Journal of Vision (2019) 19(4):8, 1–27

A computational-observer model of spatial contrast sensitivity: Effects of wave-front-based optics, cone-mosaic structure, and inference engine

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

Haomiao Jiang
Department of Electrical Engineering,
Stanford University, Stanford, CA, USA

Xiaomao Ding
Department of Psychology, University of Pennsylvania,
Philadelphia, PA, USA

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

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

We present a computational-observer model of the human spatial contrast-sensitivity function based on the Image Systems Engineering Toolbox for Biology (ISETBio) simulation framework. We demonstrate that ISETBio-derived contrast-sensitivity functions agree well with ones derived using traditional ideal-observer approaches, when the mosaic, optics, and inference engine are matched. Further simulations extend earlier work by considering more realistic cone mosaics, more recent measurements of human physiological optics, and the effect of varying the inference engine used to link visual representations to psychophysical performance. Relative to earlier calculations, our simulations show that the spatial structure of realistic cone mosaics reduces the upper bounds on performance at low spatial frequencies, whereas realistic optics derived from modern wave-front measurements lead to increased upper bounds at high spatial frequencies. Finally, we demonstrate that the type of inference engine used has a substantial effect on the absolute level of predicted performance. Indeed, the performance gap between an ideal observer with exact knowledge of the relevant signals and human observers is greatly reduced when the inference engine has to learn aspects of the visual task. ISETBio-derived estimates of stimulus representations at various stages along the visual pathway provide a powerful tool for computing the limits of human performance.

Introduction

Newton's work on the nature of light, some four centuries ago, initiated the quantitative understanding of vision. Since that time much has been learned about light, retinal image formation, fixational eye movements, and photon-initiated excitations in the cone photoreceptors (Bowmaker, Dartnall, & Mollon, 1980; Wyszecki & Stiles, 1982; Rodieck, 1998; Engbert & Kliegl, 2004; Martinez-Conde, Macknik, & Hubel, 2004; Artal, 2015). Work continues to clarify how photoreceptor excitations are transformed into photocurrent and then to retinal and cortical signals that mediate visual perception (Baylor, Nunn, & Schnapf, 1984; Wandell, 1995; Meister & Berry, 1999; Pugh & Lamb, 2000; Angueyra & Rieke, 2013; Li et al., 2014).

All visual stimuli pass through the optics and retina, giving these structures a prominent role in defining the limits of vision. For example, the three-dimensional nature of human color vision can be understood in terms of the three types of cone photopigments that absorb light (Brindley, 1960; Wandell, 1995). Also,

Citation: 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*(4):8, 1–27, https://doi.org/10.1167/19.4.8.

https://doi.org/10.1167/19.4.8 Received November 21, 2018; published April 3, 2019 ISSN 1534-7362 Copyright 2019 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

critical aspects of human pattern sensitivity depend on physiological optics (Robson, 1966; Campbell & Robson, 1968; Williams, 1985; Banks, Geisler, & Bennett, 1987). Quantification of human color and pattern sensitivity are critical for the imaging industry, including the design of cameras, displays, and printers; understanding the biological basis of visual sensitivity gives us confidence in the generality of the results and enables the diagnosis and targeted treatment of blinding disease.

Equally important, many aspects of visual perception are not explained by the initial stages of visual encoding. For example, human judgments of material appearance, the ability to recognize objects, and stereovision depend on brain circuits that integrate information across space, time, and the two eyes. Attempts to understand these circuits can nonetheless benefit from a quantitative understanding of the initial encoding, as this determines the information available for perceptual inferences made by the brain.

Although our understanding of many properties of visual encoding may in principle be quantified using explicit computational models, putting such models to use in the practice of vision science is currently daunting. The relevant information is spread across a large literature, and integrating this information for a particular project typically requires a substantial effort. We developed the Image Systems Engineering Toolbox for Biology (ISETBio, http://isetbio.org) to make the relevant computations and data more accessible. ISETBio is an open-source software system that provides an image-computable model of the first stages of visual encoding.

Image-computable means that the calculations begin with a quantitative description of the visual image. An important special case supported by ISETBio is planar images presented on a computer display or the printed page. ISETBio also includes support for a more general case, in which the input is defined using a three-dimensional description that includes the location and shape of scene elements as well as the spectral properties of each scene element. In this more general case, the retinal image is derived from the scene representation using ray-tracing methods (Pharr & Humphreys, 2010).

Computable methods are important because they can characterize visual representations for conditions that are beyond the reach of analytic formulations. When coupled with an inference engine that links the computed representations to performance on specific visual tasks, such as an ideal observer (De Vries, 1943; Rose, 1948; Tanner & Swets, 1954; Geisler, 1984, 1989), computable methods can assess limits on performance and characterize the information available to brain circuits. We use the term *computational-observer analysis* to describe image-computable methods combined with an inference engine (Lopez, Loew, Murray, & Goodenough, 1992).

This article describes extensive updates to the previous versions of ISETBio (Farrell, Jiang, Winawer, Brainard, & Wandell, 2014; Jiang et al., 2017). We review and validate the updates by showing that the predictions for a computational observer implemented in ISETBio agree with analytical calculations of ideal-observer pattern sensitivity derived in prior work (Geisler, 1984; Banks et al., 1987). We then explore how individual differences in human optics and cone mosaic affect predicted human performance. Finally, we analyze how performance varies with the parameters and architecture of the inference engine. In particular, we consider inference engines designed for pattern detection, and compare support-vector-machine (SVM) methods, in which decision rules are learned by observing labeled response instances, with ideal-observer methods, in which the decision rule is computed analytically based on exact knowledge of the stimulus and the response noise statistics.

Pattern-sensitivity analysis is but one of many potential applications of ISETBio. We hope that making the software open-source and freely available will help others to develop analyses in new application areas.

ISETBio overview

ISETBio computations are organized into a series of extensible methods that model the critical stages of visual encoding, from the visual scene through the optics, cone mosaic, and inference engine (Figure 1). ISETBio scene methods represent the visual scene and enable calculations based on this representation. Scene methods include ways to represent visual stimuli consisting of a spatiotemporal pattern specified as the spectral radiance emitted at each location and time on a flat screen. In this case, which is the one we use in this article, the scene specification can be in terms of RGB values and is coupled to display calibration data, most importantly the spectral radiance of each of the display primaries (Figure 1A). The ability to represent stimuli presented as images on a flat display is important for modeling many psychophysical experiments. We have also implemented methods that support modeling of psychophysical stimuli presented in Maxwellian view (Tuten et al., 2018) and modeling of 3-D scenes using quantitative computer-graphics calculations (Lian et al., 2018).

The spectral irradiance incident at the retina is calculated from the scene representation using ISETBio optical image methods (Figure 1B). This computation accounts for critical physiological optics factors including pupil size, wavelength-dependent blur, and

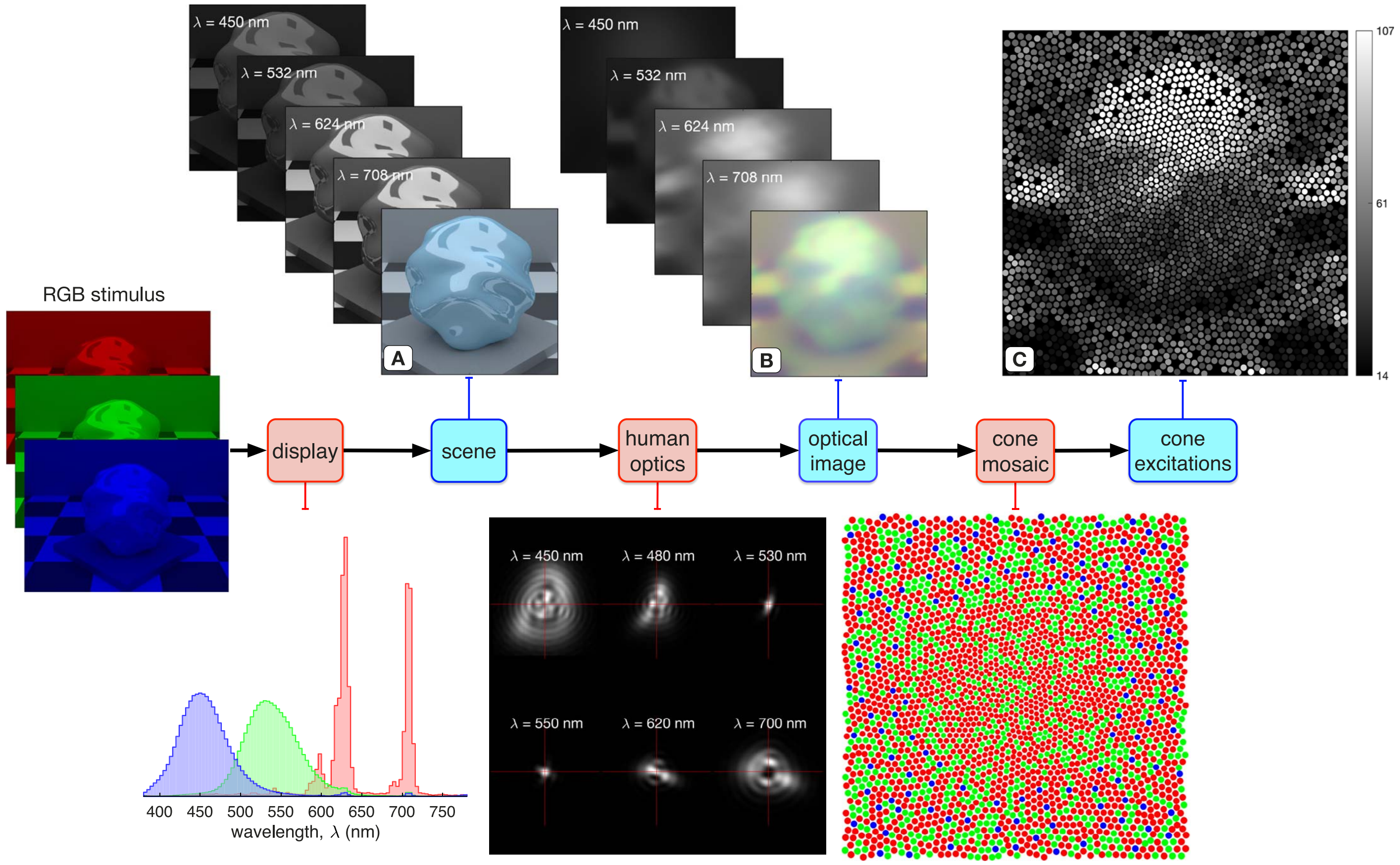

Figure 1. Flowchart of computation in ISETBio. (A) The visual stimulus, in this case an image on an RGB display, is represented as an ISETBio scene, which represents the emitted radiance at a set of wavelengths. Here, the spectral power distributions of the display primaries (lower portion of the figure) and the pixel spatial sampling are used to convert stimulus RGB values to the spatial-spectral radiance. An RGB rendition of the scene is depicted in front of the spectral-radiance stack. In the calculations reported in this article, wavelengths are sampled between 380 and 780 nm with a 5-nm spacing, but here only a subset of the sample wavelengths is shown. (B) ISETBio optical image methods transform the scene to the retinal spectral irradiance. These methods blur the scene spatial radiance using a set of wavelength-dependent, shift-invariant point-spread functions (example for one individual subject shown in the lower portion of the figure) and account for spectral transmission through the lens. Spectral transmission through the macular pigment is handled as part of the computation of cone excitations. C. ISETBio cone-mosaic methods compute the number of cone excitation events, which are coded in grayscale. S-cones appear dark, as they are excited much less than the L- and M-cones because of selective absorption of short-wavelength light by the ocular media. In the mosaic shown (lower image), cone density decreases and cone aperture increases with eccentricity, and there is a central region free of S-cones.

wavelength-dependent transmission through the crystalline lens. The optical image calculations used in this article account for the on-axis wave-front aberrations of the eye's optics, measured using a wave-front sensor (Thibos, Hong, Bradley, & Cheng, 2002), which determine a set of wavelength-dependent, shift-invariant point-spread functions.

The spatial pattern of cone excitations is computed from the retinal irradiance using ISETBio cone-mosaic methods. These methods transform the spectral irradiance at the retina into cone excitations (Figure 1C). The cone-mosaic methods include parameters which control factors such as the relative number of L-, M-, and S-cones; the existence and size of an S-cone-free zone in the central fovea; the cone spacing, inner-segment aperture size, and outer-segment length; the cone photopigment density; and the macular pigment density. These parameters all affect the number of cone excitations.

In a contrast-sensitivity experiment, the subject discriminates between a spatially uniform pattern (null stimulus) and a cosinusoidal grating pattern (test stimulus). In ISETBio, we use the term *inference engine* to describe methods that link the computed visual-system responses to psychophysical performance. Inference-engine methods make decisions in a simulated visual task based on the stimulus representation at different processing stages along the visual pathway. Figure 2A and 2B depicts the retinal-image contrasts seen by the L-, M-, and S-cones for, respectively, the

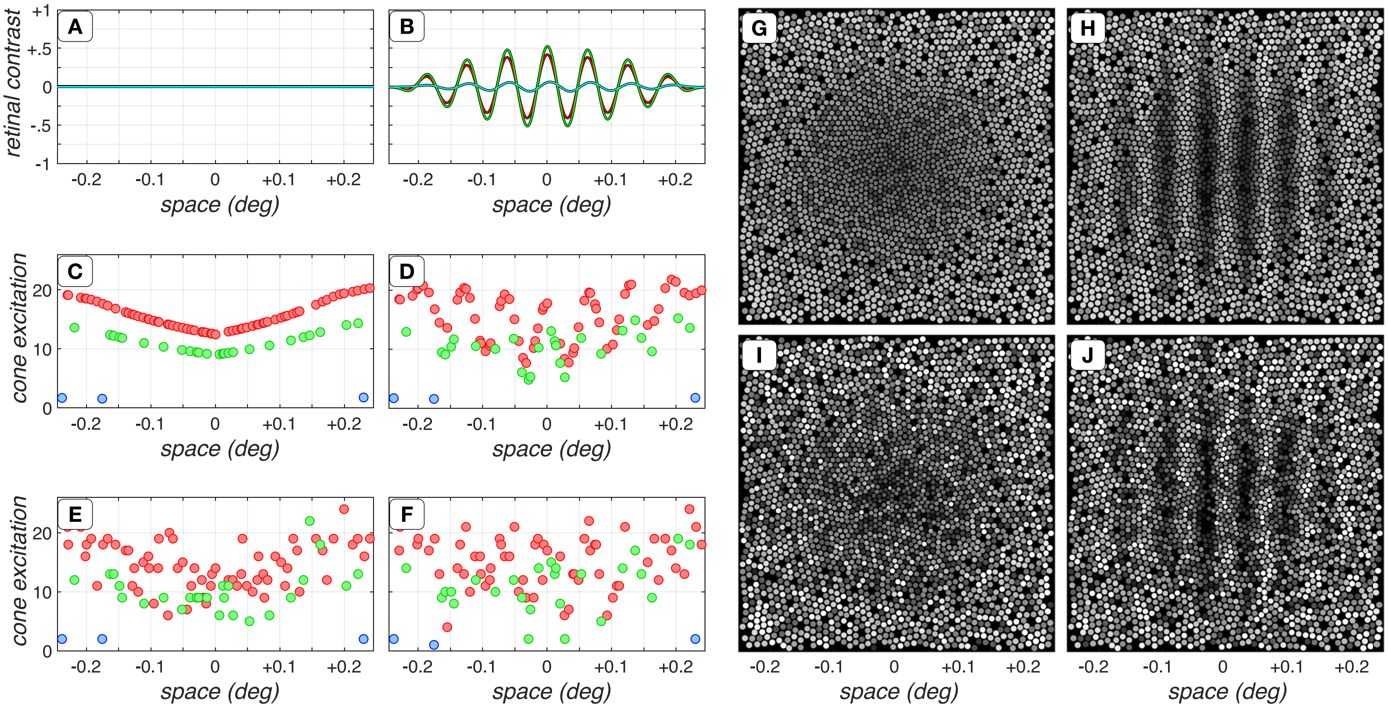

Figure 2. Stimulus representations in ISETBio. Representations of a uniform field (null stimulus) and of a 16-c/° 100% Michelson contrast cosinusoidal grating (test stimulus) are depicted in paired panels. (A–B) Retinal contrast along the horizontal meridian for the null and test stimulus, respectively. These spatial contrasts are depicted as seen by the L-cones (red), M-cones (green), and S-cones (cyan). (C–D) Mean cone excitation (number of photon-absorption events within a 5-ms time bin) for cones along the horizontal meridian to the null and test stimulus, respectively. Red, green, and blue disks indicate L-, M-, and S-cones, respectively. (E–F) A single excitation instance of cones along the horizontal meridian to the null and test stimulus, respectively. (G–H) Mean cone excitation pattern of the entire mosaic to the null and the test stimulus, respectively. (I–J) A single excitation instance of the mosaic to the null and the test stimulus, respectively.

null stimulus and for a 16-c/° grating test stimulus with 100% Michelson contrast. Note that for this stimulus, aberrations reduce the retinal-image L- and M-cone contrast by a factor of 2 relative to the stimulus image, whereas the retinal S-cone contrast is reduced by a factor of 10. Aberrations also shift the spatial phase of the retinal S-cone contrast with respect to those of L- and M-cone contrasts (Figure 2B).

