Higher-contrast requirements for recognizing low-pass–filtered letters

MiYoung Kwon

Department of Psychology, University of Minnesota, Minneapolis, MN, USA
Schepens Eye Research Institute, Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

Gordon E. Legge

Department of Psychology, University of Minnesota, Minneapolis, MN, USA

Kwon and Legge (2011) found that high levels of letter recognition accuracy are possible even when letters are severely low-pass filtered (0.9 cycles per letter). How is letter recognition possible with such severe reduction in the spatial resolution of stimulus letters? Clues may come from understanding the possible interaction between contrast and spatial resolution in letter recognition. Here, we asked what the effect is on the contrast threshold for detecting and recognizing letters as the spatial-frequency cutoff of letters is reduced (in cycles per letter). We measured contrast thresholds of seven normally sighted subjects for detecting and recognizing single letters of the English alphabet. The letters were low-pass filtered with several cutoff frequencies (0.9–3.5 cycles per letter, including unfiltered letters). We found that differences in contrast thresholds between detection and recognition increased substantially with decreasing cutoff frequency. We also incorporated the human contrast sensitivity function into an ideal observer model and found qualitatively good agreement between the pattern of performance for the model and our human subjects. Our findings show that the human visual system requires higher contrast for letter recognition when spatial resolution is severely limited. Good agreement between the model and human subjects shows that the greater contrast requirement for recognizing low-pass letters is due to a reduction in the information content of the letters rather than a change in human visual processing. The reduction in stimulus information may be due to increasing stimulus similarity associated with a reduction in spatial-frequency cutoff.

Introduction

Letter recognition is thought to rely on the shape and arrangement of individual features of a letter. A great deal of research has focused on identifying a set of pattern features (such as line segments and curves) mediating letter recognition (Blommaert, 1988; Bouma, 1971; Dunn-Rankin, 1968; Fiset et al., 2008; Geyer, 1970; Geyer & DeWald, 1973; Gibson, 1969; Laughery, 1969; Luce, 1963; Rumelhart & Siple, 1974; Townsend, 1971). Insight into the nature of these features comes from the observation that high levels of letter recognition accuracy are possible even when letters are severely blurred by low-pass spatial-frequency filtering (Kwon & Legge, 2011; Loomis, 1990). These studies showed that people can achieve 80% correct for recognition of 1 out of 26 when the letters are low-pass filtered with the cutoff frequency of 0.9 cycles per letter (CPL). This low resolution has an equivalent sampling rate ($<2 \times 2$) that would allow discrimination only among fewer than 16 patterns if the samples were binary (Shannon, 1948).[1] Furthermore, the cutoff frequency of 0.9 CPL is considerably lower than the known optimal spatial frequency for letter recognition (e.g., Chung, Legge, & Tjan, 2002; Gervais, Harvey, & Roberts, 1984; Ginsburg, 1980; Majaj, Pelli, Kurshan, & Palomares, 2002; Oruc & Landy, 2009; Parish & Sperling, 1991). It has been reported that the peak of the spatial-frequency band most useful for letter recognition ranges from about 1.7 cycles/letter for tiny characters (0.16°) to 7.7 cycles/letter for huge characters (16°; Majaj et al., 2002).

How is letter recognition possible with such severe reduction in the spatial resolution of stimulus letters? Clues may come from understanding the possible interaction between contrast and spatial resolution in letter recognition. Legge, Rubin, and Luebker (1987)

studied the role of contrast in reading speed and found that as letter size approached the acuity limit, more contrast was needed to achieve a criterion reading speed. An interaction is also found in letter recognition in peripheral vision, in which as acuity declines, images of letters (Melmoth & Rovamo, 2003; Rovamo & Melmoth, 2000) and faces (Mäkelä, Näsänen, Rovamo, & Melmoth, 2000; Melmoth, Kukkonen, Mäkelaä, & Rovamo, 2000) need to be scaled not only in size but also in contrast to match recognition performance in central vision. The interaction between contrast and spatial resolution of letters has been recognized in clinical practice, and there has been interest in measuring visual acuity with low-contrast charts. For example, the Regan Letter Chart, a low-contrast letter-acuity chart, has been used to assess the effects of reduced contrast on visual acuity (Regan, 1988; Regan & Neima, 1983).

Our interest in understanding letter recognition under conditions of low-resolution viewing is motivated by real-world applications. Examples include reading near the acuity limit (highway signs at a great distance) or coping with fog, low-resolution display rendering, refractive error, or low vision. To our knowledge, no study has addressed contrast requirements for recognizing letters with low resolution. Our primary goal is to examine the impact on letter recognition of the interaction between the contrast of letters and the spatial resolution with which they are rendered. More specifically, as the spatial-frequency cutoff of letters is reduced (in cycles per letter), what is the effect on the contrast threshold for detecting and recognizing letters?

A second goal is to determine whether this interaction helps us to understand differences in letter recognition between central and peripheral vision or between upper and lowercase letters. Differences in the shape of the human contrast sensitivity function (CSF) mean that letter stimuli in peripheral vision have reduced spatial-frequency representations for neural processing. Uppercase letters may be thought to be rendered with lower resolution than lowercase letters because they possess fewer distinguishable spatial features (e.g., no ascenders or descenders). The idea that some stimulus conditions are more vulnerable to low spatial resolution than others is indeed substantiated by our related study (Kwon & Legge, 2011). In that study, we found that to achieve reliable letter recognition (80% accuracy), spatial resolution of letters had to be higher (manifested as a larger minimum spatial-frequency requirement) in peripheral (1.06 cycles per letter) than central vision (0.9 cycles per letter) and for uppercase (1.14 cycles per letter) than lowercase letters (0.9 cycles per letter). In the current experiments, we measured contrast thresholds for detecting and recognizing single letters in central and

peripheral vision, drawn at random from the 26 lower and uppercase letters of the English alphabet. The letters were low-pass filtered (blurred) with various cutoff frequencies. We used the size of the gap between detection and recognition contrast thresholds for letters as a measure of the contrast requirement in the letter recognition task. We did so because the gap between detection and recognition thresholds would reflect the true contrast requirement for recognition after factoring out any possible differences in detection threshold induced by different stimulus conditions. Comparing recognition and detection thresholds in this way also provided us with a useful way of understanding the performance of the ideal-observer model and the similarity model to be discussed later.

A third goal of the present study is to examine whether the greater contrast requirement for recognition of letters at low spatial resolution, if any, is inherent in the stimulus or an intrinsic property of human visual processing. An ideal-observer model, a theoretical device that yields the best possible performance for a given task via a strategy of choosing the maximum posterior probability (Green & Swets, 1966; Tanner & Birdsall, 1958), is a quantitative method for demonstrating the stimulus constraints on performance. Some limitations on early visual processing such as visual acuity and contrast sensitivity can be thought of as transforming the stimulus input, producing an "equivalent" visual input. For instance, the human CSF puts a lower bound on contrast that can be detected. For modeling purposes, it can be useful to treat these early sensory limitations as transformations of the stimulus input. Previous studies of letter recognition have demonstrated that incorporating the human CSF in an ideal-observer model can explain aspects of the spatial-frequency characteristics of human letter recognition (Chung et al., 2002; Kwon & Legge, 2011; Watson & Ahumada, 2008). We incorporated a human CSF into the ideal-observer model (we now call it the CSF-noise-ideal-observer model). We tested this model for the tasks of low-pass–filtered letter detection and recognition for comparison with human performance.

Method

Subjects

Seven subjects were recruited from the University of Minnesota campus. They were all native English speakers with normal or corrected-to-normal vision. The mean acuity (Lighthouse distance acuity chart) was −0.11 logMAR (Snellen 20/16), ranging from −0.24 (Snellen 20/11) to 0.02 (Snellen 20/21). The mean LOG

contrast sensitivity (Pelli-Robson chart) was 1.74, with a range from 1.65 to 1.90. Subjects received either monetary compensation or class credit for their participation. The experimental protocols were approved by the Internal Review Board of the University of Minnesota, and written informed consent was obtained from all subjects prior to the experiment.

Stimuli

Contrast definitions

The stimulus contrast is expressed as the Weber contrast defined to be

$$C_i = \frac{L_i - L_0}{L_0}, \tag{1}$$

where L_i is the luminance of the ith pixel of an image and L_0 is the mean luminance of the stimulus image. Once we define the contrast of L_i as Weber contrast, then the root-mean square (RMS) contrast of the image is expressed follows:

$$C_{RMS} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} C_i^2}, \tag{2}$$

where m is the number of pixels in an image.

