time is between the short and long adaptive timescales, meaning the signal disappears before the negative, adaptive part of the temporal response can take effect. This is analogous to the situation yielding Eq. (14), and so again $\chi_k \approx S_2 \tilde{T}$. It is not seen in Fig. 2(c), since for all these curves $\sigma = \sqrt{2}$. Finally, when $\sigma \ll \tilde{T} \ll 1/\Gamma$ (which requires that $\Gamma \ll 1/\sigma$) the gained information saturates to a constant value given by

$$\chi_k \approx S_1. \quad (17)$$

This saturation of information gain is due to the observation time being long enough for adaptation to occur, but not long enough for the retinal motion to circumvent this adaptation by changing the signal received by the photoreceptors. This can be observed for the pink curve in Fig. 2(c), this being the only one with a small enough value of Γ to allow this regime to arise. As expected, the information plateau spans $\sqrt{2} \ll \tilde{T} \ll 100$, with the asymptotic information rate discussed earlier emerging for $\tilde{T} \gg 100$, in this case given by Eq. (13).

As noted above, Fig. 3(a) presents a phase plot indicating the values of Γ and $\tilde{T}$ where each regime of behavior in χ_k applies. This delineates the region for $\tilde{T} < 1/\Gamma$, where the response is dominated by transients arising from temporal modulation of the external stimulus via $A(t)$, and the region for $\tilde{T} > 1/\Gamma$ where the eye motion begins to have significant effect on information gained, and where a peak in χ_k emerges at $\Gamma = 1$.

Having determined the behavior of χ_k across this phase space, we can naturally establish the regions in which fixational eye movements have a beneficial, detrimental, or neutral effect on the acquisition of visual information, as shown in Fig. 3(b). To construct this diagram, at each presentation time $\tilde{T}$, we compare χ_k with its value $\chi_{k,\Gamma=0}$ at $\Gamma = 0$ that corresponds to zero eye motion. When $\chi_k > \chi_{k,\Gamma=0}$, then eye motion can be said to enhance the transmitted information, and so is beneficial. When $\chi_k < \chi_{k,\Gamma=0}$, then eye motion is detrimental to transmitted signal, while equality $\chi_k = \chi_{k,\Gamma=0}$ implies a neutral effect. This procedure gives rise to the regions sketched in Fig. 3(b). Given the long-running debate as to whether fixational eye movements have a negative, blurring, impact [12,64] or usefully modify the spatiotemporal structure of the stimulus [22,33–36,38,65,66], it is striking that all three of these regions exist.

To understand what this phase diagram implies for human vision, we convert to dimensional quantities, taking the adaptation time $t_a = 40$ ms, the diffusion constant $D = 0.02$ arcmin²/ms, and the shape parameter $\sigma = \sqrt{2}$. This results in the phase diagram shown in Fig. 3(c) in terms of stimulus spatial frequency, f_s , in cycles/arcmin and presentation time in milliseconds. We have also assumed a blurring

function of the form $|\hat{F}_k| \sim e^{-\frac{1}{2}\lambda^2 k^2}$, with $\lambda \approx 1$ arcmin. Taking $\frac{1}{2}\lambda^2 k^2 \approx 1$ as the criterion for blur having a significant effect, we find that blur eliminates the information at high spatial frequencies with $f_s \gtrsim 0.2$ cycles/arcmin, this being the hatched region in Fig. 3(c).

The practical consequence of the above is that for presentation times that are long compared to the adaptive timescales, that is, $T \gtrsim 50$ ms with the assumed parameter values, corresponding to $\tilde{T} > \sigma$, there is clear benefit from eye motion within the green portion of Figs. 3(b) and 3(c), with a pronounced peak in the gained information emerging at $\Gamma = 1$ for the magenta curves in Fig. 2(a) and an associated hotspot in Fig. 3(a). The height of the peak is proportional to the presentation time, with a shape given by Eqs. (13)–(15). Once again, when $\Gamma \ll 1/\tilde{T}$ the eye motion becomes insignificant and the gain of information saturates due to adaptation, according to Eq. (17). The overall shape of the emergent peak is consistent with the prior numerical analysis [33,55,63]. Not only do we reproduce this peak, but our theory provides a simple prediction for where it occurs. The peak exists for long presentation times because, once adaptation has set in, it is possible to have retinal motion that is neither irrelevant, nor purely decohering, but rather of appropriate magnitude to offset the adaptive effects. This is at the heart of the peak's location. It occurs when $\Gamma = 1$, that is, when the retinal motion induces temporal variation matched to the frequency of the bandpass adaptive filter. It should be highlighted that, as per our definition of Γ , the optimal eye motion is inherently stimulus-dependent, since the induced time variation is a product of both the eye's motion and the spatial variation of the stimulus. This will be important for our discussion of detection thresholds in Sec. IV.

On the other hand, for presentation times shorter than ~ 50 ms, that is, $\tilde{T} \lesssim \sigma$ with our choice of physical parameters, eye motion can only be deleterious toward signal perception. This is illustrated by the cyan curve in Fig. 2(a). On such timescales, adaptation is incomplete, and so the retinal motion is either insignificant or simply impeding the integration of the signal. The former produces the horizontal lines and flat contours to the left of Figs. 2(a) and 3(a), respectively, with the gained information following Eq. (14) or Eq. (16). In the latter case, information gain falls according to Eq. (15), leading to the gradients of -1 on the right of Fig. 2(a), and correspondingly gradients of 1 in the contours to the right of Fig. 3(a). This decrease occurs once the retinal motion is significant over the presentation window, $\Gamma \gg 1/\tilde{T}$, and dominates the adaptive timescales $\Gamma \gg \sigma$. Again, this pair of requirements means that for short presentations the eye motion is either irrelevant or dominating the decoherence of the signal; there is no regime in which they provide a benefit. This is made apparent in the red portion of Figs. 3(b) and 3(c), and can also be seen in the monotonic behavior of the contours in Fig. 3(a) for $\tilde{T} < \sigma$ and the lower three curves in Fig. 2(a), for which the same condition holds.

