Spatial-frequency and contrast properties of reading in central and peripheral vision

Susana T. L. Chung
School of Optometry, University of California, Berkeley, CA, USA
Department of Psychology and Neuroscience Graduate Program, University of Southern California, Los Angeles, CA, USA

Bosco S. Tjan

In this study, we examined the effects of contrast and spatial frequency on reading speed and compared these effects between the normal fovea and periphery. We found that when text contrast was low, reading speed demonstrated spatial-frequency tuning properties, with a peak tuning frequency that partially scaled with print size. The spatial-frequency tuning disappeared when text contrast was 100%. The spatial-frequency tuning and scaling properties for reading were largely similar between the fovea and the periphery, and closely matched those for letter identification. Just as for the task of letter identification, we showed through an ideal-observer analysis that the spatial-frequency properties for reading could be primarily accounted for by the physical properties of the word stimuli combined with human observers’ contrast sensitivity functions.

Keywords: reading, letter identification, spatial frequency, contrast, ideal observer


Introduction

Reading is an everyday task. Although most of us do so effortlessly, reading is a multistage process that involves converting visual symbols into phonological/linguistic representations (Gelb, 1963). As such, reading performance can be influenced by early visual as well as higher level linguistic factors. In addition, because normal reading involves sequences of fixation, saccadic and regression eye movements, reading performance is also limited by eye movement control. Possibly due to the complexity of the reading process as a whole, to date, we still do not have a complete understanding of the underlying mechanism of reading.

Many models have been developed to account for reading. The majority of these models focus on the eye movement control in reading while, at the same time, take into account factors such as lexical access, attention, or other cognitive influences. For example, Rayner and his colleagues developed the E-Z Reader model that provides a viable framework for understanding how linguistic and oculomotor variables affect eye movements in reading (Rayner, Reichle, & Pollatsek, 1998; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 1999). O’Regan postulated that words are identified most efficiently when the eye lands on the optimal viewing position within a word after a saccade (e.g., O’Regan, Lévy-Schoen, Pynte, & Brugaillère, 1984; Vitu, O’Regan, & Mittau, 1990). The optimal viewing position, subject to the influence of lexical structure, is close to the position within a word that minimizes word ambiguity arising from incomplete recognition of the letters in the word (e.g., Clark & O’Regan, 1999; Stevens & Grainger, 2003). These frameworks have undoubtedly enhanced our understanding of how eye-movement control and other high-level factors affect reading performance, but they do not take into consideration how early sensory factors or the quality of the visual input might limit reading speed. We learned from clinical experience that people with impaired vision due to eye diseases often complain of reading difficulties, implying that we cannot ignore the effects of early sensory factors such as poor acuity, the presence of a central scotoma, and a reduction of contrast sensitivity on reading.

To our knowledge, to date, the only model of reading that takes into account early sensory factors is Mr. Chips, a computer algorithm developed by Legge (Legge, Hooven, Klitz, Mansfield, & Tjan, 2002; Legge, Klitz, & Tjan, 1997). Mr. Chips is an ideal-observer model that makes optimal use of three sources of information: visual, lexical, and oculomotor information. With respect to Mr. Chips, the visual information refers to the bits of information transmitted through the visual span—the number of high-resolution letter slots on the retina that afford perfect recognition accuracy of letters. The visual span serves as the front-end limitation on reading in this model. By changing the number of these high-resolution letter slots, or by incorporating non-seeing letter slots simulating the effect of scotomas, the amount of visual
information available to the model will vary. As such, Mr. Chips provides a theoretical foundation for understanding the impact of visual field loss on reading.

These and similar frameworks model reading performance when the reading text condition is optimal, e.g., when the text is of high contrast and the print size is not a limiting factor. Under these conditions, reading speed reaches a maximum value and does not depend on text contrast or print size. When the reading text condition is suboptimal, e.g., when text contrast is low or the print size is close to the acuity limit, reading speed shows a greater dependence on contrast and print size (Chung, Mansfield, & Legge, 1998; Legge, Rubin, & Luebker, 1987). Such a dependency is relevant for people with impaired vision, because most of them have an acuity and/or contrast-sensitivity deficit that require them to read close to their acuity or contrast threshold. Because slow reading is the primary complaint of patients with impaired vision seeking low vision rehabilitation (e.g., Bullimore & Bailey, 1995; Elliott et al., 1987; Kleen & Levoy, 1981), it is important for us to understand how early sensory factors, such as acuity and contrast sensitivity, limit reading for conditions that have not yet reached the maximum reading speed. The first goal of this study was to examine the effects of two stimulus properties—spatial frequency and contrast—on reading performance, for observers with normal vision.

Among patients with impaired vision, those with central field loss who thus have to rely on their peripheral vision to read seem to have more reading difficulty than those with intact central field (Legge, Rubin, Pelli, & Schleske, 1985). Although magnification offers some help, previous studies have shown that the maximum reading speed is always higher in the fovea than in the periphery, even when print size is not the limiting factor (e.g., Chung et al., 1998; Latham & Whitaker, 1996). To help us understand the differences between the fovea and periphery in relation to the task of reading, it is important for us to establish how other stimulus factors affect foveal and peripheral reading speed. The second goal of this study was to examine whether or not the effects of spatial frequency and contrast on reading differ between central and peripheral vision.

To provide a better understanding of how contrast sensitivity across spatial frequencies may limit reading, we borrowed a model that we previously developed to account for human performance of letter identification—the CSF-limited ideal-observer model (Chung, Legge, & Tjan, 2002). An (unqualified) ideal observer is a Bayesian classifier that is optimal by some criterion (typically maximizing the expected accuracy) for performing a given task with a particular stimulus set. An ideal-observer model is an ideal observer formulated with respect to a set of explicitly stated limiting factors pertaining to a real observer. A key advantage of using an ideal-observer model over other modeling techniques is that one can be certain that the behaviors of the model are entirely due to an interaction between the limitations explicitly assumed for a real observer and the stimulus-level information existed in the stimulus set for performing the task. Chung, Legge et al. (2002) found that by limiting an ideal observer for letter identification with an equivalent input noise (Pelli & Farell, 1999) that had an amplitude spectrum the same shape as the inverse of a human contrast sensitivity function (CSF) measured at a given eccentricity, they were able to account for the peak frequency and bandwidth of the observed spatial-frequency tuning functions for letter identification at the given eccentricity. The model was also able to account for the dependency between peak tuning frequency and letter size and has been subsequently applied to account for spatial-frequency tuning properties of letter identification for observers with amblyopia (Chung, Levi, Legge, & Tjan, 2002) and for letter crowding (Chung & Tjan, 2007). Because the CSF is often taken to represent the basic measurement of the spatial resolution of the visual system, the CSF-limited ideal-observer model essentially states that the spatial-frequency tuning properties for letter identification are defined by the human visual system’s sensitivity to different spatial frequencies and the amount of stimulus-level information (letter-identity information) carried by the stimulus at different spatial frequencies.

