

Modeling acuity for optotypes varying in complexity

Andrew B. Watson

NASA Ames Research Center, Moffett Field, CA, USA



Albert J. Ahumada

NASA Ames Research Center, Moffett Field, CA, USA



Watson and Ahumada (2008) described a template model of visual acuity based on an ideal-observer limited by optical filtering, neural filtering, and noise. They computed predictions for selected optotypes and optical aberrations. Here we compare this model's predictions to acuity data for six human observers, each viewing seven different optotype sets, consisting of one set of Sloan letters and six sets of Chinese characters, differing in complexity (Zhang, Zhang, Xue, Liu, & Yu, 2007). Since optical aberrations for the six observers were unknown, we constructed 200 model observers using aberrations collected from 200 normal human eyes (Thibos, Hong, Bradley, & Cheng, 2002). For each condition (observer, optotype set, model observer) we estimated the model noise required to match the data. Expressed as efficiency, performance for Chinese characters was 1.4 to 2.7 times lower than for Sloan letters. Efficiency was weakly and inversely related to perimeter complexity of optotype set. We also compared confusion matrices for human and model observers. Correlations for off-diagonal elements ranged from 0.5 to 0.8 for different sets, and the average correlation for the template model was superior to a geometrical moment model with a comparable number of parameters (Liu, Klein, Xue, Zhang, & Yu, 2009). The template model performed well overall. Estimated psychometric function slopes matched the data, and noise estimates agreed roughly with those obtained independently from contrast sensitivity to Gabor targets. For optotypes of low complexity, the model accurately predicted relative performance. This suggests the model may be used to compare acuities measured with different sets of simple optotypes.

Keywords: acuity, ideal observer, optotypes, Chinese characters, Sloan letters

Citation: Watson, A. B., & Ahumada, A. J. (2012). Modeling acuity for optotypes varying in complexity. *Journal of Vision* 12(10):19, 1–19, <http://www.journalofvision.org/content/12/10/19>, doi:10.1167/12.10.19.

Introduction

Visual acuity is generally regarded as a test of the spatial resolution of the eye, involving optical, neural, and cognitive components. Until recently, however, the precise manner in which these components combined to yield a particular acuity was not known. Advances in measurement of the aberrations of the eye (Charman, 2005) have made it possible to calculate the retinal image produced by an arbitrary acuity target (Artal, 1990). Starting from this image, several authors have proposed models of the complete acuity task (Beckmann & Legge, 1996; Dalimier & Dainty, 2008; Dalimier, Pailos, Rivera, & Navarro, 2009; Nestares, Navarro, & Antona, 2003; Watson & Ahumada, 2008). All of these models include optical filtering of the optotype targets, some form of neural processing, noise, and a final template matching operation to identify the target. Beckmann and Legge (1996) used a monochromatic point-spread function formula of Navarro, Artal, and Williams (1993) followed by sampling by the cones and an ideal observer limited by Poisson receptor noise. Nestares et al. (2003) included retinal sampling, cortical filtering and sampling, and a process of Bayesian estimation of the

pattern templates. They predicted effects of defocus for a single observer, with moderate success. Watson and Ahumada (2008) proposed a simpler model in which sampling effects are neglected, neural processing is simulated by a single filter, and the templates are assumed to be the “neural images” produced by optical and neural filtering of the optotypes. They simulated results from Cheng, Bradley, and Thibos (2004) for four observers viewing acuity targets through a set of 67 distinct aberration conditions (combinations of defocus, astigmatism, and spherical aberration), and found good agreement with measured acuity data. Using a similar model, Dalimier and Dainty (2008) successfully predicted effects of higher-order aberrations as a function of light level. And in a more recent report, Dalimier et al. (2009) employed a model like that of Nestares et al. (2003) to predict effects of defocus on visual acuity in 11 eyes from which monochromatic wavefront aberrations had been measured. They found excellent agreement between average data and average predictions.

In this report we investigate the ability of the model of Watson and Ahumada (2008) to account for variations in acuity that result from changes of optotype. There are many different sets of acuity optotypes in wide use (Bailey & Lovie, 1980; Ferris &

Bailey, 1996; Moutray, Williams, & Jackson, 2008; Williams, Moutray, & Jackson, 2008). These include Sloan letters, the Landoldt C, the Snellen E, numbers in various fonts, pictograms, and letters in non-Roman alphabets. It would be useful to understand, from a theoretical point of view, any variations in acuity that result from the optotype selection (Dobson, Maguire, Orel-Bixler, Quinn, Ying; Vision in Preschoolers (VIP) Study Group, 2003; Grimm, Rassow, Wesemann, Saur, & Hilz, 1994; Pointer, 2008). There are, at present, no principled methods of translating between acuities measured with different optotypes. Empirical calibration is possible (Jackson & Bailey, 2009) but suffers from the need for a suitably large population of appropriately selected and corrected observers. More generally, one would like measurements of acuity that are independent of optotype. Beyond these practical goals, we were interested in whether the model was capable of predicting variations in acuity with optotype. This is a useful test of the generality of the model.

Acuity data

The data that we simulate here consist of letter identifications of English and Chinese characters by six observers, fluent in both languages. They were collected in the course of another study (Zhang, Zhang, Xue, Liu, & Yu, 2007). In that study the Chinese characters were further divided into six sets, based on the computed complexity of the individual characters. Because the data address a broad range of character complexity, and because the data were all collected from the same set of observers, they provide a unique opportunity for the test of acuity models.

Aberration data

Zhang et al. (2007) did not measure the wavefront aberrations of the six observers, so we cannot directly simulate the optical part of their performance. Instead we have made use of wavefront aberration measurements collected from a large population of normal human observers (Thibos, Hong, Bradley, & Cheng, 2002). In this way, a population of behaviors can be simulated for each observer, and statistics computed from that population.

Acuity model

Our acuity model has been described in detail elsewhere (Watson & Ahumada, 2008); here we provide a brief review, illustrated in Figure 1. In the acuity task that we model, an optotype is selected from a fixed set,

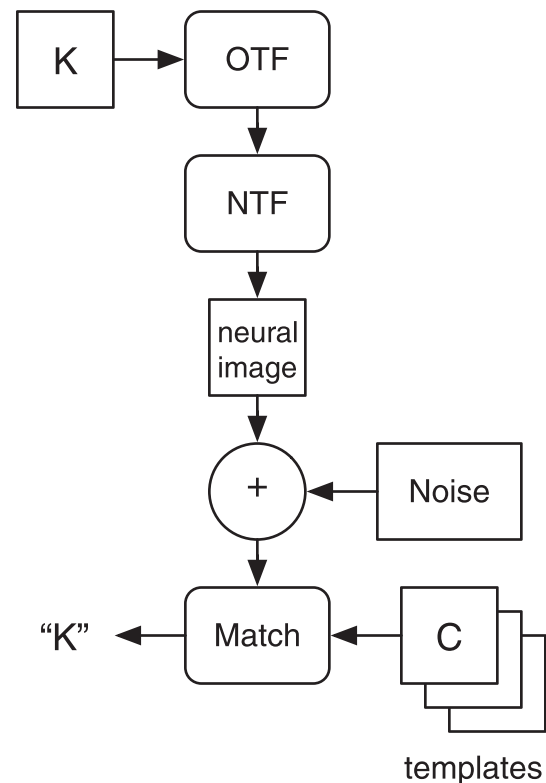


Figure 1. Outline of the template matching model for letter acuity (Watson & Ahumada, 2008). Figure courtesy of that paper.

rendered at a selected size, and presented to an observer. The stimulus image is transformed into a retinal image by the optics of the eye, which we characterize by an optical transfer function (OTF). The retinal image is transformed into a neural image by a neural filter, which we characterize by a linear neural transfer function (NTF). The variabilities inherent in both optical and neural systems are represented by white Gaussian noise, with standard deviation σ , added to the neural image. The noise power spectral density N is a key parameter of the model and is defined as

$$N = \sigma^2 A \quad (1)$$

where A is the area in square degrees of a single pixel in the model. We usually express N in the logarithmic unit of dBB, as explained in the Appendix.

The observer compares the noisy sample neural image to a set of templates. These templates are the complete set of neural images of the optotypes at the selected size. The observer selects the template that is closest, in the geometric or squared error sense, to the sample neural image. This model is an instance of an ideal observer of signals known exactly.

Watson and Ahumada (2008) considered variants of this model that included spatial uncertainty and suboptimal template matching algorithms, but found those manipulations to have little effect. Here we

	1	2	3	4	5	6	7	8	9	10
1	C	D	H	K	N	O	R	S	V	Z
2	个	文	么	山	少	之	十	力	人	天
3	州	月	占	只	无	区	以	世	风	任
4	条	名	抓	电	各	好	取	医	非	多
5	益	售	很	资	建	做	学	深	张	事
6	势	数	想	续	验	率	常	群	策	领
7	需	感	题	路	德	增	解	就	赛	整

Figure 2. Optotypes used in the experiments of Zhang et al. (2007). Each row is one set, and the set index shown at left. Chinese character optotypes increase in complexity with set index number. Optotypes are also indexed within a set, as shown above.

consider only the ideal model with optimal templates and without uncertainty.

Methods

Letter recognition data

The data analyzed in this report have been described in a previous publication (Zhang et al., 2007). The data describe letter identifications for the optotypes shown in Figure 2. These consisted of seven optotype sets, each with 10 elements. We assign each set an index number for reference. The first set consisted of the 10 Sloan letters; the remaining sets consisted of Chinese characters. Using an adaptation of the objective stroke frequency complexity measure (Majaj et al., 2002), the authors selected six sets of Chinese characters. Within each set complexity was nearly constant, while complexity increased with set index.

The raw data were provided to us by the first author of Zhang et al. (2007). Each data record consists of a decision associated with an observer, an optotype index, an optotype set index, an optotype height, a session index, and a repetition index within the session. Each decision consists of an optotype index (selection of one optotype within the set).

A possibly different set of six sizes was used with each observer and optotype set, but these sizes were drawn from a total set of 32 sizes, consisting of letter

heights ranging from 2.7 to 7.92 arcmin. Throughout this report we will use the word size as synonymous with letter height.

