# A computational observer model of spatial contrast sensitivity: Effects of photocurrent encoding, fixational eye movements, and inference engine

**Nicolas P. Cottaris** Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

**Brian A. Wandell** Department of Psychology, Stanford University, Stanford, CA, USA

**Fred Rieke** Department of Physiology & Biophysics, University of Washington, WA, USA

**David H. Brainard** Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

**We have recently shown that the relative spatial contrast sensitivity function (CSF) of a computational observer operating on the cone mosaic photopigment excitations of a stationary retina has the same shape as human subjects. Absolute human sensitivity, however, is 5- to 10-fold lower than the computational observer. Here we model how additional known features of early vision affect the CSF: fixational eye movements and the conversion of cone photopigment excitations to cone photocurrents (phototransduction). For a computational observer that uses a linear classifier applied to the responses of a stimulus-matched linear filter, fixational eye movements substantially change the shape of the CSF by reducing sensitivity above 10 c/deg. For a translation-invariant computational observer that operates on the squared response of a quadrature-pair of linear filters, the CSF shape is little changed by eye movements, but there is a two fold reduction in sensitivity. Phototransduction dynamics introduce an additional two fold sensitivity decrease. Hence, the combined effects of fixational eye movements and phototransduction bring the absolute CSF of the translation-invariant computational observer to within a factor of 1 to 2 of the human CSF. We note that the human CSF depends on processing of the retinal representation by many thalamo-cortical neurons, which are individually quite noisy. Our modeling suggests that the net effect of post-retinal noise on contrast-detection performance, when considered at the neural population and behavioral level, is quite small: The inference mechanisms that determine the CSF, presumably in cortex, make efficient use of the information carried by the cone photocurrents of the fixating eye.**

## Introduction

The spatial contrast sensitivity function (CSF) is a fundamental characterization of human vision: It specifies the amount of contrast required for a visual system to detect sinusoidal contrast modulation at different spatial frequencies. The human CSF is single-peaked, increasing with spatial frequency to about 3 to 5 c/deg (cycles per degree) and then declining steadily until the resolution limit, near 60 c/deg for a very well-refracted eye (Robson, 1966; Campbell Robson, 1968; Kelly, 1977). The falling limb of the human CSF is parallel to that of an ideal observer who makes optimal use of the information carried by the excitations of the cone photoreceptors in a model foveal retinal mosaic (Banks, Geisler, & Bennett, 1987). This alignment indicates that blurring by the eye's optics and cone apertures play an important role in limiting human contrast sensitivity. The absolute sensitivity of the ideal observer CSF, however, greatly exceeds that of human observers. This leads to the question of what visual mechanisms, not included in the ideal observer calculations, account for the lower sensitivity.

The ideal observer uses a decision rule that requires exact knowledge of the visual stimulus. Our recent work (Cottaris, Jiang, Ding, Wandell, & Brainard, 2019) relaxes this assumption by modeling an observer that learns the decision rule from labeled stimulus-response data. Indeed, it is to highlight this difference that we use the term *computational observer* (Farrell et al.,

Citation: Cottaris, N. P., Wandell, B. A., Rieke, F., & Brainard, D. H. (2020). A computational observer model of spatial contrast sensitivity: Effects of photocurrent encoding, fixational eye movements, and inference engine. *Journal of Vision*, *20*(7):17, 1–25, https://doi.org/10.1167/jov.20.7.17.

https://doi.org/10.1167/jov.20.7.17 Received September 5, 2019; published July 21, 2020 ISSN 1534-7362 

2014; Jiang et al., 2017; Cottaris et al., 2019). Using open-source and freely available software (ISETBio), we confirmed the essential features of the classic ideal observer results. The software extends the ideal observer work, allowing us to explore how variations in optics, cone mosaic structure, and choice of learned decision model (*inference engine*) affect the spatial CSF (Cottaris et al., 2019). Accounting for these factors, which affect the encoding of every visual stimulus, did not change the shape of the falling limb of the computational observer CSF. There remained, however, a 5- to 10-fold difference in absolute sensitivity between the computational and human CSFs.

The present article extends the analysis further into the visual system. We incorporate two additional factors into the ISETBio simulations that influence visual encoding: (a) spatial uncertainty introduced by fixational eye movements and (b) sensitivity regulation and noise introduced by the conversion of cone photopigment excitations into cone photocurrent. Fixational eye movements, which include slow drifts and microsaccades, translate the retinal image with respect to the cone mosaic and introduce spatial stimulus uncertainty. Drifts translate the retinal image along Brownian motion-like curved paths that change direction frequently, with instantaneous mean velocities in the range of 30 to 90 arc min/s (Cherici, Kuang, Poletti, & Rucci, 2012). Microsaccades occur between periods of drift, every 500 to 2000 ms, and induce very fast retinal image translations with speeds in the range of 4 to 100 deg/s (Martinez-Conde et al., 2009). Inference mechanisms in the visual system must confront this uncertainty, which may be beneficial for certain visual processes (Martinez-Conde et al., 2006; Engbert, 2006) or not (Kowler Steinman, 1979). Although our model of fixational eye movements includes the ability to model both drift and microsaccades, here we only examine the effect of drift, as the modeled stimulus duration is sufficiently short (100 ms) that microsaccades rarely occur.

The conversion of cone photopigment excitations into cone photocurrent (phototransduction) introduces nonlinear amplification and compression of the cone excitation signal (Endeman & Kamermans, 2010) and additive noise that is largely stimulus-independent (Angueyra & Rieke, 2013). As mean light level increases, the effect of these two factors exceeds the uncertainty caused by Poisson noise in the cone photopigment excitations. Additional effects are also introduced by phototransduction, such as background-dependent changes in the temporal dynamics of the cone photocurrent response and asymmetries between increments and decrements (Endeman & Kamermans, 2010; Angueyra, 2014).

To foreshadow our main result, for the inference engines we considered, the addition of fixational drift and phototransduction into the analysis pipeline closes much of the overall gap between computational and human CSFs while mostly retaining the agreement in the shape of sensitivity falloff as spatial frequency increases. We say mostly, because the combined effect of fixational eye movements and phototransduction has a small dependence on spatial frequency. Thus, although the CSF of a human subject depends on processing by many thalamic and cortical neurons, which are individually quite noisy, our computational modeling suggests that the net effect of this noise on contrast detection, when considered at the neural population level, is quite small: The inference mechanisms that determine the CSF, presumably in cortex, make efficient use of the information available from the cone photocurrents of the fixating eye.

## Overview of computational model of early vision

Evaluating the significance of a wide array of visual system factors requires computational modeling. To meet this challenge, we are developing ISETBio (Farrell et al., 2014; Jiang et al., 2017; Cottaris et al., 2019) and related software packages (Lian et al., 2019) as open-source software resources.[1,2,3] The software enables specification of visual scene radiance (including both three-dimensional scenes and stimuli presented on planar displays), modeling the transformation of scene radiance through the eye's optics to the retinal image, calculation of photopigment excitations in the retinal cone mosaic, modeling of phototransduction within the cones, simulation of fixational eye movements, and implementation of inference engines for relating visual representations to performance on psychophysical tasks. The computational pipeline for scenes represented on planar displays through to the level of cone photopigment excitations is described in detail elsewhere (Cottaris et al., 2019). The work here describes extensions to include fixational eye movements, which translate the retinal image over the cone mosaic, and a model of phototransduction that converts quantal cone photopigment excitation events into current flow through the cone outer segment membrane.

An overview of ISETBio's computational pipeline is depicted in Figure 1. In the present study, visual stimuli are specified initially as display RGB values (Figure 1A) and then transformed into radiance maps at a set of wavelengths (Figure 1B). This transformation is based on display calibration information, including the spectral radiance emitted by each of the display primaries (Figure 1AB). The gray-scale image stack in Figure 1B represents the emitted spatial radiance maps at different wavelengths (multispectral scene), and

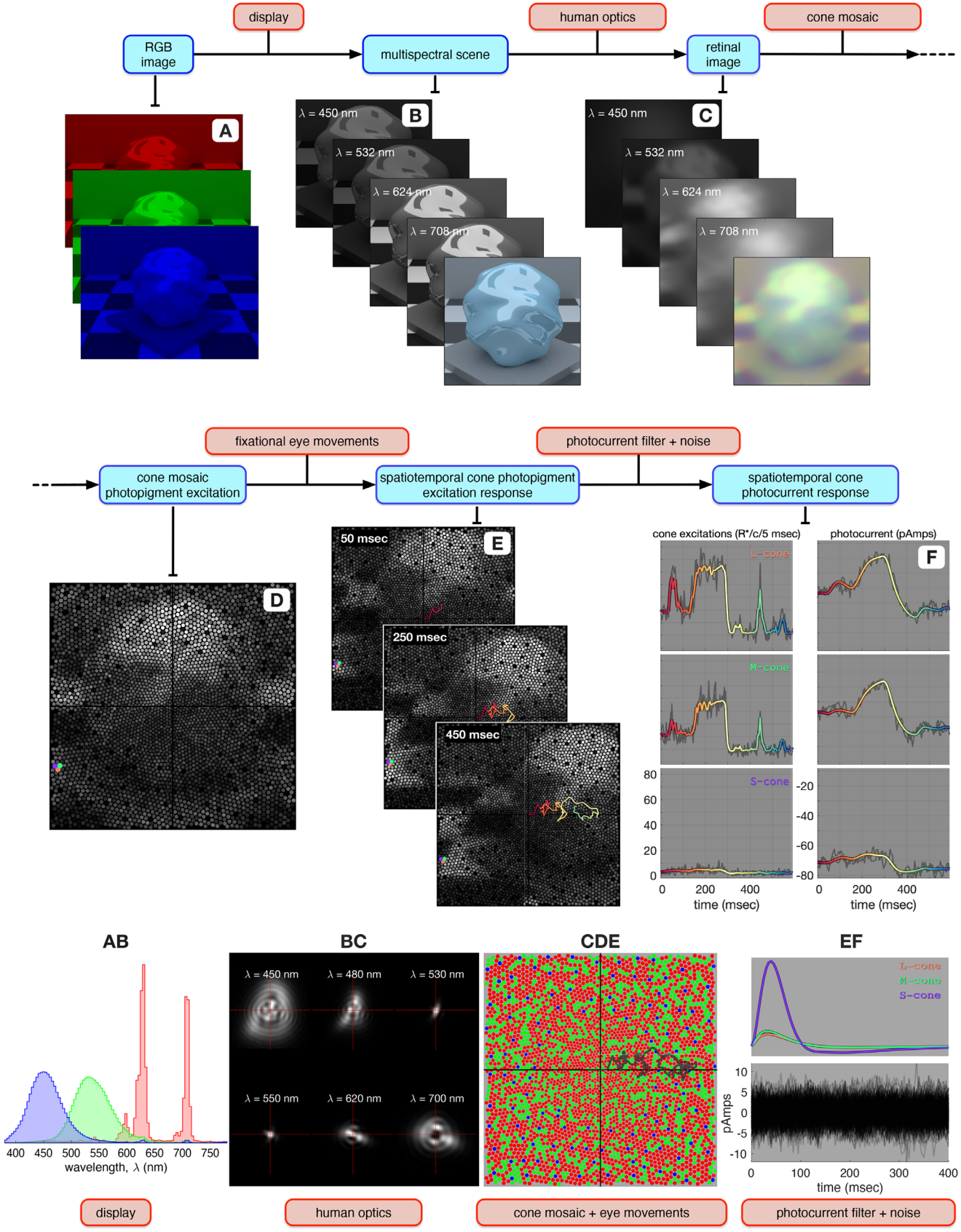


Figure 1. Flowchart of ISETBio computations. (A) The visual stimulus, here an image on an RGB display. (B) The corresponding multispecral scene, consisting of a spectral stack of spatial radiance maps at the sample wavelengths. The image at the front of the stack represents an RGB rendition of the scene. The A → B computation uses the spectral power distributions of the display primaries

→

←
(panel AB). (C) The corresponding retinal image for a single fixation location, consisting of a spectral stack of spatial irradiance maps at the sample wavelengths. The image at the front of the stack represents an RGB rendition of the retinal image. The yellow tint is due to spectral filtering by the lens. The B → C computation blurs the scene radiance using a set of shift-invariant point spread functions (panel BC), typical of human foveal vision. (D) The corresponding cone mosaic photopigment excitation response, also at a single fixation location. The pink, cyan, and magenta disks highlight an L-, an M-, and an S-cone, respectively. The C → D computation uses an eccentricity-based cone mosaic model (panel CDE) in which L- (red disks), M- (green disks) and S- (blue disks) cones spectrally integrate and spatially sample the retinal irradiance. (E) Spatiotemporal cone photopigment excitation response depicted as a temporal stack of cone mosaic photopigment response maps. The temporal dynamics introduced in the D → E computation are due to fixational eye movement paths generated via a human ocular drift model, which translates the retinal image with respect to the cone mosaic. The color coded line superimposed in panel E depicts eye position during a 600-ms long fixational eye movement trajectory, with color changing from red to yellow to green to blue, as time increases from 0 to 600 ms. (F) Conversion of noise-free photopigment excitation responses during the fixational eye movement trajectory (color-coded lines in the left panels) to noise-free photocurrent responses (color-coded lines in the right panels) for the three selected cones. This computation uses biophysically-derived photocurrent impulse responses which depend on the mean cone excitation level (top panel of EF). Gray lines in the left and right panels depict noisy instances of the corresponding cone photopigment and cone photocurrent responses. 256 instances of photocurrent noise are depicted in the bottom panel of EF.

the image at the front of the stack depicts an RGB rendition of the scene.

Maps of the spatial spectral irradiance incident on the retina, or the retinal image,[4] are calculated from the scene data using ISETBio methods that employ a model of the human eye's optics (Cottaris et al., 2019). This model accounts for factors such as on-axis wavefront aberrations, which determine a set of wavelength-dependent, shift-invariant point spread functions (Figure 1BC), pupil size, and wavelength-dependent transmission through the crystalline lens. The stacked gray-scale images in Figure 1C represent retinal irradiance at different wavelengths, and the image at the front of the stack depicts an RGB rendition of the retinal image.

For a single fixation location, the spectral irradiance on the retina is transformed into a static spatial pattern of cone photopigment excitation responses using cone mosaic methods that employ an eccentricity-varying cone mosaic model (Cottaris et al., 2019). This model accounts for the relative number of L, M, and S cones; cone spacing; inner segment aperture size and outer segment length; cone photopigment density; and macular pigment density, all varying with retinal eccentricity (Figure 1CDE). The photopigment excitation response map across the cone mosaic in response to the retinal image of Figure 1C is depicted in Figure 1D, for a single fixation location.