Contrast was defined for both filtered and unfiltered images as follows: After filtering, each pixel value of the filtered image was converted into a value of C_i by the Equation 1 (i.e., constructing a contrast function for each letter image). The nominal contrast of the filtered image was defined as the maximum contrast among all the pixel contrast values. This *max C_i* value was used for the purpose of measuring contrast thresholds for detection and discrimination tasks. Then, the RMS contrast of an image for given threshold level (i.e., the RMS contrast of letter "x" was used to define the given threshold level contrast) was computed and used for all of the data analysis and plotting. The reason for using the nominal contrast (defined as Weber contrast) for threshold measurement rather than RMS contrast is that rendering an image at a specific contrast level is more straightforward with the nominal contrast.

Letter images

The 26 lowercase and uppercase letters of the English alphabet were used (Courier font), with x-height of 1° (31 pixels) at the viewing distance of 60 cm. The letter images were constructed in Adobe Photoshop (version 8.0) and MATLAB (version 7.4). A black single letter was generated on a uniform gray background of 400 × 400 pixels.

Image filtering

The images were blurred using a third-order Butterworth low-pass filter in the spatial-frequency domain. Cutoff frequencies of the filter ranged from 0.9 CPL to 3.5 CPL depending on task and stimulus conditions. The filter function is

$$f = \frac{1}{\left(1 + \left(\frac{r}{c}\right)^{2n}\right)}, \tag{3}$$

where r is the radial frequency, c is the cutoff spatial frequency, and n is the filter's order.

The filter's response function is shown in Figure 1. Figure 2 shows samples of filtered and unfiltered letter images.

Image display on screen

To present the letter images on the monitor, we mapped the values of the contrast function to the corresponding luminance values of the monitor, and then each luminance value was mapped to the corresponding 256 gray levels using a lookup table. The 0 value of the contrast function was mapped to the mean luminance of the monitor (52 cd/m^2).

The stimuli were generated and controlled using Matlab (version 7.4) and Psychophysics Toolbox extensions (Mac OS X; Brainard, 1997; Pelli, 1997), running on a Mac Pro computer. The display was a 19-inch CRT monitor (refresh rate: 75 Hz; resolution: 1152 × 870). Stimuli were rendered with a video card with 8-bit input resolution and 14-bit output resolution using Cambridge Research System Bits++. Luminance of the display monitor was linearized using an 8-bit lookup table in conjunction with photometric readings from a Minolta CS-100 chroma meter. The image luminance values were mapped onto the values stored in the lookup table for the display.

Procedure

Measuring contrast detection thresholds for letters

Contrast detection thresholds were measured with a temporal two-alternative forced-choice staircase procedure. A 3-down 1-up staircase rule was adopted, yielding a threshold criterion of 79.4% correct (Wetherill & Levitt, 1965), and the step size of the staircase was 1 dB. The geometric mean of 10 staircase reversals was taken as the contrast threshold for each staircase run. Detection thresholds were obtained from letter images with different spatial-frequency cutoffs ranging from 0.9 CPL to 3.5 CPL and also unfiltered images. Slightly different sets of cutoffs were used for different testing conditions (see Table 1). This was because the
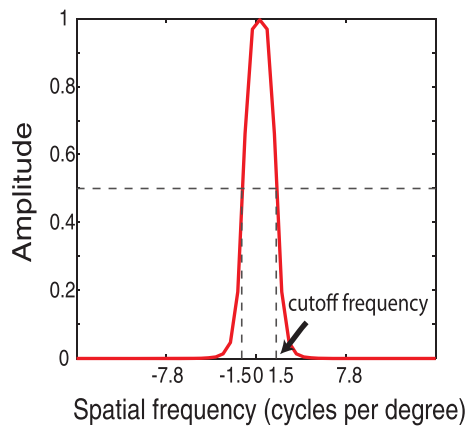
Figure 1. The response function of the third-order Butterworth filter with cutoff frequency of 1.5 cycles per degree, equivalent to 1.5 cycles per letter for a 1° letter size.

minimum cutoffs for a criterion level of recognition varied across conditions (Kwon & Legge, 2011).

In each trial, there were two 150-ms intervals each marked by an auditory beep, separated by 500 ms, one containing the stimulus, selected at random from a to z or A to Z. The subjects' task was to judge which stimulus interval contained a letter image by pressing one of two keys. (They did not have to identify the letter.) Auditory feedback was given whenever a wrong answer was made. Subjects were given a series of practice trials before the experimental test.

Measuring contrast recognition thresholds for letters

Recognition contrast thresholds were measured with the same staircase procedure just described for the detection task. In each trial, subjects were presented with a stimulus letter, randomly selected from a to z or A to Z for 150 ms at a given location on a display screen. Next, the display was set to average luminance, and after a brief pause (500 ms), 26 thumbnail versions of the letter images (56 × 56 pixels in size) appeared on the screen. The subject identified the target stimulus by clicking on one of these 26 thumbnail images. To prevent subjects from using an image-matching strategy, a different font (Arial) and the other lettercase were used for the thumbnail images. Auditory feedback was given whenever a wrong answer was made.
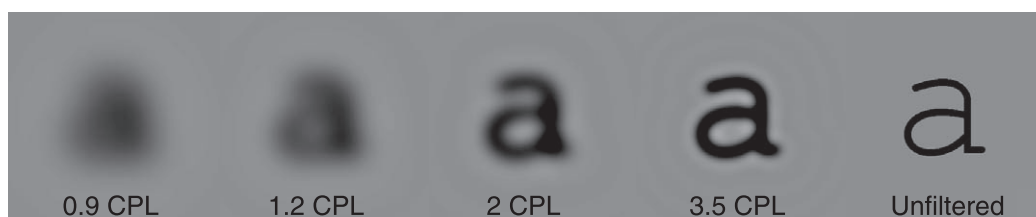
Subjects were given a series of practice trials before the experimental test.

For the peripheral viewing condition (in both detection and recognition tasks), a small cross in the center of the stimulus served as a fixation mark to minimize eye movements throughout the experiment. A chin rest was also used to reduce head movements.

The experiment consisted of 36 blocks: 4 (fovea) or 5 (peripheral) cutoffs × 2 lettercases (upper and lowercase) × 2 tasks (detection and recognition) tested over 2 separate days. All subjects participated in all the conditions. The order of blocks was counterbalanced across subjects.

Results

Contrast thresholds for detection and recognition of letters

Figure 3 shows plots of RMS contrast threshold as a function of filter cutoff spatial frequency for both detection (black circles) and recognition (red squares). Recognition required higher RMS contrast thresholds than detection. Detection thresholds were nearly constant across different cutoff frequencies, except for a noticeable threshold elevation for low-cutoff frequencies in central vision (up to 14% change) and for high-cutoff frequencies in the periphery (up to 20% change). This pattern of results can be accounted for by a shift in the peak of the CSF from a higher spatial frequency (~four cycles per degree) in the fovea to a lower spatial frequency (~one cycle per degree) in the periphery. Unlike the detection data, recognition thresholds decreased substantially with increasing cutoff frequency and reached asymptote near 2 CPL. Both detection and recognition thresholds were larger in peripheral than central vision.

Ratio of recognition to detection thresholds

The gap between detection and recognition RMS contrast thresholds was quantified as the ratio of

Figure 2. Samples of an unfiltered letter and low-pass–filtered letters with varying cutoff frequencies.

Lettercase	Lowercase		Uppercase	
Visual field	Fovea	Periphery	Fovea	Periphery
Cutoff Frequency (CPL)	0.9	NA	1.1	NA
	1.2	1.2	1.5	1.5
	2	2	2	2
	3.5	3.5	3.5	3.5
	Unfiltered	Unfiltered	Unfiltered	Unfiltered

Table 1. Spatial-frequency cutoffs used for different stimulus conditions.

recognition to detection RMS contrast thresholds equivalent to their difference on a logarithmic scale.