In the current study, we replaced the letter stimuli used in the previous studies with the list of words used in our reading task. We asked if the same front-end limitation—an equivalent input noise in the shape of an inverted CSF—could account for the spatial-frequency tuning properties for reading. We note that the task of identifying letters differs from that of reading. We argue that there are at least two reasons for the model to be applicable to the task of reading. First, letters are the fundamental building blocks of text, and there is evidence indicating that the performance for identifying a word could be predicted based on the performance for identifying the constituent letters (Pelli, Farell, & Moore, 2003). In a recent study, Pelli and Tillman (2007) reported that letter information alone accounts for 75% of the variance of reading speed. Therefore, factors that limit letter identification are likely to limit reading as well. If so, then models for letter identification might be used to account for reading performance. Second, even though the performance measurement differs between the two tasks, viz., contrast threshold for letter identification and oral reading speed for reading, contrast and reading speed are related in that reading speed improves with text contrast up to a critical contrast, after which reading speed remains at the maximum level (Legge et al., 1987). Thus, at least over the range of text contrast where reading speed improves with text contrast, reading speed and letter identification should be limited by similar factors. From this perspective, the third goal of this study was to examine how reading relates to letter identification. We addressed this goal by comparing the empirically determined spatial-frequency tuning properties of human observers for the
tasks of reading and letter identification and by testing whether a CSF-limited ideal-observer model could satisfactorily account for the spatial-frequency properties of reading, as is the case for letter identification. We predicted that when reading speed is measured for text contrast close to or below the critical contrast, the spatial-frequency properties of reading should resemble those for letter identification. However, when reading speed is measured for text contrast above the critical contrast, the task of letter identification will not be as closely coupled with reading and the spatial-frequency properties of reading would not resemble those for letter identification.

Methods

To address the goals of this study, we measured oral reading speed for spatially filtered text at two contrast levels—one close to the critical contrast and one at the maximum contrast (100% contrast), for a range of print sizes, at the fovea and 10° inferior visual field. The results for different letter sizes will help us determine if the spatial-frequency tuning properties of reading scale with letter sizes. Such scaling, representing the degree of scale dependence, is an important property for letter identification and one that we rely upon when making comparisons between the tasks of letter identification and reading, and between human performance and predictions based on the CSF-limited ideal-observer model.

Reading speed measurement

Oral reading speed was determined using single sentences. On each trial, a single sentence was chosen randomly from a pool of 2630 sentences. Sentences were identical to those used in Chung (2002) and Chung et al. (1998). Each sentence contained between 8 and 14 words (mean = 11 ± 1.7 words) and included only words that were among the 5000 most frequently used words in written English, according to word-frequency tables derived from the British National Corpus. Words were rendered in Times Roman and were presented left-justified on our display, using the rapid serial visual presentation (RSVP) paradigm, i.e., words were presented one at a time in rapid succession, each for a fixed exposure duration. For each testing condition (print size × contrast × eccentricity, see below for details), we used the Method of Constant Stimuli to present sentences at six word exposure durations. The number of words read correctly was recorded for each sentence. We then fit a cumulative-Gaussian function to each set of data relating the percentage of words read correctly as a function of exposure duration, from which we derived our reading speed based on the word exposure duration that yielded 80% of the words read correctly.

Apparatus

Stimuli were generated using a Visual Stimulus Generator (VSG) 2/5 graphics board (Cambridge Research, Rochester, UK) controlled by a Dell Precision 650 workstation and presented on a Sony 24" color graphics display monitor (Model# GDM-FW900). The resolution of the display was 1280 × 960 pixels, at a frame rate of 80 Hz. The temporal dynamics of the display was verified using a photo-detector and an oscilloscope. The background of the display was set at mid-gray (73 cd/m²) so that we could present filtered text with pixels darker and brighter than mid-gray. Unfiltered text was rendered as brighter than mid-gray. The luminance of the display was linearized and calibrated using the VSG OptiCAL software so that the minimum and maximum luminance values of the display were 2 and 144 cd/m², respectively. A video-based eye tracker (Cambridge Research, Rochester, UK), mounted on a head and chin rest, was used for monitoring eye positions during testing at 10° inferior visual field.

Filtered letters

To examine the dependence of reading speed on spatial frequency, we measured reading speed for text letters that were digitally filtered using a set of eight raised cosine-log filters (Alexander, Xie, & Derlacki, 1994; Chung, Legge et al., 2002; Chung, Levi, & Legge, 2001; Chung, Levi et al., 2002; Chung & Tjan, 2007; Peli, 1990). The bandwidth (full-width at half-height) of the filters was 1 octave. The peak object spatial frequencies of these filters ranged between 0.88 and 10 c/letter, in half-octave steps. The filter gain (G) is given by

\[
G(f) = \frac{1}{2} \left( 1 + \cos \left( \frac{\pi \log(f) \log(p)}{\log(c) \log(p)} \right) \right),
\]

where \( p \) represents the center frequency of the filter, corresponding to the frequency at the peak gain (1.0) and \( c \) represents the cut-off frequency at which the amplitude of the filter drops to zero. Figure 1 shows the word “here” filtered through the set of filters and the resulting power spectra.

