

# Predicting visual acuity from wavefront aberrations

Andrew B. Watson

NASA Ames Research Center, Moffett Field, CA, USA



Albert J. Ahumada Jr.

NASA Ames Research Center, Moffett Field, CA, USA



It is now possible to routinely measure the aberrations of the human eye, but there is as yet no established metric that relates aberrations to visual acuity. A number of metrics have been proposed and evaluated, and some perform well on particular sets of evaluation data. But these metrics are not based on a plausible model of the letter acuity task and may not generalize to other sets of aberrations, other data sets, or to other acuity tasks. Here we provide a model of the acuity task that incorporates optical and neural filtering, neural noise, and an ideal decision rule. The model provides an excellent account of one large set of evaluation data. Several suboptimal rules perform almost as well. A simple metric derived from this model also provides a good account of the data set.

**Keywords:** Sloan letters, letter identification, pattern recognition, Zernike polynomials, autorefraction

**Citation:** 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/8/4/17/>, doi:10.1167/8.4.17.

## Introduction

It is now possible to routinely measure the monochromatic aberrations of the human eye (Cheng, Barnett, et al., 2004; Cheng, Himebaugh, Kollbaum, Thibos, & Bradley, 2004). This means that we can precisely specify the retinal image produced by an arbitrary target, and yet, remarkably, we cannot predict the visual acuity that will result from a given set of aberrations. In this paper, we present a model that closes this gap and that connects wavefront aberrations to visual acuity.

One reason to seek a prediction of acuity from aberrations is the possibility of automated objective measurement of visual acuity and of automated prescription of spherocylindrical corrections. However, it has been shown that correcting the spherical and cylindrical components of the aberrations (equivalent to minimizing the RMS error of the wavefront) does not provide best acuity. Thus, these automated procedures must await a more sophisticated metric that can predict acuity from an arbitrary set of aberrations.

In this paper, we evaluate the performance of several simple metrics that predict letter acuity from wavefront aberrations. To evaluate the metrics, we make use of a set of previously published data (Cheng, Bradley, & Thibos, 2004).

One reason for our interest in this problem is our recent proposal for a standard model for foveal contrast detection (Watson & Ahumada, 2005). We were interested to know whether that model, suitably extended with aberrated optics, could provide the basis for predicting acuity from wavefront aberrations.

Our interest in this subject also derives from our general interest in human pattern identification. As relatively

simple, small, highly learned patterns, letters have long served as canonical subject for pattern identification studies and models (Bouma, 1971; Gervais, Lewis, Harvey, & Roberts, 1984; Gibson, Osser, Schiff, & Smith, 1963; Parish & Sperling, 1991; Pelli, Palomares, & Majaj, 2004; Solomon & Pelli, 1994; Watson & Fitzhugh, 1989). We have previously developed simple image classification models, limited by early constraints such as contrast sensitivity, that we hoped to test in the context of known optical limitations.

## Previous studies

Guirao and Williams (2003) explored seven metrics to predict image quality and thus refraction from wavefront aberrations. Two of their metrics were so-called “pupil plane” metrics, which consisted of calculations on the wavefront over the pupil. The remaining five were “image plane” metrics consisting of calculations in the retinal image plane. For six eyes, they collected aberration data and subjective refraction in the same apparatus. For each metric and each eye, they found the combination of sphere and cylinder that would maximize the metric. They found that their image plane metrics provided good predictions (mean error  $\approx 0.1$  D), while the pupil plane metrics yielded poorer predictions, whose quality declined as higher order aberrations increased. The same general pattern (with somewhat larger average error) was found in a second population of 146 observers, from whom separate aberration and refraction data were collected in different settings.

Cheng, Bradley, et al. (2004) collected acuity measurements from observers viewing computationally aberrated Sloan letters and compared the results to the predictions of

31 metrics. Their data clearly confirmed that the presence of fourth order aberrations altered the effect of defocus; notably, the amount of defocus required for best acuity depended upon the amount of higher order aberration. They also determined that several of their metrics correlated well with measured letter acuity.

Thibos, Hong, Bradley, and Applegate (2004) evaluated the ability of the same 31 metrics, plus two additional metrics, to predict best refraction. They began with aberration maps from 200 subjectively well-corrected eyes (Thibos, Hong, Bradley, & Cheng, 2002). For each metric and each map, they simulated subjective refraction by computationally varying spherical and cylindrical corrections so as to maximize predicted acuity. They then compared the resulting simulated correction with the actual correction, which was always zero because the eyes were well corrected. They found that five metrics were reasonably accurate (mean corrections near zero) and precise (low variance among corrections).

Applegate, Marsack, Ramos, and Sarver (2003) measured letter acuity for three observers in the presence of 34 distinct mixtures of two Zernike modes whose total RMS error was fixed at  $0.25 \mu\text{m}$ . Despite a constant wavefront RMS error, performance varied by up to 0.2 LogMAR units. LogMAR is defined here as

$$\text{LogMAR} = \log_{10} \left( \frac{h}{5} \right), \quad (1)$$

where  $h$  is Sloan letter height in minutes of arc.

Marsack, Thibos, and Applegate (2004) compared these experimental results to computed values of the 31 metrics discussed above. Their best metric (VSOTF) accounted for 81% of the variance; altogether, six of the metrics accounted for 70% or more.

## Metrics and models

In this paper, we adopt a distinction introduced previously between metrics and models (Watson & Ahumada, 2005). The former consist of formulas or computations that merely describe the quantitative relationship, while the latter include mechanistic statements that attempt to explain why the relationship exists. Models serve a scientific purpose, while metrics may have a practical advantage due to their greater simplicity.

In assessing metrics and models of the relation between aberrations and acuity, we distinguish four types. At one extreme are metrics that incorporate only optics. An example is the RMS error of the wavefront. Metrics that add some description of post-optical (neural) processing form the next group. All of the models considered above are of these first two types (Thibos et al., 2004). More complete are models or metrics that include the specific acuity targets, such as Sloan letters, in the calculation.

And still more complete are models that also simulate the specific task undertaken in the acuity measurement: identification of individual letters. Here we develop a model and a metric, both of which fall into this last category.

The model and the metric we propose are both designed to predict letter acuity from wavefront aberrations. The first model is a Monte Carlo simulation of a decision process. This model has several variants, one of which is an ideal observer limited by optics, neural filtering, and neural noise. The second metric is a deterministic calculation involving optics, letters, and a hypothetical neural CSF. We will consider in turn the details and results of our model and our metric.

## Data

### Specifying aberrations

The aberrations of the eye are conveniently described using the now-standard normalized Zernike expansion [American National Standards Institute (ANSI), 2004; Thibos, Applegate, Schwiegerling, & Webb, 2002]. This consists of a set of polynomial basis functions (also called modes) defined over the pupil; a particular aberration state is defined by the set of coefficients, and the wavefront is given by a linear combination of the bases weighted by the coefficients. Each basis is defined by two indexes, which are the order and frequency (also called the radial order and meridional frequency). The conventional notation for a basis of order  $n$  and frequency  $f$  is  $Z_n^f$ . We will also sometimes use the list notation  $\{n, f\}$  or when associated with a coefficient  $\{n, f, c\}$ . Where several modes are present, we represent them as a list of lists,  $Z = \{\{n_1, f_1, c_1\}, \{n_2, f_2, c_2\}, \dots\}$ . Defocus and astigmatism are determined by second order modes  $\{2, 0\}$  and  $\{2, \pm 2\}$ , respectively. The reader is referred to Thibos, Hong, et al. (2002) for a more detailed discussion of the Zernike polynomials.

One measure of the magnitude of a complete set of aberrations is the RMS error of the wavefront, which is equal to the square root of the sum of the squares of the coefficients. Another measure is the equivalent defocus,  $M_e$ , measured in diopters, and given by

$$M_e = \frac{4\pi\sqrt{3}\text{RMS}}{\text{pupil area}}, \quad (2)$$

where RMS is measured in  $\mu\text{m}$ , and pupil area is given in  $\text{mm}^2$  (Thibos, Hong, et al., 2002).

We note that “equivalent defocus” is a possibly misleading term since two aberrations with the same equivalent defocus may have very different blur point spreads and yield

very different acuities. We adopt it only as a means of indexing various aberrations and do so only because it is the index used by Cheng, Bradley, et al. (2004).

## Acuity versus aberrations

The data we used in this research have been described previously (Cheng, Bradley, et al., 2004). They consist of measurements of acuity (LogMAR) for each of 67 particular wavefront aberrations. Each aberration was produced by computing the aberrated image, which was then viewed through an interference filter to yield luminance of  $264 \text{ cd/m}^2$  at 556 nm. Appropriate optics were used to ensure that specific controlled aberrations could be introduced. The test objects were Sloan letters. Four observers participated. To simplify matters, the observers will be identified throughout by color names (Red, Green, Blue, Brown) and in figures by the corresponding color. Two observers (Red and Green) viewed a set of 45 aberrations, the other two (Blue and Brown) viewed a different set of 22 aberrations.

The aberration conditions consisted of three levels (0.1, 0.21, 0.5 D) of a higher order aberration, combined with

Observer	Low	High
Green and Red	{2, -2}	{4, -2}
	{2, 2}	{4, 2}
Brown and Blue	{2, 0}	{4, 0}

Table 1. Aberration conditions used for each observer.

seven or eight levels of a lower order aberration (defocus or astigmatism), as specified in Table 1. In addition, zero aberration conditions were measured for observers Red, Green, and Blue.