Figure 2C and 2D depicts the mean excitation level (within a 5-ms window) of cones along the horizontal meridian for the null and test stimuli, respectively. The mean cone excitation increases with eccentricity because of changes in cone aperture with eccentricity. The excitations of the same cones to a single stimulus instance, obtained by adding Poisson noise to the mean excitations, are depicted in Figure 2E and 2F, respectively. The mean excitation of the entire cone mosaic to the null and test stimulus is depicted in Figure 2G and 2H, respectively, whereas Figure 2I and 2J depicts a single instance of the cone-mosaic excitation. Note that it would be challenging to discriminate between the two stimuli by looking at single response instances of just a few cones, as can be seen by inspecting Figure 2E and 2F. Spatial integration across the cone mosaic will improve performance, as can be appreciated by visual comparison of Figure 2I and 2J. It must also be noted that the responses in Figure 2 are excitations to a suprathreshold (100%) contrast grating, not to a grating at contrast threshold, and ultimately classification performance cannot exceed the limits imposed by the Poisson noise inherent in these excitations.

In this article, we consider inference engines that model a two-alternative forced-choice version of the contrast-sensitivity experiment, and we use response instances at the level of cone-mosaic excitations to predict the probability of correct discrimination between gratings and a uniform field. Performance is limited according to how well the inference engine is matched to the task (the classifier's calculation efficiency; Barlow, 1964; Pelli, 1990), as well as the difference between the representations of the stimuli relative to those of noise—that is, trial-by-trial fluctuations in the representations. As already noted, Poisson noise is inherent to cone excitations and is a critical limiting factor for performance at this stage of encoding.

Results

Pattern-sensitivity validation

Complex software requires explicit testing of the individual components (unit testing), component communication (integration testing), and the overall system (validation). The ISETBio software includes a number of such tests, as well as methods to check that new software methods do not invalidate previously established tests (regression testing). In this section we describe validation testing of a complex computation that utilizes key ISETBio methods. We show that the ISETBio implementation—including stimulus definition, physiological optics, and cone excitations—matches the precise analytical calculation performed by Banks et al. (1987) for an ideal observer's contrast sensitivity to known spatial harmonic patterns (signal-known-exactly). This test is designed to provide confidence in the basic implementation and the validity of the subsequent explorations of how physiological optics, the cone mosaic, and the inference engine influence human pattern sensitivity.

We computed ISETBio contrast-sensitivity functions (CSFs) using parameters that matched those used by Banks et al. (1987). These included a 2-mm pupil diameter, a point-spread function (PSF) derived from early line-spread-function measurements (Campbell & Gubisch, 1966; Geisler, 1984), a regularly spaced hexagonal cone mosaic comprising an approximation of L- and M-cones in a 2:1 ratio, cone center-to-center spacing of 3 μm, and a cone inner-segment aperture of 3 μm. There was one small difference within the cone mosaic. Banks et al. calculated for a mosaic in which all cones were of the same type, each with a luminance spectral sensitivity of 2L +M; we modeled a mosaic consisting of distinct L- and M-cones in a 2:1 ratio.

Performance (probability correct, P_{correct}) was estimated for each grating contrast and spatial frequency separately. We simulated a two-alternative forced-choice task, using an ideal-observer classifier that selects which of the two alternative stimulus sequences (test–null or null–test) was more likely to generate the observed cone excitations. The test stimulus was a spatial grating of known contrast, frequency, and position; the null stimulus was a spatially uniform field. The simulated duration of the test and null stimulus was 100 ms on each trial, with the stimuli presented in random order. The ideal observer's performance was calculated analytically given knowledge of the sequence of mean number of excitations during the 100-ms intervals and the assumption of Poisson noise. Responses were binned using 5-ms bins. This choice is irrelevant in the present simulations, which model static stimuli in the absence of eye movements, and the results are indeed the same if a single 100-ms bin is used for each interval. The choice of 5-ms bins is to allow direct comparison with future work which will include fixational eye movements (Cottaris, Rieke, Wandell, & Brainard, 2018). The psychometric function (P_{correct} as a function of stimulus contrast) for each spatial pattern was fitted with a cumulative Weibull (Kingdom & Prins, 2010; http://www.palamedestoolbox.org), and threshold was computed as the stimulus contrast corresponding to $P_{\text{correct}} = 0.7071$. Contrast sensitivity is the reciprocal of threshold contrast.

Figure 3A compares CSFs for three mean luminance levels (3.4, 34, and 340 cd/m). The solid lines show the ideal-observer CSFs obtained by Banks et al. (1987), digitized from their figure 2. Disks depict the CSFs computed using ISETBio with its implementation of the ideal-observer inference engine. The ISETBio-derived CSF agrees with that of Banks et al. across all spatial frequencies and luminance levels. In addition, the sensitivity ratios between the different mean luminance levels (bottom panel of Figure 3A) cluster around the ratios $\sqrt{10}$ and $\sqrt{1/10}$, as expected from the square-root law for Poisson-limited sensitivity (De Vries, 1943; Rose, 1948). We take this agreement as an important system validation of the ISETBio implementation.

As a further check, we assessed the impact of pupil diameter (2 mm vs. 3 mm) on the CSFs computed using ISETBio (Figure 3B). The 2-mm pupil is used for the comparison with calculations and psychophysical data reported by Banks et al., as their data were collected using a 2-mm artificial pupil. The 3-mm pupil is used because it is more appropriate for natural viewing of the stimuli. When a 30-year old observer views an adapting field of 50° and 34 cd/m binocularly, the expected pupil diameter is 3.4 mm; when the adapting luminance is 100 cd/m, the expected pupil diameter is 3.0 mm (Watson & Yellott, 2012). For Poisson signals, sensitivity should increase with the square root of retinal irradiance, and since retinal irradiance is proportional to the square of pupil diameter, we expect the sensitivity ratios for the 3-mm vs. the 2-mm CSF to be 1.5 across all spatial frequencies. This is confirmed to a good approximation, which further validates the software implementation. Note that for this test we did not change the optical PSF, although a change would be expected in a simulation aimed at fully understanding the impact of a change in pupil size.

Taken together, these computations ground the ISETBio ideal-observer implementation in the analytical literature and validate the use of ISETBio for exploring how changes in visual-system parameters affect the estimated upper bound for the spatial CSF.

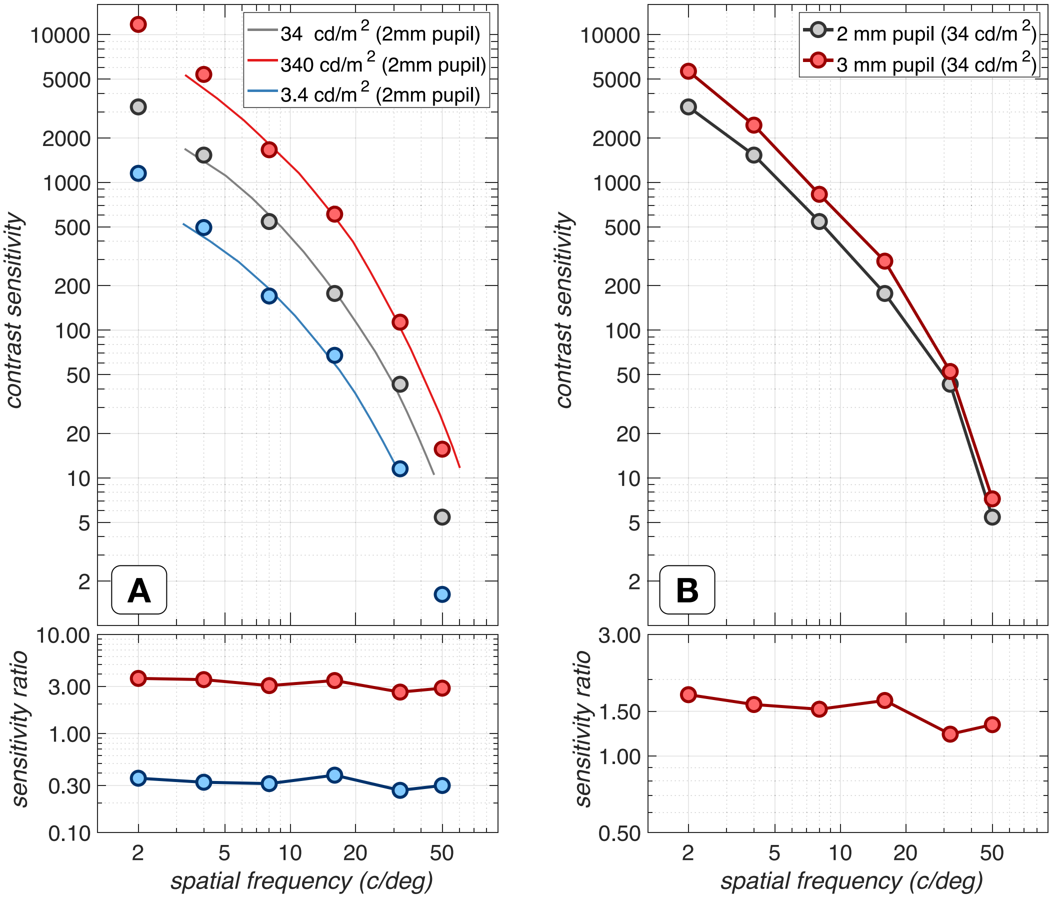

Figure 3. ISETBio validations. A. Validation against Banks et al. (1987). The top plot depicts contrast-sensitivity functions (CSFs) for a 2-mm pupil diameter and three mean luminance levels. Solid lines depict the ideal-observer CSFs, digitized from Banks et al., and disks depict the CSF values calculated using ISETBio with matched parameters. The ratios of contrast sensitivities between the 3.4- and 34-cd/m mean luminances (blue) and between the 340- and 34-cd/m mean luminances (red) are shown in the bottom plot. B. Validation with respect to pupil size. Top plot depicts the ISETBio ideal-observer CSFs for 3-mm (gray) and 2-mm (red) pupil diameters. Other parameters matched those of Banks et al., and the optical PSF was held constant across this comparison. The bottom plot shows the ratio of the 3- and 2-mm contrast sensitivities.

Cone mosaic

The Banks et al. (1987) psychophysical data were collected using a constant number of grating cycles across changes in spatial frequency. Thus the spatial extent of the stimuli was larger for lower spatial frequencies. Banks et al. employed a constant-density cone mosaic. For the human retina, however, cone density declines as a function of eccentricity; this decline is particularly rapid across the central fovea (Curcio, Sloan, Kalina, & Hendrickson, 1990). To explore how a change in cone density affects the spatial CSF for constant-cycle stimuli, we developed new methods to implement realistic cone mosaics (described in detail later, under Cone mosaics). These cone-mosaic methods retain the approximately hexagonal cone packing of central retina while decreasing the cone density with eccentricity. We compared the ideal-observer CSF calculations for the regularly-spaced hexagonal L/M-cone mosaic of Banks et al. to those obtained using different eccentricity-dependent mosaics. For these calculations, we simulated a mean stimulus luminance of 34 cd/m^2 and a 3-mm pupil. The optical PSF matched the one used by Banks et al.

Results from this analysis are depicted in Figure 4. The mosaic employed by Banks et al. is depicted in Figure 4A. It contained only L- and M-cones in a 2:1 ratio, with hexagonal packing at 3-μm cone spacing and a 3-μm cone inner-segment diameter. Three eccentricity-dependent cone mosaics are depicted in Figure 4B–4D. The mosaic shown in Figure 4B also consisted of only L- and M-cones in a 2:1 ratio, but cone density decreased according to the measurements of Curcio et al. (1990), and the cone inner-segment diameter was 1.6 μm. A second eccentricity-dependent cone mosaic consisted of L-, M-, and S-cones in the ratio 0.62:0.31:0.07, with S-cones starting to appear at eccentricities >0.1° (Figure 4C). A third eccentricity-

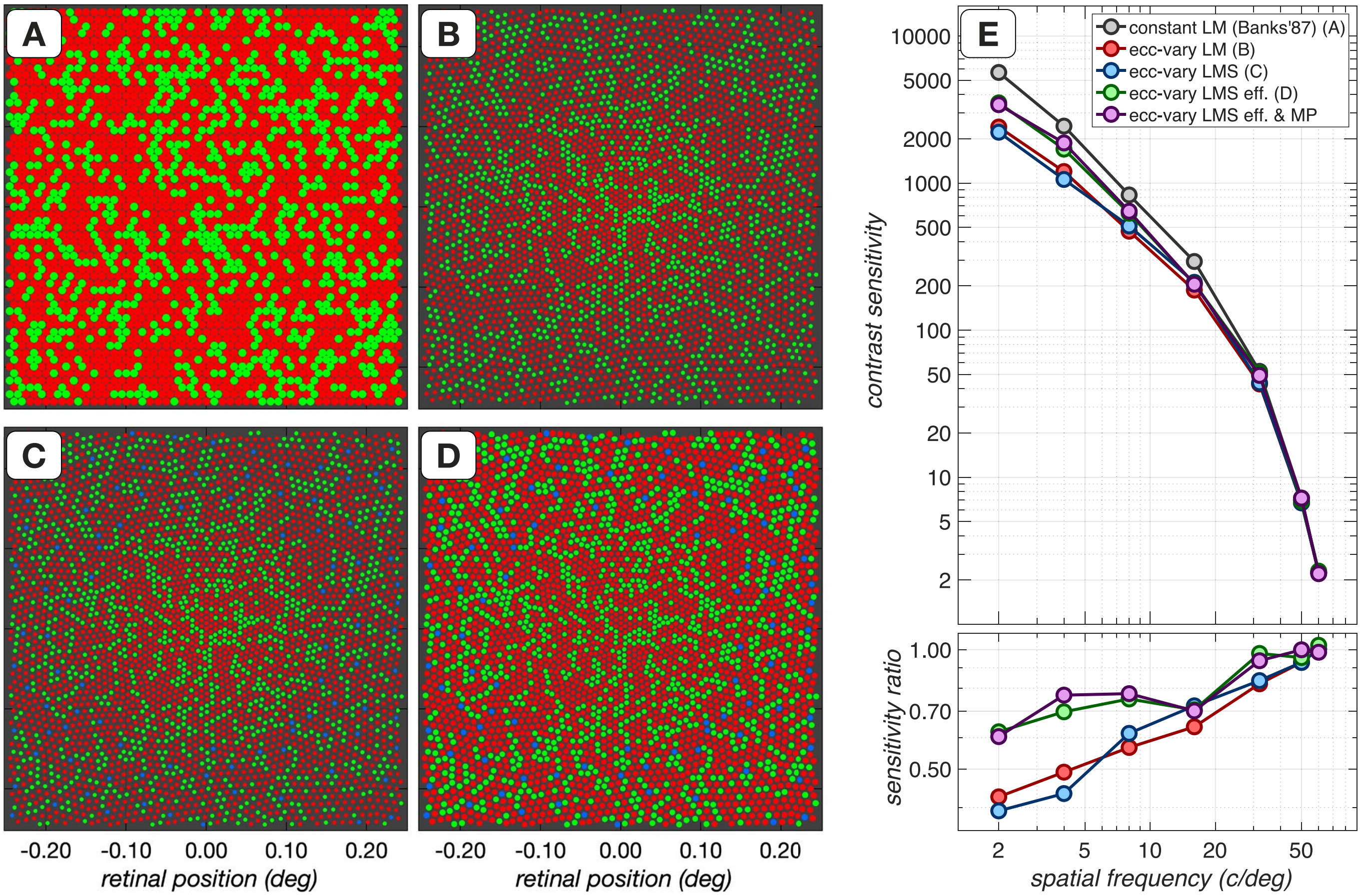

Figure 4. Effects of cone mosaic. (A–D) Central 0.5° × 0.5° of the mosaics used. (A) The mosaic used by Banks et al. (1987) with a regular hexagonal cone packing with 3-μm cone spacing, 3-μm cone aperture, and cones with a luminance spectral sensitivity. We replicated these parameters but used L- and M-cones in a 2:1 ratio. (B) A mosaic with eccentricity-dependent cone density with only L- and M-cones in a 2:1 ratio. Cones in this mosaic have foveal values for inner-segment diameter and outer-segment length, independent of eccentricity. (C) A mosaic with eccentricity-dependent cone density, foveal cone inner-segment diameter, and outer-segment length, consisting of L-, M-, and S-cones. (D) A mosaic with eccentricity-dependent cone density and cone inner-segment diameter/outer-segment length, also with L-, M-, and S-cones. In the mosaics depicted in (B–D), cones at zero eccentricity are separated by 2 μm, with a corresponding peak theoretical cone density of 287,675 cones/mm^2. This is near the high end of the cone-density range in human subjects (100,000–324,000 cones/mm^2) reported by Curcio et al. (1990). The aperture-to-cone-spacing ratio is 0.79, close to the 0.82 value suggested by Miller and Bernard (1983) and Curcio et al. In (C–D), the L-:M-:S-cone ratio is 0.62:0.31:0.07, with a central region free of S-cones, and S-cone spacing outside of this central region constrained to be relatively regular. (E) Contrast-sensitivity functions for different mosaics computed for a 3-mm pupil and the point-spread function used by Banks et al. Gray, red, blue, and green disks depict the contrast-sensitivity functions for the mosaics shown in (A–D). Magenta disks depict the contrast-sensitivity function computed for a variant of the mosaic shown in (D), in which cone excitations were corrected for the effect of varying macular pigment density with eccentricity.

dependent cone mosaic, depicted in Figure 4D, had, in addition, eccentricity-dependent changes in the cone inner-segment diameter and outer-segment length. Computation details are provided under Eccentricity-dependent cone-efficiency correction.