We adopted a nominal contrast definition to define the contrast of filtered text, as in our previous studies (Chung, Legge et al., 2002; Chung, Levi et al., 2002; Chung & Tjan, 2007). The contrast of a filtered word referred to the Weber contrast (see Equation 2) of its unfiltered version, as we did not rescale the contrast of any word after filtering. Hence, a word filtered with a band-pass filter centered at a high frequency and the same word filtered at a low center frequency are considered to have the same nominal contrast, even though the two filtered words may contain different amounts of contrast energy. As an
example, all the filtered versions of the word “here” in Figure 1 have a nominal contrast of 1.0 since the unfiltered version has a Weber contrast of 1.0. This definition of contrast for a filtered letter is mathematically convenient for computing the integration of contrast signals across spatial frequencies (Nandy & Tjan, 2008, Appendix C). This is because a contrast threshold measured in units of nominal contrast combines both the physical contrast of the filtered stimulus required for the task and the amount of contrast energy within each spatial-frequency band.

**Experimental design**

This study consisted of three parts. In the first part, we determined the smallest print size that supports maximum reading speed (Chung et al., 1998; Mansfield, Legge, &
— the critical print size (CPS). Here, we defined CPS with respect to unfiltered text, although CPS could be determined for text filtered with each of the band-pass filters. Critical print sizes were determined for each observer at the fovea and at 10° eccentricity so that we could subsequently test our observers at a range of print sizes expressed as multiples of each observer’s CPS. The use of multiples of CPS to represent print size facilitates comparison across observers. To determine the CPS for each observer, we first measured RSVP reading speed for five print sizes at each eccentricity. Unfiltered text rendered at 100% contrast was used. As expected, reading speed increased with print size up to the CPS and remained at the maximum reading speed for print sizes larger than the CPS (see Figure 2A for a sample set of data). We fit each set of data using a two-line fit (on log–log axes) where the intersection of the two lines represents the CPS. The log–log slope of the first line was constrained to 2.32, based on the empirical finding that the slope of the first line did not vary systematically with eccentricity, and averaged 2.32 across all the curve fits in a previous study (Chung et al., 1998). The slope of the second line was constrained to zero. The value of the CPS was then used to determine the physical print sizes (1.2 to 16 times the CPS) subsequently used in the rest of the experiment.

In the second part of the experiment, we determined the critical contrast required for maximum reading speed, for a range of print sizes from 1.2 to 16× CPS, and at each of the two eccentricities (fovea and 10° eccentricity). To determine the critical contrast, for each print size tested, reading speed was measured for unfiltered text at six contrast levels spanning a range of 1.1 to 1.5 log units. Contrast was defined as Weber’s contrast:

\[
\text{Text contrast} = \frac{\text{Text luminance} - \text{Background luminance}}{\text{Background luminance}}.
\]

In general, reading speed increased with text contrast up to the critical contrast and remained at the maximum reading speed for contrast above the critical value (see Figure 2B for a sample set of data). We also used a two-line fit (on log–log axes) to fit each set of reading speed vs. contrast data where the intersection of the two lines represents the critical contrast. Here, we allowed the slope of the first line to freely vary while constraining the slope of the second line to zero. Critical contrast is defined as the lowest text contrast that could still support maximum reading speed (Legge et al., 1987).

In the third part of the experiment, we measured reading speed for text filtered with our set of band-pass filters, with center frequency ranging between 0.88 and 10 c/letter, for each of the print sizes used. To test our predictions that the spatial-frequency properties of reading resemble those of letter identification at a text contrast close to the critical contrast, but not at high text contrast, we measured reading speed for the filtered text at two...
contrast levels. One of the contrast levels corresponded to $2.5 \times$ the critical contrast of unfiltered text (Experiment 1) and the other at the maximum contrast (100%, Experiment 2), which was at least several times above the critical contrast in all cases. We chose $2.5 \times$ the critical contrast as the “low” contrast level because we measured critical contrast using unfiltered text. Filtered text always contains less contrast energy than its unfiltered version; therefore, the critical contrast for filtered text is likely to be at a higher physical contrast than the unfiltered version. Based on pilot data, we found that $2.5 \times$ the critical contrast for unfiltered text was an adequate approximation of the critical contrasts for the filtered text, for the range of spatial-frequency filters used.

**Peripheral testing and eye movement monitoring**

During testing at $10^\circ$ inferior visual field, a long, green horizontal line was presented on the display as a fixation target. Words (presented one at a time using RSVP) were presented at $10^\circ$ vertically below this line. Observers were asked to fixate along this line and not to make any vertical eye movements, although horizontal eye movements along the line were allowed, as in our previous studies (e.g., Chung, 2002; Chung et al., 1998).

To ensure proper fixation so that text was presented at the intended eccentricity ($10^\circ$ inferior visual field), we monitored observers’ fixation using a video-based eye tracker. Prior to the beginning of each block of trials (18 trials/sentences per block), the observer’s eye positions were calibrated using a 9-point fixation grid presented on the display. This calibration provided information about the magnitude of the observer’s eye movements. The Video Eyetracker Toolbox, supplied by the manufacturer, provided information as to whether or not the calibration was acceptable. This calibration process was repeated if necessary until the calibration was acceptable.

Following the eye movement calibration and before each trial, the observer was asked to fixate the fixation line for 1 s. Vertical eye positions were sampled at 50 Hz during this 1-s period. A trial could only be initiated if the eye positions were maintained within a vertical window of $\pm 0.5^\circ$ from the fixation line; otherwise, the process of sampling vertical eye positions was repeated. Vertical eye positions were sampled continuously for the duration of the trial. At the end of each trial, the proportion of eye-position samples that drifted outside the vertical window of $\pm 0.5^\circ$ from the intended fixation position was calculated. Trials in which the proportion of such eye positions exceeded 5% of the total number of eye-position samples were discarded. Approximately 15–20% of the trials were discarded and repeated.

**Human observers**

Five native English speakers with normal vision aged between 22 and 31 participated in this study. Three of them participated in both Experiments 1 ($2.5 \times$ critical contrast) and 2 (100% contrast). All had (corrected) acuity of 20/16 or better in both eyes and were either emmetropic or wore contact lenses to correct for their refractive errors. All observers had prior experience in other psychophysical studies that involved the use of peripheral vision. Written informed consent was obtained from each observer after the procedures of the experiment were explained and before the commencement of data collection. All observers practiced the task of reading filtered text in central and peripheral vision using RSVP for at least one session (about an hour) before data collection commenced. Data from the practice sessions were not included in this paper.