Visual acuity was measured by collecting psychometric functions for letter identification as a function of letter size. Ten presentations of each of eight to ten logarithmically spaced letter sizes were used for each function. From each psychometric function, acuity was estimated as the size yielding 55% correct. The acuity data from this experiment are plotted in Figure 1. Each curve shows the variation in acuity due to variation in the amount of second order aberration, added to a fixed amount of fourth order aberration. Each curve shows a minimum—the defocus or astigmatism at which acuity is best—but the

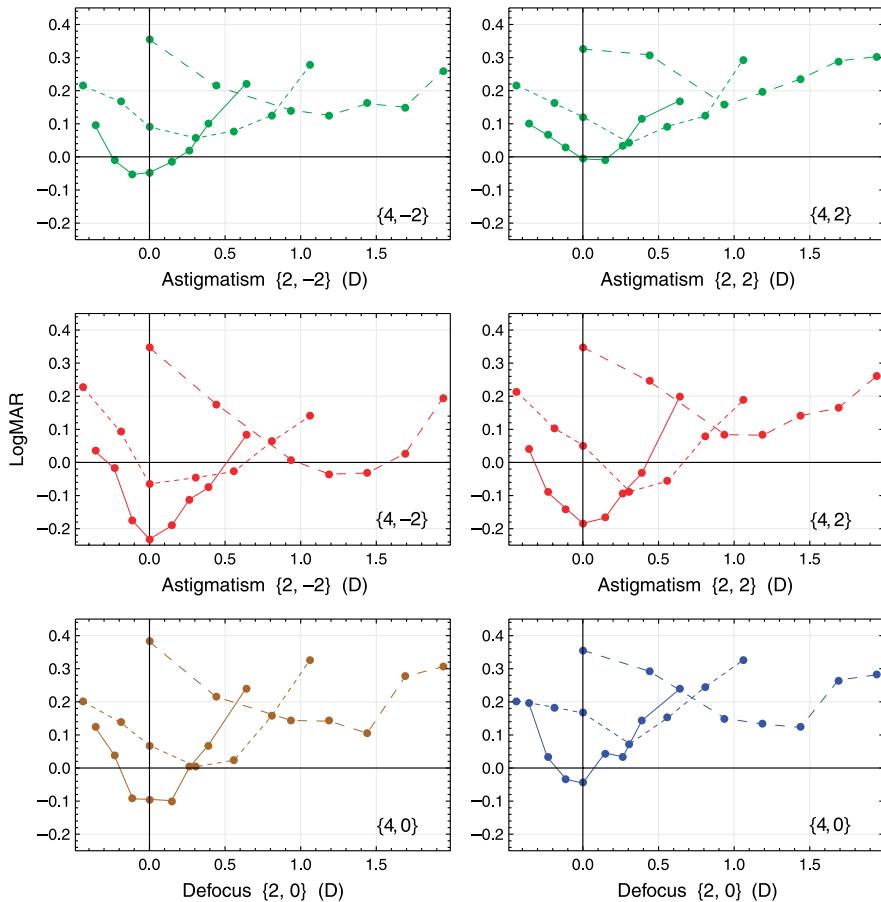


Figure 1. Acuity as a function of defocus or astigmatism in the presence of fixed fourth order aberrations. Line dashing indicates different amounts of the fourth order aberration (solid: 0.1 D, small dash: 0.21 D, large dash: 0.5 D). The colors identify individual observers. Data from Cheng, Bradley, et al. (2004).

positions of these minima clearly vary with the amount of higher order aberration (indicated by different dashing).

Not shown in Figure 1, but used in the subsequent calculations, are the three data points for zero aberration obtained from observers Red, Green, and Blue; these had values of  $-0.23$ ,  $0.036$ ,  $-0.076$  LogMAR, respectively. We were unable to obtain data for one additional condition from Cheng, Bradley, et al. (2004), consisting of pure astigmatism  $\{2, 2\}$  combined with coma  $\{3, -1\}$ . It is unlikely that these 12 data points would have a large effect on our general conclusions.

## Acuity model

All of the acuity models that we have considered are instances of what is generally called “template matching.” An overview of this general model is shown in Figure 2. In this example, the image of a letter “K” is first filtered by the optical transfer function (OTF) and then by a neural transfer function (NTF) to form the *neural image*, which is then perturbed by additive noise. In our simulations, the noise was always zero-mean Gaussian white noise. The noisy neural image is then compared to a set of template images, one for each candidate letter, and the closest match selected.

The algorithms we considered varied in two aspects: the matching rule and the templates. The matching rules that we have considered are maximum probability (ideal observer), minimum distance, and maximum normalized correlation. The templates we have considered are the original letter images, the images after filtering by diffraction-limited

optics and neural filter, and the images after filtering by aberrated optics and the neural filter (the neural images).

As will be noted below, we considered varying degrees of spatial uncertainty of the templates—in other words, we considered multiple positions of each template. However, we did not consider other forms of uncertainty, such as orientation or size.

## General methods

The models described here are implemented in the digital image domain. Inputs and outputs of the model at each stage (except the final classification) are discrete, finite digital images. The images were  $256 \times 256$  pixels, with a nominal resolution of 313.91 pixels/degree, so the images subtended 0.815 degrees. Computations were conducted in floating point and were implemented in the Mathematica programming language (Wolfram, 2003).

## Letter images

We created images of Sloan letters at LogMAR values from  $-0.6$  to  $0.7$  in steps of 0.05. The images were created using the rendering capabilities of Mathematica, and the Sloan PostScript font provided by Denis Pelli (<http://www.psych.nyu.edu/pelli/software.html>). In the simulations, the image pixel values were scaled to real numbers in the range  $\{0, 1\}$ . The letters were black (0) on a white (1) background.

We have provided a file called **SloanLetterImages** that contains all of the Sloan letter images used in this project. This file is described further in Appendix A.

## Optical transfer function

The optical transfer function (OTF) for a given aberration state was computed using standard methods (Artal, 1990), as illustrated in Figure 3. Starting from a set of Zernike coefficients, we first created a discrete image of the wavefront aberration, WA. We then computed the generalized pupil image, defined as

$$GP(x, y) = P(x, y) \exp \left[ \frac{i2\pi}{\lambda} WA(x, y) \right], \quad (3)$$

where  $P(x, y)$  is the pupil aperture image defined as 1 within the pupil and 0 elsewhere. The point-spread image is then computed as the squared modulus of the Fourier transform of the generalized pupil image. In order to obtain a desired resolution of the point-spread image, the generalized pupil image may first be embedded in a larger image of zeros. The OTF is obtained as the discrete Fourier transform (DFT) of the point-spread image. The

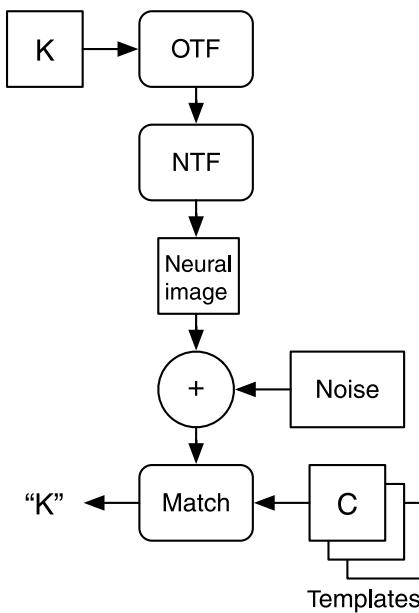


Figure 2. Template matching model for letter acuity.

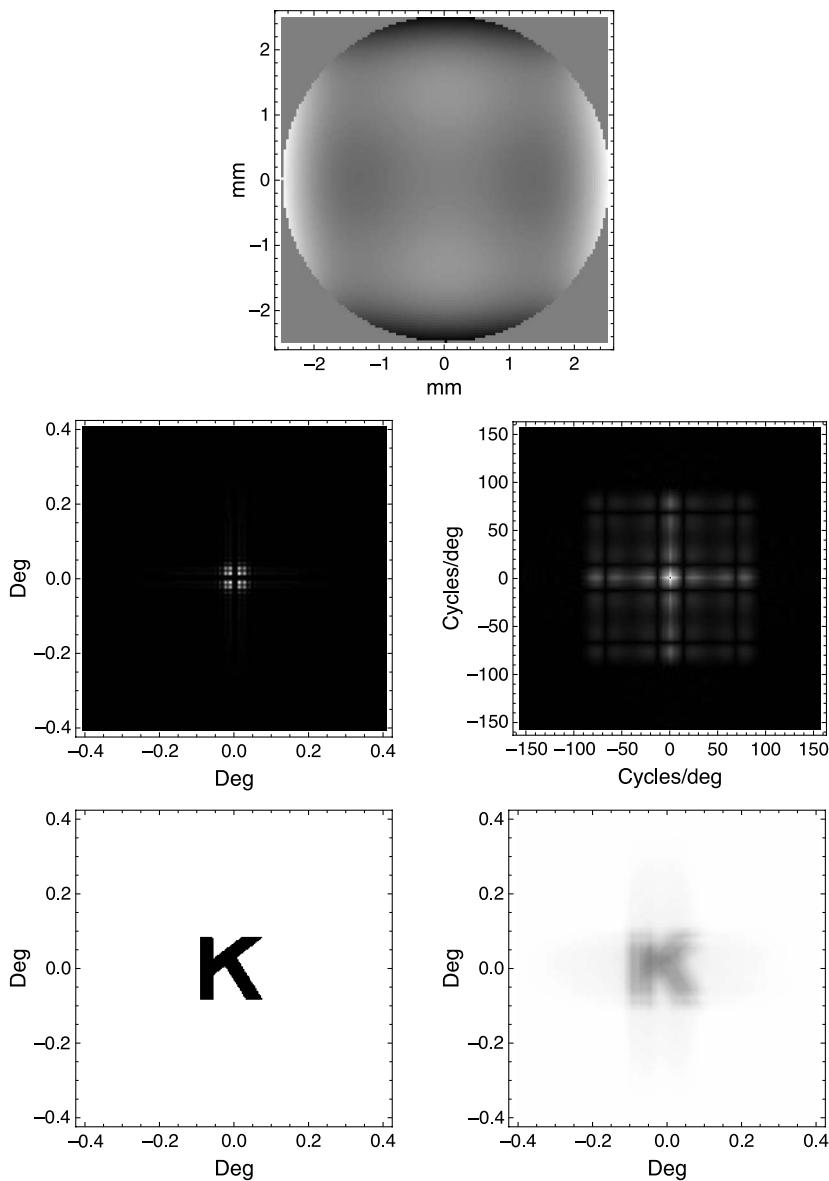


Figure 3. Computation of aberrated retinal images. (Top) wavefront aberration image; (middle) point-spread image and OTF; (bottom) original letter (LogMAR = 0.3) and aberrated image of letter. The aberrations pictured here are one of the conditions from the experiment of Cheng, Bradley, et al. (2004). They consist of 0.5 D in each of primary astigmatism and secondary astigmatism. In list notation, the aberrations are  $\{\{2,2,0.451\}, \{4,2,0.451\}\}$ .

letter image is then convolved with the point-spread image to obtain the retinal letter image. This convolution is implemented by multiplication of the OTF and the DFT of the letter image, followed by an inverse DFT. In all of the simulations in this paper, the pupil diameter is set to 5 mm. This is the value used to compute the images in the experiment of Cheng, Bradley, et al. (2004).<sup>1</sup>

## Neural transfer function

In the model pictured in Figure 2, the overall contrast sensitivity function is the product of the optical and neural transfer functions (OTF and NTF). The NTF can thus be

derived by dividing the CSF by the OTF. We have recently proposed a standard form for the contrast sensitivity function for foveal contrast detection (Watson & Ahumada, 2005). This function incorporates both a radial component and an oblique-effect component. We adopt this function here as the standard contrast sensitivity function (SCSF), using the parameters of Standard A from Table 5 of that report.