We conducted an analysis of variance (ANOVA) on threshold ratio: 4 (cutoff frequency: 1.2, 2, 3.5, unfiltered) × 2 (visual field: fovea, periphery) × 2 (lettercase: lower, upper) repeated-measures ANOVA with cutoff frequency, visual field, and lettercase as within-subject factors.[2]

There were significant main effects of cutoff, $F(3, 18) = 34.04$, $p < 0.001$, visual field, $F(1, 6) = 14.33$, $p = 0.009$, and lettercase, $F(1, 6) = 10.20$, $p = 0.019$, on threshold ratio. The ratio increased substantially with decreasing cutoff frequency (Figure 4). For example, for the foveal lowercase condition, the ratio increased from 1.41 (±0.07) for the unfiltered letters to 8.93 (±1.07) for the most blurred letters (i.e., 0.9 CPL). The
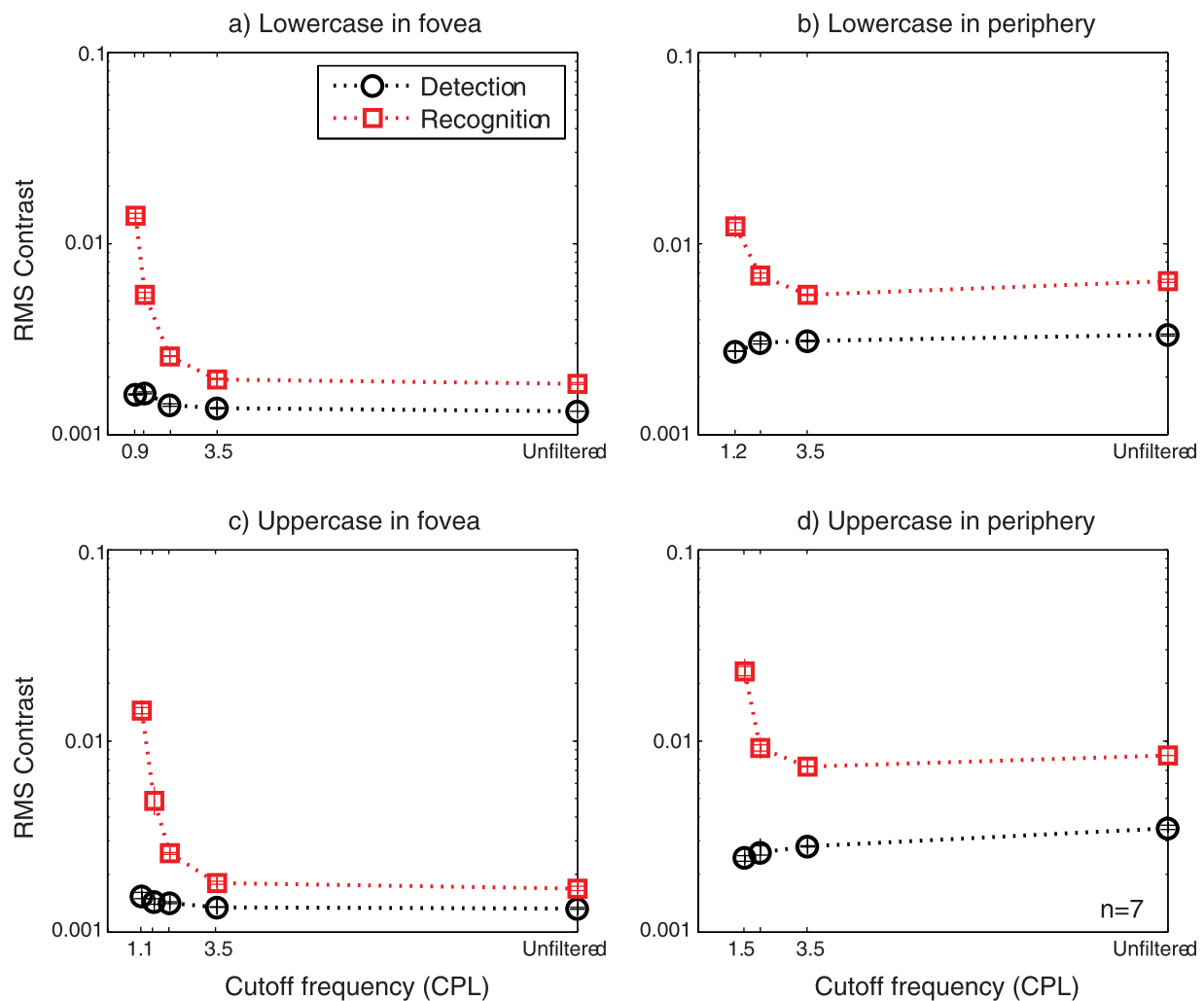

Figure 3. Mean threshold RMS contrast for letter detection and recognition ($n = 7$) as a function of cutoff spatial frequency. Error bars show ±1 standard error of the mean (SEM).
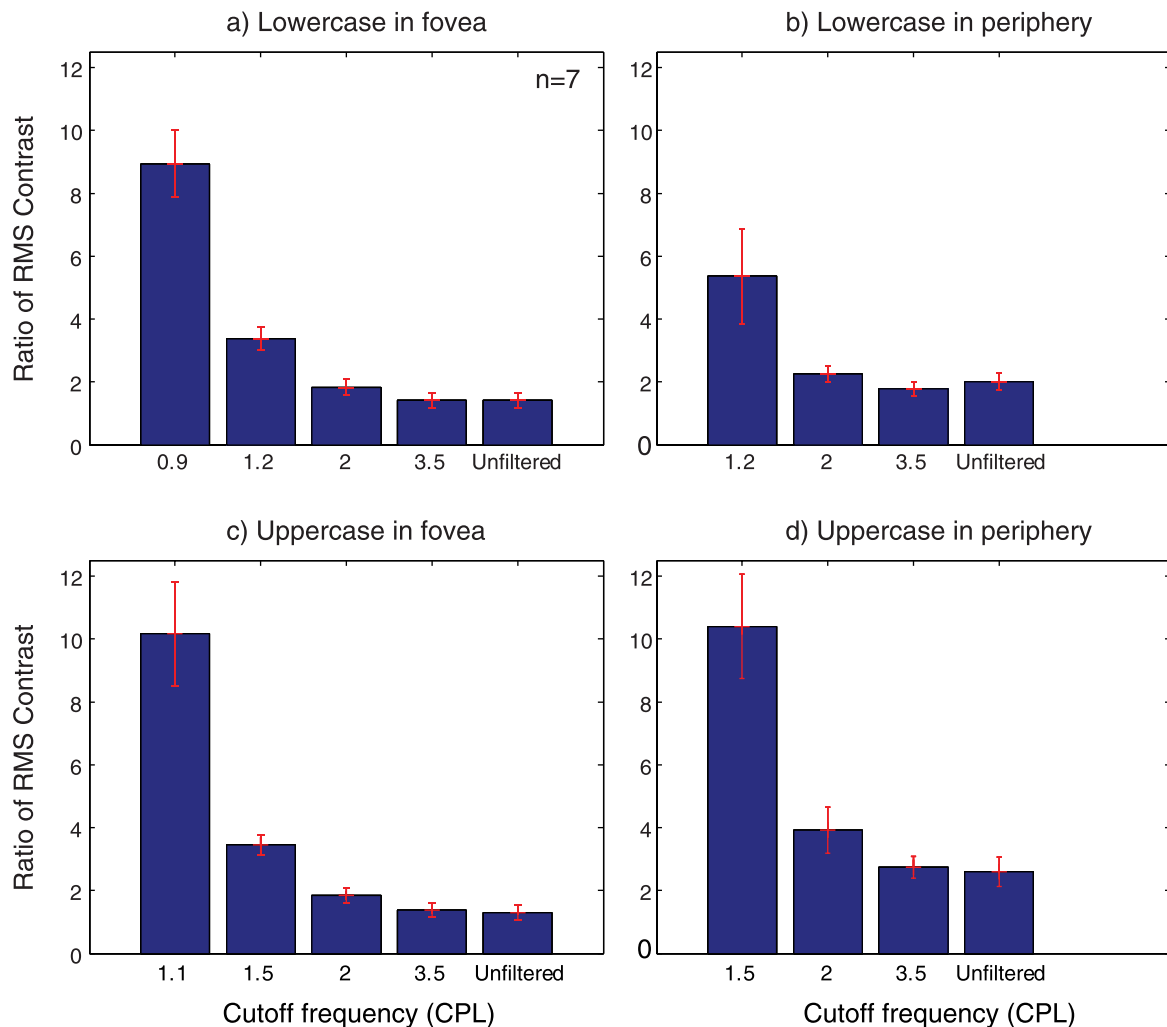
Figure 4. Mean ratios of recognition to detection RMS contrast thresholds as a function of cutoff spatial frequency ($n = 7$). Error bars show ± 1 SEM.

ratios were larger in peripheral than central vision and also larger for uppercase than lowercase letters.

We also found significant two-way interactions between cutoff frequency and lettercase, $F(3, 18) = 5.81$, $p = 0.006$, between cutoff frequency and visual field, $F(3, 18) = 10.48$, $p < 0.001$, and between visual field and lettercase, $F(1, 6) = 8.14$, $p < 0.029$. The differences in the ratios between the two lettercases and between the two visual field locations were more pronounced at lower cutoff frequencies.