**CSF-limited ideal-observer analysis**

The task for the ideal-observer model is single-word identification. The CSF-limited ideal-observer model differs from an optimal Bayesian ideal observer in that we (1) incorporated a linear filter between the stimulus and the ideal observer and (2) added white noise to the signal after the filter (see Figure 3). The linear filter has a modulation transfer function with a shape identical to a human CSF$^2$ and it is scaled such that the peak gain is 1.0. These two front-end limitations are equivalent to a single additive Gaussian noise source with an amplitude spectrum of the inverse of the linear filter (Pelli & Farell, 1999). In other words, we model the limited spatial resolution of the human vision as a form of internal noise (Ahumada & Watson, 1985). The model mimics human performance for the word-identification task by combining the contrast signal in the stimulus that is relevant for pattern discrimination through a limited spatial resolution. The ideal observer sitting behind the front-end limitations can be formulated from first principle by maximizing the a posteriori probability (MAP) of the stimulus being a particular word given the image of a word seen through the noisy front-end. The likelihood function of such a formulation is that of a multivariate white Gaussian noise density function centered (with the mean) at an image of a word filtered by the CSF filter, and the prior probability of a word is defined by the word frequency in the reading.
text. The details of this information are given in Appendix A of Chung, Legge et al. (2002).

An alternative formulation of the CSF-limited ideal-observer model is given in Appendix B of Chung, Legge et al. (2002). This alternative has two advantages. First, it simplifies the calculation for determining the spatial-frequency tuning function of the model as a function of stimulus size. Second, it makes explicit that the observed tuning function is a combination of the distribution of pattern-discriminatory contrast signals in the stimulus (identity information) across spatial frequencies and the spatial resolution of a human observer.

This alternative formulation has two steps. In the first step, we used a white-noise Bayesian ideal observer (Tjan, Braje, Legge, & Kersten, 1995), which is essentially the CSF-limited ideal-observer model without the front-end CSF filter, to estimate the contrast thresholds for identifying band-pass-filtered words (bandwidth = 1 octave) for a range of center frequencies. The reciprocal of the ideal-observer contrast threshold as a function of spatial frequency, which we referred to as the word sensitivity function (WSF, Figure 7), represents the distribution of word-identity information in the stimulus across the spatial-frequency spectrum. The stimulus set for the current study consisted of all the words that appeared in the entire pool of 2630 sentences. There were a total of 2225 words used in the sentence set, with frequency of occurrence ranging between 1 and 1183. The word frequencies of this sentence set were incorporated in the white-noise ideal observer as the prior distribution. We measured the contrast threshold for this ideal observer at the accuracy criterion of 80% (same accuracy criterion for defining reading speed in human observers). The shape and horizontal position of this function, when plotted as log contrast sensitivity vs. log spatial frequency, is independent of the arbitrary internal noise level of the white-noise ideal observer (Tjan et al., 1995). This is because any elevation in the internal noise can be compensated for by an elevation in the stimulus contrast of an equal factor to maintain a constant signal-to-noise ratio at the given accuracy criterion. This simply results in an upward shift of the function in log sensitivity units. Furthermore, our simulation showed that the shape of the WSF is insensitive to the choice of the accuracy criterion, at least within the range of 50–90% (corrected for guessing). As a result, we normalized the peak contrast sensitivity of the WSF to 1.0. For letter identification, we have derived that the ordinate of the letter sensitivity function (derived similarly as for the WSF except that the stimuli was a set of 26 lowercase letters), LSF, is linearly related to bits of information transmitted per contrast unit per octave (Chung, Legge et al., 2002, Appendix B). Here, we extend the model to words and suggest that the WSF represents the distribution of the word-identity information across spatial-frequency bands.

The second step is to quantify the spatial resolution of the human observers in terms of human observers’ CSFs for discriminating the orientation (horizontal vs. vertical) of sine-wave gratings, windowed within a fixed-size, large-field Gaussian envelope. CSFs were measured for each observer who participated in Experiment 1, at the fovea and 10° eccentricity. For single letter identification, we have shown that by multiplying the LSF with the human observers’ CSFs, we could obtain the spatial-frequency tuning functions for the CSF-limited ideal-observer model (Chung, Legge et al., 2002, Appendix B). For different letter sizes, the LSF shifts along the horizontal frequency axis, which when multiplied by the same CSF (fixed for a given eccentricity), yields different spatial-frequency tuning functions. These functions closely match those of the human observers for letter identification in terms of the peak tuning frequencies and bandwidths (Chung, Legge et al., 2002). They differ only in sensitivity by a scaling factor, which may be due to the fact that human CSF is not entirely attributable to an additive correlated Gaussian noise. Here, we replaced the LSF by the WSF, such that the tuning function is equal to the product of WSF and CSF. This observer model has no free parameter that may affect its spatial-tuning properties once the word-identity information (WSF) and the human observers’ CSFs are given.

The spatial-frequency tuning functions of the CSF-limited ideal-observer model for “reading” are measured in terms of contrast sensitivity, as is the case for measuring the human spatial-frequency tuning functions for letter identification. To relate these tuning functions of the CSF-limited ideal-observer model to those for reading by human observers measured in terms of reading speed, we observe that for text contrast below the critical contrast, reading speed is monotonically related to contrast (Legge et al., 1987). According to our hypothesis that the spatial-frequency tuning for reading is determined by the same set of limiting factors as those for the spatial-frequency tuning for letter identification, for physical contrast below the critical contrast (or nominal contrast near the critical contrast), the peak tuning frequency should be identical regardless of whether we measure reading speed or contrast sensitivity or whether the task is reading or letter identification. The bandwidth and amplitude of the tuning function, however, will depend on what we measure. In principle, we can use a function of reading speed vs. contrast (e.g., Figure 2B) as a transfer function to convert a tuning function for letter or word recognition with sensitivity measured in contrast units to a tuning function for reading with sensitivity measured in units of reading speed. The relationship between spatial-frequency tuning functions for reading and letter/word identification depends strictly on our hypothesis that spatial-frequency tuning for reading and letter/word identification are limited by the same factors.

