

# Retina-V1 model of detectability across the visual field

Chris Bradley

University of Texas at Austin, Austin, TX, USA



Jared Abrams

University of Texas at Austin, Austin, TX, USA



Wilson S. Geisler

University of Texas at Austin, Austin, TX, USA



**A practical model is proposed for predicting the detectability of targets at arbitrary locations in the visual field, in arbitrary gray scale backgrounds, and under photopic viewing conditions. The major factors incorporated into the model include (a) the optical point spread function of the eye, (b) local luminance gain control (Weber's law), (c) the sampling array of retinal ganglion cells, (d) orientation and spatial frequency-dependent contrast masking, (e) broadband contrast masking, and (f) efficient response pooling. The model is tested against previously reported threshold measurements on uniform backgrounds (the ModelFest data set and data from Foley, Varadharajan, Koh, & Farias, 2007) and against new measurements reported here for several ModelFest targets presented on uniform, 1/f noise, and natural backgrounds at retinal eccentricities ranging from 0° to 10°. Although the model has few free parameters, it is able to account quite well for all the threshold measurements.**

## Introduction

Detection and discrimination of spatial patterns is fundamental to almost every task involving the human visual system. Despite decades of research, there have been few attempts to develop models that can predict the detectability of an arbitrary target appearing in arbitrary backgrounds at any position in the visual field. The goal here is to present such a model that is both practical and testable. We focus on the classic task in which both the target and the location at which the target will appear are known to the observer. This task is perhaps the most studied task in vision science, and hence, there is a vast amount of knowledge relevant for developing models. Indeed, there are many specific models proposed to account for detection and discrimination over narrow ranges of conditions (e.g., Arnou & Geisler, 1996; Foley, 1994; Foley et al., 2007;

Goris, Putzes, Wagemans, & Wichmann, 2013; Morone & Burr, 1988; Watson & Ahumada, 2005; Watson & Solomon, 1997; Watt & Morgan, 1985; Wilson & Bergen, 1979). Our aim is not to incorporate all existing knowledge into a grand model nor to compete with existing models designed for a narrow range of conditions but rather to combine what appear to be the most important factors identified in the spatial vision literature into a streamlined model for which it is practical to generate predictions rapidly for arbitrary backgrounds and targets at arbitrary retinal locations.

There are several reasons for attempting to develop such a model. The first is to determine how well known spatial vision mechanisms can account for the detectability of targets in natural backgrounds. This is important because a central goal of vision science is to predict the visual performance of organisms in their natural environment. Given the relative simplicity of detection and discrimination tasks, they may be the best place to make progress toward this goal. The second reason is to provide a foundation for rigorous analysis and modeling of more complex tasks, such as visual search. The third reason is that such a model may be of value in human factors and medical assessment applications.

The proposed model is based largely on known anatomical and physiological factors, and hence, there are relatively few free parameters. To estimate some of these parameters and test the model for foveal detection on uniform backgrounds, we fitted the ModelFest data set, which consists of detection thresholds measured in 16 observers for 43 different targets (see Watson & Ahumada, 2005). To estimate the remaining parameters and test the model in the more general case, we measured and then fitted detection thresholds in three observers for 1/f noise backgrounds (which have the power spectrum of natural images; Burton & Moorehead, 1987; D. J. Field, 1987) and natural image backgrounds for three

Citation: Bradley, C., Abrams, J., & Geisler, W. S. (2014). Retina-V1 model of detectability across the visual field. *Journal of Vision*, 14(12):22, 1–22, <http://www.journalofvision.org/content/14/12/22>, doi:10.1167/14.12.22.

ModelFest targets at several eccentricities. In what follows, we first describe the model, then the psychophysical measurements, and finally the predictions for the psychophysical measurements.

## The retina-V1 (RV1) model

There are orders of magnitude more photoreceptors and primary visual cortex neurons than there are retinal ganglion cells. This fact alone suggests that the optic nerve (the population of retinal ganglion cells) may be the main bottleneck for spatial pattern detection information in the human visual system. In other words, there are likely to be sufficient neural resources in early cortical areas to encode the ganglion cell responses from the target and the background. Also, the target and background are most compactly represented in the ganglion cell responses (fewest numbers of neurons), and much is known about the anatomy and physiology of the retina. These observations motivated us to anchor a model of target detectability on a model of the retinal ganglion cell responses. In the model, cortical mechanisms also play important roles. One is to limit the spatial frequency and orientation content in the ganglion cell responses that mask detectability of the target. The other is to pool the neural responses in order to make perceptual decisions.

The RV1 model has two major components: a “retinal” component and a “cortical” component (Figure 1). The retinal component is grounded in the known anatomy and physiology of the eye, and the cortical component is grounded in known properties of neurons in the primary visual cortex as well as empirical relationships from the psychophysical literature.

The retinal component simulates the responses of the mid-ganglion cells (P cells) in the human/primate retina. It includes the average optical point spread function ( $psf$ ) of the human eye (Navarro, Artal, & Williams, 1993); local luminance gain control ( $G_L$ ), which enforces Weber’s law for detection on uniform backgrounds (for reviews, see Hood, 1998; Hood & Finkelstein, 1986); the average spatial sampling density of mid-ganglion cells in the human retina (Curcio & Allen, 1990; Dacey, 1993; Drasdo, Millican, Katholi, & Curcio, 2007); and the receptive field properties of mid-ganglion cells in the nonhuman primate retina (Croner & Kaplan, 1995; Derrington & Lennie 1984). We focus on the mid-ganglion cell pathway because of evidence that it is responsible for detection performance under conditions of low to moderate temporal frequency (Merigan, Katz, & Maunsell, 1991; Merigan & Maunsell, 1993). These

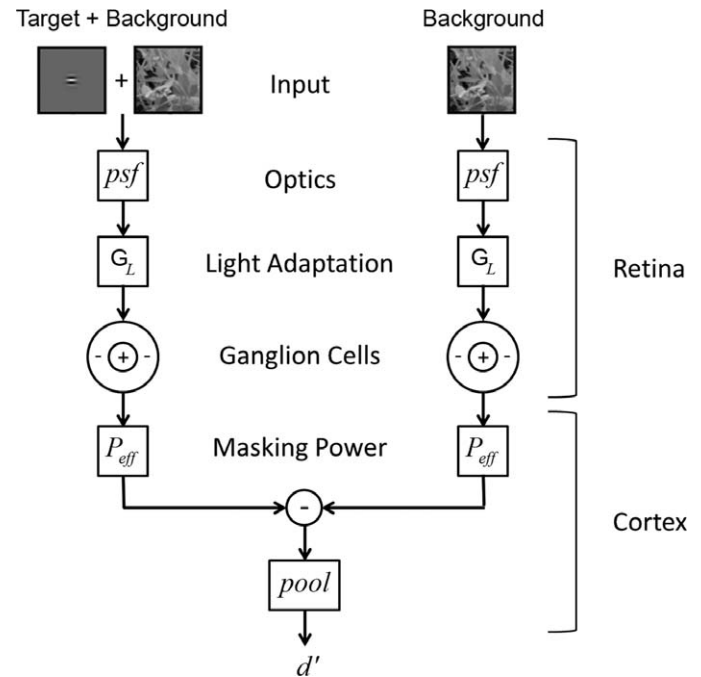


Figure 1. Schematic of The RV1 model of detection.

conditions include the case of interest here: static stimuli presented for the duration of a typical eye fixation (150–400 ms).

The cortical component simulates the spatial pattern masking effect of the background as well as the final pooling of responses that determines the predicted detectability of the target ( $d'$ ). The spatial pattern masking is represented by an effective total contrast power ( $P_{eff}$ ) that is the sum of three components: a baseline component, a narrowband component, and a broadband component.

The narrowband component is computed assuming filtering matched to the average spatial frequency and orientation bandwidth of neurons in the monkey primary visual cortex (for reviews, see De Valois & De Valois, 1988; Geisler & Albrecht, 1997; Palmer, Jones, & Stepnoski, 1991; Shapley & Lennie, 1985), which are generally consistent with estimates from the psychophysical literature (for reviews, see De Valois & De Valois, 1988; Graham, 1989; 2011). The broadband component is consistent with the contrast normalization effects observed in cortical neurons (Albrecht & Geisler, 1991; Carandini & Heeger, 2012; Carandini, Heeger, & Movshon, 1997; Geisler & Albrecht, 1997; Heeger, 1991, 1992; Sit, Chen, Geisler, Miikkulainen, & Seidemann, 2009) and evidenced in the psychophysical literature (Foley, 1994; Goris et al., 2013; Watson & Solomon, 1997). We assume that the effective total contrast power acts as an equivalent noise power in the computation of  $d'$  (Burgess & Colborne, 1988; Eckstein, Ahumada, & Watson, 1997a; Lu & Doshier, 1999, 2008). This enforces the psychophysical rule that

threshold contrast power increases linearly with background contrast power for white noise backgrounds (Burgess, Wagner, Jennings, & Barlow, 1981; Legge, Kersten, & Burgess, 1987) and for 1/f noise backgrounds (Najemnik & Geisler, 2005). This concludes a brief summary of the model; we now provide more details.

## Retinal image

The input image is either the background image alone  $I_B(\mathbf{x})$  or the sum of the target and background images  $I_T(\mathbf{x}) + I_B(\mathbf{x})$ , where we have simplified the notation by letting  $\mathbf{x} = (x, y)$ . Until the final steps of the model, the operations are effectively linear, and hence, the target and background can be processed separately. The retinal images of the target and background are computed by convolving the target and background images with an appropriate optical psf:

$$T(\mathbf{x}) = I_T(\mathbf{x}) * \text{psf}(\mathbf{x}) \quad (1)$$

$$B(\mathbf{x}) = I_B(\mathbf{x}) * \text{psf}(\mathbf{x}) \quad (2)$$

In the current implementation, we use the average human psf in the fovea reported in Navarro et al. (1993). The convolution is computed in the Fourier domain using their reported modulation transfer function:  $MTF(f) = 0.78 \exp(-0.172 f) + 0.22 \exp(-0.037 f)$ . The optical psf broadens (blur increases) with retinal eccentricity but is relatively constant out to  $10^\circ$  eccentricity, the largest eccentricity measured in the present study. This component of the model could be easily adjusted for greater eccentricities or to take into account individual differences in optics.

## Ganglion cell sampling and the magnification principle

There is strong evidence that each different type of retinal ganglion cell forms a mosaic such that the dendritic branches and the receptive fields of the cells in the mosaic tile the retinal image with no gaps. Furthermore, for each cell type, the percentage overlap of the receptive fields is approximately constant and independent of retinal eccentricity (for a recent review, see G. D. Field & Chichilnisky, 2007). This result suggests a tight link between the anatomical spacing of retinal ganglion cells and the size of their receptive fields. We exploit this fact to reduce the number of parameters in the model.

We use the average ganglion cell density reported by Drasdo et al. (2007) (six human eyes) to generate a mosaic of midget ganglion cells. The results reported by Drasdo et al. (which are based on a reanalysis of the

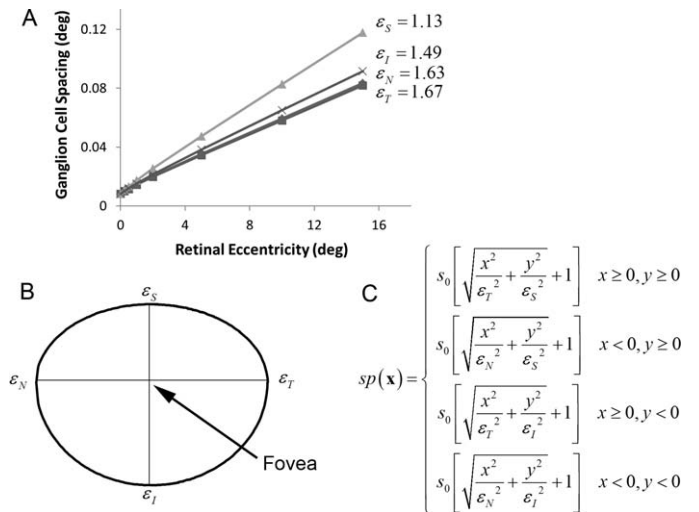


Figure 2. Midget ganglion cell spacing (in degrees) in the human retina. (A) Ganglion cell spacing (1/square-root of density) in the four cardinal directions of the visual field, assuming one midget ganglion cell for each cone in the center of the fovea (which sets the  $y$ -intercept). This one “midget ganglion cell,” which can respond positively and negatively, represents an on-and-off pair of ganglion cells (data from Drasdo et al., 2007.) (B) To generate a ganglion cell mosaic, we assume that, in each quadrant, the contours of constant spacing fall on an ellipse. This specific contour shows the retinal locations at which the spacing between midget ganglion cells is twice what it is in the center of the fovea. Thus, in the upper vertical direction, the spacing doubles at about  $1.1^\circ$  of eccentricity, but in the horizontal directions, it does not double until about  $1.6^\circ$ . (C) The equation that defines the spacing function:  $s_0$  is the spacing in the center of the fovea, and  $\epsilon_N$ ,  $\epsilon_T$ ,  $\epsilon_I$ , and  $\epsilon_S$  are the eccentricities in the four cardinal directions where spacing reaches twice  $s_0$ .

data in Curcio & Allen, 1990) describe the combined falloff for all types of ganglion cells. However, Dacey (1993) reports that the falloff in midget ganglion cell density in humans tracks that reported by Curcio and Allen over the first  $15^\circ$  eccentricity, the range of interest here. Thus, we assume that human midget ganglion cell density from  $0^\circ$  to  $15^\circ$  is proportional to the ganglion cell density reported by Drasdo et al.