There were six observers, seven optotype sets, 10 optotypes within each set, six heights per observer and optotype set, and five replications in each session. In this paper we identify the observers by the colors used to depict their data: Red, Green, Blue, Gray, Orange, and Purple. Three of the observers (Blue, Gray, Orange) completed 10 sessions, two (Red, Green) completed five sessions, and one (Purple) completed seven sessions. In total there were 98,700 decisions.

In the experiment, optotypes were dark characters on a white background of luminance 89 cd/m². Pupil diameter was not specified or controlled. Based on the mean age of the observers (22.8 years) and the relationship between luminance, age, display size, and pupil diameter (Stanley & Davies, 1995; Watson & Yellott, 2012; Winn et al., 1994), we assumed a diameter of 6 mm. Further details on our selection of pupil size are provided in the Appendix.

Wavefront aberration data

As noted above, optical aberrations for the six observers were unknown, so we replaced them with aberration sets from 200 normal human eyes of 100 observers collected by the Indiana Aberration Study (IAS) (Thibos et al., 2002). IAS aberration data for a 6-

mm pupil were provided to us by Larry Thibos and consisted of Zernike coefficients for the first 36 modes.

Calculation of retinal images

Optotype images were rendered at the center of a 256×256 image. Details of character rendering are provided in the [Appendix](#). The image was filtered by an optical transfer function (OTF) for a given eye. The OTF was computed using standard methods (Watson & Ahumada, 2008) from the wavefront aberrations for a given eye, defined in terms of Zernike coefficients. In these calculations, we used the following parameters: image size (pixels) = 256, image size (degrees) = 0.6784, pupil size = 6 mm, center wavelength = 555 nm. Although the wavefront data we used were recorded with monochromatic light, we computed from them a polychromatic OTF for white light, using the methods described in the [Appendix](#).

The retinal image was then filtered by a NTF. This NTF was defined by an EmG (exponential minus Gaussian) function (Watson & Ahumada, 2005), multiplied by an oblique effect filter (OEF) with standard parameters (Watson & Ahumada, 2005). Details on calculation of the NTF are provided in the [Appendix](#).

The OTF and NTF were then multiplied together to create a neuro-optical filter (NOTF). Individual optotype images from a given set at a given size were rendered at the center of a 256×256 image and filtered by the NOTF. These constitute the neural images for that set at that size.

Since performance of the acuity model depends only on the noise and the matrix \mathbf{C} of cross-correlations among the neural images ([Equation 2](#)), we accelerated computation of the model by precomputing the matrix \mathbf{C} for each optotype set and size. Each matrix was 10×10 , and there were seven optotype sets, 32 sizes for each set, and 200 eyes, so the result was a data structure with dimensions $200 \times 7 \times 32 \times 10 \times 10$.

Simulation of one trial

The behavior of the acuity model can be simulated as follows. Let the optotypes be indexed $1, \dots, K$, and let s_k be the neural image for the optotype indexed by k . Let \mathbf{C} be the $K \times K$ matrix of cross-correlations among the K neural images,

$$\mathbf{C}_{j,k} = s_j \odot s_k \quad j, k = 1, \dots, K. \quad (2)$$

where \odot indicates the sum of the pixel-by-pixel product of the two images (the dot product of the two images regarded as vectors). Let \mathbf{e} be the vector consisting of the diagonal of this matrix, corresponding to the

energies of the neural images. Let σ be the standard deviation of the Gaussian noise added to each pixel of the neural image. Let k be the index of the letter presented. Then we consider the vector \mathbf{g}

$$\mathbf{g}_j = \mathbf{C}_{j,k} + \mathbf{m}(\sigma, \mathbf{C}) - \frac{1}{2} \mathbf{e}_j \quad (3)$$

where \mathbf{m} is a random vector of length K , constructed as described in Watson and Ahumada (2008). The observer locates the largest entry of \mathbf{g} , and returns its index j as the index of the optotype identified. This algorithm corresponds to the behavior of an ideal observer of a signal known exactly.

Simulation of acuity

To estimate acuity from this model, we select a value of σ and conduct T trials for each optotype. The optotype size in a given presentation is controlled by an adaptive Quest procedure (Watson & Pelli, 1983). This procedure analyzes past trials and sets the current size to the current estimate of acuity. The procedure is customized so that K trials, one per optotype, are presented at a given size before a new size is selected. The Quest method provides a highly efficient way of estimating acuity from the model. To illustrate the simulation, we provide a [demonstration](#) in the [Appendix](#) in which the reader can select a set of optotypes and a noise level.

Simulation of confusion matrices

To simulate the complete confusion matrix, we select a value of σ , and for each letter size used for that observer, we conduct T trials for each optotype. The result consists of a list of confusion matrices, indexed by letter size, for each observer.

Results

Letter recognition data

A summary of the letter recognition data of Zhang et al. (2007) in terms of percent correct versus size is shown in [Figure 3](#). Note that one observer (Blue) is considerably more acute than the others. Note also that the curves move to the right as set number increases, showing that the more complex characters require a larger size for equivalent performance. The data are plotted against a log size axis since we observed that the curves were more equal in slope when plotted this way.

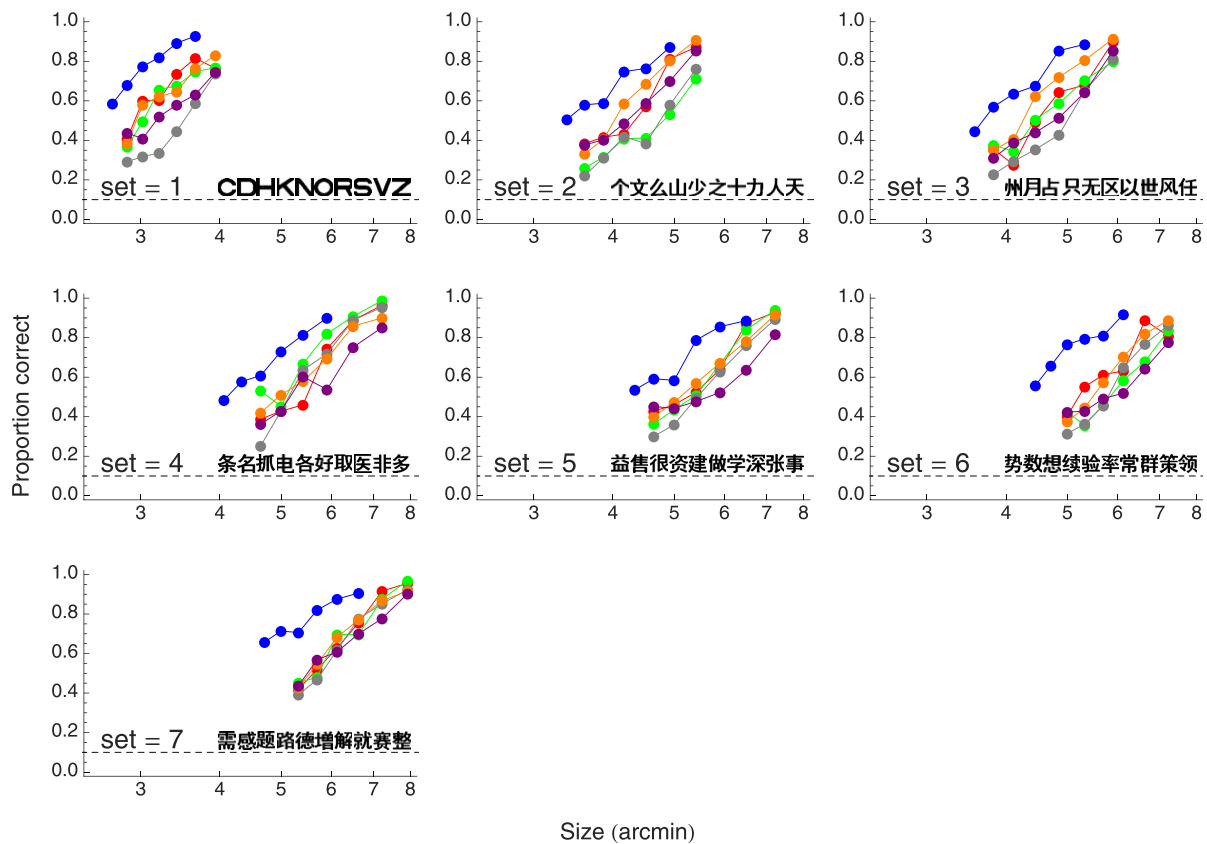


Figure 3. Proportion correct versus optotype size for six observers and seven optotype sets. The colors indicate individual observers. Optotype sets are shown in each panel. Data from Zhang et al. (2007). The dashed line is chance performance.

We fit the data for each observer and optotype set (each curve in Figure 3) with a cumulative normal distribution function (we also tried a Weibull function, which fit about as well). In the fitting, size was expressed as log arcmin. Following Zhang et al., we define acuity as the letter size yielding a proportion of 0.669 correct. The fit allows us to estimate this acuity and also a slope of the psychometric function. In Figure 4 we show example fits for two cases.

Thresholds for each observer and optotype set are plotted in Figure 5. The mean and standard deviation are shown in the companion figure. Mean size threshold increases substantially from set 1 (Sloan) to sets 2–7 (Chinese), with a mean increases of 62%. Within the Chinese character set (2–7) there is an increase with set number, and thus with stroke frequency, but little change among the three most complex sets. Observer Blue thresholds are well below

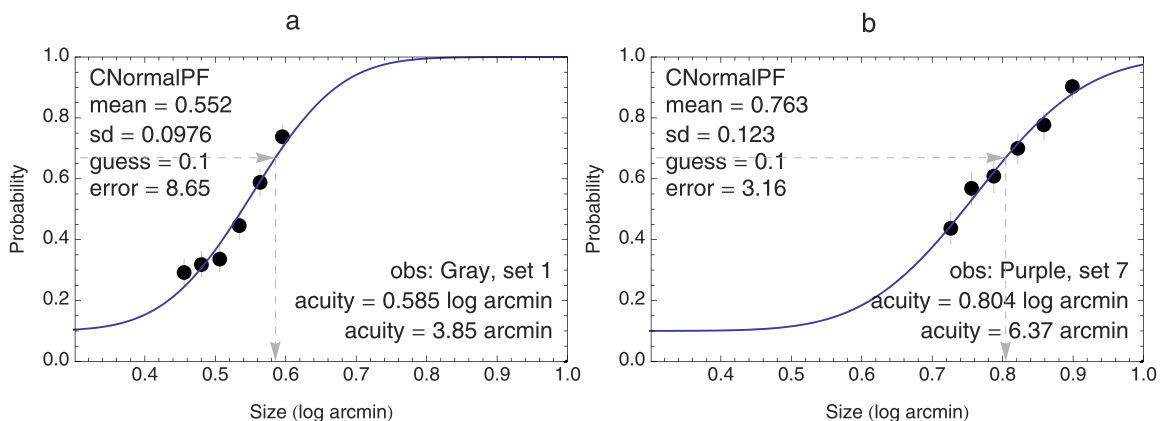


Figure 4. Example of cumulative normal fit to data for two observers and two optotype sets. The dashed gray line indicates threshold. Parameters of the fit are printed on the left in each panel, and on the right are the estimated acuity, observer, and set index.