Figure 4E shows the effect of these cone-mosaic properties on the ideal-observer CSF. The CSF plotted in gray replots the simulation of the Banks et al. (1987) constant-density mosaic from Figure 3B. The red, blue, and green plots show the CSFs obtained with the eccentricity-dependent mosaics. Note that at low spatial frequencies these deviate systematically from the CSF of the constant-density mosaic. The size of the deviation is quantified in the sensitivity-ratio plots in the bottom panel. The effect of mosaic density is most pronounced for the two mosaics with constant inner-segment diameter and outer-segment length. For these mosaics, the drop in relative sensitivity occurs because in the constant-cycle paradigm, low-frequency stimuli extend further into the periphery, where cone density is lower. This leads to lower total cone excitations in response to the stimuli compared to the constant-density mosaic, and thus lower ideal-observer sensitivity. The addition of S-cones has a negligible effect.

The drop in sensitivity for the low spatial frequencies is mitigated for the mosaic that implements an increase in cone inner-segment diameter and a decrease in outer-segment length with increasing eccentricity (Figure 4D). The net effect of the change in these factors is to increase the number of excitations per cone as eccentricity increases, partially offsetting the reduction in total excitations caused by reduction in cone density. Even for this mosaic, however, there is a notable decrease in low-spatial-frequency sensitivity compared to the constant-density-mosaic CSF.

In the four mosaics depicted in Figure 4A–4D, the macular pigment density does not change with eccentricity. To examine the effect of eccentricity-dependent changes in macular pigment density, we generated an even more realistic mosaic, whose properties were identical to those of the mosaic displayed in Figure 4D except that computation of cone excitations was corrected for the effect of macular pigment changes with eccentricity. This correction is described under Eccentricity-dependent macular pigment-density correction. The purple disks in Figure 4E depict the resulting CSF. As can be seen, incorporating realistic values for the macular pigment at different eccentricities does not have a significant impact on the achromatic CSF. This is expected, as the main effect of the macular pigment is on the excitation rates for S-cones, which are sparse enough not to play an important role. For most of the remaining calculations, we use the eccentricity-dependent mosaic shown in Figure 4D with the eccentricity-dependent macular pigment correction.

Optics

There have been significant improvements in the ability to measure the optical quality of the eye since early measurements of the human line-spread function (Westheimer & Campbell, 1962; Campbell & Gubisch, 1966). In particular, wave-front aberration measurements in individual human eyes (Liang & Williams, 1997) enable calculation of the corresponding PSFs (Thibos, Ye, Zhang, & Bradley, 1992; Goodman, 2005; Watson, 2015).

We examined how wave-front-aberration-based PSFs affect the derived CSF and contrasted this to the CSF derived by Banks et al. (1987). To do so we used the Thibos et al. (1992) data set, which includes a good sample of on-axis wave-front aberration measurements. However, as Thibos et al. (1992) point out, direct averaging of the PSFs (or of the Zernike polynomial coefficients) results in a PSF that differs qualitatively from any of the underlying measurements: The averaging process removes the idiosyncratic PSF structure found in most eyes. In addition, the optical modulation transfer function (MTF; the absolute value of the complex optical transfer function) obtained from the average of the individual-eye Zernike coefficients is sharper than the average of the optical MTFs obtained from the same set of coefficients. This happens because the mean Zernike coefficient for defocus is near zero; some subjects have positive defocus whereas others have negative defocus, and these cancel when the Zernike coefficients are averaged. Given these issues, we decided to compute CSFs based on PSFs from five individual subjects selected to cover the range of PSFs reported by Thibos et al. (2002). The selection process is described in detail under Selecting representative Thibos subjects. The results of this analysis are shown in Figure 5.

Figure 5A depicts the PSF used by Banks et al. (1987), and Figure 5B–5F depicts the PSFs (at 550 nm) of the five subjects selected from the Thibos data set. Here, all PSFs were computed assuming a 3-mm pupil. Note that the PSF used by Banks et al. has no dependence on wavelength, whereas the wave-front-derived PSFs account for both higher order aberrations and longitudinal chromatic aberration. Figure 5G compares the ideal-observer CSF obtained with the PSF used by Banks et al. to the ideal-observer CSFs obtained using optics of the five selected Thibos subjects. Note that all CSFs agree at low spatial frequencies, but the wave-front-derived CSFs fall off less rapidly than the Banks et al. CSF for spatial frequencies above 5 c/°. This difference is substantial at frequencies above 30 c/°, approaching a factor of 5 at 60 c/°. The higher sensitivity arises because the wave-front-derived PSFs (Figure 5B–5F) are somewhat narrower than the PSF used by Banks et al. (Figure 5A).

It should also be noted that wave-front-derived PSFs are typically rotationally nonsymmetric, and we have found that this asymmetry results in CSFs that are rotationally symmetric at low spatial frequencies and become progressively less so as spatial frequency is raised (typically beyond 16 c/°; data not shown). Overall, these results show that variations in optics may lead to considerable individual variation in the CSF at high spatial frequencies.

In addition to wave-front-derived PSFs, we also examined other wavelength-dependent optics models: the model developed by Navarro, Artal, and Williams (1993), which was based on a double-pass method with a 4-mm pupil (larger than the 3-mm pupil conditions we study), and the model developed by Marimont and Wandell (1994), which was based on a monochromatic MTF reported by Williams, Brainard, McMahon, and Navarro (1994) for a 3-mm pupil. The CSFs based on these two models (data not shown) drop more rapidly

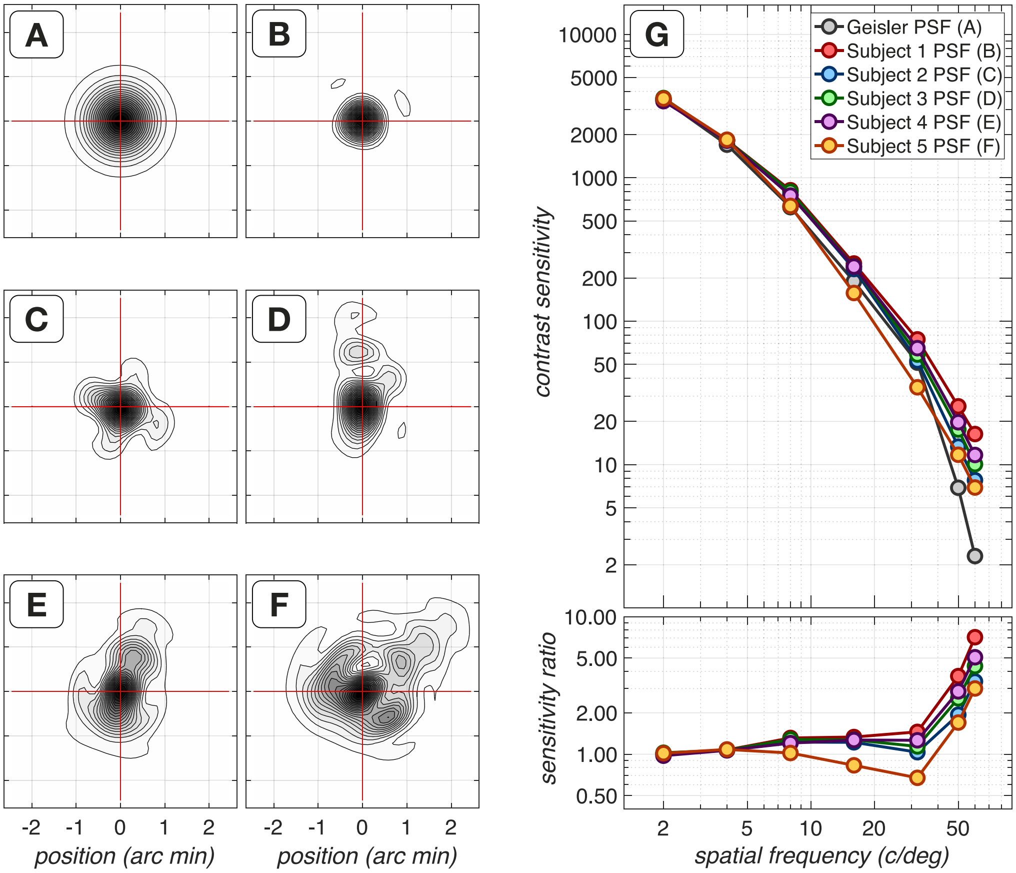

Figure 5. Effects of optics. (A–F) Contour plots of the point-spread functions (PSFs) used at 550 nm. Note that the Banks et al. (1987) PSF, displayed in (A), is identical across all wavelengths. In contrast, the five individual Thibos et al. (2002) subject wave-front-aberration-derived PSFs displayed in (B–F) vary with wavelength. We take the PSF of Subject 3, depicted in (D), to represent typical human optical quality. (G) Contrast-sensitivity functions for five individual PSFs, compared with that obtained using the PSF of Banks et al. For these calculations we used the eccentricity-dependent density and efficiency LMS-cone mosaic with corrections for the eccentricity-dependent reduction in macular pigment density.

with spatial frequency than those based on the Thibos et al. (2002) optics, and faster than the optics used by Banks et al. (1987), but in the larger view these effects are not large.

In summary, CSFs derived based on modern measurements suggest that typical observer optics enable a higher sensitivity at high spatial frequencies than the Banks et al. (1987) estimate. We selected the PSF of Subject 3 as a "typical" human PSF. All calculations from this point on were conducted using that PSF.

Inference engine

The ideal-observer calculations reported thus far characterize the information available in the mosaic excitations when the spatiotemporal dynamics of the mean response and the statistics of the noise for the test and null stimuli are known exactly. This analysis provides an upper bound on performance, but the signal-known-exactly assumption is unlikely to match the computations of the neural mechanisms that process the cone-mosaic signals. Therefore, it is important to examine how performance is affected by inference engines that learn suboptimal decision rules.

Toward this end, we employed inference engines based on an SVM classifier (Scholkopf & Smola, 2002; Manning, Raghavean, & Schutze, 2008), which uses labeled response instances to learn the parameters of a hyperplane that separates the visual representations of the test and null stimuli. Here, the visual representation is at the cone excitations and trial-to-trial variability is due to Poisson noise, and we have focused on that case. Variability can also arise due to other factors, such as fixational eye movements, fluctuations in pupil size and accommodation, and noise in the neural representation at sites central to the cone excitations.

Figure 6 illustrates the idea underlying the SVM-based inference engines in the context of our two-alternative forced-choice paradigm. Two scenes, one specifying the test stimulus S_t and one specifying the

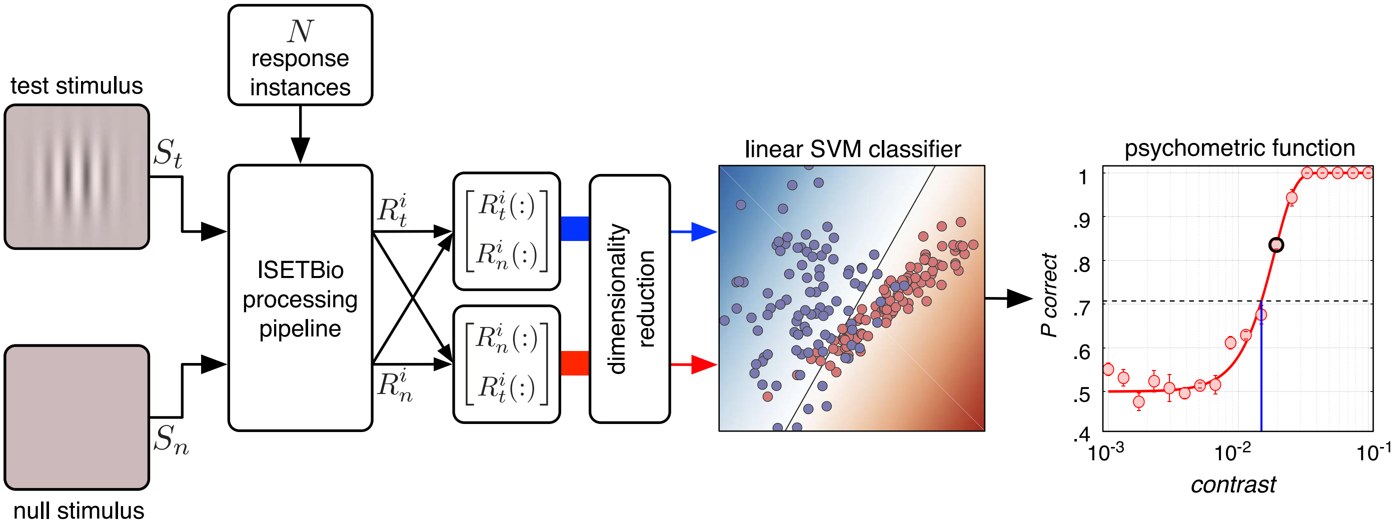

Figure 6. Illustration of inference engine based on a support-vector machine. Scenes describing the test stimulus S_t (top left) and the null stimulus S_n (bottom left) are constructed. Each is run through the ISETBio pipeline multiple times to produce N instances of cone-mosaic responses to each stimulus, R_t^i and $R_n^i, i = 1 \ldots N$. Each response instance includes an independent draw of Poisson isomerization noise. To simulate a two-alternative force-choice paradigm, composite response vectors are formed, with the response component to the test stimulus followed by the response component to the null stimulus, and vice versa. A dimensionality-reduction algorithm may be used to extract a low-dimensional feature set from these composite responses; in this illustrative example, a two-dimensional set is shown. The data are divided into training and evaluation sets. The training set is used to train a linear support-vector-machine classifier which learns the parameters of a hyperplane (shown as black line) that optimally separates instances of the two stimulus orders (null–test, red; test–null, blue). The performance of the classifier—its probability P_{correct} of correctly identifying the stimulus order—is then obtained on the evaluation set. This process is repeated for a series of stimulus contrasts, leading to a simulated psychometric function from which threshold is extracted. The black disk in the plotted psychometric function shows performance for the classifier illustrated in the figure. Threshold contrast (indicated by the blue line) is taken as the contrast that corresponds to $P_{\text{correct}} = 0.7071$ (black dashed line), based on a fit of a cumulative Weibull function to the simulated psychometric function.

null stimulus S_n are processed through the ISETBio simulation pipeline. A total of N response instances are computed for each of the test and null stimuli, $R_t^i, R_n^i,\ i = 1 \ldots N$. The samples for each stimulus differ because of Poisson noise. The sample data are divided into two sets, one used for training and the other for evaluation (held out data). Response vectors are formed by concatenating null and test excitations in the order of the two possible types of trials (test–null or null–test). For computational efficiency, a dimensionality-reduction algorithm may be used to extract a low-dimensional representation of responses; two dimensions are illustrated in Figure 6 (red and blue data points). A linear SVM classifier is trained to derive a separating hyperplane (black line) which maximizes the separation between the two types of trials, and the classifier's accuracy is evaluated on the held-out data. The entire procedure is repeated for a range of stimulus contrasts, defining a psychometric function (P_{correct} as a function of stimulus contrast). A cumulative Weibull function is fitted to the data, and the contrast level at $P_{\text{correct}} = 0.7071$ is considered the threshold. We compared the sensitivity of the ideal observer to that of different SVM-based computational-observer inference engines.