A time-varying spatiotemporal pattern of cone photopigment excitation responses is obtained using a model of fixational eye movements that takes into account the dynamics of human fixational drift as described by Engbert & Kliegl (2004) and is described in detail in the Methods (Modeling fixational eye movements). ISETBio cone mosaic methods translate the retinal image with respect to the cone mosaic along the eye movement trajectories. This generates a temporal sequence of cone photopigment excitation responses, three frames of which are depicted in the stacked plot of Figure 1E. The superimposed color-varying line depicts the eye movement trajectory up to the time of each frame (red-yellow-green-blue as time increases from 0 to 600 ms), and the full trajectory during a simulated 600-ms period is depicted by the gray line in Figure 1CDE.

Noise-free, cone photopigment excitation responses are transformed into noise-free cone photocurrent responses by temporal convolution with the cone photocurrent impulse response. Photocurrent impulse response functions are derived using a biophysically based model of phototransduction, which is based on work by Angueyra (2014) and is described in detail in the Methods (Modeling photocurrent generation). The gain and temporal dynamics of these impulse response functions depend on the background cone excitation level. For the stimulus depicted in Figure 1, there are large differences between the L-/M- and S-cone photocurrent impulse responses (Figure 1EF) due to the much weaker background excitation of S-cone photopigment by the stimulus. This results in a weaker S-cone adaptation level and therefore a higher S-cone photocurrent gain. The colored lines in the left panels of Figure 1F depict the noise-free cone photopigment excitation responses of three highlighted neighboring cones: an L cone in the top plot, an M cone in the middle plot, and an S cone in the bottom plot. The corresponding noise-free photocurrent responses are depicted by the color-coded lines in the right panels of Figure 1F. Note that temporal integration of cone photopigment excitation responses greatly attenuates fast transients (e.g., compare L and M cone traces in left and right panels of Figure 1F at around 450 ms), and that the strong S-cone photocurrent impulse response amplifies the mean photocurrent responses of S cones

relative to their photopigment excitation responses. In the final step of the simulations, photocurrent noise with Gaussian amplitude distribution and temporal spectral density matched to those of photocurrent responses in primate cones (Angueyra & Rieke, 2013) is added to the noise-free photocurrent responses; this generates the noisy photocurrent response instances depicted as gray lines in the left panel of Figure 1F.

# Results

## Cone mosaic response dynamics

On each trial in a two-interval forced-choice spatial contrast sensitivity experiment, the subject is presented with two stimulus intervals: One contains a spatially uniform pattern (null stimulus), and one contains a sinusoidal grating pattern (test stimulus). The subject reports which interval contains the test. For each spatial frequency, contrast varies across trials and percent correct is measured as a function of contrast. From such data, the contrast corresponding to a criterion percent correct is taken as detection threshold, and sensitivity is given by the reciprocal of threshold contrast.

Figure 2 depicts examples of cone photopigment excitation and photocurrent mosaic responses to a 16 c/deg, 100% contrast, 100-ms test stimulus, with a mean luminance of 34 cd/m$^2$. The mosaic's mean cone photopigment excitation response map in the absence of fixational eye movements is shown in Figure 2A. Mean cone excitation responses increase with eccentricity (brighter values) because cone aperture increases with eccentricity. Figure 2B depicts four instances of fixational eye movement trajectories, each lasting for 150 ms. Different eye movement trajectories start at random locations, but the trajectories are constrained so that their centroids are all at the origin.

Figures 2C0–C4 depict differential[5] spatiotemporal cone photopigment excitation responses for cones lying along the horizontal meridian of the mosaic during different single noisy response instances, and the corresponding spatiotemporal cone photocurrent responses are depicted in Figures 2D0–D4. The responses depicted in Figures 2C0 and Figure 2D0 were obtained in the absence of fixational eye movements. A clear spatiotemporal modulation can be seen during the stimulus presentation duration (0–100 ms) in the cone photopigment response. The stimulus-induced spatiotemporal modulation is somewhat blurred over time in the cone photocurrent response, and the overall modulation is more noisy than that in the cone photopigment excitation response.

The differential spatiotemporal cone photopigment excitation responses depicted in Figures 2C1–2C4 were obtained for the four fixational eye movement trajectories depicted in Figure 2B. Note the jitter and clear spatiotemporal response modulation that is due to the translation of the retinal image along the horizontal axis and that the corresponding photocurrent

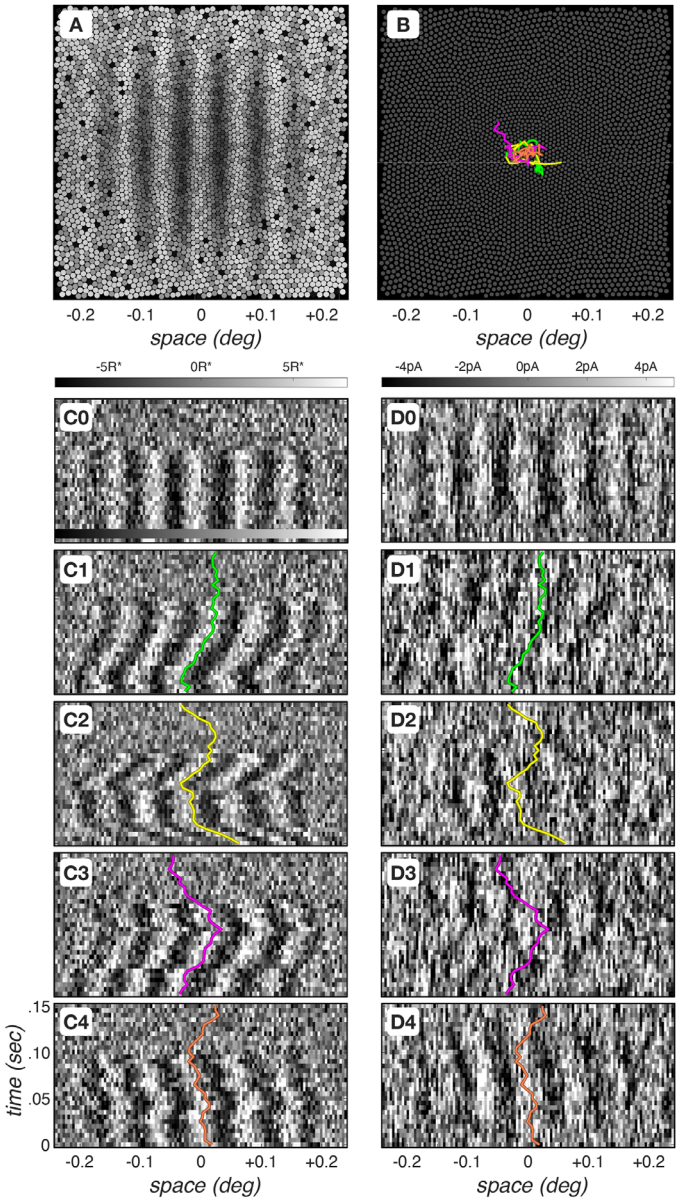


Figure 2. Spatiotemporal dynamics of cone mosaic photopigment excitation and photocurrent responses. (A) Mean cone photopigment excitation mosaic response to a 16-c/deg, 100% contrast, 34- cd/m$^2$ mean luminance grating, flashed for 100 ms. Excitation level is depicted by the gray scale value. S-cones are weakly excited and appear black. This is primarily due to selective absorption of short-wavelength light by the lens and macular pigment. (B) Four instances of fixational eye movements, each computed for a period of 150 ms. (C0) Differential (test–null) spatiotemporal response maps

→

responses, depicted in Figures 2D1–2D4, have weaker stimulus-induced spatiotemporal modulations. Two factors contribute to this. First, convolution of the jittered cone photopigment excitation response with the photocurrent impulse response smears responses in time. Second, the photocurrent response signal-to-noise ratio (SNR) is lower than the SNR of the cone photopigment excitation response,[6] as can be seen by comparing Figures 2C0 and 2D0.

## Impact of fixational eye movements

We begin our computational assessment of contrast sensitivity by examining the impact of fixational drift at the level of cone photopigment excitations. The gray disks in Figure 3 depict the contrast sensitivity function (CSF) in the absence of fixational eye movements. Performance is estimated using a computational observer that employs a linear support vector machine (SVM) classifier operating on the output of a stimulus-matched spatial pooling filter (template), which linearly sums cone responses over space at every time instant (Cottaris et al., 2019) and which we term the *SVM-Template-Linear* computational observer. This CSF serves as a baseline for assessing the impact of fixational eye movements.

The CSF computed in the presence of drift fixational eye movements using the same SVM-Template-Linear observer is depicted by the red disks in Figure 3A. Note the dramatic loss in sensitivity as spatial frequency exceeds 10 c/deg. Indeed, a threshold cannot be obtained beyond 24 c/deg. For this computational observer, fixational eye movements cause significant misalignment between the retinal image and the observer's stimulus-matched filter. This decreases the SNR of the observer's filter response in a spatial-frequency-dependent manner, leading to the rapid falloff in classifier performance with increasing spatial frequency.

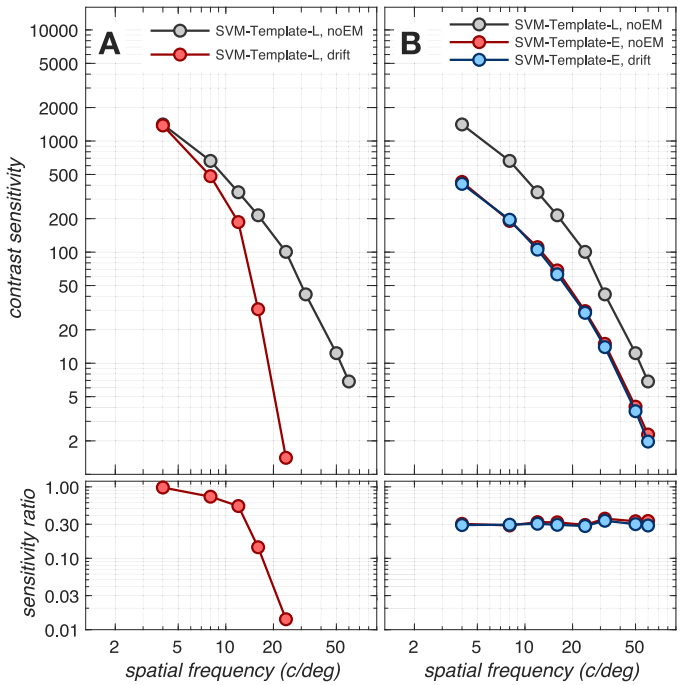


Figure 3. Impact of fixational eye movements. Top panels: Contrast sensitivity functions (CSFs) at the level of cone photopigment excitations, computed for a 3 mm pupil, typical subject wavefront-based optics, and eccentricity-based cone mosaics (Cottaris et al., 2019). Bottom panels: Ratios of CSFs with respect to the reference CSF, which is computed using the SVM-Template-Linear observer in the absence of fixational drift (gray disks). (A). CSFs computed using the SVM-Template-Linear computational observer in the absence (gray disks) and presence (red disks) of fixational drift. (B) CSFs computed using the SVM-Template-Energy observer in the absence (red disks) and presence (blue disks) of fixational drift. Gray disks in panel B replicated from panel A.

A computational observer that is less susceptible to the effects of retinal image jitter can be constructed by employing a pair of stimulus-matched spatial pooling filters that have a spatial quadrature relationship (see Appendix Figure A2, panels B1 and B2) and whose outputs are squared (Greene et al., 2016). We term this observer the *SVM-Template-Energy* computational observer. This computation, often referred to as an energy computation, was introduced in the literature to explain retinal and cortical neuron responses that are independent of stimulus spatial phase within the neuron's receptive field (Hochstein & Shapley, 1976; Emerson, Bergen & Adelson, 1992; Ohzawa et al., 1990). Our SVM-Template-Energy computational observer applies a linear SVM classifier to the energy responses. A similar approach was used by Kupers et al. (2019), who employed linear SVM classifiers operating

←

of photopigment excitation for cones positioned along the horizontal meridian, in the absence of fixational eye movements. Each column represents a single cone. Gray-scale color bar denotes excitation level in $R^* \times (\text{cone})^{-1} \times (5\ \text{ms})^{-1}$. (D0) Differential spatiotemporal response maps of photocurrent excitation for cones positioned along the horizontal meridian, also in the absence of fixational eye movements. Gray-scale color bar denotes response level in pico Amperes (pA). (C1–C4) Differential spatiotemporal cone photopigment excitation maps during the four fixational eye movement trajectories depicted in panel B. Colored lines depict the horizontal component of the corresponding eye movement trajectory. Note that time increases from bottom to top of each panel (D1-D4). Differential spatiotemporal cone photocurrent excitation response instances during the four fixational eye movement trajectories depicted in panel B.

on the Fourier power spectra of the two-dimensional time-varying cone absorption responses.

The CSFs derived using the SVM-Template-Energy computational observer in the absence and presence of fixational eye movements are depicted in Figure 3B by the red and blue disks, respectively. Note that the SVM-Template-Energy derived CSFs are nearly identical in the presence and absence of fixational eye movements, demonstrating that the sharp performance decline with spatial frequency can be eliminated when complex cell-like spatial energy mechanisms are used. This performance improvement at high spatial frequencies comes at a cost, however: an overall sensitivity drop by a factor of 2.5–3.0 across the entire frequency range independent of whether fixational eye movements are present or not (blue and red disks), as seen by comparison with the SVM-Template-Linear observer in the absence of eye movements (gray disks).

## Impact of phototransduction

Next, we examined how phototransduction impacts contrast sensitivity. To isolate performance changes due to phototransduction alone, we computed CSFs in the absence of fixational movements using the SVM-Template-Linear observer. The results are depicted in Figure 4. Note that the transformation in stimulus representation from cone photopigment excitations to cone photocurrents reduces contrast sensitivity by a factor of 2.0–2.5, with slightly more reduction at lower spatial frequencies. This spatial frequency effect is due to the increased down regulation of photocurrent response gain at more eccentric retinal locations, where the cone excitation response is stronger due to the enlarged cone aperture diameters. More eccentric retinal locations come into play because the experiment we are modeling employed a fixed number of grating cycles, resulting in larger stimulus extents at lower spatial frequencies.

Note that sensitivity loss at the cone photocurrent stage depends on the mean luminance of the background, the pupil size, and the stimulus duration. For example, as stimulus mean luminance increases, there is sensitivity loss at the level of photocurrent compared to excitations, due to increased downregulation of photocurrent gain. This loss occurs in the presence of a constant photocurrent noise for excitation rates up to 50k–100k $R^* \times \text{cone}^{-1} \times \text{s}^{-1}$.[7] Also, cone photopigment excitation responses to short duration stimuli are attenuated more than responses to longer- duration stimuli due to temporal convolution with the photocurrent temporal impulse response. The effects of mean luminance, pupil size, and stimulus duration on the computational CSF are depicted in Appendix Figures A3, A4, and A5.

