Available online at www.sciencedirect.com
ScienceDirect
Vision Research 46 (2006) 4646–4674
Vision Research
www.elsevier.com/locate/visres

Feature detection and letter identification

Denis G. Pelli *, Catherine W. Burns, Bart Farell, Deborah C. Moore-Page [1]

Institute for Sensory Research, Syracuse University, Syracuse, New York, NY 13210, USA

Received 21 April 2006

Abstract

Seeking to understand how people recognize objects, we have examined how they identify letters. We expected this 26-way classification of familiar forms to challenge the popular notion of independent feature detection (“probability summation”), but find instead that this theory parsimoniously accounts for our results. We measured the contrast required for identification of a letter briefly presented in visual noise. We tested a wide range of alphabets and scripts (English, Arabic, Armenian, Chinese, Devanagari, Hebrew, and several artificial ones), three- and five-letter words, and various type styles, sizes, contrasts, durations, and eccentricities, with observers ranging widely in age (3 to 68) and experience (none to fluent). Foreign alphabets are learned quickly. In just three thousand trials, new observers attain the same proficiency in letter identification as fluent readers. Surprisingly, despite this training, the observers—like clinical letter-by-letter readers—have the same meager memory span for random strings of these characters as observers seeing them for the first time.

We compare performance across tasks and stimuli that vary in difficulty by pitting the human against the ideal observer, and expressing the results as *efficiency*. We find that efficiency for letter identification is independent of duration, overall contrast, and eccentricity, and only weakly dependent on size, suggesting that letters are identified by a similar computation across this wide range of viewing conditions. Efficiency is also independent of age and years of reading. However, efficiency does vary across alphabets and type styles, with more complex forms yielding lower efficiencies, as one might expect from Gestalt theories of perception. In fact, we find that efficiency is inversely proportional to perimetric complexity (perimeter squared over “ink” area) and nearly independent of everything else. This, and the surprisingly fixed ratio of detection and identification thresholds, indicate that identifying a letter is mediated by detection of about 7 visual features.

Keywords: Letters; Noise; Reading; Alexia; Simultanagnosia; Ideal observer; Letter identification; Feature detection; Efficiency; Complexity; Features; Letter-by-letter reading; Object recognition; Pattern recognition; Categorization; Decision; Probability summation

1. Introduction

“Explain, explain,” grumbled Étienne. “If you people can’t name something you’re incapable of seeing it.” (Cortázar, 1966, *Hopscotch*).

This paper describes a series of experiments in object recognition—specifically, letter identification—designed to characterize the computation that mediates visual identification of familiar objects. How is the spatially distributed stimulus information combined to produce an identification? And what role does the remembered alphabet play in identifying a letter (Höffding, 1891)?

Existing theory provides surprisingly little guidance. Notions of thresholds and channels, developed in the context of grating detection experiments (e.g., Campbell & Robson, 1968; Graham, 1989; Watson & Ahumada, 2005), seem to apply equally well to letter identification (Alexander, Xie, & Derlacki, 1994; Gold, Bennett, & Sekuler, 1999a; Parish & Sperling, 1991; Solomon & Pelli, 1994), but give no hint as to how the crucial choice is made

* Corresponding author. Present address: Psychology and Neural Science, New York University, 6 Washington Place, New York, NY 10003, USA. Fax: +1 214 995 4701.

E-mail addresses: denis.pelli@nyu.edu (D.G. Pelli), katey@hypernet.com (C.W. Burns), bart_farell@isr.syr.edu (B. Farell), deborah.moore-page@hitchcock.org (D.C. Moore-Page).

[1] Present address: Department of Ophthalmology, Chesire Medical Center, Dartmouth-Hitchcock, 590 Court Street, Keene, NH 03431-1798, USA.