SVM-PCA inference engine

The first type of SVM-based inference engine we used reduces the dimensionality of the signals in the full cone mosaic to 60 by projecting response vectors to the space of the 60 principal components derived from the entire data set. We call this the SVM-PCA engine.

SVM-Template inference engines

The second type of SVM-based inference engine reduces the dimensionality of the signals in the full cone mosaic to 20—the number of 5-ms time bins within the 100-ms presentation time—by taking the inner product of the mosaic response at each time bin with a spatial-pooling template. We call this engine the SVM-Template-Linear inference engine. The spatial-pooling template for each spatial frequency is derived from the contrast profile of the test stimulus at that spatial frequency, as described under Inference engine in the Methods (Figure 13). We also examined another

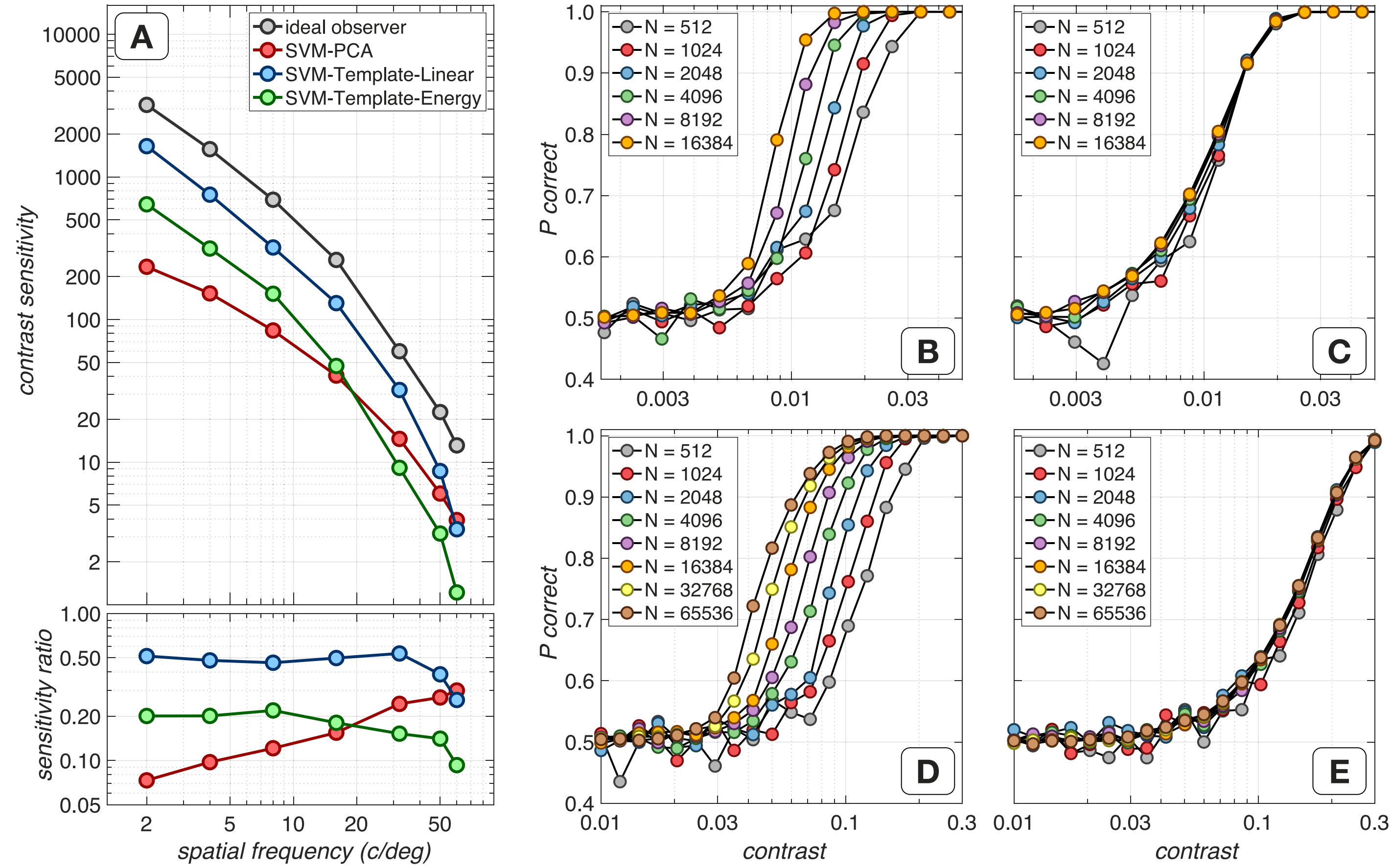

Figure 7. Effect of inference engine and training set size. (A) Effects of different inference engines on the contrast-sensitivity function. In these simulations we used the typical subject points-spread function (Subject 3 from Figure 5/Table 1), the LMS-cone mosaic with eccentricity-dependent cone density/efficiency/macular pigment density, and a data set consisting of 1,024 response instances. Note that the various support-vector-machine-based inference engines are 2–15 times less sensitive than the ideal-observer signal-known-exactly inference engine. (B–C) Psychometric functions for the SVM-PCA and SVM-Template-Energy inference engines, respectively, for the 8-c/° stimulus computed using training data sets of different sizes (512–16,384 instances). (D–E) Psychometric functions for the SVM-PCA and SVM-Template-Energy inference engines, respectively, for the 32-c/° stimulus, computed using training data sets of different sizes (512–65,536 instances). The psychometric curves in (B–E) were obtained using the mosaic shown in Figure 4C and the typical subject point-spread function.

variant of the SVM-Template engine, which employed not one but two spatial-pooling templates, with the first template being matched to the stimulus contrast profile (as in the SVM-Template-Linear inference engine) and the second one being derived from the stimulus spatial quadrature, here a sine-phase grating. In this inference engine, the inner products of the mosaic response with each of the two templates are computed, and the resulting responses are squared and then summed. This inference engine is inspired by the energy model of V1 complex cell receptive fields (Ohzawa, DeAngelis, & Freeman, 1990; Emerson, Bergen, & Adelson, 1992), and we call it the SVM-Template-Energy inference engine.

Figure 7A depicts the performance of the different inference engines we examined. In these simulations we used a 3-mm pupil, the typical human PSF (Subject 3 of Figure 5), and the cone mosaic with eccentricity-dependent cone density, efficiency, and macular pigment. We make a number of observations. First, as expected, the performance of SVM classifiers is worse than the performance of the ideal observer, with sensitivity ratios that vary between 0.07 and 0.5 across the spatial-frequency range (Figure 7A, bottom panels). Second, the sensitivity ratios of both SVM-Template inference engines are roughly constant with spatial frequency, whereas the ratio of the SVM-PCA inference engine varies with spatial frequency. Third, the performance of the SVM-PCA inference engine is worse than that of the SVM-Template-Energy inference engine for low spatial frequencies, but better for spatial frequencies above 16 c/°. And fourth, the SVM-Template-Linear engine is 2 to 3 times more sensitive than the SVM-Template-Energy engine.

These performance differences between the different inference engines are consistent with the amount of

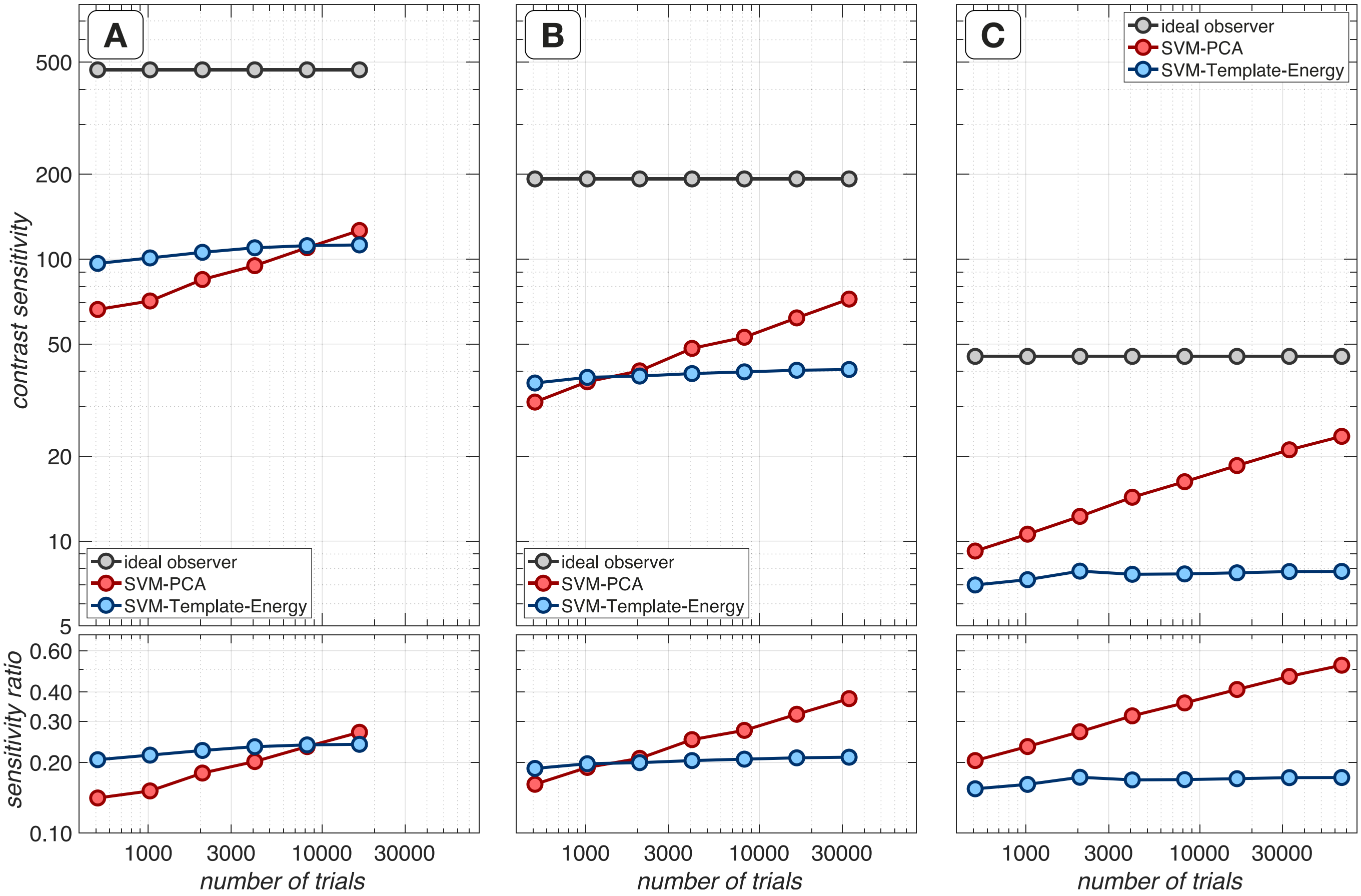

Figure 8. Dependence of computed contrast thresholds on the number of training response instances. Data for the ideal observer and the SVM-PCA and SVM-Template-Energy inference engines are depicted in gray, red, and blue disks, respectively. (A) Spatial frequency: 8 c/°. (B) Spatial frequency: 16 c/°. (C) Spatial frequency: 32 c/°.

information that these engines must learn from the training set. Unlike the ideal observer, the SVM-based inference engines have to learn the structure of the noise and the optimal criterion to apply to the underlying decision variable. Moreover, the SVM-PCA inference engine has no information regarding the stimulus, whereas the SVM-Template-Linear inference engine has knowledge of the stimulus spatial structure but not its contrast. Further, the SVM-Template-Energy inference engine has only partial knowledge of the stimulus spatial structure—the energy operation removes information regarding stimulus spatial phase.

The spatial-frequency dependence of the SVM-PCA inference engine's performance may be due to an interaction between stimulus dimensionality and learning. At low spatial frequencies, the activated mosaics are large and the response vectors have a high dimensionality. In this case, the SVM-PCA classifier might be inefficient when we use only 1,024 response instances to extract the principal components and train the SVM. On the other hand, 1,024 instances of the smaller dimensionality responses to higher spatial-frequency stimuli appears sufficient to train a good classifier, resulting in a relative increase in performance with spatial frequency. We suspect that the performance of both SVM-Template inference engines is approximately constant with spatial frequency, relative to the ideal observer, because these engines are provided with information about the spatial structure of the stimuli.

To investigate further, we examined the performance of the SVM-PCA and SVM-Template-Energy inference engines as a function of training set size for two spatial frequencies: 8 and 32 c/°. The psychometric curves of the SVM-PCA inference engine depend strongly on the data set size, shifting to the left as the data set size increases (Figure 7B and 7D). The performance of the SVM-Template-Energy inference engine is relatively stable, changing only slightly with the size of the training set (Figure 7C and 7E).

Figure 8 quantifies the effect of training-data set size on the computed contrast sensitivity for three stimuli (8, 16, and 32 c/°), for the ideal observer, the SVM-PCA, and the SVM-Template-Energy computational observers. Ideal-observer performance does not depend on the number of trials because it is computed

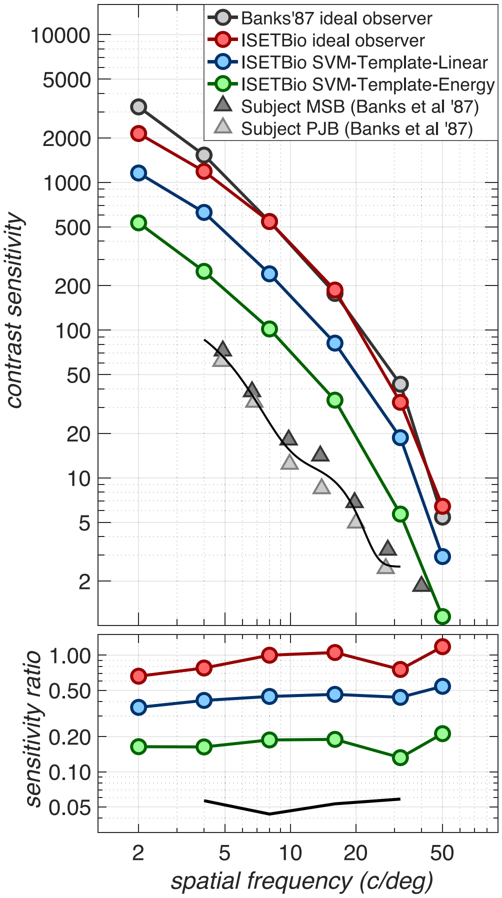

Figure 9. Comparison of computation-observer-derived contrast-sensitivity functions (CSFs) to CSFs measured in humans. All CSFs are for 2-mm pupils. The ideal-observer CSF was derived using the parameters of Banks et al. (1987) and is shown in gray disks (replotted from Figure 3A). The red disks depict the CSF derived using the eccentricity-dependent mosaic (Figure 4D) with eccentricity-dependent macular pigment corrections, the typical wave-front-based optics (Figure 5E), and the ideal-observer inference engine. This CSF exhibits a modest relative sensitivity decrease at the lowest spatial frequencies but is otherwise close to that computed by Banks et al. A twofold drop in sensitivity occurs when the inference engine is switched to the SVM-Template-Linear inference engine (blue disks), and this drop increases to fivefold for the SVM-Template-Energy inference engine (green disks). The CSFs measured in real subjects by Banks et al. are shown in triangles, and the black line depicts the mean of these subjects' CSFs, estimated by fitting the subject data with a double exponential curve. The CSF measured in human subjects is lower than the SVM-Template-Energy CSF by a factor of 3–4.

analytically based on full knowledge of the mean responses and the noise distribution for each spatial frequency and contrast. For the SVM-PCA computational observer, which must learn both the mean responses and the statistics of the noise, performance increases with number of training trials, presumably because the generalizability of the separating hyperplane increases with more data. For the SVM-Template-Energy observer, however—whose spatial-pooling operation reduces uncertainty regarding the mean responses—performance is relatively stable with the number of trials, consistently about 20% that of the ideal-observer inference engine. SVM-PCA performance exceeds SVM-Template-Energy performance after 8,000, 1,500, and 500 trials, respectively, for the stimuli at 8, 16, and 32 c/°. So when the number of trials is high enough for the response dimensionality, SVM-PCA performs better than SVM-Template-Energy, since no information is thrown away by the spatial-pooling mechanism. Note that spatial pooling matched to the contrast profile of the test stimulus is not the ideal spatial pooling for the case of Poisson noise, where the appropriate template depends on stimulus contrast (Geisler, 1989).