However, this falloff in signal perception may not be of great consequence if it occurs for spatial wavelengths that are already suppressed due to optical blur or finite size of receptors, incorporated in our description through the blurring function F in Eq. (1). As noted above, we take this blurring

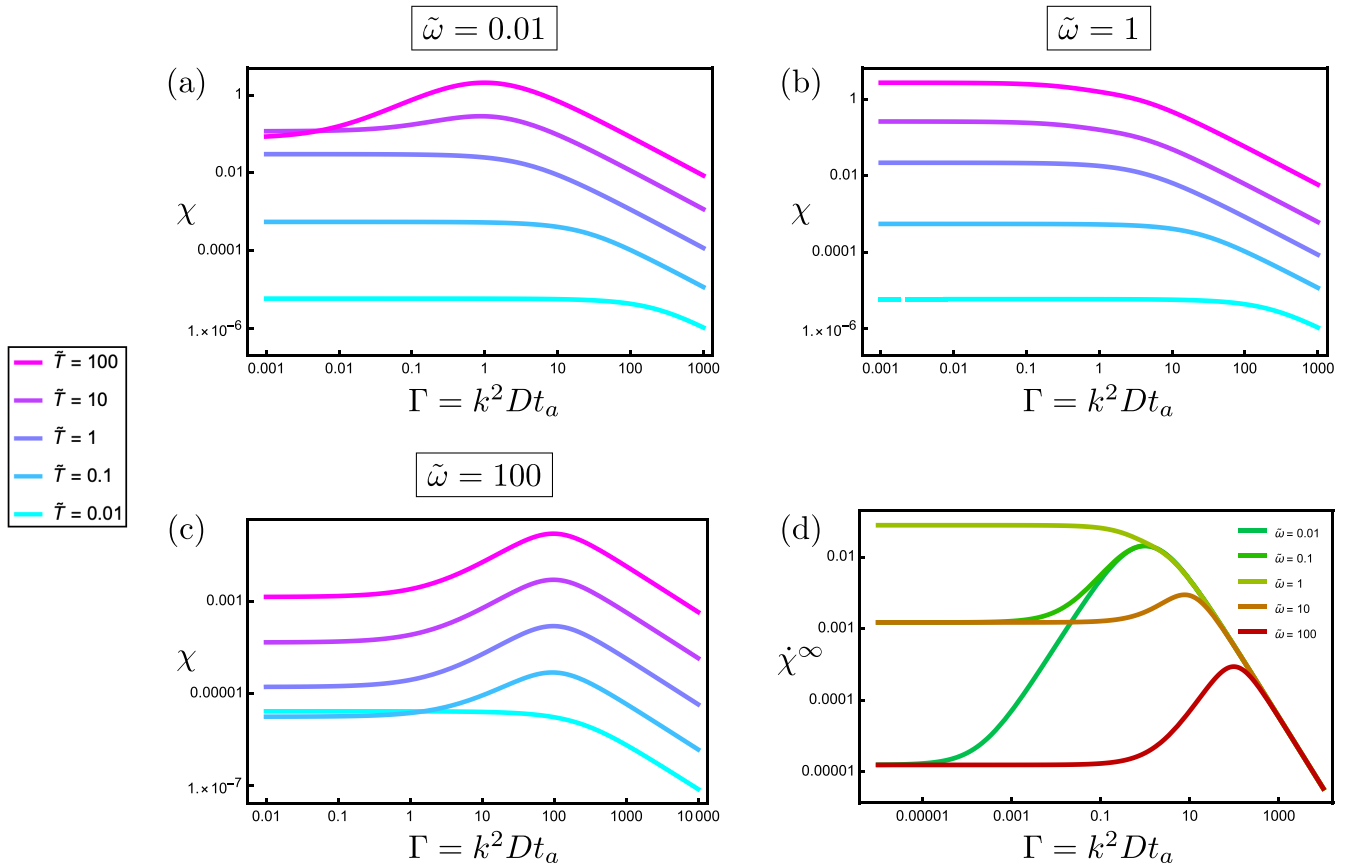


FIG. 4. The retinal information gained from a signal modulated sinusoidally at frequency $\omega = \tilde{\omega}/t_a$ with diffusive ocular drift. (a)–(c) For the colored curves, $\sigma = \sqrt{2}$ and $\tilde{T}$ varies from 0.01 (cyan) to 100 (magenta) in powers of 10. (d) The asymptotic rates of information gain for $\sigma = \sqrt{2}$ and $\tilde{\omega}$ varying from 0.01 (green) to 100 (red) in powers of 10.

function to have the form $|\hat{f}_{\mathbf{k}}| \sim e^{-\frac{1}{2}\lambda^2 k^2}$, with $\lambda \approx 1$ arcmin. As shown in Fig. 3(c), this means that with parameter values appropriate to human vision, the region in which eye motion is detrimental is essentially covered by the region in which information is already lost due to blur. This means that the magnitude of diffusion observed in human vision [25,55] is optimal to provide informational benefit, given the presence of optical blur and hints at another instance of biological systems achieving physically imposed performance limits [13,14].

B. Sinusoidal modulation

We now consider a stimulus whose amplitude is modulated sinusoidally, such that $A(t) = \cos(\omega t)$. The resulting behavior depends on how the oscillation frequency compares to the adaptation time, so we find it convenient to define a nondimensional oscillation frequency $\tilde{\omega} = \omega t_a$, and to consider a range of values of $\tilde{\omega}$. Figure 4 shows the impact of modulation frequency on the information gain. The corresponding expression for $\chi_{\mathbf{k}}$ and its derivation is given in Sec. V of the Supplemental Material [49]. For very short presentation times, so that $\tilde{\omega}\tilde{T} \ll 1$, $A(t)$ is approximately constant during the presentation time and so the results match those of the previous section. Otherwise, there are several key differences compared to the constant amplitude case. One minor difference is that the information gained does not always increase monotonically with presentation time $\tilde{T}$, since an alternating

positive and negative signal can cancel when passed through the temporal filter [42]. Typically the nonmonotonic behavior occurs for small Γ , where retinal motion is insignificant, and at $\tilde{\omega}\tilde{T} \approx 1$, where the effects of alternating signal become apparent, as seen, e.g., in Figs. 4(a) and 4(c).

It is helpful to first consider the effect of sinusoidal modulation in the absence of eye motion ($\Gamma = 0$), before discussing how eye motion perturbs the result. Here, the signal modulation provides a nonconstant signal to the temporal response, so that perfect adaptation no longer occurs. In the limit of long presentation times, $\tilde{T} \gg 1$, this gives rise to a constant rate of information acquisition, so that

$$\chi_{\mathbf{k}} \approx \frac{(\sigma^2 - 1)^2 \tilde{\omega}^2}{4(\sigma^2 + \tilde{\omega}^2)(1 + \sigma^2 \tilde{\omega}^2)} \tilde{T}. \quad (18)$$

This is maximized at $\tilde{\omega} = 1$ (and is almost constant for $1/\sigma \ll \tilde{\omega} \ll \sigma$), indicating that the transmission of information through the temporal response is optimized when the oscillation frequency is matched to the adaptation time. In this case, the oscillations of C_{AA} in Fig. S3 from the Supplemental Material [49] match well the positive and negative portions of C_{KK} in Fig. 1(g). Then, given that the signal modulation is optimally overcoming adaptation, adding eye motion can only make things worse. Hence, in Fig. 4(b), $\chi_{\mathbf{k}}$ is always a decreasing function of Γ , transitioning from a constant value toward $\chi_{\mathbf{k}} \sim \tilde{T}/\Gamma$ above a critical value of Γ (which is $\Gamma = 1$ for $\tilde{T} > 1$, or $\Gamma = 1/\tilde{T}$ for $\tilde{T} < 1$).

