

# Letter identification and the Neural Image Classifier

Andrew B. Watson

NASA Ames Research Center, Moffett Field, CA, USA



Albert J. Ahumada

NASA Ames Research Center, Moffett Field, CA, USA



**Letter identification is an important visual task for both practical and theoretical reasons. To extend and test existing models, we have reviewed published data for contrast sensitivity for letter identification as a function of size and have also collected new data. Contrast sensitivity increases rapidly from the acuity limit but slows and asymptotes at a symbol size of about 1 degree. We recast these data in terms of contrast difference energy: the average of the squared distances between the letter images and the average letter image. In terms of sensitivity to contrast difference energy, and thus visual efficiency, there is a peak around ¼ degree, followed by a marked decline at larger sizes. These results are explained by a Neural Image Classifier model that includes optical filtering and retinal neural filtering, sampling, and noise, followed by an optimal classifier. As letters are enlarged, sensitivity declines because of the increasing size and spacing of the midjet retinal ganglion cell receptive fields in the periphery.**

## Introduction

In this article, we explore and model some aspects of human visual pattern classification. By *pattern*, we mean a fixed spatial distribution of luminance contrast.<sup>1</sup> Experimentally, pattern classification is defined by procedures in which an observer is presented with one of a finite set of candidate patterns and is asked to identify which of the candidates has been presented. The set is typically small, the patterns are typically simple (although that term has no formal definition), and the observer is typically well trained on the set. A canonical example is letter identification.

Our interest in pattern classification arises in part from our belief that it exposes a fundamental step in the complete process of visual information gathering. Just as contrast detection, motion estimation, estimation of disparity, and chromatic discrimination are fundamental operations in low-level vision, so too is the ability to classify simple shapes. Object identification

and scene understanding are much more complex operations, which presumably rely on understanding the relations among many articulated parts, but it is difficult to envision how they could operate without a preliminary identification of those parts.

Pattern classification is also a powerful method with which to study transmission of information within the early stages of the visual system. Because the set of patterns is finite and well defined, it is possible to compute the performance of an ideal observer who uses all available information (Tanner, 1961). Performance of ideal observers with specific early losses of information can also be computed. These model observers can be compared with human performance (Banks, Geisler, & Bennett, 1987; Banks, Sekuler, & Anderson, 1991; Geisler, 1989; Watson, 1987).

Additional reasons for our interest in pattern classification are the many applied questions to which it is central. An obvious example is measurement and interpretation of visual acuity, which relies on classification of letters or other optotypes (Watson & Ahumada, 2012). Many aspects of reading can be understood through an understanding of letter identification (Legge, 2006). Legibility of text and other symbols is likewise dependent on pattern classification, and thus labeling, signage, and visual interface design could all benefit from a better understanding of this process (Castro & Horberry, 2004). Finally, we are interested in quantifying the end-to-end performance of electro-optical visual imaging systems, such as surveillance or other remote viewing systems; predicting human pattern classification through these systems is a promising performance metric (Watson, 2011).

In this report, we focus on one particular aspect of letter identification: the effect of letter size. Size is of interest in many applications, but it is also of theoretical interest because it addresses the spatially anisoplanatic nature of the visual field. As letters become larger, they impinge more on the peripheral visual field, which differs markedly from the fovea in its spatial attributes. We will test whether a model of

Name	Year	ms	$L_0$ (cd/m <sup>2</sup> )	Color	Font	Weight	Symbols	Display	%	Obs	x-height	Eyes	Log $D_c$
Ginsburg	1978	$\infty$	68	Black	Helvetica	Bold	12	Print	50	1	False	2	−0.992
Blommaert and Timmers	1987	64	150	Black	Eurostile	Bold	26	Print	50	2	True	2	−1.725
Legge et al.	1987	$\infty$	300	Black	Sloan	Plain	10	CRT	75	3	False	2	−0.946
Strasburger et al.	1991	100	62	White	Zeile	Plain	10	CRT	67	4	False	2	−1.801
Alexander et al.	1994	249	25.3	Black	Sloan	Plain	10	CRT	71	1	False	1	−1.549
Pelli and Farell	1999	200	50	White	Bookman	Bold	26	CRT	64	2	True	2	−1.154
McAnany and Alexander	2006	35	60	White	Sloan	Plain	10	CRT	80	3	False	1	−2.402
Aparicio et al.	2010	$\infty$	200	Black	Sloan	Plain	10	Print	66	4	False	2	−0.946
Watson et al.	2015	221.6	116	White	Sloan	Plain	10	LCD	75	3	False	2	−1.299

Table 1. Nine studies of contrast threshold for identification of alphanumeric symbols as a function of size. *Notes:* Obs indicates the number of observers; % indicates percentage correct at threshold; x-height indicates whether letter size was defined by the height of a lowercase letter “x”; eyes indicates binocular or monocular viewing; and Log  $D_c$  indicates the log contrast difference energy of the set of symbols at a size of 1 degree and the specified duration (or 500 ms, if duration was indefinite). See text and Appendices 1 and 2 for additional details.

peripheral processing can account for variations in performance with letter size.

We will base our analyses initially on eight previously published studies of contrast thresholds for letter identification as a function of size. We will also collect some similar data of our own, to allow direct comparison with simple contrast sensitivity measurements on the same observers and apparatus.

The model that we will explore is image based. By this, we mean that the input is an actual letter image, rather than some more abstract representation. This constraint obliges us (and competing models) to explicitly represent all stages in the classification process. Our model is also operational: It actually performs letter classifications. This constraint is also important to ensure that we are in fact modeling the relevant behavior. Details of our model will be given below, but in summary, it is a template model limited by noise, by visual optics, and by the size and spacing of midget retinal ganglion cells (mRGC).

A number of image-based operational models of letter identification have been proposed (Beckmann & Legge, 2002; Chung, Legge, & Tjan, 2002; Gold, Bennett, & Sekuler, 1999a; Nestares, Navarro, & Antona, 2003; Parish & Sperling, 1991; Watson & Ahumada, 2008, 2012; Watson & Fitzhugh, 1989). All of these are template models, and indeed we are unaware of any published operational image-based model of letter identification that is not based on templates. Our model shares many similarities with these models, as will be discussed later, but has two novel features. The first is that the optical component is based on a new model for the average observer and can be tuned for a given pupil diameter (Watson, 2013), which can in turn be computed from the display conditions (Watson & Yellott, 2012). The second and more significant feature is eccentricity-dependent spa-

tial filtering and sampling by the mRGC. The local size and spacing of mRGC are computed from a new formula (Watson, 2014).

## Letter identification data

Table 1 lists nine studies in which contrast thresholds were collected for identification of alphanumeric symbols as a function of size (Alexander, Xie, & Derlacki, 1994; Aparicio et al., 2010; Blommaert & Timmers, 1987; Ginsburg, 1978; Legge, Rubin, & Luebker, 1987; McAnany & Alexander, 2006; Pelli, Burns, Farell, & Moore-Page, 2006; Strasburger, Harvey, & Rentschler, 1991). The final study (Watson) refers to data collected for this report. The studies varied in their methods, including exposure duration, symbol color (black or white), the luminance of the background  $L_0$ , the font, the number of symbols, the percentage correct that defined threshold, monocular or binocular viewing, and the type of display. The values of these parameters are provided for each study in Table 1. In all studies, trials were blocked by size, except in the study by Ginsburg (1978), in which an eye chart ordered by size was used.

The data from these studies are summarized in Figure 1a, plotted as contrast sensitivities (inverse contrast thresholds). To make the summary data more legible, we have averaged over observers within each study, using the method described in Appendix 1. There is substantial variability between studies, in part because of the variation in conditions. For example, the lowest sensitivities are for the briefest durations (McAnany & Alexander, 2006) whereas the highest are for the longest durations (Aparicio et al., 2010; Legge et al., 1987). But despite the variations, a general pattern

is evident. The dashed gray line in the figure illustrates the behavior of a simple ideal observer: Sensitivity increases in proportion to size (the vertical position of the curve would be determined by the power spectral density of the noise, but here it is arbitrary). The data, on the other hand, show an initial slope that is much greater than 1, followed by a flattening and a decline beyond about 1 deg.

An alternative representation of these data is in terms of contrast difference energy. An ideal observer classifying a set of  $M$  images  $c_m$  will base its decision on the vector distance between the test image and each candidate image in the set. Performance for the set will depend on the squared distances between members of the set (Watson & Ahumada, 2008). A summary metric representing these differences is the average squared distance between members of the set (Ahumada & Watson, 2013; Dalimier & Dainty, 2008). This is equivalent to the average squared distance between each member and their average (Ahumada & Watson, 2013). Thus, a useful metric for the set is the contrast difference energy

$$D_c = \frac{dx dy dt}{M-1} \sum_{m=1}^M \|c_m - \bar{c}\|^2 \quad (1)$$

where  $\bar{c}$  is the average image,  $dx$  and  $dy$  are the width and height of a pixel, and  $dt$  is the duration, or the integral of the square of the temporal waveform. In this expression, we have omitted the spatial image coordinates  $x$  and  $y$ .

Expressing identification thresholds in terms of contrast difference energy has two advantages: It automatically takes into account differences in duration and font weight and styling, and it displays performance variations relative to an ideal observer. To transform the data, we first computed the “unit” contrast difference energy of the set of symbols used in each study, at a nominal size of 1 deg and a duration appropriate to the study (see Appendix 2 for details and Table 1 for values). The contrast difference energy of each point in Figure 1a can then be determined by multiplying the unit contrast difference energy by the square of the contrast and by the square of the letter size in deg.

A convenient unit for contrast energy or contrast difference energy is dB (Watson & Ahumada, 2005; Watson & Solomon, 1997). This is given by

$$dB(D_c) = 10 \log_{10}(D_c) + 60. \quad (2)$$

This is a decibel measure, adjusted so that 0 dB approximates the minimum visible contrast energy for a sensitive human observer (Watson, Barlow, & Robson, 1983). Thresholds can be expressed as dB, whereas sensitivity can be expressed as  $-dB$ . The data in Figure 1a are presented again in Figure 1b as

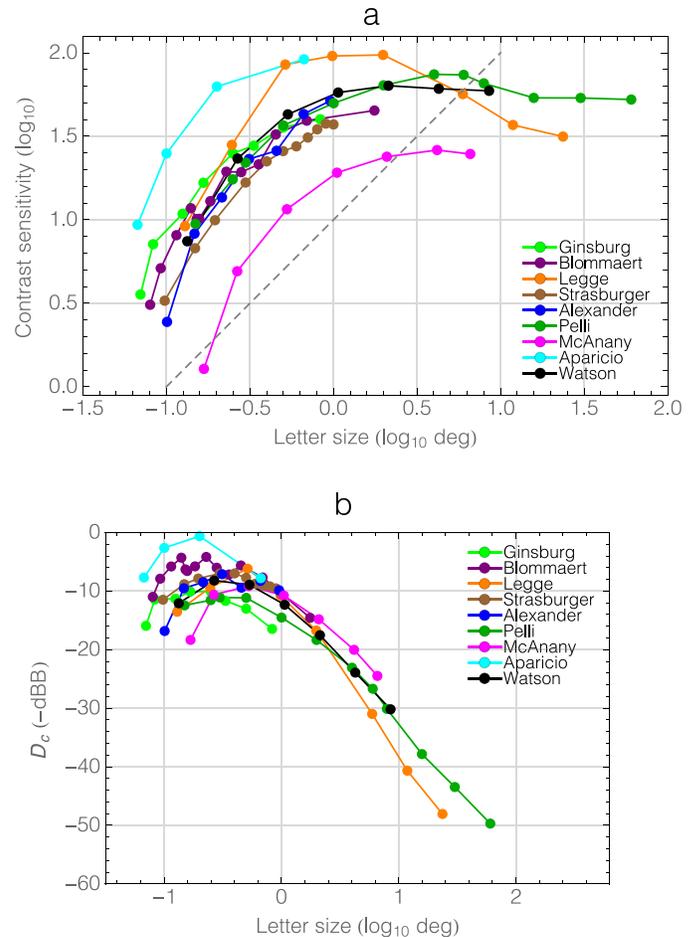


Figure 1. (a) Contrast sensitivities for identification of alphanumeric symbols as a function of size. Average data from nine studies are shown. (b) The same data replotted as contrast difference energies.

contrast difference energy sensitivities in  $-dB$ . Note that the scatter of the data has been reduced considerably. With respect to the remaining variation, we note that even after compensating for duration and contrast energy, the studies differed in specific observers, font complexity, percentage correct, number of symbols, background luminance, pupil diameter, and contrast polarity. All of these are likely to have some effect on contrast sensitivity for identification. The largest variations are for the smallest symbols, where variations in acuity will have their effect.

Rather than attempting to fit the data for one study or one observer, we will fit the ensemble of data. In the case of data collected in our own lab, where all relevant experimental parameters are known, we will fit results for individual observers.

For an ideal observer, limited only by the inherent noise in the signal, contrast energy sensitivity would be constant when expressed in dB. Considering the data in Figure 1b, we note that the sensitivities first rise with increasing size, reaching a peak at a size around  $1/4^\circ$ ,

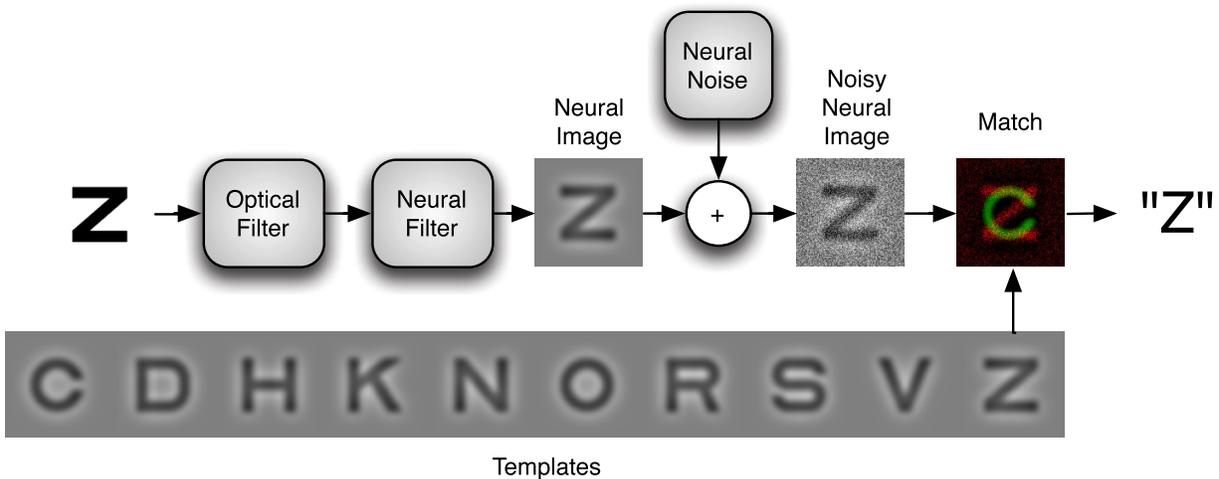


Figure 2. Neural Image Classifier. In this example, we show identification of Sloan letters.

and then fall continuously at larger sizes. We will argue below that the initial rise is due to the escape from optical and neural blur, whereas the later decline is due to the decline in resolution of the periphery.