However, since the SCSF incorporates both optical and neural transfer functions, it is necessary to remove the optical component since that will be separately specified by the OTF of the model.

We accomplish this by constructing a mean optical transfer function (MOTF), making use of a published

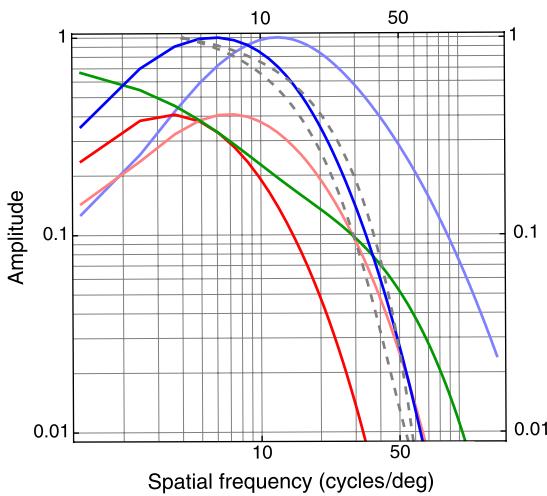


Figure 4. Derivation of the neural transfer function (NTF) (blue) from the standard CSF (SCSF) (red) and mean optical transfer function (MOTF) (green). Versions of the SCSF and NTF for a frequency scale of 2 are also shown in pink and light blue, respectively. Gray dashed curves are from Campbell and Green (1965), obtained by fitting cubic polynomials to their data for two observers and normalizing each curve at 4 cycles/deg.

formula that includes parameters for age and pupil size (IJsspeert, van den Berg, & Spekreijse, 1993). We assumed an age of 35 years and a pupil size of 5 mm. The resulting MOTF is pictured in Figure 4.

To obtain the neural transfer function NTF, we divide the SCSF by the MOTF. These three functions are pictured in Figure 4. The NTF has been normalized by its maximum value to give a peak gain of 1.

### Frequency scale

The MOTF and SCSF are based on population averages and may not be accurate for the observers we are simulating here. Further, different observers may have different NTFs. To investigate possible variations in the neural transfer functions of different observers, we introduced a parameter  $\phi$ , the frequency scale, that multiplies the two parameters  $f_0$  and  $f_1$  of the SCSF (Watson & Ahumada, 2005). This has the effect of shifting the SCSF horizontally in the log-log coordinates of Figure 4. This in turn has the effect of shifting the NTF to higher values. Higher values of the frequency scale correspond to higher values of acuity. Examples of the SCSF and NTF for a frequency scale of  $\phi = 2$  are shown in pink and light blue, respectively, in Figure 4.

This method of varying the NTF is somewhat arbitrary, and other variations might produce a better fit to the data. It has the advantage of providing a single parameter that controls perhaps the most important aspect of the NTF: its sensitivity at high spatial frequencies. For comparison, we also plot in Figure 4, as dashed gray curves, the NTFs derived by Campbell and Green for two subjects from

visibility of interference fringes (Campbell & Green, 1965). These are close to our NTF for  $\phi = 1$ .

### Neural images

Neural images were computed from letter images by filtering by the OTF and the NTF. The filtering was computed digitally in the frequency domain, using discrete sampled versions of the filters.

### Matching rules

#### Ideal observer

An ideal observer limited by optics, internal noise, and spatial uncertainty can be implemented with a particular matching rule. The rule computes the Bayesian posterior probability of each candidate letter, possibly shifted by integer pixel amounts horizontally and vertically, and selects the largest (Watson & Fitzhugh, 1989). The templates used by this model are the neural images of the aberrated letters.

We specify the spatial uncertainty in terms of the prior probability of integer pixel spatial shifts of the test image relative to the template, specified by a probability density  $u(\mathbf{x})$ . We considered varying amounts of spatial uncertainty, but report results only for two special cases: no uncertainty or complete uncertainty.

Categorization problems can often be analyzed in terms of discriminants: functions on the data indexed by category alternatives, the largest of which identifies the selected category (Duda & Hart, 1973). We can think of each discriminant function as computing a score for one alternative and the alternative with the highest score wins. The discriminant for the ideal observer can be written as

$$g_j = \log \left[ \sum_{\mathbf{x}} u(\mathbf{x}) \exp \left( \frac{1}{\sigma_n^2} s_j(\mathbf{x}) \otimes (s_k(\mathbf{x}) + n(\mathbf{x})) \right) \right] - \frac{1}{2\sigma_n^2} \| s_j \|^2, \quad (4)$$

where  $s_k$  and  $s_j$  are the sample and the candidate letter neural images, respectively,  $\otimes$  is the cross-correlation operator, and  $n$  is the Gaussian noise with standard deviation  $\sigma$ . This quantity is computed for each candidate letter  $j$ , and the value of  $j$  for which  $g_j$  is largest identifies the letter. A summary of model notation is provided in Appendix E.

### Minimum distance

If we again consider the templates as the neural images of the candidate letters and regard the neural images as points in a space with dimensions equal to the number of pixels, then we can consider a matching rule that selects the template closest to the noisy neural image of the test

letter. We have only considered this rule for the case in which the templates are the neural images of the candidate letters ( $t_j = s_j$ ), in which case the determinant for this minimum distance rule is given by

$$g_j = \max \left[ s_j(\mathbf{x}) \otimes (s_k(\mathbf{x}) + n(\mathbf{x})) \right] - \frac{1}{2} \| s_j(\mathbf{x}) \|^2. \quad (5)$$

This expression is derived in [Appendix B](#). The expression also allows for spatial shifts of the template (spatial uncertainty), although in this case all possible shifts are treated equally.

### Normalized correlation

Normalized correlation is a widely used pattern matching algorithm (Duda & Hart, [1973](#)). The normalization is applied so that patterns will be matched on the basis of shape rather than mean value. In the present context, it is an interesting rule because it allows us to consider templates other than the aberrated neural images.

The discriminant for normalized correlation is

$$g_j = \max \left[ \bar{t}_j(\mathbf{x}) \otimes (s_k(\mathbf{x}) + n(\mathbf{x})) \right], \quad (6)$$

where  $\bar{t}_j$  is the normalized template for the letter indexed by  $j$ . As discussed below, this may be the aberrated neural image, the original letter, or a diffraction-limited neural image. Note that the result of cross-correlating the test and candidate images is itself an image, in which the value at each pixel reflects the correspondence of the two images when one is shifted by the coordinates of that pixel. Taking the maximum selects the value at the shift with the greatest correspondence. Thus, this rule also accommodates spatial uncertainty, although the uncertainty here is uniform over the image. An uncertainty function could be introduced here, as in the ideal matching rule, but we have not done so.

Comparison of [Equation 6](#) with [Equation 5](#) shows that normalized correlation is an instance of minimum distance when the templates are normalized.

### Templates

For the normalized correlation rule, we considered three types of templates: the original letters, the neural images of the aberrated letters, or the neural images of the letters filtered by diffraction-limited optics. The reasoning behind these three choices was as follows. The aberrated neural image is appropriate if the observer is able to learn a template from prior experience of the letter under the optics in question. This is also the template used by the ideal observer. The original letter would be appropriate if the observer retained a template, perhaps learned at a larger size at which optical degradation would be minimal, and mentally scaled down to the appropriate size. The diffraction-limited case corresponds to an

Identifier	Rule	Templates	Uncertainty
ID	Ideal	Aberrated	Zero
IU	Ideal	Aberrated	Infinite
DA	Distance	Aberrated	Infinite
XA	Cross-correlation	Aberrated	Infinite
XD	Cross-correlation	Diffraction limited	Infinite
XL	Cross-correlation	Letters	Infinite

Table 2. Models and their attributes.

observer who has learned the letter under diffraction-limited conditions. It may be viewed as a case intermediate between the other two.

In [Table 2](#), we summarize the six models that we have considered, which vary in terms of uncertainty, templates, and matching rule. Each model is labeled with a particular identifier for future reference.

### Simulating acuity performance

We simulated performance using Monte Carlo methods. First a specific model was defined, in terms of aberrations, matching rule, templates, uncertainty, and noise level. We then conducted individual trials, in which a test letter of a particular size was randomly selected, filtered by optical and neural transfer functions, and matched to the specified templates. The decision of the model was then judged as correct or incorrect. From trial to trial, we used a QUEST adaptive procedure (Watson & Pelli, [1983](#)) to vary the letter size. Letter size varied from LogMAR = −0.6 to 0.7 in steps of 0.05. One hundred twenty-eight trials were collected for each estimate of acuity.

Data were fit by a Weibull function with a  $\beta$  (slope) of 4 and a  $\gamma$  (lower asymptote) of 0.1 (Watson & Solomon, [1997](#)). The value of  $\beta = 4$  was determined from a preliminary simulation, as described in [Appendix D](#). From the fit, acuity was defined as the value of LogMAR at which probability correct was  $P = 0.66$ . An example of one simulation is shown in [Figure 5](#). This acuity estimation procedure was applied to each of the 67 wavefront aberration conditions used by Cheng, Bradley, et al. ([2004](#)).

This simulation is similar to the experimental method used by Cheng, Bradley, et al. ([2004](#)), although they used a method of constant stimuli with 10 trials at each of ten-letter sizes spaced 0.1 LogMAR apart and used a threshold criterion of  $P = 0.55$ .

### Estimating noise level

Performance of the acuity model is governed by the amount of neural noise  $\sigma_n$ . We explored various shortcuts to estimation of  $\sigma_n$  but ultimately determined that

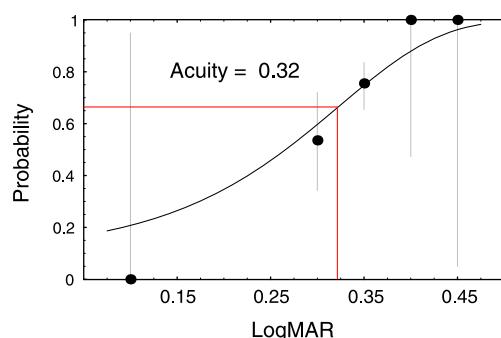


Figure 5. Results of QUEST staircase to estimate acuity for the ideal observer with no uncertainty ( $\sigma_u = 0$  deg). The noise level was  $\sigma_n = -1$  dB. The points indicate proportion correct; the vertical gray lines are 95% confidence limits based on binomial variance. The red lines indicate the estimate of acuity as the LogMAR for which probability correct is 0.66.

exhaustive testing of a range of alternative values was required. Thus, for each model, we tested a sequence of values of  $\sigma_n$  that bracketed the best fitting value.

## Fast method

For the special case of the ideal observer with zero spatial uncertainty (ID), a fast method of computing the simulations is available. This method can also be used to compute approximations to the other models.