The symbols in Figure 2A plot one over the square root of the density (the linear spacing) of ganglion cells as a function of eccentricity in the four cardinal directions (nasal, temporal, inferior, superior), assuming that there is one midget ganglion cell (with a linear receptive field) for each cone in the center of the fovea. The spacing of cones in the center of the fovea,  $s_0$ , is approximately 30 arcsec ( $0.0083^\circ$ ), and thus, the assumed density of midget ganglion cells in the center of the fovea is  $120 \text{ cells}/^\circ$ . In reality, there is one on and one off midget ganglion cell for each cone, and hence, the actual density is approximately  $240 \text{ cells}/^\circ$ . How-

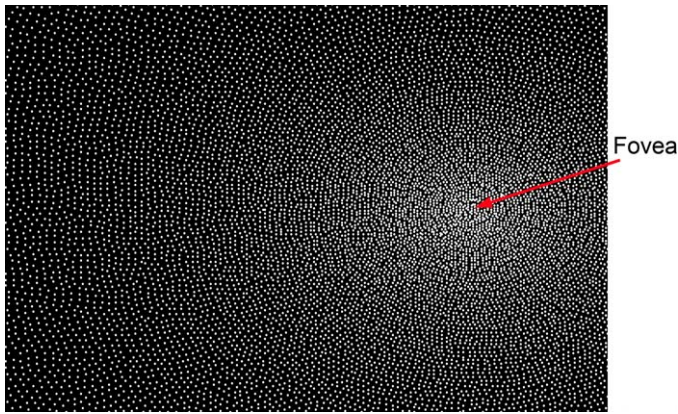


Figure 3. Part of the midget ganglion cell mosaic generated from the human anatomical data in Figure 2. Each dot represents the location of the center of a ganglion cell receptive field. This mosaic is generated from the equation in Figure 2C, using the algorithm given in the Appendix.

ever, with little loss of precision, we represent the pair of on and off cells by a single linear receptive field that produces positive and negative responses. (In the current model, we ignore differences in the density and receptive field sizes of on and off ganglion cells [Dacey & Peterson, 1992].) As can be seen, midget ganglion cell spacing increases approximately linearly with a slope that depends on direction in the visual field.

We use these data to generate a ganglion cell mosaic. In particular, we assume that the contours of constant spacing in each quadrant of the visual field fall on an ellipse (Figure 2B). Thus, the spacing function is given by the equation in Figure 2C, where  $\varepsilon_N$ ,  $\varepsilon_T$ ,  $\varepsilon_L$ , and  $\varepsilon_S$  are the eccentricities in the four cardinal directions at which the spacing between ganglion cells reaches twice what it is in the center of the fovea. This spacing is then used to generate the ganglion cell mosaic, a portion of which is shown in Figure 3. The specific algorithm used to generate the mosaic is given in the Appendix. The algorithm produces a mosaic that satisfies the spacing function and does not have any observable artifacts. We represent the mosaic by the function  $samp(\mathbf{x})$ .

Once the ganglion cell mosaic is specified, we then enforce the magnification principle by assuming that the receptive field properties (center and surround size) of the simulated ganglion cells scale with the spacing between ganglion cells in the mosaic. Thus, for each property, there is only a single free parameter, a scale factor, that applies to all eccentricities.

## Light adaptation

Retinal light adaptation mechanisms maintain pattern detection and discrimination sensitivity by keeping

the responses of neurons within their limited dynamic ranges. The primary effects of light adaptation can be summarized as a multiplicative luminance gain control (the signal is scaled by the inverse of the average luminance). An important perceptual effect of retinal light adaptation is Weber's law: contrast threshold on uniform backgrounds is approximately constant independent of background luminance. To include luminance gain control, we compute the local average luminance at each retinal location. Let  $g_a(\mathbf{y}; \mathbf{x})$  be a 2-D Gaussian (with a volume of 1.0) centered on retinal location  $\mathbf{x}$ . Then, the local average luminance at  $\mathbf{x}$  is

$$L(\mathbf{x}) = \sum_{\mathbf{y}} B(\mathbf{y})g_a(\mathbf{y}; \mathbf{x}) \quad (3)$$

It is plausible that local retinal luminance gain is set by neural populations having receptive fields that increase in size with retinal eccentricity, but less is known about these populations in primates, and hence, for simplicity, we assume the standard deviation ( $SD$ ) of the Gaussian is fixed:  $\sigma_L(\mathbf{x}) = \sigma_L$ . Thus, the effect of light adaptation is represented by a single parameter. The local luminance gain is  $G_L(\mathbf{x}) = 1/L(\mathbf{x})$ . Note that when the background is uniform then luminance gain is the same at all retinal locations (because the Gaussian has a volume of 1.0). To handle low light levels in which Weber's law fails, a constant  $L_0$  can be added to the denominator, but for the conditions of interest here, that was not necessary.

We note that there is also global light adaptation due to slower mechanisms (pupil response, photoreceptor adaptation) that adjust the retina to the overall ambient light level in the environment. However, here we focus on stimuli with which the global average luminance is fixed (displays in which the average luminance is fixed across conditions), and hence, we ignore the effects of global light adaptation.

## Ganglion cell responses

The spatial receptive fields of midget ganglion cells (and the corresponding P cells in the lateral geniculate nucleus) are often approximated by a difference of 2-D Gaussians (Croner & Kaplan, 1995; Derrington & Lennie, 1984; Rodieck, 1965). Using this approximation, let  $g_c(\mathbf{y}; \mathbf{x})$  and  $g_s(\mathbf{y}; \mathbf{x})$  be 2-D Gaussians representing the center and surround mechanisms of a midget ganglion cell at retinal location  $\mathbf{x}$  (equations are in the Appendix). The response of ganglion cells to the background alone is given by

$$r_B(\mathbf{x}) = samp(\mathbf{x}) \sum_{\mathbf{y}} G_L(\mathbf{y})B(\mathbf{y})D(\mathbf{y}; \mathbf{x}) \quad (4)$$

where  $D(\mathbf{y}; \mathbf{x})$  is a difference of Gaussians:  $D(\mathbf{y}; \mathbf{x}) = w_c g_c(\mathbf{y}; \mathbf{x}) - (1 - w_c)g_s(\mathbf{y}; \mathbf{x})$ . For the conditions of

interest here, the target contributes little to the local luminance, and hence, the response to the target plus background is simply the sum of the responses to the target and background, and the response of the ganglion cells to the target is given by

$$r_T(\mathbf{x}) = \text{samp}(\mathbf{x}) \sum_{\mathbf{y}} G_L(\mathbf{y}) T(\mathbf{y}) D(\mathbf{y}; \mathbf{x}) \quad (5)$$

We assume that the magnification principle holds, and thus, the *SD* of the center mechanism is given by  $\sigma_c(\mathbf{x}) = k_c \text{sp}(\mathbf{x})$  and the surround mechanism by  $\sigma_s(\mathbf{x}) = k_s \text{sp}(\mathbf{x})$ . We see then that three parameters,  $w_c$ ,  $k_c$ , and  $k_s$ , describe the receptive field properties of all the ganglion cells.

### Effective masking power

Masking in the RV1 model is represented by an effective contrast power  $P_{\text{eff}}$  that is the weighted sum of three components (see Figure 4): a baseline component  $P_0$ , a narrowband component  $P_{nb}$ , and a broadband component  $P_{bb}$ :

$$P_{\text{eff}} = P_0 + k_b w_b P_{nb} + k_b (1 - w_b) P_{bb} \quad (6)$$

where  $k_b$  sets the overall strength of pattern masking, and  $w_b$  sets the relative strength of the narrowband and broadband components.

The baseline component is a constant that represents the masking when the background is uniform. This component includes the effect of spontaneous ganglion cell activity, decision noise, and other factors not dependent on the spatial pattern of the background.

The narrowband component is the power in the ganglion cell response to the background that drives the population of primary visual cortex (V1) neurons responding to the target, and thus, it is target dependent. In other words, the narrowband component represents the fact that neurons in V1 are simultaneously selective to spatial frequency and orientation and thus will filter out background power in the ganglion cell responses that does not activate the population of V1 neurons activated by the target. In computing the narrowband component, we assume that the spatial frequency selectivity of V1 neurons is approximately Gaussian in log frequency (a log Gabor function) with a bandwidth that averages 1.5 octaves (De Valois, Albrecht, & Thorell, 1982; Geisler & Albrecht, 1997) and that the orientation selectivity is approximately Gaussian on a circle with a bandwidth that averages  $40^\circ$  (De Valois, Yund, & Hepler, 1982). These log Gabor and Gaussian functions are defined in the Appendix.

The first step in computing the narrowband component is to obtain the filtered ganglion cell responses, which are given by

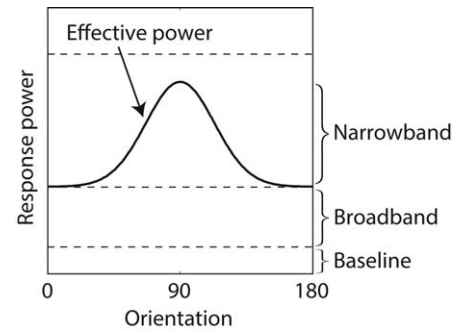


Figure 4. Effective masking power of responses to a  $1/f$  noise background for a Gabor target. Shown is a cross-section of the average masking power as a function of orientation at one spatial frequency (solid curve).

$$r_{nb}(\mathbf{x}) = \text{samp}(\mathbf{x}) \sum_{\mathbf{y}} G_{bc}(\mathbf{y}; \mathbf{x}) f_T(\mathbf{y}; \mathbf{x}) \quad (7)$$

where  $G_{bc}(\mathbf{y}; \mathbf{x})$  is the continuous (unsampled) ganglion cell center response to the background, and  $f_T(\mathbf{y}; \mathbf{x})$  is the target-specific filter that removes the background power in the ganglion cell responses that do not drive the cortical neurons that encode the target. To determine the target-specific filter, we (a) take the Fourier transform of the ganglion cell center response to the target alone, (b) convert to log polar coordinates (log frequency vs. orientation), (c) convolve (in the frequency domain) with a function that is the product of the amplitude spectrum of a log Gabor (bandwidth 1.5 octaves) and a Gaussian function in orientation (bandwidth  $40^\circ$ ), and (d) convert back into standard spatial frequency axes and take the inverse Fourier transform. We convert to log polar coordinates so that the cortical filters at all log frequencies and orientations have the same shape, allowing simple convolution in step (c) (Watson & Solomon, 1997, use a similar trick). In this version of target-dependent filtering, we did not include the effect of the ganglion cell surround because the lowest frequency (DC) is automatically removed by the log-Gabor cortical filtering. However, we have preliminary results that include both center and surround, and the quality of the Model predictions is similar.

The narrowband component is given by

$$P_{nb}(\mathbf{x}) = \sum_{\mathbf{y}} r_{nb}^2(\mathbf{y}) E_T(\mathbf{y}; \mathbf{x}) \quad (8)$$

where  $E_T(\mathbf{y}; \mathbf{x})$  is the blurred spatial envelope of the target, and the blurring depends on retinal location (i.e., envelope size increases with eccentricity). The filtered ganglion cell responses are weighted by the blurred envelope of the target under the plausible assumption that only the background power falling within some spatial neighborhood of the target will have a masking effect. The envelope is defined to be the

2-D Gaussian (with arbitrary covariance matrix) that best fits the absolute value of the target (see Appendix). The blurred envelope is obtained by convolving the envelope with a 2-D Gaussian having a  $SD$  of the ganglion cell center  $\sigma_c$  (see Appendix).