## Neural Image Classifier

The model that we will consider to explain these data is illustrated in Figure 2. It is an extension of a template model that we have used previously to explain letter recognition limited by optical and neural factors (Watson & Ahumada, 2008, 2012). We call it the Neural Image Classifier (NIC) because it is an optimal pattern classifier (Duda & Hart, 1973) that operates at the level of the neural image (Robson, 1980), specifically the neural image defined by the output of the mRGC. A summary of notation for the model is provided in Appendix 9.

A letter of a certain size is presented to the observer. It passes through optical and neural filters to yield the neural image. Gaussian white neural noise is added to produce a noisy neural image, which is then matched in turn to each of a fixed number of templates. The templates are the neural images of the candidate letters of a fixed size. The template with the highest cross-correlation is reported as the letter identified.

Elsewhere we have developed mathematical techniques to allow efficient Monte Carlo simulation of this model (Watson & Ahumada, 2008, 2012). In this report, we describe elaborations of the model to include (a) an optical filter derived from a population of observers or from an individual eye, (b) space-variant filtering by mRGC, (c) space-variant noise that results from varying density of mRGC receptive fields, and finally (d) variations in identification efficiency as a function of target size. The optical and neural

processing are similar to those in a model recently proposed by Bradley, Abrams, and Geisler (2014), discussed at greater length in the Discussion section. We also extend the model to deal with binocular viewing, in which the two eyes may have different optical properties. The details of this model and its extensions will be described in the sections below. Details of the model implementation are provided in Appendix 3.

### Optical filter

Each image is first blurred by a filter that simulates the action of the human visual optics. This filter was computed from a formula recently proposed to describe the polychromatic (white light) mean optical modulation transfer function (MTF) for a large population of healthy, well-corrected human eyes (Watson, 2013). The formula depends on pupil size. The MTF for a pupil size of 5 mm is shown in Figure 3. When available, the actual optical filter of a specific individual eye was used in place of the population average formula.

This filter does not include subsequent optical filtering by the waveguide properties of the cone inner segments. That is instead presumed to be incorporated in the neural filter, described next.

### Neural filter

The next element of the model is a filter that simulates the linear receptive field of mRGC in the human retina. However, unlike the standard definition of the receptive field, this operates on the retinal image, after blurring by the eye's optics. The mRGC receptive field has long been modeled as a difference of

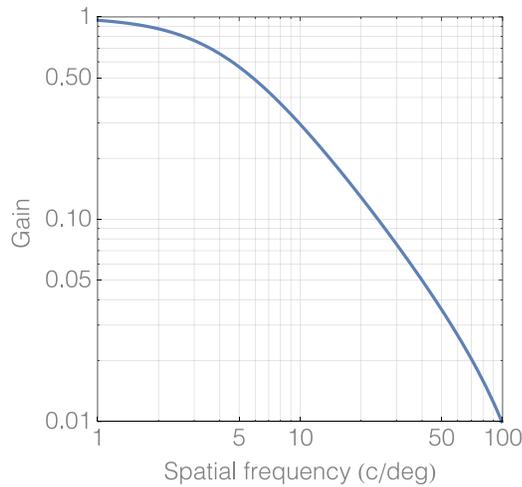


Figure 3. MTF of the optical filter at pupil diameter 5 mm.

Gaussians (Enroth-Cugell & Robson, 1966; Enroth-Cugell, Robson, Schweitzer-Tong, & Watson, 1983; Rodieck, 1965), and we adopt this model as well, based in part on the fitting exercise described below. To describe these mechanisms, we introduce the following Gaussian kernel function,

$$k_{gauss}(r, s) = \frac{1}{s^2} \exp\left(-\pi\left(\frac{r}{s}\right)^2\right) \quad (3)$$

Here,  $r$  indicates distance from the receptive field center in degrees and  $s$  is a scale factor, also in degrees. The function has a normalizing factor to ensure that the volume under the function is 1, which in turn ensures that when transformed into a transfer function in the frequency domain, it will have a peak gain of 1 at 0 cycles/degree.

The mRGC receptive field is then modeled as the difference between a narrow center Gaussian kernel and a broader surround Gaussian kernel.

$$k_{mRGC}(r, s_c, s_s, a) = k_{gauss}(r, s_c) - ak_{gauss}(r, s_s) \quad (4)$$

where the parameters are the center scale  $s_c$ , the surround scale  $s_s$ , and the surround attenuation  $a$ . Because of the normalization of the individual kernels,  $a$  is also the ratio of volumes of center and surround and, in the frequency domain, the gain at 0 cycles/degree. We have explored the use of two other kernel types for the center mechanisms (exp and sech), as described in Appendix 4. All three kernels worked well, with Gaussian and sech about equal. Because of historical precedent and mathematical simplicity, we have selected the Gaussian.

The preceding equation defines the shape of the mRGC receptive field at the center of the visual field (the *visual center*). At eccentric locations, we assume that the receptive fields are scaled in proportion to their spacing. This provides constant mRGC coverage (the

product of receptive field diameter and spacing) throughout the retina, as found by Dacey (1993). Thus, at a location  $\mathbf{x}$  in the visual field, the mRGC receptive field would be

$$k_{mRGC}(r, s_c, s_s, a, \mathbf{x}) = k_{gauss}(r, \delta(\mathbf{x})s_c) - ak_{gauss}(r, \delta(\mathbf{x})s_s) \quad (5)$$

where  $r$  now indicates distance from  $\mathbf{x}$ , and  $\delta(\mathbf{x})$  is the local scale, which we equate to the mRGC receptive field spacing at location  $x$ , relative to that at the visual center  $\mathbf{x} = \{0,0\}$ . Below, we describe a formula for the mRGC receptive field spacing as a function of location in the visual field.

In our model, we consider only the midget retinal ganglion cells and ignore the other classes of ganglion cell, notably the parasol cells that contribute to the magnocellular pathway. We do so because we believe that the midget cells underlie the spatial pattern vision for stationary or slowly changing patterns. The midget cells are the most numerous class: They comprise as much as 90% foveally and perhaps 40% peripherally of all ganglion cells (Dacey, 1993; Drasdo, Millican, Katholi, & Curcio, 2007). This suits them for tasks demanding high spatial resolution. Furthermore, destruction of the midget cells in primates reduces contrast sensitivity by almost a factor of 10 at both high and low spatial frequencies, provided that the temporal frequency is low (Merigan & Eskin, 1986).

We also employ only one size of retinal ganglion cell at each eccentricity. In empirical measurements of ganglion cell receptive fields or dendritic fields, there is some scatter at one eccentricity (Dacey, 1993), but it is unclear whether this is functional or fortuitous. The scatter is quite small at small eccentricities, and essentially zero at the fovea, where each mRGC (center) is driven by only a single cone.

## Local scale

Elsewhere, we have developed a formula to specify the spacing of mRGC receptive fields as a function of location in the visual field (Watson, 2014). The formula is based on anatomical estimates of cone and ganglion cell densities along the four principal meridians (Curcio & Allen, 1990; Curcio, Sloan, Kalina, & Hendrickson, 1990), an assumption that each cone drives exactly two mRGC at the fovea, and estimates of the proportion of ganglion cells that are midget ganglion cells as a function of eccentricity (Drasdo et al., 2007). We have extended the formula to arbitrary retinal locations and to arbitrary locations in the binocular visual field. Density computed from the formula is illustrated in Figure 4a. From a peak density of 29,609 cells degree<sup>-2</sup>,

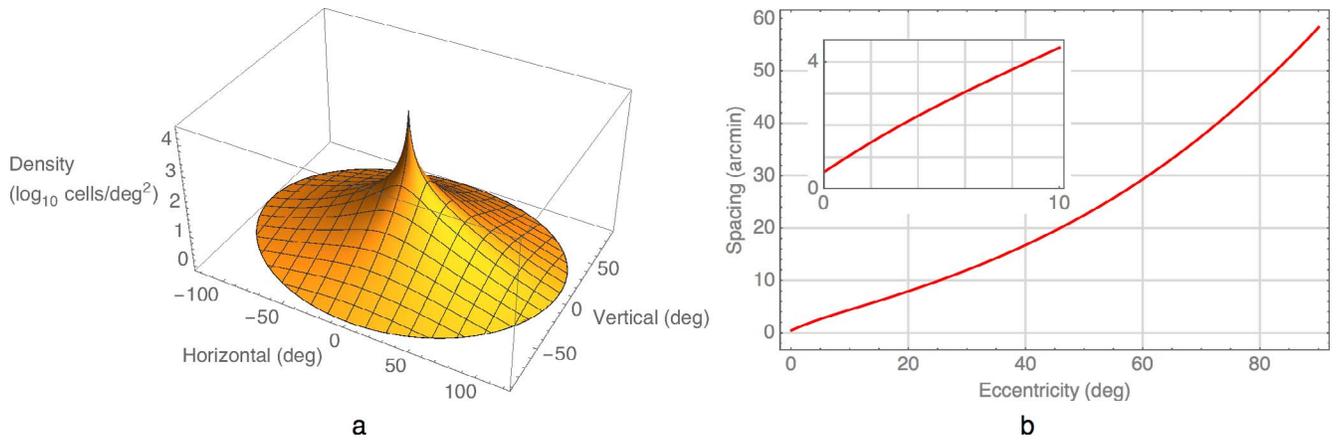


Figure 4. (a) Midget retinal ganglion cell density over the binocular visual field, computed from the formula of Watson (2014). (b) Midget retinal ganglion cell spacing in the on- or off-center lattice along the horizontal meridian of the binocular visual field. The inset shows the central 10 degrees.

it declines by more than four log units at an eccentricity of 90 deg.

The mRGC population consists of overlapping on- and off-center lattices. The calculated density includes both lattices. Assuming approximately hexagonal packing of the cells and equal numbers of on- and off-center cells, spacing within either the on-center or off-center lattice will be  $2 \cdot 3^{-1/4} d^{-1/2}$ , where  $d$  is density in cells degree<sup>-2</sup>. This spacing is plotted in Figure 4b for the horizontal meridian of the binocular visual field. Because the local scale function  $\delta(\mathbf{x})$  is equal to the relative spacing, we can write

$$\delta(\mathbf{x}) = \sqrt{d(0)/d(\mathbf{x})} \quad (6)$$

### Space-variant filtering

To apply the space-variant filtering due to retinal ganglion cells, we have implemented and extended the methods developed by Perry and Geisler (2002). The essence of their method is that a spatially varying filter can be approximated by a spatially varying linear combination of spatially invariant filtered images. Specifically, we first compute a local scale image of the same dimensions as the image to be filtered, using the local scale function  $\delta(\mathbf{x})$  described above in Equation 6. We then compute a set of filtered images in which the spatial scale of the filter increases by some factor (e.g., 2) between each member of the set. For each pixel in the output image, we first determine the value of the corresponding pixel in the local scale image. We then determine the pair of filtered images whose scales enclose that value. The pixel in the output image is a linear combination of the corresponding pixels in those two images, weighted by where the local scale lies in the interval between their scales.

Our extensions allow an arbitrary filter kernel (e.g., exponential and hyperbolic secant) and arbitrary increments in scale between the spatially invariant images. We have used this method to effectively convolve the space-variant receptive field of Equation 5 with each letter image. In general, we have used scale increments of factors of two.

### Noise and efficiency

After optical and neural filtering, the image is perturbed by additive Gaussian noise. We assume that at the photopic levels and relatively long durations we are considering, the noise is dominated by a constant output noise of the mRGC. If the noise is constant for each cell, then the variance per unit area is inversely proportional to the cells per unit area. In that case, the standard deviation of the equivalent noise in the image domain will increase in proportion to the spacing of the mRGC. To simulate this would require the addition of noise fields in which the standard deviation increased with eccentricity. To simplify our computations, we have instead used space-invariant noise and attenuated the image contrast in proportion to the inverse of spacing. This approximation relies on the notion that performance depends essentially on the ratio between signal and noise. In the remainder of this article, we will call this *eccentric attenuation*. It is implemented by dividing the filtered target image by the local scale function  $\delta(\mathbf{x})$ .