If we extrapolate SVM-PCA performance with the number of trials, we approach ideal-observer performance after 16.0 million, 4.9 million, and 1.6 million trials, respectively, for the stimuli at 8, 16, and 32 c/°. Computing such large numbers of trials of high-dimensionality signals is prohibitive in terms of computational resources. Therefore, given the relative stability of the SVM-Template engines with respect to the size of the training data set, we decided to employ these classifiers for the full simulation in which we compare performance of computational to human observers.

Comparison of computational and human observers' performance

Figure 9 compares the performance of our computational observers to the human psychophysical data reported by Banks et al. (1987). For these comparisons, all CSFs were computed for a 2-mm pupil to match the psychophysics, and the various ISETBio CSFs were derived using our most realistic eccentricity-based mosaic and the optics of our typical Thibos subject (Subject 3, but with a PSF computed from the wave-front aberrations for a 2-mm pupil). We make two main observations.

First, there are modest changes in relative sensitivity between our realistic mosaic and optics (red disks) and our replication of the Banks et al. mosaic and optics (gray disks). These are a reduction at the lowest spatial frequencies, due to the eccentricity-varying cone density, and a slight increase at the highest spatial frequency examined (here, 50 c/°) due to the wave-front-based optics.

Second, the use of computational observers (blue and green disks) results in a major decrease in sensitivity across the spatial-frequency range, around twofold for the SVM-Template-Linear engine and fivefold for the SVM-Template-Energy engine. As mentioned before, these results are consistent with the amount of information provided to the different inference engines. Therefore, data-driven inference engines which learn suboptimal decision rules reduce sensitivity, bringing the computed CSFs closer to human measurements.

Discussion

The limits of spatial resolution

We applied the ISETBio computational methods to clarify how specific properties of the cone mosaic, the physiological optics, and the inference engine limit pattern resolution. Our results can be summarized as follows.

First, the spatial structure of the cone mosaic is an important factor in limiting the CSF. The CSF is commonly measured using gratings that cover different amounts of the mosaic, and the change in the cone density across the mosaic is dramatic for typical variation in size with frequency. This variation is partly compensated by the change in cone aperture, but even so there remain differences between computations based on uniform and eccentricity-dependent mosaics.

Second, modern wave-front measurements indicate better human optics than earlier measurements, and all else being equal, incorporating the wave-front measurements leads to less attenuation in the ideal-observer CSF at high spatial frequencies.

Third, the choice of inference engine has a large effect on the absolute level of performance. Certain choices bring the computational observer into closer agreement with measured performance (SVM classifiers). Other choices show that more information is available in principle (e.g., the signal-known-exactly ideal observer). The idea that behavior can be completely described based on the visual representation at the cone mosaic is, of course, wrong. But the ability to calculate the information available to an ideal or computational observer at specific stages of visual processing provides useful benchmarks to clarify the aspects of performance that require explanation in terms of other factors.

Future directions

We are currently investigating the impact of additional ISETBio computational modules on pattern sensitivity. These include models of fixational eye movements and of the nonlinear transformation from cone excitations to photocurrent (Cottaris et al., 2018). Further assessments of optical factors beyond the shift-invariant optics models employed in the present work, such as the effect of changes in the PSF with eccentricity (Polans, Jaeken, McNavv, Artal, & Izatt, 2015), are also planned. ISETBio also includes methods based on computer graphics and ray tracing that quantify the retinal images of three-dimensional scenes (Lian et al., 2018). Indeed, these ray-tracing methods allow for computation of the retinal image from the scene specification via a model eye, and will also enable us to consider off-axis pupils and model transverse chromatic aberration, as well as to model accommodation and depth perception.

Retinal and cortical visual processing transforms the cone excitations in many ways that affect visual performance. ISETBio is designed to be extensible, and the current implementation contains placeholders for models of multiple parallel mosaics of retinal bipolar and ganglion cells. For example, understanding the limits of color sensitivity may be accessible through these calculations. Opponent processing of signals from different cone classes is a key step in color coding (Stockman & Brainard, 2010), and quantifying how this combination is implemented in neural circuits remains elusive. Implementing image-computable models of bipolar and ganglion cells may also clarify where key gaps exist in our current knowledge of how these cells operate.

Applications

The enormous growth of the imaging industry is based on the ability to design and implement new optical and electronic devices; during this process, designers inevitably turn to vision science for guidance in setting parameters. Critical information includes the standard color observer (Judd & Wyszecki, 1975; Wyszecki & Stiles, 1982) and knowledge of pattern resolution (Geisler & Banks, 1995; Wandell, 1995; Watson & Ahumada, 2004, 2005) and position resolution (Westheimer, 1981; Klein & Levi, 1985; Jiang et al., 2017). The computational methods in ISETBio integrate quantitative models of scenes and display devices and are useful in supporting the design and evaluation of new imaging devices.

Medicine is a second important application area. As treatments for partial sight restoration become feasible, for example through gene therapy or retinal prostheses, it will be important to understand the degree to which the additional information provided to the nervous system by these technologies supports performance. The ability to use simulations to model the information carried by restored representations and understand the upper limits on visual performance available from them should facilitate the design of therapies that can ameliorate partial and full blindness (Cottaris & Elfar, 2005; Jiang, Wandell, & Farrell, 2015; Beyeler, Boynton, Fine, & Rokem, 2018; Golden et al., 2019).

Machine learning

An important step to understanding human performance is characterizing possible mechanisms that underlie the visual system's capacity for flexible and effective performance across a wide range of tasks. Human observers underperform ideal observers. One reason for this is that human observers do not have access to all the information about the stimulus available to an ideal observer. Therefore, characterizing performance using different learned decision rules is important. Although comparison to psychophysical CSF data alone may not distinguish between all plausible decision rules, at the very least we may be able to rule out candidate decision rules, such as ones that are sufficiently inefficient that they predict performance below human levels.

In this article we take only a small step in this direction, by exploring rather simple forms of learned rules, some of which (the template-based versions) share with the ideal observer the fact that knowledge about the stimulus is provided. We started with the SVM approach because it is relatively simple and because its learning is known to have good convergence properties. Moreover, spatial pooling by weighted sums is an approximate model for the receptive-field properties of several neuronal populations (Movshon, Thompson, & Tolhurst, 1978; Andrews & Pollen, 1979; Shapley, Kaplan, & Soodak, 1981; Wandell, 1995). Such inference engines implement a decision variable that is a weighted sum of the representational input (here, cone excitations). In addition, the linear classifiers are learned from training data. Basing decisions on the weighted sum of cone signals that are learned by linear classifiers may approximate the inference engines used by real observers' neural processing better than the ideal-observer calculation.

Striking to us is the substantial drop in performance observed when we learn aspects of the decision rule. This observation motivates future work that explores a wider range of learned decision rules. Other strategies that implement suboptimal decision rules are possible, such as spatial pooling based on banks of primary visual-cortex receptive fields, or classification images (Ahumada, 2002; Murray, 2011), in which spatial-summation templates are learned from a subset of the responses.

Another inference-engine approach that may be explored is deep neural networks. Several machine-learning successes use convolutional neural networks to analyze images (Kriegeskorte, 2015). The architecture and parameters of these networks offer inspiration about how to model cortical circuits, and conversely there are opportunities to explore how findings from cortical circuits might be used to implement artificial neural networks (Khaligh-Razavi & Kriegeskorte, 2014; Yamins et al., 2014). A limitation in the interaction between vision science and machine learning arises from the stimulus representation. Convolutional neural networks are typically trained using digital image values (RGB), which have no biological basis. The machine-learning work can be more closely integrated with biology by training on inputs comprising realistic visual signals. The ISETBio simulations are well suited for converting RGB images into retinal responses that serve as more biologically realistic inputs to train artificial neural networks.

Theory and computation

Theory is how we develop a principled understanding of complex systems. Computational models built on theoretical principles can provide additional insights about the impact of specific system components and deviations from the ideal. Coordination between theory and computation arises in many fields. Rocket design incorporates Newton's gravity formulation as well as computational models of material properties, friction, and heat. Telecommunications systems incorporate Shannon's information theory as well as information about switching times, conduction delays, and circuit noise.

In vision science, ideal-observer theory informs us how to conceptualize the inputs and decision variables that define system performance. With the enormous growth of computational power, this formal theory—which inevitably involves many approximations—can be extended to account for specific system characteristics. Modeling the impact of these system components is important for bridging basic discovery and applications, say for display engineering or medicine.

Methods

Stimulus

The simulated scenes were designed to match the stimuli used by Banks et al. (1987): cosinusoidal patterns windowed using a half-cosine spatial modulation which spans 7.5 cycles of the grating. The windowing makes the spatial extent of each stimulus inversely proportional to its spatial frequency, a choice motivated by the observations that contrast sensitivity increases with extent up to a critical size and the critical size is approximately constant when expressed in terms of stimulus cycles (Howell & Hess, 1978).

ISETBio scenes are spatially sampled spectral radiances. The stimuli employed by Banks et al. are specified in colorimetric units (x, y chromaticity and luminance). A spectral representation is necessary to model chromatic aberration, inert pigments, and absorption of light by the three classes of cone photoreceptors. To promote the colorimetric specification to spectral radiance, we simulated the scenes as arising from a typical color cathode-ray tube from the era when their article was published. The critical display information is the R-, G-, and B-channel spectral power distributions, the RGB display quantization, and the pixel spatial sampling. Because our interest here is not the effect of display properties per se, we modeled a cathode-ray tube with 18-bit linear control of the R, G, and B primary intensities. We also set the pixel spatial sampling to be inversely proportional to the stimulus spatial frequency; consequently, all stimuli were represented on a 512×512 spatial grid mapped onto the corresponding retinal region in a manner that took the stimulus size into account. The spectra we model differ somewhat from those in the Banks et al. experiment, as that experiment was performed using a monochrome cathode-ray tube with a P4 phosphor.

Retinal image

Physiological optics transform the scene spectral radiance to the retinal image (spectral irradiance). The transformation can be conveniently grouped into two parts. First, the scene radiance is transformed to an idealized retinal spectral irradiance. This transformation accounts for the pupil diameter (which controls the amount of light entering the eye), the stimulus distance and posterior nodal distance of the lens (which controls the retinal image magnification; Holst, 1989), and the lens pigment spectral transmittance, which reduces retinal irradiance in a wavelength-dependent manner (Stockman, Sharpe, & Fach, 1999). Second, the idealized retinal spectral irradiance is convolved with a wavelength-dependent PSF. The PSF is determined by monochromatic and chromatic aberrations of the optics as well as diffraction. Blurring by the wavelength-dependent PSF produces the retinal image.

In general, there are three types of optical aberrations: monochromatic aberrations, longitudinal chromatic aberration (LCA), and transverse chromatic aberration (TCA). Monochromatic aberrations produce complex deformations in the retinal image that vary between individuals. LCA is a wavelength-dependent defocus which occurs due to the wavelength-dependent refractive index of the ocular media. It is consistent across individuals and amounts to about 2.2 diopters of defocus across the spectrum in the range of 400–700 nm (Bedford & Wyszecki, 1957; Thibos et al., 1992; Marimont & Wandell, 1994; Cottaris, 2003). TCA causes a wavelength-dependent shift in the position and magnification of the retinal image. It results from changes in the index of refraction of the optical elements, combined with misalignment of these components. TCA varies between individuals and between the eyes of a given individual (Marcos, Burns, Moreno-Barriusop, & Navarro, 1999; Harmening, Tiruveedhula, Roorda, & Sincich, 2012). Because the optical axis of the eye is not always centered with its visual axis, TCA can be observed at the fovea. In some individuals, TCA can be more significant than LCA, whereas in other individuals it can be minimal (Marcos et al., 1999). In the present work we model monochromatic aberrations and LCA. We neglect TCA, as well as changes with wavelength in wave aberrations other than defocus (Marcos et al., 1999). In addition, we neglect light scatter due to the ocular media (Vos, 2003) and the Stiles–Crawford effect (Stiles & Crawford, 1933; Westheimer, 2008).

We model monochromatic aberrations using the first 15 Zernike polynomials, which were measured in a population of 200 human eyes (Thibos et al., 2002). From a set of Zernike polynomials, we can compute the wave-front aberration map at the in-focus wavelength (550 nm), and from this the corresponding PSF (Goodman, 2005; Watson, 2015). To generate the PSF for any other wavelength, we add a defocus term $d(\lambda)$ to the Zernike polynomials according to the formula given by Howarth and Bradley (1986):

$$d(\lambda) = 633.46 \times \left(\frac{1}{\lambda_{\text{focus}} - 214.1} - \frac{1}{\lambda - 214.1} \right),$$

where $\lambda_{\text{focus}} = 550$ nm. The computed PSF is translated in space so that its center of mass at the in-focus wavelength is centered at the origin. This is done so as to eliminate performance differences due to off-centered PSFs.

Selecting representative Thibos subjects

For the CSF simulations we wanted to choose wave-front-based optics for a typical subject. As noted previously, using a wave-front function based on the mean of the Zernike coefficients across all subjects is not satisfactory because the cancelation of positive and negative defocus coefficients in the averaging leads to a higher optical MTF than is observed in most subjects. At the same time, deriving typical optics directly from the mean MTF is not straightforward, because the MTF does not completely determine the spatial

OSA no.	Aberration name	Subject 1 (98)	Subject 2 (132)	Subject 3 (54)	Subject 4 (194)	Subject 5 (21)
0	Piston	0.5693	0.1946	−0.2132	0.0748	−0.4462
1	Vertical tilt	0.0782	−0.0137	0.2232	−0.0396	0.2318
2	Horizontal tilt	−0.2519	0.1446	−0.0162	0.0955	0.1148
3	Oblique astigmatism	0.0152	−0.0380	0.0366	−0.0187	−0.0295
4	Defocus	0.0000	−0.0042	0.0182	0.0821	0.1054
5	Vertical astigmatism	−0.0145	−0.0472	−0.0432	−0.0252	0.0158
6	Vertical trefoil	0.0079	−0.0188	−0.0437	−0.0046	0.0148
7	Vertical coma	0.0136	−0.0058	0.0515	0.0057	0.0176
8	Horizontal coma	−0.0071	0.0014	−0.0131	0.0019	−0.0082
9	Oblique trefoil	−0.0054	0.0006	0.0055	0.0126	0.0142
10	Oblique quadrafoil	0.0015	−0.0039	−0.0013	−0.0006	−0.0052
11	Oblique secondary astigmatism	−0.0037	0.0037	0.0008	0.0036	−0.0119
12	Spherical	−0.0019	0.0106	0.0076	0.0032	−0.0058
13	Vertical secondary astigmatism	−0.0020	−0.0032	−0.0106	0.0004	−0.0010
14	Vertical quadrafoil	0.0038	−0.0039	−0.0022	−0.0025	−0.0066

Table 1. Zernike coefficients for the five Thibos subjects. These are taken from the 3-mm-pupil data set of Thibos et al. (2002). The numbers within the parentheses next to each subject's number correspond to the index of the "OU" field in the data set, which contains left and right eyes for the population of 100 subjects. Data for pupils at 4.5, 6, and 7.5 mm are available for these subjects in the full Thibos et al. data set.

structure of the PSF, so additional assumptions are required.