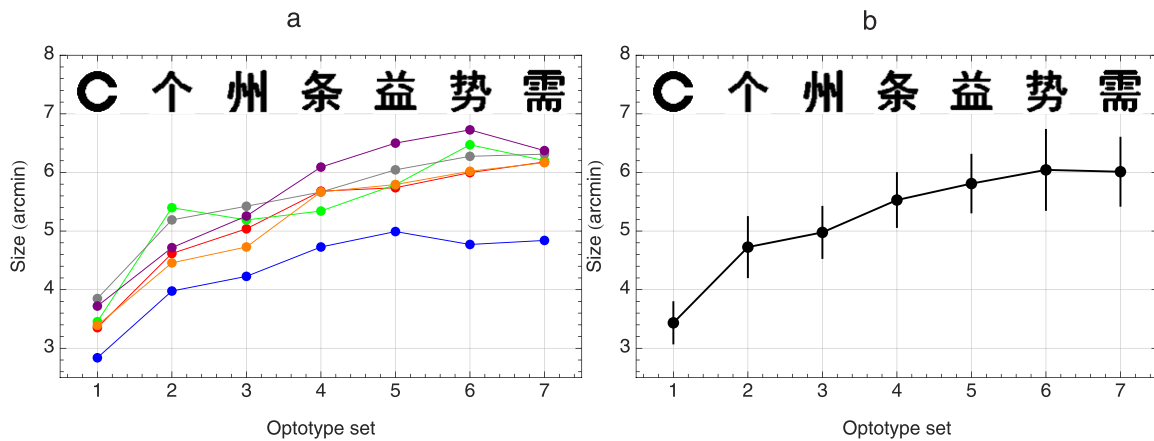


Figure 5. Threshold letter size for six observers and seven optotype sets. Colors are for individual observers. The mean and standard deviation are shown in the panel on the right. One letter from each set is shown.

the mean and also show less of an effect of set number.

When fitting the data we also estimated the width of the psychometric function (the standard deviation of the cumulative normal distribution, as a function of log arcmin). This parameter had a mean value of 0.113. It did not vary much as a function of observer or optotype set, as shown in Figure 6. The most sensitive observer (Blue) and the least sensitive observer (Purple) have values slightly greater than the mean.

The invariance with respect to optotype of psychometric function slope, on a log size abscissa, is a useful fact that is exploited in the design of acuity tests. When tests space optotype size uniformly in logMAR units, they are also spacing them uniformly in just-noticeable-differences.

Simulation of acuity with fixed noise

To assess the performance of the model, we first fixed the noise parameter σ at 0.24 ($N = -3.931$ dBB)

and used a Quest adaptive procedure based on 1,024 trails/letter to locate the threshold size for each optotype set (see the Appendix for details). From earlier simulations, we determined that this noise value would approximate the human data for Sloan letters. We have repeated this simulation for each of the 200 eyes of the IAS. We also provide a demonstration in the Appendix to illustrate the process of estimating acuity for a given set of optotypes.

All 200 outcomes are pictured in Figure 7a, along with the mean in red. Four of the curves fall outside of the figure, illustrating cases for which the threshold size was greater than the largest available size in the Zhang et al. (2007) experiment. The figure shows the range of threshold sizes that might be predicted from the model for a population of normal eyes.

In Figure 7b the mean and standard deviation have been reproduced (red), along with the human data (black). For these simulations, a noise level ($N = -3.931$ dBB) was chosen to approximate the data for optotype set 1 (Sloan letters). The figure shows that, with one fixed noise value set to agree for set 1, set 2 is also

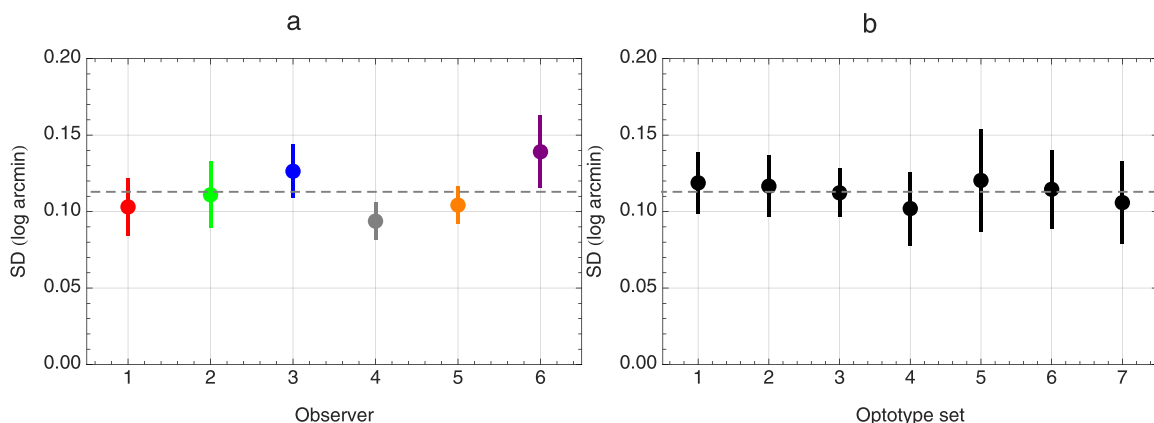


Figure 6. Mean and standard deviation of psychometric function width for each observer and optotype set. The mean is shown by the dashed gray line.

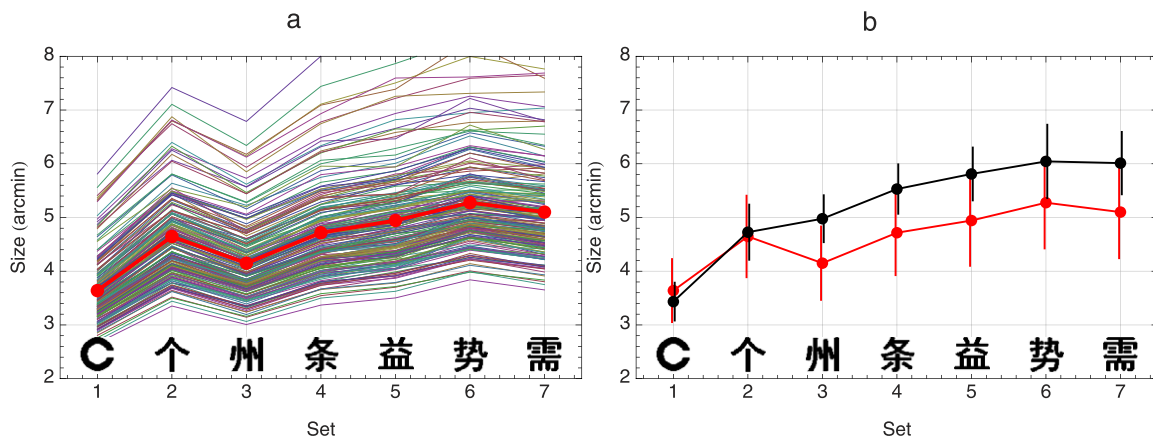


Figure 7. (a) Model size threshold versus optotype set for 200 eyes. The mean is shown in red. The noise level was $N = -3.931$ dB. (b) Mean model size thresholds (red) compared to data (black). In both cases, error bars show plus and minus one standard deviation.

accounted for, but the data move above the predictions for the more complex sets. In other words, the model accounts for some, but not all of the rise in threshold size with set number (or stroke frequency). Indeed the model actually predicts a decline in threshold from set 2 to set 3, in spite of an increase in stroke frequency. The separation between the two curves is a measure of the portion of the rise in threshold size, with set number, that is not accounted for by the model.

Although in Figure 7b we used a noise value that yielded model performance approximately matched to data for the Sloan letters (set 1), we could have used a noise suitable to match set 3, in which case the more complex sets 3–7 would have been well fit, while the predictions for the simpler sets 1 and 2 would have been too high. The point is that with a single noise value, one cannot fit both simple and complex sets. We could fit all points with two noise parameters, for the two sets, but that would not allow a predictive model because we would not know with a new set which parameter to use.

Considering just the predictions in Figure 7b, can we explain the pattern of results? At least three things are changing across the various sets: stroke width varies (it is thinner for the Chinese characters), the number of strokes per unit area, and thus the mean spatial frequency increases, across sets 2–7, and the amount of information in the target increases. The first two factors will reduce performance, while the last will increase performance, but in each case it is difficult to say by how much. The ideal observer shows us quantitatively how all the factors combine to yield performance.

We also compared measured and simulated psychometric function width (the inverse of slope). This comparison is shown in Figure 8. As in Figure 6, psychometric function width is quantified by the standard deviation of the cumulative Gaussian as a function of log size. For the model (red) and data

(black) we show the mean and a ribbon that encloses plus and minus one standard deviation. For the data, the standard deviations are computed over the six observers. For the model, they are computed over the 200 IAS eyes. The width values for the model appear slightly higher (shallower slopes), but the overall agreement between model and data is reassuring. Uncertainty tends to steepen psychometric functions relative to the ideal observer (Pelli, 1985), and this may explain the small difference in slope.