The spatial-frequency tuning properties of a human observer can be summarized by an important property—how the spatial frequency corresponding to the peak of the spatial-frequency tuning function, the peak tuning frequency,
changes with letter size. The dependence of peak tuning frequency on letter size, a representation of the degree of scale dependence, is an important feature of letter identification (Chung, Legge et al., 2002; Chung, Levi et al., 2002; Majaj, Pelli, Kurshan, & Palomares, 2002) and can be represented by a plot of peak tuning frequency, expressed as retinal frequency in c/deg, as a function of the reciprocal of letter size, also in units of c/deg. The data can usually be described using a power function (a straight line fit to the data on log–log coordinates), where the exponent is related to the degree of scale dependence, i.e., whether we use the same or different spatial-frequency mechanisms to identify letters of different sizes. This function is referred to as the spatial-frequency scaling function. An exponent of 1 implies perfect size scaling or size invariance. In other words, when letter size in c/deg increases, the peak tuning frequency in c/deg also increases by the same magnitude such that the peak tuning frequency is scaled with letter size. In contrast, an exponent of 0 indicates a complete lack of size scaling, or that peak tuning frequency in c/deg is completely independent of letter size.

Chung, Legge et al. (2002) and Majaj et al. (2002) found that the exponent of the scaling function lies between 0.6 and 0.7 for letter identification, which suggests partial scale dependence. Chung et al. further showed that this partial scaling could be accounted for by the CSF-limited ideal-observer model. According to the model, the position of the observed letter “channel” is co-determined by the right limb of the CSF and the left limb of the LSF. Because of the finite half-bandwidth of the right limb of the CSF, a horizontal shift of the LSF, corresponding to a different letter size, yields a smaller shift in the peak tuning of the observed “channel”, thus leading to a partial scaling of tuning frequency with letter size.

In this study, we computed the peak tuning frequencies of the CSF-limited ideal-observer model, for the same range of print sizes tested in the human observers at the fovea and 10° eccentricity.

Results

Experiment 1: Text contrast = 2.5× critical contrast

Reading speed (words per minute, wpm) measured at the fovea is plotted as a function of the center frequency of the band-pass filter for the four observers in Figure 4, with print size as a parameter. Similar measurements obtained at 10° inferior visual field are shown in Figure 5. For all observers and all letter sizes, the reading speed vs. spatial frequency plot demonstrates spatial-tuning characteristics such that reading speed reaches a maximum at some intermediate text spatial frequency and falls for lower or higher text spatial frequencies. To describe the spatial-tuning characteristics of the data, we fit each data set using a parabolic function, symmetrical on log–log coordinates, as given by the following equation:

\[
\log(r(f)) = \log(a) - \frac{4}{\log(2)\sigma^2} \left( \log(f) - \log(p) \right)^2, 
\]

where \(r(f)\) is reading speed for text filtered at a center frequency \(f\), \(a\) represents the maximum reading speed, \(p\) is the peak tuning frequency at which maximum reading speed occurs, and \(\sigma\) is the bandwidth of the tuning function in octaves. This function is essentially the same one we used previously to describe the spatial-tuning characteristics of letter identification for single and crowded letters (Chung, Legge et al., 2002; Chung, Levi et al., 2002; Chung & Tjan, 2007), except that the dependent variable is reading speed instead of contrast sensitivity.

Although we observed spatial-frequency tuning characteristics for all print sizes, the tuning functions are not identical. Figures 4 and 5 show that the peak tuning frequency progressively shifts toward higher object spatial frequency, in units of c/letter, when print size increases at the fovea (repeated measures ANOVA: \(F_{(df=4,12)} = 60.25\), Greenhouse–Geisser adjusted \(p = 0.0002\)) and 10° eccentricity (repeated measures ANOVA: \(F_{(df=2,6)} = 76.65\), Greenhouse–Geisser adjusted \(p = 0.0009\)). Averaged across the four observers, the peak tuning frequency shifted from 1.7 to 3.4 c/letter when print size increased from 1.2 to 16× CPS at the fovea, and 3.4 to 4.9 c/letter when print size increased from 1.2 to 4× CPS at 10° inferior visual field. This shift in peak tuning frequency with print size is reminiscent of the finding for letter identification, in which the peak tuning frequency also demonstrates a progressive shift toward higher object spatial frequency with increased letter size, for single or crowded letters (Chung, Legge et al., 2002; Chung, Levi et al., 2002; Chung & Tjan, 2007). When the peak tuning frequency is converted to retinal frequency in c/deg and plotted as a function of the reciprocal of letter size, the exponent of the resulting spatial-frequency scaling function is approximately 0.6–0.7, indicating only partial scale dependence. Here, to quantify the scale dependency of reading, we plot the spatial-frequency scaling function for reading in Figure 6. Data are pooled across all observers and are plotted separately for the fovea (Figure 6A) and 10° eccentricity (Figure 6B). A power function fit to each data set yielded an exponent of 0.70 ± 0.02 at the fovea and 0.60 ± 0.05 at 10° eccentricity. These values are very similar to those obtained for identification of single or crowded letters (Chung, Legge et al., 2002; Chung, Levi et al., 2002; Chung & Tjan, 2007; Majaj et al., 2002). However, even when the exponents are comparable between the reading speed and letter identification data,
the two sets of data can still be offset vertically. Therefore, in Figure 6, we also include the data for single letter identification (small gray symbols), replotted from Chung, Legge et al. (2002). Clearly, the data for single letter identification closely match those for reading speed at the fovea. At 10° eccentricity, there seems to be a shift in spatial scale between the reading speed and the letter identification data. The significance of the good match of the data between reading speed and single letter identification at the fovea and the demonstrated shift in spatial scale at 10° eccentricity will be addressed in the Discussion section.