The larger gap between recognition and detection thresholds for low-cutoff frequencies makes clear that recognition of very low-resolution letters has a higher contrast requirement. Threshold ratios were greater in peripheral than central vision and for uppercase than lowercase letters.

The average ratio of recognition to detection thresholds across lettercases and visual fields for the unfiltered letters was 1.8 (± 0.6). This value is close to the mean value of 1.7 found by Pelli, Burns, Farell, and

Moore-Page ([2006](#)), who also measured contrast detection and recognition thresholds for letters. Their value was based on an average across many fonts for single letters and short words. These authors used the threshold ratio in their theoretical derivation of the number of feature detectors for letter recognition. They found that the recognition/detection ratio was constant at 1.7 across sets of letters of different complexity and used this result to conclude that identifying 1 of 26 letters (at 64% correct) is based on 7 ± 2 feature detections.

The plots of the threshold ratio in [Figure 4](#) indicate that the ratio is constant for high-cutoff frequencies and grows for cutoff frequencies below a critical value. We refer to this critical point as the "contrast-dependent" critical cutoff frequency for letter recognition (hereafter *critical cutoff frequency* for convenience). We use this term because it refers to the low-pass cutoff frequency required to recognize letters at minimum contrast.
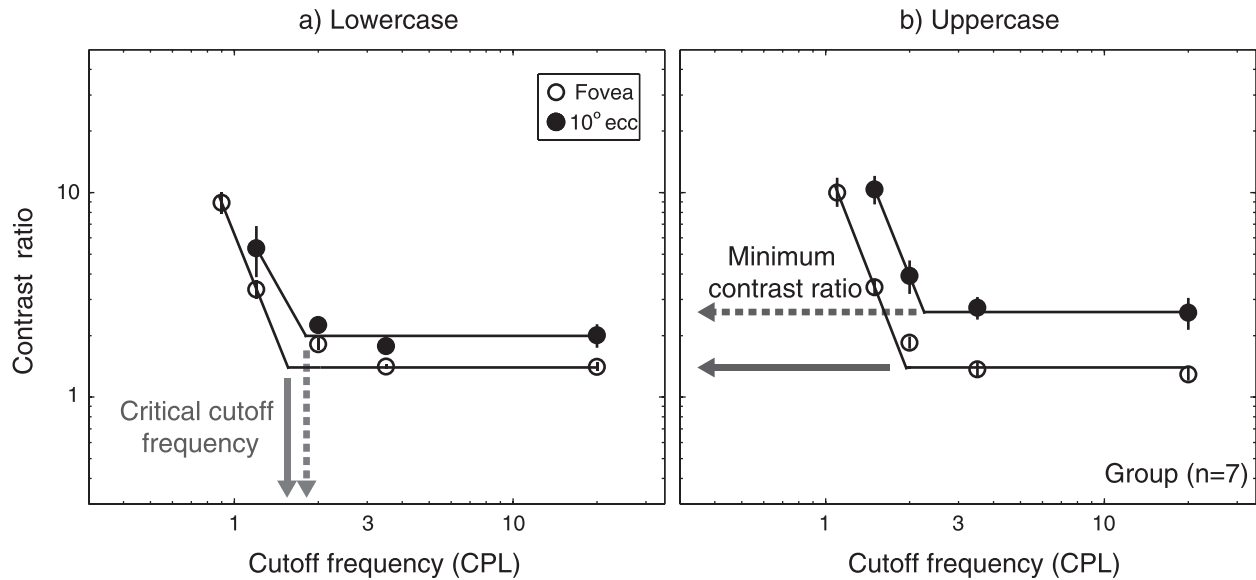
Figure 5. Mean contrast ratio as a function of cutoff spatial frequency ($n = 7$). (a) Lowercase letters. (b) Uppercase letters. Each panel contains two data sets: one from the fovea (open circles) and the other from the periphery (closed circles). Error bars show ±1 SEM. Data were fitted with the two-limbed function. The horizontal arrows indicate estimated minimum contrast ratios. The vertical arrows indicate estimated critical cutoff frequencies.

To estimate this critical cutoff frequency, we fitted the graphs of contrast ratio versus cutoff frequency (Figure 4) with a two-limbed function (Equation 4). This function contains a rising straight line and a horizontal straight line (Figure 5). The X-coordinate of the intersection point is called the *contrast-dependent critical cutoff frequency*. The Y-coordinate is called the *minimum contrast ratio*.

$$Y = b, \qquad \text{if } X \geq c \qquad\qquad (4)$$

$$Y = a*X + b - a*c, \qquad \text{if } X < c,$$

where Y is contrast ratio, X is spatial-frequency cutoff frequency, and a, b, and c are free parameters.

As shown in Figure 5, the average critical cutoff frequency across subjects for lowercase letters in the fovea was 1.47 CPL (±0.04), and the minimum (asymptotic) contrast ratio was 1.56 (±0.05). Table 2 summarizes estimated critical cutoff frequencies and minimum contrast ratios for our four stimulus conditions. Note that values of both critical cutoff frequency and minimum contrast ratio increased from central to peripheral vision (e.g., $t(6) = -3.04$, $p = 0.023$), and also

from lowercase to uppercase letters (e.g., $t(6) = -4.46$, $p = 0.004$).

Modeling

Model overview

We now describe the implementation of the CSF-noise-ideal-observer model (hereafter the model) for letter detection and recognition. This model is a computational device that includes an optimal recognition decision rule and takes into account an internal source of noise and the shape of the human CSF. This model is very similar to the CSF-ideal-observer model described by Chung et al. (2002). If the information content of the stimulus limits human performance rather than inherent human visual processing, we would expect the pattern of results of the model to be similar to our empirical findings.

Many prior studies comparing human performance to an ideal observer have included noise-perturbed stimuli. In principle, the ideal observer is able to

	Lowercase		Uppercase	
	Fovea	Periphery	Fovea	Periphery
Critical cutoff frequency (CPL)	1.47 (±0.05)	1.80 (±0.06)	1.95 (±0.18)	2.25 (±0.12)
Minimum contrast ratio	1.56 (±0.04)	1.99 (±0.15)	1.46 (±0.13)	2.64 (±0.40)

Table 2. Critical cutoff frequencies yielding minimum contrast ratios ($n = 7$). *Notes:* The numbers in parentheses indicate ±1 SEM.
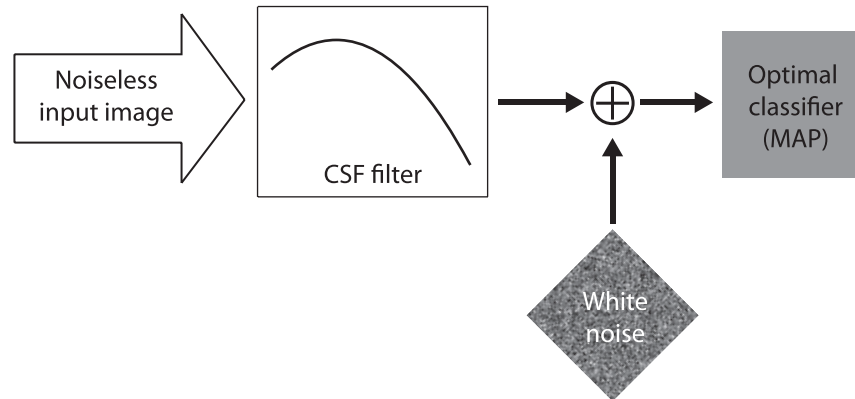
Figure 6. A schematic diagram of the CSF-noise-ideal-observer model.

perform with 100% accuracy if the stimulus is noise free. The noise is added to the stimulus to ensure that even an optimal decision maker will make incorrect decisions and drop below perfect performance. In our study, no noise was added to the stimulus. Instead, we constructed the model to have a source of additive noise following stimulus encoding and prior to the decision process, which may mimic the internal noise of human visual processing. The amplitude of this noise was a parameter in the model. The noise parameter was set by varying its value to equate overall performance levels between the model and our human subjects for the unfiltered lowercase foveal condition. Our goal was then to determine whether the model and humans show the same pattern of threshold changes across stimulus cutoff frequency, central and peripheral vision, and upper and lowercase letters.

The model has a CSF filter[3] and an additive noise source situated between the stimulus and an optimal classifier as depicted in Figure 6. The CSF filter is a linear filter with a shape identical to a human CSF. The model performs letter detection and recognition in a procedure mimicking the human psychophysical task. In each trial, the model is presented with a stimulus (the contrast function associated with the low-pass–filtered letter stimulus). This function is then filtered by the CSF, and independent samples of Gaussian noise are added to each pixel. The model knows the probabilities of the possible signals and the statistics of the added Gaussian noise.