Estimating noise from proportion correct

An alternative approach to understanding the role of optotype set on threshold size is to estimate the best fitting model noise parameter N for each set by fitting the model to the proportion correct data. To do this, for each set, observer, and eye, we selected a range of N values. For each, we simulated 1,000 trials for each letter at each of the sizes used by Zhang et al. (2007) for

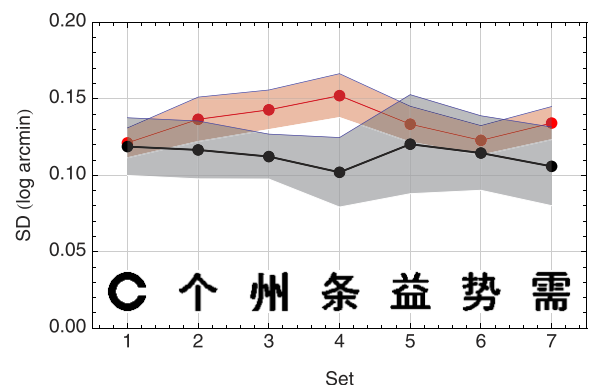


Figure 8. Comparison of psychometric function widths for the data (black) and the model (red). Mean plus and minus one standard deviation are shown for each.

that set. The error between model and data, defined as the log of the likelihood ratio, expressed as

$$\text{error} = \sum_{\text{size}} [p_{\text{data}} \log(p_{\text{model}}) + (1 - p_{\text{data}}) \log(1 - p_{\text{model}})] \quad (4)$$

was computed for each noise value (a small constant was substituted for zero values of p_{model} to prevent overflow). An interpolating function was then used to estimate the N value yielding the minimum error. The accuracy of this method was confirmed by generating simulated data from the model with a known N , and then estimating the value of N . This yielded a total of $6 \times 200 = 1,200$ noise estimates for each set. The average of these 1,200 values is shown in Figure 9.

The noise levels divide roughly into two groups: for sets 1 and 2, they are around -3 to -5 dB, while for sets 3–7 they are about 3 dB greater. Consistent with Figure 7b, these results show that it is possible to account for sets 1 and 2 with approximately the same noise value, but that the other sets require a significantly larger value. As expected, these estimated noise values correspond closely to the discrepancy between the two curves in Figure 7b.

Estimating noise from confusion matrices

To this point we have examined the data and predictions in terms of proportion correct. This is the only measure that matters in estimation of threshold size or acuity. But the data of Zhang et al. (2007) also allow us to examine the pattern of errors, in terms of the confusion matrix. This a matrix showing, for each letter presented (row), the number of times a particular letter was reported (column).

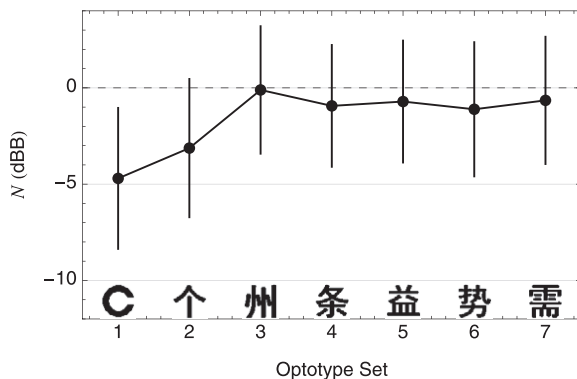


Figure 9. Estimated noise values for optotype sets. Each point is the mean over estimates for six observers and 200 eyes. Error bars show plus and minus one standard deviation. Standard errors are smaller than the point size.

The complete data for each condition (observer, optotype set, size) is one confusion matrix. The proportion correct is a single number derived from the matrix (the ratio of entries on the main diagonal to the total entries in the matrix). The data for a single observer and optotype set is a collection of confusion matrices, one for each size used with that observer. An example of such a collection is shown in Figure 10.

Rather than fitting to the proportion correct, as we have done above, we can fit to the complete confusion matrix. For each set, observer, and IAS eye, we selected a range of noise values. For each, we simulated 1,000 trials for each letter at each of the sizes used by Zhang et al. (2007) for that set. We then computed the error between model and data over the complete set of sizes, again using Equation 4. The validity of this statistic was again tested by creating and fitting sets of simulated data.

An interpolating function was then used to estimate the noise value yielding the minimum error. This yielded a total of $6 \times 200 = 1,200$ noise estimates for each set. The average of these 1,200 values is shown in Figure 11. In the same figure for comparison we reproduce in red the estimates of N obtained from the proportion correct, shown in Figure 9.

The noise levels again divide roughly into two groups: for sets 1 and 2, they are around -3 to -5 dB, while for sets 3–7 they are a about 3 dB greater. Consistent with Figure 7b, these results show that it is possible to account for sets 1 and 2 with approximately the same noise value, but that the other sets require a significantly larger value. Not surprisingly, these estimated noise values correspond closely to the discrepancy between the two curves in Figure 7b.

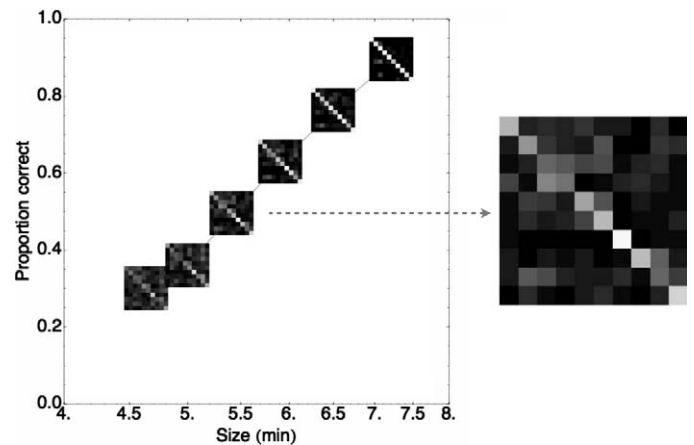


Figure 10. Example of confusion matrices for observer Gray for optotype set 5 at six sizes. The matrices are positioned at the corresponding proportion correct, which is given by the proportion of trials lying on the main diagonal in each matrix. One matrix is expanded for clarity.

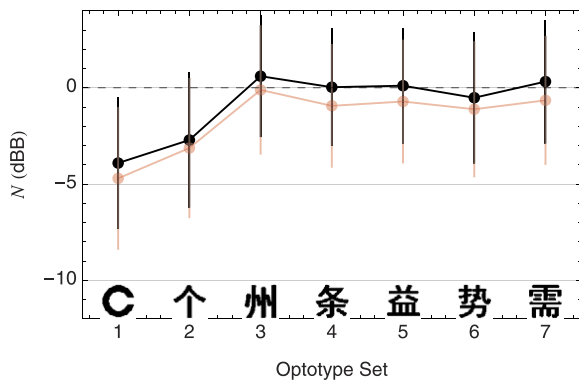


Figure 11. Model noise for each optotype set estimated from confusion matrices. The faint red points are the estimates from percent correct, reproduced from Figure 9.

The estimates obtained from the full confusion matrix (black) are systematically but only slightly higher than those obtained from just the proportion correct (red). This may be due to a small but consistent positive bias that we find in simulations of our method of estimation from confusion matrices. But in general the two estimation methods yield results that are in close agreement.

Efficiency and complexity

Efficiency is the performance of an actual observer relative to an ideal observer. It is computed as the ratio of squared signal/noise ratios of the two observers (Pelli, Burns, Farell, & Moore-Page, 2006; Tanner & Swets, 1954). Thus the noise estimates in Figure 11 can be converted to estimates of efficiency relative to the Sloan letters by computing N_l/N_k , where N_k is the estimated noise for set k . The inverse of efficiency is plotted as black points in Figure 12. The normalized values range from 1 (for the Sloan letters, by definition) to between 1.4 and 2.7 for the Chinese characters.

Pelli et al. (2006) have examined efficiency for identification of letters from various alphabets embedded in noise. They found efficiency for one set of Chinese characters that was about 2.1 times lower than that for Sloan letters, consistent with the differences shown here.¹ It should be noted, however, that the letters here, at size threshold, are some 13 (Chinese) to 32 (Sloan) times smaller than those used by Pelli et al.

Pelli et al. (2006) also found that efficiency was nearly inversely proportional to the average perimetric complexity of the letter set. Perimetric complexity is defined for binary images as the length of inner and outer perimeters of the foreground, squared and divided by the area of the foreground. For binary digital images, composed of discrete pixels, a strict definition of the perimeter would consist of a city-block

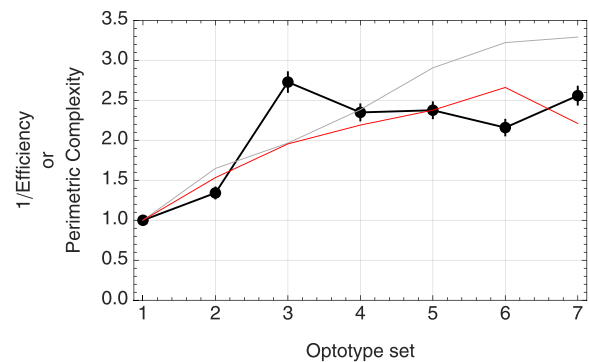


Figure 12. Inverse of efficiency (normalized to Sloan letters) as a function of optotype set (black). Raw perimetric complexity is shown by the gray line; the red line shows visual perimetric complexity at 1/6 the acuity distance.

path along the exposed edge of each pixel in the perimeter (Watson, 2012). Complexity calculated in this way is shown by the gray line in Figure 12. This measure of complexity does not precisely mirror the changes in efficiency, but does show a rough agreement.

Perimetric complexity is defined for binary images and is a sensible measure when letters are large relative to the visual point-spread. However at the acuity limit, characters are severely blurred by the eye's optics and possible neural filtering, which makes the measure problematic. This is because the blurred letters are far from binary and may have lost much detail. We have developed a metric of visual perimetric complexity that attempts to deal with these challenges, by first filtering the letter in an appropriate way and then binarizing the result (Watson, 2012).

It is instructive that when this metric is applied to the optotypes used here, at the acuity size, all optotypes have a visual perimetric complexity of approximately 1. This is the minimum theoretical value, and is the result obtained from a circular blob. The severely blurred optotypes retain sufficient gray-scale information to be identified, but their perimeters (as we have defined it) are uninformative. We conclude that perimetric complexity is not a sensible measure for targets near the acuity limit.

To illustrate this problem, in Figure 13 we show the calculation of visual perimetric complexity for one optotype from set 7 when viewed at twice the acuity size.

However, for illustrative purposes, we have computed visual perimetric complexity at a size 6 times larger than the threshold, and the results are shown by the red curve in Figure 12. The point is only to show that blurring has a selective effect on the more complex optotypes.

With no filtering (Figure 12, gray line), the complexity increases with set number. This is expected.