The magnitude of the noise is a model parameter and is characterized by the power spectral density  $N$  degree<sup>-2</sup> s, given by

$$N = \sigma^2 dx dy dt \quad (7)$$

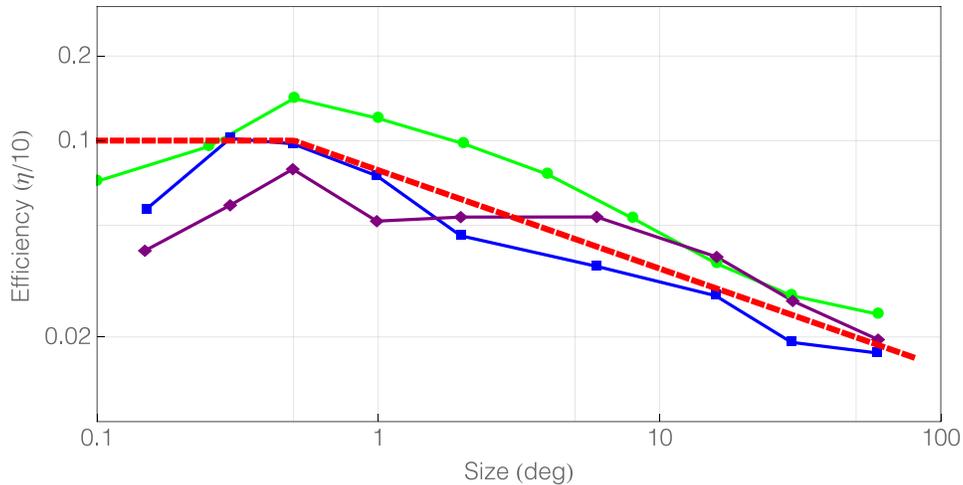


Figure 5. Efficiency versus size. Points are letter identification data (Pelli et al., 2006). The red dashed curve is the relative efficiency of the model (Equation 8) divided by 10.

where  $\sigma^2$  is the noise pixel variance,  $dx$  and  $dy$  are the width and height of a pixel in degrees, and  $dt$  is the duration of the noise sample in seconds.

In the template model, performance is limited by early filtering, by the power spectral density of the noise and by the efficiency  $\eta$ . Efficiency is the ratio of the ideal threshold contrast energy to the actual threshold contrast energy (Pelli, 1990). A value less than 1 indicates that the calculation performed by the brain is less than ideal. One source of inefficiency is that the templates may not be perfect copies of the neural images. Other possible sources are fixational eye movements that require the brain to integrate information over multiple locations during the exposure duration, spatial uncertainty, and inefficient combination of features (Pelli et al., 2006). Because we are not directly measuring the neural noise, we cannot estimate efficiency, and our estimate of  $N$  is effectively the ratio  $N/\eta$ . We will use this observation to interpret differences in estimated  $N/\eta$  to in fact be differences in  $\eta$ .

Pelli has shown that efficiency declines with target size (Pelli & Farell, 1999). His data are reproduced in Figure 5, along with a simple function we have used to approximate this effect. Our function for relative efficiency as a function of size  $S$  has a value of 1 up to  $0.5^\circ$  and then declines with a log-log slope of  $-0.35$ , as shown in Equation 8.

$$\eta' = \text{Max}\left(1, \frac{S}{0.5}\right)^{-0.35} \quad (8)$$

Although we have no operational mechanisms for this variation in efficiency, we include it in our predictions because our targets vary in size. In our simulations, we have estimated size by the largest diameter of the convex hull (Feret diameter) of the binarized image of the target.

## Estimating NIC parameters

The NIC model has four parameters. Three parameters are properties of the mRGC: the center scale  $s_c$ , the surround scale  $s_s$ , and the surround attenuation  $a$  (Equation 4). The fourth parameter is the power spectral density  $N$ . Here, we estimate values for the parameters by fitting the complete model to a subset of data from the ModelFest experiment (Carney et al., 2000; Watson & Ahumada, 2005).

The ModelFest data consist of luminance contrast thresholds for 43 spatial contrast patterns, each defined within a  $2.13 \times 2.13$  deg square, presented within a larger uniform background. The contrast of each stimulus varied as a Gaussian function of time with a standard deviation of  $1/8$  s. When computing contrast energy, this is equivalent to a duration of 222 ms. We use these data because they include a range of spatial frequencies and sizes, they are freely available, and they were collected with relatively well-defined and modern methods from a sizable population of observers (16). Here, we use only the average thresholds for patterns 1 to 14, which consist of Gabor functions of several sizes and spatial frequencies, and patterns 26 to 29, which consist of Gaussians of various sizes.

It is important to note that the NIC model can predict both identification and detection results. Predictions for two-alternative forced-choice detection by the NIC model are particularly simple. In the ModelFest project, thresholds were defined as 84% correct in the 2AFC task, which corresponds to  $d' = 2$ . At that threshold, the contrast energy  $E_V$  of the neural image is given by

$$E_V = 2N \quad (9)$$

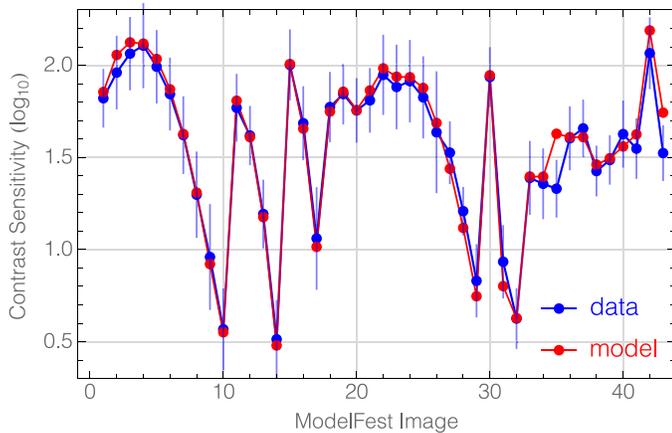


Figure 6. Fit of the NIC model to ModelFest data. Parameters were estimated based on thresholds for images 1 to 14 and 26 to 29.

where  $N$  is the noise power spectral density (Ahumada & Watson, 2013).

We varied the three mRGC parameters of the model and the quantity  $N$  to minimize the RMS error between the logs of data and predictions. We have estimated the pupil diameter for the ModelFest experiment to be 5 mm (Appendix 5), so we used that value in the optical filter.

In Figure 6, we show the best fit at 5-mm pupil. Although we fit to only a subset of the patterns (1–14, 26–29), predictions are quite good for the entire set. The estimated parameters are  $s_c = 1.3$  arcmin,  $s_s = 9.137$  arcmin ( $s_s/s_c = 7.031$ ),  $a = 0.878$ , and  $\log_{10}N = -6.44$ . The error of fit is  $\text{RMS} = 0.295$ . A range of parameters give similar results, so alternative values may be adopted in the future based on additional psychophysical or physiological data. In Appendix 4, we describe supplementary fits with alternative pupil sizes and center kernels.

## Predicting letter identification

We generated predictions for the Sloan font, using the ModelFest equivalent duration of 222 ms and a 4-mm pupil. Although we used a 5-mm pupil for the parameters estimated above, in the letter identification simulations, we used a pupil diameter of 4 mm, because that was a better estimate of the pupil size used in the letter identification studies (Appendix 6). Details of the implementation of the model, and simulations of letter identification, are given in Appendix 3.

Because the model consists of a sequence of processing stages, we can conduct the simulations with each stage added in turn to illustrate its separate effect. In Figure 7, we show the ensemble data, along with the model, as individual stages are added. The first panel

(a) shows the result with no filtering and homogeneous sampling. This flat line (a constant contrast energy threshold) is the behavior of an ideal observer, limited only by a particular amount of noise. In these simulations, we set the noise PSD  $N$  to the value estimated from the ModelFest data. This value will be discussed in greater detail below.

In Figure 7b, we introduce the optical filtering. As expected, it reduces sensitivities for the smallest letters but has no effect on very large letters. Introduction of the neural filtering by the mRGC center (Figure 7c) causes further attenuation for the small letters, because it introduces additional blur, but has no effect on the largest letters, even though the neural blur increases with eccentricity. The identification of the largest letters is presumably carried by very low spatial frequencies, which are not attenuated by the center mechanism. However, introduction of the surround mechanism (Figure 7d) does produce substantial attenuation ( $\sim 10$  dB) for the largest letters. This subtraction of the surround signal does attenuate the very low frequencies that underlie identification of the largest letters.

As noted earlier, the effective local noise (power spectral density,  $N$ ) in the model is inversely proportional to the density of the mRGC receptive fields. We simulate this with fixed noise and eccentric attenuation by the local scale image (see the Noise and efficiency section). This effect is added in Figure 7e, which has an even more substantial effect at the larger letter sizes because they impinge on more sparsely sampled peripheral retina.

In Figure 7f, we introduce the size-dependent efficiency of Equation 8. This further attenuates sensitivity at larger sizes and brings the predictions close to the data ensemble. The discrepancies that remain appear to be that predictions are too low at the smallest letter sizes.

Because of the variability among studies, we have not attempted to fit precisely any one set of data. But it is worthwhile considering what changes to the predictions would be produced by changes in the model parameters. Improvements in optical quality, either through a smaller pupil or fewer aberrations, would elevate relative sensitivity to small letters. It is known that the average ModelFest observer is of somewhat low acuity. Thus, the estimate of the mRGC center scale may be too large to simulate “average” behavior. Like better optical quality, a smaller center scale would also raise relative sensitivity at small sizes.

## New data

In the predictions above, we have estimated parameters from detection data obtained from 16 observers in six different labs (the ModelFest study) and used them

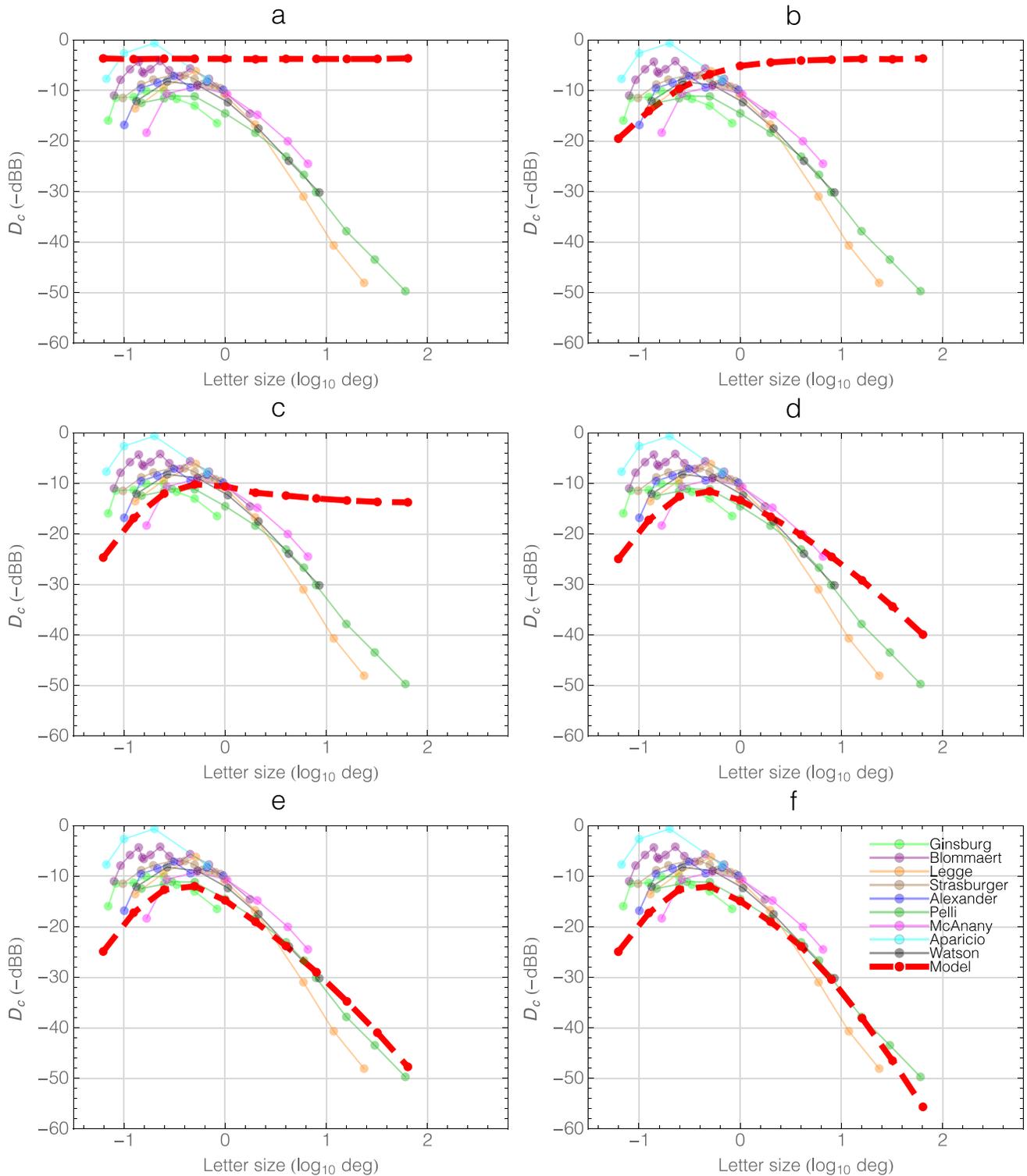


Figure 7. Letter identification data and model predictions. In each panel, an additional stage is added: (a) ideal observer, (b) optical filter, (c) mRGC center, (d) mRGC surround, (e) mRGC noise, (f) size-dependent efficiency.

to predict letter identification data from an ensemble of 23 different observers in nine different labs. The predictions also made use of average optical transfer function derived from 200 eyes, none of which participated in the detection or identification studies.

Given the opportunities for discrepancies based on age, procedures, and individuals, the accuracy of the predictions is encouraging. However, our conclusions would be more compelling if optical, detection, and identification data were obtained in a single lab from a

single set of observers. This was accomplished for three observers as described in the following sections. Additional details are provided in Appendix 7.

## Observers

Data were collected from three observers, C. V. R., P. M. Z., and L. R. W., with ages of 37, 22, and 25 years. Observer P. M. Z. wore glasses, C. V. R. and L. R. W. did not. C. V. R. is emmetropic, whereas L. R. W. ordinarily wears glasses with a prescription of OD 0.5,  $-0.5$  at  $103^\circ$ , OS 0,  $-0.75$  at  $66^\circ$ . The observers were naive as to the purposes of the experiment. Experimental protocols were approved by the NASA Human Research Institutional Review Board. Subjects gave informed consent before testing. All research conformed to the Declaration of Helsinki.

## Optical measurements

Wavefront aberrations for a 4-mm pupil were measured from both eyes of each observer, using the same optical correction (eyeglasses or not) used in psychophysical data collection. Details of optical measurements are provided in Appendix 8. We selected the 4-mm pupil based on our estimate of pupil diameter for the conditions of our experiment (Appendix 6). We zeroed the defocus components, assuming that the observers were accommodated to our display. The remaining Zernike coefficients were used to compute the point spread function (PSF) for each eye of each observer, as shown in Figure 8. These PSFs were used to filter the target images separately for the two eyes.