CSF-limited ideal-observer analysis

To compare human observers’ data with the predictions based on the CSF-limited ideal-observer model, we multiplied the WSF (Figure 7) with the human observers’ CSFs (Figure 8) to obtain the predicted spatial-frequency tuning functions. Figure 9 plots the peak tuning frequency of these tuning functions derived for the CSF-limited ideal-observer model as a function of letter size. Each symbol represents the peak tuning frequency of one such tuning function obtained for one letter size. We fit a straight line on log–log axes, representing the spatial-

Text contrast = 2.5x Critical contrast

Figure 4. Reading speed (wpm) is plotted as a function of the center frequency of the band-pass filter used to generate the text. Each panel shows data measured for five print sizes (1.2–16× CPS, represented by different size of symbols) for an individual observer, tested at the fovea. Text contrast was 2.5× the critical contrast. Smooth curves through data points represent the best-fit parabolic functions (see text for details) from which we derived the peak tuning frequency to be plotted in Figure 6. Error bars represent ±1 SEM.
frequency scaling function, to each data set obtained at the fovea and 10° eccentricity (red solid lines). For comparison, the spatial-frequency scaling functions for our human observers for text contrast equivalent to 2.5× the critical contrast are included in the figure (black lines). At the fovea, the spatial-frequency scaling function fit to the CSF-limited ideal-observer model data yields a shallower exponent than that for human observers. The exponent was 0.51 ± 0.01, compared with 0.70 ± 0.02 for human data. At 10° eccentricity, the exponents for the CSF-limited ideal-observer model and human data were more comparable—0.56 ± 0.04 for the model and 0.60 ± 0.05 for the human data. It is worth noting that the CSF ideal-observer model consistently yields spatial-frequency scaling functions with exponents ranging between 0.5 and 0.6, shallower than those for the human observers, for the tasks of letter identification and reading (Chung, Legge et al., 2002; Chung, Levi et al., 2002; Chung & Tjan, 2007). We shall return to the significance and interpretation of these exponent values in the Discussion section.

Experiment 2: Text contrast = 100%

Reading speed (wpm) measured for text contrast of 100% is plotted as a function of the center frequency of the band-pass filter for the four observers in Figures 10 (fovea) and 11 (10° eccentricity). A comparison of these

![Text contrast = 2.5x Critical contrast](image_url)

Figure 5. Reading speed (wpm) is plotted as a function of the center frequency of the band-pass filter used to generate the text for the same four observers as in Figure 4, with the exceptions that data shown here were obtained at 10° eccentricity and that only three print sizes were tested (1.2– 4 × CPS). Details of the figure are as in Figure 4.
A key difference in the spatial-frequency tuning characteristics revealed by data with those shown in Figures 4 and 5 is that the spatial-frequency tuning is much broader when the text contrast was 100%. In fact, for large print sizes, reading speed remains virtually constant across a wide range of spatial frequencies and does not fall off at the highest spatial frequency (c/letter). This lack of spatial-frequency tuning for large print sizes occurs at both the fovea and 10° eccentricity and clearly illustrates that the spatial-frequency properties of reading are not

Figure 7. The word sensitivity function (WSF) derived from the CSF-limited ideal-observer model, representing the distribution of word-identity information across spatial frequency for our list of words, is plotted as black circles. The smooth black curve through the circles represents the best-fit asymmetric parabolic function fit to the data. For comparison, the letter sensitivity functions (LSF) for single (red) and crowded (green) letters are shown. To facilitate comparisons between the shape and the horizontal position of these functions, we normalized the peak sensitivity of each curve to 1.0.

Figure 6. Peak tuning frequency for reading, converted to retinal frequency in c/deg, is plotted as a function of print size, expressed as nominal letter frequency in c/deg, for the (A) fovea and (B) 10° eccentricity. Each symbol represents the peak tuning frequency obtained from one of the spatial-tuning functions shown in Figures 4 and 5, each corresponding to one print size. Data from the four observers are represented by different shapes of symbols. The straight line in each panel represents the best-fit power function (on log-log axes) fitted through the human observers' data, with the exponent of the power function given next to the fitted line. For comparison, human data for identifying single letters, replotted from Chung, Legge et al. (2002), are included as small gray filled symbols. Panel C combines the data for the fovea and 10° eccentricity.
identical for suprathreshold (in terms of contrast relative to the critical contrast for reading) and near-threshold text. A simple explanation of the lack of spatial-frequency dependence of reading speed at high contrast is that reading speed is limited by other factors at this regime. Nevertheless, this observation is predictable based on how reading speed changes with contrast (Legge et al., 1987).

**Discussion**

The first goal of this study was to examine the effects of contrast and spatial frequency on reading performance. By measuring RSVP oral reading speed for text filtered with different bands of spatial frequencies at two contrast levels, we found that reading speed shows a strong dependence on spatial frequency at low text contrast. Specifically, reading speed is the highest for some intermediate spatial frequency that depends on letter size and drops for lower or higher spatial frequencies. However, the dependence of reading speed on spatial frequencies varies with text contrast. At high text contrast, reading speed is virtually independent of spatial frequency for a wide range of frequencies and only drops at very low and/or very high spatial frequencies.

The second goal of this study was to examine if the spatial-frequency tuning characteristics for reading are similar between the fovea and the periphery, after...
accounting for the differences in the CPS. When print size is expressed as multiples of CPS, we found that the peak tuning frequencies are higher at 10° eccentricity than at the fovea, when expressed in object frequency. For instance, when print size increases from 1.2 to 4× CPS, peak tuning frequency shifts from 1.7 to 2.5 c/letter at the fovea, and 3.4 to 4.9 c/letter at 10° eccentricity—almost a factor of two higher than the corresponding values at the fovea! However, this finding does not imply that the spatial-frequency scaling properties of reading are vastly different between the fovea and the periphery. When spatial frequency and letter size are expressed in absolute units (c/deg), the relationship between peak tuning frequency and letter size for the fovea and 10° eccentricity roughly falls on the same line (Figure 6C). In fact, the difference between the fovea and periphery in terms of scale dependence is small—the slope of the spatial-frequency scaling function is $0.70 \pm 0.02$ at the fovea and $0.60 \pm 0.05$ at 10° eccentricity. Therefore, it appears that a single framework with the same front-end limitations as defined in the CSF-limited ideal-observer model may account for the spatial-frequency properties of reading regardless of eccentricity. The partial scale dependence (the peak tuning frequency scales with letter size with an exponent between 0.6 and 0.7) means that the human visual system relies more on high spatial frequencies when reading small than large print.