Computing models that include spatial uncertainty means considering all possible integer pixel shifts of each template, which obliges us to generate a full noise image and to compute the full cross-correlation with each template on each trial of the simulation.

The calculations are greatly simplified if we consider only one possible shift for each template: either zero shift or that which yields the highest correlation with the neural image. This latter variant is not itself a plausible uncertain model since the observer could not know in advance what template shift to use, but it may nonetheless be a reasonable approximation, since the best shift with noise may frequently equal the optimal shift without noise. The details of how to compute this fast approximation are given in [Appendix C](#).

Note that this approximation is exact in one case: the zero shift variant corresponds exactly to the ideal observer with zero uncertainty. We have verified that in this case the two methods give nearly identical results (they do not give exactly equal results because both are Monte Carlo simulations).

We note also that the approximations for the ideal observer model are in fact identical to the approximations for the distance model since the two models differ only in their treatment of uncertainty.

For the matched filter models with uncertainty (IU, XA, DA), both approximations are very close to the full

simulation, but slightly more efficient, by about 0.5 dB for shifted templates and about 1.5 dB with unshifted templates.

For the unmatched filter models, the approximation is not as close. For XD, the shifted template approximation is about 5 dB more efficient, and the unshifted about 1 dB better still. The RMS errors for the approximation are about 0.02 larger.

## Results

### Comparing model and data

Previous evaluations of models have generally been content to report estimates of linear correlation between data and model. This measure is problematic for several reasons. First, it does not acknowledge the variance among observers, which necessarily affects and limits the maximum obtainable correlation. Second, correlation measures only the degree of linear covariation, not the absolute agreement between corresponding values of model and data. For example, despite a high correlation, the linear relation between data and model could have an intercept far from zero and a slope far from unity.

A better absolute measure of performance might be the RMS error between model and data. But here the presence of several observers with possibly different average acuities presents a problem. A solution in the present case is to optimize the one parameter of the model ( $\sigma_n$ ) separately for each observer. In [Figure 6](#), we plot for each model the RMS error for each observer as a function of  $\sigma_n$ . For all of these results, the frequency scale  $\phi = 1$ . Note that each point in this figure is based on 128 Monte-Carlo trials at each of 67 aberration conditions. From repeated measures of a number of the conditions (see model XD in [Figure 6](#)), we estimate the standard deviation of the RMS values to be 0.0024 LogMAR.

For each observer and model, a curve has been fit to the results to provide a continuous estimate of the minimum and the location of the minimum. These curves are shown in [Figure 6](#), along with the locations of the minima shown by colored arrows. This method of estimating the minimum RMS error for a given model and observer will be used in the remainder of this paper.

### Matched and unmatched templates

In [Figure 6](#), there is a clear difference in the error functions between the models for which the templates are the aberrated images (ID, IU, DA, XA) and those for which they are not (XD, XL). We will call these “matched” and “unmatched” templates, respectively. For the former, the error functions are nearly symmetrical

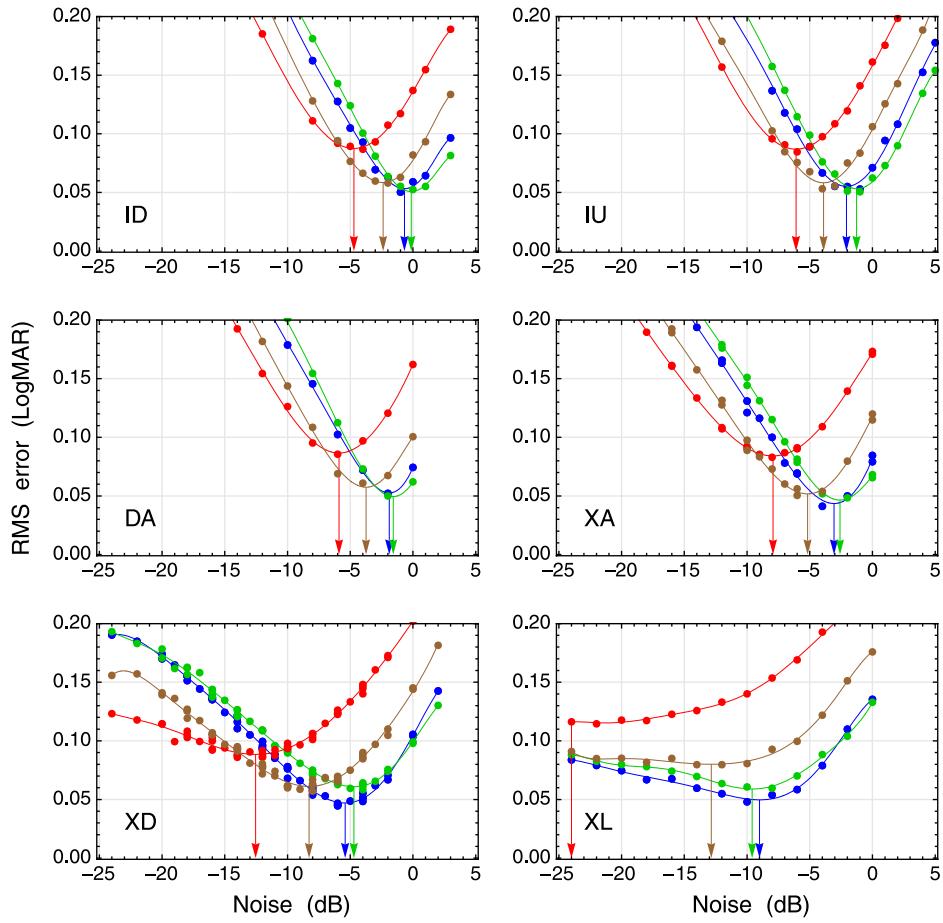


Figure 6. RMS error for each observer as a function of the neural noise  $\sigma_n$ . Observers are indicated by standard colors. A continuous curve has been fit to each set of points, and the location of the minimum is shown by an arrow. For all of these results, the frequency scale  $\phi = 1$ .

about a clear, deep minimum. For the latter, the curve is much more shallow to the left of the minimum. This difference may be due to the fact that, for unmatched templates, as noise tends towards zero, the matching rule is not guaranteed to converge to a correct decision. Examination of model predictions as noise is reduced shows that, especially for model XL and observer Red, performance asymptotes at about  $-16$  dB.

The partition of models into matched-template and unmatched-template classes is reinforced by the correlations between results for the various models, as shown in Figure 7. These are correlations between the LogMAR predictions from each model when at the noise level yielding the best fit to the data. Correlations within matched or unmatched groups are always above 0.96, while those between matched and unmatched range between 0.71 and 0.85 (the darker region in the table).

## Estimated noise

Considering a single curve in Figure 6 for one model and one observer, the minimum of the curve reflects the

quality of fit, while the location of the minimum reflects the best estimate of the noise parameter  $\sigma_n$ . In Figure 8, we plot the best estimates of  $\sigma_n$  as a function of model, with each observer again represented by their designated color.

	ID	DA	IU	XA	XD	XL
ID	0.98	0.98	0.97	0.8	0.71	
DA	0.98		0.98	0.97	0.81	0.73
IU	0.98	0.98		0.97	0.8	0.72
XA	0.97	0.97	0.97		0.85	0.78
XD	0.8	0.81	0.8	0.85		0.96
XL	0.71	0.73	0.72	0.78	0.96	

Figure 7. Correlation between models. The gray level in each cell is linear with the value.

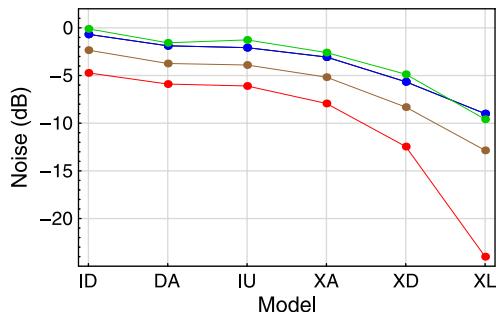


Figure 8. Estimates of noise parameter  $\sigma_n$  for each model and observer.

### Differences among observers

For a given model, it is clear that each observer has a different estimate of  $\sigma_n$ . In general, these estimates are ordered in agreement with the empirical differences in sensitivity: Red is the most acute observer, and Green and Blue are the least acute (see Figure 1).

### Differences among models

While a small estimated noise is symptomatic of an acute observer, it is also indicative of an inefficient matching rule and templates. Intuitively, we are able to add more noise and get the same performance, when a model is efficient. We can directly compare the efficiencies of the various models with a re-plotting of the results in Figure 8. In Figure 9, we plot the relative efficiency of each model compared to the ideal. These values are obtained from the square of the ratio of the estimated noise values for the two models compared. The four observers are very similar. The ideal observer with uncertainty (IU) and the distance model (DA) are about 74% as efficient as the ideal (ID), followed by cross-correlation rules with aberration (XA, 53%), diffraction-limited (XD, 27%), and letter templates (XL, 9%), respectively.

We should note that this estimate of efficiency is not a direct comparison of the performance of two models but is rather a comparison of two models that are fit to a

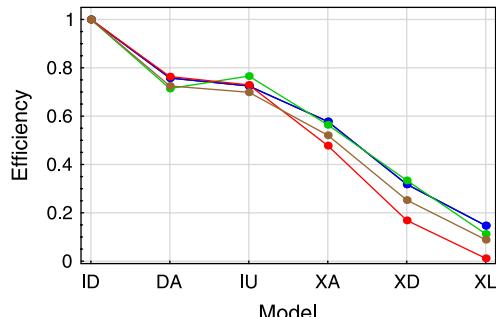


Figure 9. Relative efficiency of the models.

common set of human observer data. To the extent that the fits are good, the measure will approach a conventional efficiency.

### Effect of frequency scale

The results we have discussed so far have used the neural transfer function derived from a standard CSF (Figure 4). In light of the large variations in the quality of fit and acuity, among the observers, we considered whether some of this variation might be due to variations in the individual CSFs and, in particular, to the high-frequency fall-off of these functions. To manipulate this fall-off, we varied the frequency scale parameter  $\phi$  (see Frequency scale section). Initially, we explored this parameter with the ID model, using the fast method for which it is an exact solution. Each condition was simulated with 512 trials. For each model and observer, the results at a broad range of noise levels were analyzed as in Figure 6 to estimate the minimum RMS error. The results are shown in Figure 10.

The figure shows that the fit does indeed depend significantly on the frequency scale, and that the optimum is not at  $\phi = 1$  for any observer. For observers Red and Brown, the optimum is near to  $\phi = 2$ . For observers Green and Blue, the optimum is near to  $\phi = 1.3$ .