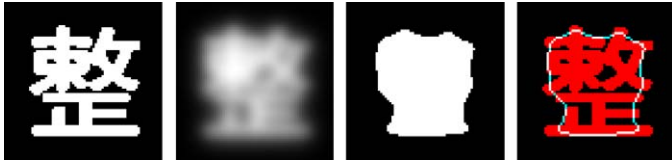


Figure 13. Calculation of visual perimetric complexity for one optotype at twice its acuity size. The images show: original optotype; after visual filtering; after binarization; location of the perimeter, relative to the original optotype. The visual perimetric complexity is 1.35.

As noted earlier, the optotype sets were designed by Zhang et al. (2007) to increase in stroke frequency with set number, and those authors report that stroke frequency and perimetric complexity were highly correlated ($r = 0.956$). However, this unfiltered complexity does not match the measured efficiencies, which are nearly constant for the more complicated Chinese characters. But this flattening at higher set numbers is somewhat mirrored by the filtered complexities (Figure 12, red line).

In summary, variations in measured efficiency are roughly mirrored by increases in perimetric complexity, but not when perimetric complexity is computed on appropriately blurred optotypes. Perhaps the variations are related to some other measure of image complexity, but that measure has not yet been defined.

Correlation with confusion matrices

In a recent article, Liu, Klein, Xue, Zhang, and Yu (2009) proposed a model of human letter recognition based on a small set of measurements (derived from geometric moments) on the letter images. To evaluate their model, and compare it to other models, they applied it to the empirical confusion data collected by Zhang et al. (2007) that we have used here.

To evaluate their model, they relied primarily on Pearson correlation between off-diagonal elements of empirical and model confusion matrices. To enable comparison of our model with theirs, we have attempted to follow their correlation procedures exactly. This has not always been possible. First, they describe their data as consisting of “approximately 110,000 trials,” while the data we were provided consist of only 98,700 trials. Thus our two data sets may not correspond exactly. Second, they indicate that they created a set of seven empirical confusion matrices by combining results across observers, but selecting only data from sizes that yielded a percent correct of between 54% and 60%. They published these matrices (converted to response proportions). However, applying these rules, we find two optotype sets (3 and 4) that

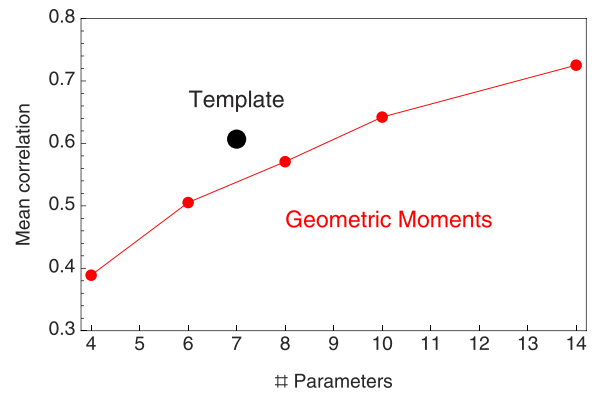


Figure 14. Mean correlation between model and data confusion matrices for the present model (Template) and a series of models described by Liu et al. (2009) (Geometric Moments).

yield no data in the requisite band. For those sets that do yield data, the matrices we find are slightly different from those published. To allow us to proceed, with these caveats, we have determined the selection band that gives the best agreement (in the least squared error sense) between the published matrices and our own. This band is 40% to 75%. Using this band, we constructed mean confusion matrices for each optotype set by averaging over observers.

We then found for each optotype set the value of N that maximized the correlation between model and data. This corresponds to a model with seven parameters, one for each optotype set. The average correlation for this optimized model is plotted against number of parameters in Figure 14 as the single black point. Liu et al. (2009) considered several variants of their model, differing in number of parameters. We plot their average correlations against number of parameters in Figure 14. It is evident that our model lies above the red curve, and thus fits better than a geometric moment model with a comparable number of parameters.

Comparison with detection data

Further evidence for the plausibility of the template model of acuity can be sought in a comparison of the absolute level of performance in the acuity task and in a simple detection task. As an example, we consider detection of a Gabor. Again assuming an ideal observer, the proportion correct in a 2AFC task can be shown to be

$$P(c) = \Phi\left(\sqrt{\frac{E}{2N}}\right) \quad (5)$$

where Φ is the cumulative distribution function of the standard normal density and E is the energy of the neural image. We can compute values of E using average

threshold contrasts for a Gabor target from the ModelFest project (Watson & Ahumada, 2005), combined with the neural and optical transfer functions employed above. Using the ModelFest value of $P(c) = 0.84$, we can estimate the corresponding value of N . We can repeat this exercise for each of the 200 eyes in the IAS, and produce a distribution of estimates of N . This is shown in Figure 15 for Gabor functions of 4, 8, and 16 cycles/deg, each of constant one octave bandwidth. These are stimuli 12, 13, and 14 from ModelFest experiment (Watson & Ahumada, 2005). We also show as a red arrow the mean value estimated for the Sloan letters.

Considering the breadth of the distribution, this represents rather close agreement between the values of N values estimated in these two very different ways and provides further support for the acuity model proposed here.

However, there are a number of differences in methods and subject populations that complicate this comparison. Two of these differences favor letter identification. First, the observers employed by Zhang et al. (2007) were younger (mean age 23 vs. 39 years). Second, letter identifications were performed with unlimited duration, while the Gabor targets had a brief Gaussian time course with a standard deviation of 0.125 s. The effect of this duration difference on noise estimates cannot be known precisely, but could be as much as a factor of 2 (Watson, 1979).

Favoring Gabor detection, we have observed in this paper that more complex targets yield larger estimates of noise, and the Sloan letters are arguably more complex than the Gabor targets. Indeed elsewhere Pelli (2011) has argued that to the visual system, the Gabor may be the simplest target, with an efficiency of 20%, while large Sloan letters have a lower efficiency of 10%. Though Sloan letters at the acuity limit may be simpler than large letters (cf. Figure 13) they may nonetheless be more complex than the Gabor, which would lead to larger noise estimates for the letter targets. Since these three caveats work in opposite directions, they may cancel, yielding the relatively good agreement we find

here between noise estimates for Gabor detection and letter identification.

In Figure 15 the mean noise estimates for Gabor detection rise as the patterns become smaller and of higher spatial frequency. One possible explanation for this is spatial uncertainty. As the near-threshold patterns become smaller and of higher spatial frequency, their position becomes less certain. The optotypes do not suffer as much from this effect, because they are all of high contrast. A better comparison would be to a contrast increment experiment with Gabors. A second explanation is that the NTF we have used here is more sensitive to high spatial frequencies than the one actually derived from the ModelFest data (for reasons explained elsewhere [Watson & Ahumada, 2008], it was shifted by a factor of two to higher spatial frequencies). To match the lower sensitivities of ModelFest observers at high spatial frequencies, a higher value of noise must be assumed.

Discussion

Acuity and complexity for the ideal observer

One of the predictions of our ideal observer model is that threshold letter size generally increases with complexity, as shown by the red curve in Figure 7b. Why might this be so? The more complex optotype sets contain more information, and on that basis might be expected to be more discriminable than the simpler sets, and thus have a lower threshold. But of course that larger amount of information is conveyed in strokes and patterns that must be more crowded together, if size is held constant, and thus be represented with, on average, higher spatial frequencies. If the area must be expanded to allow each feature to be resolved, then we would expect the increase in size that is in fact predicted. These two factors, information and resolu-

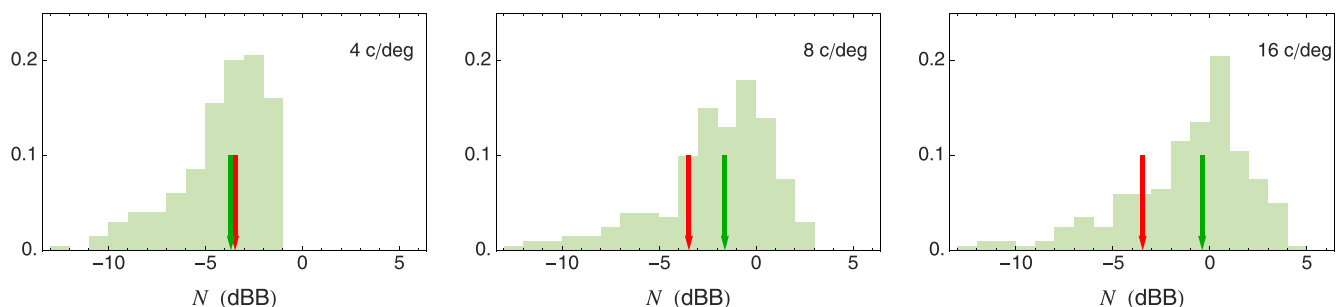


Figure 15. Distribution of estimates of noise N for detection of one octave Gabor targets, assuming the same neural and optical transfer functions used in simulation of letter identification. The green arrow is the median of the distribution, and the red arrow is the median value estimated from the confusion data for Sloan letters (Figure 11).

tion, presumably compete to produce the particular gyrations that we see in [Figure 7b](#).

We also note that the Chinese characters in sets 2–7 employed narrower stroke widths than in the Sloan characters. As shown by [Zhang et al. \(2007\)](#) this reduces acuity, and our model likely manifests this same effect, providing an additional explanation for the difference in predictions between Sloan and Chinese characters.

Efficiency and complexity

One of the main results of our analysis is that efficiency of optotype identification declines with complexity for targets at the acuity limit. This compares to a comparable result obtained previously for letter targets much larger than the acuity limit ([Pelli et al., 2006](#)). In their discussion of efficiency and complexity, [Pelli et al. \(2006\)](#) argued that the reduction in efficiency could be the result of independent detection of multiple simple features. But there are several other equally plausible explanations.

One is the idea that the templates are imperfect, and that the amount of imperfection increases with complexity. This would likely be the case for any plausible model of learning and memory for spatial templates. Imperfection of the template would result in a decline in efficiency, since the ideal observer employs a perfect template. Elsewhere [McIlhagga and Paakkonen \(1999\)](#) have explored the behavior of noisy or imperfect templates.