The spatial-frequency tuning properties of reading, at least at low text contrast, is reminiscent of the spatial-frequency tuning properties of letter identification. The third goal of this study was to relate the spatial-frequency properties of reading to those of letter identification. We showed that they are closely related based on two observations. First, we compared the spatial-frequency scaling functions for reading obtained from the current study with those for letter identification obtained in Chung, Legge et al. (2002). Figure 6A shows that the two are indistinguishable at the fovea. A similar finding has been shown by Majaj, Liang, Martelli, Berger, and Pelli (2003) who used noise masking to derive the “channel” for reading at the fovea. At 10° eccentricity in the periphery, the exponent of the scaling functions for letter identification and reading are almost identical; however, the scaling function for letter identification shows a small offset to a lower spatial frequency compared with that for reading by 0.31 octave (Figure 6B).

The second observation comes from the comparison of the human data with the CSF-limited ideal-observer model. Different shapes of symbols represent the use of CSF from different human observers (same as shown in Figure 6). The dashed (fovea) and solid (10° eccentricity) black lines are the spatial-frequency scaling functions for the human observers, replotted from Figure 6.

Figure 9. Similar to Figure 6, with the exception that each symbol (in red) represents the peak tuning frequency derived from a spatial-frequency tuning function of the CSF-limited ideal-observer model. Different shapes of symbols represent the use of CSF from different human observers (same as shown in Figure 6).
model. For letter identification, the scaling function of the CSF-limited ideal-observer model matched extremely well with the human data at 10° eccentricity in the periphery (Chung, Legge et al., 2002). In contrast, at the fovea, despite the very similar slope of the scaling functions between the CSF-limited ideal-observer model (0.51) and the human data (0.56), there was a scale shift of 0.34 octave between the two sets of data. In the current study, we found a similar pattern of results—the scaling function of human observers well matches that of the CSF-limited ideal-observer model at 10° eccentricity but that match is not as good at the fovea (Figure 9). These observations suggest that the spatial-frequency properties of letter identification might define the spatial-frequency properties of reading when contrast is a limiting factor. In this regime, the CSF-limited ideal-observer model provides a reasonable account for the spatial-frequency properties of reading.

**Early sensory limitations on reading**

As we mentioned in the Introduction section, to date the only model on reading that takes into account early sensory limitation is Mr. Chips (Legge et al., 2002, 1997). Here, we applied the CSF-limited ideal-observer analysis...
to model reading speed. The CSF-limited ideal-observer model and Mr. Chips are complementary when it comes to accounting for the early sensory limitations on reading.

Mr. Chips makes optimal use of three sources of information: visual, lexical, and oculomotor information to perform its mission of making a saccade that “minimizes uncertainty about the current word”. It obtains visual information by sampling text through three different types of letter slots—clear slots that allow letters to be identified, opaque slots in which spaces can be distinguished from letters but letters cannot be identified, and blind spots in which there is no vision (Legge et al., 1997). The related visual span theory (Legge et al., 2007; Legge, Mansfield, & Chung, 2001) and Mr. Chips state that as long as we have the knowledge of the bits of information transmitted by each letter slot, we can predict the reading speed. Although Mr. Chips provides an invaluable framework for our understanding of how visual information could affect reading speed, we still do not know what determines the bits of information transmitted per letter slot. Here, we argue that the bits of information transmitted by any given letter slot is governed by the pattern-discriminatory contrast signals of the text (the letter or word sensitivity functions as measured by an ideal observer) combined with the human observer’s general sensitivity to visual stimuli (the CSF). In fact, the CSF-limited ideal-observer model can be used to translate the image quality of a letter to number of bits of

Figure 11. Similar to Figure 10 with the exception that data shown here were obtained at 10° eccentricity.
information (Chung, Legge et al., 2002, Appendix B). At high contrast, the limit is not the number of bits per letter slot but the number of letter slots within the visual span. As a result, reading speed becomes independent of contrast but changes with the size of the visual span. At low contrast, the image quality becomes more of a limiting factor than the number of letter slots within the visual span. In other words, for the bilinear log reading speed vs. log contrast function, Mr. Chips determines the maximum reading speed and the elbow point, while the CSF-limited ideal-observer model and Mr. Chips co-determine the reading speed to the left of the elbow, where reading speed depends on contrast.

**Reading vs. letter identification**

Clearly, the task of reading involves more than simply letter identification. Yet, there is evidence demonstrating the importance of letter information for reading. Pelli et al. (2003) convincingly showed that human efficiency for word identification is inversely proportional to word length. They also showed that despite the well-known “word superiority effect”, human word identification performance never exceeds the performance predicted by strictly letter-based models. These findings are consistent with the “recognition by parts” hypothesis (Biederman, 1987) where in the case of words, the “parts” refer to the letters. More recently, Pelli and Tillman (2007) compared the contributions of letter, word and sentence information on reading speed and found that letter information alone accounts for 75% of the variance of reading speed. If reading relies so much on the information of the constituent letters, then it should have the same front-end limitations as letter identification. Here, we showed that the spatial-frequency properties for reading are indeed very similar to those for letter identification. Furthermore, using the CSF-limited ideal-observer model, we showed that both tasks are limited by the same front-end limitations, viz., the physical properties of the stimuli and the spatial resolution of the human observers. Note that we used sentences that have contextual cues for our reading task. Given that contextual cues are considered as high-level, non-sensory factors, in this study, we do not make a distinction between word recognition and sentence reading.