In light of these results, we completed additional simulations for the remaining models, using the slow but accurate method, at values of  $\phi = 1.5$  and 2. These results are shown in Figure 11 along with the earlier results at  $\phi = 1$  and the ID results from Figure 10. It should be noted that the ID results (black points) are more accurate since they are derived from 512 trials per condition, while the other model results are based on only 128 trials per condition.

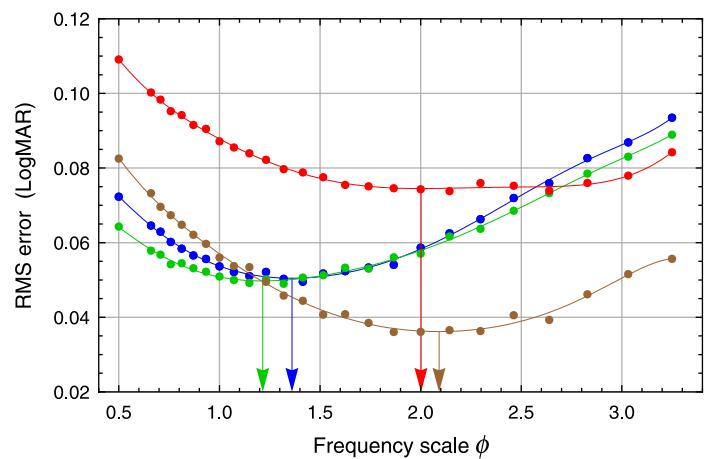


Figure 10. Fit of the ID model as a function of the frequency scale. The curves are interpolating polynomials. The arrows show the estimated minimum for each observer.

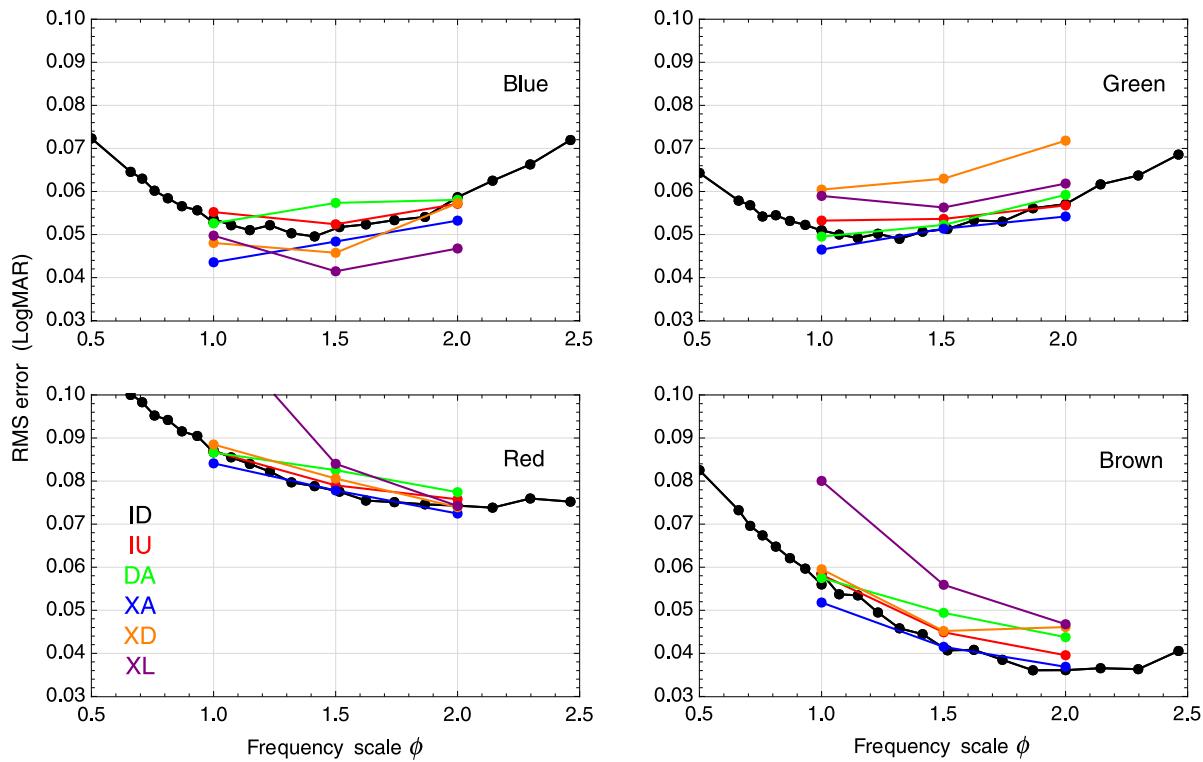


Figure 11. Fit of all models as a function of frequency scale. Each panel is for a different observer, indicated in the upper right. Each model is represented by a different color.

As a general observation, the other models mimic the behavior of the ID model with respect to the frequency scale  $\phi$ . The best fitting frequency scale appears to be approximately the same for all models.

Comparison of the red and the black points shows that spatial uncertainty has little effect on the fit of the ideal model.

For three of the observers (Red, Green, and Brown), the ID model is generally the best fit. For these observers, the other matched template models (IU, DA, XA) fit almost as well, while one or both of the unmatched template models (XD and XL) fit substantially worse. For one observer (Blue), there is an indication that the unmatched template models (XD and XL) may fit somewhat better. This may be a genuine difference in recognition strategy between observers.

Although it is not the absolute best fit for every observer, the ID provides a near-to-best fit for three of the four observers. To summarize the overall performance of that model, we consider its results for near-to-optimal frequency scales of  $\phi = 1.32$ , for observers Blue and Green and 2 for observers Red and Brown.

One view of the overall performance of a model is a plot of measured versus simulated values of LogMAR. This plot is given for the model ID, for the noted values of  $\phi$ , in Figure 12. The correlation between the two sets of values is 0.913. For comparison, the best correlation reported by Cheng, Bradley, et al. (2004) was 0.85. The total RMS is 0.056 LogMAR. The metrics considered by

Cheng, Bradley, et al. (2004) did not report RMS error since they did not attempt prediction of absolute LogMAR values. Correlation coefficients and RMS values for the group and for the individual observers are shown in Table 3.

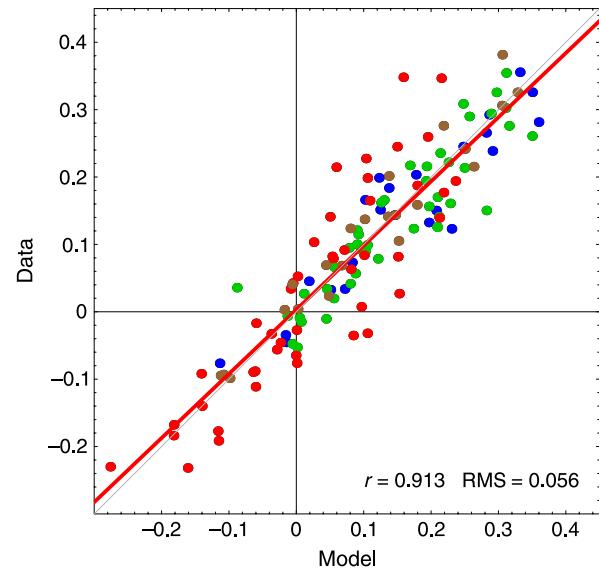


Figure 12. Data versus ID model. For each observer, the best-fitting value of noise  $\sigma_n$  and frequency scale  $\phi$  has been used. The red line is the best fitting straight line fit to the points by minimizing error in the data dimension. It has the form  $y = 0.003 + 0.952x$ . The observers are indicated by the standard colors.

Observer	$\phi$	RMS	$r$
Group		0.056	0.913
Blue	1.32	0.046	0.933
Green	1.32	0.048	0.905
Red	2	0.074	0.868
Brown	2	0.034	0.971

Table 3. Quality of fit for the ID model for the four observers and the group.

We can also plot the experimental data along with the simulated results, for each observer using the optimal value of noise and the near-optimal value of  $\phi$ . These results are pictured for the ID model in Figure 13, which shows graphically how well the model tracks the variations in acuity with aberration.

## Comments on model behavior

For the various matched template models, the best predictions (as in Figure 13) are very similar. Once equated for sensitivity to noise, they behave almost identically. As a group, their predictions are generally flatter (show less effect of aberrations) than the data and shift vertically with variations in noise. In other words, LogMAR acuity is the same function of noise at every aberration condition.

For the unmatched templates, the prediction curves tend to be deeper (show more of an effect of aberration) than the matched template models, and in addition, the depth is a function of noise. This is illustrated in Figure 14, which shows XD predictions for the same aberration conditions at two noise levels: -24 and 0 dB. At the low noise level,