A second possible explanation is uncertainty. As noted by [Zhang et al. \(2007\)](#) primary school students must learn at least 2,500 Chinese characters, while there are only 26 letters in the Roman alphabet, and only 10 letters in the Sloan subset. In the acuity task, the observers select from among a fixed small set of alternatives. The sessions for various sets and sizes were randomly interleaved. To perform like the ideal, the observers of [Zhang et al. \(2007\)](#) would have to accurately learn and maintain separate subsets for each of the six sets of Chinese character optotypes. To the extent that they do not, performance and efficiency will suffer. Because the more complex characters are likely to be less frequently used, and thus less familiar, they may suffer more.

Reduced familiarity with the Chinese characters may also lead to less perfect templates. Although the subjects of [Zhang et al. \(2007\)](#) had considerable experience with both English and Chinese alphabets, the much larger number of Chinese characters may reduce the learning of each one. [Pelli et al. \(2006\)](#) have shown that at least some alphabets are learned rapidly, but their results also show that with increasing training learning slows, but does not stop.

Limitations of perimetric complexity

The measure of perimetric complexity has become a popular method of quantifying the complexity of visual targets ([Pelli et al., 2006](#)). However, as already noted, it is defined only for binary images and becomes particularly problematic for small, highly blurred targets near the acuity limit ([Watson, 2012](#)). The complexity of the original target will be irrelevant if it is removed by visual filtering (cf. [Figure 13](#)). Perimetric complexity, as a visual measure, also does not make much sense in a theory that imagines the simplest feature, with the lowest complexity, to be a Gabor function. That function is not even defined for a binary image, and its binary renderings would have rather high perimetric complexity. We do not yet have a clear understanding of how to measure visual complexity, but until we do we suggest that perimetric complexity, if used at all, be used only for large binary targets in which visual filtering is a minor effect.

Simulating a population of observers

In our simulations we have made use of a large population of empirically measured eyes in order to accurately predict population behavior. We could have instead computed an “average” eye and computed only a single simulation for this “average” observer. The problem with this approach is that there is no single correct way to compute the average eye. As noted by [Thibos et al. \(2002\)](#), one could average the OTFs, or the PSFs, but neither, when used in a simulation, is guaranteed to yield the mean performance of the ensemble of eyes. Averaging the PSFs, in particular, will give a mean PSF that is much more blurred than the typical PSF. Likewise averaging the OTFs or the PSFs will lead to a mean that is much more symmetrical (and thus without phase shifts) than any individual PSF or OTF. For these reasons we have elected to simulate each eye separately and consider the distribution of results. This has the added advantage of indicating the expected variation in the population of the simulated performance. We believe this population simulation methodology is a profitable approach that should be adopted more often.

Implications for selection and calibration of optotypes

One motive for this study was the hope that the acuity model could provide a means of designing, selecting, and calibrating optotypes. As noted, there are many sets in wide use and frequent attempts to create new sets that serve a particular need. With existing sets,

there is a need to calibrate: to determine the relation between its results and those of some canonical set, such as the Sloan letters. With new sets, there is a need to calibrate as well and also to confirm that its results are not too far from those of the canonical standard. The model might provide other benefits as well, such as selecting elements of the set that are sufficiently distinct from one another, and designing tests that have low variability.

However, we have shown that the model cannot account for differences in performance that are associated with large changes in complexity. This might suggest that the model cannot serve the calibration function. But in practice, optotype sets are invariably simple. Indeed, that appears to be an implicit design principle. And we note that the model did a reasonable job of predicting relative performance of Sloan letters and the simplest subset of Chinese character optotypes. Thus, though further research on this point is warranted, the model may indeed serve a valuable function in calibration, design, and selection of simple practical optotypes.

Templates versus features

Pelli et al. (2006) have argued that letter identification relies on detection and identification of features, followed by feature binding. This continues a long tradition of feature-based models (Geyer & DeWald, 1973; Gibson, Osser, Schiff, & Smith, 1963; Yeh & Eriksen, 1984). These theories have suffered from an inability to identify the actual features involved. Liu et al. (2009) have made an effort to remedy this problem by proposing a particular set of luminance contrast patterns as the elementary features involved in letter identification, at least near the acuity limit.

Our model also continues a tradition, that of template models (Blommaert, 1988; Gervais et al., 1984; Loomis, 1990; Watson & Fitzhugh, 1989). What we have shown here is that, based at least on the pattern of results in the confusion matrix, a template model performs better than the feature model of Liu et al. (2009). While template matching may be an implausible model for the visual identification of more elaborate and less stereotyped patterns and objects, it cannot yet be rejected as a model for identification of letters.

Conclusions

Relative acuity performance on different optotype sets (Sloan letters and six sets of Chinese characters)

from a single set of observers cannot be predicted by an acuity model consisting of an ideal observer limited only by noise, optical filtering, and neural filtering.

Efficiency for the acuity model should be independent of optotype set, but efficiencies for Chinese character optotype sets were between 1.3 to 2.7 times lower than for Sloan letters.

Decreased efficiency of Chinese character optotypes is loosely associated with increases in complexity, as measured by perimetric complexity.

Perimetric complexity is a poor measure of complexity of small, blurred, gray-scale targets near the acuity limit.

The acuity model does provide a reasonable account for the relative performance of Sloan letters and the simplest set of Chinese character optotypes.

The acuity model may be useful in calibrating acuities measured with different simple optotypes.

Estimates of internal noise for the acuity task and for detection of a Gabor target are consistent, but this conclusion must be qualified by differences in observer populations, target complexity, and target duration.

Psychometric functions for the acuity model and human observers were similar in slope.

The acuity model provided a better fit to the data, as measured by correlation of confusion matrices, than did a feature model based on geometric moments.

Acknowledgments

We thank Dr. Cong Yu and Jun-Yun Zhang for providing the Chinese character optotypes. We thank Larry Thibos for providing the wavefront data. Thanks to Jack Yellott for urging a clearer look at the pupil reflex. This work was supported by NASA Space Human Factors Engineering WBS 466199 and Office of Naval Research Award N0001411WX21610. ABW is employed by NASA, which has patent interests in this work.

Commercial relationships: Yes.

Corresponding author: Andrew B. Watson.

Email: andrew.b.watson@nasa.gov.

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

Footnote

¹ Pelli et al. (2006) used a set of 26 Chinese characters, but only 10 Sloan letters. It is unknown how the set size might affect efficiency.

References

- 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(8), 1374–1381. [PubMed]
- Bailey, I. L., & Lovie, J. (1980). The design and use of a new near-vision chart. *American Journal of Optometry and Physiological Optics*, 57(6), 378.
- Beckmann, P. J., & Legge, G. E. (1996). Psychophysics of reading—XIV. The page navigation problem in using magnifiers. *Vision Research*, 36(22), 3723–3733. [PubMed]
- Blommaert, F. J. J. (1988). Early-visual factors in letter confusions. *Spatial Vision*, 3(3), 199–224.
- Charman, W. N. (2005). Wavefront technology: Past, present and future. *Contact Lens and Anterior Eye*, 28(2), 75–92.
- 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://www.journalofvision.org/content/4/4/7>, doi:10.1167/4.4.7. [PubMed] [Article]
- 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 and Image Science*, 25(8), 2078–2087. [PubMed]
- Dalimier, E., Pailos, E., Rivera, R., & Navarro, R. (2009). Experimental validation of a Bayesian model of visual acuity. *Journal of Vision*, 9(7):12, 1–16, <http://www.journalofvision.org/content/9/7/12>, doi:10.1167/9.7.12. [PubMed] [Article]
- Dobson, V., Maguire, M., Orel-Bixler, D., Quinn, G., Ying, G. S.; Vision in Preschoolers (VIP) Study Group. (2003). Visual acuity results in school-aged children and adults: Lea symbols chart versus Bailey-Lovie chart. *Optometry and Vision Science*, 80(9), 650–654.
- Ferris, F., III, & Bailey, I. (1996). Standardizing the measurement of visual acuity for clinical research studies: Guidelines from the Eye Care Technology Forum. *Ophthalmology*, 103(1), 181.
- Gervais, M. J., Lewis, O., Harvey, J., & Roberts, J. O. (1984). Identification confusions among letters of the alphabet. *Journal of Experimental Psychology*, 10(5), 655–666.
- Geyer, L. H., & DeWald, C. G. (1973). Feature lists and confusion matrices. *Perception and Psychophysics*, 14(3), 471–482.
- 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* (pp. 1–20). Cooperative Research Program No. 639, Office of Education.
- Grimm, W., Rassow, B., Wesemann, W., Saur, K., & Hiltz, R. (1994). Correlation of optotypes with the Landolt ring—a fresh look at the comparability of optotypes. *Optometry and Vision Science*, 71(1), 6–13.
- Jackson, J. A., & Bailey, I. L. (2009). Visual acuity: The effect of optotype and test chart design. [Abstract]. *Proceedings, Meeting of American Academy of Optometry*, 95521.
- Liu, L., Klein, S. A., Xue, F., Zhang, J.-Y., & Yu, C. (2009). Using geometric moments to explain human letter recognition near the acuity limit. *Journal of Vision*, 9(1):26, 1–18. <http://journalofvision.org/9/1/26/>, doi:10.1167/9.1.26. [PubMed] [Article]
- Loomis, J. M. (1990). A model of character recognition and legibility. *Journal of Experimental Psychology: Human Perception and Performance*, 16(1), 106.
- Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research*, 42(9), 1165–1184. [PubMed]
- Marcos, S., Burns, S. A., Moreno-Barriusop, E., & Navarro, R. (1999). A new approach to the study of ocular chromatic aberrations. *Vision Research*, 39(26), 4309–4323.
- McIlhagga, W., & Paakkonen, A. (1999). Noisy templates explain area summation. *Vision Research*, 39(2), 367–372. [PubMed]
- Moutray, T. N., Williams, M. A., & Jackson, A. (2008). Change of visual acuity recording methods in clinical studies across the years. *Ophthalmologica*, 222(3), 173–177.
- Navarro, R., Artal, P., & Williams, D. R. (1993). Modulation transfer of the human eye as a function of retinal eccentricity. *Journal of the Optical Society of America. A, Optics and Image Science*, 10(2), 201–212. [PubMed]
- Nestares, O., Navarro, R., & Antona, B. (2003). Bayesian model of Snellen visual acuity. *Journal of the Optical Society of America. A, Optics and Image Science*, 20(7), 1371–1381. [PubMed]
- Pelli, D. (2011). Visual sensitivity explained. *Perception*, 40(ECVP Abstract Supplement), 15.
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination.