The model optimizes performance for detection or recognition by choosing the maximum posterior probability of the signal arising from 1 of the 26 possible stimulus images given the noisy input stimulus (Green & Swets, 1966; Tanner & Birdsall, 1958; Tjan, Braje, Legge, & Kersten, 1995). The computation of the maximum posterior probability is equivalent to minimizing the Euclidean distance between the noisy input stimulus and its stored noiseless templates, often called *template matching*. But note that the algorithms are

slightly different for detection and recognition tasks (details are provided in Appendix A). This numerical analysis was done through Monte Carlo simulations.

The model's contrast thresholds for detection and recognition were obtained via the same staircase procedure (~79.4% accuracy) used for human observers. The contrast of a stimulus image was defined and computed in the same way as those used for the human observers (see the Method section). The contrast thresholds were measured with both lowercase and uppercase letters in foveal and peripheral viewing conditions using the identical cutoff frequencies used for human observers.

Our model used empirical CSFs from two subjects who also participated in our detection and recognition tasks. The CSFs were obtained from a detection task using a vertical sinewave-grating (the center of the patch was in cosine phase) with cosine envelope (subtending 1.4° visual angle at the viewing distance of 60 cm) and stimulus duration of 150 ms. The CSFs were measured at the fovea and at 10° in the lower visual field (Figure 7). The CSFs were measured with similar spatial and temporal stimulus characteristics used in our empirical letter detection and recognition tasks.

Comparison of model and human results

Figure 8 shows plots of the model's RMS contrast thresholds as a function of cutoff frequency for both detection (black circles) and recognition (red squares). The model's behavior is qualitatively similar to human behavior. The model required larger RMS contrast thresholds for recognition than detection. Like human observers, recognition thresholds for the model increased with decreasing cutoff frequency. Also similar to humans, recognition thresholds of the model were higher in peripheral than central vision and were higher
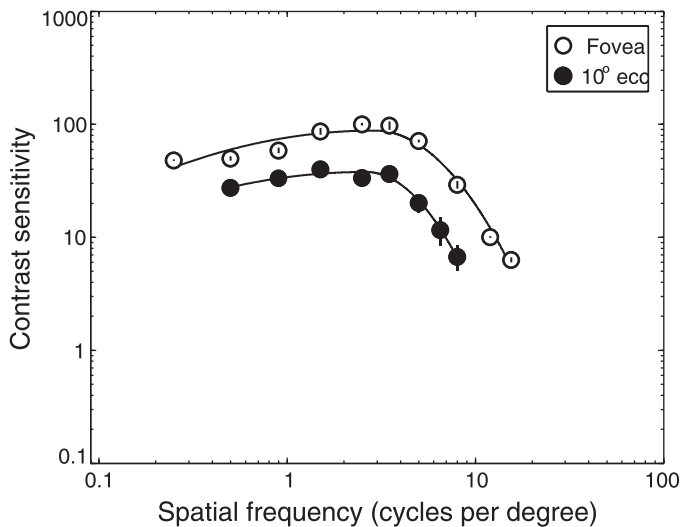
Figure 7. The CSF of human observers (averaged across two subjects) at the fovea (open circles) and at 10° lower visual field (closed circles). The CSFs were obtained from a detection task using a vertical sinewave-grating (the center of the patch was in cosine phase) with cosine envelope (subtending 1.4° visual angle at the viewing distance of 60 cm) and stimulus duration of 150 ms. The CSFs were measured with similar spatial and temporal stimulus characteristics used in our empirical letter detection and recognition tasks.

for uppercase than lowercase letters. One noticeable difference between the model and human observers is that detection thresholds of the model decreased with decreasing cutoff frequency. This is probably due to decreasing stimulus uncertainty associated with blurry letters.[4]

Consistent with human observers, the ratio of recognition to detection thresholds increased substantially with decreasing cutoff frequency. For example, for the foveal lowercase condition, the model's ratio increased from 1.46 (± 0.02) for the unfiltered letters to 6.77 (± 0.11) for the most blurred letters (0.9 CPL; Figure 9). The corresponding human ratios were 1.41 (± 0.07) for the unfiltered letters and 8.93 (± 1.07) for the most blurred letters (i.e., 0.9 CPL).

Like human observers, the model also exhibited its maximum ratio for the peripheral uppercase letters with the cutoff frequency of 1.5 CPL (10.40 $\pm$ 1.65) and its minimum ratio for the unfiltered foveal lowercase letters (1.46 $\pm$ 0.02). The corresponding human ratios were 10.23 (± 0.16) and 1.41 (± 0.07), respectively.

Relative to detection threshold, humans seem to be as effective as an ideal model in recognizing unblurred letters but less able to use the information in blurry letters; humans need a greater increase in contrast to recognize the blurry letters, as demonstrated by the

steeper rise in the human threshold ratio at the lowest cutoff frequencies (Figure 9).

Discussion and conclusions

In the present study, we demonstrated that as spatial resolution for rendering letters decreases, the visual system requires higher contrast for letter recognition. This means that there is a larger gap between contrast thresholds for letter detection and recognition. The gap between these two thresholds was quantified as the ratio of RMS contrast thresholds for letter recognition and detection. We found larger recognition/detection contrast ratios for letters with lower cutoff frequencies. The contrast ratios were even larger for stimulus conditions that suffer from poor spatial resolution (peripheral vision) and fewer distinguishable spatial features (uppercase letters). To our knowledge, this is the first empirical evidence showing a higher contrast requirement for letters with low spatial resolution.

Previous studies (e.g., Legge et al., 1987) have shown that reading and letter recognition are usually highly tolerant to contrast reduction for a wide range of letter size (0.25°–2°). They showed that for normally-sighted individuals, reading rate is little affected by the contrast of letters over a 1-log-unit range from about 100% to 10%. But the current results imply that even for letters in this size range, there is a higher contrast requirement when the letters have very low spatial resolution.

Our results may be relevant to people with central-field loss from macular degeneration who must rely on their peripheral vision to read. We found that there is an even higher contrast requirement for peripheral vision than central vision for letters rendered in low resolution. As mentioned earlier in the Introduction, the need for a larger contrast reserve for peripheral viewing has been demonstrated by previous studies (Mäkelä et al., 2000; Melmoth et al., 2000; Melmoth & Rovamo, 2003; Rovamo & Melmoth, 2000). Those authors measured contrast sensitivity for face and letter identification to see if foveal and peripheral performance would become equivalent by magnification of image size only. They found that to achieve equivalent performance, both the size and the contrast of the image needed to increase in the periphery.