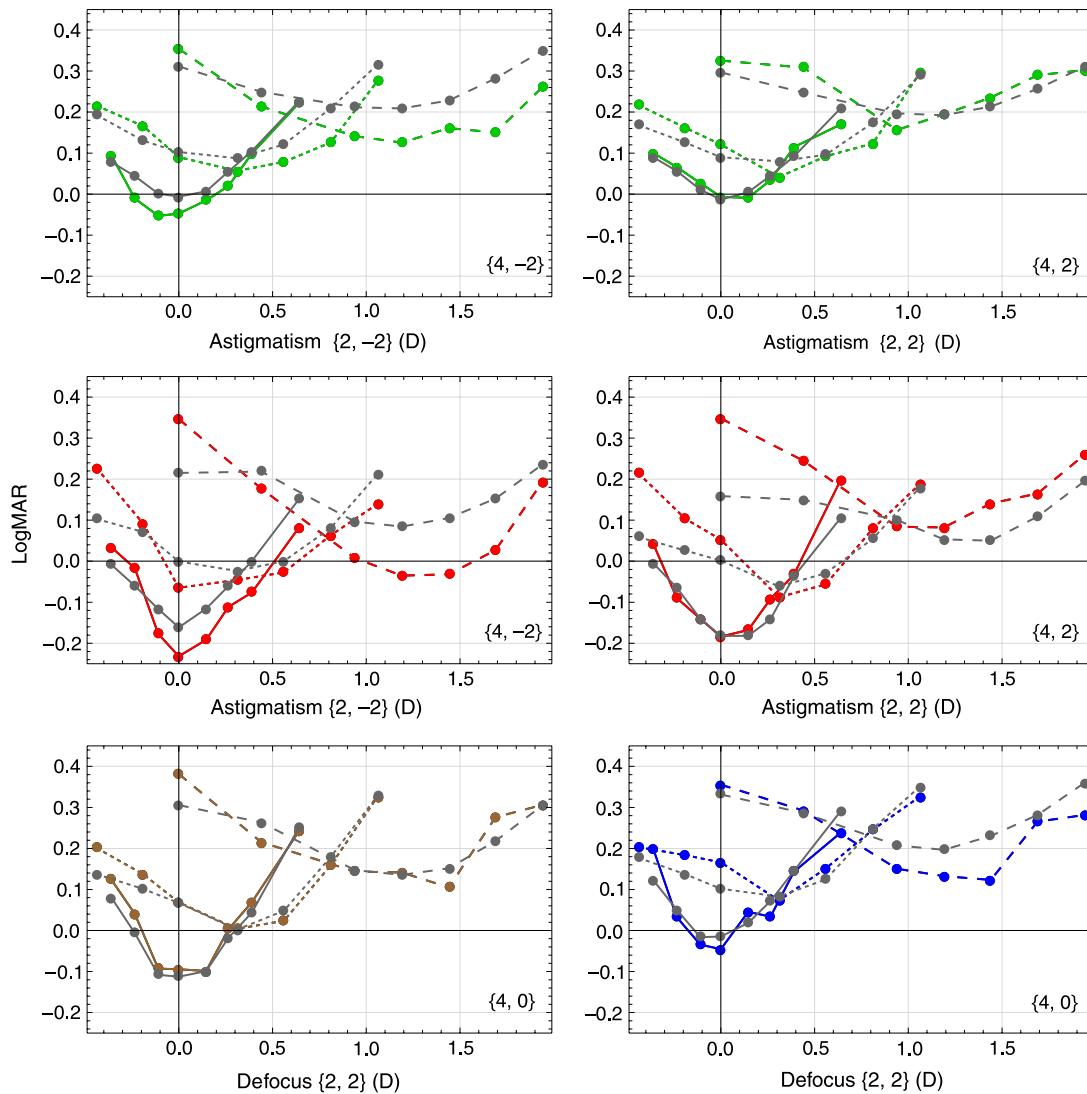


Figure 13. Data (color) and model (gray) for near best fitting values of noise and frequency scale  $\phi$  for each observer for the ID model. Total RMS error is 0.056 LogMAR.

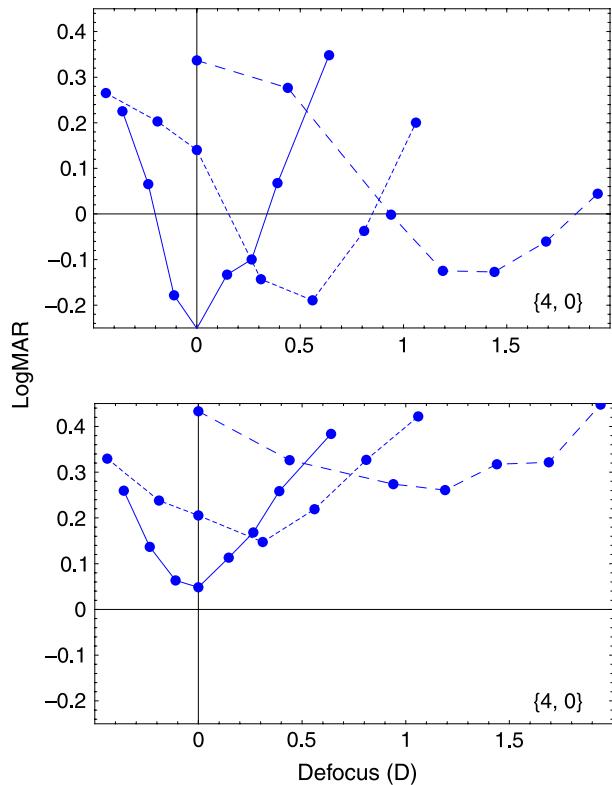


Figure 14. Predictions of XD model for defocus conditions at noise levels of  $-24$  (top) and  $0$  dB (bottom).

the predictions for defocus-only (solid line) cover a range of about  $0.6$  LogMAR, while at the high noise level, the range is about  $0.33$ . This deepening effect does not occur for model XL, perhaps because of the asymptotic behavior at low noise levels discussed above.

The fact that the XD model shows a greater effect of aberration is perhaps not surprising. In that model, as aberrations are added or intensified, the signals received depart more from the templates. In contrast, for the matched template models, increased aberration reduces performance only because some components of both signal and template become lost in the noise.

## Acuity metric

As noted above, we define a metric as a simple calculation that mimics the data to a reasonable degree, and that may have practical value. Here we describe the design of a metric that transforms an arbitrary wavefront aberration into a LogMAR acuity value. In describing this design, it should be noted that we are providing only a recipe, and some intuition, not a formal model.

The starting point for this metric is the set of neural images generated for the ten Sloan letters of a particular

size by a particular aberration. These can be computed using the methods described in the [Acuity model](#) section. We write these neural images  $s_j$ , where  $j$  indexes the individual Sloan letter. We then consider a matrix  $r_{j,k}$  consisting of the dot products of each neural image with each other. We normalize these dot products by the modulus of each neural image,

$$r_{j,k} = \frac{s_j \cdot s_k}{\|s_k\|}. \quad (7)$$

These normalized values are akin to the values used by the normalized correlation model, or the ideal observer model, when spatial uncertainty is absent. We imagine that each value is perturbed by uncorrelated Gaussian noise with standard deviation  $\sigma$  (in fact, if the noise is derived from noise at the input, there would be correlations, but we ignore that here for simplicity). If the letter with index  $j$  is presented, then the observer would select the entry in the  $j$ th row that is largest. The probability that the correct column is selected is equal to the probability that its entry is larger than each of the incorrect entries. To compute this probability, it is useful to first compute the difference between each column entry ( $r_{j,k}$ ) and the one corresponding to the correct answer ( $r_{j,j}$ ) and to divide these by the standard deviation  $\sigma$ ,

$$d_{j,k} = \frac{r_{j,k} - r_{j,j}}{\sigma}. \quad (8)$$

Then the probability of being correct can be written

$$P_j = \int_{-\infty}^{\infty} f(x) \prod_{k \neq j} F(x + d_{j,k}) dx, \quad (9)$$

where  $f$  and  $F$  are the normal density and distribution functions, respectively.

The logic behind this expression is that we compute, for a value  $x$ , the probability that the correct entry equals  $x$ , and that all the other entries are less than  $x$ . We then integrate this over all possible values of  $x$ . We note that in practice the integral may be taken over the range  $\{-3, 3\}$  without great loss in accuracy. This calculation is illustrated in [Figure 15](#).

We also note that another approximation is available. Consider the individual probabilities of each incorrect entry being less than the correct entry. If these were independent, then the probability of being correct would just be their product:

$$\hat{P}_j = \prod_{k \neq j} F(d_{j,k}). \quad (10)$$

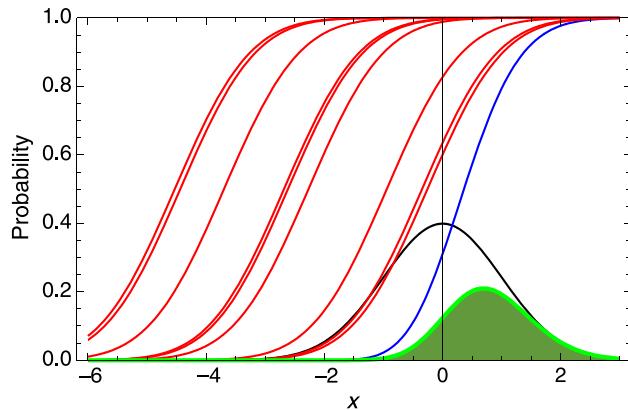


Figure 15. Illustration of the acuity metric calculation of the probability of a correct response to the presentation of one letter ([Equation 9](#)). The black curve is the Normal probability density  $f(x)$ , corresponding to the probability distribution of the cross-correlation between the neural image of the letter presented and its corresponding correct template. The red curves are the normal probability cumulative distribution functions, shifted to the locations of the cross-correlations between the neural image and the nine incorrect templates  $F(x + d_{jk})$ . The product of the red curves is shown by the blue curve, corresponding to the probability that none of the incorrect templates yields a cross-correlation greater than the specified value. The green curve is the product of black and blue curves. Its integral is the probability correct—in this example,  $P_{jk} = 0.38$ .

Even though these probabilities are certainly not independent, this approximation is reasonably close to the more accurate formula in [Equation 9](#).

Finally, the predicted percent correct in the acuity experiment is the average over all rows (letters),

$$P = \bar{P}_j. \quad (11)$$

Note that this quantity is a function of the letter size. We can compute these values for the range of sizes used in the experiment to produce a psychometric function, yielding probability correct versus LogMAR, as shown in [Figure 16](#).

In the actual experiment, thresholds were defined as the LogMAR value at which the probability correct was 0.55. Accordingly, for each aberration, the LogMAR value yielding the criterion of 0.55 is read from the psychometric function. This procedure is illustrated by the red lines in [Figure 16](#) for the particular case of zero aberration.

To evaluate this metric, we have applied it to all of the data of Cheng, Bradley, et al. (2004) discussed earlier in this paper. We explored variations in the frequency scale  $\phi$  and the standard deviation  $\sigma$ . We observed that a range of parameter values gave good fits. The best fit was obtained at  $\phi = 2$ ,  $\sigma = 0.38$ , at which point the RMS

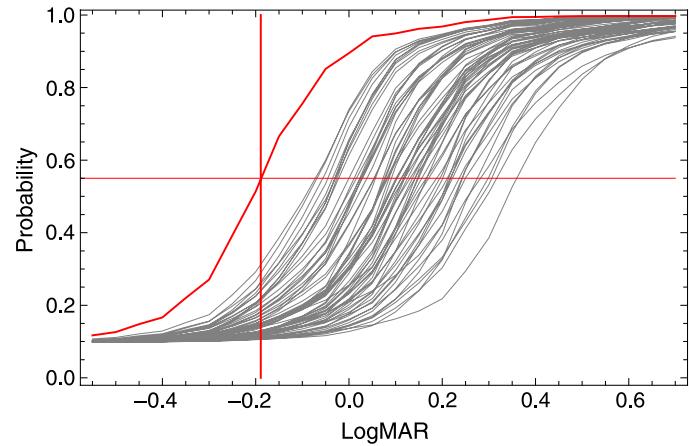


Figure 16. Probability correct versus LogMAR computed by the acuity metric. Each curve is for one aberration. The red curve is for the zero aberration condition, and the red lines illustrate estimation, for that aberration, of the LogMAR required to yield 0.55 correct. Other parameters:  $\phi = 2$ ,  $\sigma = 0.38$ .

error was 0.07. The RMS error is plotted against frequency scale in [Figure 17](#). In this figure, each point may have a different value of the noise parameter  $\sigma$ . Note that the optimal value of the frequency scale  $\phi$  is roughly the same as that for our two more acute observers (Red and Brown), as estimated by the model.