To deal with this issue we computed CSFs using optics from five sample eyes from the Thibos et al. (2002) data set that were chosen to span the range of measured optical quality. The PSFs of these subjects are depicted in Figure 5, where we use the term *subject* to refer to a specific eye of a particular subject. The subjects are referred to as Subjects 1, 2, 3, 4, and 5; we selected Subject 3 as typical. The Zernike coefficients for these five subjects are provided in Table 1.

We selected the subjects by ranking the entire population of 200 eyes measured by Thibos et al. and choosing five subjects whose scores span the range of computed scores. Subject ranking was done as follows. First, we computed the singular-value decomposition of all subject PSFs, separately for each wavelength. This provided a basis set for representing the PSFs at each wavelength. We then projected each subject's PSF and the mean Zernike-coefficient PSF to the basis set, separately for each wavelength. A PSF matching score was computed for each subject based on the mean (over wavelengths) root-mean-square error between that subject's projection coefficients and the projection coefficients of the mean Zernike-coefficient PSF. Then we computed the mean MTF (absolute value of the optical transfer function) across all subjects. An MTF matching score was computed for each subject as the mean (over wavelengths) root-mean-square error between that subject's MTF and the mean MTF. The calculations were done for a 3-mm pupil. Subjects were ranked according to their PSF score (Figure 10), and the five representative subjects were selected as follows. Subject 1 was selected because their PSF best resembled the PSF obtained from the mean of the Zernike coefficients. Note that this subject's MTF score is very low. Subject 2 also has a high PSF score but a much higher MTF score. Subject 3 has PSF and MTF scores of similar magnitude. This is the subject we take to represent typical human optics. Subjects 4 and 5 have

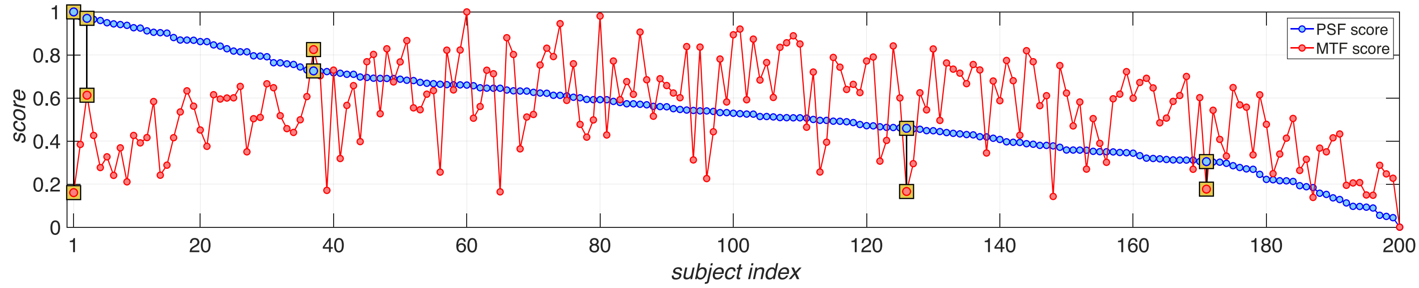

Figure 10. Selecting representative subjects. Scores on point-spread and modulation transfer functions for the population of the 200 Thibos subjects sorted according to their point-spread-function score. The five representative subjects are indicated by the black squares which connect their scores, with Subjects 1–5 running from left to right in the figure.

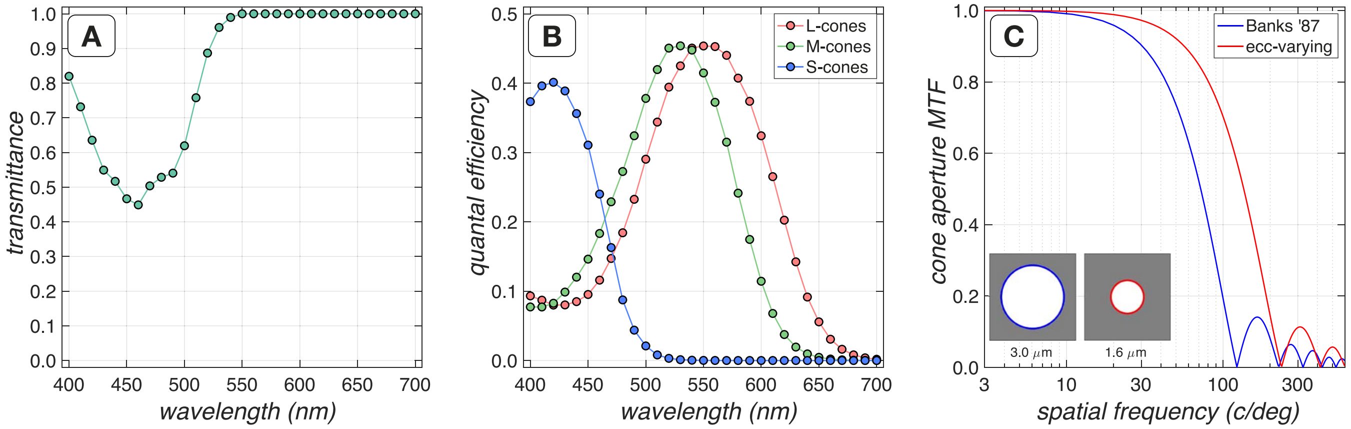

Figure 11. Components of the cone excitation response model. (A) Foveal macular pigment transmittance as a function of wavelength, $T_{\text{macular}}(\lambda)$. (B) Foveal spectral quantal efficiencies $q_c(\lambda)$ for $c = L$-, M-, S-cones. (C) Foveal cone-aperture filters (inset) and corresponding modulation transfer functions for the inner-segment diameters considered in this article.

progressively worse PSF scores and low MTF scores (Figure 10).

Computation of cone excitations (photon isomerizations)

The main factors that determine how the retinal image $RI(x, y, \lambda)$ is transformed into a pattern of cone photoisomerization rates are the macular pigment, which differentially absorbs short-wavelength photons; the spectral quantal efficiency, or spectral absorptance, of the cone photopigment, which controls the proportion of incident photons that get absorbed by the photopigment; the cone aperture diameter, which determines the photon-collecting area of a cone and also acts as a spatial low-pass filter; and the cone lattice, which controls the spatial sampling of the retinal irradiance image.

The foveal macular pigment transmittance, $T_{\text{macular}}(\lambda)$, is depicted in Figure 11A. Minimum transmittance is 0.45 at 460 nm. The foveal spectral quantal efficiencies (absorptances) of different cone classes $q_c(\lambda)$, with $c = \{L, M, S\}$, are depicted in Figure 11B and are computed based on the Stockman–Sharpe normalized absorbance values $SS_c(\lambda)$ (Stockman et al., 1999; Stockman & Sharpe, 2000), as

$$q_{c_k}(\lambda) = q_{\text{peak}} \times \left(1 - 10^{-OD_{c_k} \times SS_{c_k}(\lambda)}\right), \quad (1)$$

where q_{peak} is the peak cone quantal efficiency (0.667 for all cone types) and OD_{c_k} is optical density of cone type c_k (0.5 for L- and M-cones and 0.4 for S-cones). These values are within the range of optical densities reported (0.29–0.91 for L-cones, 0.36–0.97 for M-cones; Renner, Knau, Neitz, Neitz, & Werner, 2004).

Cones exhibit waveguide properties (Enoch, 1961), according to which light incident on the cone inner segment is guided to the outer segment, where it gets absorbed. To model this, we employed a spatially uniformly weighted circular averaging filter $A(x, y)$, whose diameter corresponds to the inner-segment diameter and whose volume is 1. In the Banks et al. (1987) mosaic, the inner-segment diameter is 3 μm, whereas in the ISETBio mosaics it is 1.6 μm in the fovea. The aperture filters and corresponding MTFs for these mosaics are depicted in Figure 11C. Note that the MTF at 60 c/° is 0.63 for the Banks et al. mosaic and 0.89 for the eccentricity-varying cone mosaics. Although we varied the size of the inner-segment diameter with eccentricity when we computed cone excitations, we used a constant inner-segment diameter (foveal value) when computing blur by the cone apertures. This choice was made for reasons of computational efficiency. We have verified, however, that using the mean aperture value across all cones in a mosaic produces essentially indistinguishable contrast-sensitivity curves, as the effects of the optical PSF dominate the effects of the aperture.

To compute the spatial distribution of the cone excitation rate $CER_c(x, y)$ for each cone class c, the retinal image $RI(x, y, \lambda)$ was first filtered with the macular pigment transmittance $T_{\text{macular}}(\lambda)$, multiplied by the corresponding spectral quantal efficiency $q_c(\lambda)$ and integrated numerically over wavelength. The result was then spatially convolved with the cone aperture $A(x, y)$:

$$CER_c(x,y) = \left(\int_{\lambda} RI(x,y,\lambda) \times T_{\text{macular}}(\lambda) \times q_c(\lambda)\delta\lambda\right) \odot A(x,y). \quad (2)$$

To compute the cone-mosaic excitations we estimate the mean count of excitation events for each cone within the simulation time interval, here $\tau = 5$ ms.

Specifically, for a cone k of class c_k located at coordinates (x_k, y_k), the mean count of cone excitation events $\overline{CE}(k)$ within τ ms is computed by spatially sampling the continuous function $CER_{c_k}(x, y)$ at $(x = x_k, y = y_k)$, multiplying by the cone inner-segment area α and the time interval τ:

$$\overline{CE}(k) = CER_{c_k}(x = x_k, y = y_k) \times \alpha \times \tau. \quad (3)$$

Finally, an excitation response instance i for the kth cone, $CE^i(k)$, is generated by sampling from a Poisson distribution whose mean is equal to the mean count of excitation events:

$$CE^i(k) = \text{Poisson}\left(\overline{CE}(k)\right). \quad (4)$$

Eccentricity-dependent macular pigment-density correction

To model eccentricity-dependent variation in macular pigment density, we replaced $T_{\text{macular}}(\lambda)$ in Equation 2 with

$$T_{macular}(x, y, \lambda) = 10^{-OD_{\text{macular}}(x,y) \times SS_{\text{macular}}(\lambda)}, \quad (5)$$

where $SS_{\text{macular}}(\lambda)$ is the spectral sensitivity of the macular pigment and $OD_{\text{macular}}(x, y)$ is a factor describing the eccentricity-dependent variation in the optical density of the macular pigment, modeled as by Putnam and Bland (2014):

$$OD_{\text{macular}}(x, y) = OD_{\text{macular}}(0, 0) \times \frac{3.6028}{x^2 + y^2 + 3.6028}. \quad (6)$$

In Equation 6, $OD_{\text{macular}}(0, 0)$ is the optical density of the macular pigment at the fovea, which is set to 0.35 (Putnam & Bland, 2014).

Eccentricity-dependent cone-efficiency correction

This computation of isomerization does not take into account the fact that as eccentricity increases, inner-segment area increases (Curcio et al., 1990) and outer-segment length decreases (Banks, Sekuler, & Anderson, 1991; Jonnal et al., 2017). We approximated these effects by defining an eccentricity-dependent correction factor b_k for the kth cone located at (x_k, y_k), defined as

$$b_k = b^{IS}(x_k, y_k) \times b^{OS}_{c_k}(x_k, y_k), \quad (7)$$

where

$$b^{IS}(x_k, y_k) = \frac{\alpha(x_k, y_k)}{\alpha(0, 0)}$$

is the correction factor required to account for the change in inner-segment area $\alpha(x_k, y_k)$ at location (x_k, y_k) relative to its foveal value $\alpha(0, 0)$. The quantity $b^{OS}_{c_k}(x_k, y_k)$ is the correction factor required to account for the decrease in outer-segment length for cone class c_k at location (x_k, y_k) relative to its foveal value, and is computed as the mean value of $b^{OS}_{c_k}(x_k, y_k, \lambda)$ over the wavelength parameter λ, with

$$b^{OS}_{c_k}(x_k, y_k, \lambda) = \frac{\text{quantal efficiency of cone } c_k \text{ at } (x_k, y_k)}{\text{quantal efficiency of cone } c_k \text{ at } (0, 0)} = \frac{1 - 10^{-OD^e_{c_k}(x_k, y_k) \times SS_{c_k}(\lambda)}}{1 - 10^{-OD_{c_k} \times SS_{c_k}(\lambda)}} \quad (8)$$

and

$$OD^e_{c_k}(x_k, y_k) = OD_{c_k} \times \frac{\text{outer segment length at } (x_k, y_k)}{\text{outer segment length at } (0, 0)}. \quad (9)$$

Ideally, $b^{OS}_{c_k}(x_k, y_k, \lambda)$ should be applied within the integral of Equation 2. For ease of computation, we use the mean value over all wavelengths of $b^{OS}_{c_k}(x_k, y_k, \lambda)$ and apply the b_k correction factor (Equation 7) to the mean count of excitation events $\overline{CE}_k$ to update the quantity computed by Equation 3:

$$\overline{CE}_k \leftarrow \overline{CE}_k \times b_k. \quad (10)$$

This allows us to compute a cone excitation count which takes into account the eccentricity-dependent changes in cone efficiency due to changes in inner-segment aperture and outer-segment length. In these computations, we assume that photopigment concentration and extinction coefficients remain constant across eccentricity, and we ignore photopigment bleaching, which is small at these light levels (Rushton & Henry, 1968).

Cone mosaics

We examined two types of hexagonal cone mosaics: the regularly spaced hexagonal mosaic used by Banks et al. (1987), in which cone density is constant across all eccentricities with a spacing of 3 μm and an inner-segment diameter of 3 μm, and eccentricity-dependent mosaics, in which cone density varies with eccentricity. In eccentricity-dependent mosaics, the desired cone spacing at the foveola is 2 μm. This corresponds to a theoretical peak cone density of 287,675 cones/mm^2, which is near the high end of the cone-density range in human subjects (Curcio et al., 1990). In practice, the eccentricity-based mosaics used in the primary calculations, which are synthesized stochastically as described later, had peak cone sensitivities in the range of 270,353–290,448 cones/mm^2. The ratio of cone diameter to inner-segment aperture was 0.79 across all