The broadband component is the power in the ganglion cell responses that contributes to masking but is not spatial frequency and orientation dependent. Such a broadband component is consistent with divisive contrast gain control (normalization) observed in cortical neurons (Albrecht & Geisler, 1991; Carandini & Heeger, 2012; Carandini et al., 1997; Geisler & Albrecht, 1997; Heeger, 1991, 1992; Sit et al., 2009) and with the psychophysical literature (Foley, 1994; Goris et al., 2013; Watson & Solomon, 1997). The broadband component is given by

$$P_{bb}(\mathbf{x}) = \sum_{\mathbf{y}} [r_B(\mathbf{y}) - r_0]^2 E_T(\mathbf{y}; \mathbf{x}) \quad (9)$$

where  $r_0$  is the response of a ganglion cell to a uniform background, which is a constant that depends only on the relative weight of center and surround:  $r_0 = 2w_c - 1$ . Subtraction of  $r_0$  guarantees that  $P_{bb}$  is zero for uniform backgrounds. Note that  $P_{nb}$  is also zero for uniform backgrounds because the spatial frequency tuning of the cortical neurons is log Gabor (which goes to zero at zero spatial frequency). Generally, the masking power of the background is greatest when the weight on the narrowband component is zero (upper dashed line in Figure 4) and least when the weight on the broadband component is zero (i.e., when the solid curve touches the baseline in Figure 4).

## Pooling

We assume that the pooled response is given by the following formula:

$$r_{pooled} = \frac{\sqrt[\rho]{\sum_{\mathbf{x}} |r_T(\mathbf{x})|^\rho}}{\sigma_{eff}} \quad (10)$$

where  $\rho$  is a pooling exponent, and  $\sigma_{eff} = \sqrt{P_{eff}}$  is the effective masking contrast. If one regards the effective masking contrast as an equivalent noise (Burgess & Colborne, 1988; Eckstein et al., 1997a; Lu & Doshier, 1999, 2008), then  $r_{pooled}$  can be regarded a signal-to-noise ratio. In this case, if the pooling exponent is 2.0, then Equation 10 is the standard formula for optimal pooling of statistically independent Gaussian signals (“ $d'$  summation”). Following others (Graham, 1977; Quick, 1974; Watson, 1979; Watson & Ahumada, 2005), we allow the pooling exponent to be greater than 2.0 (which is suboptimal) although for the current

model the estimated exponent is only slightly larger, 2.4 (see later).

## Detectability and contrast threshold

The last step is to specify the relationship between the pooled response, detection threshold, and detectability. For the purpose of predicting detection performance, we define the contrast of the target in the standard way as the target amplitude (peak gray level) divided by the mean background gray level of the whole screen. We define detection threshold  $c_t$  to be the contrast of the target at which the signal-to-noise ratio given by Equation 10 is equal to 1.0, which corresponds to 69% correct. In other words, the predicted contrast threshold is the solution to the equation

$$r_{pooled}(c_t) = 1 \quad (11)$$

Although this equation gives the threshold, another parameter  $\beta$  is required to predict the steepness of the psychometric function. Specifically, we assume that detectability has the form

$$d'(c) = r_{pooled}^\beta(c) \quad (12)$$

Note that at threshold,  $d'(c_t) = r_{pooled}(c_t) = 1$ . In the model, for a given target and background, the pooled response is linear with target contrast and hence it is easy to show that

$$d'(c) = (c/c_t)^\beta \quad (13)$$

Using the usual formula from signal detection theory (Green & Swets, 1966), the predicted psychometric function is given by

$$pcorr(c) = \Phi\left[\frac{1}{2}d'(c)\right] = \Phi\left[\frac{1}{2}(c/c_t)^\beta\right] \quad (14)$$

where  $\Phi(z)$  is the standard normal integral function (this assumes optimal criterion placement).

Note that for expository purposes we regard  $\sigma_{eff}$  as an equivalent noise. However, it could also be regarded as a deterministic gain control, which would make  $r_{pooled}^\beta(c)$  a deterministic signal. A constant late decision noise would then also give Equations 12 through 14.

In sum, the contrast thresholds predicted by the model are determined by only eight parameters. Five of these parameters,  $k_c$ ,  $k_s$ ,  $w_c$ ,  $P_0$ , and  $\rho$ , determine the predicted contrast thresholds for uniform backgrounds at all retinal locations. The additional three parameters,  $\sigma_L$ ,  $k_b$ , and  $w_b$ , determine predicted thresholds for more complex backgrounds. A ninth parameter  $\beta$  is needed for predicting values of detectability ( $d'$ ) that do not correspond to the 69% correct threshold.

## Implementation

Although the RV1 model is relatively simple conceptually, programming an efficient implementation is nontrivial, especially if one would like to rapidly compute detectability for all possible target locations and/or fixation locations for a wide range of targets and backgrounds. The primary difficulty is that all the linear weighted summations (except the optics) are shift variant (they change with location relative to the point of fixation). To make the computations efficient, we use multiresolution stacks. Specifically, we fix the target and background images at a canonical location, centered at  $\mathbf{x} = (0,0)$ , and then convolve each image separately with a series of Gaussians having *SDs* that incrementally increase in powers of two. This set of images forms a stack of successively blurred images, each corresponding to a particular discrete *SD* (resolution). We precompute and save these stacks for each target and background image to be processed. For each target image, we also precompute and save the target-specific spatial frequency filter corresponding to each level of the target stack. Once these stacks are computed and stored, they can be interpolated to rapidly determine the local luminance function, the ganglion cell target response function, and the ganglion cell effective background response function for any target location and fixation location. Specifically, each fixation location and target location specifies a spatial region of the background as well as the spatial coordinates of the samples (ganglion cells) covering that region. The location of a sample specifies a particular continuous *SD* (resolution). That resolution will fall between two neighboring resolutions in the stack. The value at the sample location is obtained by linearly interpolating between the two values in these neighboring resolution images. This procedure provides a close approximation to the exact calculations. A MatLab implementation of the model is available at <http://natural-scenes.cps.utexas.edu/>.

## Parameter estimation

To estimate model parameters, we minimized the squared error between the measured and predicted contrast thresholds expressed in log units (dB). Let  $c_i$  be the observed contrast threshold (in dB) for condition  $i$ , and let  $\hat{c}_i(\boldsymbol{\theta})$  be the predicted contrast threshold for parameters  $\boldsymbol{\theta}$ . We minimize the sum of the squared errors  $S(\boldsymbol{\theta})$ , and thus,  $\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} S(\boldsymbol{\theta})$ . When the background is fixed (e.g., a uniform background), this minimization is straightforward. However, when the background randomly varies from trial to trial (the 1/f noise and natural backgrounds), it is not practical to generate a predicted model response for each trial for

each vector of parameter values evaluated during the parameter search.

To handle the case of variable backgrounds, we use the following procedure. First, we pick a random background patch for each background condition and then obtain estimates  $\hat{\boldsymbol{\theta}}_1$ . Once these estimates are obtained, we generate the predicted threshold  $\hat{c}_{ij}(\hat{\boldsymbol{\theta}}_1)$  for each specific background patch  $j$  in each condition  $i$ . Then, for each condition, we rank order the thresholds and select the patch having the median threshold. Let this patch be  $j_i$ . We then estimate the parameters again, where the fixed patch for condition  $i$  is  $j_i$ . These estimates are  $\hat{\boldsymbol{\theta}}_2$ . We repeat this process until the estimated parameters converge (usually just a couple of iterations). Simulations show that this procedure is effective in finding the optimal parameters.

Once the optimal parameters are estimated, a predicted threshold is computed for every background patch in every condition. The predicted threshold for a particular condition is the average of the predicted thresholds for all the background patches in that condition.

## Experiment: Measurement of detectability in 1/f noise and natural backgrounds

The goal of our model is to accurately predict detection thresholds for localized targets in arbitrary natural backgrounds at arbitrary locations in the visual field. To test the accuracy of the predictions, we measured contrast detection thresholds in a single-interval forced choice (yes/no) paradigm for three target stimuli (Gabor, Gaussian, and Edge) presented at four retinal eccentricities ( $0^\circ$ ,  $2.5^\circ$ ,  $5^\circ$ , and  $10^\circ$ ) along the horizontal meridian in the right visual field in three different types of background (uniform, 1/f noise, and natural image). The 1/f noise and natural image backgrounds were presented at RMS contrast levels of 7.5% and 15%. The yes/no task was used because it is more typical of real-world tasks in which one is not given the opportunity to compare the image with and without the target present. Also, more like natural tasks, the sample of 1/f noise and natural image background was different in each trial. Thresholds were measured for three observers (two were authors).

The targets were chosen because they represent three broad categories of targets: narrowband in frequency and orientation (Gabor), broadband in spatial frequency and orientation (Gaussian), and narrowband in orientation and broadband in frequency (Edge).

The background types were chosen to vary in the degree of similarity to natural backgrounds. Natural

backgrounds are extraordinarily complex, differing from uniform backgrounds along a number of different dimensions. One dimension is the shape of the average amplitude spectrum, which typically falls off inversely with spatial frequency (D. J. Field, 1987). Thus, as a first approximation to natural backgrounds, we used random noise backgrounds that have a  $1/f$  amplitude spectrum.

The  $1/f$  noise backgrounds are isotropic and stochastically stationary across space. However, natural backgrounds tend to vary across space in luminance, contrast, spatial frequency, orientation, and phase structure. Our second (closer) approximation to natural backgrounds was to include the spatial frequency, orientation, and phase structure but to control the variations in local luminance and contrast. To do this, we adjusted the gray scale histograms of natural images to match those of  $1/f$  noise with 7.5% and 15% contrast (see Stimuli). These “Gaussianized” natural images appear remarkably naturalistic (see Figure 13A in Discussion), and comparing detection performance in  $1/f$  noise with that in Gaussianized natural images allows us to isolate the effects of spatial frequency, orientation, and phase structure.

In future studies, we plan to measure detection thresholds in unaltered natural backgrounds, but we focused first on Gaussianized backgrounds because they are more useful for testing our model. Because of the large variations in local luminance and contrast in natural images, there are many trials, even for a fixed-amplitude target, in which the target will be either trivially detectable or trivially impossible to detect. Performance in such trials is easier for a model to predict, making unaltered natural images less useful.

## Stimuli

Eight-bit gray scale images were displayed on a calibrated monitor (Sony Trinitron, GDM-FW900) at a resolution of  $1920 \times 1080$  pixels and a frame rate of 60 Hz noninterlaced. The monitor was placed 168 cm from the eyes, and all stimuli were displayed at 120 pixels/°. The graphics card lookup table was set to produce 256 linear steps in luminance with a mean luminance of  $18 \text{ cd/m}^2$ .

There were three target stimuli in our Experiment: Gabor, Gaussian, and Edge. The Gabor was horizontal at  $4 \text{ c}/^\circ$  in cosine phase and had a bandwidth of one octave. The Gaussian had a  $SD$  of 8.43 arcmin. The Edge was horizontal and windowed with a Gaussian having a  $SD$  of  $0.5^\circ$ . These three targets were taken from the ModelFest stimulus set (ModelFest stimuli #12, #27, and #30, see Watson & Ahumada, 2005).

Targets were presented at the center of a  $512 \times 512$  background located within a larger mean luminance

background ( $18 \text{ cd/m}^2$ ,  $1920 \times 1080$ ). The target contrast is defined to be the difference between the peak and background luminance divided by the background luminance ( $18 \text{ cd/m}^2$ ). Depending on the background condition, the  $512 \times 512$  background was either set to mean luminance or randomly selected from either large  $1/f$  noise images ( $1280 \times 1280$ ) or from one of 10 large ( $4284 \times 2844$ ) “Gaussianized” natural images. In all conditions, the pixels on the edge of the  $512 \times 512$  background were set to black; this created a 1-pixel-wide box that cued the location of the background under all conditions. Detection measurements were obtained for uniform backgrounds and for  $1/f$  noise and natural backgrounds of 7.5% and 15% RMS contrast (i.e., five background conditions).

The natural images were randomly selected from a set of 1,200 calibrated natural images (available at [www.cps.utexas.edu/natural\\_scenes](http://www.cps.utexas.edu/natural_scenes)), and both the  $1/f$  noise and natural images were converted to eight-bit gray scale. The natural images were “Gaussianized” by matching their gray scale histograms to a  $1/f$  noise image. The first step was to rank-order the pixels in each image according to gray level from smallest to largest. Note that for each specific gray level, the fraction of pixels having that gray level will differ between the two images. The goal was to make the fraction of pixels at each gray level in the natural image the same as that in the  $1/f$  noise image. This was done in the second step by applying the following mapping:  $g_i = f_j$ , where  $g_i$  is the gray level of the natural image pixel having rank order  $i$  out of a total of  $N$  pixels,  $f_j$  is the gray level of the  $1/f$  noise pixel having rank order  $j$  out of a total of  $M$  pixels, with  $j = \lceil iM / N \rceil$ . (Note that  $N > M$ , and  $\lceil x \rceil$  is the “ceiling” function.) This mapping preserved the spatial frequency, orientation, and phase structure of natural images but allowed us to select patches from Gaussianized natural images with similar mean luminance and contrast as patches selected from our large  $1/f$  noise images. Specifically, for each randomly selected  $512 \times 512$  patch of  $1/f$  noise used in the Experiment, we randomly selected a patch of Gaussianized natural image having approximately the same mean luminance (the mean luminance differed by a maximum of  $1.45 \text{ cd/m}^2$ ); the RMS contrasts of the two patches were set to the same value (i.e., 7.5% or 15%).