- Journal of the Optical Society of America. A, Optics and Image Science*, 2(9), 1508–1532.
- Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, 46(28), 4646–4674.
- Pointer, J. S. (2008). Recognition versus resolution: A comparison of visual acuity results using two alternative test chart optotype. *Journal of Optometry*, 1(2), 65–70.
- Ravikumar, S., Thibos, L. N., & Bradley, A. (2008). Calculation of retinal image quality for polychromatic light. *Journal of the Optical Society of America. A, Optics and Image Science*, 25(10), 2395–2407. [PubMed]
- Stanley, P. A., & Davies, A. K. (1995). The effect of field of view size on steady-state pupil diameter. *Ophthalmic & Physiological Optics: The Journal of the British College of Ophthalmic Opticians*, 15(6), 601–603. [PubMed]
- Tanner, W. P., Jr., & Swets, J. A. (1954). The human use of information. I. Signal detection for the case of a signal known exactly. *Transactions of the IRE, PGIT-4*, 213–221.
- 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 and Image Science*, 19(12), 2329–2348.
- Thibos, L. N., Ye, M., Zhang, X., & Bradley, A. (1992). The chromatic eye: a new reduced-eye model of ocular chromatic aberration in humans. *Applied Optics* 31(19), 3594–3600.
- Watson, A. B. (1979). Probability summation over time. *Vision Research*, 19, 515–522.
- 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., & 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., & 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. [PubMed] [Article]
- Watson, A. B., Barlow, H. B., & Robson, J. G. (1983). What does the eye see best? *Nature*, 302(5907), 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(2), 113–120.
- Watson, A. B., Taylor, M., & Borthwick, R. (1997). Image quality and entropy masking. *SPIE Proceedings*, 3016, 2–12.
- 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/12/10/12/>, doi:10.1167/12.10.12.
- Williams, M. A., Moutray, T. N., & Jackson, A. J. (2008). Uniformity of visual acuity measures in published studies. *Investigative Ophthalmology & Visual Science*, 49(10), 4321–4327, <http://www.iovs.org/cgi/content/abstract/49/10/4321>. [PubMed] [Article]
- Winn, B., Whitaker, D., Elliott, D. B., & Phillips, N. J. (1994). Factors affecting light-adapted pupil size in normal human subjects. *Investigative Ophthalmology & Visual Science*, 35(3), 1132–1137, <http://www.iovs.org/cgi/content/abstract/35/3/1132>. [PubMed] [Article]
- Wolfram Research: Mathematica, Technical and Scientific Software. Internet site: www.wolfram.com/. (Accessed September 26, 2011).
- Yeh, Y., & Eriksen, C. W. (1984). Name codes and features in the discrimination of letter forms. *Perception and Psychophysics*, 36(3), 225–233.
- Zhang, J.-Y., Zhang, T., Xue, F., Liu, L., & Yu, C. (2007). Legibility variations of Chinese characters and implications for visual acuity measurement in Chinese reading population. *Investigative Ophthalmology & Visual Science*, 48(5), 2383–2390, <http://www.iovs.org/cgi/content/abstract/48/5/2383>. [PubMed] [Article]

Appendix

Ideal observer

Here we provide a brief derivation of the formula used to simulate the performance of the ideal observer (Equation 3). Recall that the optotypes are indexed 1, . . . , K . Let k be the index of the optotype presented. The resulting neural image s_k is corrupted by the

addition of a noise image n with standard deviation σ . The ideal observer considers which of the template neural images s_j is closest to the signal plus noise image $s_k + n$; that is, it seeks the index j for which the following is minimized

$$d_j = \|s_j - (s_k + n)\|^2. \quad (6)$$

Expanding this expression for the distance, we have

$$d_j = \|s_j\|^2 - 2s_j \odot s_k - 2s_j \odot n + \|s_k + n\|^2 \quad (7)$$

where \odot indicates the sum of the pixel-by-pixel product of the two images (the dot product of the two images regarded as vectors). Note that the last term is just an additive constant, that does not depend on the index j . Thus minimizing the distance is equivalent to maximizing the quantity

$$g_j = s_j \odot s_k + s_j \odot n - \frac{1}{2}\|s_j\|^2. \quad (8)$$

We call the quantity g_j the discriminant. It is convenient to define \mathbf{C} as the $K \times K$ matrix of cross-correlations among the K neural images,

$$\mathbf{C}_{j,k} = s_j \odot s_k \quad j, k = 1, \dots, K. \quad (9)$$

Then we can rewrite the discriminant as

$$g_j = \mathbf{C}_{j,k} + s_j \odot n - \frac{1}{2}\mathbf{C}_{j,j}. \quad (10)$$

Note that the noise term $s_j \odot n$ is a vector of length K . When conducting Monte Carlo simulations, rather than constructing on each trial a new random image n with possibly millions of pixels, it is sufficient to directly construct a vector of length K , provided that its elements have the correct correlation. A method for constructing such a vector from the matrix \mathbf{C} and the noise standard deviation σ is described in Watson and Ahumada (2008). We write that vector $\mathbf{m}(\sigma, \mathbf{C})$. Finally, the discriminant can be written

$$\mathbf{g}_j = \mathbf{C}_{j,k} + \mathbf{m}(\sigma, \mathbf{C}) - \frac{1}{2}\mathbf{C}_{j,j}. \quad (11)$$

The observer locates the largest entry of \mathbf{g} , and returns its index j as the index of the optotype identified. This algorithm corresponds to the behavior of an ideal observer of a signal known exactly.

Pupil diameter

Pupil diameter has a substantial effect on the optical performance of the eye. Larger pupils produce a smaller diffraction-limited point-spread function, but admit more aberrations. In order to simulate the optical performance of the observers in the experiment of Zhang et al. (2007) it is necessary to select an

appropriate pupil diameter. Elsewhere we have derived a unified formula for pupil diameter that includes luminance, age, and adapting field size (Watson & Yellott, 2012).

The mean age of observers in Zhang et al. (2007) was 22.8 years. Display maximum luminance (and adapting background) was 89 cd/m². The display was viewed binocularly in a “dimly lit room.” Letter size was varied by changing the viewing distance, which ranged from 4.1 to 9.6 m. Display pixels were square with a width of 0.189 mm, yielding resolutions from 378.6 to 886.5 pixels/deg. The display subtended 2,048 by 1,536 pixels, so adapting field areas varied from 21.9 to 4 deg². Using these values in our unified formula gives predicted pupil diameters of between 6.5 and 5.6 mm (Watson & Yellott, 2012). Accordingly, we used a pupil diameter of 6 mm in our simulations.

IAS eyes

Because the optical characteristics of the observers of Zhang et al. (2007) were not recorded, we made use instead of a set of aberration coefficients measured for a population of 200 of healthy eyes in 100 observers in the IAS (Thibos et al., 2002). A file of the coefficients was provided to us (Thibos, personal communication). For the 6-mm pupil diameter used in this report, the file provided coefficients of modes 0–35 at a measurement wavelength of 633 nm.

Polychromatic optical transfer functions

The psychophysical data we analyzed were collected with polychromatic (white light) targets, whereas the wavefront aberration data we used in our simulations were recorded with monochromatic light at 633 nm. Under simplifying assumptions, it is possible to generate a polychromatic PSF or OTF from a set of monochromatic aberrations (Ravikumar, Thibos, & Bradley, 2008). The standard assumption is that only the defocus term varies with wavelength, due to longitudinal chromatic aberration. This assumption is based on aberration measurements at multiple wavelengths that show only small changes in terms other than defocus (Marcos, Burns, Moreno-Barriusop, & Navarro, 1999). Following procedures described previously (Nestares et al., 2003; Ravikumar et al., 2008), we computed monochromatic OTFs for a series of wavelengths centered on 555 nm, at intervals of 10 nm, extending from 405 to 695 nm. In each case, the Zernike coefficient corresponding to was computed based on a published formula (Thibos, Ye, Zhang, & Bradley, 1992),

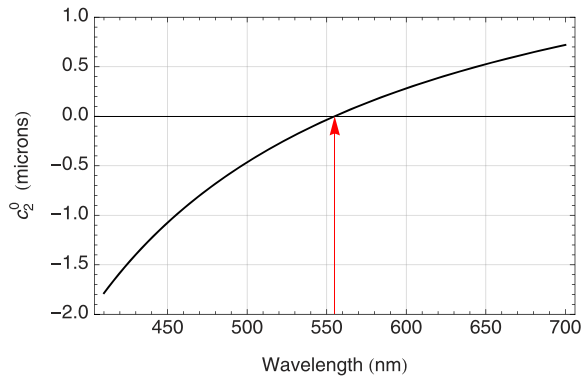


Figure A1. Change in the Zernike defocus coefficient as a function of wavelength due to longitudinal chromatic aberration.

$$D_{589}(m) = p - \frac{q}{m - c} \quad (12)$$

where D is defocus in diopters, m is the wavelength in micrometers, and p , q , and c are parameters ($p = 1.68524$, $q = 0.63346$, $c = 0.21410$). This describes absolute defocus for an eye in focus at ~ 589 nm. For an eye in focus at m_0 , the defocus at other wavelengths m will be

$$D(m, m_0) = D_{598}(m) - D_{589}(m_0). \quad (13)$$

This can be converted to a Zernike defocus coefficient in micrometers by the formula

$$c_2^0 = \frac{(m, m_0)d^2}{16\sqrt{3}} \quad (14)$$

where d is the pupil diameter in mm. This function is illustrated in [Figure A1](#) for the case of focus at 555 nm and a 6 mm pupil.

With defocus coefficients computed in this way, and added to the existing set of coefficients for each eye, a set of monochromatic OTFs was constructed for each wavelength. These OTFs were weighted by values of the photopic luminosity function at the corresponding wavelengths (normalized by their sum), and added together to create the polychromatic OTF.

Computing NTF

In [Watson and Ahumada \(2008\)](#), the NTF was derived by dividing a standard contrast sensitivity function by a hypothetical mean optical transfer function. Here we simplify the process by creating a function that approximates the previous result. First, we computed an array of discrete samples of the NTF as described in [Watson and Ahumada \(2008\)](#). We used a frequency scale value of 2, as defined in that publication. We then fit an EmG (exponential minus Gaussian) function ([Watson & Ahumada, 2005](#)) to those samples. The best fitting parameters were: $f_0 = 33.3573$, $f_1 = 5.37916$, $g = 3.32619$, $loss = 0.923853$. We then multiplied this function by the OEF with standard parameters ([Watson & Ahumada, 2005](#)). This function was then normalized to a peak value of 1. The final result was then used to compute the NTF component of the NOTF. This function is illustrated in [Figure A2](#).