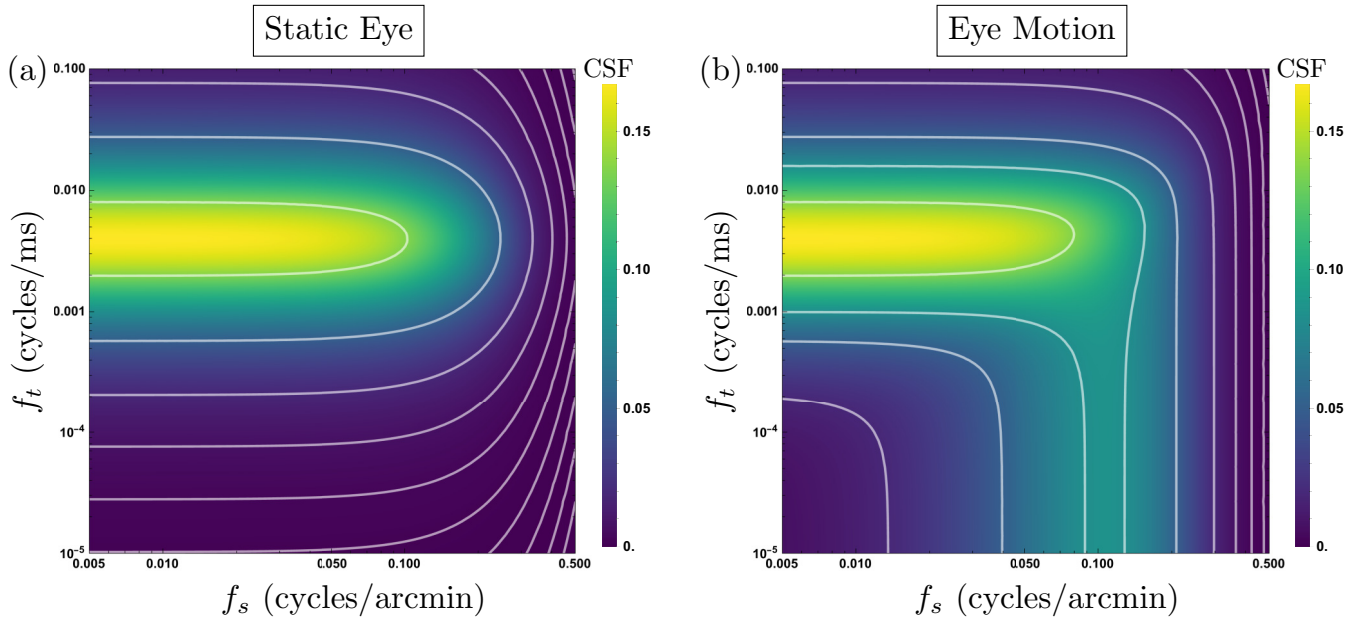


FIG. 5. Spatiotemporal contrast sensitivity functions. The contrast sensitivity function, given in Eq. (20), is presented as a function of spatial and temporal stimulus frequencies, f_s and f_t , for (a) no eye motion, and (b) $D = 0.02 \text{ arcmin}^2/\text{ms}$. In both cases, the adaptation time has been taken as $t_a = 40 \text{ ms}$, the shape parameter as $\sigma = \sqrt{2}$, and the blur length scale as $\lambda = 1 \text{ arcmin}$. The white lines correspond to contours of the contrast sensitivity function.

For smaller frequencies $\tilde{\omega} \ll 1/\sigma$, the modulation is too slow to optimally overcome adaptation. In this case, eye movement still helps, so that a peak in χ_k at $\Gamma = 1$ emerges when $\tilde{T} > 1$, as can be seen in Fig. 4(a). This is qualitatively similar to the constant amplitude case in the previous section, except that in the limit of low eye movement, $\Gamma \rightarrow 0$, the plateau in χ_k continues to increase when $\tilde{T} > 1$ because of the signal modulation.

Finally, for high-oscillation frequencies, $\tilde{\omega} \gg \sigma$, the signal modulation is more rapid than the integration time of the fast component of the temporal response. Hence, the temporal response averages over the modulated signal, giving a low output in the absence of eye movement. In this case, eye movement can help, not by overcoming adaptation, but rather by spatially decorrelating the signal faster than the amplitude modulation. This occurs when the decay of S_k in Fig. 1(f) matches well the oscillations of C_{AA} in Fig. S3 from the Supplemental Material [49], cutting out the negative portion of C_{AA} . The optimum is when the rate of eye movement matches the signal modulation, giving a peak in χ_k at $\Gamma = \tilde{\omega}$ that emerges for $\tilde{T} > 1/\tilde{\omega}$, as illustrated in Fig. 4(c).

Hence, for long presentation times $\tilde{T} > 1$ we find that eye motion helps both when $\tilde{\omega} < 1/\sigma$ (when the motion continues to help overcome adaptation) and for $\tilde{\omega} > \sigma$ (when the eye movement offsets the signal modulation). For $\tilde{\omega} \approx 1$, the eye movement is no longer beneficial. With eye motion now included, the long time behavior is given by

$$\chi_k \approx \frac{(\sigma^2 - 1)^2}{4(\sigma^2 + 1)[(\Gamma + \sigma)^2 + \tilde{\omega}^2]} \times \left[1 + \frac{(1 + \Gamma\sigma)(\Gamma^2 + \tilde{\omega}^2 - 1)}{(1 + \Gamma\sigma)^2 + (\sigma\tilde{\omega})^2} \right] \tilde{T}, \quad (19)$$

the corresponding rates being the curves shown in Fig. 4(d). Recalling that the contrast sensitivity is proportional to $\sqrt{\chi_k}$, Eq. (19) and Fig. 4(d) can be considered to represent the contributions of eye movement, signal modulation, and temporal response to the spatiotemporal contrast sensitivity function [58–61,63]. The symmetry between a frequency and its inverse for small Γ is a consequence of our choice of temporal response function, although the power laws for small and large frequencies are generic. A similar symmetry is apparent in Fig. 2(b).