## Procedure

Psychometric functions were measured in a single-interval, blocked, forced choice paradigm in which the observer judged whether a target was present or absent at the center of the background. Each psychometric function was based on at least 240 trials, collected in separate sessions of 120 trials each. For a given



condition in a session, four blocks of 30 trials each were run in descending order of target contrast. Eye position was monitored using an Eyelink 1000 eye tracker. If eye position deviated by more than  $1^\circ$  from the fixation dot, the trial was discarded and another trial added to the block.

Each 30-trial block began with a standard nine-point calibration procedure for the eye tracking. After the calibration procedure, the observer was required to hold fixation on a fixation dot for each of the 30 trials in the block. Each trial began with a 500-ms interval in which the background location was cued with a 1-pixel-wide black square outlining the background area. In conditions in which the target location was the center of the fovea, the fixation dot was extinguished 100 ms before onset of the test stimulus. The test stimulus consisted of a 250 ms presentation of either background or background-plus-target. At the end of this interval, there was a 2-s response window (mean luminance background) during which the observer could signal “target present” or “target absent” by pressing one of two buttons. Failure to respond led to the trial being replaced with a new one; this occurred less than 1% of the time. Feedback was given at the end of the 2-s response window with a high tone representing “correct” and a low tone representing “incorrect.” The next trial began immediately after feedback was given.

Psychometric functions were measured for 60 separate conditions (three stimuli  $\times$  four eccentricities  $\times$  five background conditions). The psychometric functions with uniform and  $1/f$  noise backgrounds were measured in a random order. Then the psychometric functions for the Gaussianized natural backgrounds were measured in a random order.

### Fitting psychometric functions and thresholds

As mentioned earlier, we used a yes/no task because it is more typical of natural conditions. Performance in all forced choice tasks can be influenced by criterion bias, but yes/no tasks are often thought to be more susceptible. Therefore, for each condition, we obtained maximum likelihood estimates of the threshold ( $c_t$ ), steepness parameter ( $\beta$ ), and criterion ( $\gamma$ ). Consistent with Equation 14, the probability of a hit is given by

$$P_h = \Phi\left(\frac{1}{2}\left(\frac{c}{c_t}\right)^\beta - \gamma\right) \quad (15)$$

and the probability of a false alarm by

$$P_{fa} = \Phi\left(-\frac{1}{2}\left(\frac{c}{c_t}\right)^\beta - \gamma\right) \quad (16)$$

Thus, the log likelihood of all the responses from a condition is

$$\begin{aligned} \ln L(c_t, \beta, \gamma) = & \sum_{i=1}^n N_h(c_i) \ln P_h(c_i) \\ & + N_m(c_i) \ln(1 - P_h(c_i)) \\ & + N_{fa}(c_i) \ln P_{fa}(c_i) \\ & + N_{cr}(c_i) \ln(1 - P_{fa}(c_i)) \end{aligned} \quad (17)$$

where  $n$  is the number of contrast levels of the target, and  $N_h(c_i)$ ,  $N_m(c_i)$ ,  $N_{fa}(c_i)$ , and  $N_{cr}(c_i)$  are the numbers of hits, misses, false alarms, and correct rejections, for contrast level  $c_i$ . We first estimated the parameters by maximizing Equation 17. We found that the values of the steepness parameter were consistent across conditions (see Results) and that there were no systematic variations in the criterion across conditions for a given observer. Thus, the final thresholds for each observer were obtained by setting the steepness parameter to the average across all subjects and conditions, setting the criterion to the average across conditions for that subject, and then finding the maximum likelihood estimate of the thresholds using Equation 17. Importantly, the pattern of thresholds was robust across different versions of this analysis (including ignoring criterion effects and only analyzing percent correct).

## Results

Maximum likelihood fits of Equations 15 and 16 to the psychometric data were used to obtain the estimated contrast threshold  $c_t$  for each of the 60 conditions. Figure 5 plots the square of the estimated contrast thresholds (threshold power) as a function of the square of the background contrast (background power). The open circles represent the average thresholds of three observers for three target stimuli (columns) presented at four retinal eccentricities (colors) in  $1/f$  noise and Gaussianized natural backgrounds (rows). The colored lines are linear fits to the data (not Model predictions). Note the thresholds measured in uniform backgrounds (background contrast of zero) are the same in both rows of plots and that the vertical scales are different for the different targets. The estimated criterion (bias) for the three observers in units of  $d'$  were 0.362 (JSA), 0.228 (CKB), and 0.277 (SPS).

Two principles of masking are suggested by these plots: (a) threshold contrast power increases linearly as a function of background contrast power, and (b) the slope of the best fitting line increases as a function of

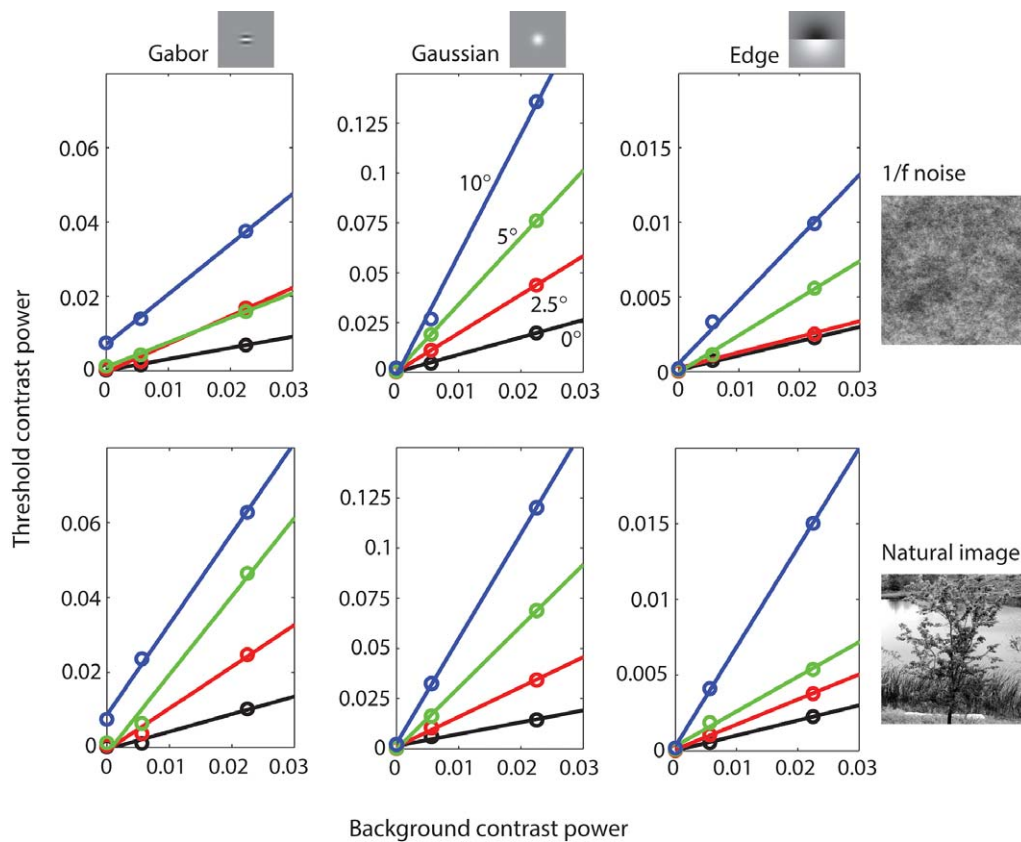


Figure 5. Detection threshold measurements for three different targets at four different eccentricities as function of background contrast power for  $1/f$  noise and natural backgrounds. Data points are the average of three observers. The solid lines are best fitting linear functions.

retinal eccentricity. Figure 6 shows more clearly how well our data are described by linear masking functions. In this figure, the data from each subject has been normalized so that the linear fits have an intercept of 0 and a slope of 1; also shown are the average thresholds (Figure 6a). If threshold contrast power is a linear function of background contrast power then the normalized data points should fall on a line of slope 1 through the origin (black line).

Although not easily seen in Figure 5, the intercepts of the masking function also increase with retinal eccentricity. Figure 7 plots the intercept as a function of retinal eccentricity for each type of target. The intercepts tend to increase exponentially with eccentricity (solid curves).

In general, the thresholds in  $1/f$  noise and in natural backgrounds are similar (see Figure 5). However, for the Gabor and Edge targets, masking was somewhat greater in the natural backgrounds. This can be seen clearly in Figure 8, which plots threshold in  $1/f$  noise as a function of threshold in natural backgrounds separately for each target. The points for the Gaussian target fall near or slightly below the diagonal, but the points for the Gabor and Edge target fall above the diagonal. Even though the

points do fall off the diagonal, they fall roughly on straight lines, indicating that thresholds in  $1/f$  noise and natural backgrounds differ approximately by a fixed proportionality constant that depends on the target.

The slopes of the psychometric functions were fairly constant over the 60 conditions. Figure 9a shows that the steepness parameter varies little across the three types of target. Figure 9b plots the average steepness parameters of the three targets for all background and retinal eccentricity conditions. On average, there is a slight trend for the parameter to increase with eccentricity. Overall, the average steepness parameter is 1.685. We take this average parameter value to be the estimate of  $\beta$  in the RV1 model.

## Model predictions

In fitting the model to the estimated thresholds we use Equation 11 and then Equation 14 if we need to predict thresholds for a different criterion percentage correct (e.g., 82% rather than 69%). In what follows, we estimate a subset (five) of the parameters by fitting the

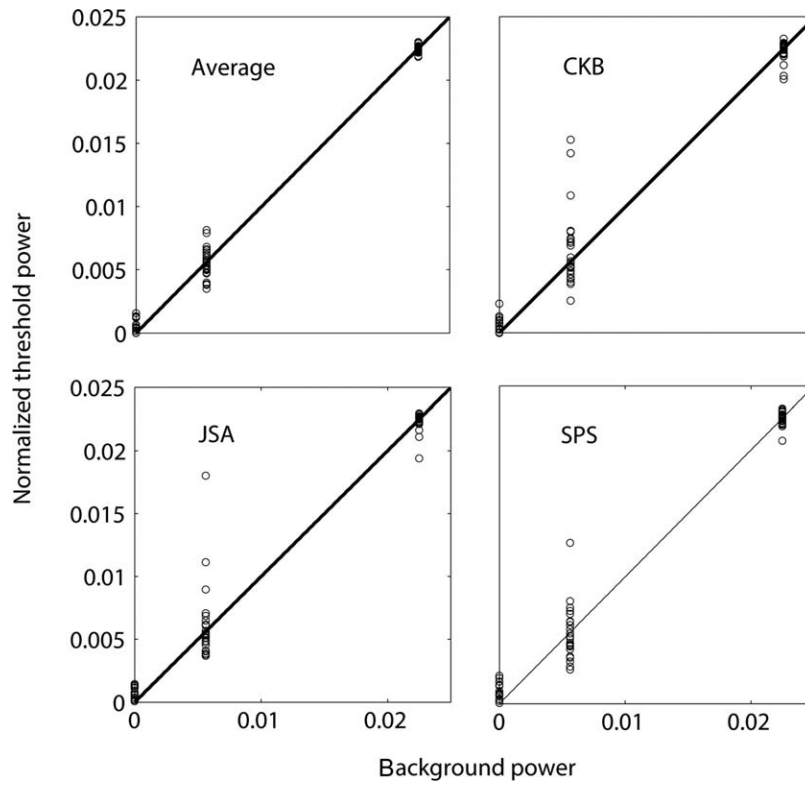


Figure 6. Normalized contrast threshold power as a function of background contrast power for all experimental conditions.