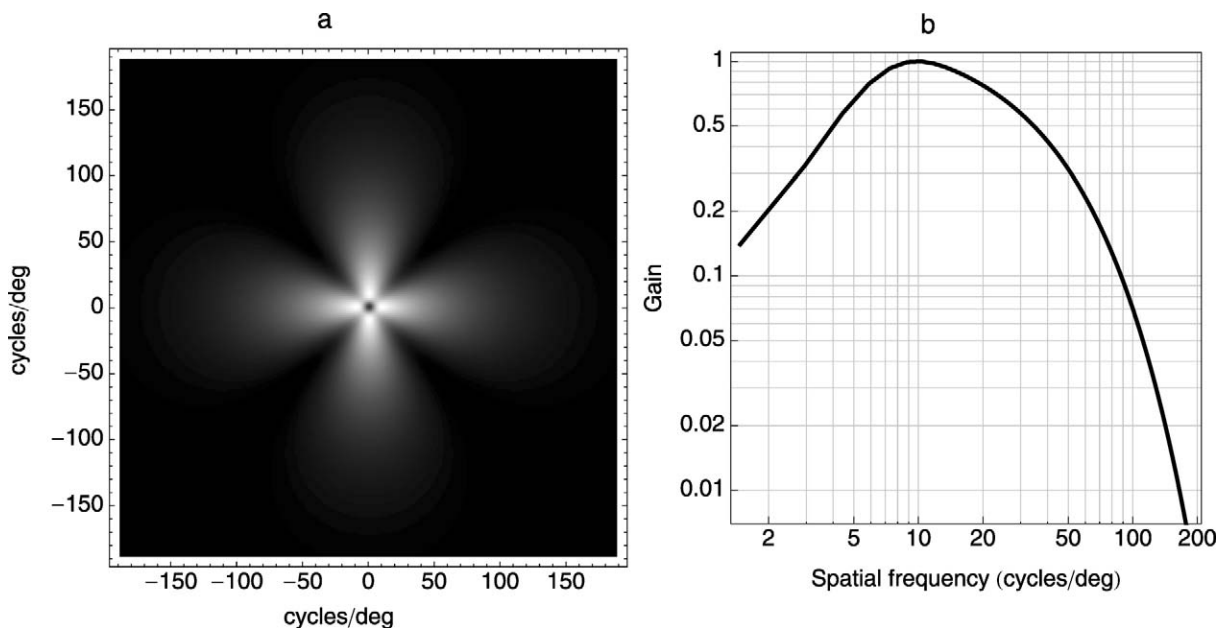
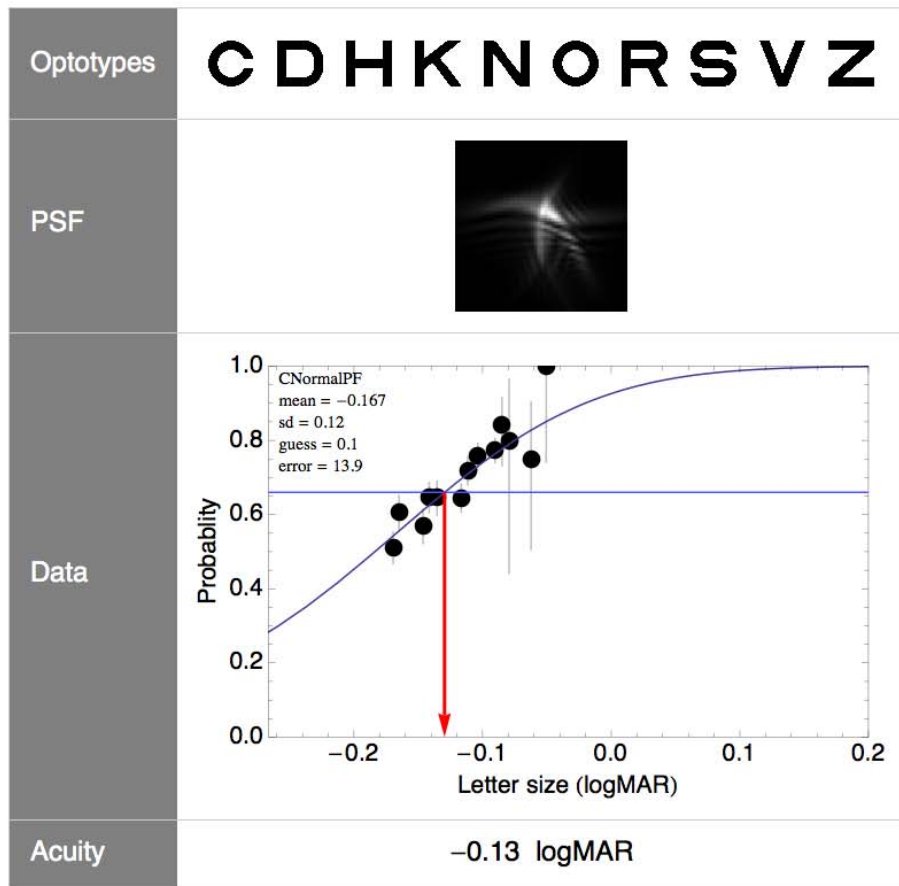


Figure A2. NTF used in the simulations. The filter is shown on the left, on the right is a cross section (half of one row) at vertical frequency = 0.

Acuity Model Demonstration



Optotype set 1 2 3 4 5 6 7

Model eye 1 2 3 4 5 6 7 8 9 10

Trials/optotype 100 500 1000 2000 5000 10000

Quest jitter 0 1 2 3 4 5 6 7 8 9 10

Target probability 0.5 0.55 0.66 0.75 0.82

Boost PSF image?

This demonstration illustrates a model of visual acuity. It shows measurement of visual acuity for a particular set of optotypes. A Quest procedure is used to vary the size of the optotypes to find the size that yields a particular percent correct (target probability). The simulation makes use of wavefront aberration data for ten eyes from the Indiana Aberration Study (Thibos, Hong, Bradley & Cheng ,2002). The demonstration allows the user to select the optotype set, a particular eye, the number of trials/letter in the simulation, the Quest jitter (that determines how widely trials are distributed over size), and the target probability. The optotype set, the pointspread function, and the simulated data are shown. The PSF images are either shown with a constant gain, or each one boosted to maximum contrast (for display only). This demonstration accompanies the article: Watson, A. B., & Ahumada, A. J., Jr. (2012). Modeling acuity for optotypes varying in complexity. *Journal of Vision*. Date of this version: 2012/07/28.

Figure A3. [Demonstration](#) of estimating model acuity using Quest.

Resizing optotypes

Optotypes were provided to us in the form of binary (0 or 1) digital images, with sizes approximately 40 × 40 pixels for Sloan letters and 50 × 50 pixels for Chinese

characters. In our simulations, these letter images were scaled to obtain the complete set of 32 sizes. To obtain each size, the original images were magnified or minified using the Mathematica ImageResize operator ([Wolfram Research, Inc.](#)). Each character is centered in

a 256×256 background image with a constant value of 1.

Estimating model acuity

To estimate acuity from the model we use the Quest adaptive procedure (Watson & Pelli, 1983). We select an optotype set and an IAS eye and precompute the matrix C (Equation 2) for each of the 32 letter sizes used by Zhang et al. (2007). We simulate K trials, one per optotype, at the middle size. Based on the model performance, Quest estimates the location of the acuity threshold, expressed as a likelihood function of the 32 letter sizes. The mode of the likelihood function is selected as the new size, and another K trials are presented. We again estimate a new location for threshold, and this process continues until T are completed for each letter. The resulting data are then fit by a normal distribution function of log size, using a maximum likelihood method (Watson, 1979), to yield an estimate of the letter size yielding the target probability correct. The Quest method provides a highly efficient way of estimating acuity from the model.

To illustrate this method, we provide a demonstration of an actual estimation of one acuity (Figure A3). The reader can select a set of optotypes and an IAS eye. One can also set several parameters of the method, including trials per optotype T , the target probability, and the Quest jitter. The last value is the width of a uniform distribution from which a number is drawn that is added to the selected test location at each step, in order to spread out the trials over more of the psychometric function. In actual use, this parameter was set to zero.

Computation of power spectral density

At times in this paper we have made use of the parameter σ to describe the standard deviation of the independent normally distributed noise samples added to each pixel in the simulated neural image. From the point of view of simulation, this is a convenient parameter. However, because this noise is in the

domain of the neural image, the value estimated for a given set of data will depend on (a) the spatial resolution of the simulation, and (b) the normalization of the neural transfer function.

There is no unique method of normalizing the neural transfer function. The optical transfer function, which converts light into light, and does not alter the total quantity of light, has a natural gain of 1 at DC (0 cycles/deg in horizontal and vertical frequency). For the neural transfer function, we have normalized it to its peak, as shown in Figure A2. Thus the combined transfer function, the product of neural and optical transfer functions, will have a peak gain that is typically somewhat less than 1, depending on the amount of optical attenuation around the peak of the NTF (around 5 cycles/deg). For the complete set of 200 IAS eyes, the mean is 0.414581 and the standard deviation is 0.096861.

An expression that characterizes the noise in a manner that is independent of spatial resolution is provided by the power spectral density, given by the expression

$$N = \sigma^2 dx dy \quad (15)$$

where dx and dy are the width and height of a single pixel. As an example, the estimate of σ for Sloan letters derived from confusion data is 0.257908. In the simulation, $dx = dy = 0.00264993$ deg (0.158996 arcmin). Thus,

$$N = 0.257908^2 \times 0.00264993^2 = 4.67 \times 10^{-7}. \quad (16)$$

Contrast energy in dBB

A convenient unit for the expression of contrast energy thresholds or visual power spectral densities is dBB (Watson & Ahumada, 2005; Watson, Taylor, & Borthwick, 1997). This is given by

$$dBB(x) = 10 \log_{10}(x) + 60. \quad (17)$$

This is a decibel measurement of energy or power, adjusted so that 0 dBB approximates the minimum visible contrast energy for a sensitive human observer (Watson, Barlow, & Robson, 1983). In the example of Equation 16, the result would be -3.306 dBB.