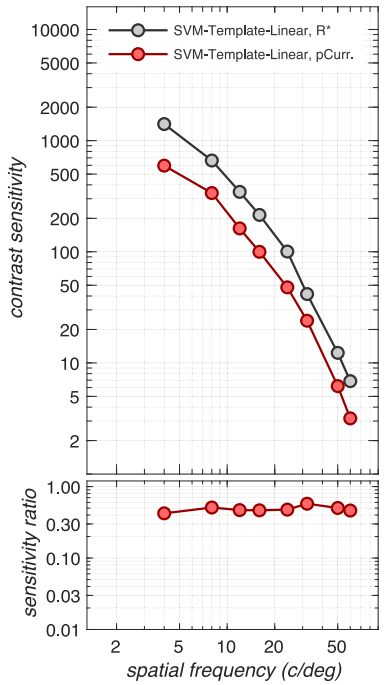


Figure 4. Impact of phototransduction. Contrast sensitivity functions computed at the level of cone photopigment excitations (gray disks) and at the level of cone photocurrents (red disks) in the absence of fixational eye movements. Here, CSFs were obtained using the SVM-Template-Linear computational observer for a 3 mm pupil, typical subject wavefront-based optics, and eccentricity-based cone mosaics. The transformation from cone photopigment excitations to cone photocurrent results in a sensitivity loss of a factor of 2 to 2.5. Note that this sensitivity loss is specific to mean light level (here 34 cd/m$^2$) and stimulus duration (here 100 ms). Stimuli presented at different adapting light levels and/or for different durations will be affected differently (see Appendix Figures A3 and A5).

## Combined effect of fixational eye movements and phototransduction

Figure 5 depicts the combined effect of phototransduction and fixational eye movements, using the SVM-Template-Linear (Figure 5A) and the SVM-Template-Energy (Figure 5B) inference engines. We previously showed that the SVM-Template-Energy computational observer is effective at mitigating the effect of fixational eye movements when applied at the level of cone excitation responses (Figure 3). When it is applied at the level of cone photocurrent responses, it is less effective, as can be seen by the small spatial-frequency-dependent performance loss above 8 c/deg (compare red disks to blue disks in Figure 5B). The reduced efficiency of the energy computation at discounting fixational jitter in high spatial frequency occurs because the photocurrent impulse response temporally integrates the spatially

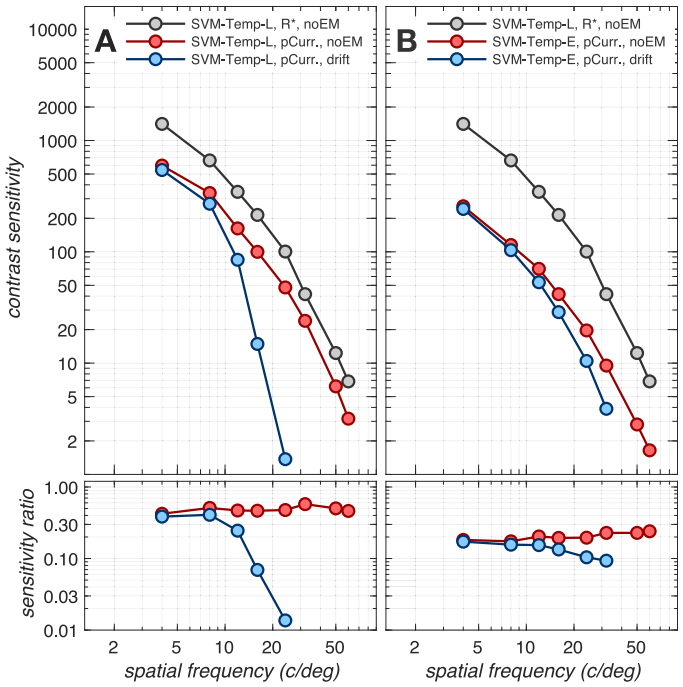


Figure 5. Combined effect of fixational eye movements and phototransduction. Contrast sensitivity functions for a 3 mm pupil, typical subject wavefront-based optics, and eccentricity-based cone mosaics. (A) CSFs computed using the SVM-Template-Linear-based computational observer. (B) CSFs obtained with the SVM-Template-Energy computational observer. Gray disks: Reference CSF computed at the level of cone excitations in the absence of fixational eye movements. Red disks: CSFs computed at the level of photocurrent in the absence of fixational eye movements. In A, red disks are replotted from Figure 4. Blue disks: CSFs computed at the level of photocurrent in the presence of fixational eye movements.

jittered cone excitation responses (see Figures 2C1–2C4, 2D1–2D4), reducing the signal to noise ratio *before* the energy computation can discount the spatial jitter.

Nonetheless, spatial pooling via the energy computation is beneficial relatively to spatial pooling via a linear computation for maintaining performance at higher spatial frequencies as can be seen by comparing the blue disks in Figure 5A to the blue disks in Figure 5B. Further, as we show in the Appendix section (Impact of spatial pooling mechanism), performance at higher spatial frequencies can be improved by using *ensembles* of energy mechanisms whose centers are spatially offset and which tile a region larger than the spatial extent of the stimulus (see Appendix Figure A2).

Overall, our results indicate that the combined effects of photocurrent encoding, fixational eye movements, and the energy-based computational observer reduce performance by a factor of 5 to 10 compared to the performance at the level of cone photopigment excitations alone. The reduction is largely independent of spatial frequency, with the spatial frequency dependence caused by the temporal integration of spatially jittered cone photopigment excitation signals during phototransduction.

## Comparison of computational and human observer performance

To compare our computational-observer contrast sensitivity functions to human psychophysical sensitivity (Banks et al., 1987), we repeated the simulations for a 2 mm pupil to match the psychophysics. The results are depicted in Figure 6. Here, the CSF at the level of cone photopigment excitations in the absence of fixational eye movements, computed using the ideal (Poisson, analytically based) observer inference engine, provides a reference (gray disks). The performance at the cone photopigment excitations but now assessed using our computational, SVM-Template-Linear observer is depicted by the red disks. Performance at the level of cone photocurrent, also in the absence of fixational eye movements using the SVM-Template-Linear observer, is depicted by the blue disks. Performance at the level of cone photocurrent in the absence of fixational eye movements, but now computed using the SVM-Template-Energy observer, is depicted by the green disks. Finally, performance at the level of photocurrent in the presence of fixational eye movements, computed using the SVM-Template-Energy observer, is depicted by the magenta disks. The performance of two human observers measured by Banks et al. (1987) is depicted by the gray triangles.

Note that at low to mid spatial frequencies (4–16 c/deg), the performance of the computational observer is within a factor of about 2 of the human observer performance. As spatial frequency increases beyond 16 c/deg, performance of the computational observer drops somewhat more rapidly than that of the two human observers of Banks et al. (1987), so that the computational and human CSFs are coming into agreement for spatial frequencies above about 20 to 30 c/deg. Thus, at the higher spatial frequencies, the early vision factors we model here account for the absolute contrast sensitivity of the human observers.

The middle panel of Figure 6 depicts CSF ratios with respect to the reference CSF (gray disks), that is, accumulated loss due to all processing stages up to the current stage, whereas the bottom panel of Figure 6 depicts CSF ratios with respect to the CSF computed at the previous processing stage, that is, performance loss at each processing stage. So in the bottom panel, red disks depict performance reduction due to having to learn the noise statistics (analytical vs. template-based computational observer), which results in the computational observer performance hovering between 80% and 90%[8] of the ideal observer performance. Blue disks depict loss at the cone photocurrent generation

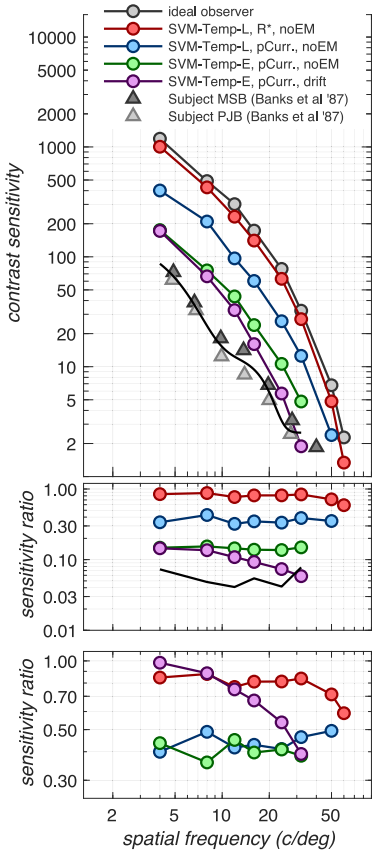


Figure 6. Comparing human and computational CSFs. Top panel: spatial CSFs computed for a 2 mm pupil, matching the psychophysical conditions of Banks et al. (1987). The curves through the different-colored disks show estimated CSFs that separate out the effects of a variety of physiological factors. The gray disks depict the CSF at the level of cone photopigment excitations in the absence of fixational eye movements using the ideal observer, analytically based inference engine, which has exact knowledge of the stimulus and Poisson noise statistics. This CSF serves as the reference CSF. Red disks depict the CSF also at the level of cone photopigment excitations and in the absence of fixational eye movements but now derived using the SVM-Template-Linear computational observer. Blue disks depict the CSF at the level of cone photocurrent in the absence of eye movements, also using the SVM-Template-Linear observer. Green disks depict the CSF also at the level of cone photocurrent and in the absence of eye movements but now using the SVM-Template-Energy observer. Magenta disks depict the CSF at the level of photocurrent in the presence of fixational eye movements using the SVM-Template-Energy observer. Triangles depict the CSFs measured in two subjects by Banks et al. (1987). Middle panel: CSF ratios with respect to the reference CSF (gray disks). These data capture accumulated loss due to all processing stages up to the current one. Bottom panel: CSF ratios with respect to the CSF computed at the previous processing stage. These data capture sensitivity loss due to different processing stages: red disks: loss due to employing a computational observer, blue disks: loss due to photocurrent generation, green disks: loss due to using an energy computational observer, magenta disks: loss due to fixational eye movements. The photocurrent and the energy inference engine each contribute to a 2.0- to 2.5-fold sensitivity loss across the spatial frequency range. Fixational eye movements contribute an additional loss which ranges from a factor of 1.0 to a 2.5 as spatial frequency increases from 4 to 32 c/deg.

→

stage, which results in a performance that is around 40% to 50% of the performance at the level of cone photopigment excitations. Green disks depict loss due to the energy computational observer, which results in a performance around 40% of the performance of the linear computational observer. Finally, magenta disks depict loss due to fixational eye movements, which results in a performance that ranges from 100% to 40% of the performance obtained in the absence of fixational movements, as spatial frequency increases from 4 to 32 c/deg.

More elaborate inference engines, for example, using an ensemble of energy spatial pooling mechanisms, can mitigate some of the performance loss in the high spatial frequency regime due to fixational eye movements (see Appendix section, Impact of spatial pooling mechanism). Also note that a good agreement in performance between our computational observer and that of human observers was found in a separate set of simulations where we computed CSFs for a variety of stimulus sizes (see Appendix section, Impact of spatial summation).

# Discussion

## Benefits and drawbacks of fixational eye movements

In the absence of fixational eye movements, post-receptoral processes render real human observers functionally blind to stationary objects (Riggs et al., 1953), and recent studies have demonstrated that fixational eye movements can improve the precision of vision at intermediate (10 c/deg) spatial frequencies (Rucci et al., 2007) and near the resolution limit (Ratnam et al., 2017). In contrast, our work suggests that fixational eye movements reduce sensitivity at high spatial frequencies. Because of this contrast, the relation between our work and theoretical work that has considered the role of fixational eye movements is worth discussion.

Rucci and colleagues (Rucci et al., 2007; Kuang et al., 2012; Boi, Poletti, Victor, & Rucci, 2017), observed that fixational eye movements reformat the spatiotemporal power spectrum of the stimulus. Under certain assumptions about post-receptoral processing,

this reformatting can be beneficial in terms of the transfer of information between the stimulus and subsequent post-receptoral mechanisms. Their theory is based on the observation that the temporal dynamics of ocular drift redistributes the $1/f^2$ power of static natural images by progressively boosting power at spatial frequencies up to 10 to 15 c/deg at nonzero temporal frequencies. This redistribution can improve detection performance for high spatial frequency targets if subsequent detection mechanisms are bandpass tuned to intermediate temporal frequencies (Boi et al., 2017). It can also ease the problem of decorrelating visual signals for efficient coding by capacity-limited post receptoral mechanisms.

In contrast, our computational observer simulations do not show that fixational drift enhances performance at high spatial frequencies. The difference between our observations (fixational eye movements reduce performance at high spatial frequencies) and those of Rucci and colleagues (fixational eye movements improve performance at high spatial frequencies) probably arises because of a few key differences between the two studies.

First, we explicitly model spatio-spectral low-pass spatial filtering by the eye's optics, isomerization noise, and the dynamics of phototransduction. These factors are not taken into consideration in the computation of the spatiotemporal power distribution by Rucci et al. (2007; Kuang et al., 2012; Boi et al., 2017). Second, their calculations are based on the pixel-level representation of the stimulus. Importantly, the low-pass temporal filtering embodied in the conversion of cone excitations to cone photocurrent is not accounted for in their analysis. This temporal filtering, when combined with fixational eye movements, reduces effective contrast at high spatial frequencies, which post-receptoral processing cannot undo. The potential advantage of reformatting the spatiotemporal power spectrum must be large enough to overcome this loss.

Third, by starting their analysis with consideration of the power spectrum, their calculations exclude phase information (Rucci et al., 2007; Kuang et al., 2012; Boi et al., 2017). Our comparisons of the SVM-Template-Linear and SVM-Template-Energy observers show that removing spatial phase information produces translation invariance in the face of fixational eye movements but comes at a cost in sensitivity when applied to the case where there are no fixational eye movements. When considering possible functional benefits of fixational eye movements, the underlying cost of applying requisite translation-invariant decision models should also be taken into account.

Fourth, we model a 100-ms stimulus because we are comparing with the data of Banks et al. (1987). During this short time interval, there is not a significant reduction in the photocurrrent response amplitude (see Figure 8E), limiting the opportunity for drift transients to enhance the response. Boi et al. (2017) report that contrast sensitivity enhancement at high spatial frequencies in the presence of fixational drift is nearly absent for stimuli presented for 100 ms and that a progressive enhancement occurs during prolonged stimulus exposure (800 ms).

Fifth, in our simulations, stimuli at progressively higher spatial frequencies have progressively reduced spatial extent, maintaining a constant number of cycles. In the studies by Rucci and colleagues (Rucci et al., 2007; Kuang et al., 2012; Boi et al., 2017), stimuli at all spatial frequencies were matched in spatial extent. Fixational eye movements may have a different effect in these two types of stimuli.