0042-6989/$ - see front matter
doi:10.1016/j.visres.2006.04.023

among the 26 letters: a–z. Previous work on letter recognition yields some insights, such as ruling out high-threshold theories of letter identification (Loomis, 1982). From detection theory, we know that the optimal solution to the seemingly trivial task of identifying one of 26 exactly known patterns in white noise is simple template matching (e.g., Van Trees, 1968). However, people and animals effortlessly identify objects even when the number of possibilities is vast, and we have no idea how this is done. Indeed, the range of choice is so vast that the choosing is akin to a creative act. Recording erroneous letter identifications produces a confusion matrix. Understanding the similarity space it describes might reveal the perceptual units and the algorithm generating letter misidentifications (Blommaert, 1988; Bouma, 1971; Briggs & Hocevar, 1975; Cattell, 1896; Cavanagh, personal communication; Gervais, Harvey, & Roberts, 1984; Geyer, 1977; Gibson, Osser, Schiff, & Smith, 1963; Loomis, 1982; Townsend, 1971a, 1971b; Townsend & Ashby, 1982; Townsend, Hu, & Evans, 1984; Watson & Fitzhugh, 1989). Those letter confusion studies had various goals, but in our attempts to understand letter identification we come away with little more than the observation that letter image correlation is predictive of letter confusion: similar letters are confused. We return to this in Section 4. Alignment and interpolation theories for the recognition of distorted two-dimensional and rotated three-dimensional objects reduce the larger problem to two smaller ones: alignment and template matching (Belongie, Malik, & Puzicha, 2002; Olshausen, Anderson, & Van Essen, 1993; Riesenhuber & Poggio, 2000; Simard, LeCun, Denker, & Victorri, 2001; Thompson, 1917; Ullman, 1989). The success of that two-step approach in machine recognition of handwriting—separately aligning and matching each stroke (Edelman, Flash, & Ullman, 1990)—suggests that eliminating the need for alignment—by using printed letters of fixed size and position—might allow a pure study of the template matching aspect of human visual identification. Psychological theories (e.g., Gibson, 1969) talk about "features" as the elements of an intermediate representation used for identification, but it is far from clear which parts of a letter should be called its elementary features. Perhaps, for an experienced reader, entire letters are elements (see Pelli, Farell, & Moore, 2003). Finally, the theory of signal detectability, by defining the ideal observer, provides a most useful tool, making it possible to meaningfully compare human efficiency across tasks and stimuli (Appendix A).

Thus we began this study with little idea of how observers identify letters. Indeed it is partly because we had no idea how to tackle the general problem of object recognition that we chose identification of printed letters, perhaps the simplest case that preserves the choice among many complex forms, hoping that we might come to understand something here that would illuminate the general problem. We were impressed by the compelling evidence that visual detection is mediated by many independent feature detectors ("channels"), each responding to a different component of the image (Graham, 1989), but detection simply requires the observer to answer the question "Is the screen blank?" which seemed remote from our identification task, "What letter is it?" Feature detection has explained identification of gratings (Graham, 1980; Watson, Thompson, Murphy, & Nachmias, 1980; Watson & Robson, 1981; Thomas, 1985), but gratings are all geometrically similar, a laboratory curiosity, whereas letters are familiar objects with distinctive shapes. Despite its successes with gratings, feature detection seemed an inappropriate first step in recognizing the objects of everyday life. So we thought we would begin by disproving it, to set the stage for a better theory.

Independent detection of image components is the central idea in what is variously called channel theory, probability summation, and feature detection. (We define *feature* as an image, or image component, and suppose that there are several possible features, so that the signal to be identified can be described as a sum of independently detectable features.) The independence is expressed mathematically, in the theory, by assuming that the probability of failing to detect a signal made up of many features is equal to the product of the probabilities of failing to detect each feature. This has become the null hypothesis of visual perception, an extremely simple model that accounts for much that we know. It must be disproved before seriously entertaining any more complicated explanation.

We believed that this model was appropriate for explaining grating detection, a trivial task with simple signals. But it struck us as an unreasonable part of an explanation for more substantial tasks that the visual system does impressively well, like identifying letters. Letters have been optimized for reading, and years of reading presumably allow our visual system to optimize itself to see letters extremely well. So we supposed that people might identify letters efficiently. Theory of signal detectability tells us that the ideal observer identifying one of many possible signals will compare the noisy stimulus with templates of each possible signal (Appendix A). This corresponds to evaluating the output of a weighting function ("receptive field") applied to the image. Efficient letter identification demands use of receptive fields that each match a whole letter. If, instead, the visual system uses receptive fields that each pick up a feature that carries only a small fraction of the energy of the letter, making individual yea/nay decisions about each feature, then the observer's threshold for the whole letter will be limited by the energy of each feature, a small fraction of the energy of the whole letter. Demonstrating that human observers are highly efficient at identifying letters would rule out independent feature detection as an explanation for their performance, unless letters themselves are features.