In Figure 9, we have computed the radial MTF for each eye for each observer. This provides a convenient summary for the optical performance of each eye and also allows us to compare the results to the formula used previously in this article to compute the MTF for an average observer (Watson, 2013). That formula is shown by the red curve in Figure 9. For observer C. V. R., one eye is well below the formula, whereas the other is slightly above. For observer P. M. Z., both eyes are near but slightly below the formula. For the observer who did not wear their usual spectacle correction (L. R. W.), the MTFs for both eyes are well below the formula. These results give us additional confidence in the formula as a useful benchmark and substitute when optical data are unavailable.

## Gabor and Gaussian detection

We measured contrast thresholds for 14 of the ModelFest stimuli, consisting of 10 Gabor functions of

constant size (standard deviation of 0.5 deg) and increasing spatial frequency and one Gaussian of the same size. These correspond to ModelFest stimuli with index numbers 1 to 10 and 26 (Watson & Ahumada, 2005). The logic for this selection is that the Gabor sensitivities characterize the center mechanism, whereas the Gaussian characterizes the balance. The data for the three observers are shown in Figure 10.

## Estimating NIC parameters

The binocular model was applied to the 14 ModelFest stimuli, and parameters were optimized to minimize the error between predictions and data for each observer. Data and model are shown for reach observer in Figure 10. The estimated parameters are shown in Table 2.

For two observers, the center scale  $s_c$  is about 0.4 arcmin. For observer L. R. W.,  $s_c$  is more than twice as large. This may be due to residual defocus in this observer, who did not wear her spectacle correction during psychophysical testing. It may also be a genuine variation in center scale, because foveal cone density, and thus mRGC density, varies substantially among observers. For example, Curcio et al. (1990) found foveal cone density differences as large as 3.3, which would be consistent with spacing and thus scale differences as large as 1.8. Sekiguchi, Williams, and Brainard (1993) estimated a neural PSF with a full-width at half-height of 0.71 arcmin, which corresponds to a center scale of 0.75 arcmin, close to that of L. R. W.

For technical reasons, we constrained the scale ratio  $s_s/s_c$  to below 20. For two observers, this was the value returned. The value was not well constrained by the data. All three observers' estimates of  $a$  were a little greater than 0.9 and of  $\log N$  were a little less than  $-6.2$ .

The most salient differences between the ModelFest mean observer (MFM) and the three new observers is in the value of the center scale  $s_c$ . The MFM value is about 1.5 times as large as that for L. R. W. (who wore no glasses) and about 3.3 times as large as for C. V. R. and P. M. Z. One explanation is that the ModelFest observers were on average older. It is also likely that some blur was introduced by the displays used. It has been previously remarked that the MFM has rather poor acuity (Ahumada & Scharff, 2007; Watson & Ahumada, 2008). The other notable difference is the smaller value of  $a$ . This may be the result of artifactually high ModelFest sensitivity to the Gaussian stimulus (Ahumada & Scharff, 2007).

## Letter identification

Contrast thresholds for letter identification were collected from the same three observers on the same

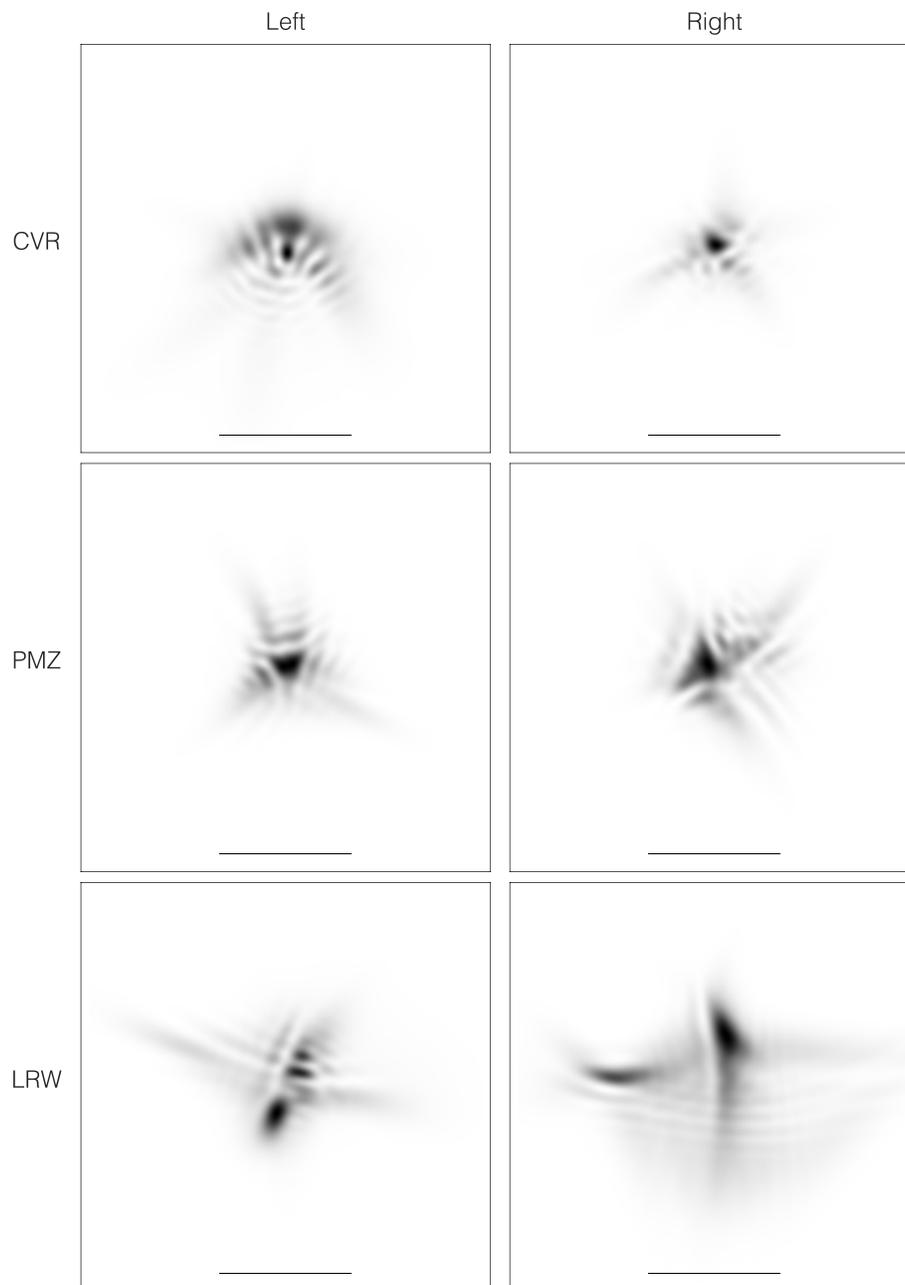


Figure 8. Point spread functions for three observers. The horizontal line in each panel is 5 arcmin in length.

experimental apparatus (Appendix 7). The results are shown in Figure 10. They are very similar to the historical data, as may be seen by their mean, plotted as “Watson” in Figure 1.

### Extending the model to two eyes

Our previous simulations treated the observer as possessing a single cyclopean eye. Here, we have applied distinct optical filters for each eye to each stimulus. In addition, we have applied appropriate

monocular mRGC filtering for each eye. This will differ for the two eyes because of the nasal-temporal asymmetry of the mRGC density (Curcio & Allen, 1990; Watson, 2014). We have then concatenated the left- and right-eye images to form a single binocular neural image. This image now serves as the input to the classification process. Both templates and sample images are now binocular. The large literature on binocular combination is outside the scope of this article, but we note that this method will yield the frequently observed quadratic combination of contrasts (Legge, 1984).

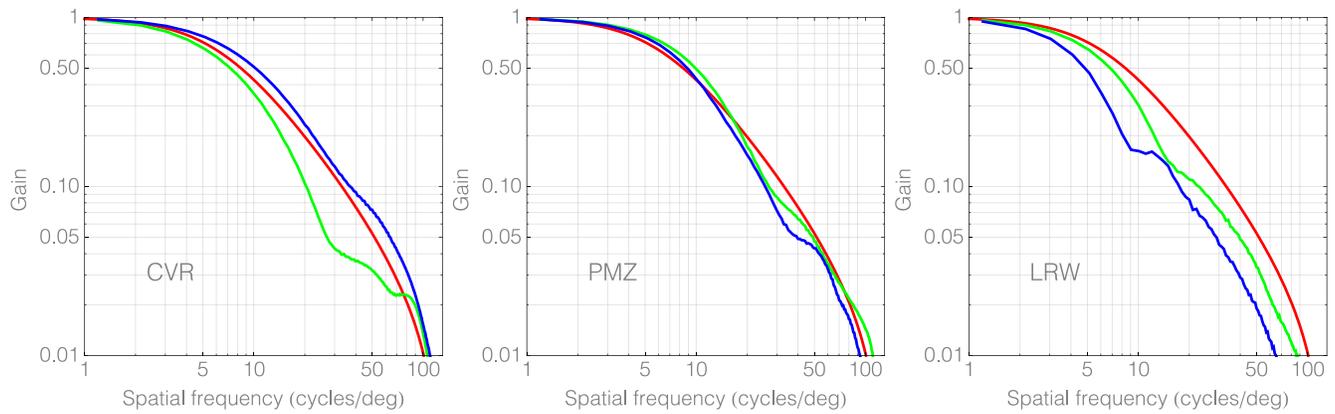


Figure 9. Radial MTFs for three observers. Green and blue curves are for left and right eyes, respectively. The red curve is the formula of Watson (2013) for a pupil diameter of 4 mm.

### Model predictions

Using the binocular version of the NIC model, and the parameters estimated from detection of Gabor functions, we have computed predictions for contrast thresholds for letter identification as a function of size. These are shown along with the experimental data in Figure 11.

For one observer (L. R. W.), the predictions are quite accurate. Note that there are no free parameters in this prediction and that the estimates of model parameters obtained from Gabor thresholds are for targets about  $1^\circ$  in size (ModelFest stimuli), whereas the data predicted here extend to letter sizes as small as  $1/8^\circ$  and as large as  $8^\circ$ .

For the other two observers (C. V. R. and P. M. Z.), the predictions are accurate at large letter sizes but too low at small letter sizes. For observer C. V. R., at the smallest letter sizes, the predicted sensitivity is about a factor of 2 (6 dBB) lower than the measured sensitivity.

For observer P. M. Z., the discrepancy is similar but smaller. This pattern of results resembles a tendency evident in Figure 7f; however, predictions here are based on model estimates obtained for individual observers. We now consider possible reasons for this discrepancy for two of the observers.

### Efficiency

As noted above, an undetermined parameter in these simulations is the efficiency of the pattern matching process. We have effectively combined (multiplicatively) the efficiency into our estimate of noise power  $N$ . One interpretation of the discrepancy in Figure 10 is that (a) overall efficiency for letter identification is higher than that for Gabor detection and (b) efficiency declines more rapidly with eccentricity than our model assumes. At least the first part of this explanation seems unlikely, because it has been argued that Gabor

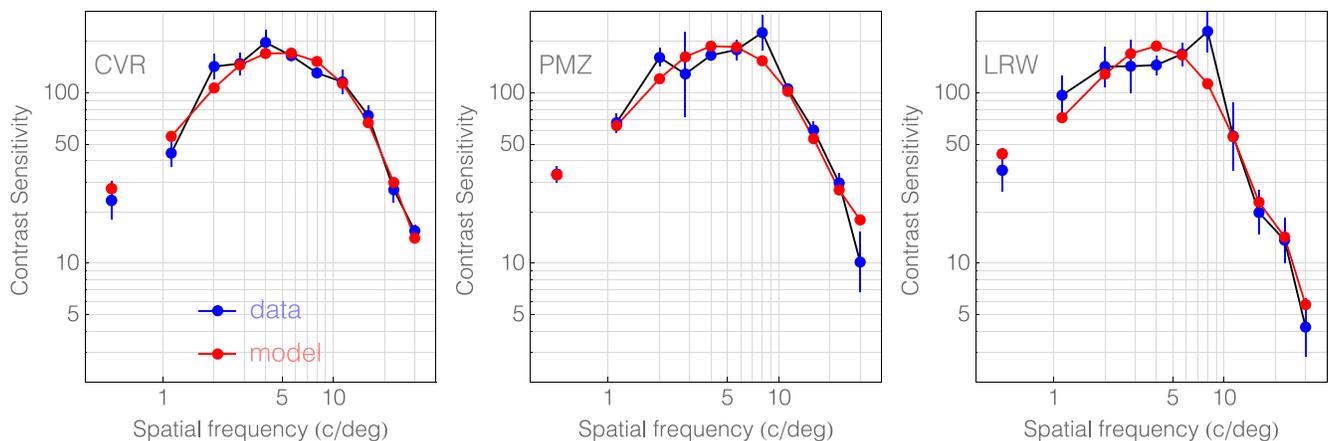


Figure 10. Contrast sensitivity for three observers and best-fitting versions of the binocular NIC model with measured PSF for each eye. The Gaussian target (index 26) is plotted as a separate point at the left.

Observer	$s_c$	$s_s/s_c$	$a$	$\log_{10}N$	RMS
C. V. R.	0.381	20	0.944	-6.23	1.23
P. M. Z.	0.404	20	0.933	-6.25	1.3
L. R. W.	0.862	8.62	0.921	-6.41	1.18
MFM	1.3	7.03	0.878	-6.44	0.295

Table 2. Estimated NIC model parameters for three observers. *Note:* For comparison, we also include the ModelFest mean (MFM) observer estimated earlier.

detection may have an efficiency as high as 0.2 (D. G. Pelli, personal communication, June 26, 2014), whereas efficiency for letter identification (for Sloan letters) has been estimated at about 0.1 (Pelli & Farell, 1999). Pelli has also shown that efficiency can vary strongly with the complexity of the targets (Pelli et al., 2006). For this reason, it should be acknowledged that the near correspondence of our data and predictions is in part a coincidence; had we chosen a very different font with a very different efficiency, the separation between data and predictions would presumably have been larger.