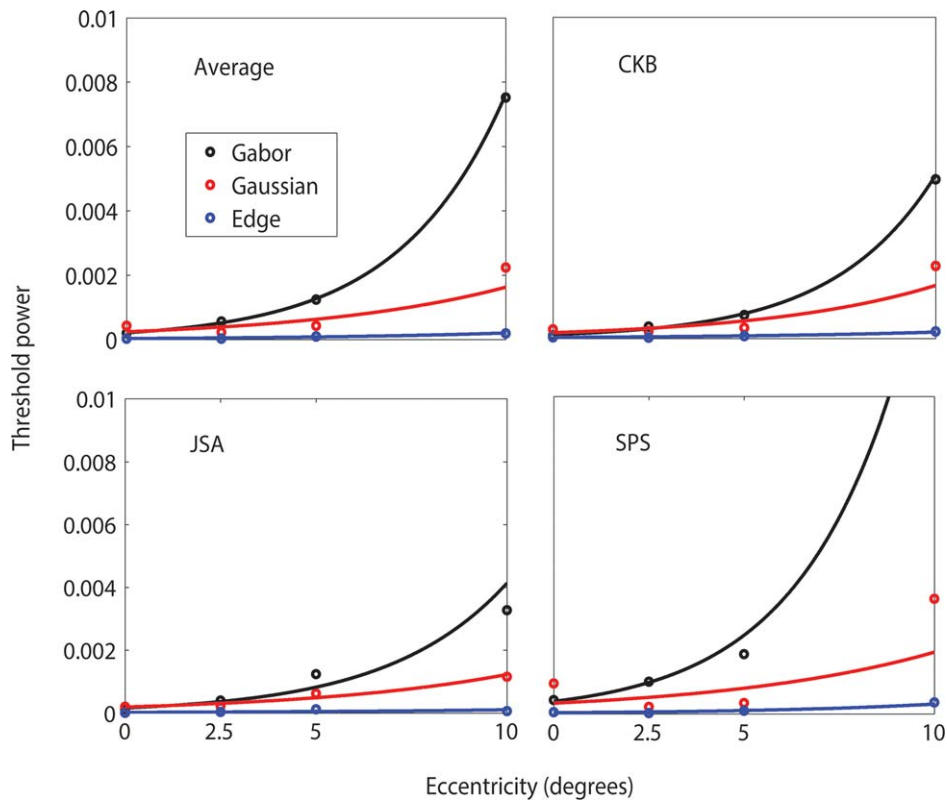


Figure 7. Threshold contrast power as function of eccentricity when background is uniform. Solid curves are best fitting exponential functions.

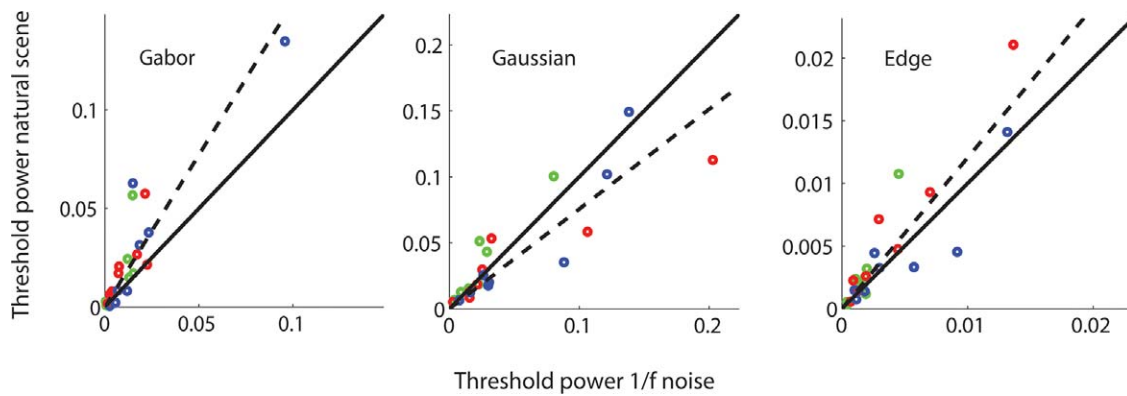


Figure 8. Threshold in Gaussianized natural images as a function threshold in 1/f noise for all conditions for the three subjects: CB (green symbols), SS (blue symbols), and JA (red symbols). Dashed curve is best fitting line through the origin.

model to the average thresholds (based on 16 observers) reported in the ModelFest study. Keeping these parameters fixed (except for  $P_0$  that scales all thresholds up and down), we estimate the remaining three parameters by fitting the average thresholds from the present Experiment. Finally, in the Discussion section, we keep the parameters fixed (except for  $P_0$ ) and generate predictions for the results of Foley et al. (2007).

Figure 10 shows the predictions of the RV1 model (red line) to the ModelFest data set (black points), which is the average contrast thresholds of 16 observers across 10 labs for foveal detection of 43 target stimuli in a uniform background. The thresholds in the ModelFest set are based on an 82% correct criterion. The thresholds are plotted in dB units. Thus, a 6-dB difference in threshold corresponds to a factor of two in contrast threshold. The RMS error of the model is comparable to the RMS error of the better models tested by Watson and Ahumada (2005). Recall that these predictions depend on five parameters: three that control the relative size and strength of the center and surround, a pooling exponent, and a baseline noise

parameter. The values of the parameters are given in the figure caption.

Figure 11 shows the predictions (solid curves) of the model for the present Experiment (plotted in dB units rather than contrast power as in Figure 5). The open circles show the average power contrast thresholds of the three observers for three target stimuli (columns) as a function of retinal eccentricity for three background contrasts (colors) in 1/f noise and Gaussianized natural backgrounds (rows). The plotted thresholds for detection in the uniform background (black open circles) are the same in both rows. In order to maximize compatibility with the ModelFest data, these thresholds are also based on the 82% correct criterion. In fitting these data, we kept the parameters values obtained from fitting the ModelFest data with the exception that we allowed the baseline noise parameter  $P_0$  to change. The only effect of the baseline noise parameter is to shift the predictions for uniform backgrounds vertically on the dB axis. Although we allowed the baseline noise parameter to vary, the estimated value was well within the range of individual

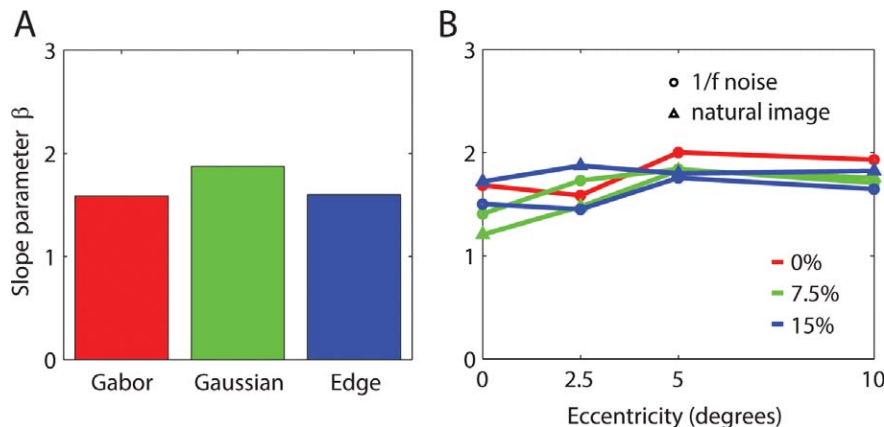


Figure 9. Psychometric function slope parameter. (A) Average slope parameter values for three types of target. (B) Average slope parameter as function of eccentricity for type of background and background contrast.

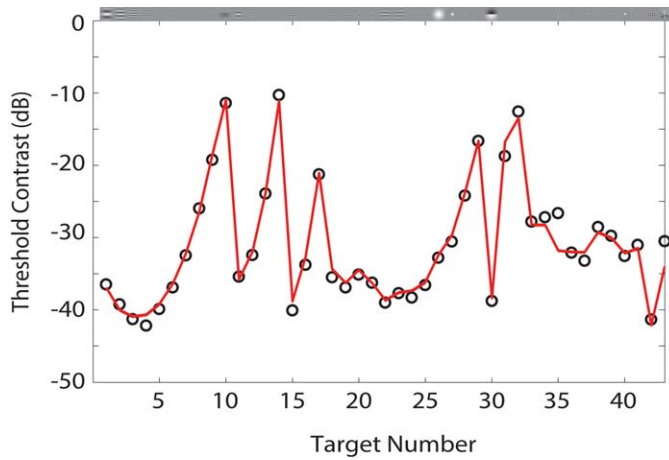


Figure 10. Predictions for ModelFest data set. Data points are average contrast thresholds of 16 observers in 10 labs for 43 different targets. The solid curve is the prediction of the RV1 model. Parameter values:  $k_c = 1$ ,  $k_s = 10.1$ ,  $w_c = 0.53$ ,  $\rho = 2.4$ ,  $\beta = 1.685$ ,  $P_0 = 1.4E-3$ . Note that changing  $P_0$  translates the entire predicted curve vertically on the logarithmic (dB) scale. Threshold contrast in dB =  $20\log(c_t)$ , where here  $c_t$  is the 82% correct contrast threshold (RMS error = 1.09 dB).

differences for that parameter in the ModelFest data set.

Recall that the model has three additional parameters for predicting thresholds in nonuniform backgrounds: overall pattern masking strength, the relative weight of the narrowband and broadband components, and the spatial area for local luminance gain control. Again, the values of the estimated parameters are given in the figure caption.

As can be seen, the model captures most of the variance in the thresholds but is qualitatively more accurate for the Gaussian and Edge targets than for the Gabor target. Note that the foveal thresholds for the three targets on the uniform background are similar to those in the ModelFest data set (our targets correspond to ModelFest stimuli #12, #27, and #30 in Figure 10).

## Discussion

We describe an attempt to develop a computationally efficient (practical) model that can predict the detectability of spatially localized targets presented at arbitrary retinal locations in arbitrary backgrounds, in which the target and retinal location are known to the

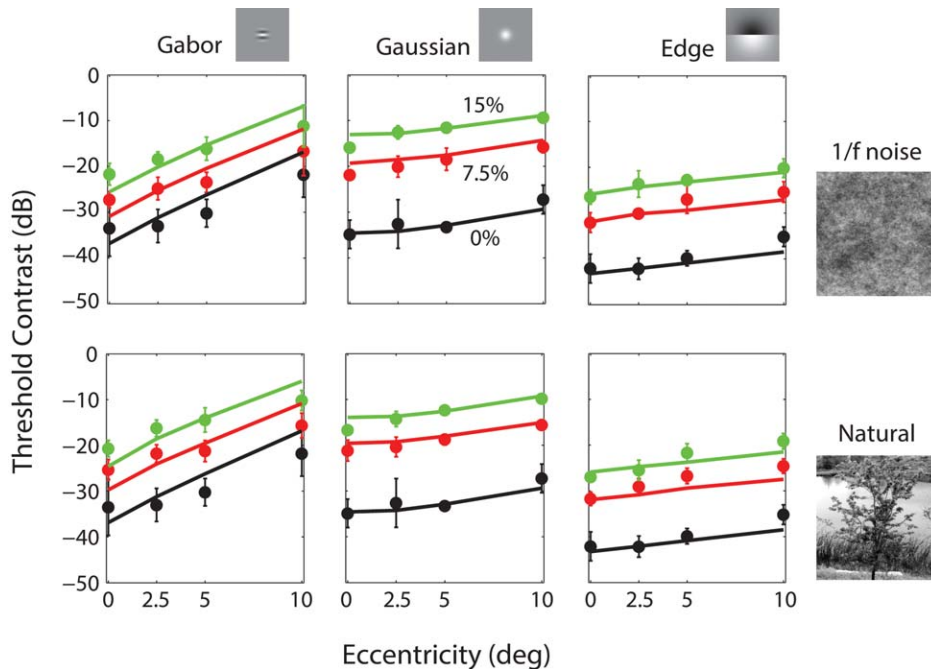


Figure 11. Predictions for the data from the present Experiment. Data points are the average contrast thresholds of three observers for 60 different conditions (three targets  $\times$  four eccentricities  $\times$  three background contrast levels for two kinds of background; the 0% background contrast is a uniform field, and hence, the black points are the same in the upper and lower plots). Error bars represent  $\pm 2$  standard errors (across observers). Parameter values:  $k_c = 1$ ,  $k_s = 10.1$ ,  $w_c = 0.53$ ,  $\rho = 2.4$ ,  $\beta = 1.685$ ,  $P_0 = 4.45E-4$ ,  $\sigma_l = 1^\circ$ ,  $k_b = 25$ ,  $w_b = 0.962$ . The first six parameters are the same as in Figure 10 except for  $P_0$ , which accounts for (modest) differences in overall sensitivity between groups of observers (RMS error = 2.27 dB).

observer. The model is based directly on known optical, retinal, and V1 properties of the human/primate visual system, and hence, there are only nine free parameters: the spatial extent and relative weight of ganglion cell center and surround mechanisms; the spatial extent of local luminance gain control; the strengths of the baseline, broadband, and narrowband masking power; a response-pooling exponent; and a parameter controlling the slope of the psychometric function. We find that the model is computationally efficient and does a respectable job of predicting detection performance for a wide range of targets on uniform,  $1/f$  noise, and natural image backgrounds for retinal eccentricities ranging from  $0^\circ$  to  $10^\circ$ .