To our chagrin, our letter identification results eventually compelled us to reconsider and reaffirm feature detection, in Section 4. It is a strong result, just not what we expected. In the meantime, lacking theoretical guidance, we present our empirical investigation of how efficiency for letter identification depends on alphabet, experience, and viewing conditions, i.e. contrast, duration, size, and eccentricity. Like Loomis (1982, 1990), we found that the letters in some alphabets are easier for people to identify than those in others. The notion of complexity turns out to be very helpful in understanding this.

1.1. Complexity of form

What limits the observer's efficiency when identifying patterns? It is well known that we can apprehend only a little in a glimpse (e.g., Miller, 1956a; Sperling, 1960; Woodworth, 1938), yet it is far from clear in what units one ought to specify the size of this bottleneck (Baddeley, 1986; Farell & Pelli, 1993; Nakayama, 1990; Näsänen, Kukkonen, & Rovamo, 1993, 1994, 1995; Olshausen et al., 1993; Verghese & Pelli, 1992). There is a long tradition of measuring visual span for a row of letters, which we discuss below ("Memory span"). Here we ask what perceptual limit might apply to a single object. Images that exceed our span of apprehension often seem more complex. In fact, complexity is defined in part by the psychological effort it demands. According to Webster's Third New International Dictionary, *complex* means "having many interrelated parts, patterns, or elements and consequently hard to understand fully." Latin[2] letters, a–z, seem less complex than Chinese characters, so we imagined that they might be more easily identified. To explore the effect of complexity on human performance, we tested fluent readers with several typeface families (i.e. fonts[3]) and styles (serif, sans serif, and decorative script; plain and bold; proportional and uniform letter spacing[4]) of the Latin alphabet, various non-Latin traditional alphabets and scripts (Arabic, Armenian, Hebrew, and Devanagari), a 26-character subset of Chinese, and two artificial alphabets of random checkerboard patterns. We also created "alphabets" made up of the 26 most common three- and five-letter English words. All the typefaces and alphabets are shown in Fig. 1.

We measured *perimetric complexity:* inside-and-outside perimeter, squared, divided by "ink" area,[5] p^2/a (Attneave & Arnoult, 1956). (They called it "dispersion"; its reciprocal, a/p^2, is called "compactness" in the pattern recognition literature, Ullman, 1995.) For stroked characters, complexity is roughly four times the length of the stroke divided by the width of the stroke, i.e., four times the aspect ratio of the untangled stroke (i.e. $p^2/a = 4(r+1)^2/r \approx 4r$, where r is the aspect ratio). We chose this measure because it tends to capture how convoluted a character is, and is easily computed, independent of size, and additive, i.e., the perimetric complexity of two equal-area adjacent objects (considered as one) is equal to the sum of their individual complexities. As defined here, it is only applicable to binary (i.e. two color) images. Perimetric complexity is only weakly correlated with observers' subjective ratings of complexity, for which the best single predictor is number of turns in the outline (Arnoult, 1960; Attneave, 1957). However, Krauskopf, Duryea, and Bitterman (1954) found luminance thresholds for identifying various equal-area shapes to be inversely related to perimeter, which suggests that perimeter, or the size-invariant p^2/a, might help explain letter identifications.

1.2. Experience

Reading is a natural experiment in human cortical plasticity. Reading an hour a day for 40 years, our older observers have read hundreds of millions of words and a billion letters: $5\ \text{letter/word} \times 250\ \text{word/min} \times 60\ \text{min/h} \times 1\ \text{h/day} \times 365\ \text{day/yr} \times 40\ \text{yr} = 1.1 \times 10^9$ letters. Given the dramatic cortical reorganization produced by major changes in sensory stimuli in primates (Gilbert & Wiesel, 1992; Pons, Garraghty, & Mishkin, 1988; Thompson, Berger, & Madden, 1983), we thought that reading a billion letters might modify the visual cortex in ways that would greatly improve letter identification. So we tested letter identification by fluent readers of English, 6–68 years old.