## Position uncertainty

The NIC model assumes that the observer knows the exact location of the target and can apply each template at the appropriate position in the neural image. When the observer is uncertain about the position, efficiency will decline relative to the ideal classifier (Michel & Geisler, 2011). This is undoubtedly one reason why human empirical efficiencies are less than 1. However, identification may be less uncertain than detection. The reason is that identification is based on the differences between targets. The parts they share do not contribute to the classification. But the parts they share may serve as a reference to reduce uncertainty.

In this same vein, the contrast threshold for identification is typically higher than that for detection. But detection is sufficient to reduce spatial uncertainty.

Thus, at threshold for identification, there may be sufficient contrast to reduce spatial uncertainty. The spatial uncertainty explanation for the discrepancy might also explain why it is diminished at the larger sizes, where spatial uncertainty plays a smaller role. Note that spatial uncertainty is one possible source of inefficiency.

## Cortical noise

One aspect of the discrepancy is that empirical sensitivity declines more rapidly with size than predicted. Note that in these predictions, we have estimated parameters by fitting detection data for small ( $\sim 1$  degree) foveal targets and combined that with a model for the density of mRGC as a function of eccentricity to predict identification of large targets, extending well into the periphery. We have assumed that the predominant noise is at the level of the neural image, which means effectively at the output of the mRGC. Sensitivity declines for large targets in part because of the effective increase in noise in the periphery, due to the sparsity of the mRGC (Figure 7e). If additional noise is introduced at higher (e.g., cortical) levels, which also show declining density with eccentricity (Schira, Wade, & Tyler, 2007), then a more rapid decline with eccentricity would be expected.

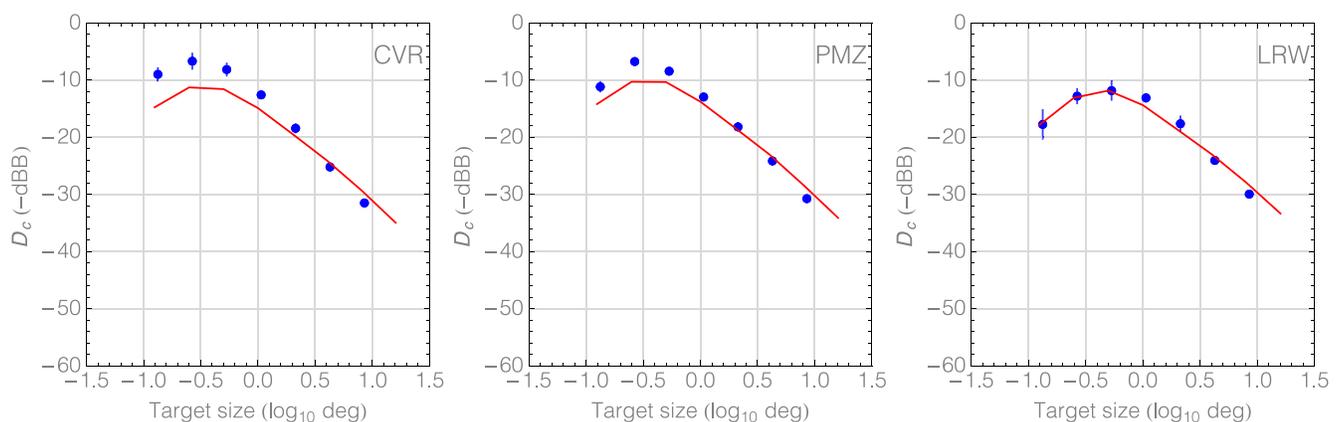


Figure 11. Contrast difference energy sensitivities for letter identification as a function of size for three observers. The blue points are data. The red curve is the prediction of the NIC model.

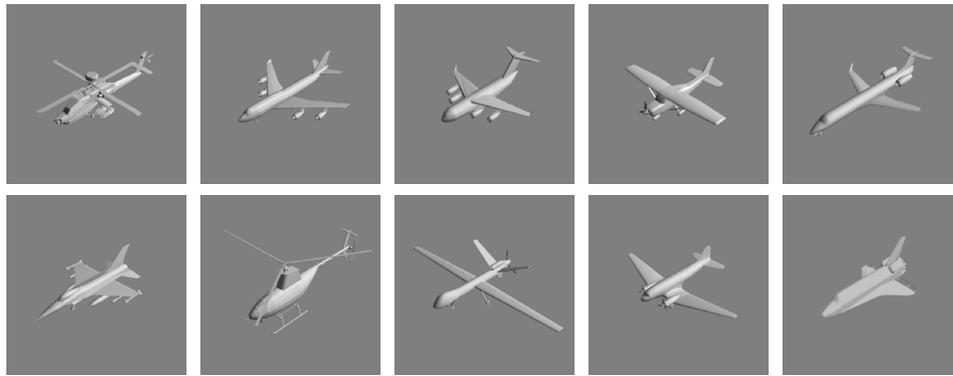


Figure 12. Aircraft images used in the identification experiment.

### Aliasing in the periphery

Another reason why sensitivity might decline more rapidly with eccentricity than predicted is that in the periphery sampling density is well below the optical limit that would prevent aliasing. Thibos, Still, and Bradley (1996) have provided compelling empirical evidence for peripheral aliasing. We hope to include aliasing in a future version of the NIC model. One puzzle is why peripheral aliasing does not result in a clear decline in efficiency with eccentricity (Pelli & Farell, 1999), although that result extends only to 5 deg eccentricity.

### Identification of aircraft

To evaluate the generality of the NIC model, we have conducted an additional experiment in which our three observers identified images of aircraft, as shown in Figure 11. We selected 10 images to match the number of Sloan letters in the letter identification experiment. These images differed from the letter images in several ways. First, they were grayscale images, rather than binary images, with portions both brighter than and darker than the background. Second, they were arguably more complex than the letter images. Third, they were presumably less familiar to our observers than the letter images.

The contrast threshold for aircraft identification was collected from the same three observers who participated in the letter identification experiment. The images were created from 3D graphics models using methods described previously (Watson, Ramirez, & Salud, 2009). The images were equated for size (total number of nonbackground pixels) and contrast energy. We used six image sizes from 128 to 2,048 pixels (1.07 to 17.07 deg), in steps of a factor of 2. These limits were imposed by our display and resolution but still encompassed five octaves of size. The target occupied approximately half the width or height of the image, as shown in Figure 12. Other details of the methods and

procedures were identical to those for the letter identification experiment (Appendix 7).

Results are shown in Figure 13. Similar to the case of letter identification, energy sensitivity rises to a peak, here about 1 degree, and declines at larger sizes. The results for the three observers are quite similar to one another. In each panel, we also show the customized predictions for each observer, based on the optical parameters of each eye and retinal parameters fit to the Gabor data. One modification was made to the implicit model parameters: to match (by eye) the vertical position of the data, the efficiency was reduced by a factor of three relative to that for detection of Gabors.

These results show that the NIC model works well for targets more complex than letter images but also suggests that absolute predictions will depend on efficiency, which in turn may depend on the complexity of the targets (Pelli et al., 2006). It has also been shown that efficiency depends on the familiarity and degree of learning of the targets (Doshier & Lu, 1998; Gold, Bennett, & Sekuler, 1999b; Pelli et al., 2006), and our letters are undoubtedly more familiar than our aircraft images.

### Discussion

#### Summary

We have described a model of human visual pattern classification whose primary elements are optical filtering, filtering by mRGC receptive fields (mRGCf), variation in mRGCf size and density with eccentricity, noise at the ganglion cell output, and ideal pattern classification. In the first part of this article, we estimated parameters for this model from detection thresholds for Gabor targets extracted from the ModelFest data set. We then used the parameterized model to predict a large collection of historical data sets of contrast thresholds for letter identification as a function of size. The predictions were reasonably

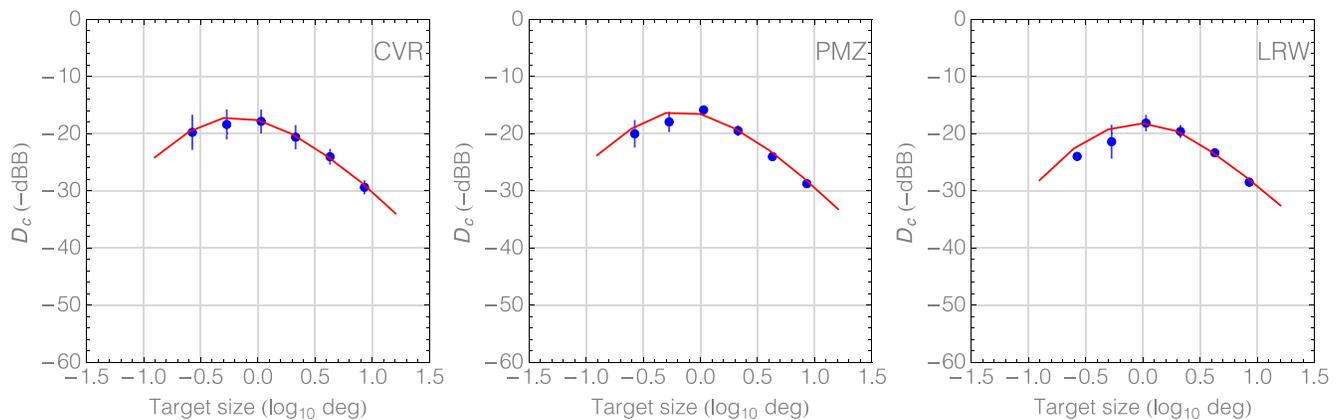


Figure 13. Contrast thresholds for aircraft identification. Target size is defined as half the image width. The red curve is the model prediction with efficiency one-third that for Gabor detection.

accurate, given the heterogeneity of observers, labs, and methods.

In the second part of the article, we made new measurements of both Gabor detection and letter identification, using a single set of observers, displays, and methods. Here the predictions were accurate for one observer but showed modest systematic discrepancies for two other observers. For the latter, sensitivity was accurately predicted for large letters but underpredicted for small letters. We have suggested several possible explanations for this discrepancy.

The NIC model provides a good account of the general pattern of results for letter identification as a function of size. Expressed as contrast difference energy, sensitivity peaks at about  $\frac{1}{4}$  degree and falls steeply at both smaller and larger sizes. The decline at smaller sizes is due to optical blur and neural blur contributed by the mRGC layer and depends largely on properties of that layer in the visual center. The decline at larger sizes is due primarily to action of the surround and progressively greater neural blur and noise at peripheral locations and secondarily to a decline in efficiency with size.

### Contrast sensitivity function

Since the work of Selwyn (1948), Schade (1956), Lowry and Depalma (1961), and Campbell and Robson (1968), the contrast sensitivity function has been a critical measurement of visual sensitivity (Robson, 1993) and a critical component of models of early vision. Yet two conceptual problems have persisted. The first is that there is no consensus explanation of its form and basis, although even Selwyn attributed it to “the structure of the retina.” The second problem is that there is no consistent empirical definition of the function, because it depends strongly on the size of the test pattern. Indeed, authors

have devised formulas in which this dependence is incorporated parametrically (Barten, 1999). The model we have presented here provides a unifying representation and explanation for contrast sensitivity. It is no longer understood as a one-dimensional function, describing the inverse of the contrast detection threshold as a function of the spatial frequency of a grating target. It is instead understood as the contrast energy required to achieve a specified level of classification performance and is governed primarily by optical blur, mRGC filtering and the eccentricity-dependent nature of that filtering, and an internal noise with a particular power spectral density. From this, one can compute arbitrary contrast sensitivity functions for gratings of arbitrary size and location in the visual field.

### Beckmann and Legge (2002)

Beckmann and Legge (2002) constructed a particularly detailed simulation of letter identification, including optical filtering, filtering by the photoreceptor aperture, sampling by a photoreceptor lattice, and template matching. Their model is one of the few to explicitly include parameters (e.g., cone density, cone aperture, and optical point spread) that vary with eccentricity, based on anatomical data. They compared threshold letter size for human and model at foveal and peripheral locations ( $0^\circ$ ,  $5^\circ$ , and  $20^\circ$ ). From  $0^\circ$  to  $20^\circ$ , threshold size increased by a factor of 13.7 for the human observers, whereas for the model, the factor was only 1.7. The modest effect of eccentricity on the ideal is surprising, given that the model cone density varied by a factor of 667 (and spacing by 25.8) over that range of eccentricity. But a primary difference from our model is that they considered only “preneural” limitations and thus did not include, as we did, mRGC receptive fields, sampling, and noise.



Figure 14. Neural images for small and large versions of an aircraft target. Image sizes are 1 deg (left) and 16 deg (right). These are the approximate limits used in our aircraft experiment (Figure 13).

### Identification of small and large patterns

Some studies suggest that small and large patterns are identified based on different sorts of features (Majaj, Pelli, Kurshan, and Palomares, 2002; Oruç & Landy, 2009). Both of these studies suggest that large letters are identified by their edges, whereas small letters are identified by the coarse strokes. This might appear inconsistent with our model, in which the neural image always provides the template. However, the neural images for small and large letters are quite different, because of the optical and retinal filtering. This is illustrated in Figure 14, in which we show the neural images for the same aircraft target when the image is either 1° or 16° in width. In the first case, the neural image is blurred but consists of large strokes brighter than the background. In the second, the neural image consists primarily of edges, which are attenuated as they extend into the periphery. Thus, the neural images for small and large targets do in fact agree with these earlier results.

### Equivalent noise

In the NIC model, energy sensitivity declines for large letters in part because the effective noise increases in the periphery, due to the greater sparsity of the mRGCf and the assumption that the predominant noise is at the output of the ganglion cells. The success of our predictions encourages this view of visual noise, at least under the conditions of our experiments, such as moderate mean luminance and long durations.

By measuring the point at which external noise begins to have an effect on thresholds, it is possible to estimate the so-called equivalent noise  $N_{eq}$ . This is an estimate of the internal noise, in units of the external

noise. In Figure 15, we show empirical estimates of  $N_{eq}$  for letter identification at various sizes (Pelli & Farell, 1999). In the same figure, we show an estimate of equivalent noise in the NIC model, computed as the inverse of the mean density over the letter area. The vertical position is arbitrary. It appears that much, although not all, of the equivalent noise can be attributed to the ganglion cells. The portion not accounted for at the largest sizes may be due to cortical noise or it may be that peripheral ganglion cells are inherently noisier than those in the fovea.