There are also differences in the modeling. We do not include limits imposed by known post-receptoral mechanisms. These are the spatiotemporal filtering applied by circuits in the retina, as well as bandwidth limits on the transmission of information between the retina and the cortex. The work of Rucci and colleagues focuses on the relation between these factors (e.g., spatiotemporal filtering by ganglion and cortical neurons) and the effect of fixational eye movements. We are planning to extend our modeling to explicitly include post-receptoral mechanisms (e.g., retinal ganglion cells), at which point it will be of interest to re-examine the effects of fixational eye movements. Modeling both the factors we consider here and those underlying the thinking of Rucci and colleagues seems likely to further clarify the costs and benefits of fixational eye movements and how they depend on what is taken as given about post receptoral processing.

A different theoretical framework for understanding the beneficial effects of fixational eye movements observed experimentally has recently been proposed by Anderson, Olshausen, Ratnam, and Roorda (2016). Their work considers inference engines that seek to simultaneously estimate both the visual stimulus and the eye movement path, thus minimizing the effects of spatial uncertainty introduced by the eye movements. They show potential advantages of fixational eye movements, particularly if cone sampling density is low relative to the spatial structure in the stimulus. In that case, the sweep of the mosaic across the retinal image samples stimuli more finely than a stationary retina. It is possible that this inference engine retains or even increases sensitivity in the presence of fixational eye movements. Such inference engines might also help explain the spatial stability of our perceptual representations in the face of fixational eye movements. Note, however, that the inference engine described by Anderson et al. (2016) requires short-term storage and cannot be implemented before the cone-bipolar synapse; such inference engines thus remain subject to the information loss caused both by optical blurring and by the temporal smearing of

spatially-jittered (due to fixational drift) cone excitation responses by the low-pass filter of phototransduction (Figure 2). Further modeling of the effect of these factors on the performance of their algorithm would be interesting.

## Other inference engines

Our conclusions about computational observer performance are tied to specific choices of inference engines to study: In principle, other inference engines could achieve better or worse performance. A more extended study of the effect of inference engines, particularly in the face of fixational eye movements, would be interesting but beyond the scope of the current article. Indeed, as noted just above, we think that an important direction for future work is to couple the study of sophisticated inference engines with model visual systems such as the one we present here, which incorporate the known limits of visual encoding. We outline some additional possible directions of investigation in the next few paragraphs. The interested reader is also directed to the discussion in Kupers et al. (2019), who developed an inference engine based on the power spectrum of the cone excitations.

As an example, inference engines based on deep convolutional networks have a lower performance loss than the SVM-Template-based inference engines we studied (Reith & Wandell, 2019). Another possibility is that performance might be increased if multiple inference mechanisms were employed in parallel. For example, at low to mid spatial frequencies (2-8 c/deg), where retinal jitter due to eye movements causes essentially no performance degradation when a linear pooling-based inference engine is used, it would be detrimental to use energy-based inference engines that have reduced overall sensitivity. At very high spatial frequencies, however, it would be beneficial to use ensembles of energy-based mechanisms (see Appendix section, Impact of spatial pooling mechanism). A hybrid inference engine that relies more on linear summation mechanisms for low to mid spatial frequency stimuli and ensembles of nonlinear summation mechanisms for high spatial frequency stimuli could be examined.

As another approach, the uncertainty introduced by fixational eye movements could be handled by the visual system using mechanisms that discard phase information in different ways than our V1 complex cell-like energy-based mechanism. For example, an inference engine might employ an ensemble of spatially shifted stimulus-matched pooling templates that, unlike our present strategy, make a decision based on a classifier operating not on the ensemble of responses from all pooling mechanisms but instead on the single maximal response across all the mechanisms, selected at each time point. Such an approach would tend to choose the mechanism whose position best overlapped with the translating stimulus at each time point and would in effect estimate the eye movement trajectory and apply that estimate to compute a time-varying spatial pooling. The general performance of uncertain observers has been examined previously (Pelli, 1985; Geisler, 2018). An alternative along these lines would be to use the statistics of the fixational eye movement trajectories to construct a signal-known statistically, maximum likelihood type of observer.

Finally, in the other direction, our inference engines employ a stimulus-matched template, so that they do not have to learn the structure of the stimulus from examples. It is possible that inference engines provided with less a priori information about the stimulus would exhibit reduced performance.

## Central visual processing stages

Our work analyzes how the information loss imposed by early stages of visual processing limits the human spatial CSF. The inclusion of nonlinear spatial summation-based inference engines already introduces post-receptoral processing elements. A full account must also include explicit modeling of the retinal and cortical circuitry (Wassle, 2004; Lennie & Movshon, 2005) and, in the case of free-viewing of natural scenes, realistic visual search strategies (Najemnik & Geisler, 2005). The close agreement between our computational observer CSF and human performance suggests that the post-receptoral circuitry efficiently transmits stimulus information from the photoreceptors to decision mechanisms. Confirming this efficiency experimentally and understanding how it is achieved remains an important goal. Recent work by Horwitz (2020) illustrates how such experimental investigation may be approached for the case of contrast detection, although differences in stimulus conditions preclude a precise comparison of our results and that work.

Computational modeling connecting cortical models to psychophysical performance is a second approach to understanding visual performance. For example, Goris et al. (2013) employed a neural population approach, which included a cortex-like stimulus encoding stage (a population of spatial frequency bandpass-tuned units), a nonlinear transducer stage (broadly tuned divisive inhibition followed by an expansive nonlinearity and additive Gaussian noise with activation-dependent amplitude), and a stimulus decoding stage (maximum likelihood inference engine). The Goris et al. (2013) model contained eight free parameters whose values were determined by fitting the model to a set of behavioral performance measurements, which included the Campbell and Robson (1968) contrast sensitivity

function. These parameters allow the model to mimic the combined information loss from multiple components of early vision, as well as any imposed by the cortex. The phenomenological model has the advantage of simplicity, but it does not separately characterize specific visual components. There is an interesting possibility of combining the two approaches, using the known early vision factors to constrain and shape the input to computational models of cortical processing.

### Caveats

We list some limitations of our modeling of early vision. First, we do not account at a fine scale for eccentricity-based changes in photocurrent sensitivity. In our present simulations, the amplitude and dynamics of the light-regulated photocurrent impulse response are determined based on the average activation of each cone type (L, M, and S) across the entire mosaic. This approximation improves computational efficiency and affects our simulations in two ways. Central cones, which have the smallest apertures, have lower excitation levels than the mean excitation level across the entire mosaic, and therefore, for these cones, the modeled photocurrent impulse response is more heavily regulated than what it would have been if their actual cone excitations were used. Since the model photocurrent noise is stimulus independent, the SNR for central cones is slightly lower than it ought to be. The opposite holds for peripheral cones: Their photocurrent responses have higher SNR than they ought.

Second, foveal cones are considerably slower than peripheral cones (Sinha et al., 2017), so the temporal integration of cone excitation signals in the presence of fixational eye movements would have less of an impact for peripheral stimulus locations than is captured by our simulations, which assume foveal cone photocurrent dynamics at all mosaic locations. The combined effect probably results in a lower overall sensitivity, and we suspect that our CSF estimates would be a little higher if we took eccentricity-based changes in photocurrent sensitivity and dynamics into account.

Third, our simulations employed stimulus-independent photocurrent noise. Work by Angueyra Rieke (2013) has shown that noise is not entirely independent of the background cone excitation level. Low-frequency noise components (1–10 Hz) are subject to the same adaptational gain reduction as the mean photocurrent response, whereas high-frequency noise components (> 100 Hz) are less affected by background light level. So in real cone mosaics, in which peripheral cones have larger apertures than central cones, the noise spectrum changes shape with eccentricity. Our simulations do not capture this effect.

Fourth, our simulations match the size of the cone mosaic to the stimulus size, which in turn varies with spatial frequency. This simplification was chosen for computational efficiency. However, this poses a problem for the highest spatial frequency stimuli, for which part of the retinal image can be brought out of the field of view of the cone mosaic in the presence of fixational eye movements. The Appendix section, Impact of spatial pooling mechanism, presents an analysis of performance for different spatial pooling mechanisms that extend beyond the stimulus spatial support.

Finally, the fixational eye movement model computes drift eye movement trajectories whose mean velocity is 60 arc min/s (Figure 7C), near the mean of the velocity distribution across a number of human observers (Cherici, Kuang, Poletti, & Rucci, 2012). However, Cherici et al. (2012) report that trained observers have significantly lower mean drift velocities (30 arc min) as opposed to naive observers, who can have mean drift velocities up to 90 arc min/s, and therefore trained observers have narrower fixation spans. The observers employed by Banks et al. (1987) were the authors themselves and were certainly well trained. Our computational observer performance would likely improve in the higher spatial frequency regime if we employed fixational eye movements whose velocity matched the low end of the distribution of velocities reported by Cherici et al. (2012).

## Summary and conclusion

We extended the ISETBio computational observer model of the human spatial contrast sensitivity to incorporate fixational eye movements and the transformation of quantal cone photopigment excitations to cone photocurrent.[9] Our analysis indicates that fixational eye movements abolish sensitivity above 10 c/deg for a computational observer that employs a stimulus-matched, linear pooling template. Energy-based computational observers eliminate the sharp performance decline at high spatial frequencies but at the cost of an overall decrease in sensitivity. The decrease in overall sensitivity in the absence of eye movements should not be surprising—to achieve translation invariance, energy-based observers ignore the stimulus spatial phase, using less stimulus information than the linear summation observer.

Phototransduction-induced sensitivity regulation and additive noise further decrease sensitivity by a factor of ∼2. Combining the effects of fixational eye movements, photocurrent encoding, and energy-based computational observers brings the computational observer performance to levels that are within a factor of 1 to 2 of human sensitivity, depending on spatial frequency. This analysis indicates that the sensitivity

loss observed in human performance relative to the sensitivity of an ideal observer operating on cone photopigment excitations in the absence of fixational eye movements can largely be accounted for by cone photocurrent encoding of spatially jittered cone excitation responses and the inference engine employed for the computational inference engines we considered. This leaves little room for additional sensitivity loss as the signal is processed by the neurons in the thalamus and cortex.

# Methods

## Modeling optics and cone mosaic excitation

ISETBio computations begin with a quantitative description of the visual stimulus, here, a spatio-temporal pattern specified as the spectral radiance emitted at each location and time on a flat screen. The spectral irradiance incident at the retina (retinal image) is computed by taking into account human optical factors such as pupil size, wavelength-dependent blur, on-axis wavefront aberrations, and wavelength-dependent transmission through the crystalline lens. Cone mosaic excitations are computed from the spectral retinal irradiance using naturalistic cone mosaics that model the relative number of L, M, and S cones; the existence and size of an S-cone free zone in the central fovea; cone photopigment density; and the variation with eccentricity in cone spacing, inner segment aperture size, outer segment length, and macular pigment density. Performance on a two-alternative forced-choice detection task is assessed using SVM-based binary classifiers that operate on the output of mechanisms that pool cone responses over space using linear and nonlinear (energy-based) stimulus-derived pooling schemes. Detailed descriptions of all elements of this computational pipeline and estimates of how different elements impact the spatial contrast sensitivity function can be found in Cottaris et al. (2019). In the following sections, we describe the ISETBio modeling of two additional elements of early vision, fixational eye movements and phototransduction, whose impact on the spatial contrast sensitivity function is examined in the present article.

## Modeling fixational eye movements

The fixational eye movement model in ISETBio includes a drift and a microsaccade component. The drift component is generated by a delayed feedback loop mechanism based on work by Mergenthaler and Engbert (2007), which generates eye movement paths using a generalized Brownian motion process. The microsaccade component injects abrupt shifts in eye position with the frequency, speed, and amplitude of the trajectories drawn from published distributions of microsaccades (Martinez-Conde et al., 2009). Saccade direction is based on heuristics that aim to maintain fixation while also avoiding recently visited positions (Martinez-Conde et al., 2006; Engbert, 2006). In the present work, which simulates presentation of a 100-ms stimulus, we only engage the drift component, as microsaccades typically occur only every 500 to 1,000 ms (Martinez-Conde et al., 2009).

### *Drift fixational eye movement generation model*

Drift eye movement trajectories during fixation of steady targets resemble generalized Brownian motion (Engbert & Kliegl, 2004). In a generalized Brownian motion process, the mean squared displacement, $\overline{\Delta p^2}$, at a time lag, $\Delta T$, relative to an arbitrary time point, $t_i$, computed as $\overline{\Delta p^2} = \overline{\|p_{i+1} - p_i\|^2} = \overline{\|p(t_i + \Delta T) - p(t_i)\|^2}$, is proportional to $\Delta T^K$, where $K$ is a real number between 0 and 2. In a purely Brownian process, the position at time step $i + 1$, $p_{i+1}$, is given by

$$p_{i+1} = p_i + \eta_i \quad (1)$$

where $p_i$ is the position at time step $i$, and $\eta_i$ is a normally distributed random variable with zero mean. In such a process, the sequence of spatial displacements is uncorrelated and $K = 1$. Using diffusion analysis, Engbert and Kliegl (2004) showed that fixational eye movements differ from pure Brownian motion, exhibiting correlations over two time scales. Over short time scales (2 to 30 ms), the eye has a tendency to continue to move in the current direction (persistent behavior), resulting in correlated displacements and $K > 1$. Over longer time scales (100 to 500 ms), there is a tendency to reverse direction (anti persistent behavior), resulting in uncorrelated displacements and $K < 1$.

Differences in eye position dynamics from those of a purely Brownian process affect fixation span and may affect performance. To simulate a combination of persistent and anti persistent dynamics, we generate fixational eye movements using the delayed random walk model proposed by Mergenthaler and Engbert (2007). In this model, the position at time step $i + 1$, $p_{i+1}$, is given by

$$p_{i+1} = p_i + w_i + \eta_i \quad (2)$$

where

$$w_i = \chi_i + (1 - \gamma) \times w_{i-1} - \lambda \times \tanh(\epsilon \times w_{i-\tau}) \quad (3)$$

In Equation 3, the autoregressive term, $(1 - \gamma) \times w_{i-1}$, generates the persistent behavior at short time scales; the negative delayed feedback term, $-\lambda \times \tanh(\epsilon \times$

| *Symbol* | *Description* | *Value* |
|---|---|---|
| $\gamma$ | Control gain | 0.25 |
| $\mu_\chi$ | Control noise $\chi$, mean | 0.000 |
| $\sigma_\chi$ | Control noise $\chi$, standard deviation | 0.075 |
| $\mu_\eta$ | Position noise $\eta$, mean | 0.000 |
| $\sigma_\eta$ | Position noise $\eta$, standard deviation | 0.35 |
| $\lambda$ | Feedback gain | 0.15 |
| $\epsilon$ | Feedback steepness | 1.1 |
| $\tau_x$ | Feedback delay ($x$ position) | 0.07 |
| $\tau_y$ | Feedback delay ($y$ position) | 0.04 |

Table 1. Values of the drift fixational eye movement model parameters used in this study.