## Uniform backgrounds

The predictions of the RV1 model for uniform backgrounds are determined by six parameters. One of these parameters,  $\beta$ , was determined from the average steepness of the psychometric functions in the Experiment reported here. We find  $\beta = 1.685$ , which we note corresponds to a Weibull slope parameter of 2.13. The remaining five parameters can be estimated from the detection thresholds measured on uniform backgrounds in the fovea. To estimate these parameters we fit the ModelFest data set, which consists of foveal detection thresholds measured for 43 different targets on 16 observers in 10 different laboratories. The fit of the model to the ModelFest data is good, comparable to (slightly worse than) the best nonphysiologically based models (see Watson & Ahumada, 2005).

The estimated parameters for the midget ganglion cell receptive fields are reasonably consistent with the anatomy and physiology of the primate retina. Our psychophysical estimate of the  $SD$  of the ganglion cell center mechanism  $\sigma_c$  is almost exactly equal to the spacing between the (on or off) midget ganglion cells, which in the central visual field is approximately equal to the spacing between the photoreceptors (about a half minute of arc). This is consistent with the anatomical finding that in the central visual field a midget ganglion cell synapses with one midget bipolar cell, which synapses with one cone photoreceptor. The measured width of center mechanisms with single-unit recording is larger than a single cone, but the larger size is expected because of the effect of the optical psf; the measured center mechanism should be the convolution of the physiological center mechanism and the optical psf. Croner and Kaplan (1995) report that in the central  $5^\circ$  the median  $SD$  of the center mechanism is  $0.03^\circ$  and of the surround mechanism is  $0.18^\circ$  (about six times larger than the center). We computed the effective center  $SD$ s for our model and find that they range from  $0.021^\circ$  at  $0^\circ$  eccentricity to  $0.038^\circ$  at  $5^\circ$  eccentricity,

spanning the value reported by Croner and Kaplan. Similarly, the effective surround  $SD$  for the model ranges from  $0.077^\circ$  (3.6 times larger than center) at  $0^\circ$  eccentricity to  $0.3^\circ$  (7.9 times larger) at  $5^\circ$  eccentricity. Finally, Croner and Kaplan report that the relative weight on the center mechanism  $w_c$  is about 0.64 whereas our estimate is 0.53. Thus, we also find greater weight for the center mechanism but not by as large a factor.

The ModelFest data set only contains measurements made in the center of the fovea. In the present Experiment, we made measurements for three of the ModelFest targets at four eccentricities (black circles in Figure 11) and obtained reasonable predictions (solid curves) without altering parameters except that we allowed the baseline masking power  $P_0$  to change from  $1.4E-3$  to  $4.5E-4$  to account for modest differences in overall sensitivity among different groups of observers.

As a further test of the model, we generated predictions for the detection thresholds reported in Foley et al. (2007). In their experiment 1, Foley et al. measured thresholds for vertical  $4\ c/^\circ$  Gabor targets at retinal eccentricities ranging from  $-5^\circ$  to  $5^\circ$  along the horizontal meridian. In three observers, thresholds were measured for a cosine-phase Gabor having an envelope  $SD$  of  $0.25^\circ$ . In two other observers, thresholds were measured for a sine-phase Gabor having an envelope  $SD$  of  $0.18^\circ$ . The symbols in Figure 12A show the average thresholds. The solid curve shows the prediction of the RV1 model without altering parameters except for the baseline masking power (see figure caption). In their experiment 2, Foley et al. measured thresholds in the fovea for  $4\ c/^\circ$  Gabor targets in cosine phase (Figure 12B), sine phase (Figure 12C), and antic cosine phase (Figure 12D) for various areas and aspect ratios in two observers. The blue symbols show the thresholds for Gabor targets with a radially symmetric envelope. In this case, the horizontal axis gives the  $SD$  of the envelope in all directions. The red symbols show the thresholds for Gabor targets that are elongated parallel to the orientation of the grating. In this case, the horizontal axis gives the  $SD$  of the envelope in the parallel direction, with the  $SD$  in the perpendicular direction fixed at  $0.25^\circ$ . The green symbols show the thresholds for Gabor targets that are elongated perpendicular to the orientation of the grating. In this case, the horizontal axis gives the  $SD$  in the perpendicular direction, with the  $SD$  in the parallel direction fixed at  $0.25^\circ$ . The solid curves show the predictions of the RV1 model.

We have only evaluated the predictions of the model out to  $10^\circ$  eccentricity. However, if it works well over this range, then the literature suggests that it would apply over a wider range (Peli, Yang, & Goldstein, 1991).

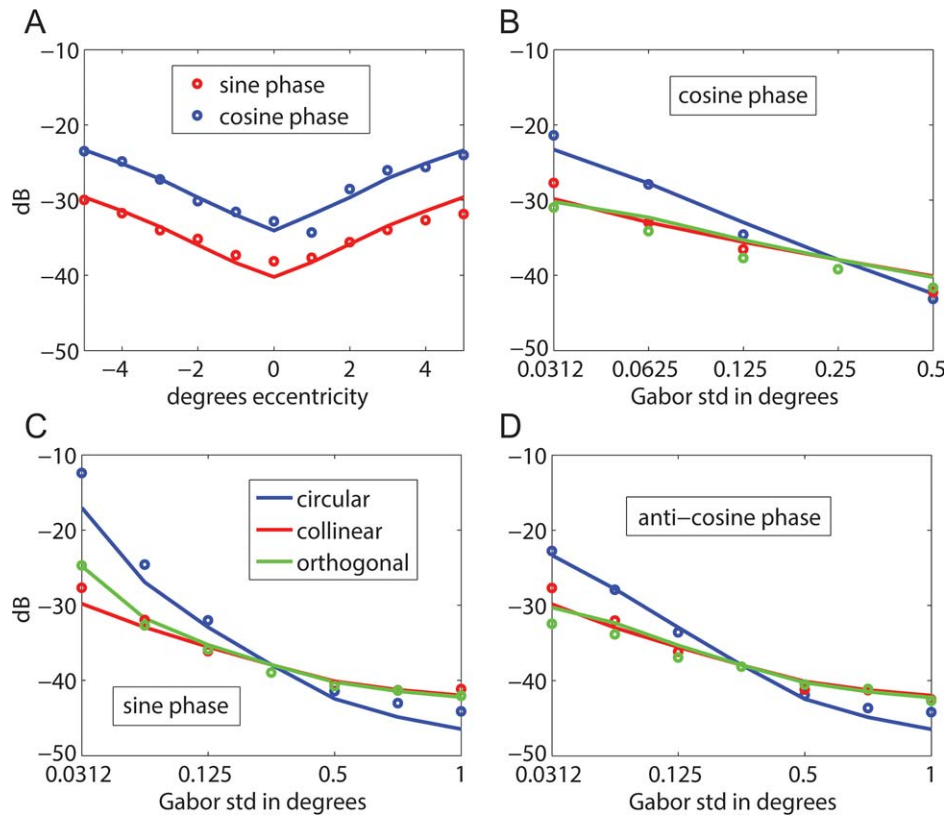


Figure 12. Predictions for data from Foley et al. (2007). (A) Threshold as a function of eccentricity for 4  $c/^\circ$  radially symmetric Gabor targets in sine phase (envelope  $SD = 0.25^\circ$ ; two observers;  $P_0 = 5.4E-4$ ) and cosine phase (envelope  $SD = 0.18^\circ$ ; three observers;  $P_0 = 1.3E-3$ ). (B–D) Threshold as a function of envelope  $SD$  for 4  $c/^\circ$  Gabor targets with a circular envelope, an envelope elongated collinear with the grating, and an envelope elongated orthogonal to the grating (two observers;  $P_0 = 9.2E-4$ ). Solid curves are the predictions of the RV1 model with the same parameters as in Figure 10 except for  $P_0$ , which accounts for (modest) differences in overall sensitivity between groups of observers (RMS error = 1.28 dB).

The relatively good fit of the model to all the uniform background data and the reasonable agreement of the estimated parameters with retinal anatomy and physiology suggest that optical and retinal factors may be the primary factors causing the variation in detection thresholds across different targets on uniform backgrounds. This is not implausible given that the optic nerve is arguably the major information transmission bottleneck in the visual pathway, making it possible for cortical circuits to process the ganglion cell responses with relatively constant efficiency across the different targets. The largest errors (underestimates) of the thresholds in Figure 10 occur for the two spatially complex targets (binary noise, #34, and cityscape, #43), for which it is reasonable to expect reduced central efficiency in pooling all the relevant features.

### Patterned backgrounds

The predictions of the RV1 model for patterned backgrounds depend on three additional parameters. To estimate these remaining parameters and provide a

further test of the model, we measured psychometric functions for Gabor, Gaussian, and Edge targets at four different eccentricities in uniform backgrounds in  $1/f$  noise backgrounds and in natural backgrounds whose gray scale histogram has been adjusted to match that of  $1/f$  noise. The predictions are good but slightly poorer for the Gabor target than for the Gaussian and Edge targets (see Figure 11). It is interesting to note, however, that the average thresholds reported by Foley et al. (2007) for the Gabor target (Figure 12a) increase slightly faster with eccentricity in better agreement with the RV1 model.

Perhaps the most remarkable result is that the model does about as well predicting detection thresholds in Gaussianized natural backgrounds as it does in  $1/f$  noise backgrounds and that the thresholds for the two kinds of background are similar. The background masking effects in the model are entirely based on the narrowband and broadband power in the ganglion cell responses, not on the specific phase structure, which differs greatly between the natural image and  $1/f$  noise backgrounds. Perhaps the trial-to-trial variation in the backgrounds is hiding the effect of the phase structure.

That is, thresholds may be similar in the two types of background only because in some trials the phase structure helps detection and in other trials it hurts detection. However, if this were true, then one might expect shallower psychometric functions for natural backgrounds. In fact, the slope parameter of the psychometric functions is similar for uniform,  $1/f$  noise, and Gaussianized natural backgrounds (see Figure 9). It would appear that for Gaussianized natural backgrounds, the complex phase structure of natural backgrounds has, practically speaking, a relatively minor effect on detection thresholds.

A limitation of our test of the RV1 model for patterned backgrounds is that it is based on data for only three different targets. However, note that the pattern-masked thresholds for these three targets tend to parallel (on a log scale) the thresholds obtained on a uniform background (see Figure 11). This suggests that the pattern-masked thresholds for other ModelFest targets would also tend to parallel those obtained on a uniform background. Thus, it seems likely that the predictions of the RV1 model would be of similar accuracy for the other ModelFest targets given the accuracy of its predictions for the other ModelFest targets on uniform backgrounds.

### Components in the model and components not in the model

The RV1 model contains a number of different components, and they each play an important role in the predictions. The optical psf has a substantial effect on the shape of the contrast sensitivity function (CSF) (especially the high-frequency falloff) and on how rapidly thresholds rise with eccentricity; thresholds for high-frequency targets would rise more rapidly without the effect of the optics because the effective ganglion cell center size would grow more rapidly.

Obviously, the discrete sampling function has a big effect. The number of samples declines rapidly with eccentricity, and hence, the maximum amount of retinal image information transmitted by the ganglion cells for high-frequency and broadband (e.g., natural or  $1/f$  noise) images drops rapidly.

The continuous variation in ganglion cell receptive field size with the sample spacing is also important. For example, consider the CSF in the fovea. In Figure 10, the thresholds for stimuli 1 through 10 give the CSF for targets with a fixed spatial extent, and the thresholds for stimuli 11 through 15 give the CSF for targets with a fixed numbers of cycles. These CSFs are not well approximated by a difference of Gaussians (Watson & Ahumada, 2005), which is the shape of the ganglion cell receptive fields. The relatively accurate prediction of the RV1 model is due in part to the fact that there is a

distribution of ganglion cell receptive field sizes falling under the stimuli.

In agreement with the masking literature (Eckstein, Ahumada, & Watson, 1997b; Foley, 1994; Watson & Solomon, 1997), we find that both the narrowband and the broadband masking components are important. If the parameters are estimated with the weight on the narrowband component set to zero ( $w_b = 0.0$ , see Equation 6), then the predictions are substantially worse. Conversely, if the parameters are estimated with the weight on the broadband component set to zero, then predictions are also substantially worse. Although the estimated weight is much higher on the narrowband component ( $w_b = 0.962$ ) than the broadband component ( $1 - w_b = 0.038$ ), they both play an important role. In fact, the average total masking power due to the broadband and narrowband components is about equal across the three targets:  $w_b P_{nb} \cong (1 - w_b) P_{bb}$ . More specifically, the average ratio of narrowband to broadband masking power is smallest for the Gabor target (0.165), intermediate for the Edge target (0.98), and largest for the Gaussian target (2.27). Although we find that both components are important in the current version of the model, the result may depend on how the target-dependent filter is computed. It is perhaps also worth emphasizing that broadband and narrowband components have no effect on the predictions for uniform backgrounds.