A variant of this metric is one in which we begin not with the dot products of pairs of unshifted neural images, but rather with the largest dot product that can be obtained through some shift of one of the images. This is analogous to spatial uncertainty. In [Figure 17](#) we also show, in red, the fit of this variant. The performance is similar, but not quite as good as that for the standard version.

The best version of the metric yielded a correlation of 0.86. We note that other metric parameter values yield

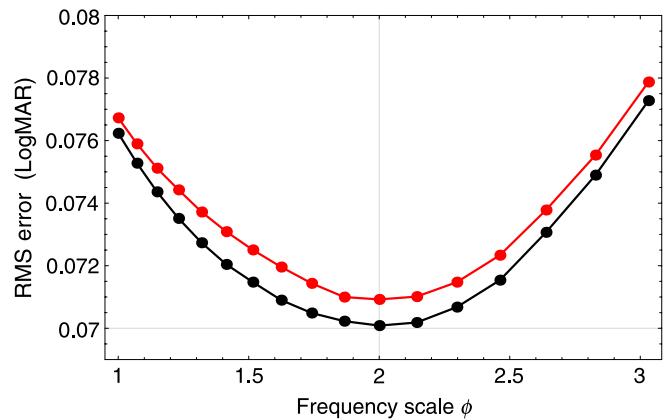


Figure 17. Minimum RMS error as a function of frequency scale for fits of the acuity metric. The black points are the standard version of the metric; the red points are a version with optimally shifted images.

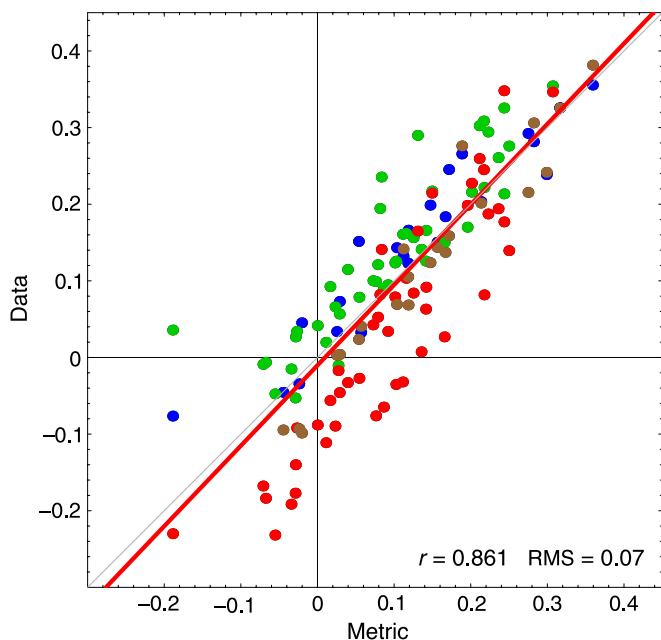


Figure 18. Correlation between metric and data. Observers are indicated by standard colors. The RMS error is 0.07.

correlations as high as 0.874, but we believe RMS error is the better measure of the performance of the metric. The predictions of this metric are shown in Figure 18.

We note that the fit of this metric is slightly poorer than that of the model, but this is due in part to the fact that the model allowed different frequency scales and different noise values for each observer. Even so, it is worth noting that for individual observers, the metric yields correlations of 0.95, 0.88, 0.91, and 0.97.

To allow others to experiment with our metric, we plan to provide a portable version as described in Appendix F.

## Discussion

In this study, we evaluated the fit of a set of template-matching models, incorporating both optical and neural filtering, to a set of acuity versus aberration data. In general, the class of models we have considered fit very well. Relative to prior metrics, which attained a correlation of 0.85 (Cheng, Bradley, et al., 2004), the best model here attains a correlation of 0.913, and indeed all models tested (including our metric) attain correlations of over 0.86. The best RMS error we have obtained is about 0.056 LogMAR. We cannot directly compare this RMS value to prior fitting results as those metrics did not attempt to generate absolute LogMAR predictions, which are required for the RMS calculations.

The success of the ID model means that we now have a successful operational account of how particular

aberrations reduce visual acuity. This account may prove useful in future study of the effects of specific aberrations.

In terms of RMS error, the ID model appears to provide the best overall fit, but models IU, XA, and DA fit almost as well. The unmatched template models XD and XL fit slightly or appreciably less well for three of the four observers but better than ID for one observer (Blue). This discrepancy may reflect a genuine difference in strategy among observers. Note that the matched template models assume that the observer has “learned” the aberration, so that they are able to construct accurate aberrated templates. Elsewhere, strong evidence has been provided that observers do adapt to their own distinctive aberrations (Artal, Guiarao, Berrio, & Williams, 2001). Perhaps observer Blue has not learned the experimental aberrations as well as the other observers.

One of the observers (Red) had an RMS error twice as large as the best fit observer. Observer Red was also characterized by high acuity and a more pronounced effect of aberration on acuity. The differences in fit among observers were quite general across models. We do not yet fully understand why this particular observer was less well fit by our models.

The neural transfer functions that fit best differed among our observers. Two observers (Red and Brown) required a function that passed considerably higher frequencies ( $\phi = 2$ ) than for the other two observers ( $\phi = 1.32$ ). Both sets of observers were more acute than our “standard” observer ( $\phi = 1$ ). It is perhaps not surprising that observers should differ in this regard, given the known variations in cone density (Roorda & Williams, 1999). With respect to the standard observer, it was derived from a population of 16 observers of unknown age and with uncertain optical aberrations (Watson & Ahumada, 2005). We must also acknowledge that the data from Cheng, Bradley, et al. (2004) used here come from only four observers, and we do not know where they lie relative to the larger population.

In contrast to the metrics described by Cheng, Bradley, et al. (2004), our models, and indeed our metric, are process models that begin with a stimulus image and proceed to a letter identification. Their success shows that such process models and metrics can in fact provide a good account of the data. The virtues of process models here are threefold: (1) they assist in understanding the mechanisms that may operate in human performance, (2) they can be applied to new types of aberrations, and (3) they can be applied to completely new sets of optotypes or indeed different identification tasks altogether.

Another process model applied to an acuity task was described by Nestares et al. (Nestares, Navarro, & Antona, 2003). While similar in many respects to the present effort, their model was considerably more complex, involving retinal sampling, a pyramid decomposition into Gabor channels, and identification based on Bayesian inference on the channel outputs. They also report that simpler models (like those employed here) fail to simulate the effects of various optical degradations, which appears

to be contrary to our results. While we certainly cannot rule out a role for spatial frequency channels in human letter identification, from a practical point of view it is useful to know that they are not required to obtain good predictions.

Spatial uncertainty may be required in a model for two reasons. The first is that the observer may be uncertain. The second is that if the templates are unmatched, then the aberration may cause a spatial shift in the letter, and the template must be shifted to obtain the best match. For matched templates, this extra shifting is not required because both signals and templates are shifted equally. In one pair of models (ID and IU), we directly tested the effect of including spatial uncertainty. The two models give very similar results (except for a modest difference in efficiency), which suggests that for matched template models, spatial uncertainty does not have a significant effect.

Although our models fit quite well, systematic deviations from the data remain, suggesting room for improvement. Among the plausible features that we have not included are imperfect or noisy templates (McIlhagga & Pääkkönen, 1999).

In the [Introduction](#) section, we discussed the difference between models and metrics. One goal of this project was evaluate various models in order to lay the groundwork for a useful metric. We found that a metric based on the XA model worked quite well, yielding an RMS error of 0.07. Because this metric is derived from a plausible model, and because it depends upon the actual letter images rendered by a particular aberration, we expect it to be much more robust to changes in aberration type or changes in optotype than prior metrics, which were fit only to a particular set of aberrations and did not include either a process model or the actual acuity optotypes.

## Conclusions

An ideal observer model that incorporates both optical and neural transfer functions provides an excellent account of acuity in the presence of various aberration mixtures.

Other template-matching models work almost as well. For one observer, a model with templates consisting of diffraction-limited letters or the letters themselves performed better than the ideal observer model.

The contrast sensitivity function defined for the so-called Spatial Standard Observer (Watson & Ahumada, 2005) provides a successful starting point for the construction of the neural transfer function. However, all observers required that this function be shifted to higher frequencies to obtain the best fit.

A simple metric, based on the ideal observer or normalized correlation model but not requiring Monte Carlo simulations, also provides a good prediction of the data.

## Appendix A: File of letter images

With this report, we have provided a file called [SloanLetterImages](#) that contains all of the Sloan letter images used in this project. The letter images are arranged in a sequence in which letters vary first alphabetically, then by size. Each image is 256 × 256 pixels, with a nominal resolution of 313.91 pixels/degree, thus subtending 0.815 degrees. The letter sizes vary from LogMAR of −0.6 to 0.7 in steps of 0.05, for a total of 27 sizes. Within each image, each pixel is represented by a single byte representing an unsigned integer, and the ordering of bytes with respect to pixels is left to right, top to bottom. The images are provided in a single binary file of  $10 * 27 * 256 * 256 = 17,694,720$  bytes.