The full spatiotemporal contrast sensitivity function, $\text{CSF}(\Gamma, \tilde{T})$, is attained by incorporating blur, which, as in the construction of Fig. 3(c), we take to be of the form $|\hat{f}_k| \sim e^{-\frac{1}{2}\lambda^2 k^2}$, giving

$$\text{CSF}(\Gamma, \tilde{T}) = \left\{ \frac{(\sigma^2 - 1)^2}{4(\sigma^2 + 1)[(\Gamma + \sigma)^2 + \tilde{\omega}^2]} \times \left[1 + \frac{(1 + \Gamma\sigma)(\Gamma^2 + \tilde{\omega}^2 - 1)}{(1 + \Gamma\sigma)^2 + (\sigma\tilde{\omega})^2} \right] \right\}^{1/2} e^{-\frac{1}{2}\lambda^2 k^2}. \quad (20)$$

This is shown in Fig. 5 as a function of spatial frequency, f_s , in cycles/arcmin and temporal frequency, f_t , in cycles/ms. In converting to dimensional units, we have used the same parameter values as in Fig. 3(c), namely, $t_a = 40 \text{ ms}$, $\sigma = \sqrt{2}$, and $\lambda = 1 \text{ arcmin}$. The case with no eye motion, $D = 0$, is shown in Fig. 5(a) and exhibits a bandpass response in temporal frequency, owing to the adaptive linear response, and low-pass behavior in spatial frequency, due to blur.

The case with eye motion is illustrated in Fig. 5(b) for $D = 0.02 \text{ arcmin}^2/\text{ms}$. The contrast sensitivity function is now bandpass in both temporal and spatial frequencies. When the temporal frequency is below its bandpass value, the optimal

spatial frequency is constant. However, for higher temporal frequencies, the optimal spatial frequency is shifted upward, as indicated by the outward bowing contour in Fig. 5(b). This behavior mirrors that shown in Fig. 4(d), where $\dot{\chi}^\infty$ peaks at $\Gamma = 1$ for $\tilde{\omega} < 1$, but for $\tilde{\omega} > 1$ the peak moves to the right, occurring at $\Gamma \approx \tilde{\omega}$.

We have considered the contrast sensitivity function in the limit of infinite presentation time, that is, using $\dot{\chi}^\infty$, but we can infer the effect of finite presentation times from Figs. 4(a)–4(c). Reducing the presentation time flattens the rise in χ_k , such that the peak is diminished relative to its value for low Γ . In the context of Fig. 5, this corresponds to filling in the region where both f_s and f_t are low. For sufficiently short presentation times, the bandpass spatial response vanishes, as indicated by the cyan curves in Figs. 4(a)–4(c).

The key message from Fig. 5 is that, by converting spatial variation in a stimulus into temporal variation in the photoreceptor signal, eye movements can transmute the adaptive bandpass response for temporal frequencies into a bandpass response for spatial frequencies. Even though it has been established that retinal ganglion cells also produce a spatial bandpass effect [67–71], it is striking that such mechanisms are not required to reproduce the features of the spatiotemporal contrast sensitivity function; this can be done with eye motion alone.

With this eye-motion-mediated connection between temporal and spatial adaptation in hand, we are now in a position to consider whether the apparent scale-dependent adaptation seen in experiments of Barlow [41] may also be explained by eye movements.

IV. DETECTION THRESHOLDS

In the previous section, we determined the impact of retinal motion on the eye's ability to gain information from a signal with a given spatial frequency. We now extend this to consider the impact of this movement on the threshold amplitude for detection of a target, in which many spatial frequencies are superposed. In doing this, we employ a model of detection probability, p , proposed by Watson [42]. The underlying principle of the model is that detection of a stimulus requires only that the stimulus be detected above the background by at least one of the available receptors, at one or more times during the observation window. As shown in Sec. I D of the Supplemental Material [49], this model gives a detection probability of

$$p = 1 - \exp \left[-A^2 \frac{\alpha}{4\pi^2} \int d^2\mathbf{k} |\hat{F}_k|^2 |\hat{\chi}_k|^2 \chi_k \right], \quad (21)$$

where α is a normalization constant and we have described the stimulus in terms of a spatial form $\hat{\chi}_k$ and an overall intensity A . The stimulus intensity A can be varied to find the threshold for detection, which, from Eq. (21), is

$$A_{\text{th}} \propto \left[\frac{1}{4\pi^2} \int d^2\mathbf{k} |\hat{F}_k|^2 |\hat{\chi}_k|^2 \chi_k \right]^{-1/2}. \quad (22)$$

We consider the experimental paradigm employed by Barlow [41], in which detection thresholds were measured for circular stimuli of two different sizes: a large target of radius 177 arcmin and a small one of radius 3.5 arcmin.

For both targets, it was found that the threshold intensity initially decreased with the inverse of the presentation time. For large stimuli, the threshold intensity then did not change with further increases in presentation time, due to adaptation. However, for a small stimulus the threshold intensity continued to decrease, but at a reduced rate.

To explain this behavior, it has been suggested that the intrinsic retinal response has a form of temporal adaptation with some spatial dependence, such that low spatial frequencies are perfectly adapted, but high spatial frequencies are imperfectly adapted [42]. Such a retinal response cannot be modeled within the simple framework presented in this paper, as presented in Sec. II A, using a separate temporal response function K and spatial blurring function F . Instead, this explanation requires that the functions K and F should be combined into a nonfactorizable spatiotemporal retinal response function (i.e., this explanation requires that the intrinsic spatiotemporal contrast sensitivity function of the retina cannot be separated into the product of spatial and temporal functions). It is certainly possible that the retinal ganglion cells produce such a complex spatiotemporal response, for example, by giving rise to multiple channels of spatial response, each with their own temporal behavior [42,72]. However, we shall show that all features of the experiments can be explained without invoking such scale-dependent adaptation. The necessary spatiotemporal complexity arises naturally within our minimal model, via the interaction of retinal motion with an adaptive temporal response function.

For a signal with constant intensity within a circle of radius R , and zero outside this, a standard calculation gives $\hat{\chi}_k = \frac{2\pi R}{k} J_1(kR)$, with J_1 a Bessel function of the first kind. Additionally, we once again assume an exponential blurring function, $|\hat{F}_k|^2 \sim e^{-\lambda^2 k^2}$. Consequently,

$$A_{\text{th}} \propto \left[2\pi R^2 \int_0^\infty \frac{1}{k} e^{-\lambda^2 k^2} [J_1(kR)]^2 \chi_k dk \right]^{-1/2}. \quad (23)$$

The integral is a sum over χ_k , the rate of gaining information from a wave vector $\mathbf{k}$, weighted by the prevalence of that wave vector in the signal and the blurring function. In Fig. 6, we plot this threshold intensity A_{th} as a function of stimulus presentation time T , both for (a) the large target of radius 177 arcmin and (b) the small one of radius 3.5 arcmin. Here, we use the same parametrization as in previous figures, i.e., adaptation time $t_a = 40$ ms, shape parameter $\sigma = \sqrt{2}$, and blur length scale $\lambda = 1$ arcmin, but we show results for a range of diffusion constants D from 0.001 25 min²/ms (cyan) to 0.04 min²/ms (magenta).