In the present version of the NIC model, we have assigned all of the noise to the output of the ganglion cells. It is likely that under various conditions, other sources of noise, including quantum fluctuations in the signal itself, will dominate or at least contribute. In principle, this more general model can be accommodated within the framework of the current model.

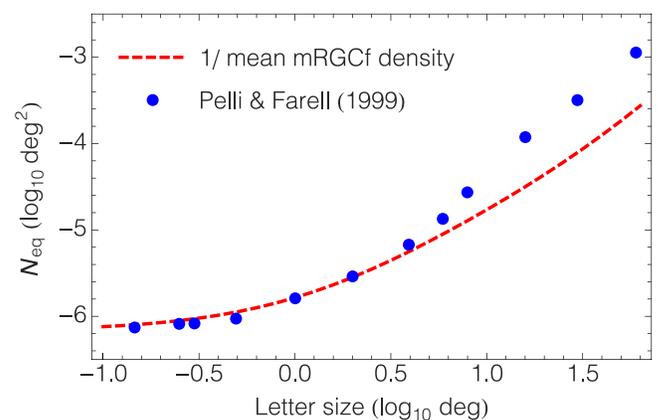


Figure 15. Empirical equivalent noise for letter identification (blue) estimated by Pelli and Farell (1999) and theoretical mRGCf noise derived from mean density (red).

## General model of detection and discrimination

The NIC described in this report is a refinement of models we have previously presented to account for visibility of spatial patterns (Watson & Ahumada, 2005). The enhancements provided by this model are primarily an explicit role for neural noise, the ability to predict arbitrary classifications (rather than just detection), and a more realistic explanation for variations with eccentricity based on the density of mRGC. We believe that this model could serve as a useful “first resort” in accounting for a broad class of visual experiments. In particular, the model provides predictions for visibility of luminance contrast targets of arbitrary size, shape, and location in the visual field. In a subsequent report, we will compare these predictions to published results.

## Limitations of the model

It is important to acknowledge the limitations of this model. The model offers no explanation for the modest efficiency ( $\sim 0.1$ ) of even the most efficient classifications. Likewise, we do not have a satisfying account for the variations in efficiency between classes of patterns (e.g., letters vs. aircraft). Complexity has been proposed as a determinant, but there is no accepted way of computing visual complexity (Watson, 2012). Promising results on the question of efficiency have been obtained recently using models of pattern learning (Ziskind, Hénaff, LeCun, & Pelli, 2014).

The model is also a “single channel model” and thus does not account for various phenomena, such as inefficient summation, often attributed to models that incorporate spatial frequency channels (Graham & Nachmias, 1971; Graham, Robson, & Nachmias, 1978). We have avoided the complexity of channels in an effort to see whether a simple model limited only by optical and retinal constraints could account for the essential aspects of the data. A similar model that also includes channels, but is concerned primarily with detection, has been proposed recently by Bradley et al. (2014).

Further, our model does not include masking (Foley, 1994; Watson & Solomon, 1997) but see Bradley et al. (2014) for an effort to include masking in a similar model. The NIC model also does not include either crowding, which will influence identification of peripheral targets when other patterns are nearby (Strasburger, Rentschler, & Juttner, 2011; Whitney & Levi, 2011), or aliasing, which should hamper identification of peripheral targets relative to predictions based only on a change of scale (Thibos et al., 1996). Finally, we note that this model accounts for only the

simplest sort of visual recognition: classification among a small set of fixed patterns.

## Summary

We have gathered data from eight published studies of contrast thresholds for letter identification as a function of size.

We have constructed an NIC model that performs ideal classification of visual targets, limited by optical blur, retinal filtering, and retinal noise.

The optical component of the model may be supplied by a recent formula for the mean human optical MTF for a given pupil diameter (Watson, 2013; Watson & Yellott, 2012).

The neural component of the model was provided by a recent formula for the size and spacing of the retinal ganglion cells as a function of position in the visual field (Watson, 2014).

When calibrated by Gabor detection data from the published literature (Watson & Ahumada, 2005), the NIC model gives a good account of the letter identification data.

We have also collected new data for three observers. These include optical measurements, Gabor detection thresholds, and letter identification thresholds.

When calibrated by the optical measurements and Gabor detection data for the new observers, the NIC model gives a good account of the letter identification data, although some discrepancies are evident.

We have also collected identification thresholds for aircraft images. These are also predicted well by the NIC model but with an efficiency about one-third that for detection of Gabors or identification of letters.

*Keywords: contrast, detection, identification, classification, letter identification, peripheral vision, contrast difference energy, noise, efficiency, optics*

## Acknowledgments

We gratefully acknowledge the assistance of Dr. Austin Roorda in the collection of wavefront aberration data. We thank Jeffrey Mulligan for helpful comments. This work supported by the NASA Space Human Factors Research Project WBS 466199.

Commercial relationships: Both authors are inventors of U.S. Patent 8408707 B1, Prediction of Visual Acuity From Wavefront Aberrations.

Corresponding author: Andrew B. Watson.

E-mail: [andrew.b.watson@nasa.gov](mailto:andrew.b.watson@nasa.gov).

Address: NASA Ames Research Center, Moffett Field, CA, USA.

## Footnote

<sup>1</sup> We define patterns in terms of contrast, rather than luminance, because we believe that the absolute luminance is discounted very early in vision and that, as a consequence, observers effortlessly generalize patterns across different luminance backgrounds. By use of the term *distribution*, we allow that the overall contrast might vary but that the relative spatial distribution remains constant.

## References

- Ahumada, A., & Scharff, L. (2007). Lines and dipoles are efficiently detected. *Journal of Vision*, 7(9), 337, <http://www.journalofvision.org/content/7/9/337>, doi:10.1167/7.9.337. [Abstract]
- Ahumada, A. J., & Watson, A. B. (2013). Visible contrast energy metrics for detection and discrimination. *Proceedings of the SPIE*, 8651, 86510D–86510D, <http://dx.doi.org/10.1117/12.2009383>.
- Alexander, K. R., Xie, W., & Derlacki, D. J. (1994). Spatial-frequency characteristics of letter identification. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 11, 2375–2382.
- Aparicio, J. A., Arranz, I., Matesanz, B. M., Vizmanos, J. G., Padierna, L., González, V. R., et al. (2010). Quantitative and functional influence of surround luminance on the letter contrast sensitivity function. *Ophthalmic and Physiological Optics*, 30, 188–199.
- Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Res*, 27, 1915–1924.
- Banks, M. S., Sekuler, A. B., & Anderson, S. J. (1991). Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling. *Journal of the Optical Society of America. A, Optics and image science*, 8, 1775–1787, <http://www.ncbi.nlm.nih.gov/pubmed/1744774>.
- Barten, P. G. J. (1999). *Contrast sensitivity of the human eye and its effects on image quality*. Bellingham, WA: SPIE Optical Engineering Press.
- Beckmann, P. J., & Legge, G. E. (2002). Preneural limitations on letter identification in central and peripheral vision. *Journal of the Optical Society of America. A, Optics and image science*, 19, 2349–2362, <http://www.ncbi.nlm.nih.gov/pubmed/12469729>.
- Blommaert, F. J. J., & Timmers, H. (1987). Letter recognition at low contrast levels: Effects of letter size. *Perception*, 16, 421.
- Bradley, C., Abrams, J., & Geisler, W. S. (2014). Retina-VI 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. [PubMed] [Article]
- Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197, 551–566.
- Carney, T., Klein, S. A., Tyler, C. W., Silverstein, A. D., Beutter, B., Levi, D., . . . Eckstein, M. P. (1999). The development of an image/threshold database for designing and testing human vision models. In B. Rogowitz & T. Pappas (Eds.), *Human vision, visual processing, and digital display IX* (Vol. 3644, pp. 542–551). Bellingham, WA: SPIE.
- Carney, T., Tyler, C. W., Watson, A. B., Makous, W., Beutter, B., Chen, C.-C., . . . Klein, S. A. (2000). Modelfest: Year one results and plans for future years. *Human Vision, Visual Processing, and Digital Display IX*, 3959, 140–151.
- Castro, C. N., & Horberry, T. (2004). *The human factors of transport signs*. Boca Raton, FL: CRC Press.
- Chung, S. T. L., Legge, G. E., & Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Research*, 42, 2137–2152.
- Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *Journal of Comparative Neurology*, 300, 5–25.
- Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of Comparative Neurology*, 292, 497–523, <http://www.ncbi.nlm.nih.gov/pubmed/2324310>.
- Dacey, D. M. (1993). The mosaic of midget ganglion cells in the human retina. *Journal of Neuroscience*, 13, 5334–5355, <http://www.ncbi.nlm.nih.gov/pubmed/8254378>.
- Dalimier, E., & Dainty, C. (2008). Use of a customized vision model to analyze the effects of higher-order ocular aberrations and neural filtering on contrast threshold performance. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 25, 2078–2087, [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=18677370](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18677370).
- Dosher, B. A., & Lu, Z.-L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceed-*

- ings of the National Academy of Sciences, 95, 13988–13993.
- 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, <http://www.ncbi.nlm.nih.gov/pubmed/17320143>.
- Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: John Wiley.
- Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology (London)*, 187, 517–552.
- Enroth-Cugell, C., Robson, J. G., Schweitzer-Tong, D., & Watson, A. B. (1983). Spatio-temporal interactions in cat retinal ganglion cells showing linear spatial summation. *Journal of Physiology (London)*, 341, 279–307.
- Foley, J. M. (1994). Human luminance pattern mechanisms: Masking experiments require a new model. *Journal of the Optical Society A*, 11, 1710–1719.
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, 96, 267–314.
- Ginsburg, A. P. (1978). Visual information processing based on spatial filters constrained by biological data. *Aerospace Medical Research Laboratory Report*, amrl-tr-78-129, 1&2.
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999a). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, 39, 3537–3560, <http://www.ncbi.nlm.nih.gov/pubmed/10746125>
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999b). Signal but not noise changes with perceptual learning. *Nature*, 402, 176–178, <http://www.ncbi.nlm.nih.gov/pubmed/10647007>.
- Graham, N., & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple-channels models. *Vision Research*, 11, 251–259.
- Graham, N., Robson, J. G., & Nachmias, J. (1978). Grating summation in fovea and periphery. *Vision Research*, 18, 815–825.
- Legge, G. E. (1984). Binocular contrast summation—II. Quadratic summation. *Vision Research*, 24, 385–394.
- Legge, G. E. (2006). *Psychophysics of reading in normal and low vision*. Mahwah, NJ: Lawrence Erlbaum.
- Legge, G. E., Rubin, G. S., & Luebker, A. (1987). Psychophysics of reading—V. The role of contrast in normal vision. *Vision Research*, 27, 1165–1177, <http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=0003660667>.
- Lowry, E. M., & Depalma, J. J. (1961). Sine-wave response of the visual system. I. The mach phenomenon. *Journal of the Optical Society of America*, 51, 740–746, <http://www.opticsinfobase.org/abstract.cfm?URI=josa-51-7-740>
- Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research*, 42, 1165–1184, [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=11997055](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11997055).
- McAnany, J. J., & Alexander, K. R. (2006). Contrast sensitivity for letter optotypes vs. gratings under conditions biased toward parvocellular and magnocellular pathways. *Vision Research*, 46, 1574–1584.
- Merigan, W. H., & Eskin, T. A. (1986). Spatio-temporal vision of macaques with severe loss of P<sub>β</sub> retinal ganglion cells. *Vision Research*, 26, 1751–1761.
- Michel, M., & Geisler, W. S. (2011). Intrinsic position uncertainty explains detection and localization performance in peripheral vision. *Journal of Vision*, 11(1):18, 1–18, <http://www.journalofvision.org/content/11/1/18>, doi:10.1167/11.1.18. [PubMed] [Article]
- Nestares, O., Navarro, R., & Antona, B. (2003). Bayesian model of Snellen visual acuity. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 20, 1371–1381, [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=12868641](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12868641).
- Oruç, I., & Landy, M. S. (2009). Scale dependence and channel switching in letter identification. *Journal of Vision*, 9(9), 4, 1–19, <http://www.journalofvision.org/content/9/9/4>, doi:10.1167/9.9.4. [PubMed] [Article]
- Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research*, 31, 1399–1415, [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=1891827](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1891827).
- Pelli, D. G. (1990). The quantum efficiency of vision. In C. B. Blakemore (Ed.), *Vision: Coding and efficiency*. Cambridge, UK: Cambridge University Press.
- Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, 46, 4646–4674, <http://www>.