We also measured how new alphabets are learned in the first place. Our observers included young children (ages 3–4), who learned English (i.e. Latin) letters, and adults, who learned foreign and artificial alphabets.

Learning an alphabet lays down memories for letters. We wondered whether the memories of our "trained" observers—who had learned to accurately identify letters from a previously unfamiliar alphabet—were effective for purposes other than identifying letters, the task they were trained on. We measured novice, trained, and fluent observers' memory span for the letters of several alphabets, both traditional and artificial, as a function of experience.

[2] A linguist would say that English and most other modern European languages use the modern *Latin* (or *Roman*) alphabet, a–z, with minor variations, like ç and ñ. However, to a typographer, *roman* denotes a particular class of typefaces (upright, not italic). Both perspectives, linguistic and typographic, are relevant to our research. Modern typefaces appeared in Venice, around 1470 (e.g., Nicholas Jenson and Aldus Manutius). The uppercase letters in these typefaces are based on ancient Roman inscriptions, but the lowercase letters are based on humanistic script of the Italian Renaissance, which in turn was based on the Caroline minuscule (named after Charlemagne) that appeared about 700 AD (Fairbank, 1952). While we expect our results with English letters to generalize to other instances of the Latin alphabet, e.g., French or Spanish, which include accents and special characters, those decorations would have been unfamiliar to most of our English-speaking observers. Thus, to be precise, we will mostly refer to the "English" alphabet used in our experiments.

[3] Before desktop publishing, a "font" was a specific size of a particular typeface, e.g., 12 point Times italic, corresponding to a case of metal type, which was the unit of purchase. Following the new fashion of desktop publishing, we use the word "font" to refer to an entire typeface family, e.g., Times, including all the styles and sizes, which is the new unit of purchase.

[4] Most of our experiments presented one letter at a time, making letter spacing seemingly moot, but typewriter fonts, like Courier, are designed to be uniformly spaced, so their letter shapes (e.g., wide i's) are markedly different than those of the proportionally spaced typesetting fonts, like Bookman, used for most printed reading matter.

[5] "Ink" area is the number of inked pixels times the area of a pixel.