There are some well-known components that are not included in the RV1 model. One is a component that would produce the oblique effect; foveal detection thresholds tend to be higher for gratings oriented along the diagonals (Campbell, Kulikowski, & Levinson, 1966; McMahan & MacLeod, 2003). This effect is most likely cortical in origin (McMahan & MacLeod, 2003). We left out this component because the underlying anatomy and neurophysiology are not well understood and because including the oblique effect produces only minor improvements in prediction accuracy for the stimuli tested here (Watson & Ahumada, 2005). However, it would not be difficult to include in the model.

A second missing component is one that would produce the dipper effect: When the masker has the same (or nearly the same) shape as the target, then the detection threshold reaches a minimum (dips) when the contrast of the masker is itself at or near detection threshold (Legge & Foley, 1980). The dipper effect has been modeled with an accelerating (or threshold) nonlinearity prior to late noise (e.g., Foley, 1994; Goris et al., 2013; Legge & Foley, 1980; Watson and Solomon 1997). We left out this component because it would reduce the computational efficiency of the RV1 model (which depends on linearity) and because the dipper effect is likely to occur relatively infrequently under natural conditions. The dipper effect would be potentially present only in low background contrast regions,



and it is reduced or disappears if the target and masker differ in shape, which they generally do in  $1/f$  noise or natural images.

A third missing component is one that would produce some of the stronger crowding effects: Identification of targets can be strongly suppressed by the presence of surrounding objects or textures that are sufficiently similar to the target (for a review, see Levi, 2008). We did not try to include explicit crowding mechanisms because the current aim is to predict detection rather than identification performance (detection is a special case of identification). However, it is interesting that the RV1 model is able to predict detection in natural backgrounds without including the kinds of mechanisms (extended feature or texture integration) thought to underlie crowding. For example, natural backgrounds are filled with edges of various scales and orientations, yet threshold for the edge target across the visual field is accurately predicted from only the background power falling under the envelope of the target (note the envelope expands slightly with eccentricity, see Appendix). Like crowding paradigms, doesn't detection in this case involve identifying whether the specific target is present as opposed to whatever other edge shape or object might be at that location? (Recall that in the present yes/no task, the observer does not get to compare target + background with background alone.) Perhaps the success of the model is because in natural scenes (and in  $1/f$  noise) the target is on average not very similar to the background surrounding the target. This raises the question: How important are crowding effects when looking for specific targets in natural scenes? If one takes an arbitrary target and adds it at a random location in a natural image, then does the target tend to be sufficiently similar to the surrounding background for crowding effects to be strong relative to the more local masking effects? It may be possible to answer this question by analysis of natural image statistics. Of course, in some natural cases, crowding effects are known to be very important (e.g., reading) and in other cases are likely to be very important (e.g., detecting animals, which often mimic the backgrounds in their natural habitat).

## Comparison to previous models

The RV1 model borrows heavily from previous models of pattern detection (as indicated by the references in earlier sections) but has several unique features. First, the model directly incorporates physical measurements of the optics of the eye and of the anatomy and physiology of retinal ganglion cells, extending an earlier attempt to do this (Arnou & Geisler, 1996). This approach exploits known physical and physiological constraints and hence reduces the

number of free parameters. Second, there are few models of pattern detection that explicitly model the variation in spatial resolution across the visual field. Indeed most models focus exclusively on detection in the fovea, which reduces their generality and utility. Third, the model takes into account the spatial frequency and orientation selectivity of cortical populations (channels) by applying a target-dependent filter (that varies with eccentricity) to the modeled ganglion cell responses. This approach allows for very efficient computation while still representing the information processing carried out by the (very large) cortical population. Fourth, the implementation of the model makes extensive use of Gaussian stacks, which make it possible to rapidly generate predictions for arbitrary locations across the visual field even though the visual system is highly inhomogeneous (shift variant).

There are a number of previous pattern masking models that include narrowband spatial frequency channel masking together with broadband contrast masking (e.g., M. P. Eckstein et al., 1997b; Foley, 1994; Goris et al., 2013; Rohaly, Ahumada, & Watson, 1997; Watson & Solomon, 1997), and some of these have been applied to detection of targets in natural backgrounds (e.g., M. P. Eckstein et al., 1997b; Rohaly et al., 1997). These models differ from the present model in that they explicitly represent the spatial frequency channels rather than implicitly with a target-dependent filter, and they do not explicitly represent the variation in spatial resolution across the visual field. As mentioned above, the models of Foley (1994), Goris et al. (2013), and Watson & Solomon (1997) include an accelerating nonlinearity to account for the dipper effect, which the present model does not.

Another class of pattern masking model is related to ideal detection in noise (Burgess, 2011; Burgess et al., 1981; M. P. Eckstein et al., 1997a; Legge et al., 1987; Myers & Barrett, 1987; Zhang, Pham, & Eckstein, 2006). These models typically involve first characterizing the statistical properties of the backgrounds plus any assumed internal noise properties or constraints and then deriving a model (ideal) observer that is optimal given those statistical properties and constraints. This is a more principled approach that often yields nearly parameter-free predictions and can provide deeper insight into neural computation (Geisler, 2011). This approach has been extensively developed in the area of medical imaging perception (e.g., see Burgess, 2011; Samei, 2010; Zhang et al., 2006). The RV1 model does not directly consider the statistical properties of backgrounds and hence is not an ideal observer model; however, it borrows from this approach by regarding the baseline, narrowband, and broadband masking effects as a combined equivalent noise in a signal-detection framework. In the future, an ideal observer analysis that includes the biological

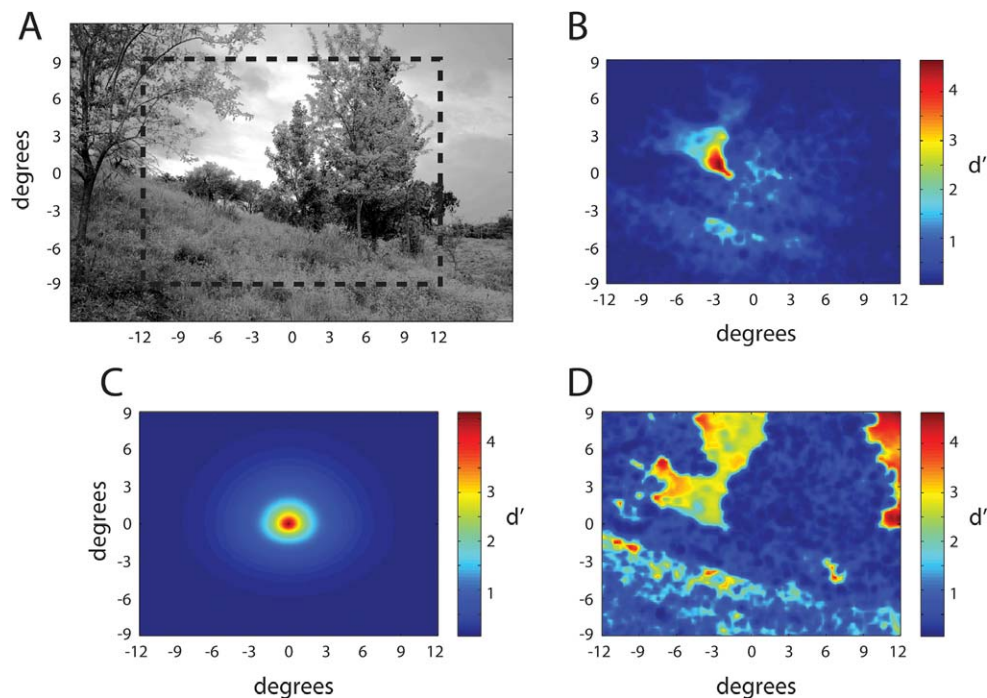


Figure 13. Detectability maps for a 4  $c/^\circ$  Gabor target. (A) Gaussianized natural image. (B) Detectability ( $d'$ ) of the target at all locations within the dashed box in (A) given fixation at the center of the scene. (C) Detectability of the target presented in the center of the scene for all possible fixation locations within the dash box in (A). (D) Detectability of the target at all locations within the dash box in (A) given fixation at the location of the target. The target contrast was fixed within each map but was set so that  $d'$  reached a maximum of 4.5.

constraints represented in the RV1 model may provide deeper insights into, and new predictions for, the neural computations underlying pattern detection in natural scenes. Nonetheless, the RV1 model may prove useful because it is (a) based directly on known biological constraints, (b) contains few parameters, (c) is extensible, (d) takes images of the background and target as input and produces a predicted performance ( $d'$ ) or predicted response (yes/no) as output, (e) can generate predictions across the visual field for arbitrary backgrounds, and (f) is computationally efficient.

## Detectability maps

The computational efficiency of the RV1 model makes it possible to generate maps of target detectability across the visual field for arbitrary backgrounds. Figure 13A shows a Gaussianized natural image that is  $24^\circ$  across. Figure 13B through D illustrates three different types of detectability ( $d'$ ) map.

Figure 13B shows the  $d'$  map for all possible locations of a fixed-contrast 4  $c/^\circ$  Gabor target given fixation in the center of the image (0,0). As can be seen,  $d'$  is predicted to be higher near the fixation point but also to vary greatly depending on the background content at the target location. Such  $d'$  maps for target

location may be useful in making predictions for a single fixation search, in which the observer's task is to detect the target during a single brief presentation when the location of the target is uncertain. For example, these *location  $d'$  maps* could be used to determine the best possible search performance, assuming perfect parallel processing of all potential target locations. This is a critical baseline analysis for interpreting the results of visual search and attention experiments (e.g., see Eckstein, 2011; Geisler & Cormack, 2011).

Figure 13C shows the  $d'$  map for all possible fixation locations given that the target location is at the center of the image. In this case,  $d'$  falls smoothly away from the target location. This *fixation  $d'$  map* is closely related to the conspicuity area—the spatial region around a target where it can be detected in the background (Bloomfield, 1972; Engel, 1971; Geisler & Chou, 1995; Toet, Kooi, Bijl, & Valeton, 1998). The conspicuity area can be defined as the area of the region where  $d'$  exceeds some fixed criterion. Previous studies (Geisler & Chou, 1995; Toet et al., 1998) have shown that there is a strong negative correlation (on the order of  $-0.8$  to  $-0.9$ ) between the conspicuity area and the time it takes humans to locate the target even in natural scenes (Toet et al., 1998). This is a powerful result of theoretical importance and of potential practical value. But to be of practical value, one must know the

conspicuity area for the particular target at its particular location in the background. The RV1 model might prove useful for estimating conspicuity areas without having to directly measure them in preliminary psychophysical experiments.

Finally, Figure 13D shows the  $d'$  map for all possible target locations when the observer is directly fixating the target. Such *foveal  $d'$  maps* could be used to determine the best possible search accuracy for a given target in a given background given unlimited search time.

*Keywords:* spatial vision, detection, masking, peripheral vision, ganglion cells, natural images

## Acknowledgments

We thank John Foley for comments and providing the data in Figure 12. We also thank Steve Sebastian. Supported by NIH Grants EY02688 and EY11747.

Commercial relationships: none.

Corresponding author: Wilson S. Geisler.

Email: geisler@psy.utexas.edu.

Address: University of Texas at Austin, Austin, TX, USA.