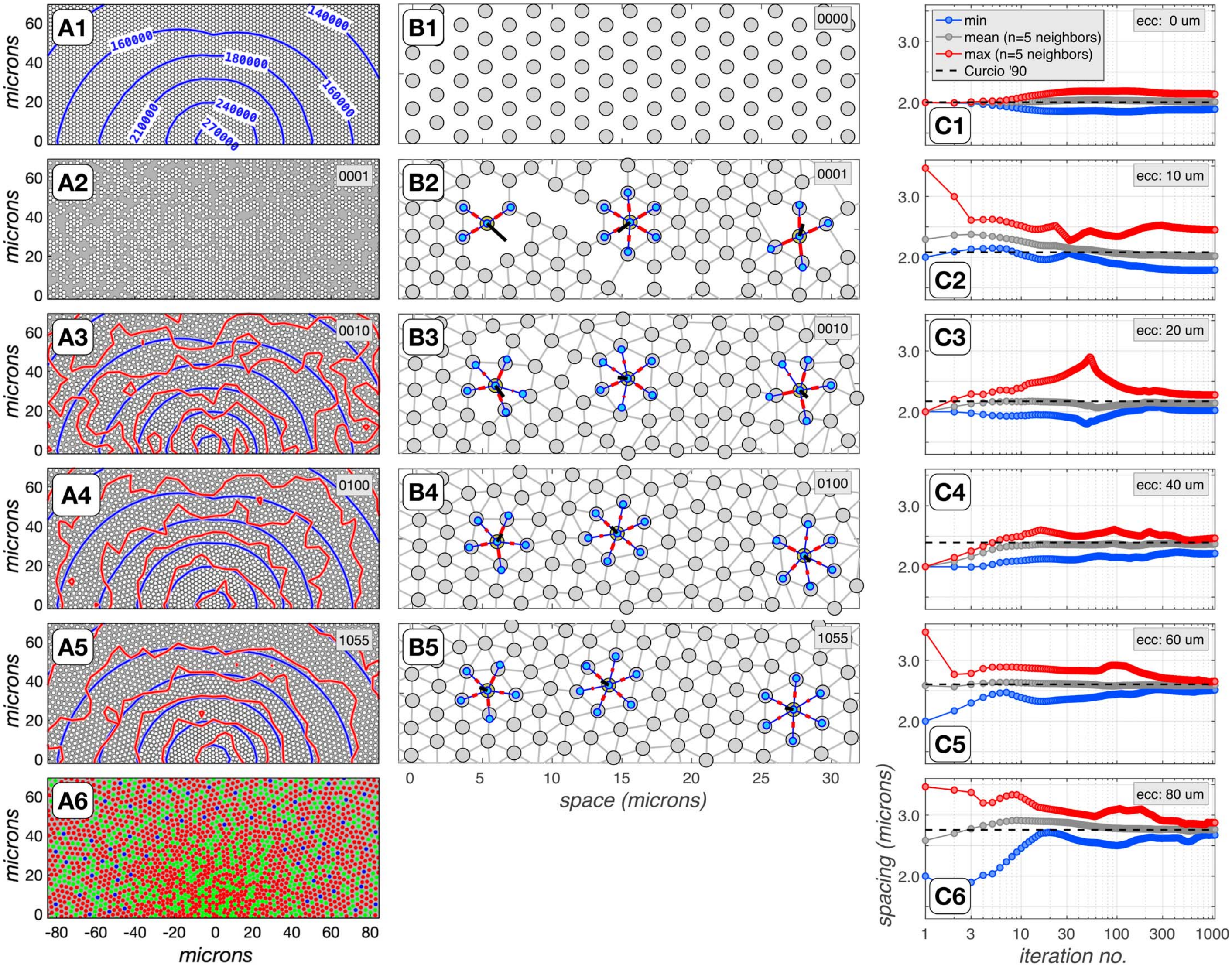

Figure 12. Generation of approximately hexagonal mosaics with eccentricity-varying density. (A1–A6) Snapshots of the mosaic at different iteration stages. (A1) Initialization with a regular hexagonal lattice of the highest density. Blue lines depict isodensity contours (in cones/mm^2) for the desired density profile (Curcio et al., 1990). (A2) Probabilistic subsampling according to the desired density (Iteration 1). (A3–A5) Iterative lattice adjustment. Red lines depict the isodensity contours for the actual mosaic. (A6) Cone-type labeling with L-, M-, and S-cones depicted in red, green, and blue disks, respectively. (B1–B5) High-resolution snapshots of the lattice-adjustment process. The red line segments depict the mutually repulsive forces between select cone pairs, with segment length denoting force magnitude, and the thick black lines depict the net forces, which determine cone movement. The blue disks represent the desired spacing, and the gray line segments the actual spacing. (C1–C6) Analysis of minimum, mean, and maximum cone spacing (within five neighbors) for exemplar cones positioned at horizontal distances of 0, 10, 20, 40, 60, and 80 μm from the fovea. Note that the minimum, mean, and maximum cone spacing are all converging toward the desired cone spacing, which is indicated by the dashed lines.

eccentricities, close to the 0.82 suggested by Miller and Bernard (1983; see also Curcio et al., 1990).

We developed a novel approach for generating eccentricity-dependent hexagonal mosaics. In this approach, a cone mosaic is initialized using a regular hexagonal lattice with node spacing equal to the foveal cone separation $\sigma_{0,0} = 2$ μm (Figure 12A1 and 12B1). In the first iteration, the lattice is subsampled, and a node located at (x, y) is eliminated with an eccentricity-dependent probability

$$P_{\text{reject}} = 1 - \frac{1}{\left(\sigma_{x,y}/\sigma_{0,0}\right)^2}, \quad (11)$$

where $\sigma_{x,y}$ is the desired cone spacing at (x, y), taken from Curcio et al. (1990). The subsampled spatial mosaic approximates the desired eccentricity-dependent cone density, but cone coverage is nonuniform, with regions without any cones (Figure 12A2 and 12B2). To improve the uniformity of the mosaic, an iterative procedure is used (Persson, 2005). In this approach, a cone and its neighboring cones are subjected to simulated movement driven by mutually repulsive forces. The magnitude of the repulsive force between a target cone i and a neighboring cone j is given by

$$\left|F_j^i\right| = \begin{cases} k \times \left(\sigma_{i,j}^{\mathrm{desired}} - \sigma_{i,j}^{\mathrm{actual}}\right) & \text{if } \sigma_{i,j}^{\mathrm{actual}} < \sigma_{i,j}^{\mathrm{desired}}, \\ 0 & \text{otherwise.} \end{cases} \qquad (12)$$

In Equation 12, $\sigma_{i,j}^{\mathrm{desired}}$ is the desired cone spacing at the midpoint between these cones, $\sigma_{i,j}^{\mathrm{actual}}$ is their actual separation, and k is set to a value >1. Therefore, when the spacing between two cones is smaller than the desired spacing, a positive force is generated which tends to pull these cones apart, whereas when the spacing is larger than the desired one, no force is acting between them. The mutually repulsive forces spread cones around the mosaic, filling in regions with no cones (Figure 12A2–12A5 and 12B2–12B5). Cones moved outside of the mosaic boundary are forced back inside the boundary. To avoid irregularities at the mosaic edges, the extent of the boundary is usually 20% larger than the width of the mosaic to be generated.

Cone position is updated based on the net force from its K neighbors, $F_{net}^i = \sum_{j=1}^{K} F_j^i$, where the K neighboring cones are determined by Delaunay triangularization. The update rule is $p^i \leftarrow p^i + \delta \cdot F_{net}^i$, where δ is the update step, set to $0.2 \times \sigma_{0,0}$. The iterative position-adjustment process is terminated when nodes move less than a threshold value. Snapshots of the mosaic at iterations 10, 100, and 1,055 are depicted in Figure 12A3–12A5, along with the isodensity contour lines of the achieved and the desired cone-density profiles.

In the final step, cones are assigned a type L, M, or S, depending on the specified L/M/S-cone density property, as well as the desired S-cone mosaic properties, such as a minimum distance between neighboring S-cones and the size of a central region free of S-cones (Figure 12A6).

A good agreement in the density profile is obtained at convergence (here, 1,055 iterations). Figure 12C1–12C6 depicts a detailed analysis of neighboring cone spacing for six exemplar cones positioned at horizontal distances of 0, 10, 20, 40, 60, and 80 microns from the fovea. Note that the mean cone spacing (across five neighboring cones) always converges to the desired cone spacing as the number of iterations approaches 1,000. The minimum and maximum (across five neighboring cones) spacing are also closely matched to the desired cone spacing.

An alternative method to generate eccentricity-based cone mosaics has been proposed by Bradley, Abrams, and Geisler (2014). That method uses various heuristics to position cones along isodensity contours around the fovea, ensuring that cones are no closer than the spacing implied by a density model. A direct comparison of how mosaics generated by these two methods compare to real mosaics remains an interesting topic for future work (see also Cooper, Wilk, Tarima, & Carroll, 2016).

Inference engine

An ideal inference engine for cone excitations modeled as Poisson processes is constructed from knowledge of the mean isomerization counts to the test and null stimuli. This signal-known-exactly calculation defines an upper bound on the information that can be extracted. For more realistic calculations—such as accounting for uncontrolled fixational eye movements—the cone excitations have additional uncertainty, and across trials the noise is no longer Poisson. With these additional terms, a simple closed-form mathematical expression describing the cone excitation signals across trials may be beyond our reach.

For the general case, it is possible to choose an inference engine that learns from training examples. In this study we use SVMs (Scholkopf & Smola, 2002; Manning et al., 2008) that learn a linear classifier (Figure 6). A general challenge in the implementation concerns the high dimensionality of the cone excitations. In this study, for example, the smallest cone mosaic had 7,460 dimensions (20 time bins × 373 cones) and the largest had 1,400,040 dimensions (20 time bins × 70,002 cones). To efficiently train inference engines based on SVM linear classifiers, we applied dimensionality-reduction techniques. In the present study, we examined three different approaches.

SVM-PCA classifier

For the first dimensionality-reduction technique, we computed the first 60 principal components of the two composite responses $\left[R_t^i; R_n^i\right]$ and $\left[R_n^i; R_t^i\right]$. The principal-components analysis was performed separately for each spatial frequency and contrast. Binary SVM classification was performed on the projections of the response instances into the space spanned by the 60 principal-components analyses. We refer to this classifier as the SVM-PCA classifier.

SVM-Template-Linear classifier

In the second dimensionality-reduction technique, we employed spatial pooling of cone responses via a weighting kernel, or template, $V(k), k = 1 \ldots M$, where M is the number of cones in the mosaic. The spatial profile of $V(k)$ was derived from the spatial contrast modulation of the test stimulus: The weight associated with cone k was the spatial contrast of the test stimulus at

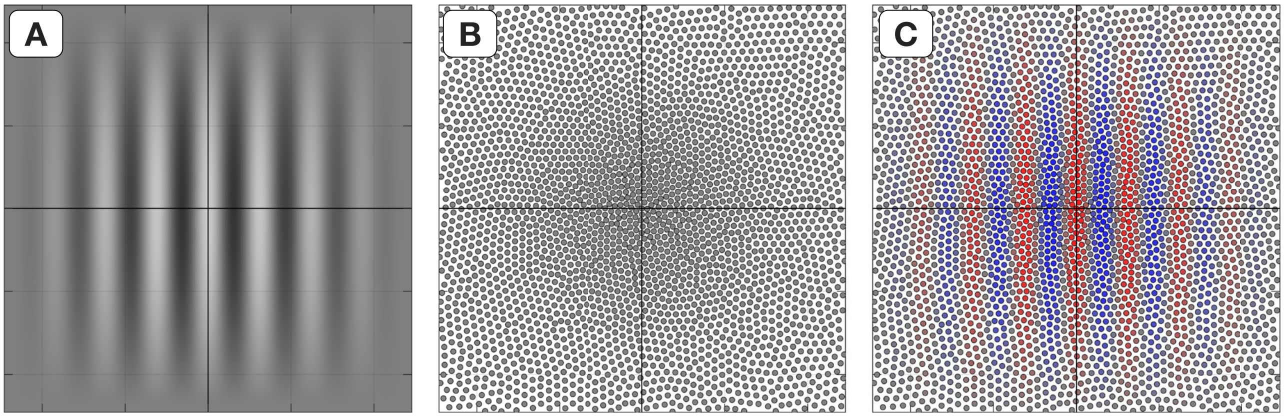

Figure 13. Stimulus-matched spatial-pooling template for the SVM-Template-Linear inference engine. (A) The spatial contrast modulation for the 16-c/° stimulus. (B) The cone mosaic used. (C) The spatial-pooling kernel, or template, for this stimulus and this mosaic, with which cone responses are weighted before pooled. Each disk corresponds to a cone; the color of the disk indicates the weight value—red for positive, blue for negative—and the color saturation indicates the weight strength.

the location corresponding to the spatial position of that cone (Figure 13). Note that use of a stimulus-matched template of this sort would be optimal if mean cone responses were perturbed only by independent identically distributed Gaussian noise. For the Poisson-noise model considered here, there is no single spatial-pooling template that is optimal across stimulus contrasts.

Spatially pooled responses were computed as follows. Given a set of null- and test-stimulus response-instance vectors $R^i_n(k,\tau), R^i_t(k,\tau), i = 1\ldots N, k = 1\ldots M, \tau = 1\ldots T$, we computed the mean response, over the N instances and T time bins, of each cone k to the null stimulus, $\bar{R}_n(k)$. This mean response to the null stimulus was subtracted from both the test and the null response-instance vectors, and the inner product between the mean-subtracted cone excitations and the template was taken to simulate spatial pooling using the $V(k)$ template:

$$R^i_{pool,n}(\tau) = \sum_{k=1}^{M} \left(R^i_n(k,\tau) - \bar{R}_n(k)\right) \times V(k), \tau = 1\ldots T$$

$$R^i_{pool,t}(\tau) = \sum_{k=1}^{M} \left(R^i_t(k,\tau) - \bar{R}_n(k)\right) \times V(k), \tau = 1\ldots T.$$

The spatially-pooled responses $R^i_{pool,n}(\tau)$ and $R^i_{pool,t}(\tau)$ were used to train the SVM linear classifier. In the present article, in which we concentrate on cone excitation responses, $R^i(k,\tau)$ is the quantity $CE^i(k,\tau)$ (Equation 4).

SVM-Template-Energy classifier

This classifier also used spatial pooling but did so via a pair of weighting kernels $V(k)$ and $V_Q(k)$. The $V_Q(k)$ kernel was derived from the contrast modulation of the spatial-quadrature version of the test stimulus. The responses of these spatial-pooling mechanisms were squared and summed, to yield an energy response

$$R^i_{E,t}(\tau) = \left(R^i_{pool,t}(\tau)\right)^2 + \left(R^i_{poolQ,t}(\tau)\right)^2, \tau = 1\ldots T.$$

The spatially pooled energy responses $R^i_{E,n}(\tau)$ and $R^i_{E,t}(\tau)$ were used to train the SVM linear classifier.

ISETBio sample code

Code for all computations in this article is available at https://github.com/isetbio/isetbio and https://github.com/isetbio/IBIOColorDetect. An introductory script is displayed in Figure 14, and a more extensive version can be found at https://github.com/isetbio/IBIOColorDetect/tree/master/tutorials/recipes/CSFpaper1.

Keywords: computational model, computational observer, contrast-sensitivity function, optics, cone mosaic, machine learning, simulation, visual performance, retinal processing

Acknowledgments

Supported by the Simons Foundation Collaboration on the Global Brain Grant 324759 and Facebook Reality Labs. We thank Jennifer Maxwell for help with software validation and documentation, and for feedback on the manuscript.

Commercial relationships: none.
Corresponding author: Nicolas P. Cottaris.

```
1  %% Generate a display for presenting stimuli and place it at a viewing distance of 57 cm
2  presentationDisplay = displayCreate('LCD-Apple', 'viewing distance', 0.57);
3
4  %% Specify an achromatic Gabor stimulus
5  stimParams = struct(...
6      'spatialFrequencyCyclesPerDeg', 10, ... % 10 cycles/deg
7      'orientationDegs', 45, ...             % 45 degree grating
8      'widthDegs', 0.4, ...                  % 0.4 x 0.4 deg patch size
9      'contrast', 0.8,...                    % 80% Michelson contrast
10     'meanLuminanceCdPerM2', 40);           % 40 cd/m2 mean luminance
11
12 %% Generate an ISETBio scene describing this stimulus
13 scene = generateStimulusScene(stimParams, presentationDisplay);
14
15 %% Realize this scene on the chosen display
16 realizedScene = realizeSceneInDisplay(scene, presentationDisplay);
17
18 %% Generate wavefront-aberration derived human optics
19 opticalImage = oiCreate('wvf human');
20
21 %% Compute the retinal image of the stimulus
22 opticalImage = oiCompute(opticalImage, realizedScene);
23
24 %% Generate an eccentricity-based hexagonal cone mosaic
25 sampleFactor = 7;
26 coneMosaic = coneMosaicHex(sampleFactor, ...  % hexagonal-to-rectangular coords sampling factor
27    'fovDegs', stimParams.widthDegs, ...       % match mosaic width to stimulus size
28    'eccBasedConeDensity', true, ...           % cone density varies with eccentricity
29    'eccBasedConeQuantalEfficiency', true, ... % cone quantal efficiency varies with eccentricity
30    'integrationTime', 10/1000, ...            % 10 msec integration time
31    'maxGridAdjustmentIterations', 50);        % terminate lattice adjustment after 50 iterations
32
33 %% Compute cone excitations for nTrials stimulus presentations
34 nTrials = 3; nTimeBins = 1;                    % 3 response instances, each with one 10 ms time bin
35 eyeMovementPath = zeros(nTrials, nTimeBins, 2); % zero eye movements
36 coneExcitations = coneMosaic.compute(opticalImage, 'emPath', eyeMovementPath);
```

Figure 14. ISETBio script for generating cone excitation responses to a scene rendered on a particular display. Line 2: Generate a presentation display. Lines 5–10: Specify parameters for an achromatic Gabor stimulus. Line 13: Generate an ISETBio scene describing the stimulus. Line 16: Generate an ISETBio scene describing the stimulus as realized on the chosen display. Line 19: Generate wavefront-aberration-derived human optics. Line 22: Generate the retinal image of the Gabor stimulus. Lines 25–31: Generate a hexagonal, eccentricity-based cone mosaic. Lines 35–36: Compute three cone excitation response instances with zero eye movements.

Email: cottaris@psych.upenn.edu.
Address: Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA.