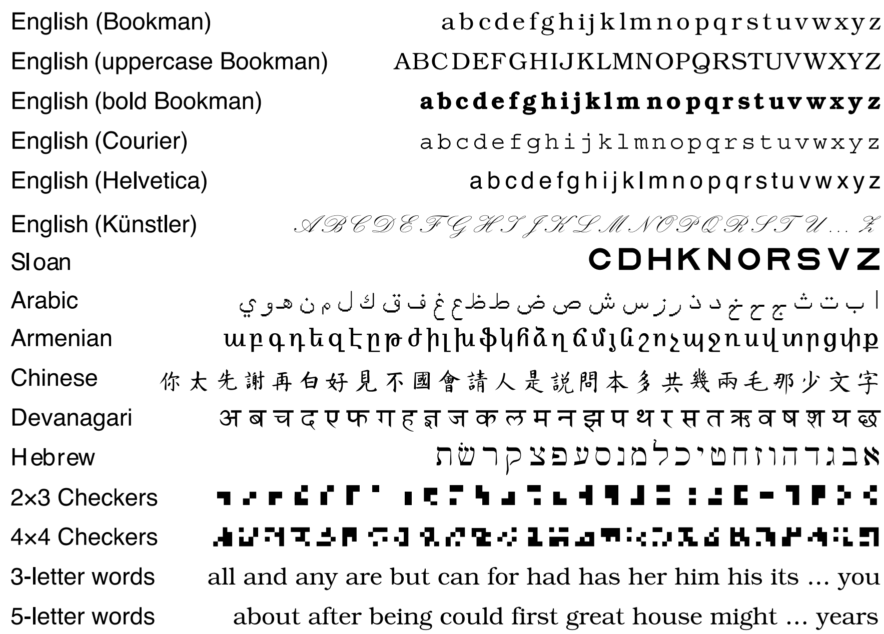

Fig. 1. The fonts and alphabets (and scripts) that we used. English and many other modern European languages use the modern Latin[2] alphabet, a–z, with minor variations, like ñ. Sloan is a special font designed by Louise Sloan for eye charts and contains only the 10 letters shown (NAS-NRC Committee on Vision, 1980). The 26 Chinese characters were scanned from a primer (DeFrancis, 1976). Devanagari is the script used to write Hindi and Sanskrit. The 2×3 and 4×4 Checkers fonts contain 26 random 2×3 and 4×4 checkerboards, as explained in Section 2. The words are the 26 most common three- and five-letter words (Kučera & Francis, 1967).

If the newly learned letters represent the addition of new elementary patterns to our observers' visual vocabulary (Gibson, 1969), then one might expect the observers' memory span for these letters to be greatly increased. If, once learned, each character is perceived as an elementary "chunk," then more characters might squeeze through the information bottleneck of a glimpse (Miller, 1956a).

1.3. Efficiency

Comparisons across tasks and stimuli that vary in difficulty (e.g., identifying letters in different alphabets) are facilitated by expressing the results as *efficiency*, the ratio of thresholds of ideal and human observers. The ideal observer is defined mathematically, and is implemented in software. When asked to identify a noisy stimulus, the ideal observer chooses from the possible signals the one most similar to the stimulus. The ideal's threshold captures the intrinsic difficulty of the task. Efficiency, being a ratio of ideal and human thresholds, strips away the intrinsic difficulty, leaving a pure measure of human ability. This efficiency is directly analogous to the thermodynamic quantity of the same name, which is a ratio of energies. For letters, threshold contrast energy is the product of squared threshold contrast and the average "ink" area (in deg^2), i.e., integrated contrast power (Burgess, Wagner, Jennings, & Barlow, 1981; Parish & Sperling, 1991; Pelli, 1981, 1985). ("Ink" area is the same for both human and ideal, so it drops out of the efficiency ratio.) Efficiency cannot exceed one. Of all possible observers, the ideal observer uses the least contrast energy to do the task. Human observers need more contrast energy to perform as well as the ideal. Appendix A derives the ideal observer for classification of signals in white noise and shows that its performance depends only on the probability and covariance of the signals and the power spectral density of the noise.

2. Methods

On each trial, the observer was briefly shown a faint letter in visual noise (Fig. 2a) and was then asked to select the letter from a display of all the letters in the alphabet (Fig. 2b).

2.1. Observers

Our observers ranged in age from 3 to 68 years. All had normal or corrected-to-normal acuity and were tested binocularly, with their correction, if any. For a given alphabet, each of our observers was either a *fluent* reader of literature (and did primary school[6]) in that alphabet, a "novice," or (after practice) a "trained" letter identifier. *Novice* observers had little or no experience with the alphabet in any context. *Trained* observers had extensive laboratory experience (>2000 trials) with an alphabet but could not read any language printed in that alphabet.

2.2. Identification task

The observer viewed a gamma-corrected grayscale computer monitor (Pelli & Zhang, 1991), normally from a distance of 60 cm. (Exceptions from normal conditions are noted in the text and the figure captions.)

[6] An exception: Observer SS is classified as "fluent" in Hebrew though she did primary school in English, not Hebrew. She learned Hebrew in kindergarten, continued to take Hebrew language courses throughout primary and secondary education, and for two years in college, and now teaches Hebrew.