$w_{i-\tau}$), generates the anti persistent behavior at longer time scales; and $\chi_i$ represents a Gaussian random noise process with zero mean, which provides the driving signal for $w_i$. The values of the model parameters that we used in the present study are listed in Table 1 and are taken from Mergenthaler and Engbert (2007). The model generates fixational eye movement paths with a resolution of 1 ms, with $x$ and $y$ position coordinates computed independently.

### *Dynamics of model drift fixational eye movements*

Various properties of the eye movement paths generated by the model are illustrated in Figure 7. Figure 7A illustrates the $x$ and $y$ components of 1,024 drift trajectories, each lasting for 150 ms. The contour plot in Figure 7 B depicts the fixation span computed over all those traces, which is defined as the spatial probability distribution with which the eye is at each position during the analyzed period, and the superimposed black line depicts a single eye movement trajectory. The distribution of instantaneous drift velocities, computed by taking the time derivative of the low-passed position signal (low-pass filter: 41-sample Savitzky-Golay filter), is depicted in Figure 7C. Notice that the model velocities span the range of velocities measured in human observers measured by Cherici et al. (2012) (depicted by the vertical lines). The spectra of the $x$ and $y$ position trajectories are displayed in Figure 7D, in red and blue lines, respectively. The displacement of eye position as a function of time lapsed from any time point during the trajectory is depicted by the solid line in Figure 7E. Note that for short time scales (2 to 30 ms), displacement is larger from what would be observed if the motion were purely Brownian (dashed line), whereas for longer time scales (100–500 ms), anti persistent dynamics minimize this difference.

## Modeling photocurrent generation

Our cone photocurrent response model consists of two stages. In the first stage, a biophysically based model of the phototransduction cascade transforms a time-varying sequence of photon excitation rate to a photocurrent temporal response. The model is a modified version of the canonical model of phototransduction (Pugh & Lamb, 1993; Hateren, 2005) with parameters based on population recordings from primate cone photoreceptors (Angueyra, 2014). Although we have implemented the full nonlinear conversion between time-varying excitation rate and photocurrent, the full calculation is compute-intensive because a small time step (0.1 ms) is required in order to accurately simulate the differential equations that govern the phototransduction cascade. Here we use the full calculation to determine a linear approximation that is valid for near-threshold perturbations around a mean luminance. That is, the full phototransduction model is used to compute a photocurrent impulse response function, which is defined as the outer segment membrane current in response to a cone excitation delta function superimposed on a constant cone excitation background rate. This biophysically derived impulse response function is specific to the stimulus mean cone excitation, and stimuli of different mean luminances and chromaticities will have different photocurrent impulse responses. The derived photocurrent impulse response is downsampled to the time step of the simulations, here 5 ms, and subsequently convolved with the sequence of the mean cone photopigment excitations, which are also computed every 5 ms, to derive the noise-free photocurrent response. The second stage of the model adds a stochastic component that captures noise in the phototransduction cascade. The noise has Gaussian amplitude distribution and a power spectrum that is matched to that of primate cone photoreceptors (Angueyra & Rieke, 2013). The computed photocurrent captures the characteristics of primate cone responses (Angueyra, 2014) for a range of adaptation levels, 0 to 30,000 $R^* \times cone^{-1} \times s^{-1}$. In this range, photopigment bleaching is less than 2%, assuming a half-bleaching constant of 6.4 log $R^* \times cone^{-1} \times s^{-1}$, and can therefore be ignored. The half-bleaching constant was estimated from the value of 4.3 log Trolands provided by Rushton and Henry (1968).

### *Biophysically based model of the phototransduction cascade*

In darkness, there is a constant inflow of $Na^+$ and $Ca^{+2}$ ions into the photoreceptor outer segment via cyclic guanosine monophosphate (cGMP)–gated channels, many of which are open due to the high concentration of intracellular cGMP. cGMP is constantly being produced by the enzyme guanylate cyclase (GC). The constant inflow of $Na^+$ and $Ca^{+2}$ ions into the photoreceptor creates a negative current. This negative current hyperpolarizes the cone membrane,

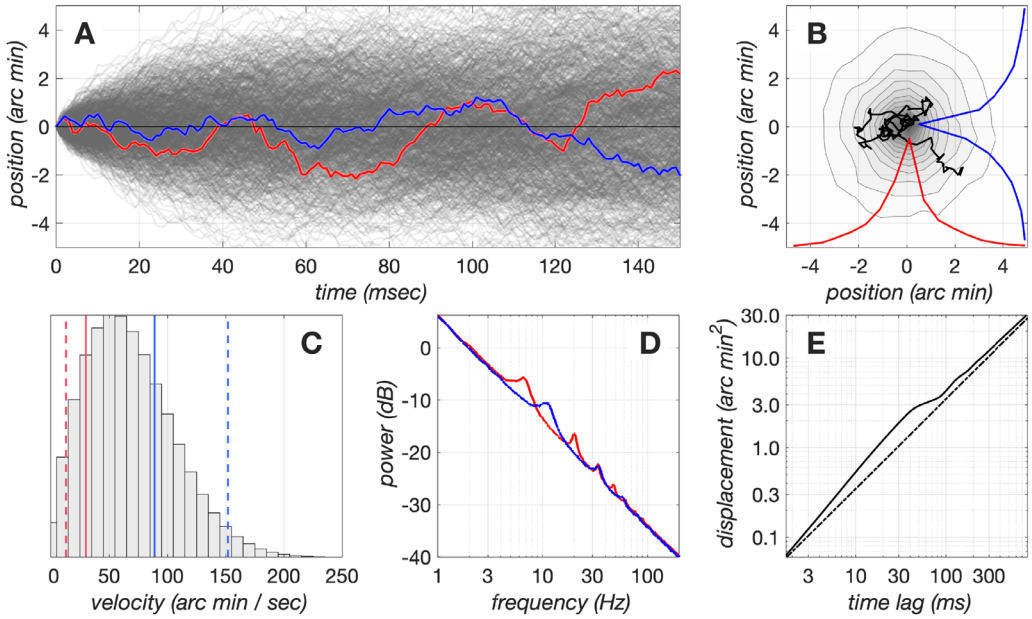


Figure 7. Dynamics of drift fixational eye movements generated by our model. (A) *x* and *y* components of a set of 1,024 eye movement trajectories during a period of 150 ms. For visualization purposes, the paths always start at (0,0), but in the simulations, each path starts at a random location, and the centroid of each path is constrained to be at (0,0). Red and blue lines denote the *x* and *y*, respectively, trajectories of a single eye movement trajectory. (B) The contour plot depicts the fixation span of the set of 1,024 eye movement paths depicted in A, and the red and blue lines depict the marginal distributions of *x* and *y* eye positions of this data set. The superimposed black line depicts the eye movement path whose *x* and *y* trajectory components are depicted in A as the red and blue traces. (C) Distribution of instantaneous velocities in the examined set of 1,024 eye movement trajectories. The red and blue solid lines depict the mean drift velocities from 2 human subjects who had the lowest and highest velocity, respectively, in a pool of 12 subjects (Cherici, Kuang, Poletti, & Rucci, 2012). The red and blue dashed lines represent $-1\sigma$ and $+1\sigma$ of the drift velocity distributions for these subjects, respectively. (D) Power spectra of the *x* and *y* drift trajectory components. The spectral peaks are due to the oscillatory behavior of the modeled delayed negative feedback mechanism, which has slightly different delays for the *x* and *y* drift components. (E) Displacement analysis of model drift eye movements. The solid line depicts the mean squared displacement, $\overline{\Delta p^2}$, as a function of time lag for the examined set of eye movement paths. The dashed line depicts $\overline{\Delta p^2}$ for a purely Brownian process. Notice the persistent behavior between 2 and 30 ms, which causes the eye to diffuse more than it would have if it were under the control of a purely Brownian process, and the antipersistent behavior that stabilizes eye position at longer time delays (100 to 500 ms).

which results in a continuous release of glutamate at the synapses with bipolar and horizontal cells. When a photon isomerizes an opsin molecule, it initiates a biochemical cascade that results in the activation of multiple phosphodiesterase (PDE) enzymes. The increased PDE activity hydrolizes cGMP at a higher rate than in the dark, thereby reducing the intracellular cGMP concentration, which leads to closure of cGMP channels. This blocks the entry of $Na^{+}$ and $Ca^{+2}$ ions into the cone, causing a depolarization in the membrane and a decrease in glutamate released. Our model of this process, which captures the steps between cone photopigment excitation rate and the modulation of membrane current, is illustrated in Figure 8. The implementation of the different stages is as follows.

***Opsin activation:*** Absorption of photons by photopigment molecules (Figure 8A) turns inactive opsin proteins, $R$, into their activated state, $R^*$. Activated opsin molecules are produced instantaneously with a rate that is proportional to the photon absorption rate, $A(t)$, and inactivated with a rate constant, $\rho_R$ (Figure 8B). When the light intensity is such that $A(t) < 30{,}000$ photons/cone/s, we can neglect photopigment bleaching and treat the concentration of inactive photopigment as a constant. In this regime, the production of activated opsin, $R^*(t)$, is described by

$$\frac{dR^*(t)}{dt} = g_R \cdot A(t) - \rho_R \cdot R^*(t) \quad (4)$$

where $g_R$ is a scaling constant.

***PDE concentration:*** PDE enzymes are in turn activated by activated opsin proteins with a rate $R^*(t) + \rho_E^{dark}$, where $\rho_E^{dark}$ is the spontaneous PDE activation rate in the dark, and become inactivated

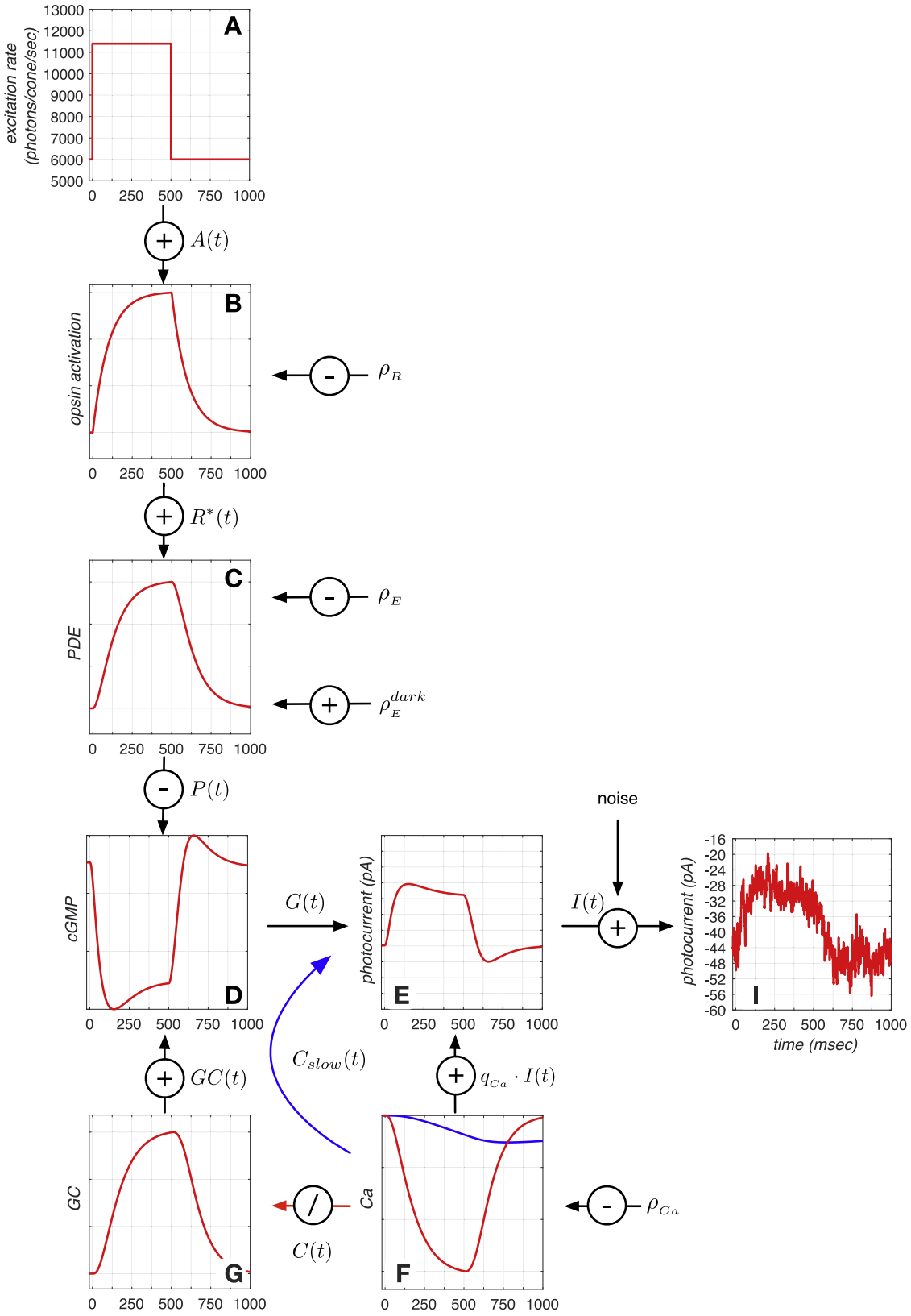


Figure 8. Phototransduction cascade model. Depicted here are the responses of the different model components to a 500 ms long light increment pulse. (A) Sequence of mean cone photopigment excitation rate. (B) Opsin activation in response to the change in cone photopigment excitation rate. (C) Activation of PDE enzymes in response to opsin activation. (D) Change in cGMP concentration, which is synthesized by CG and broken down by PDE enzymes. (E) Cone membrane current, photocurrent, which is an instantaneous function of the cGMP activation. (F) Change in intracellular $Ca^{+2}$ concentration, which is the result of ion inflow via the membrane

→

←

current and ion outflow via the $Na^{+}-Ca^{+2}$ exchanger pump. Two feedback mechanisms, both based on the intracellular $Ca^{+2}$ concentration, modify the membrane current. A slow $Ca^{+2}$ -derived signal directly regulates the membrane current (blue line), and an indirect signal (red line) modulates production of GC, shown in G, which is responsible for producing cGMP. (I) Noisy membrane current response instance generated by adding photocurrent noise to the mean membrane current response depicted in E. See text for more details.

with a rate constant, $\rho_E$ (Figure 8C). The production of activated PDE, $E(t)$, is described by

$$\frac{dE(t)}{d\tau} = R^{*}(t) + \rho_E^{dark} - \rho_E \cdot E(t) \quad (5)$$

***cGMP concentration:*** cGMP molecules are synthesized continuously due to the GC protein activity at a rate $GC(t)$ and hydrolized at a rate that is a product of PDE enzymatic activity, $P(t)$, and cGMP concentration, $G(t)$ (Figure 8D). The concentration of cGMP, $G(t)$, is described by

$$\frac{dG(t)}{dt} = GC(t) - P(t) \cdot G(t) \quad (6)$$

The rate at which GC is producing cGMP is an instantaneous function of the intracellular $Ca^{+2}$ concentration, $C(t)$:

$$GC(t) = \frac{GC_{max}}{1 + \left(\frac{C(t)}{k_{GC}}\right)^{n_{GC}}} \quad (7)$$

where $GC_{max}$ is the maximal production rate, $k_{GC}$ is the half-maximal $Ca^{+2}$ concentration, and $n_{GC}$ is an exponent that determines the steepness of the relation between $GC(t)$ and $C(t)$ (Figure 8G).