Although there does not exist a closed form for A_{th} , we can understand all its essential features from a few simple arguments. As shown in Sec. III A, at short times $\chi_k \sim \tilde{T}^2$, while at long times $\chi_k \sim \tilde{T}$, in both cases for all k . This immediately produces the short time $A_{\text{th}} \sim \frac{1}{\tilde{T}}$ behavior seen in Ref. [41], as well as an $A_{\text{th}} \sim \frac{1}{\tilde{T}^{1/2}}$ dependence at long times. Both these regimes are indicated with black dashed lines in Fig. 6. At intermediate times, there is a plateau due to adaptation causing the gain of information to stagnate. In Sec. III A, we showed that this stagnation begins after full adaptation, when $\tilde{T} \approx \sigma$, i.e., $T = t_l$, the long adaptation time (of order

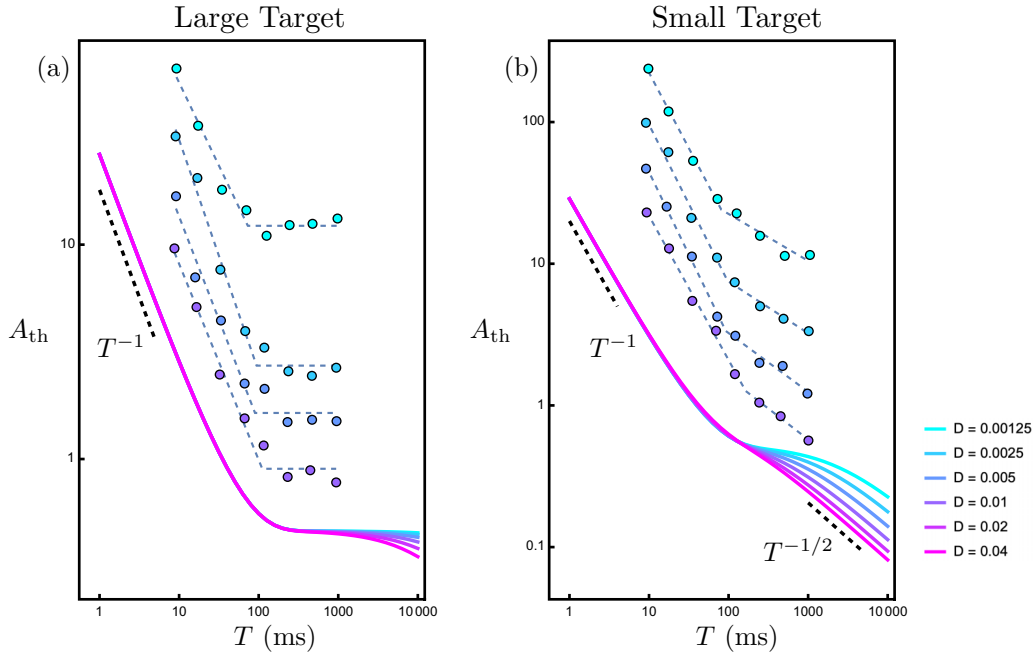


FIG. 6. The contrast detection threshold, A_{th} , for a circular stimulus of radius R (measured in arcmin) as a function of observation time, T , (in ms). The dots show the data from Barlow's experiments [41], with the dashed lines added to guide the eye, and the solid lines the prediction of Eq. (23) for (a) the large target ($R = 177$ arcmin) and (b) the small target ($R = 3.5$ arcmin). The ocular drift is taken to be diffusive with diffusion constant D (measured in min^2/ms). In both panels, D is varied in powers of 2 from $0.00125 \text{ min}^2/\text{ms}$ (cyan) to $0.04 \text{ min}^2/\text{ms}$ (magenta). The data from Barlow's experiments have been colored according to the background intensity, the logarithm of which is 7.83 (cyan), 5.94, 4.96, and 3.65 (purple). Since our theory predicts only the dependence of A_{th} with T , not its absolute value, we have applied an overall shift to the datasets to ease presentation.

60 ms for our parametrization). This stagnation is apparent in both Figs. 6(a) and 6(b), as the time at which the plateau commences. The information stagnation ends once there has been significant retinal motion to overcome adaptation, that is, when $\tilde{T} = \frac{1}{k^2 D_{\text{eff}}}$, i.e., $T = \frac{1}{k^2 D}$. Hence, irrespective of the stimulus size, we expect the plateau to shorten as the diffusion constant D increases, which is again apparent in both Figs. 6(a) and 6(b).

The crucial point is that a given eye motion does not affect all spatial scales in the stimulus equally, and this inherent scale dependence replaces the need for any scale dependence in the form of adaptation. Making the target smaller, i.e., decreasing R , puts the relevant visual information at smaller length scales, or larger spatial frequencies k . Hence, $T = \frac{1}{k^2 D}$ is smaller at the relevant scale and so decreasing R shortens the plateau, as is clear when comparing Figs. 6(a) and 6(b). Indeed, for the small stimulus size of radius 3.5 arcmin in Fig. 6(b), when D is sufficiently large ($0.01 \text{ arcmin}^2/\text{ms}$ or larger) the plateau almost entirely vanishes, so that one finds only a transition region of indeterminate slope between the regimes of $A_{th} \sim \frac{1}{T}$ and $A_{th} \sim \frac{1}{T^{1/2}}$.

Figures 6(a) and 6(b) also show the results of Barlow [41]. In these experiments, the contrast threshold for the large and small stimuli were measured for various background light intensities, shown as separate datasets on the plots. We have omitted data from the lowest light intensity, since visual response is known to be altered at low intensities [50,73,74]. Since our theory predicts only the dependence of A_{th} with T , not its absolute value, we have applied an overall shift to the

datasets to ease presentation. It is clear that with a reasonable choice of diffusion constant (say $D = 0.02 \text{ arcmin}^2/\text{s}$), and with the reasonable choices for other parameters described above, our minimal model matches the temporal behavior of the experiments extremely well. In particular, we reproduce the central result of stagnation of contrast detection threshold for the large stimulus, but that the contrast detection threshold continues to decrease for the smaller stimulus.

In practice, it would be impractical to run sufficiently long experiments to see the fully established $A_{th} \sim \frac{1}{T^{1/2}}$ regime, and indeed additional physiology may obscure it. Hence, the time axis in Fig. 6(a) is truncated in the experiments by Barlow, and we believe that what was observed is perfect adaptation for large targets, and the transition toward the $A_{th} \sim \frac{1}{T^{1/2}}$ regime for smaller targets.