In Mathematica, the images can be read in using the expression:

```
images =
  Fold [Partition,
    Import["SloanLetterImages", "Byte"], {256, 256, 10}]
```

In MatLab, the images can be read in using the expression:

```
imfid = fopen('SloanLetterImages', 'r');
images = fread(imfid); size(images);% 17694720 1
n = 256; nsize = 27; nlet = 10;
images = reshape(images, [n n nlet nsize]);
images = permute(images(:,end:-1:1,:,:,[2 1 3 4]));
image (images(:,:,3,27)); colormap(gray(256));axis
image;
```

## Appendix B: Minimum distance

Here we derive the discriminant for the minimum distance model. The noisy neural image due to presentation of letter indexed by  $k$  is

$$sn(\mathbf{x}) = s_k(\mathbf{x}) + n(\mathbf{x}). \quad (B1)$$

The distance between template displaced by  $\mathbf{v}$  and the noisy neural image can be written

$$d_{j,v} = \| t_{j,v}(\mathbf{x}) - sn(\mathbf{x}) \|^2. \quad (B2)$$

Expanding the squared modulus and recognizing an expression for cross correlation, we have

$$d_{j,v} = \| t_j(\mathbf{x}) \|^2 + \| sn(\mathbf{x}) \|^2 - t_j(\mathbf{x}) \otimes sn(\mathbf{x}). \quad (B3)$$

The minimum of this quantity over  $v$  is given by

$$\max_{\text{over } v} [t_j(\mathbf{x}) \otimes sn(\mathbf{x})] - \frac{1}{2} \| t_j(\mathbf{x}) \|^2 - \frac{1}{2} \| sn(\mathbf{x}) \|^2. \quad (B4)$$

But the final term is constant for all  $j$ , so it is sufficient to use the discriminant

$$\begin{aligned} g_j &= \max_{\text{over } v} [t_j(\mathbf{x}) \otimes sn(\mathbf{x})] - \frac{1}{2} \| t_j(\mathbf{x}) \|^2 \\ &= \max_{\text{over } v} [t_j(\mathbf{x}) \otimes (s_k(\mathbf{x}) + n(\mathbf{x}))] - \frac{1}{2} \| t_j(\mathbf{x}) \|^2 \end{aligned} \quad (B5)$$

## Appendix C: Fast simulation method

The acuity model simulations may be greatly accelerated by assuming in separate simulations, one of two possible positions for each template: either the unshifted template or the version of each template shifted to optimally match the letter presented. In this case, each of the model discriminants is based on the simpler quantity

$$c_j = t_{j,k}(\mathbf{x}) \cdot s_k(\mathbf{x}) + t_{j,k}(\mathbf{x}) \cdot n(\mathbf{x}), \quad (C1)$$

where  $\cdot$  indicates the dot product, and  $t_{j,k}(\mathbf{x})$  indicates template  $j$  unshifted or shifted optimally for neural image  $k$ . Here it is not necessary to compute the actual noise image  $n(\mathbf{x})$  or its dot product with the template; instead we create  $J$  random deviates that will have the same distribution as the results of that dot product. Specifically, consider the matrix  $\mathbf{C}$  consisting of the correlations of each neural image with each shifted template. Then consider the SVD of  $\mathbf{C}$

$$\mathbf{U} \mathbf{V} \mathbf{W}^* = \mathbf{C}, \quad (C2)$$

where  $*$  indicates the Hermitian transpose.  $\mathbf{V}$  is a diagonal matrix which contains the singular values. Then we compute the matrix  $\mathbf{M}$

$$\mathbf{M} = \mathbf{U} \sqrt{\mathbf{V}}, \quad (C3)$$

and finally the vector  $\mathbf{m}$  of  $J$  noise samples

$$\mathbf{m} = \mathbf{M} \mathbf{n}. \quad (C4)$$

Then we can write

$$c_j = t_{j,k}(\mathbf{x}) \cdot s_k(\mathbf{x}) + m_j. \quad (C5)$$

For a given set of templates, the matrix  $\mathbf{C}$  can be computed in advance for each letter size and each aberration condition.

## Appendix D: Slope of psychometric function

In our simulations, we have used a QUEST psychometric procedure and have fit the resulting data with a Weibull function. In both steps, it is advantageous to know the approximate value of the slope  $\beta$  of the psychometric function. To discover this, we conducted 1024 simulated trials of the ideal observer model with zero uncertainty (ID) on the diffraction-limited condition ( $Z = \{ \}$ ). We used a version of QUEST that scatters the trials somewhat about threshold. Strength is measured in units of LogMAR/20. The Weibull fit yields  $\beta = 3.99$ . We have conducted additional simulations for other conditions and uncertainties, with similar results ( $\beta$  between 3 and 5). In subsequent simulations, we have used a value of  $\beta = 4$  in the QUEST procedure and in subsequent fitting of the data.

## Appendix E: Notation

$g_k$	discriminant for letter $_k$
$t_k(\mathbf{x})$	template for letter $_k$
$\bar{t}_k(\mathbf{x})$	normalized template for letter $_k$
$sn(\mathbf{x})$	noisy neural image
$\otimes$	cross correlation operator
$n(\mathbf{x})$	noise image
$s_k(\mathbf{x})$	neural image for letter $_k$
$\sigma$	neural noise standard deviation
$u(\mathbf{x})$	spatial uncertainty
$\phi$	frequency scale
$f$	Normal probability density
$F$	Normal probability distribution

## Appendix F: WFAMetric

WFAMetric is a computer program to compute acuity from a wavefront aberration, defined by a set of Zernike coefficients, using the formulas provided in the Acuity

**metric** section. We are developing a portable version of this program that we hope to share with the research community. When and if we are able to do so, the program will be available at <http://vision.arc.nasa.gov/projects/mfametric/>.

## Acknowledgments

This research was supported by NASA Grants 21-131-50-20-04 and 21-131-50-10-08 and by FAA Grant 21R-711-80-1129-01. We are grateful to X. Cheng, A. Bradley, and L. Thibos for generously sharing their data.

Commercial relationships: ABW has a commercial interest in the Spatial Standard Observer.

Corresponding author: Andrew B. Watson.

Email: [andrew.b.watson@nasa.gov](mailto:andrew.b.watson@nasa.gov).

Address: MS 262-2 NASA Ames Research Center, Moffett Field, CA 94035-1000, USA.

## Footnote

<sup>1</sup>Cheng, Bradley, et al. (2004) computed images corresponding to a 5-mm pupil but had the observers view them through a 2.5-mm pupil in order to minimize intrusion of the observer's own aberrations. They state that they compensated the calculated images for the effects of the 2.5-mm pupil. However, the 2.5-mm pupil zeros frequencies beyond a limit of about 78.5 cycles/deg (half the limit passed by the 5-mm pupil), and for these frequencies no compensation is possible. Thus, the images we compute contain energy above 78.5 cycles/deg that was not present in the images seen by the Cheng, Bradley, et al. (2004) observers. However, this discrepant energy is always less than 0.5% of the contrast energy of each letter and is effectively removed by our neural transfer function, which at this frequency limit is less than 1% of its maximum.

## References

- American National Standards Institute (ANSI). (2004). *American national standard for ophthalmics—Methods for reporting optical aberrations of eyes* (vol. ANSI Z80.28). Washington, D. C.
- Applegate, R. A., Marsack, J. D., Ramos, R., & Sarver, E. J. (2003). Interaction between aberrations to improve or reduce visual performance. *Journal of Cataract and Refractive Surgery*, 29, 1487–1495. [PubMed]
- Artal, P. (1990). Calculations of two-dimensional foveal retinal images in real eyes. *Journal of the Optical Society of America A, Optics and Image Science*, 7, 1374–1381. [PubMed]
- Artal, P., Guirao, A., Berrio, E., & Williams, D. R. (2001). Compensation of corneal aberrations by the internal optics in the human eye. *Journal of Vision*, 1(1):1, 1–8, <http://journalofvision.org/1/1/1/>, doi:10.1167/1.1.1. [PubMed] [Article]
- Bouma, H. (1971). Visual recognition of isolated lower-case letters. *Vision Research*, 11, 459–474. [PubMed]
- Campbell, F. W., & Green, D. G. (1965). Optical and retinal factors affecting visual resolution. *Journal of Physiology*, 181, 576–593. [PubMed] [Article]
- Cheng, H., Barnett, J. K., Vilupuru, A. S., Marsack, J. D., Kasthurirangan, S., Applegate, R. A., et al. (2004). A population study on changes in wave aberrations with accommodation. *Journal of Vision*, 4(4):3, 272–280, <http://journalofvision.org/4/4/3/>, doi:10.1167/4.4.3. [PubMed] [Article]
- Cheng, X., Bradley, A., & Thibos, L. N. (2004). Predicting subjective judgment of best focus with objective image quality metrics. *Journal of Vision*, 4(4):7, 310–321, <http://journalofvision.org/4/4/7/>, doi:10.1167/4.4.7. [PubMed] [Article]
- Cheng, X., Himebaugh, N. L., Kollbaum, P. S., Thibos, L. N., & Bradley, A. (2004). Test-retest reliability of clinical Shack-Hartmann measurements. *Investigative Ophthalmology & Visual Science*, 45, 351–360. [PubMed] [Article]
- Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: John Wiley.
- Gervais, M. J., Harvey, L. O., Jr., & Roberts, J. O. (1984). Identification confusions among letters of the alphabet. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 655–666. [PubMed]
- Gibson, E. J., Osser, H., Schiff, W., & Smith, J. (1963). An analysis of critical features of letters, tested by a confusion matrix. In *A basic program on reading* (Cooperative Research Program No. 639). Office of Education.
- Guirao, A., & Williams, D. R. (2003). A method to predict refractive errors from wave aberration data. *Optometry and Vision Science*, 80, 36–42. [PubMed]
- IJspeert, J. K., van den Berg, T. J., & Spekreijse, H. (1993). An improved mathematical description of the foveal visual point spread function with parameters for age, pupil size and pigmentation. *Vision Research*, 33, 15–20. [PubMed]
- Marsack, J. D., Thibos, L. N., & Applegate, R. A. (2004). Metrics of optical quality derived from wave aberrations predict visual performance. *Journal of Vision*, 4(4):8, 322–328, <http://journalofvision.org/4/4/8/>, doi:10.1167/4.4.8. [PubMed] [Article]

- McIlhagga, W., & Pääkkönen, A. (1999). Noisy templates explain area summation. *Vision Research*, 39, 367–372. [\[PubMed\]](#)
- 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. [\[PubMed\]](#)
- Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research*, 31, 1399–1415. [\[PubMed\]](#)
- Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, 4(12):12, 1136–1169, <http://journalofvision.org/4/12/12/>, doi:10.1167/4.12.12. [\[PubMed\]](#) [\[Article\]](#)
- Roorda, A., & Williams, D. R. (1999). The arrangement of the three cone classes in the living human eye. *Nature*, 397, 520–522. [\[PubMed\]](#)
- Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, 369, 395–397. [\[PubMed\]](#)
- Thibos, L. N., Applegate, R. A., Schwiegerling, J. T., & Webb, R. (2002). Standards for reporting the optical aberrations of eyes. *Journal of Refractive Surgery*, 18, S652–S660. [\[PubMed\]](#)
- Thibos, L. N., Hong, X., Bradley, A., & Applegate, R. A. (2004). Accuracy and precision of objective refraction from wavefront aberrations. *Journal of Vision*, 4(4):9, 329–351, <http://journalofvision.org/4/4/9/>, doi:10.1167/4.4.9. [\[PubMed\]](#) [\[Article\]](#)
- Thibos, L. N., Hong, X., Bradley, A., & Cheng, X. (2002). Statistical variation of aberration structure and image quality in a normal population of healthy eyes. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 19, 2329–2348. [\[PubMed\]](#)
- 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/5/9/6/>, doi:10.1167/5.9.6. [\[PubMed\]](#) [\[Article\]](#)
- 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. [\[PubMed\]](#)
- Watson, A. B., & Solomon, J. A. (1997). Psychophysica: Mathematica notebooks for psychophysical experiments. *Spatial Vision*, 10, 447–466. [\[PubMed\]](#)
- Wolfram, S. (2003). *The mathematica book* (5th ed.). Champaign, IL: Wolfram Media.