***$Ca^{+2}$ concentration:*** The intracellular $Ca^{+2}$ concentration, $C(t)$, depends on two factors: the $Ca^{+2}$ inflow though the open cGMP-gated outer segment membrane channels and the $Ca^{+2}$ outflow via $Na^{+}$ – $Ca^{+2}$ exchanger pumps (Figure 8F), and is described by

$$\frac{dC(t)}{dt} = q_{Ca} \cdot I(t) - \rho_{Ca} \cdot C(t) \quad (8)$$

where $q_{Ca} = \frac{2 \cdot \rho_{Ca} \cdot C_{dark}}{k_{cGMP} \cdot (G_{dark})^{n_{cGMP}}}$ is the fraction of the ionic membrane current that is carried by $Ca^{+2}$ ions, and $\rho_{Ca}$ is the rate constant at which $Na^{+}$ – $Ca^{+2}$ exchanger pumps eject $Ca^{+2}$ out of the receptor. Note that $Ca^{+2}$ drives a negative feedback pathway in the outer segment, since the concentration of $Ca^{+2}$ regulates the rate at which cGMP is produced by GC (Equation 7).

***Photocurrent*:** The photocurrent, $I(t)$, is the ionic inflow of extracellular $Na^{+}$ and $Ca^{+2}$ into the photoreceptor outer segment. $Na^{+}$ and $Ca^{+2}$ enter via pores located in the outer segment plasma membrane, which remain open when cGMP molecules bind to them. The number of open cGMP-gated channels depends on the cGMP concentration, $G(t)$, and determines the amplitude of the photocurrent. In the canonical phototransduction model, $I(t)$ is an instantaneous function of $G(t)$ (Figure 8E) and is described by

$$I(t) = k_{cGMP} \cdot (G(t))^{n_{cGMP}} \quad (9)$$

where $k_{cGMP}$ and $n_{cGMP}$ are constants that determine the non linear dependence of current on open cGMP channels. In our model, $I(t)$ is regulated based on a slow $Ca^{+}2$ -derived signal, $C_{slow}(t)$, as follows:

$$I(t) = \frac{k_{cGMP} \cdot (G(t))^{n_{cGMP}}}{1 + C_{slow}(t)/C_{dark}} \quad (10)$$

where $C_{dark}$ is the $Ca^{+2}$ concentration in the dark. The $C_{slow}(t)$ signal tracks the calcium concentration, $C(t)$, filtered through a slow rate constant, $\rho_{C_{slow}} < \rho_C$, and is described by

$$\frac{dC_{slow}(t)}{dt} = \rho_{C_{slow}} \cdot (C(t) - C_{slow}(t)) \quad (11)$$

This is a second $Ca^{+2}$-based feedback pathway in the outer segment, which provides a slow adaptational mechanism that helps to capture cone responses to impulse, step, and naturalistic stimuli in the primate (Angueyra, 2014). The values of all parameters of this cone photocurrent model are listed in Table 2.

### *Photocurrent noise model*

Photocurrent noise is generated by multiplying the Fourier transform of Gaussian white noise with the sum of two spectral functions, $L_{low}(f)$ and $L_{high}(f)$, and subsequently computing an inverse Fourier transform. The $L_{low}(f)$ and $L_{high}(f)$ functions are given by

$$L_{low}(f) = \frac{\alpha_{low}}{\left(1 + (f/f_{low})^2\right)^{n_{low}}} \quad (12)$$

$$L_{high}(f) = \frac{\alpha_{high}}{\left(1 + (f/f_{high})^2\right)^{n_{high}}} \quad (13)$$

| *Symbol* | *Description* | *Value* | *Units* |
|---|---|---|---|
| $R^*(t)$ | Opsin activity | – | $s^{-1}$ |
| $g_{R^*}$ | Scaling constant for opsin activation | 12 | |
| $\rho_{R^*}$ | Rate of opsin inactivation | 10 | $s^{-1}$ |
| $E(t)$ | PDE activity | – | $s^{-1}$ |
| $\rho_E^{dark}$ | Rate of PDE activation in the dark | 700 | $s^{-1}$ |
| $\rho_E$ | Rate of PDE inactivation | 22 | $s^{-1}$ |
| $GC(t)$ | Guanylate cyclase activity | – | |
| $k_{GC}$ | Half–Maximal $Ca^{+2}$ concentration | 0.5 | M |
| $n_{GC}$ | Steepness of the relation between *GC(t)* and *C(t)* | 4 | – |
| $GC_{max}$ | Max GC activity | 11,090 | – |
| $C(t)$ | $Ca^{+2}$ concentration | – | M |
| $q_{Ca}$ | Fraction of membrane current carried by $Ca^{+2}$ ions | 0.0580 | – |
| $\rho_{Ca}$ | Rate of $Ca^{+2}$ extrusion by $Na^+$ -$Ca^{+2}$ exchange pumps | 5 | $s^{-1}$ |
| $C_{dark}$ | $Ca^{+2}$ concentration in the dark | 1 | M |
| $G(t)$ | cGMP concentration | – | M |
| $G_{dark}$ | Concentration of cGMP in the dark | 20.5 | M |
| $k_{cGMP}$ | Scaling coefficient for cGMP activity | 0.02 | $pA \cdot M^{-1}$ |
| $n_{cGMP}$ | Apparent cooperativity for cGMP activity | 3 | – |
| $I(t)$ | Ionic flow through the outer segment membrane (photocurrent) | – | pA |
| $L_{low}(f)$ | Low-frequency component of photocurrent noise spectral power distribution | – | $pA^2$ |
| $a_{low}$ | | 0.16 | – |
| $f_{low}$ | | 55 | Hz |
| $n_{low}$ | | 4 | – |
| $L_{high}(f)$ | High-frequency component of photocurrent noise spectral power distribution | – | $pA^2$ |
| $a_{high}$ | | 0.045 | – |
| $f_{high}$ | | 190 | Hz |
| $n_{high}$ | | 2.5 | – |

Table 2. Parameters of the cone photocurrent model used in the present study. Parameter values were determined by fitting the model to primate cone responses to impulses delivered in darkness, pulses on various adapting backgrounds, and naturalistic stimuli (Angueyra, 2014).

and model the low and high temporal frequency components, respectively, of the spectral power distribution of photocurrent noise recorded in macaque cone photocurrent responses. The values of the parameters of $L_{low}(f)$ and $L_{high}(f)$ are listed in Table 2. Figure 9A depicts the spectral power distribution of the generated noise (black line) along with the spectral power distributions of the low- and high-frequency noise components (red and blue lines, respectively), and Figure 9B depicts 150 ms of the generated noise, along with a histogram of the amplitude distribution accumulated over 1,024 instances.

It must be noted that our model photocurrent noise does not depend on the background cone excitation level. However, Angueyra and Rieke (2013) have shown that the low-frequency noise component, which includes contributions from extrinsic noise (noise due to fluctuations in cGMP concentration originating from the spontaneous activation of opsin and PDE) and intrinsic noise (noise in the opening and closing of cGMP channels), decreases with background levels (similarly to the gain of mean photocurrent response) with a half-desensitizing level of 4500 $R^* \times \text{cone}^{-1} \times s^{-1}$. The high-frequency noise component is mostly due to intrinsic noise and is only mildly affected by the background cone excitation rate (and with a rate that is not proportional to the inverse of the background) and has a half-desensitizing level of 17500 $R^* \times \text{cone}^{-1} \times s^{-1}$. These background-dependent noise gain effects are not included in our implementation.

*Keywords: contrast sensitivity function, computational modeling, fixational eye movements, photocurrent, phototransduction, spatial pooling, inference engine*

## Acknowledgments

Supported by the Simons Foundation Collaboration on the Global Brain Grant 324759 and Facebook Reality Labs. We thank Marty S. Banks for kindly providing the spatial summation psychophysical data set from the unpublished work of Crowell and Banks.

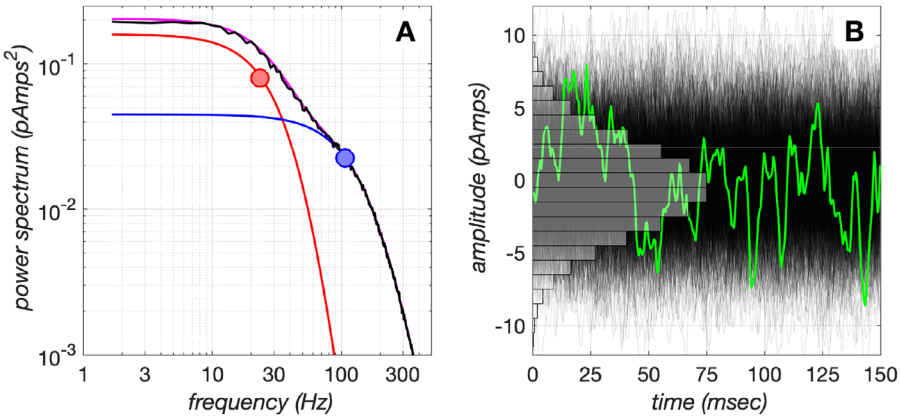


Figure 9. Photocurrent noise model. (A) Photocurrent noise is generated by shaping the spectral power distribution of Gaussian white noise to that corresponding to the sum of two spectral functions, depicted in red and blue lines, with corner frequencies of 23 Hz and 107 Hz (disks), respectively. Their sum is depicted in magenta, and the black line depicts the spectral power distribution of the realized noise. (B) A total of 1,024 instances of 150-ms duration photocurrent noise generated by the model. The green line depicts a single noise instance. The amplitude histogram of the noise is depicted in gray.

Commercial relationships: none.
Corresponding author: Nicolas P. Cottaris.
Email: cottaris@psych.upenn.edu.
Address: Department of Psychology, University of Pennsylvania, 414 Goddard Labs, 3710 Hamilton Walk, Philadelphia, PA, USA.

## Footnotes

[1]https://github.com/isetbio/isetbio.
[2]https://github.com/isetbio/ISETBioCSF.
[3]https://github.com/ISET/iset3d.
[4]In the ISETBio software, the more general term *optical image* is used to refer to the retinal image. In this article, however, we will use the term *retinal image*. We also note that the exact retinal image depends on the point of fixation and will vary over time due to eye movements.
[5]Differential responses are computed by subtracting noisy single-instance responses to the test stimulus from the mean responses to the null stimulus. Single noisy instances of the mosaic's photopigment excitation and photocurrent responses to the null and a test stimulus are provided in Appendix Figure A1.
[6]Note that the SNR of the photocurrent response changes with the background luminance due to changes in the gain of the photocurrent impulse response, without associated changes in the photocurrent noise.
[7]As cone excitation rates exceed 50k–100k $R^* \times \text{cone}^{-1} \times \text{s}^{-1}$, photocurrent noise starts to decrease with background level (Angueyra, 2014). This decrease is not implemented in our present photocurrent model, which implements a constant amplitude of photocurrent noise.
[8]In our earlier study, (Cottaris et al., 2019), we reported that the performance reduction due to having to learn the noise statistics was higher, with the computational observer performance hovering around 50% of the ideal observer performance. Here, we report that the computational observer performance is reduced to only 80% of the ideal observer performance. This reported higher performance loss in our earlier study was due to a misalignment between the retinal image and the spatial pooling mechanism in those computations, which was corrected in the present computations.
[9]An introductory script that demonstrates computation of photocurrent responses in the presence of fixational eye movements can be found at https://github.com/isetbio/ISETBioCSF/tree/master/tutorials/recipes/CSFpaper2.
[10]Subject JAC (light gray triangles) had a lower high-frequency sensitivity than subject MSB and he was not able to detect stimuli of certain patch sizes and spatial frequencies at any contrast, resulting in unmeasurable contrast sensitivity.

## References

Anderson, A. G., Olshausen, B. A., Ratnam, K., & Roorda, A. (2016). A neural model of high-acuity vision in the presence of fixational eye movements. In *2016 50th Asilomar Conference on Signals, Systems and Computers* (pp. 588–592). IEEE, doi:10.1109/ACSSC.2016.7869110.

Angueyra, J. M. (2014). *The limits imposed in primate vision by transduction in cone photoreceptors* (unpublished doctoral dissertation). University of Washington, Seattle, USA.

Angueyra, J. M., & Rieke, F. (2013). Origin and effect of phototransduction noise in primate cone photoreceptors. *Nature Neuroscience, 16*, 1692–1700. [PubMed] [Article]

Banks, M., Geisler, W., & Bennett, P. (1987). The physical limits of grating visibility. *Vision Research, 27*, 1915–1924. [PubMed] [Article]

Boi, M., Poletti, M., Victor, J. D., & Rucci, M. (2017). Consequences of the oculomotor cycle for the dynamics of perception. *Current Biology, 27*, 1268–1277. [PubMed] [Article]

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, 197*, 551–566. [PubMed] [Article]

Cherici, C., Kuang, X., Poletti, M., & Rucci, M. (2012). Precision of sustained fixation in trained

and untrained observers. *Journal of Vision, 12*, 31, doi:https://doi.org/10.1167/12.6.31. [PubMed] [Article]

Cottaris, N. P., Jiang, H., Ding, X., Wandell, B. A., & Brainard, D. H. (2019). A computational-observer model of spatial contrast sensitivity: Effects of wave-front-based optics, cone-mosaic structure, and inference engine. *Journal of Vision, 19*, 8, doi:https://doi.org/10.1167/19.4.8. [PubMed] [Article]

Emerson, R. C., Bergen, J. R., & Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research, 32*, 203–218. [PubMed] [Article]

Endeman, D., & Kamermans, M. (2010). Cones perform a non-linear transformation on natural stimuli. *Journal of Physiology, 588*, 435–446. [PubMed] [Article]

Engbert, R. (2006). Microsaccades: A microcosm for research on oculomotor control, attention, and visual perception. *Prog Brain Res. 154*, 177–192, doi:10.1016/S0079-6123(06)54009-9.