Previous studies have reported that object spatial frequencies ranging between 1 and 3 c/letter are the most useful frequencies for letter identification (e.g., Alexander et al., 1994; Chung, Legge et al., 2002; Chung, Levi et al., 2002; Chung & Tjan, 2007; Ginsburg, 1980; Majaj et al., 2002; Parish & Sperling, 1991; Solomon & Pelli, 1994). There is also evidence suggesting that the same range of object spatial frequencies is most useful for reading (Legge, Pelli, Rubin, & Schleske, 1985; Majaj et al., 2003). In this study, we found that the peak tuning frequency, representing the most sensitive band of spatial frequencies for the task of reading, fell within a range of 1.7 to 3.4 c/letter at the fovea and 3.4 to 4.9 c/letter at 10° eccentricity. The range of peak tuning frequencies we found at the fovea is in excellent agreement with previous reports, but the range at 10° eccentricity is much higher than previously reported. This effect, however, can be easily accounted for by the difference in the rate of change of single letter acuity and CPS with eccentricity. To quantify how visual performance changes with eccentricity, we often refer to the parameter $E_2$, which represents the physical properties of the stimuli and the spatial resolution of the human observers. Note that we used sentences that have contextual cues for our reading task, where reading speed depends on contrast.

**Fovea vs. periphery**

Previously, we reported that the spatial-frequency scaling functions for single letter identification are remarkably similar between human performance and the CSF-limited ideal-observer model at 10° eccentricity in the normal periphery, in terms of both the slope and the intercept of the function. At the fovea, however, human observers demonstrated a shift toward a frequency higher than that is predicted based on the CSF-limited ideal-observer model (Chung, Legge et al., 2002). Here we show similar qualitative result for the task of reading, adding another piece of evidence that early sensory limitations on reading and letter identification are similar (see above). Figure 9 shows that although the spatial-frequency scaling functions predicted based on the CSF-limited ideal-observer model well match those of the human observers, the exponents are shallower for the model prediction (0.51 and 0.56) than for human data, especially at the fovea. Previously when we applied the model to the task of letter identification, we found similarly shallower exponents for the model prediction (~0.50) than for human observers. In all instances, the discrepancies in the exponents between the model
prediction and human data are small, but the discrepancies are always more apparent for measurements obtained at the fovea than in the periphery. We do not yet know why the exponents are consistently lower for the CSF-limited ideal-observer model than for human observers, but a shallower exponent implies that a change in letter size is associated with a smaller change in the peak tuning frequencies for the CSF-limited ideal-observer model than for human observers. Given that the CSF-limited ideal-observer model is optimal with respect to the human observers’ CSF and the physical properties of the stimuli, our finding extends our previous assertion that human observers rely on suboptimal channels for the task of letter identification (Chung, Legge et al., 2002; Majaj et al., 2002) to reading, especially at the fovea.

A more intriguing observation lies in the comparison of human spatial-frequency scaling functions for the tasks of letter identification and reading, because it brings up both the similarities and differences between central and peripheral vision. In Figure 12 we compare the spatial-frequency scaling functions for three related tasks: reading (the present study) and identifying single (Chung, Legge et al., 2002) and crowded (Chung & Tjan, 2007) letters. There are two prominent features illustrated in Figure 12. First, as we have already pointed out, the spatial-frequency scaling functions are similar across tasks (reading or letter identification), at both the fovea and 10° eccentricity. Second, while the scaling functions for reading and letter identification are almost identical in the fovea, there is a small (up to 0.31 octave) yet very noticeable vertical shift between the scaling functions in the periphery. This second feature points to a distinction between central and peripheral vision and between the tasks of reading and letter identification.

A possible explanation for the small vertical shift between the scaling functions for reading and letter identification in the periphery is crowding. Given that words are made up of more than one letter, and that the letters are in close proximity to one another, the spatial properties for reading should better match those for identifying crowded letters (Chung & Tjan, 2007) than for identifying single letters (Chung, Legge et al., 2002). This is apparent in Figure 12. Note that we did not have measurement at 10° eccentricity for crowded letters in Chung and Tjan’s (2007) study, therefore, the spatial-frequency scaling function included in the right panel of Figure 12 for 10° eccentricity is a replot of that at 5° eccentricity (a spatial scale shift of 0.19 octave). A comparison of these spatial-frequency scaling functions shows that the function for reading is indeed better matched by the function for identifying crowded letters instead of the one for identifying single letters, but the effect is very small.

**Conclusions**

In this study, we showed that reading speed depends on the spatial frequency and contrast of the text. When text contrast is low, reading speed shows strong spatial-frequency dependence such that it is highest at some intermediate spatial frequency and falls off at higher or lower spatial frequencies (spatial-frequency tuning). When text contrast is high, reading speed becomes less
dependent on the spatial-frequency content of text, possibly because reading speed is then limited by other factors such as articulation speed.

Based on the comparisons of the spatial-frequency scaling functions relating the peak of spatial-frequency tuning functions with print size, we conclude that the spatial-frequency properties for reading are largely similar between central and peripheral vision. These spatial-frequency properties for reading can be accounted for by the physical properties of the word stimuli combined with human observers’ contrast sensitivity functions, just as for the task of letter identification. Indeed, we also showed that the spatial-frequency properties for reading closely match those for letter identification, lending support to a previous claim that word recognition is a letter-based process, as well as extending this claim from word recognition to reading.

Acknowledgments

This study was supported by Research Grants R01-EY012810 (STLC) and R01-EY017707 (BST) from the National Eye Institute, National Institutes of Health. We thank Yiji Lin and Hope Queener for programming support. We are grateful to Jean-Baptiste Bernard, Gordon Legge, Denis Pelli, Julian Wallace, Dion Yu, and three anonymous reviewers for their very helpful comments on the paper.

Commercial relationships: none.
Corresponding author: Susana T. L. Chung.
Email: s.chung@berkeley.edu.
Address: University of California, School of Optometry, 360 Minor Hall # 2020, Berkeley, CA 94720-2020, USA.

Footnotes

1The choice of rendering text brighter, instead of darker, than the mid-gray background was arbitrary. The important parameter was contrast. Since we used Weber’s contrast to define text contrast (see Equation 2), the absolute value of contrast for text brighter or darker than the background is the same, the only difference being the sign.

2There are many different methods for measuring the human contrast sensitivity function (CSF), each yielding a different result. For our purpose, only the relative sensitivity (i.e., the shape of the CSF) is important. Our CSFs were measured using an orientation discrimination task (vertical vs. horizontal) for sine-wave gratings that consisted of at least five cycles (bandwidth <0.3) on the display. The obtained threshold is comparable to the corresponding grating detection threshold.

References