Furthermore, Rubin and Legge (1989) studied the effect of contrast on reading performance in 19 low-vision observers with a wide range of visual disorders and degrees of vision loss. They found that unlike normally-sighted individuals, visually impaired individuals showed less tolerance to contrast reduction. Thus, as the clinical community already knows through practical experience, contrast is an important visual dimension in designing low-vision reading aides.
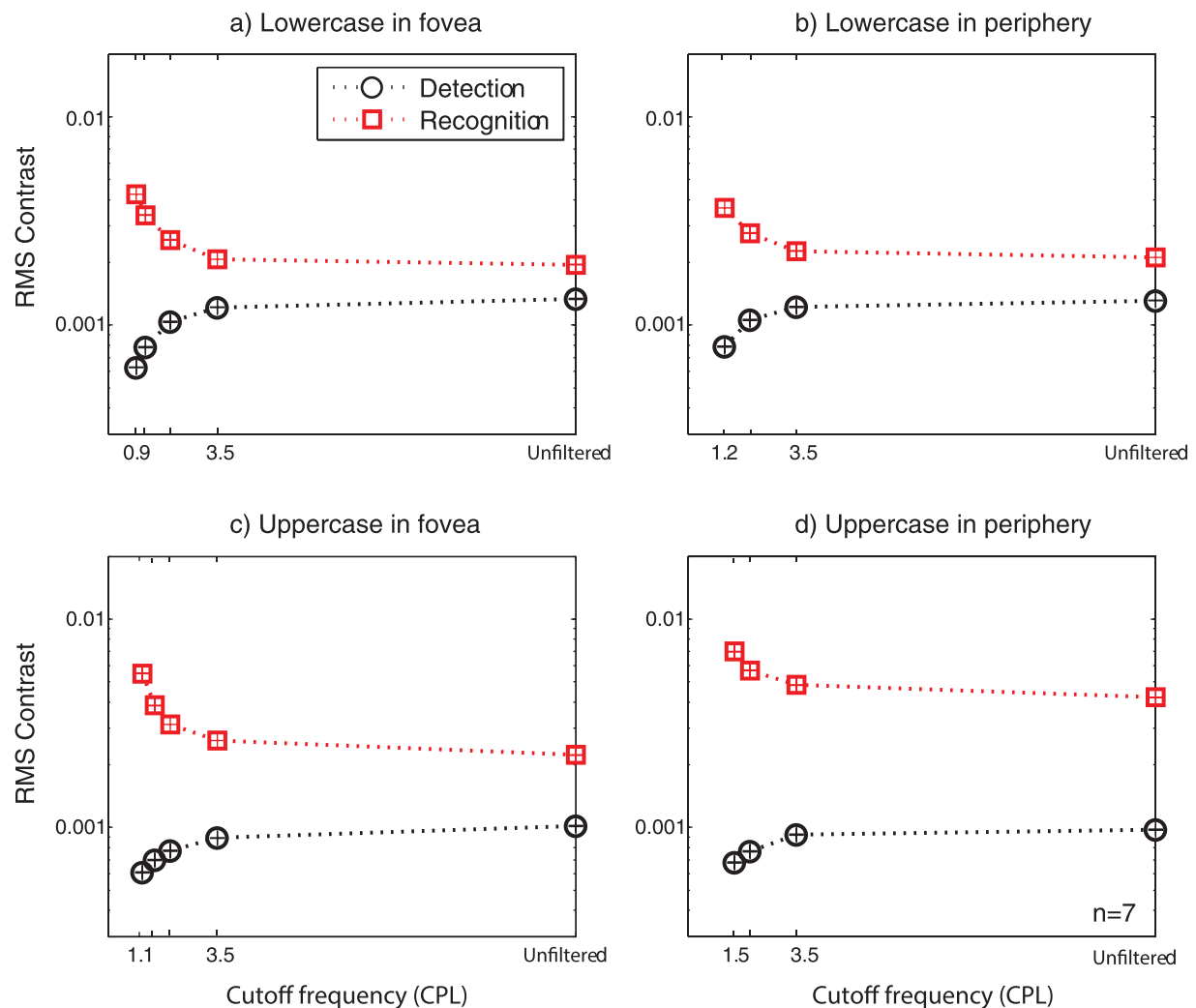
Figure 8. Mean threshold RMS contrast for letter detection (black circles) and recognition (red squares) from the model. Each mean threshold is based on 100 thresholds, each obtained from a staircase procedure using 50 reversals. Error bars show ± 1 SEM.

Using the ideal observer model, we further asked whether the higher contrast requirement for letters at low spatial resolution exhibited by human observers is due to the information content of the stimulus or due to intrinsic properties of human visual processing. The qualitatively similar pattern of results for ideal and human observers makes it likely that human performance is primarily due to the information content of the stimuli.

What stimulus properties account for the higher contrast requirement for recognizing blurry letters? Letter similarity has been studied extensively (Blommaert, 1988; Bouma, 1971; Dunn-Rankin, 1968; Geyer, 1970; Geyer & DeWald, 1973; Gibson, 1969; Laughery, 1969; Loomis, 1990; Luce, 1963; Rumelhart & Siple, 1974; Townsend, 1971). It has been shown that the more similar the letters, usually represented by overlap in their features, the more likely those letters are to be confused with each other, resulting in poor recognition

performance. Intuitively, an increase in blur level brings about an increase in similarity among letters. To quantify the similarity among letters filtered with a given cutoff frequency, we computed normalized cross-correlations between all the possible pairs of letters and obtained the mean correlation value averaged across all the pairs. In pattern recognition, normalized cross-correlation has often been used to measure similarity between sets of images (Duda & Hart, 1973; Watson & Ahumada, 2008). The circles in Figure 10a represent normalized cross-correlation values for the different spatial-frequency cutoffs and four stimulus conditions (lowercase and uppercase letters; foveal and peripheral viewing conditions). As we expected, the similarity (mean cross-correlation) increased monotonically with increasing blur level. Using regression analysis (dotted lines in Figure 10a), we found that 69% (the model) and 77% (human observers) of the variance in the
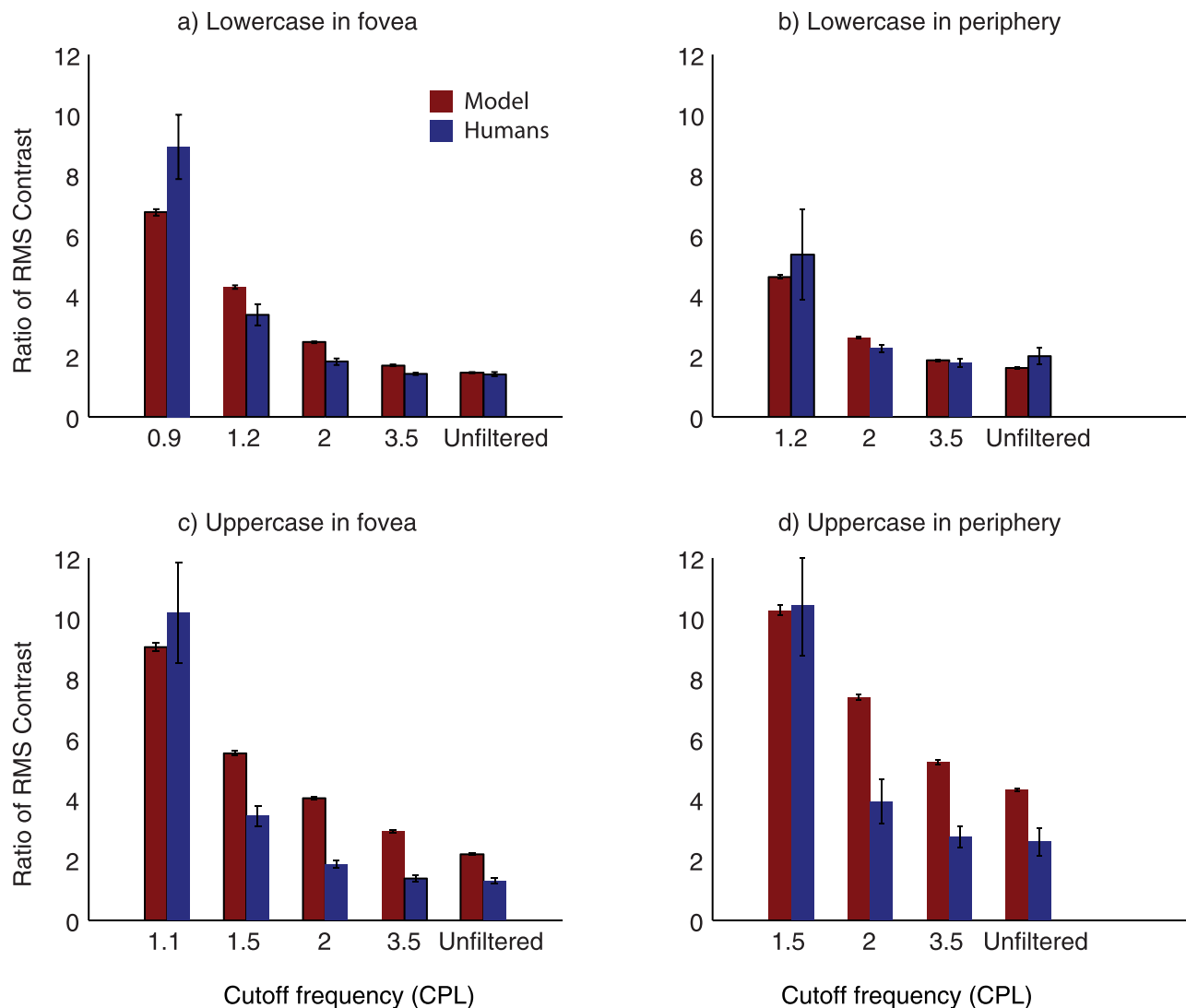
Figure 9. Mean ratios of recognition to detection RMS contrast thresholds as a function of cutoff frequency for our CSF-noise-ideal-observer model. Each panel includes human data from Figure 4 for comparison (red bar for the model; blue bar for human data). Error bars show ±1 SEM.

recognition:detection threshold ratios were accounted for by this letter similarity measure.

Why should increasing similarity require higher contrast for letter recognition? Greater similarity in the presence of a fixed level of noise means lower SNR, where *signal* refers to discriminability among letters. To achieve criterion performance associated with a given level of SNR, a decrease in signal strength (discriminability) due to increasing blur can be compensated for by an increase in the contrast of letters.