- sciencedirect.com/science/article/B6T0W-4K9C562-1/2/3d22863119565906e0ad3760a24b4880.
- Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *16*, 647–653.
- Perry, J. S., & Geisler, W. S. (2002). Gaze-contingent real-time simulation of arbitrary visual fields. *Proceedings of the SPIE*, *4662*, 57–69.
- Robson, J. (1993). Contrast sensitivity: One hundred years of clinical measurement. In R. Shapley & D. M. K. Lam (Eds.), *Contrast sensitivity* (Vol. 5, pp. 253–267). Cambridge, MA: MIT Press.
- Robson, J. G. (1980). Neural images: The physiological basis of spatial vision. In C. S. Harris (Ed.), *Visual coding and adaptability* (pp. 177–214). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell responses to visual stimuli. *Vision Research*, *5*, 583–601.
- Schade, O. H., Sr. (1956). Optical and photoelectric analog of the eye. *Journal of the Optical Society of America*, *46*, 721–739.
- Schira, M. M., Wade, A. R., & Tyler, C. W. (2007). Two-dimensional mapping of the central and parafoveal visual field to human visual cortex. *Journal of Neurophysiology*, *97*, 4284–4295.
- Sekiguchi, N., Williams, D. R., & Brainard, D. H. (1993). Efficiency in detection of isoluminant and isochromatic interference fringes. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *10*, 2118–2133, <http://josaa.osa.org/abstract.cfm?URI=josaa-10-10-2118>.
- Selwyn, E. (1948). *The photographic and visual resolving power of lenses*. Part I: Visual resolving power. *The Photographic Journal*, *88B*(6).
- Strasburger, H., Harvey, L. O., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Attention, Perception, & Psychophysics*, *49*, 495–508.
- Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5):13, 1–82, <http://www.journalofvision.org/content/11/5/13>, doi:10.1167/11.5.13. [PubMed] [Article]
- Tanner, W. P. (1961). Physiological implications of psychophysical data. *Annals of the New York Academy of Science*, *89*, 752–765.
- Thibos, L. N., Still, D. L., & Bradley, A. (1996). Characterization of spatial aliasing and contrast sensitivity in peripheral vision. *Vision Research*, *36*, 249–258, <http://www.ncbi.nlm.nih.gov/pubmed/8594823>.
- Watson, A., Ramirez, C. V., & Salud, E. (2009). Predicting visibility of aircraft. *PLoS One*, *4*, e5594, <http://dx.doi.org/10.1371/journal.pone.0005594>.
- Watson, A. B. (1987). The ideal observer concept as a modeling tool. In *Frontiers of visual science: Proceedings of the 1985 Symposium* (pp. 32–37). Washington, DC: National Academy Press.
- Watson, A. B. (2011). Video acuity: A metric to quantify the effective performance of video systems. *Proceedings, Optical Society of America, Imaging Systems Applications Topical Meeting, IMD3*, <http://www.opticsinfobase.org/abstract.cfm?URI=IS-2011-IMD3>.
- Watson, A. B. (2012). Perimetric complexity of binary digital images: Notes on calculation and relation to visual complexity. *Mathematica Journal*, *14*, <http://www.mathematica-journal.com/2012/02/perimetric-complexity-of-binary-digital-images/>.
- Watson, A. B. (2013). A formula for the mean human optical modulation transfer function as a function of pupil size. *Journal of Vision*, *13*(6):18, 1–11, <http://journalofvision.org/13/6/18>, doi:10.1167/13.6.18. [PubMed] [Article]
- Watson, A. B. (2014). A formula for human retinal ganglion cell receptive field density as a function of visual field location. *Journal of Vision*, *14*(7):15, 1–17, <http://www.journalofvision.org/content/14/7/15>, doi:10.1167/14.7.15. [PubMed] [Article]
- Watson, A. B., & Ahumada, A. J., Jr. (2005). A standard model for foveal detection of spatial contrast. *Journal of Vision*, *5*(9):6, 717–740, <http://journalofvision.org/content/5/9/6/>, doi:10.1167/5.9.6. [PubMed] [Article]
- Watson, A. B., & Ahumada, A. J., Jr. (2008). Predicting visual acuity from wavefront aberrations. *Journal of Vision*, *8*(4):17, 1–19, <http://journalofvision.org/content/8/4/17>, doi:10.1167/8.4.17. [PubMed] [Article]
- Watson, A. B., & Ahumada, A. J. (2012). Modeling acuity for optotypes varying in complexity. *Journal of Vision*, *12*(10):19, 1–19, <http://journalofvision.org/content/12/10/19>, doi:10.1167/12.10.19. [PubMed] [Article]
- Watson, A. B., Barlow, H. B., & Robson, J. G. (1983). What does the eye see best? *Nature*, *302*, 419–422.
- Watson, A. B., & Fitzhugh, A. E. (1989). Modelling character legibility. *Society for Information Display Digest of Technical Papers*, *20*, 360–363.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*, 113–120.
- Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking.

*Journal of the Optical Society A*, 14, 2379–2391, <http://josaa.osa.org/abstract.cfm?id=1940>.

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

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, 15, 160–168, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3070834/>.

Wolfram Research Inc. (2013). *Mathematica* (version 9.0 ed.). Champaign, IL: Wolfram Research, Inc.

Ziskind, A. J., Hénaff, O., LeCun, Y., & Pelli, D. G. (2014). The bottleneck in human letter recognition: A computational model. *Journal of Vision*, 14(10):1311, <http://www.journalofvision.org/content/14/10/1311>, doi:10.1167/14.10.1311. [Abstract]

## Appendix 1: Averaging observer data

To reduce the clutter of the data across the nine studies considered, we have averaged over observers within each study. For some studies, in which the same set of sizes was used for each observer, this is straightforward. In the other cases, we have used the following method.

The data for each observer consist of a set of contrast thresholds at a set of sizes. We first construct a linear interpolation between the measured points for each observer. We then construct a new set of sizes, consisting of the union of sizes used by all observers but limited to the range between the largest lower bound and the

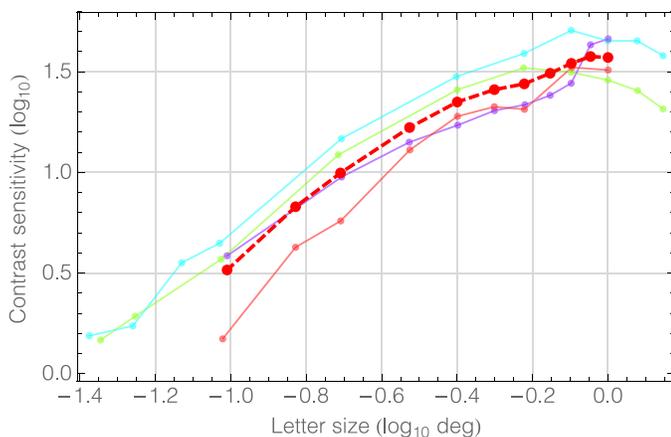


Figure A1. Example of averaging observers. The heavy red dashed line and points show the averaged data. The thin lines show the individual observer data.

smallest upper bound among the several observers. This eliminates any large or small sizes used with only a subset of the observers. We then average any groups of sizes that are separated by less than 1% of the total range. This merges sizes that are almost the same. The interpolations for the several observers are then evaluated at each of the new set of sizes, and the results are averaged over observers. An example of this process is shown for one observer in Figure A1. In this example (Strasburger, Harvey, & Rentschler, 1991), we see that some data are lost at the largest and smallest sizes.

## Appendix 2: Computing contrast difference energy for data sets

Calculation of contrast difference energy for the historical data is complicated by ambiguities in font, symbol set, letter size, monocular versus binocular viewing, and duration. For each study, we have first computed a contrast difference energy for the symbols used in the study at a nominal size of 1 degree. Contrast energy at other sizes could then be computed by multiplying by the square of nominal letter size.

### Font

Ginsburg used a “Snellen letter chart,” but the chart shown in his figure 86 includes letters not present in the Sloan font. The font appears Helvetica-like, however, so we have assumed a Helvetica bold font. Strasburger et al. (1991) used numerals in the unknown “zeile” font, but we obtained digital images of the symbols from the first author. Blommaert and Timmers (1987) used lowercase characters of a font called “Eurostile bold extended,” for which we used the Postscript font “Eurostile-BoldExtendedTwo.” Pelli and Farell (1999) do not specify their font, but based on illustrations and comparison with Pelli et al. (2006), we have assumed a lowercase Bookman bold font.

### Symbol size

In all but three studies, uppercase letters were used, and letter size was defined as letter height. In Blommaert and Timmers (1987) and Pelli and Farell (1999), a full alphabet of lowercase letters was used, and letter size was defined as “x-height”: the height of the lowercase letter “x.”

### Symbol set

In most of the studies, the 10 uppercase letters of the Sloan font were used (C, D, H, K, N, O, R, S, V, Z). In

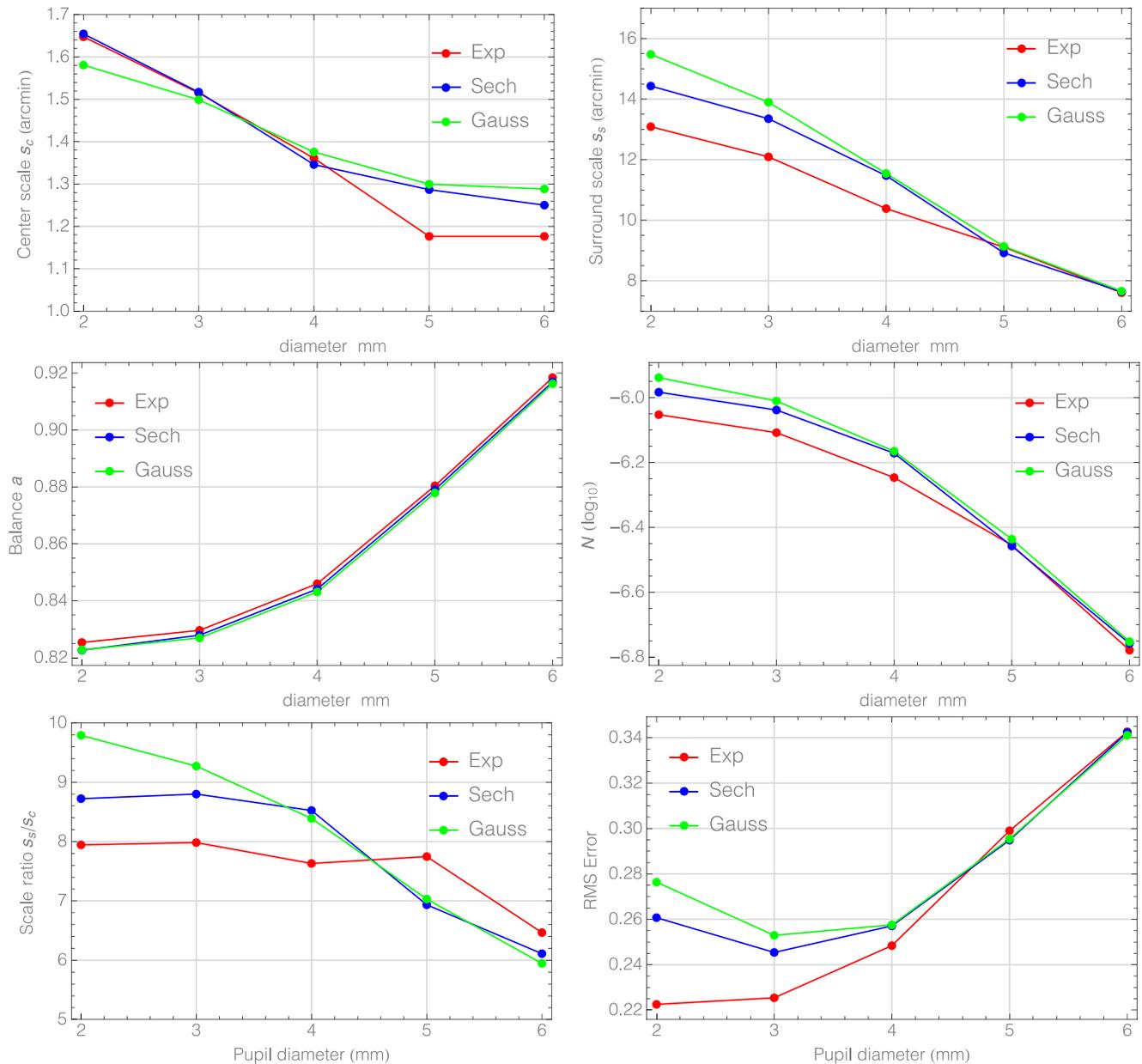


Figure A2. Parameter estimates and error of fits of NIC model to ModelFest data.

Ginsburg (1978), a single eye chart with a particular 12 distinct uppercase letters was evidently used. In Blommaert and Timmers (1987) and Pelli and Farell (1999), a full alphabet of lowercase letters was used. Strasburger et al. (1991) used the 10 numerals from 0 to 9. In each case, the contrast difference energy was determined for the corresponding set of symbols at a nominal symbol size of 1 degree.

### Duration

Most studies used fixed contrast presentations of specific finite durations, except for Alexander et al.

(1994), who used a sixth derivative of a Gaussian, with a peak temporal frequency at 2 Hz. We calculated that waveform to have an energy equivalent duration of 249 ms. Energy equivalent duration of a time waveform is defined as the duration of a pulse with the same energy as the waveform. Data collected for this article used the ModelFest time course of a Gaussian with a standard deviation of  $1/8$  s, which has an energy equivalent duration of  $\sqrt{\pi}/8 = 222$  ms. Legge et al. (1987), Ginsburg (1978), and Aparicio et al. (2010) used an indefinite duration. We have substituted an effective duration of 500 ms, because that value brought their data into rough agreement with the other studies.

## Number of eyes

Two studies used monocular viewing; the others used binocular viewing (see Table 1). We are uncertain about the study of Ginsburg (1978) but have assumed binocular viewing. Because the model used to predict these data effectively employs a single “cyclopean” eye, the monocular studies effectively have half as much contrast energy at a given contrast. We have compensated for this by reducing their computed contrast difference energy by a factor of 2 (Table 1). Note that in the model applied to the new data, a binocular model is used, and this adjustment is not necessary.

## Appendix 3: Model implementation

Here we provide some notes on the implementation of the NIC model. The model was implemented in Mathematica (Wolfram Research Inc., 2013). We use the term *image* to refer to a rectangular array of numbers, each of which specifies luminance contrast, and *target* to describe a spatially bounded collection of nonzero pixels embedded within an image. In most cases in this article, the target is a letter in a larger background image.

The model starts with a set of images that contain the targets to be identified. We typically began with images of dimensions  $1024 \times 1024$ . The images could be scaled to a smaller size, to accelerate computations.

We also supply an exposure duration. If the stimulus time waveform is not a pulse, we supply the energy equivalent duration (the pulse duration yielding the equivalent energy, equal to the integral of the square of the waveform). For the ModelFest stimuli, and the data collected for this report, the waveform was a Gaussian with a standard deviation of  $1/8$  sec, and the energy equivalent duration is  $\sqrt{\pi}/8$ .

We first compute the mean contrast energy of the set, when each image subtends 2 deg (in most cases, because the target is half the height of the image, this means the target height is 1 deg). Contrast energy is the integral of the square of contrast, integrated over space and time. For each image, it is the sum of the squares of the pixels, multiplied by the area of a pixel, and multiplied by the duration.

We also compute the mean size of the targets, when each image subtends 2 deg. This was computed using the Mathematica functions `MorphologicalComponents`, `DeleteSmallComponents`, and `ComponentMeasurements`, with option “`CaliperLength`.” This effectively measures the largest diameter of the convex hull (Feret diameter) of the target.