Engbert, R., & Kliegl, R. (2004). Microsaccades keep the eyes' balance during fixation. *Psychological Science, 15*, 431–436. [PubMed]

Farrell, J. E., Jiang, H., Winawer, J., Brainard, D. H., & Wandell, B. A. (2014). Modeling visible differences: The computational observer model. *SID Symposium Digest of Technical Papers, 45*, 352–356. [Article]

Geisler, W. S. (2018). Psychometric functions of uncertain template matching observers. *Journal of Vision, 18*, 1, doi:https://doi.org/10.1167/18.2.1. [PubMed] [Article]

Goris, R. L. T., Putzeys, T., Wagemans, J., & &Wichmann, F. A. (2013). A neural population model for visual pattern detection. *Psychological Review, 120*, 472–496. [PubMed] [Article]

Greene, G., Gollisch, T., & Wachtler, T. (2016). Non-linear retinal processing supports invariance during fixational eye movements. *Vision Research, 118*, 158–170. [PubMed] [Article]

Hateren, H. van. (2005), A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells. *Journal of Vision, 5*, 5, doi:https://doi.org/10.1167/5.4.5. [PubMed] [Article]

Hochstein, S., & Shapley, R. (1976). Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. *Journal of Physiology, 262*, 265–284.

Horwitz, G. D. (2020). Temporal information loss in the macaque early visual system. *PLoS Biology, 18*, e3000570. [Article]

Jiang, H., Cottaris, N. P., Golden, J., Brainard, D. H., Farrell, J. E., & Wandell, B. A. (2017). Simulating retinal encoding: factors influencing Vernier acuity. In *Human vision and electronic imaging* (pp. 177–181). Burlingame, CA: Society for Imaging Science and Technology.

Kelly, D. (1977). Visual contrast sensitivity. *Optica Acta: International Journal of Optics, 24*, 107–129. [Article]

Kowler, E., & Steinman, R. M. (1979). Miniature saccades: Eye movements that do not count. *Vision Research, 19*, 105–108. [PubMed] [Article]

Kuang, X., Poletti, M., Victor, J., & Rucci, M. (2012). Temporal encoding of spatial information during active visual fixation. *Current Biology, 22*, 510–514. [PubMed] [Article]

Kupers, E., Carrasco, M., & Winawer, J. (2019). Modeling visual performance differences around the visual field: A computational observer approach. *PLoS Comput Biol, 15*(5):e1007063, https://doi.org/10.1371/journal.pcbi.1007063. [PubMed] [Article]

Lennie, P., & Movshon, J. A. (2005). Coding of color and form in the geniculostriate visual pathway (invited review). *Journal of the Optical Society of America A, 22*, 2013–2033. [PubMed] [Article]

Lian, T., MacKenzie, K. J., Brainard, D. H., Cottaris, N. P., & Wandell, B. A. (2019). Ray tracing 3d spectral scenes through human optics models. *bioRxiv*. https://www.biorxiv.org/content/early/2019/03/27/589234.full.pdf, doi:10.1101/589234. [Article]

Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., & Dyar, T. A. (2006). Microsaccades counteract visual fading during fixation. *Neuron, 49*, 297–305. [PubMed] [Article]

Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., & Hubel, D. H. (2009). Microsaccades: a neurophysiological analysis. *Trends in Neurosciences, 32*, 463–475. [PubMed] [Article]

Mergenthaler, K., & Engbert, R. (2007). Modeling the control of fixational eye movements with neurophysiological delays. *Physical Review Letters, 98*, 138104. [PubMed] [Article]

Najemnik, J., & Geisler, W. S. (2005, March). Optimal eye movement strategies in visual search. *Nature, 434*, 387–391. [PubMed] [Article]

Ohzawa, I., DeAngelis, G., & Freeman, R. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science, 249*, 1037–1041. [PubMed] [Article]

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination.

*Journal of the Optical Society of America A, 2*, 1508–1532. [PubMed] [Article]

Pugh, E., & Lamb, T. (1993). Amplification and kinetics of the activation steps in phototransduction. *Biochimica et Biophysica Acta (BBA) Bioenergetics, 1141*, 111–149. [PubMed]

Ratnam, K., Domdei, N., Harmening, W. M., & Roorda, A. (2017). Benefits of retinal image motion at the limits of spatial vision. *Journal of Vision, 17*, 30, doi:10.1167/17.1.30. [PubMed] [Article]

Reith, F., & Wandell, B. (2019). A Convolutional Neural Network Reaches Optimal Sensitivity for Detecting Some, but Not All, Patterns. arXiv [cs.CV]. arXiv. http://arxiv.org/abs/1911.05055.

Riggs, L. A., Ratliff, F., Cornsweet, J. C., & Cornsweet, T. N. (1953). The disappearance of steadily fixated visual test objects. *Journal of the Optical Society of America A, 43*, 495–501. [PubMed] [Article]

Robson, J. G. (1966). Spatial and temporal contrast sensitivity functions of the visual system. *Journal of the Optical Society of America A, 56*, 1141–1142. [Article]

Rucci, M., Iovin, R., Poletti, M., & Santini, F. (2007). Miniature eye movements enhance fine spatial detail. *Nature, 447*, 851–854. [PubMed] [Article]

Rushton, W., & Henry, G. (1968). Bleaching and regeneration of cone pigments in man. *Vision Research, 8*, 617–631. [PubMed]

Sinha, R., Hoon, M., Baudin, J., Okawa, H., Wong, R. O., & Rieke, F. (2017). Cellular and circuit mechanisms shaping the perceptual properties of the primate fovea. *Cell, 168*, 413–426.e12. [Article]

Wässle, H. (2004). Parallel processing in the mammalian retina. *Nature Reviews Neuroscience, 5*, 747–757. [PubMed] [Article]

# Appendix

## Examples of noisy cone mosaic response instances

Appendix Figure A1 depicts single noisy instances of cone mosaic responses to the null stimulus (uniform field with a background of 34 cd/m$^2$), depicted in panels A and C, and a test stimulus (100% contrast, 16 c/deg grating presented on a background of 34 cd/m$^2$), depicted in panels B and D. Cone photopigment excitation responses are depicted in panels A and B and the corresponding cone photocurrent responses are depicted in panels C and D.

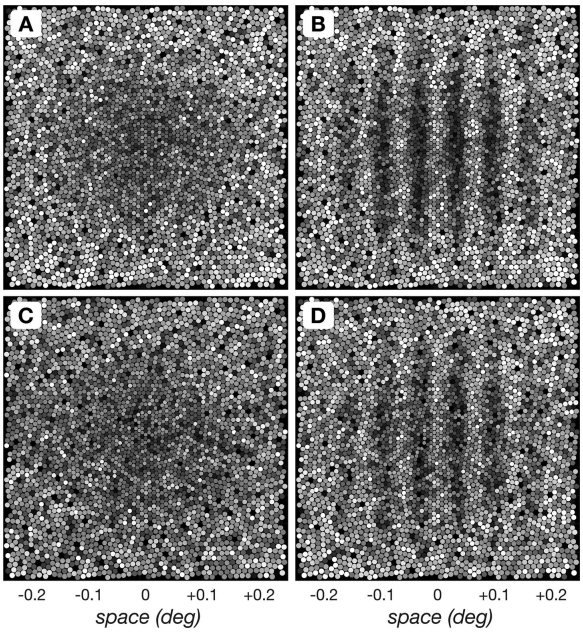


Figure A.1. Single instance cone mosaic responses to the null and a test stimulus depicted at the time of peak response. (A) Cone mosaic photopigment excitation response instance to the null stimulus (uniform field with a background of 34 cd/m$^2$). (B) Cone mosaic photopigment response instance to a test stimulus, here a 16 c/deg, 100% contrast, 34 cd/m$^2$ mean luminance grating stimulus. (C) Cone mosaic photocurrent mosaic response instance to the null stimulus. (D) Cone mosaic photocurrent mosaic response instance to the same test stimulus as in (B).

## Impact of spatial pooling mechanism

When using fixed bandwidth (constant cycles) stimuli, increasing the spatial frequency decreases the stimulus size. The spatial pooling for the computational observer also shrinks because it matches the area of the stimulus. Fixational eye movements can move the retinal image of the stimulus outside the field of view of these pooling mechanisms, thereby lowering performance.

Appendix Figure A2 depicts how performance is affected by choosing spatial pooling regions that extend beyond the stimulus. We consider two spatially-extended cone pooling schemes. The first scheme, depicted in panel B, consists of a single spatial pooling energy mechanism which is centered on the retinal image, depicted in panel A, and which spans 0.3 × 0.3 deg, a spatial region that is considerably larger than the retinal image of the stimulus, 0.16 × 0.16 deg. This energy mechanism consists of a pair of quadrature-phase spatial pooling kernels which are depicted in panels B1 (cos-phase) and B2 (sin-phase). The second scheme consists of an ensemble of spatial pooling energy mechanisms whose centers are positioned on a spatial

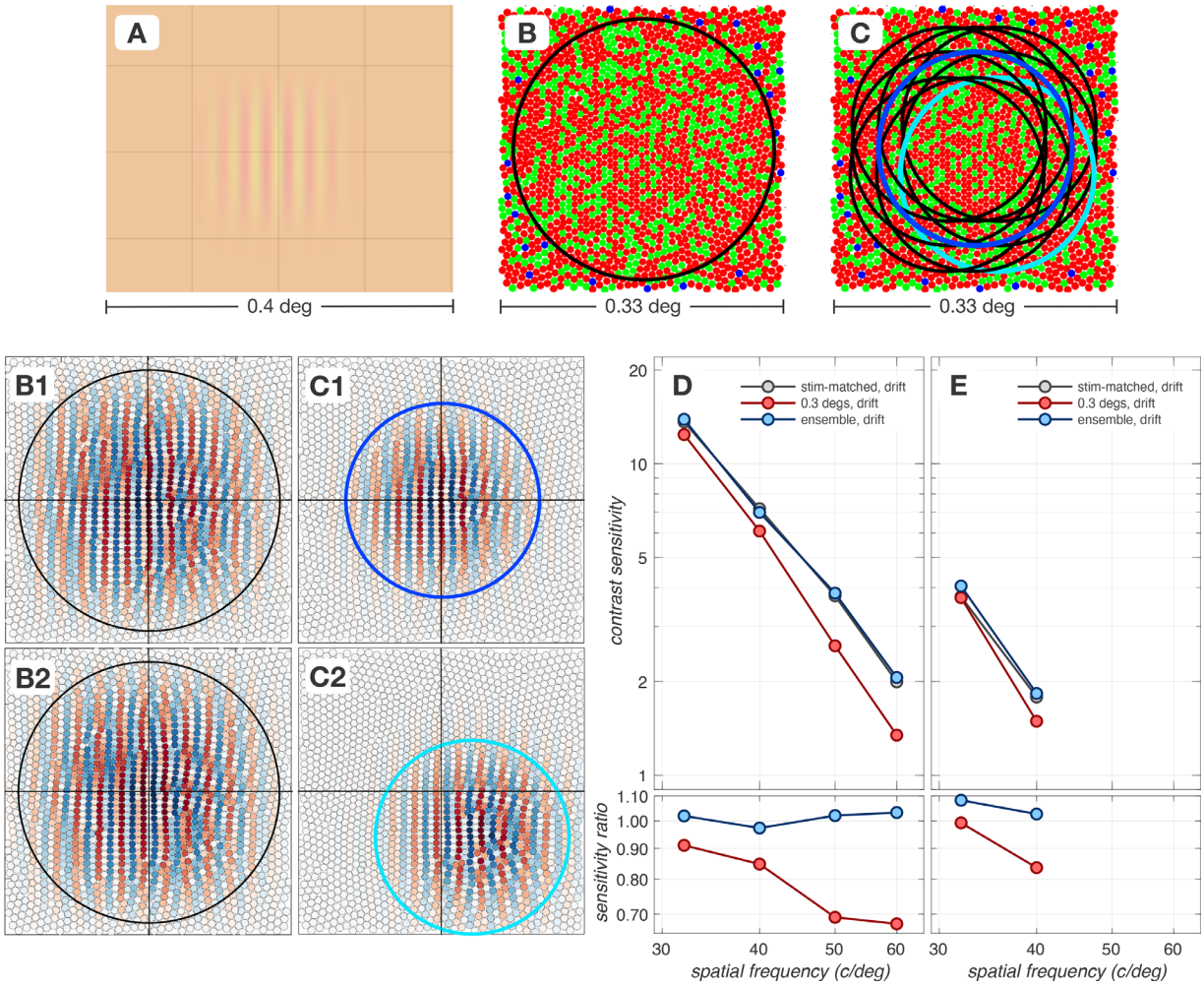


Figure A.2. Impact of spatial pooling. High spatial frequency regime of contrast sensitivity functions computed using different spatial pooling mechanisms in the presence of fixational eye movements. (A) Retinal image of a 40 c/deg test stimulus with a spatial support of 0.16 × 0.16 deg. (B) The spatial pooling extent of a single energy mechanism which pools cone responses over a 0.3 × 0.3 deg region (black circle) is depicted superimposed on the cone mosaic. (C) The spatial pooling extents of a 3 × 3 ensemble of spatially-offset energy mechanisms (each pooling over a 0.16 × 0.16 deg) with a 1.7 arc min separation are depicted in black, blue and cyan circles. (B1 & B2). The quadrature-phase pooling kernels for the single, spatially-extended energy mechanism depicted in panel B. (C1 & C2). Cos-phase pooling kernels for 2 pooling mechanisms in the 3 × 3 ensemble, outlined in blue and cyan in panel (C, D, & E). Performance of different spatial pooling mechanisms at the level of cone photopigment excitations and cone photocurrents, respectively. The CSF obtained using stimulus-matched spatial pooling filters serves as the reference CSF (gray disks). All CSFs are computed for a 3 mm pupil, typical subject wavefront-based optics and eccentricity-based cone mosaics.

grid which also extends beyond the spatial support of the retinal image. In the example shown in panel C we depict a 3 × 3 grid. For this mechanism the input to the SVM classifier is the ensemble of temporal responses of 9 spatial filters. The cos-phase pooling kernels for 2 of these mechanisms (outlined in blue and cyan in panel C) are depicted in panels C1 and C2, respectively.