The experimental observations of the qualitative difference in contrast threshold with stimulus size can thus be explained solely via the space-time coupling produced by fixational eye movements. These effects are seen even within our minimal model. It is thus not necessary to invoke complex length scale dependence of the temporal adaptive response to explain these data. However, as noted above, such conjectures exist in the literature [42,72]. This points to the need for additional experiments that would tease apart these two possibilities. A suitable investigation could involve comparing contrast detection thresholds as a function of presentation time, for stimuli that are either stabilized or unstabilized on the retina. Here, if the effects observed by Barlow are accounted for by eye movements, then with the effects of eye movement removed

via stabilization the temporal response to small stimuli should more closely match those of large stimuli.

V. DISCUSSION

We have formulated an analytic framework for determining the impact of ocular drift on the information acquired from a stimulus. Our measure of the acquired information may be derived from the mutual information between the stimulus and a minimal model for the response of the early visual system, which incorporates drift, along with blur and a bandpass temporal response. Taking the drift to be diffusive, we identify the key dimensionless parameters that underpin the interactions between the stimulus structure, temporal response, and ocular drift. These interactions are phenomenologically rich, as we summarize in a phase diagram, and have the consequence that ocular drift may be beneficial, detrimental, or negligible in its impact on the acquisition of visual information. We find that optimal performance is achieved when $k^2 D t_a \approx 1$, where k is the wave vector of the stimulus, D is the diffusion constant of the drift, and t_a is a timescale characterizing the adaptive temporal response.

Working with parameters appropriate for human vision, our model reveals that the stimulus wave vector for which the observed degree of drift has the greatest benefit and the wave vector above which information is lost due to blur are the same. This suggests that, far from merely being noise, ocular drift is tuned to maximize the gain of visual information given the constraint of optical blur. Although such a statement exists in the literature [33], our work shows concretely how this arises from different components of the model. Additionally, a key emergent principle is that retinal motion combined with a bandpass temporal response produces an effective bandpass spatial response, even though such a response is not an explicit component of the model. This is shown most fundamentally in the modeled contrast sensitivity function. Although the existence of neural spatial receptive fields with a so-called center-surround (biphasic) structure is well established [67–71], we believe it important to note that the data at the system level can be qualitatively explained without this. Furthermore, the model provides an alternative qualitative explanation for Barlow’s data on threshold detection amplitudes for stimuli of different sizes as a function of time [41]. These data have been taken as evidence for multiple spatial mechanisms that differ in their temporal response, such that smaller stimuli are postulated to recruit detection mechanisms with more sustained temporal impulse response, meaning that the integral of the linear temporal response function is not zero, but positive [42]. The model instead shows that the observed behavior can emerge from the combination of eye movements and a biphasic temporal response. Indeed, our work indicates a need for further experiments, repeating those of Barlow but with the stimulus stabilized on the retina, to distinguish the roles that ocular drift and spatiotemporal tuning of neural mechanisms play in producing the observed behavior.

There are various ways in which our calculational approach could be extended. For example, one could relax the condition that the spatial and temporal variation of stimulus be separable, or consider a prior distribution of stimuli consistent with the statistics of natural images [75–77]. In this latter context,

it would be interesting to turn the calculation on its head, such that rather than calculating the effect of a prescribed ensemble of retinal motion on the available visual information, we instead determine the optimal form of ocular drift on the basis that optimality corresponds to whitening the power spectrum [1,33,45,78].

In this paper, we have taken a minimal approach to physiological details, both by neglecting certain features and in the modeling of those we retain. However, our framework is sufficiently generic to allow more detail to be incorporated as desired. As one example of this, we assumed the FEMs to be purely diffusive yet, as noted earlier, true eye motion paths may be biased toward a preferred retinal location [51,52], and could have short term persistence and longer-term antipersistence [53,54]. There is no need to take the form of ocular drift to be diffusive and indeed, although not explored here, our theory indicates that a degree of persistence may be beneficial. The heuristic argument for this is that persistence increases the power law of the mean squared displacement for the time interval over which motion is persistent, meaning that the path structure factor $S_k(\tau)$ has a more pronounced step, i.e., decaying more slowly at short times and more rapidly at long times. This enhances the positive parts of the memory kernel autocorrelation function while allowing greater suppression of the negative parts. Initial tests suggest this tends to increase χ_k at larger wave vectors beyond the peak that emerges in Fig. 2(a), i.e., for $\Gamma > 1$. However, we believe care remains to be taken to ensure that the ensemble averages employed in deriving Eq. (3) remain valid for persistent motion with finite stimulus presentation times.

There are several other ways that the details of our model may be made more realistic within the existing overall framework. While we have taken the spatial response to be a low-pass filter originating from optical blur, there is no obstruction to incorporating a bandpass spatial response. It would then be interesting to study both separable and nonseparable forms of spatial-temporal response, particularly given the latter’s relevance to attempts to connect Barlow’s data [41] to postulated underlying physiological mechanisms, as described above. We have taken the photoreceptor array to be continuous and homogeneous, neglecting, among other things, the discrete nature of photoreceptors and the high-acuity foveola in the center of the retina. Incorporating these features into our model would provide insight into other hypotheses for the function of FEMs, such as that they act to maintain a preferred retinal location [51], or achieve super-resolution through appropriate sampling of an image [52,79].

Nonetheless, we maintain that this work does not suffer by omitting such details. This allows us to expose the basic principles underpinning the function of ocular drift in vision. Moreover, it lets us trade specificity for universality and see that, while the effects we have described apply to vision, they are not limited to that context. Motion and some form of adaptive response to the environment are near-ubiquitous signatures of life and our results indicate how these may interact nontrivially when an organism tries to learn about its surroundings. As one example, bacteria employ a bandpass filter to detect chemical gradients, which they traverse via runs interrupted by stochastic tumbles. Though the mechanisms behind the two are different, one might wonder whether these

tumblings have a similar impact on sensing chemical signals as ocular drift has on detecting visual ones.

ACKNOWLEDGMENTS

This work was supported by the UK Research And Innovation (UKRI) and the Wellcome Trust grant “A fresh look at visual sampling: How are fixational eye-movements optimised? [PhysFEM],” Grant No. EP/W023873/1. The authors

are grateful to T. McLeish, now deceased, who was instrumental in bringing together this project and collaboration.

DATA AVAILABILITY

No data were created or analyzed in this study. All Figures are simply plots of equations given in the main text or Supplemental Material [49]. Please contact the authors if further information is required.