## References

- Albrecht, D. G., & Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience*, *7*, 531–546.
- Arnou, T. L., & Geisler, W. S. (1996). Visual detection following retinal damage: Predictions of an inhomogeneous retino-cortical model. *SPIE Proceedings: Human Vision and Electronic Imaging*, *2674*, 119–130.
- Bloomfield, J. R. (1972). Visual search in complex fields: Size differences between target disc and surrounding discs. *Human Factors*, *14*, 139–148.
- Burgess, A. E. (2011). Visual perception studies and observer models in medical imaging. *Seminars in Nuclear Medicine*, *41*, 419–436.
- Burgess, A. E., & Colborne, B. (1988). Visual signal detection: IV. Observer inconsistency. *Journal of the Optical Society of America A*, *2*, 617–627.
- Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science*, *214*, 93–94.
- Burton, G. J., & Moorehead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics*, *26*, 157–170.
- Campbell, F. W., Kulikowski, J. J., & Levinson, J. Z. (1966). The effect of orientation on the visual resolution of gratings. *Journal of Physiology*, *187*, 427–436.
- Carandini, M., & Heeger, D. J. (2012). Normalization as canonical neural computation. *Nature Reviews Neuroscience*, *13*, 51–62.
- Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, *17*, 8621–8644.
- Croner, L. J., & Kaplan, E. (1995). Receptive fields of P and M ganglion cells across the primate retina. *Vision Research*, *35*(1), 7–24.
- Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *The Journal of Comparative Neurology*, *300*, 5–25.
- Dacey, D. M. (1993). The mosaic of midget ganglion cells in the human retina. *Journal of Neuroscience*, *13*, 5334–5355.
- Dacey, D. M., & Peterson, M. R. (1992). Dendritic field size and morphology of midget and parasol ganglion cells of the human retina. *Proceedings of the National Academy of Sciences, USA*, *89*, 9666–9670.
- Derrington, A. M., & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurons in lateral geniculate nucleus of macaque. *Journal of Physiology*, *357*, 219–240.
- De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 545–559.
- De Valois, R. L., & DeValois, K. K. D. (1988). *Spatial vision*. Oxford: Oxford University Press.
- De Valois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 531–544.
- Drasdo, N., Millican, C. L., Katholi, C. R., & Curcio, C. A. (2007). The length of Henle fibers in the human retina and a model of ganglion receptive field density in the visual field. *Vision Research*, *47*, 2901–2911.
- Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, *11*(5):14, 1–36, <http://www.journalofvision.org/content/11/5/14>, doi:10.1167/11.5.14. [PubMed] [Article]
- Eckstein, M. P., Ahumada, A. J., & Watson, A. B. (1997a). Visual signal detection in structured backgrounds. II. Effects of contrast gain control, background variations, and white noise. *Journal of the Optical Society of America A*, *14*, 2406–2419.

- Eckstein, M. P., Ahumada, Jr., A. J., & Watson, A. B. (1997b). Image discrimination models predict signal detection in natural medical image backgrounds. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human vision, visual processing, and digital display VIII, proc. vol. 3016* (pp. 44–56). Bellingham, WA: SPIE.
- Engel, F. (1971). Visual conspicuity, directed attention and retinal locus. *Vision Research*, *11*, 563–576.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*, 2379–2394.
- Field, G. D., & Chichilnisky, E. J. (2007). Information processing in the primate retina: Circuitry and coding. *Annual Review of Neuroscience*, *30*, 1–30.
- Foley, J. M. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A*, *11*(6), 1710–1719.
- Foley, J. M., Varadharajan, S., Koh, C. C., & Farias, M. C. (2007). Detection of Gabor patterns of different sizes, shapes, phases and eccentricities. *Vision Research*, *47*(1), 85–107.
- Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, *51*, 771–781.
- Geisler, W. S., & Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: Detection, discrimination and identification. *Visual Neuroscience*, *14*, 897–919.
- Geisler, W. S., & Chou, K. (1995). Separation of low-level and high-level factors in complex tasks: Visual search. *Psychological Review*, *102*(2), 356–1378.
- Geisler, W. S., & Cormack, L. (2011). Models of overt attention. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *Oxford handbook of eye movements*. New York: Oxford University Press.
- Goris, R. L. T., Putzes, T., Wagemans, J., & Wichmann, F. A. (2013). A neural population model for visual pattern detection. *Psychological Review*, *120*(3), 472–496.
- Graham, N. V. (1977). Visual detection of aperiodic spatial stimuli by probability summation among narrowband channels. *Vision Research*, *17*, 637–652.
- Graham, N. V. (1989). *Visual pattern analyzers*. New York: Oxford.
- Graham, N. V. (2011). Beyond multiple pattern analyzers as linear filters (as classical V1 simple cells): Useful additions of the last 25 years. *Vision Research*, *51*, 1397–1430.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Heeger, D. J. (1991). Nonlinear model of neural responses in cat visual cortex. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual perception* (pp. 119–133). Cambridge, MA: The MIT Press.
- Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*, 191–197.
- Hood, D. C. (1998). Lower-level visual processing and models of light adaptation. *Annual Review of Psychology*, *49*, 503–535.
- Hood, D. C., & Finkelstein, M. A. (1986). Sensitivity to light. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance (vol. 1)* (pp. 51–66). New York: John Wiley and Sons.
- Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, *70*, 1458–1471.
- Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A*, *4*(2), 391–404.
- Levi, D. M. (2008). Crowding—an essential bottleneck for object recognition: a mini-review. *Vision Research*, *48*, 635–654.
- Lu, Z. L., & Doshier, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *Journal of the Optical Society of America A*, *16*, 764–778.
- Lu, Z. L., & Doshier, B. A. (2008). Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review*, *115*, 44–82.
- McMahon, M. J., & MacLeod, D. I. (2003). The origin of the oblique effect examined with pattern adaptation and masking. *Journal of Vision*, *3*(3):4, 230–239, <http://www.journalofvision.org/content/3/3/4>, doi:10.1167/3.3.4. [PubMed] [Article]
- Merigan, W. H., Katz, L. M., & Maunsell, J. H. R. (1991). The effects of parvocellular lateral geniculate lesions on the acuity and contrast sensitivity of macaque monkeys. *The Journal of Neuroscience*, *11*, 994–1001.
- Merigan, W. H., & Maunsell, J. H. R. (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience*, *16*, 369–402.
- Myers, K. J., & Barrett, H. H. (1987). Addition of a channel mechanism to the ideal-observer model. *Journal of the Optical Society of America A*, *4*, 2447–2457.
- Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision - A phase-dependent energy model. *Proceedings of the Royal Society of London Series B - Biological Sciences*, *235*, 221–245.
- Najemnik, J., & Geisler, W. S. (2005). Optimal eye

- movement strategies in visual search. *Nature*, *434*, 387–391.
- Navarro, R., Artal, P., & Williams, D. R. (1993). Modulation transfer function of the human eye as a function of retinal eccentricity. *Journal of the Optical Society of America A*, *10*, 201–212.
- Palmer, L. A., Jones, J. P., & Stepnoski, R. A. (1991). Striate receptive fields as linear filters: Characterization in two dimensions of space. In A. G. Leventhal (Ed.), *The neural basis of visual function* (pp. 246–265). Boston: CRC Press.
- Peli, E., Yang, J., & Goldstein, R. B. (1991). Image invariance with changes in size: The role of peripheral contrast thresholds. *Journal of the Optical Society of America A*, *8*, 1762–1774.
- Quick, R. (1974). A vector-magnitude model of contrast detection. *Biological Cybernetics*, *16*, 65–67.
- Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, *5*, 583–601.
- Rohaly, A. M., Ahumada, Jr. A. J. & Watson, A. B. (1997). Object detection in natural backgrounds predicted by discrimination performance and models. *Vision Research*, *37*, 3225–3235.
- Samei, E. (Ed.). (2010). *The handbook of medical image perception and techniques*. New York: Cambridge University Press.
- Shapley, R. M., & Lennie, P. (1985). Spatial frequency analysis in the visual system. *Annual Review of Neuroscience*, *8*, 547–583.
- Sit, Y. F., Chen, Y., Geisler, W. S., Miikkulainen, R., & Seidemann, E. (2009). Complex dynamics of V1 population responses explained by a simple gain-control model. *Neuron*, *64*, 943–956.
- Toet, A., Kooi, F. L., Bijl, P., & Valeton, J. M. (1998). Visual conspicuity determines human target acquisition performance. *Optical Engineering*, *37*, 1969–1975.
- Watson, A. B. (1979). Probability summation over time. *Vision Research*, *19*, 515–522.
- Watson, A. B., & Ahumada, A. J. (2005). A standard model for foveal detection of spatial contrast. *Journal of Vision*, *5*(9):6, 717–740, <http://www.journalofvision.org/content/5/9/6>, doi:10.1167/5.9.6. [PubMed] [Article]
- Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A*, *14*, 2379–2391.
- Watt, R. J., & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research*, *25*, 1661–1674.
- Wilson, H. R., & Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Research*, *19*, 19–32.
- Zhang, Y. L., Pham, B. T., & Eckstein, M. P. (2006). The effect of nonlinear human visual system components on performance of a channelized Hotelling observer in structured backgrounds. *IEEE Trans Med Imaging*, *25*(10), 1348–1362.

## Appendix

### Generation of ganglion cell mosaic

First, place a ganglion cell at the center of the fovea. The algorithm then creates successive rings of ganglion cells around that location. Begin by defining  $g_{k,n}$  to be the  $k^{\text{th}}$  ganglion cell on ring  $n$ . The ganglion cell at the fovea will be  $g_{1,1}$  (the only ganglion cell on ring 1). Also, define  $C(g_{k,n})$  to be the circle centered at  $g_{k,n}$  with a radius (spacing) specified by the equation in Figure 2C (the radius depends on the retinal location of  $g_{k,n}$ ). Two rules specify how all ganglion cells on ring  $n$  are created given that ring  $n - 1$  has been completed. For each rule, there is a special case in which the fovea is the previously created ring.

Rule 1: The first ganglion cell on ring  $n$ ,  $g_{1,n}$ , is placed at the intersection furthest from the fovea between  $C(g_{k-1,n-1})$  and  $C(g_{k,n-1})$ , where  $k$  is a randomly chosen positive integer at most as large as the total number of ganglion cells on ring  $n - 1$ . The special case in which  $n - 1 = 1$  (the central ganglion cell) is handled by placing  $g_{1,2}$  at any randomly chosen point on  $C(g_{1,1})$ .

Rule 2: The  $k^{\text{th}}$  ganglion cell on ring  $n$ , for  $k > 1$ , is found by first identifying all intersections between  $C(g_{k-1,n})$  and the circles of all ganglion cells on ring  $n - 1$ . In the special case in which  $n - 1 = 1$ , there will be only two such intersections, one clockwise and the other counterclockwise from  $g_{k-1,n}$ . In this case, choose the intersection that is clockwise from  $g_{k-1,n}$  as the location for  $g_{k,n}$ . In the more general case in which  $n - 1 > 1$ , we first find the subset of intersections between  $C(g_{k-1,n})$  and the circles of all ganglion cells on ring  $n - 1$  that lie counterclockwise from  $g_{k-1,n}$ . The location of  $g_{k,n}$  is at the intersection (within this subset) that is furthest from the fovea.

MatLab code for generating the array is available at <http://natural-scenes.cps.utexas.edu/>.

### Ganglion cell receptive fields

The center and surround mechanisms at retinal location  $\mathbf{x}$  are given by:

$$g_c(\mathbf{y}; \mathbf{x}) = \frac{1}{2\pi\sigma_c^2(\mathbf{x})} \exp\left(-0.5 \frac{\|\mathbf{y} - \mathbf{x}\|^2}{\sigma_c^2(\mathbf{x})}\right) \quad (18)$$

$$g_s(\mathbf{y}; \mathbf{x}) = \frac{1}{2\pi\sigma_s^2(\mathbf{x})} \exp\left(-0.5 \frac{\|\mathbf{y} - \mathbf{x}\|^2}{\sigma_s^2(\mathbf{x})}\right) \quad (19)$$

where  $\|\cdot\|$  is the Euclidean norm (vector length).

## Cortical filters

The filtering characteristics of cortical neurons are assumed to be described by the product of a log Gabor function in spatial frequency (Gaussian on a log spatial frequency axis) and a Gaussian function in orientation, in which the log Gabor has a spatial frequency bandwidth at half height,  $b_u$ , of 1.5 octaves and the Gaussian has an orientation bandwidth at half height,  $b_\theta$ , of  $40^\circ$ . The form of the functions is as follows:

$$G_{\log}(u; u_0, b_u) = \exp\left[-\ln(16) \frac{(\log_2 u - \log_2 u_0)^2}{b_u^2}\right] \quad (20)$$

$$G(\theta; \theta_0, b_\theta) = \exp\left[-\ln(16) \frac{(\theta - \theta_0)^2}{b_\theta^2}\right] \quad (21)$$

## Target envelope

The envelope of the target was computed by first finding the parameters of a scaled two-dimensional Gaussian function  $g(\mathbf{y}; \mathbf{u}, \mathbf{\Sigma})$  that best fits the absolute value of the target:

$$\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}, \hat{k} = \operatorname{argmin}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}, k} \sum [|T(\mathbf{y})| - kg(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})]^2 \quad (22)$$

where  $\boldsymbol{\mu}$  is the mean vector,  $\boldsymbol{\Sigma}$  is the covariance matrix, and  $g(\mathbf{y}; \mathbf{u}, \mathbf{\Sigma})$  has a volume of 1.0. To obtain the envelope for a given retinal location, we then blurred this Gaussian with another Gaussian (of volume 1.0) having the size of the ganglion cell center at that retinal location  $\sigma_c(\mathbf{x})$ . Thus the envelope (which also has a volume of 1.0) is given by

$$E_T(\mathbf{y}; \mathbf{x}) = g\left(\mathbf{y}; \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}} + \begin{bmatrix} \sigma_c^2(\mathbf{x}) & 0 \\ 0 & \sigma_c^2(\mathbf{x}) \end{bmatrix}\right) \quad (23)$$