References

Ahumada, A. J., Jr. (2002). Classification image weights and internal noise level estimation. *Journal of Vision*, *2*(1):8, 121–131, https://doi.org/10.1167/2.1.8. [PubMed] [Article]

Andrews, B. W., & Pollen, D. A. (1979). Relationship between spatial-frequency selectivity and receptive field profile of simple cells. *The Journal of Physiology*, *287*, 163–176. [PubMed]

Angueyra, J., & Rieke, F. (2013). Origin and effect of phototransduction noise in primate cone photoreceptors. *Nature Neuroscience*, *16*(11), 1692–1700. [PubMed] [Article]

Artal, P. (2015). Image formation in the living human eye. *Annual Review of Vision Science*, *1*(1), 1–17, https://doi.org/10.1146/annurev-vision-082114-035905. [PubMed]

Banks, M., Geisler, W., & Bennett, P. (1987). The physical limits of grating visibility. *Vision Research*, *27*(11), 1915–1924. [PubMed] [Article]

Banks, M., Sekuler, A., & Anderson, S. (1991). Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling. *Journal of the Optical Society of America A*, *8*, 1775–1787, https://doi.org/10.1364/JOSAA.8.001775. [PubMed]

Barlow, H. (1964). The physical limits of visual discrimination. In A. C. Giese (Ed.), *Photophysiology, Vol. 2* (pp. 163–202). New York: Academic Press.

Baylor, D., Nunn, B., & Schnapf, J. (1984). The photocurrent, noise and spectral sensitivity of rods of the monkey *Macaca fascicularis*. *The Journal of Physiology*, *357*, 575–607. [PubMed] [Article]

Bedford, R. E., & Wyszecki, G. (1957). Axial chromatic aberration of the human eye. *Journal of the Optical Society of America A*, *47*(6), 564–565. [PubMed] [Article]

Beyeler, M., Boynton, G. M., Fine, I., & Rokem, A. (2018). pulse2percept: A Python-based simulation framework for bionic vision [Preprint]. bioRxiv 148015, https://doi.org/10.1101/148015.

Bowmaker, J., Dartnall, H., & Mollon, J. (1980). Microspectrophotometric demonstration of four classes of photoreceptor in an old world primate, *Macaca fascicularis*. *The Journal of Physiology*, *298*, 131–143. [PubMed] [Article]

Bradley, C., Abrams, J., & Geisler, W. (2014). Retina-v1 model of detectability across the visual field. *Journal of Vision*, *14*(12):22, 1–22, https://doi.org/10.1167/14.12.22. [PubMed] [Article]

Brindley, G. (1960). *Physiology of the retina and the visual pathway*. London, UK: Arnold.

Campbell, F. W., & Gubisch, R. W. (1966). Optical quality of the human eye. *The Journal of Physiology*, *186*(3), 558–578. [PubMed] [Article]

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *The Journal of Physiology*, *197*(3), 551–566. [PubMed] [Article]

Cooper, R. F., Wilk, M. A., Tarima, S., & Carroll, J. (2016). Evaluating descriptive metrics of the human cone mosaic. *Investigative Ophthalmology & Visual Science*, *57*(7), 2992–3001. [PubMed] [Article]

Cottaris, N. P. (2003). Artifacts in spatiochromatic stimuli due to variations in preretinal absorption and axial chromatic aberration: Implications for color physiology. *Journal of the Optical Society of America A*, *20*(9), 1694–1713. [PubMed] [Article]

Cottaris, N. P., & Elfar, S. D. (2005). How the retinal network reacts to epiretinal stimulation to form the prosthetic visual input to the cortex. *Journal of Neural Engineering*, *2*(1), S64–S90. [PubMed] [Article]

Cottaris, N. P., Rieke, F. W., Wandell, B. A., & Brainard, D. (2018). Computational observer modeling of the limits of human pattern resolution. In *OSA Fall Vision Meeting*, October, Reno, NV.

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *The Journal of Comparative Neurology*, *292*(4), 497–523, https://doi.org/10.1002/cne.902920402. [PubMed]

De Vries, H. (1943). The quantum character of light and its bearing upon the threshold of vision, the differential sensitivity and visual acuity of the eye. *Physica*, *10*(7), 553–564. [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*(2), 203–218. [PubMed] [Article]

Engbert, R., & Kliegl, R. (2004). Microsaccades keep the eyes' balance during fixation. *Psychological Science*, *15*, 431–436, https://doi.org/10.1111/j.0956-7976.2004.00697.x. [PubMed]

Enoch, J. M. (1961). Nature of the transmission of energy in the retinal receptors. *Journal of the Optical Society of America A*, *51*(10), 1122–1126. [PubMed] [Article]

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*(1), 352–356. [Article]

Geisler, W. S. (1984). Physical limits of acuity and hyperacuity. *Journal of the Optical Society of America A*, *1*(7), 775–782. [PubMed] [Article]

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*(2), 267–314. [PubMed]

Geisler, W. S., & Banks, M. S. (1995). Visual performance. In M. Bass (Ed.), *Handbook of optics: Volume 1. Fundamentals, techniques, and design* (pp. 25.1–25.55). New York: McGraw Hill.

Golden, J. R., Erickson-Davis, C., Cottaris, N., Parthasarathy, N., Rieke, F., Brainard, D. H., … Chichilnisky, E. J. (2019). Simulation of visual perception and learning with a retinal prosthesis.

Journal of Neural Engineering, *16*(2):025003, https://doi.org/10.1101/206409.

Goodman, J. W. (2005). *Introduction to Fourier optics* (3rd ed.). Greenwood Village, CO: Roberts & Co.

Harmening, W. M., Tiruveedhula, P., Roorda, A., & Sincich, L. C. (2012). Measurement and correction of transverse chromatic offsets for multi-wavelength retinal microscopy in the living eye. *Biomedical Optics Express*, *3*(9), 2066–2077. [PubMed] [Article]

Holst, G. C. (1989). *CCD arrays, cameras and displays* (2nd ed.). Bellingham, WA: SPIE Optical Engineering Press.

Howarth, P. A., & Bradley, A. (1986). The longitudinal chromatic aberration of the human eye, and its correction. *Vision Research*, *26*(2), 361–366. [PubMed] [Article]

Howell, E., & Hess, R. (1978). The functional area for summation to threshold for sinusoidal gratings. *Vision Research*, *18*(4), 369–374. [PubMed] [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. *Electronic Imaging, Human Vision and Electronic Imaging* (pp. 177–181). http://biorxiv.org/content/early/2017/02/17/109405

Jiang, H., Wandell, B. A., & Farrell, J. E. (2015). D-CIELAB: A color metric for dichromatic observers. *SID Symposium Digest of Technical Papers*, *46*(1), 231–233. [Article]

Jonnal, R. S., Iwona, G., Migacz, J. V., Mehdi, A., Zawadzki, R. J., & Werner, J. S. (2017). The properties of outer retinal band three investigated with adaptive-optics optical coherence tomography. *Investigative Ophthalmology & Visual Science*, *58*(11), 4559–4568. [PubMed] [Article]

Judd, D., & Wyszecki, G. (1975). *Color in business, science, and industry*. New York: John Wiley and Sons.

Khaligh-Razavi, S., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, *10*(11), 1–28, https://doi.org/10.1371/journal.pcbi.1003915. [PubMed]

Kingdom, F., & Prins, N. (2010). *Psychophysics: A practical introduction*. San Diego, CA: Academic Press.

Klein, S. A., & Levi, D. M. (1985). Hyperacuity thresholds of 1 sec: Theoretical predictions and empirical validation. *Journal of the Optical Society of America A*, *2*(7), 1170–1190. [PubMed] [Article]

Kriegeskorte, N. (2015). Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science*, *1*(1), 417–446, https://doi.org/10.1146/annurev-vision-082114-035447. [PubMed]

Li, P. H., Field, G. D., Greschner, M., Ahn, D., Gunning, D. E., Mathieson, K., . . . Chichilnisky, E. J. (2014). Retinal representation of the elementary visual signal. *Neuron*, *81*(1), 130–139. [PubMed] [Article]

Lian, T., Farrell, J., & Wandell, B. A. (2018). Image systems simulation for 360 camera rigs. In *IS&T Electronic Imaging Conference*, San Francisco, CA.

Liang, J., & Williams, D. (1997). Aberrations and retinal image quality of the normal human eye. *Journal of the Optical Society of America A*, *14*, 2873–2883. [PubMed] [Article]

Lopez, H., Loew, M., Murray, H., & Goodenough, D. (1992). Objective analysis of ultrasound images by use of a computational observer. *IEEE Transactions on Medical Imaging*, *11*, 496–506. [PubMed] [Article]

Manning, C. D., Raghavean, P., & Schutze, H. (2008). *Introduction to information retrieval*. Cambridge, UK: Cambridge University Press.

Marcos, S., Burns, S. A., Moreno-Barriusop, E., & Navarro, R. (1999). A new approach to the study of ocular chromatic aberrations. *Vision Research*, *39*(26), 4309–4323. [PubMed] [Article]

Marimont, D. H., & Wandell, B. A. (1994). Matching color images: The effects of axial chromatic aberration. *Journal of the Optical Society of America A*, *11*(12), 3113–3122. [Article]

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, *5*(229), 229–240, https://doi.org/10.1038/nrn1348. [PubMed]

Meister, M., & Berry, M. J. (1999). The neural code of the retina. *Neuron*, *22*(3), 435–450. [PubMed] [Article]

Miller, W. H., & Bernard, G. D. (1983). Averaging over the foveal receptor aperture curtails aliasing. *Vision Research*, *23*(12), 1365–1369. [PubMed] [Article]

Movshon, J., Thompson, I., & Tolhurst, D. (1978). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *The Journal of Physiology*, 53–77. [PubMed] [Article]

Murray, R. F. (2011). Classification images: A review. *Journal of Vision*, *11*(5):2, 1–25, https://doi.org/10.1167/11.5.2. [PubMed] [Article]

Navarro, R., Artal, P., & Williams, D. R. (1993). Modulation transfer of the human eye as a function of retinal eccentricity. *Journal of the Optical Society of America A*, *10*(2), 201–212. [PubMed] [Article]

Ohzawa, I., DeAngelis, G., & Freeman, R. (1990, August 31). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science*, *249*(4972), 1037–1041. [PubMed] [Article]

Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 3–24). Cambridge, UK: Cambridge University Press.

Persson, P. (2005). Mesh generation for implicit geometries (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Pharr, M., & Humphreys, G. (2010). *Physically based rendering: From theory to implementation* (2nd ed.). San Francisco: Morgan Kaufmann.

Polans, J., Jaeken, B., McNavv, R., Artal, P., & Izatt, J. (2015). Wide-field optical model of the human eye with asymmetrically tilted and decentered lens that reproduces measured ocular aberrations. *Optica*, *2*(2), 124–134, https://doi.org/10.1364/OPTICA.2.000124.

Pugh, E.N., Jr., & Lamb, T. (2000). Phototransduction in vertebrate rods and cones: Molecular mechanisms of amplification, recovery and light adaptation. In D. Stavenga, W. de Grip, & E. Pugh (Eds.), *Handbook of biological physics, Vol. 3: Molecular mechanisms of visual transduction* (pp. 183–255). Amsterdam, the Netherlands: Elsevier.

Putnam, C., & Bland, P. J. (2014). Macular pigment optical density spatial distribution measured in a subject with oculocutaneous albinism. *Journal of Optometry*, *7*(4), 241–245. [PubMed] [Article]

Renner, A. B., Knau, H., Neitz, M., Neitz, J., & Werner, J. S. (2004). Photopigment optical density of the human foveola and a paradoxical senescent increase outside the fovea. *Visual Neuroscience*, *21*(6), 827–834. [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]

Rodieck, R. (1998). *The first steps in seeing*. Sunderland, MA: Sinauer.

Rose, A. (1948). The sensitivity performance of the human eye on an absolute scale. *Journal of the Optical Society of America A*, *38*(2), 196–208. [PubMed] [Article]

Rushton, W., & Henry, G. (1968). Bleaching and regeneration of cone pigments in man. *Vision Research*, *8*(6), 617–631. [PubMed] [Article]

Scholkopf, B., & Smola, A. (2002). *Learning with kernels*. Cambridge, MA: MIT Press.

Shapley, R., Kaplan, E., & Soodak, R. (1981, August 6). Spatial summation and contrast sensitivity of X and Y cells in the lateral geniculate nucleus of the macaque. *Nature*, *292*(5823), 543–545, http://doi.org/10.1038/292543a0. [PubMed]

Stiles, W., & Crawford, B. (1933). The luminous efficiency of rays entering the eye pupil at different points. *Proceedings of the Royal Society of London B: Biological Sciences*, *112*(778), 428–450. [Article]

Stockman, A., & Brainard, D. (2010). Color vision mechanisms. In M. Bass, C. DeCusatis, & J. Enoch (Eds.), *The Optical Society of America handbook of optics, Volume: 3: Vision and vision optics* (pp. 1.11–11.104). New York: McGraw Hill.

Stockman, A., & Sharpe, L. T. (2000). The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research*, *40*, 1711–1737. [PubMed] [Article]

Stockman, A., Sharpe, L. T., & Fach, C. C. (1999). The spectral sensitivity of the human short-wavelength sensitive cones derived from thresholds and color matches. *Vision Research*, *39*, 2901–2927. [PubMed] [Article]

Tanner, W. J., & Swets, J. (1954). The human use of information: I. Signal detection for the case of a signal known exactly. *Transactions of the IRE Profession Group in Information Theory*, *4*(4), 213–221. [Article]

Thibos, L. N., Hong, X., Bradley, A., & Cheng, X. (2002). Statistical variation of aberration structure and image quality in a normal population of healthy eyes. *Journal of the Optical Society of America A*, *19*(12), 2329–2348. [PubMed] [Article]

Thibos, L. N., Ye, M., Zhang, X., & Bradley, A. (1992). The chromatic eye: A new reduced-eye model of ocular chromatic aberration in humans. *Applied Optics*, *31*(19), 3594–3600. [PubMed] [Article]

Tuten, W. S., Cooper, R. F., Tiruveedhula, P., Dubra, A., Roorda, A., Cottaris, N. P., … Morgan, J. I. W. (2018). Spatial summation in the human fovea: Do normal optical aberrations and fixational eye movements have an effect? *Journal of Vision*, *18*(8): 6, 1–18, https://doi.org/10.1167/18.8.6. [PubMed] [Article]

Vos, J. J. (2003). On the cause of disability glare and its dependence on glare angle, age and ocular pigmentation. *Clinical and Experimental Optometry*, *86*(6), 363–370, https://doi.org/10.1111/j.1444-0938.2003.tb03080.x. [PubMed]

Wandell, B. A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer.

Watson, A. (2015). Computing human optical point spread functions. *Journal of Vision*, *15*(2):26, 1–25, https://doi.org/10.1167/15.2.26. [PubMed] [Article]

Watson, A., & Ahumada, A. (2004). The spatial standard observer. *Journal of Vision*, *4*(8): 51, https://doi.org/10.1167/4.8.51. [Abstract]

Watson, A., & Ahumada, A. (2005). A standard model for foveal detection of spatial contrast. *Journal of Vision*, *5*(9):6, 717–740, https://doi.org/10.1167/5.9.6. [PubMed] [Article]

Watson, A., & Yellott. (2012). A unified formula for light-adapted pupil size. *Journal of Vision*, *12*(10): 12, 1–16, https://doi.org/10.1167/12.10.12. [PubMed] [Article]

Westheimer, G. (1981). Visual hyperacuity. In H. Autrum, E. R. Perl, R. F. Schmidt, & D. Ottoson (Eds.), *Progress in sensory physiology* (pp. 1–30). Berlin, Germany: Springer.

Westheimer, G. (2008). Directional sensitivity of the retina: 75 years of Stiles–Crawford effect. *Proceedings of the Royal Society of London B: Biological Sciences*, *275*(1653), 2777–2786. [PubMed] [Article]

Westheimer, G., & Campbell, F. W. (1962). Light distribution in the image formed by the living human eye. *Journal of the Optical Society of America*, *52*(9), 1040–1045. [PubMed] [Article]

Williams, D. (1985). Visibility of interference fringes near the resolution limit. *Journal of the Optical Society of America A*, *2*(7), 1087–1093. [PubMed] [Article]

Williams, D., Brainard, D. H., McMahon, M. J., & Navarro, R. (1994). Double-pass and interferometric measures of the optical quality of the eye. *Journal of the Optical Society of America A*, *11*(12), 3123–3135. [PubMed] [Article]

Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulas*. New Work: John Wiley & Sons.

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences, USA*, *111*(23), 8619–8624. [PubMed] [Article]