-
- [1] M. Rucci and M. Poletti, Control and functions of fixational eye movements, *Annu. Rev. Vision Sci.* **1**, 499 (2015).
 - [2] M. F. Land, Motion and vision: why animals move their eyes, *J. Comp. Physiol. A* **185**, 341 (1999).
 - [3] S. Martinez-Conde, S. L. Macknik, and D. H. Hubel, The role of fixational eye movements in visual perception, *Nat. Rev. Neurosci.* **5**, 229 (2004).
 - [4] R. G. Alexander and S. Martinez-Conde, Fixational eye movements, in *Eye Movement Research: An Introduction to its Scientific Foundations and Applications* (Springer, Switzerland, 2019), pp. 73–115.
 - [5] N. R. Bowers, A. E. Boehm, and A. Roorda, The effects of fixational tremor on the retinal image, *J. Vision* **19**, 8 (2019).
 - [6] M. Rolfs, R. Kliegl, and R. Engbert, Toward a model of microsaccade generation: The case of microsaccadic inhibition, *J. Vision* **8**, 5 (2008).
 - [7] A.-K. Hauperich, L. K. Young, and H. E. Smithson, What makes a microsaccade? A review of 70 years of research prompts a new detection method, *J. Eye Mov. Res.* **12**, 13 (2019).
 - [8] W. H. Miller and G. D. Bernard, Averaging over the foveal receptor aperture curtails aliasing, *Vision Res.* **23**, 1365 (1983).
 - [9] J. Jurin, *A Reply to Mr. Robins's Remarks on the Essay Upon Distinct and Indistinct Vision Published at the End of Dr. Smith's Compleat System of Opticks* (W. Innys and R. Manby, London, 1739).
 - [10] R. Waring Darwin and E. Darwin, New experiments on the ocular spectra of light and colours, *Philos. Trans. R. Soc. London Ser. I* **76**, 313 (1786).
 - [11] H. Von Helmholtz, *Helmholtz's Treatise on Physiological Optics* (Optical Society of America, Rochester, NY, 1925), Vol. 3.
 - [12] O. Packer and D. R. Williams, Blurring by fixational eye movements, *Vision Res.* **32**, 1931 (1992).
 - [13] J. Burge, Image-computable ideal observers for tasks with natural stimuli, *Annu. Rev. Vision Sci.* **6**, 491 (2020).
 - [14] W. Bialek, Physical limits to sensation and perception, *Annu. Rev. Biophys. Chem.* **16**, 455 (1987).
 - [15] S. Hecht, S. Shlaer, and M. H. Pirenne, Energy, quanta, and vision, *J. Gen. Physiol.* **25**, 819 (1942).
 - [16] D. A. Baylor, T. D. Lamb, and K.-W. Yau, Responses of retinal rods to single photons, *J. Physiol.* **288**, 613 (1979).
 - [17] A. Dey, A. J. Zele, B. Feigl, and P. Adhikari, Threshold vision under full-field stimulation: Revisiting the minimum number of quanta necessary to evoke a visual sensation, *Vision Res.* **180**, 1 (2021).
 - [18] H. B. Barlow, The size of ommatidia in apposition eyes, *J. Exp. Biol.* **29**, 667 (1952).
 - [19] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, Reading, Massachusetts, 1963), Vol. I.
 - [20] T. C. McLeish, Are there ergodic limits to evolution? Ergodic exploration of genome space and convergence, *Interface Focus* **5**, 20150041 (2015).
 - [21] E. N. Clarkson and R. Levi-Setti, Trilobite eyes and the optics of Descartes and Huygens, *Nature (London)* **254**, 663 (1975).
 - [22] E. Hering, Über die grenzen der sehschärfe, *Ber. math.-phys. Cl. D. königl. Sächs. Gesell. Wiss. Leipzig* **303**, 16 (1899).
 - [23] M. Rucci, R. Iovin, M. Poletti, and F. Santini, Miniature eye movements enhance fine spatial detail, *Nature (London)* **447**, 852 (2007).
 - [24] K. Ratnam, N. Domdei, W. M. Harmening, and A. Roorda, Benefits of retinal image motion at the limits of spatial vision, *J. Vision* **17**, 30 (2017).
 - [25] J. Intoy and M. Rucci, Finely tuned eye movements enhance visual acuity, *Nat. Commun.* **11**, 795 (2020).
 - [26] J. Nachmias, Determiners of the drift of the eye during monocular fixation, *J. Opt. Soc. Am.* **51**, 761 (1961).
 - [27] R. M. Steinman, G. M. Haddad, A. A. Skavenski, and D. Wyman, Miniature eye movement: The pattern of saccades made by man during maintained fixation may be a refined but useless motor habit, *Science* **181**, 810 (1973).
 - [28] E. Kowler, Eye movements: The past 25 years, *Vision Res.* **51**, 1457 (2011).
 - [29] E. Ahissar and A. Arieli, Seeing via miniature eye movements: A dynamic hypothesis for vision, *Front. Comput. Neurosci.* **6**, 89 (2012).
 - [30] M. Rucci, E. Ahissar, D. C. Burr, I. Kagan, M. Poletti, and J. D. Victor, The visual system does not operate like a camera, *J. Vision* **25**, 2 (2025).
 - [31] R. Ditchburn and B. Ginsborg, Vision with a stabilized retinal image, *Nature (London)* **170**, 36 (1952).
 - [32] L. A. Riggs, F. Ratliff, J. C. Cornsweet, and T. N. Cornsweet, The disappearance of steadily fixated visual test objects, *J. Opt. Soc. Am.* **43**, 495 (1953).
 - [33] M. Rucci and J. D. Victor, The unsteady eye: An information-processing stage, not a bug, *Trends Neurosci.* **38**, 195 (2015).
 - [34] L. E. Arend Jr., Spatial differential and integral operations in human vision: Implications of stabilized retinal image fading, *Psychol. Rev.* **80**, 374 (1973).
 - [35] H. L. Averill and F. W. Weymouth, Visual perception and the retinal mosaic. II. The influence of eye-movements