We also asked whether a letter recognition model incorporating a measure of similarity would exhibit a higher contrast threshold for recognizing blurry letters. Loomis (1990) described a model of letter recognition and used it to account for human recognition of tactile and visual letters from different fonts. His model generated a theoretical confusion matrix constructed from an internal representation of letter similarity, which he compared with the empirical confusion matrices generated by his subjects.

Briefly stated, in Loomis's model, stimulus encoding involves transformation of the stimulus into an internal representation via linear filtering (convolution with a point-spread function) and nonlinear filtering (a compressive nonlinear transducer at the neuronal stage). Response selection was based on a measure of similarity (Getty et al., 1979; Shepard, 1958, 1987) that is reciprocally related to the Euclidean distance between transformed stimulus letter and transformed template letter. Our implementation[5] of the Loomis model showed that decreasing cutoff frequency resulted in increasing similarity for a fixed contrast level, whereas increasing stimulus contrast at a fixed cutoff frequency resulted in decreased similarity. This means
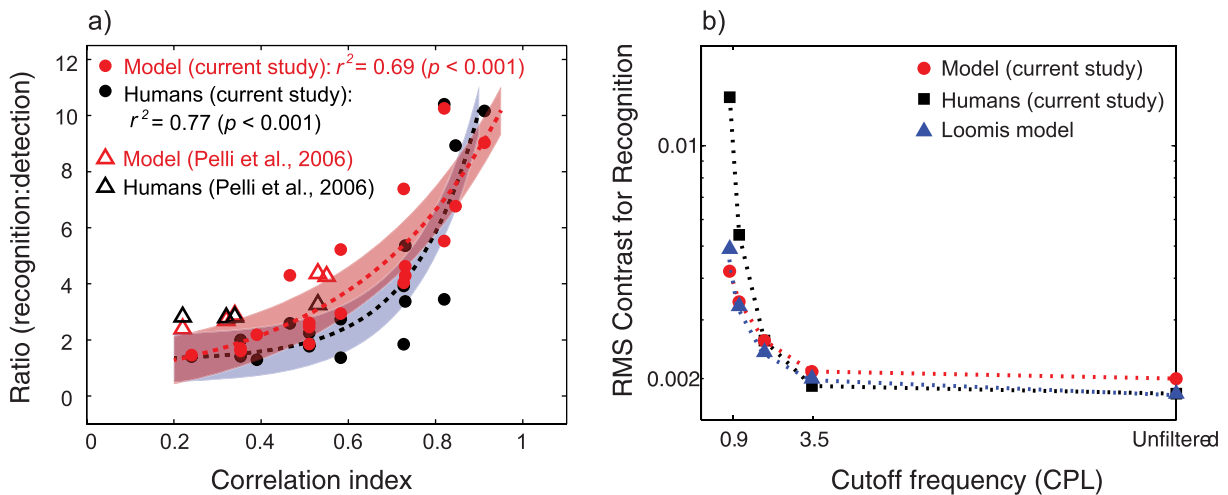
Figure 10. (a) Relationship between the correlation index and ratio of recognition to detection thresholds for the CSF-noise-ideal-observer model (red circles) and human observers (black circles) from the current study. Each circle represents a data point from each spatial-frequency cutoff used for different stimulus conditions (i.e., lowercase and uppercase letters; foveal and peripheral viewing conditions). The fitted lines indicate the regressions of the ratio of recognition to detection thresholds on the correlation index. The percentage of variance accounted for by the correlation index was $r^2 = 0.69$, $p < 0.001$, for the model (red circles), and $r^2 = 0.77$, $p < 0.001$, for our human subjects (black circles). Pelli et al. (2006) have performed a similar correlation analysis on several sets of unfiltered letters (such as bold Bookman, Bookman, Kunstler, and also five-letter words) and measured the detection and recognition threshold contrast for these stimulus sets. Panel (a) also includes the data (red triangles for the ideal-observer model; black triangles for human observers) from their study. Although they did not explicitly report the effect of correlation on the ratio of contrast thresholds for recognition and detection, we computed these ratios from human thresholds and ideal thresholds from their table 2 and have plotted them as a function of the correlation index in this panel. Despite some methodological differences (e.g., their correlation index ["overlap"] was somewhat different from ours), it is evident that their data lie near the 95% confidence interval of the regression lines for our data, confirming that increased correlation is associated with a higher recognition/detection threshold ratio. (b) Comparisons of RMS contrast thresholds for recognition from our human observers and from two models: our CSF-noise-ideal-observer model and Loomis's (1990) model. RMS recognition thresholds are plotted as a function of cutoff frequency for foveal lowercase letters. Black squares: human observers; red circles: CSF-noise-ideal-observer model from the current study; blue triangles: Loomis's (1990) model. (For details on the implementation of the Loomis model, please see Footnote 5.)

that for a constant level of recognition performance, a decrease in cutoff frequency would need to be accompanied by an increase in stimulus contrast. Figure 10b compares recognition data from our human observers, our CSF-noise-ideal-observer model, and the Loomis model. All three show a very similar pattern of performance, although the human data rise more steeply at the lowest frequency than the two models. The qualitative agreement between data and models implies that the higher contrast requirement for recognizing blurry letters is due at least in part to greater perceptual similarity among the letters.

In conclusion, our findings show that the human visual system requires higher contrast for letter recognition when spatial resolution is severely limited. Good agreement between the CSF-noise-ideal-observer model and human observers shows that the greater contrast requirement for recognizing low-pass letters is due to a reduction in the information content of the letters rather than a change in human visual processing. It is likely that increasing blur results in higher perceptual similarity of letters, requiring higher contrast for reliable letter recognition.

Footnotes

[1] In signal processing, the Nyquist rate is the maximum sampling rate that can be transmitted through a channel, which is equal to two times the highest frequency contained in the signal.

[2] Note that as shown in Figure 4, there was a slight mismatch in the number of treatment levels (cutoff frequencies) for fovea and periphery. To keep a balance (in terms of the number of treatment levels) between conditions, the lowest cutoff of each foveal condition (0.9 CPL from the uppercase condition; 1.1 CPL from the lowercase condition) was excluded from the ANOVA analysis. The exclusion is not believed to change the pattern of results to be reported below.

[3] Without causing any computational discrepancy, the CSF filter can be viewed as frequency-dependent noise. White noise is added after the CSF filter so that the signal-to-noise ratio across spatial frequencies followed the shape of the CSF. The performance of the ideal observer is determined by signal-to-noise ratio (SNR). The SNR can be modified by increasing noise spectral density or by decreasing signal level. As far as SNR is concerned, incorporating the human CSF into the ideal observer model is essentially equivalent to modifying the SNR in a frequency-dependent way by introducing frequency-dependent noise (see Chung & Tjan, 2009; Pelli, 1990, pp. 3–24), that is, additive noise whose shape follows the inverse of the CSF. We can formulate the outcome of the ideal and CSF-noise-ideal observers in terms of object signal and noise as follows: The CSF-noise-ideal observer = $f(S + N(freq) + N_0)$, (5) where S is a target signal, N_0 is constant noise, and $N(freq)$ is a frequency-dependent noise that mimics the effect on SNR of the CSF filter.

[4] Stimulus uncertainty grows larger with an increasing number of independent signals and is associated with a rise in detection threshold (Pelli, 1985). The low-pass filtering of 26 letters reduces the independence of the 26 letters by increasing image overlap. In the extreme (e.g., 0.9 CPL), the 26 letters all become a similar looking blob, and uncertainty is reduced.

[5] We implemented Loomis's model using our lowercase letter stimuli and human CSF filter. For the nonlinear transducer, we applied the square-root transformation (i.e., a power function with an exponent of 0.5) to the stimuli. In this model, the perceptual similarity was reciprocally related to the Euclidean distance (D) between stimulus i and template j. Thus, the similarity was expressed as $S(i,j) = \exp[-\gamma*D(i,j)]$. We set the parameter value of γ in the model to equate human and "ideal" RMS contrast threshold for lowercase foveal letters. For each blur level, we obtained the RMS contrast threshold that corresponds to 0.79 proportion correct for the model.

References

Blommaert, F. J. (1988). Early-visual factors in letter confusions. *Spatial Vision, 3,* 199–224.

Bouma, H. (1971). Visual recognition of isolated lowercase letters. *Vision Research, 11,* 459–474.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Chung, S. T., Legge, G. E., & Tjan, B. S. (2002). Spatial frequency characteristics of letter identification in central and peripheral vision. *Vision Research, 42,* 2137–2152.