CSFs (high frequency regime) computed in the presence of fixational eye movements using these different spatial pooling mechanisms are depicted in panels D and E of Appendix Figure A2, for cone photopigment excitation and cone photocurrent signals, respectively. Here, the CSFs computed using the stimulus-matched pooling mechanism are depicted by the gray disks and serve as reference CSFs. Red disks depict the CSFs derived using the single, spatially-extended, energy mechanism and blue disks depict the CSFs derived using the ensemble of spatially-offset energy mechanisms. Note that at the level of cone photopigment excitations (panel D), the single spatially-extended pooling energy mechanism

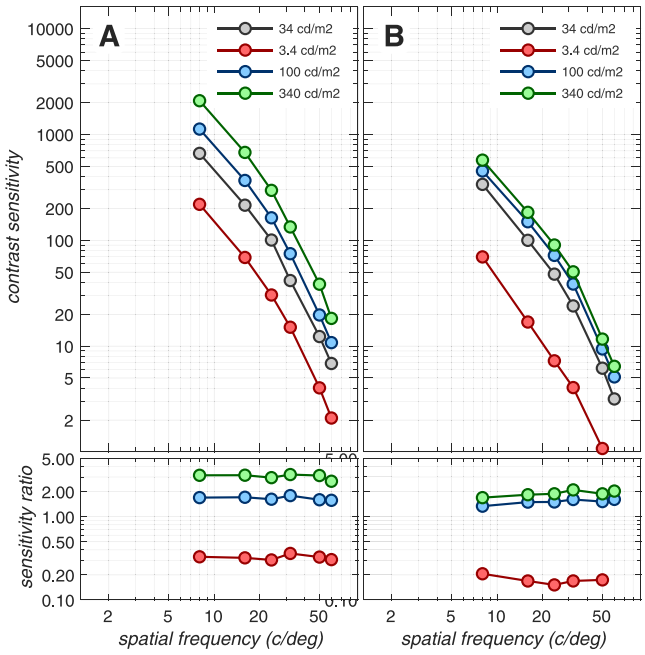


Figure A.3. Impact of background luminance. Contrast sensitivity functions computed at different background levels, based on the cone photopigment excitation response (A) and the cone photocurrent response (B). These CSFs were computed for a 3 mm pupil, typical subject wavefront-based optics, eccentricity-based cone mosaics, in the absence of eye movements and using the SVM-Template-Linear computational observer. The reference CSF (gray disks) is computed for 34 cd/m$^2$, the background used in all other computations in this paper. At the level of cone photopigment excitations (A), the ratios of sensitivity with respect to the reference CSF follows the square root law of Poisson noise limited sensitivity. At the level of cone photocurrents (B), where noise is additive and sensitivity is regulated by the phototransduction dynamics, the contrast sensitivity increase with luminance level is closer to Weber's law which implies no change in contrast sensitivity with luminance. Note that our photocurrent model has been validated against experimental data up to a luminance of about 100 cd/m$^2$, so the performance shown for 340 cd/m$^2$ represents an extrapolation.

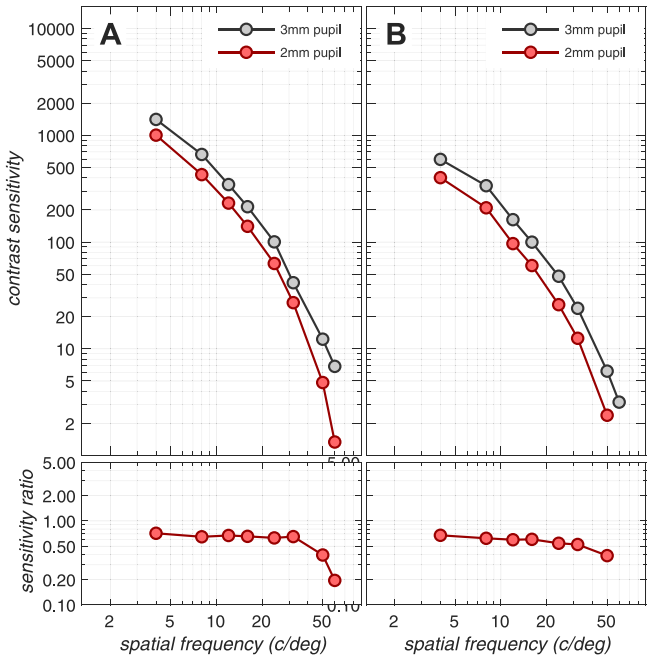


Figure A.4. Impact of pupil size. Contrast sensitivity functions computed for 2 and 3 mm pupils based on cone photopigment excitation (A) and cone photocurrent responses (B). These CSFs were computed for typical subject wavefront-based optics, eccentricity-based cone mosaics, and in the absence of eye movements using the SVM-Template-Linear computational observer. The reference CSF (gray disks) is computed for 3 mm pupil, the pupil size used in most computations in this paper. The 2 mm pupil size (red disks) was used for the calculations shown in Figure A4 for comparison to the human psychophysical data of Banks et al. (1987). At the level of cone photopigment excitations (A), the ratio of sensitivity with respect to the reference CSF cluster around 2/3, i.e., the ratio of pupil diameters. This is expected since the cone photopigment excitation is a Poisson noise signal and its sensitivity increases with the square root of retinal irradiance, and retinal irradiance is proportional to the square of pupil size. The 50 and 60 c/deg ratio significantly departs from the 2/3 ratio. This is presumably related to changes in the wavefront-aberration based optics which depend on pupil size. At the level of cone photocurrents (B), sensitivity is also reduced as the pupil diameter decreases from 3 to 2 mm by a factor that varies from 0.67 at the lowest frequency (4 c/deg) to 0.4 at the highest frequency (50 c/deg). The somewhat increased sensitivity loss at 50 c/deg is again likely related to changes in the wavefront-aberration based optics with pupil size.

(red disks) performs worse than the stimulus-matched pooling mechanism (gray disks). On the other hand, the performance of the ensemble of spatial pooling energy mechanisms (blue disks) is slightly better than the performance of both the spatially-extended and the stimulus-matched single spatial pooling mechanisms. At the level of photocurrents (panel E), the ensemble of spatial pooling filters mechanisms again slightly outperforms both the stimulus-matched and the spatially extended single spatial pooling mechanisms at 32 and 40 c/deg, whereas at spatial frequencies $>$ 40 c/deg we were not able to obtain a detection threshold for any spatial pooling mechanism in the presence of eye movements.

Overall, these results indicate that ensembles of spatially-overlapping energy mechanisms can slightly enhance performance at high spatial frequencies in the presence of fixational eye movements. The performance of the ensemble mechanisms depends on the amount of spatial overlap and the width of the component filters, but this dependence is not examined in the present work.

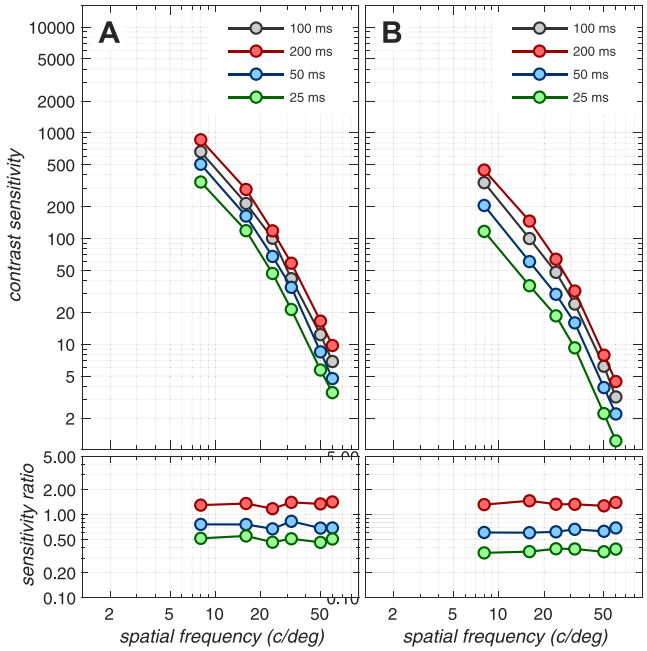


Figure A.5. Impact of stimulus duration. Contrast sensitivity functions computed for different stimulus durations, at the level of cone photopigment excitations (left) and cone photocurrents (right). These CSFs were computed for a 3 mm pupil, typical subject wavefront-based optics, eccentricity-based cone mosaics, and in the absence of eye movements using the SVM-Template-Linear computational observer. The reference CSF (gray disks) is computed for 100 ms, the duration used in all computations in this paper. Note that at the level of cone photopigment excitations, ratios of CSFs with respect to the reference CSF (100 ms) cluster around $\sqrt{2.0}$ for 200 ms, $\sqrt{0.5}$ for 50 ms, and $\sqrt{0.25}$ for 25 ms, as expected from the square root law of Poisson noise limited sensitivity. At the level of cone photocurrents, where noise is additive and sensitivity is regulated by the phototransduction dynamics, there is a more dramatic effect of stimulus duration on performance.

## Impact of background luminance

Appendix Figure A3 depicts contrast sensitivity functions for different background luminance levels at the level of cone photopigment excitations (panel A) and cone photocurrents (panel B). There is no change in the shape of the contrast sensitivity function as luminance varies at either response stage.

## Impact of pupil size

Appendix Figure A4 depicts contrast sensitivity functions for 3 and 2 mm pupils, assessed at the level of cone photopigment excitations (panel A) and cone photocurrents (panel B). Note that changes in the pupil size affect not only the retinal irradiance but also the wavefront aberration function and therefore the point spread function. The sensitivity ratios depicted in the bottom panels of Appendix Figure A4 are relatively constant for spatial frequencies up to 50 c/deg with small deviations at 50 and 60 c/deg. Therefore, except for the very high spatial frequencies, changing the pupil size from 2 to 3 mm does not impact the shape of the CSF. In this paper, we conduct most of our simulations using a 3 mm pupil size, because this is a more realistic pupil size for natural viewing of laboratory created stimuli, as discussed in Cottaris, Jiang, Ding, Wandell, and Brainard (2019). The 2 mm pupil size was used for the calculations shown in Figure A4 for comparison to the human psychophysical data of Banks, Geisler, and Bennett (1987).

## Impact of stimulus duration

Appendix Figure A5 depicts the impact of stimulus duration. At the level of the Poisson-limited cone photopigment excitations, and under the assumption of full temporal integration of responses, we expect a square-root law reduction in sensitivity as stimulus duration is decreased, and this is what is observed in panel A. At the level of cone photocurrents, depicted in panel B, we see a more dramatic reduction in sensitivity as stimulus duration is decreased. This reduction occurs because of the temporal integration of cone excitation responses during phototransduction. This effect can be observed directly in Figure 1F, where temporally-sharp modulations in the cone photopigment excitation response result in severely attenuated corresponding cone photocurrent modulations.

## Impact of spatial summation

Crowell and Banks (unpublished manuscript) measured contrast sensitivity functions for Gabor patches of different sizes, in an attempt to discover possible sources underlying the 20-fold discrepancy between ideal and human observer performance. Their rationale was that the ideal observer is allowed to combine responses across the entire stimulus patch, whereas detection mechanisms in the visual system may have a more limited spatial summation window. Crowell and Banks found that as grating patch size decreased, the ratio of ideal to human performance also decreased, bringing the ideal and human performance gap closer to 5–10 fold. We compare the performance of our computational observer to the performance of the two human observers examined by Crowell and Banks. We thank M.S. Banks for suggesting the comparison and for providing us with the unpublished manuscript.

The three panels of Appendix Figure A6 depict contrast sensitivity functions for three patch sizes

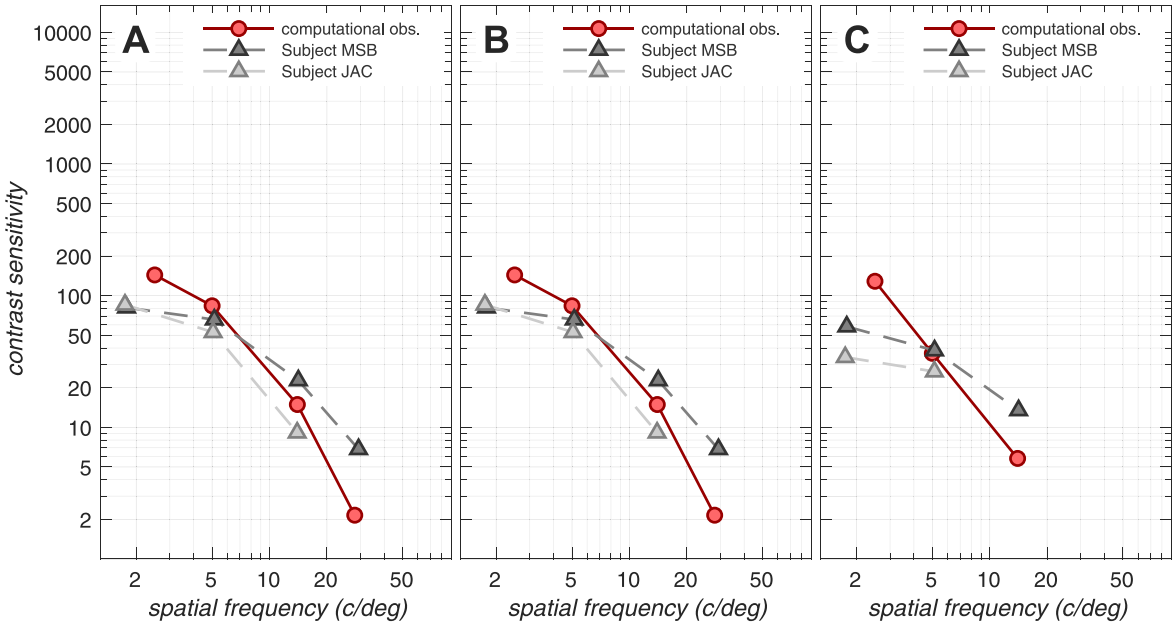


Figure A.6. Impact of spatial summation. Contrast sensitivity functions computed for different sizes of Gabor patches: (A) 3.3 cycles, (B), 1.7 cycles, (C) 0.4 cycles. Red disks depict computational observer performance assessed at the level of cone photocurrents, in the presence of fixational eye movements, and computed using the SVM-Template-Energy inference engine. Gray triangles: performance of the two subjects measured by Crowell and Banks. Subject JAC (light gray triangles) has a lower high frequency sensitivity than subject MSB, and was not able to detect stimuli of certain patch size and spatial frequencies at any contrast. The simulated stimulus conditions (matched to the experiment of Crowell and Banks) were: 100 cd/m$^2$ mean luminance, 100 ms pulsed stimulus presentation, 2.5 mm pupil size, and performance threshold set to 75% correct.

(expressed as number of grating cycles within the $2 \times \sigma$ of the Gabor stimulus) examined by Crowell and Banks. The red disks depict the performance of our computational observer assessed at the level of cone photocurrents, in the presence of fixational eye movements, and computed using the SVM-Template-Energy inference engine. The gray triangles depict the performance of the two subjects measured by Crowell and Banks. Note that, for spatial frequencies greater than 5 c/deg, the performance of our computational observer lies between the performance of the two subjects[10] for all three patch sizes. Also note that at 2.5 c/deg, the computational observer performance is significantly higher than human performance. This is presumably because the antagonistic center/surround interactions in the receptive fields of post-receptoral neurons (e.g., ganglion cells), which reduce sensitivity to low spatial frequency achromatic stimuli, are not included in our computational observer. These results provide further support for the ability of our computational observer to capture the medium to high spatial frequency sensitivity loss that occurs due to processing early in the visual system.