For a given pupil diameter and each of a defined set of image sizes in degrees, we compute the mean human

optical filter using the formula of Watson (2013) and filter the images of the set. We then compute the scale image, of the same dimensions as the target images, in which each pixel indicates the local relative spatial scale, defined as the square root of the density of mRGC receptive fields, relative to the visual center. This density is computed using the formula of Watson (2014).

Each image is then subjected to space-variant filtering, using the receptive field model (Equation 6), and where  $\delta(x)$  is provided by the scale image (Equation 7). Each image is then multiplied by the scale image. We do this to attenuate the contrast by the square root of receptive field density, so that spatially homogeneous noise can be used.

For the mean target size, scaled to degrees, we then computed the relative efficiency using Equation 8. For the set of images, we then determine the noise standard deviation that will yield identification performance at a target probability correct (we used 0.75). The returned value can also be regarded as an attenuation of the image contrast that will yield the target probability when the standard deviation is 1. Probability correct was obtained using Monte Carlo simulation of individual trials, using fast methods described previously (Watson & Ahumada, 2008, 2012). We used a search method to locate threshold. At each step of the search, it conducts 256 trials/image. It first estimates classification probability at an initial noise guess and, if necessary, increases the noise by factors of 2 until the probability is less than 1. It then decreases noise by factors of 2 until  $(\text{target probability} + \text{chance})/2 < p < 1$ . It then for eight steps fits a cumulative normal to the accumulated data and places the next step at the estimated noise that will yield the target probability. The returned value is then multiplied by the square root of the product of relative efficiency and duration and divided by the value of pixels/degree for that size. This corresponds to the predicted contrast sensitivity when the power spectral density is  $N = 1$ .

Because in the template model contrast sensitivity is linear with  $\sqrt{N}$ , predictions for any specified value of  $N$  can be obtained by multiplying the sensitivities by  $\sqrt{N}$ . On a log sensitivity axis (such as dB), this consists of a vertical shift. Thus, we can estimate  $N$  by finding the vertical shift that minimizes the separation between data and predictions.

## Appendix 4: Fits to ModelFest data

Contrast thresholds for 18 ModelFest images (Nos. 1–14, 26–29) were fit using the NIC model. These are 10 equal-sized Gabor functions, four Gabors with equal numbers of cycles, and four Gaussians with varying sizes (Watson & Ahumada, 2005). We chose these

pupil (mm)	kernel	$s_c$	$s_s$	$s_s/s_c$	$a$	$N$ ( $\log_{10}$ )	RMS
2	Exp	1.647	13.09	7.944	0.8253	−6.052	0.2225
3	Exp	1.515	12.09	7.984	0.8296	−6.108	0.2254
4	Exp	1.361	10.39	7.631	0.846	−6.246	0.2484
5	Exp	1.176	9.114	7.747	0.8804	−6.455	0.299
6	Exp	1.176	7.607	6.466	0.9184	−6.779	0.3426
2	Sech	1.324	14.43	10.9	0.8226	−5.983	0.2607
3	Sech	1.214	13.35	11	0.8279	−6.038	0.2454
4	Sech	1.077	11.47	10.66	0.8441	−6.171	0.2571
5	Sech	1.029	8.923	8.668	0.8791	−6.458	0.2949
6	Sech	1	7.639	7.638	0.9169	−6.758	0.3423
2	Gauss	1.581	15.48	9.791	0.8227	−5.938	0.2764
3	Gauss	1.499	13.9	9.273	0.8269	−6.01	0.2529
4	Gauss	1.376	11.54	8.391	0.8431	−6.165	0.2576
5	Gauss	1.3	9.137	7.031	0.8779	−6.436	0.2954
6	Gauss	1.288	7.66	5.946	0.9163	−6.753	0.3411

Table A1. Parameter estimates and RMS error of fits of NIC model to ModelFest data.

stimuli to provide a small set that spanned a range of sizes and frequencies. We minimized squared error between logs of data and predictions, weighted by the inverse variance of the data at each point. The RMS error is thus in units of standard deviations of the data. The fit optimized the four parameters  $s_c$ ,  $s_s/s_c$ ,  $a$ , and  $N$ . We considered three possible forms of center kernel  $k_s$ : Gaussian (Equation 4) as well as exponential and hyperbolic secant,

$$k_{\text{exp}}(r, s) = \frac{\pi}{2s^2} \exp\left(-\pi \frac{r}{s}\right) \quad (\text{A1})$$

$$k_{\text{sech}}(r, s) = \frac{\pi}{4Cs^2} \text{sech}\left(-\pi \frac{r}{s}\right). \quad (\text{A2})$$

We repeated the fit for each type of center kernel. We also repeated the fit for each of five pupil diameters between 2 and 6 mm.

The results of all of our fits are shown in Figure A2 and Table A1. In the figure, the center scale for the sech kernel has been multiplied by 1.25 to show agreement with the other kernels. We note that as pupil size increases, the various model parameters vary, because they are trying to account for a fixed pattern of performance of the ModelFest observers. The lowest error is for the Exp kernel with a 2-mm pupil. However, the ModelFest observers likely had pupils much larger than that, about 5 mm (Appendix 5). The decline in center scale with increasing pupil size can be understood as a tradeoff between blur contributed by optical and neural filters. As the modeled pupil enlarges, the optical blur increases; to compensate for this, the center scale decreases, which reduces neural blur. We have estimated the ModelFest mean observer pupil diameter to be 5 mm (Appendix 5). At that diameter, the fit of all three kernels is about equal.

## Appendix 5: Pupil diameter in the ModelFest experiment

The pupil diameters of the observers in the ModelFest experiment are unknown. Although all experimenters adopted a common mean luminance of  $30 \text{ cd m}^{-2}$ , no specifications were provided for the dimensions of the display or of ambient room lighting. In our own lab, which contributed 3 of the 16 observers, the display subtended  $1024 \times 768$  pixels, at 120 pixels/deg, for an area of about  $54.6 \text{ deg}^2$ . The average age of the observers was about 39 years. Using the formula of Watson and Yellott (2012), we compute a predicted average pupil diameter of about 5.15 mm.

## Appendix 6: Pupil diameter in archival letter identification experiments

The nine studies summarized in this report did not generally report pupil diameter. The one exception is Alexander et al. (1994), who used an artificial pupil of 2 mm. We can roughly estimate the pupil diameters in the other studies by using a formula relating mean luminance, display area, and observer age to pupil diameter (Watson & Yellott, 2012). Assuming a display area of  $100 \text{ deg}^2$  and an age of 39 years, and applying the formula to the reported mean luminances, we obtain pupil diameters of 4.4, 3.99, 3.66, 4.45, 2, 4.57, 4.47, 3.85, and 4.13 mm and an average of about 3.95 mm. Given the many assumptions and uncertainties

involved, we adopt a pupil diameter of 4 mm for the simulations of letter identification.

## Appendix 7: Experimental methods

This appendix describes the methods used to collect detection and identification data in our lab at NASA.

### Display

The stimuli were presented on an LCD monitor with an LED backlight (VIEWPixx Model VPX-VPX-2000A) with a resolution of  $1920 \times 1200$  pixels and a frame rate of 120 Hz. The display had a grayscale resolution of 12 bits (4,096 gray levels). To calibrate the display, we measured the gamma table: the luminance produced by each gray level. The gamma table was used to create a look-up table (LUT) that mapped from image gray levels to linear luminances with a specified contrast from  $-1$  to  $1$ . The pixel pitch of the display was 0.252 mm both horizontally and vertically. The display was viewed from a distance of 173.3 cm so that display resolution was 120 pixels/degree. The mean luminance of the display was  $116 \text{ cd m}^{-2}$ .

### Contrast and time course

For both detection and identification experiments, the contrast of the images was a Gaussian function of time, with a standard deviation of 0.125 s and a total duration of 1 s. This was the temporal contrast waveform used in the ModelFest experiments (Carney et al., 1999; Watson & Ahumada, 2005). The Gaussian could vary in peak amplitude between 0 and 1, thus determining the peak contrast of the stimulus. The contrast in each frame was determined by an appropriately constructed LUT, and the complete stimulus was controlled by a sequence of LUTs, presented at the 120-Hz frame rate. In advance of the experiment, we computed LUT sequences corresponding to Gaussians with peak contrasts ranging from 0 to  $-50$  dB in steps of 2 dB.

### Gabor and Gaussian stimuli

Gabor stimuli consisted of Gabor functions with a standard deviation of 30 arcmin and spatial frequencies of 1.12, 2, 2.83, 4, 5.66, 8, 11.3, 16, 22.6, and 30 cycles/deg. The Gabor modulation was in the vertical dimension (the stripes were horizontal). The sinusoid was in cosine phase at the center of the Gaussian. These

are ModelFest stimuli Nos. 1 to 10. In addition, we tested Gaussian stimuli, with standard deviations of 30, 8.43, 2.106, and 1.05 arcmin. These are ModelFest stimuli Nos. 26 to 29 (Watson & Ahumada, 2005).

### Letter and aircraft stimuli

Letter stimuli consisted of a set of 10 images of letters in the Sloan font. Each original image was  $1024 \times 1024$  pixels in size and was white (gray level 255) on a gray background (gray level 128). The letter target within each image was about half the width of the total image and was centered within each image. We used image sizes from 32 to 2,048 pixels in steps of a factor of 2. The upper limit was imposed by our display.

Aircraft stimuli consisted of 10 aircraft images created from 3D graphics models using methods described previously (Watson et al., 2009). The targets were equated for size (total number of nonbackground pixels) and contrast energy. We used image sizes from 64 to 2,048 pixels in steps of a factor of 2. The target occupied approximately half the width or height of the image, as shown in Figure 12.

In the experiment, the original images were scaled to a specified size using the Mathematica ImageResize operator with the option Resampling  $\rightarrow$  "Kaiser."

### General procedure

Observers viewed the display in an otherwise darkened room. A chin and forehead rest were used for comfort and to control viewing distance. Contrast thresholds were determined using a Quest adaptive procedure (Watson & Pelli, 1983). The Quest procedure controlled the peak contrast of the Gaussian LUT sequence. Each presentation interval was accompanied by an audible tone. Synthetic speech feedback was provided. Speech was used for observers' responses. The "Speakable Items" speech recognition capability of the Apple OSX 10.8.4 was used to recognize the verbal responses.

### Detection procedure

A two-interval forced-choice procedure was used. Observer verbal responses were "first" and "second." Each threshold was based on a block of 32 trials. Each observer completed at least three blocks for each target image. After the completion of a block of trials, the percentage correct as a function of contrast was fit by a log Weibull function (guessing parameter =  $1/2$ ) to estimate the contrast yielding 75% correct.

## Identification procedure

In a block of 100 trials, each of the 10 target images was presented 10 times. The order of the sequence was randomized. On each presentation, the stimulus was accompanied by a warning tone. The observer named the target with a verbal response. Synthesized speech feedback was provided (“right,” “wrong”). When the response was wrong, the correct name of the target was also spoken. For letter identification, we found that the simple names of the letters were sometimes confused by the speech recognition software, so we used the “NATO phonetic alphabet” instead ([http://en.wikipedia.org/wiki/NATO\\_phonetic\\_alphabet](http://en.wikipedia.org/wiki/NATO_phonetic_alphabet)). For aircraft identification, we assigned a specific verbal label to each aircraft (“apache,” “seven forty seven,” “c seventeen,” “cessna,” “embraer,” “f sixteen”, “fire-scout,” “reaper,” “d c 3,” “shuttle”). Observers were trained on large high-contrast targets until their performance was free of errors.

A single Quest staircase was used for the complete set of 10 images at one image size. After the completion of 100 trials, the percentage correct as a function of contrast was fit by a log Weibull function (guessing parameter = 1/10) to estimate the contrast yielding 75% correct. Each observer completed at least three blocks for each of the seven image sizes.

## Appendix 8: Wavefront aberration measurements

Wavefront aberration data were collected from four observers (A. B. W., C. V. R., P. M. Z., L. R. W.) on March 20, 2014, at the University of California with the assistance of Austin Roorda of the Department of Optometry. All observers except C. V. R. had eyeglasses. Using a custom-built Shack-Hartmann wavefront sensor, Dr. Roorda collected several images for each eye of each observer, both with and without glasses if they had them. Subsequently, Dr. Roorda selected the three best images for further processing. From these, he created Zernike coefficients of several pupil diameters, extending from the largest possible and then smaller in integer multiples of 1 mm down to 3 mm. In each case, coefficients were obtained by fitting the Zernike polynomials to the surface over the relevant diameter. The result was a collection of 222

files. Each file included coefficients up to 10th order, resulting in 66 terms.

We subsequently analyzed the data for a 4-mm pupil, because that was the presumed diameter during our psychophysical experiments. We also selected the condition (glasses or no glasses) used in the experiment. For observer P. M. Z., this was with glasses; for C. V. R. and L. R. W., this was without glasses. For each file, we first deleted zero- and first-order terms, as well as the defocus term. We then selected the one of the three replications for each eye that was closest in the mean squared sense to the average of the three replications. The result was one set of coefficients for each eye of each observer.

For each set of coefficients, we computed the polychromatic PSF and OTF for an image of size 1 deg and 256 pixels. We used a wavelength spacing of 20 nm, which previous calculations showed was sufficient (Watson, 2013). We assumed an equal energy white spectrum. The resulting PSFs are shown in Figure 8, and the radial MTFs are shown in Figure 9.

## Appendix 9: Notation

symbol	Definition	Unit
$k(r, s)$	normalized kernel of mRGC center or surround	
$r$	radial distance from receptive field center	deg
$s_c$	center kernel scale	deg
$s_s$	surround kernel scale	deg
$a$	balance, ratio of center and surround weights	
$x$	visual field location	(deg, deg)
$\delta(x)$	mRGC local scale	
$d(x)$	mRGC receptive field density	cells deg <sup>-2</sup>
$M$	Number of alternatives in classification	
$N$	Power Spectral Density of neural noise	deg <sup>-2</sup> sec <sup>-1</sup>
$D_c$	Contrast difference energy	
$E_v$	Contrast energy of neural image	
$\eta$	efficiency	
$\eta'$	relative efficiency	

Table A2. Notation used in this report.