Chung, S. T., & Tjan, B. S. (2009). Spatial-frequency and contrast properties of reading in central and peripheral vision. *Journal of Vision, 9*(9):16, 1–19, http://www.journalofvision.org/content/9/9/16, doi:10.1167/9.9.16. [PubMed] [Article]

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis.* New York: Wiley.

Dunn-Rankin, P. (1968). The similarity of lowercase letters of the English alphabet. *Journal of Verbal Learning and Verbal Behavior, 7,* 990–995.

Fiset, D., Blais, C., Éthier-Majcher, C., Arguin, M., Bub, D., & Gosselin, F. (2008). Features for uppercase and lowercase letter identification. *Psychological Science, 19,* 1161–1168.

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.

Getty, D. J., Swets, J. A., Swets, J. B., & Green, D. M. (1979). On the prediction of confusion matrices from similarity judgments. *Perception & Psychophysics, 26,* 1–19.

Geyer, J. J. (1970). Models of the perceptual process in reading. In H. Singer & R. Ruddell (Eds.), *Theoretical models and processes in reading* (pp. 47–94). Newark, DE: International Reading Association.

Geyer, L. H., & DeWald, C. G. (1973). Feature lists and confusion matrices. *Perception & Psychophysics, 14,* 471–482.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts.

Ginsburg, A. P. (1980). Specifying relevant spatial information for image evaluation and display design: An explanation of how we see certain objects. *Proceedings of the SID, 21,* 219–227.

Green, D. M., & Swets, J. A. (1966). *Signal detection*

theory and psychophysics. New York: John Wiley and Sons.

Kwon, M., & Legge, G. E. (2011). Spatial-frequency cutoff requirements for pattern recognition in central and peripheral vision. *Vision Research, 51,* 1995–2007.

Laughery, K. R. (1969). Computer simulation of short-term memory: A component decay model. In G. T. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. VI, pp. 135–398). New York: Academic Press.

Legge, G. E., Rubin, G. S., & Luebker, A. (1987). Psychophysics of reading. V. The role of contrast in normal vision. *Vision Research, 27,* 1165–1171.

Loomis, J. M. (1990). A model of character recognition and legibility. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 106–120.

Luce, D. R. (1963). Detection and recognition. In D. R. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. I, pp. 103–188). New York: John Wiley and Sons.

Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research, 42,* 1165–1184.

Mäkelä, P., Näsänen, R., Rovamo, J., & Melmoth, D. (2000). Identification of facial images in peripheral vision. *Vision Research, 41,* 599–610.

Melmoth, D. R., Kukkonen, H. T., Mäkelä, P. K., & Rovamo, J. M. (2000). The effect of contrast and size scaling on face perception in foveal and extrafoveal vision. *Investigative Ophthalmology and Visual Science, 41,* 2811–2819, http://www.iovs.org/content/41/9/2811. [PubMed] [Article]

Melmoth, D. R., & Rovamo, J. M. (2003). Scaling of letter size and contrast qualises perception across eccentricities and set sizes. *Vision Research, 43,* 769–777.

Oruc, I., & Landy, M. S. (2009). Scale dependence and channel switching in letter identification. *Journal of Vision, 9*(9):4, 1–19, http://www.journalofvision.org/content/9/9/4, doi:10.1167/9.9.4. [PubMed] [Article]

Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research, 31,* 1399–1415.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, 2,* 1508–1532.

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

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46,* 4646–4674.

Regan, D. (1988). Low contrast letter charts and sine-wave grating tests in ophthalmological and neurological disorder. *Clinical Vision Science, 2,* 235–250.

Regan, D., & Neima, D. (1983). Low-contrast letter charts as a test of visual function. *Ophthalmology, 90,* 1192–1200.

Rovamo, J. M., & Melmoth, D. M. (2000). Double scaling normalizes foveal and extrafoveal letter recognition. *Investigative Ophthalmology and Visual Science, 41,* S437.

Rubin, G. S., & Legge, G. E. (1989). Psychophysics of reading. VI. The role of contrast in low vision. *Vision Research, 29,* 79–91.

Rumelhart, D. E., & Siple, P. (1974). Process of recognizing tachistoscopically presented words. *Psychological Review, 81,* 99–118.

Shannon, C. E. (1948). *The Mathematical Theory of Commmunication* (pp. 81–96). Urbana, IL: University of Illinois Press.

Shepard, R. N. (1958). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review, 65,* 242–256.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323.

Tanner, W. P., Jr., & Birdsall, T. G. (1958). Definition of d' and η as psychophysical measures. *Journal of the Acoustic Society of America, 30,* 922–928.

Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research, 35,* 3053–3069.

Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics, 9,* 40–50.

Watson, A. B., & Ahumada, A. J., Jr. (2008). Predicting visual acuity from wavefront aberrations. *Journal of Vision, 8*(4):17, 1–19, http://www.journalofvision.org/content/8/4/17, doi:10.1167/8.4.17. [PubMed] [Article]

Wetherill, G. B., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology, 18,* 1–10.

Appendix A

Decision rules of the csf-noise-ideal-observer for letter recognition and detection tasks

The ideal observer has to solve the inverse optics problem, which is to figure out the probability of having a target letter (T_i) out of 26 possible letters (T) for a given input retinal image R, arrays of luminance pixel values.

The detection task refers to which interval contains any one of the 26 letters, whereas the recognition task refers to identification of 1 of 26 letters. We can formulate the detection task as letter recognition with two possible alternatives: letter absent versus letter present (for convenience, hereafter we refer to these two alternatives as two detection intervals).

For the detection task, in order for an observer to determine which interval contains any one of the 26 letter signals, the observer needs to consider all 26 possible letter signals for a given detection interval. Thus, we can denote an interval with the n-number of discrete letter signals by T_{ij} that is, the jth letter signal of the ith interval. Because two letter signals cannot appear at the same time (i.e., mutually exclusive), the inverse optics problem can be expressed as follows:

$$P(T_i|R) = P\left(\bigcup_j^n T_{ij}|R\right) = \sum_{j=1}^n P(T_{ij}|R) \qquad (A1)$$

Because an ideal observer knows all 26 letters and their prior probabilities, the term $P(T_i|R)$ can be solved using Bayes's rule as follows:

$$P(T_{ij}|R) = \frac{P(R|T_{ij})P(T_{ij})}{P(R)} \approx P(R|T_{ij})P(T_{ij}) \qquad (A2)$$

After removing the denominator $P(R)$, a normalizing constant, $P(T_i|R)$, can be reduced to the product of the likelihood function $P(R|T_i)$ and the prior probability $P(T_i)$ of a target signal.

Because the prior probabilities of the 26 letters in our experiment are equal, $P(T_i) = 1/26$, the problem of finding the maximum posterior probability can be expressed as a maximum likelihood function:

$$P(T_{ij}|R) \approx \max_{1 < i < 26} P(R|T_{ij}) \qquad (A3)$$

Therefore, the ideal observer's goal is to find the maximum likelihood function of a given noisy input image R for each of the possible 26 letters T_i and choose the highest possible T_i as its recognition response for a given trial.

Because the luminance noise at different pixels is a random sample from the same Gaussian distribution, the probability of the entire input image R is the product of the probabilities of all the pixels. Thus, $P(R|T_{ij})$ is equivalent to the product of the m-number of Gaussian distributions as follows:

$$\begin{aligned} P(R|T_{ij}) &= \frac{1}{(\sigma\sqrt{2\pi})^m} \exp\left(-\frac{1}{2\sigma^2}\left\|R - T_{ij}\right\|^2\right) \\ &= \exp\left(-\frac{2}{2\sigma^2}\left\|R - T_{ij}\right\|^2\right) \end{aligned} \qquad (A4)$$

The likelihood function $P(R|T_{ij})$ in Equation A4 is reduced to the exponential function after removing terms that do not depend on i or j. Therefore, $P(T_i|R)$ in Equation A1 can be re-expressed as follows:

$$P(T_i|R) = \sum_{j=1}^n \exp\left(-\frac{1}{2\sigma^2}\left\|R - T_{ij}\right\|^2\right) \qquad (A5)$$

When there is only one interval (i.e., letter recognition task), the summation sign for grouping the 26 letter signals of an interval can be dropped from Equation A5. Because the exponential function is monotonic, maximizing Equation A5 is the same as minimizing the Euclidean distance $\|R - T_{ij}\|^2$ between the input image R and a template T_{ij}. In other words, the ultimate job of the ideal observer, which is to find the maximum posterior probability, is equivalent to finding the smallest Euclidean distance between the noiseless template T_{ij} and noisy input image $\boldsymbol{R}$.