- on the displacement threshold, *J. Comp. Psychol.* **5**, 147 (1925).
- [36] W. H. Marshall and S. A. Talbot, Recent evidence for neural mechanisms in vision leading to a general theory of sensory acuity, in *Visual Mechanisms*, edited by H. Klüver (Jacques Cattell, Lancaster, PA, 1942), pp. 117–164.
- [37] R. M. Steinman and J. Z. Levinson, The role of eye movement in the detection of contrast and spatial detail, in *Eye Movements and their Role in Visual and Cognitive Processes*, edited by E. Kowler (Elsevier, Amsterdam, 1990), pp. 115–212.
- [38] E. Ahissar and A. Arieli, Figuring space by time, *Neuron* **32**, 185 (2001).
- [39] U. Tulunay-Keesey and R. M. Jones, The effect of micromovements of the eye and exposure duration on contrast sensitivity, *Vision Res.* **16**, 481 (1976).
- [40] M. Gur, Seeing on the fly: Physiological and behavioral evidence show that space-to-space representation and processing enable fast and efficient performance by the visual system, *J. Vision* **24**, 11 (2024).
- [41] H. Barlow, Temporal and spatial summation in human vision at different background intensities, *J. Physiol.* **141**, 337 (1958).
- [42] A. B. Watson, Temporal Sensitivity, in *Handbook of Perception and Human Performance*, edited by K. Boff, L. Kaufman, and J. Thomas (John Wiley & Sons, New York, 1986), Chap. 6.
- [43] M. Potters and W. Bialek, Statistical mechanics and visual signal processing, *J. Phys.* **14**, 1755 (1994).
- [44] D. A. Clark and J. E. Fitzgerald, Optimization in visual motion estimation, *Annu. Rev. Vision Sci.* **10**, 23 (2024).
- [45] W. Bialek, *Biophysics: Searching for Principles* (Princeton University Press, Princeton, NJ, 2012).
- [46] H. L. de Vries, The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye, *Physica* **10**, 553 (1943).
- [47] A. Rose, The relative sensitivities of television pickup tubes, photographic film, and the human eye, *Proc. IRE* **30**, 293 (1942).
- [48] A. Rose, The sensitivity performance of the human eye on an absolute scale, *J. Opt. Soc. Am.* **38**, 196 (1948).
- [49] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/14qd-5s7d> for all derivations of key results, key measures of gaining information about external stimuli, and discussion of alternative forms of temporal filter.
- [50] A. T. Rider, G. B. Henning, and A. Stockman, Light adaptation controls visual sensitivity by adjusting the speed and gain of the response to light, *PLoS One* **14**, e0220358 (2019).
- [51] J. L. Reiniger, N. Domdei, F. G. Holz, and W. M. Harmening, Human gaze is systematically offset from the center of cone topography, *Curr. Biol.* **31**, 4188 (2021).
- [52] J. L. Witten, V. Lukyanova, and W. M. Harmening, Sub-cone visual resolution by active, adaptive sampling in the human foveola, *eLife* **13**, RP98648 (2024).
- [53] K. Mergenthaler and R. Engbert, Modeling the control of fixational eye movements with neurophysiological delays, *Phys. Rev. Lett.* **98**, 138104 (2007).
- [54] C. J. Herrmann, R. Metzler, and R. Engbert, A self-avoiding walk with neural delays as a model of fixational eye movements, *Sci. Rep.* **7**, 12958 (2017).
- [55] X. Kuang, M. Poletti, J. D. Victor, and M. Rucci, Temporal encoding of spatial information during active visual fixation, *Curr. Biol.* **22**, 510 (2012).
- [56] J. Porter, A. Guirao, I. G. Cox, and D. R. Williams, Monochromatic aberrations of the human eye in a large population, *J. Opt. Soc. Am. A* **18**, 1793 (2001).
- [57] L. N. Thibos, X. Hong, A. Bradley, and X. Cheng, Statistical variation of aberration structure and image quality in a normal population of healthy eyes, *J. Opt. Soc. Am. A* **19**, 2329 (2002).
- [58] J. G. Robson, Spatial and temporal contrast-sensitivity functions of the visual system, *J. Opt. Soc. Am.* **56**, 1141 (1966).
- [59] F. W. Campbell and J. G. Robson, Application of fourier analysis to the visibility of gratings, *J. Physiol.* **197**, 551 (1968).
- [60] R. L. De Valois, H. Morgan, and D. M. Snodderly, Psychophysical studies of monkey vision-iii. spatial luminance contrast sensitivity tests of macaque and human observers, *Vision Res.* **14**, 75 (1974).
- [61] D. H. Kelly, Motion and vision. II. Stabilized spatio-temporal threshold surface, *J. Opt. Soc. Am.* **69**, 1340 (1979).
- [62] H. D. L. Dzn, Experiments on flicker and some calculations on an electrical analogue of the foveal systems, *Physica* **18**, 935 (1952).
- [63] A. Casile, J. D. Victor, and M. Rucci, Contrast sensitivity reveals an oculomotor strategy for temporally encoding space, *eLife* **8**, e40924 (2019).
- [64] Y. Burak, U. Rokni, M. Meister, and H. Sompolinsky, Bayesian model of dynamic image stabilization in the visual system, *Proc. Natl. Acad. Sci. USA* **107**, 19525 (2010).
- [65] M. Rucci and A. Casile, Fixational instability and natural image statistics: Implications for early visual representations, *Network: Comput. Neural Syst.* **16**, 121 (2005).
- [66] M. Boi, M. Poletti, J. D. Victor, and M. Rucci, Consequences of the oculomotor cycle for the dynamics of perception, *Curr. Biol.* **27**, 1268 (2017).
- [67] C. Enroth-Cugell and J. G. Robson, The contrast sensitivity of retinal ganglion cells of the cat, *J. Physiol.* **187**, 517 (1966).
- [68] C. Enroth-Cugell, J. Robson, D. Schweitzer-Tong, and A. Watson, Spatio-temporal interactions in cat retinal ganglion cells showing linear spatial summation, *J. Physiol.* **341**, 279 (1983).
- [69] L. J. Croner and E. Kaplan, Receptive fields of P and M ganglion cells across the primate retina, *Vision Res.* **35**, 7 (1995).
- [70] D. M. Dacey, Primate retina: cell types, circuits and color opponency, *Prog. Retinal Eye Res.* **18**, 737 (1999).
- [71] L. Diller, O. S. Packer, J. Verweij, M. J. McMahon, D. R. Williams, and D. M. Dacey, L and M cone contributions to the midget and parasol ganglion cell receptive fields of macaque monkey retina, *J. Neurosci.* **24**, 1079 (2004).
- [72] J. Kulikowski and D. Tolhurst, Psychophysical evidence for sustained and transient detectors in human vision, *J. Physiol.* **232**, 149 (1973).
- [73] D. Lamming, Contrast sensitivity, in *Vision and Visual Dysfunction*, Vol. 5, edited by J. Cronly-Dillon (Macmillan Press, London, 1991).
- [74] W. Hart Jr., The temporal responsiveness of vision, in *Adlers Physiology of the Eye, Clinical Application*, edited by R. J. Moses and W. M. Hart (The C.V. Mosby Company, St. Louis, 1987), pp. 429–457.

- [75] D. Ruderman and W. Bialek, Statistics of natural images: Scaling in the woods, *Adv. Neural Inf. Process. Syst.* **6**, 551 (1993).
- [76] W. S. Geisler, Visual perception and the statistical properties of natural scenes, *Annu. Rev. Psychol.* **59**, 167 (2008).
- [77] E. G. Wu, N. Brackbill, C. Rhoades, A. Kling, A. R. Gogliettino, N. P. Shah, A. Sher, A. M. Litke, E. P. Simoncelli, and E. Chichilnisky, Fixational eye movements enhance the precision of visual information transmitted by the primate retina, *Nat. Commun.* **15**, 7964 (2024).
- [78] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA, 1997).
- [79] J. A. Patrick, N. W. Roach, and P. V. McGraw, Motion-based super-resolution in the peripheral visual field, *J. Vision* **17**, 15 (2017).