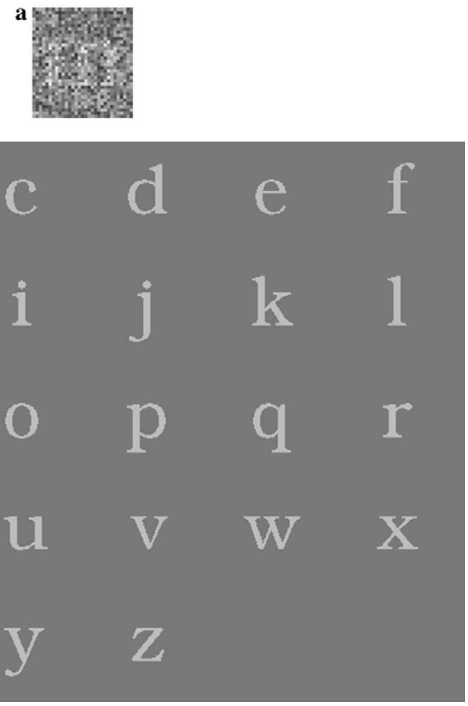

Fig. 2. (a) The stimulus used in our letter identification task: a letter in white Gaussian noise. Letters were rendered off-screen. Independent Gaussian noise was added to each pixel. The image was then scaled by pixel replication and displayed for 200 ms in all experiments, except when we measured the effect of duration (Fig. 6), and except that we tested children under 6 years old with 400 ms stimuli (Figs. 10 and 11). The illustrated font is Bookman (ITC Bookman Light), which was used in all experiments, except when measuring the effects of font and alphabet (Figs. 3, 4, 6, 9, 10, 11, and 12). The x-height for Bookman was 1 deg, except when we measured the effect of size (Figs. 7 & 8). (b) The response screen showing the alphabet from which the observer was asked to choose the letter seen in the stimulus, either with a mouse-controlled cursor (adults) or by pointing (young children, under age 6).

The background luminance was set to the middle of the monitor's range, about 50 cd/m^2. A trial began with the appearance of a fixation point on the gray background. The observer moved a mouse-controlled cursor to the fixation point and clicked the mouse button to initiate the trial. The stimulus consisted of a signal (a letter) and zero-mean white Gaussian noise (Fig. 2a) added to the steady uniform background. The signal was a letter-shaped luminance increment. Its Weber contrast c was the ratio of the increment to the background luminance. The threshold criterion was 64% correct.

This static stimulus was normally displayed for 200 ms to adults and children over age 6, or for 400 ms to young children, under 6. The longer duration for young children improved their chances of seeing the signal, as their attention and fixation wandered. (We see later, Fig. 6, that this increase in duration from 200 to 400 ms does not affect thresholds of adult observers.)

The response screen (Fig. 2b), which appeared 200 ms after the stimulus disappeared, contained all the letters of the alphabet used in that experiment, displayed at the same size, font, and contrast polarity as the letter in the stimulus. (If necessary, the size of letters on the response screen was reduced to fit the whole alphabet onto the screen.) Asked to select the signal letter, adult observers used the mouse to click on a letter of the alphabet. Young children pointed directly at the letter on the screen; the experimenter used the mouse to record their selection. After the observer responded, the correct letter was highlighted. Each correct answer was rewarded with a short beep.

2.3. Detection task

This paper is primarily concerned with identification, but Appendix B and Fig. 9 compare thresholds for detection and identification. For detection, each trial presented a letter (the signal) or a blank with equal probability. The signal letter was randomly any one of 26 (or whatever the alphabet size was), with equal probability, on each trial. The observer merely had to indicate whether a letter (the signal for this trial) or a blank was presented, by selecting from among these two alternatives shown on the response screen. In principle, showing the signal letter on the response screen might eliminate all signal uncertainty. The ideal observer uses this information to consider only the specific signal and blank hypotheses. Our observers claimed that the letter identity was helpful in deciding, but we suspect that their thresholds would have been little changed had we suppressed the letter identity from the response screen. Each detection trial was scored right or wrong and we took threshold to be the contrast for 82% correct. No feedback was given.

2.4. Letters

The letters and characters were drawn onto the screen using the typographic capabilities of Apple Macintosh software augmented by Adobe Type Manager. Three kinds of font description were used: bitmap, PostScript, and TrueType. Bitmap fonts specify the bit pattern at a few text sizes, and must be extrapolated by the operating system to generate other sizes, usually with very jagged results. PostScript and TrueType fonts contain geometric descriptions that allow good rendering of letters at any size. PostScript is rendered by Adobe Type Manager and TrueType is rendered by TrueType software built into the Macintosh operating system. All the fonts were rendered off-screen (in computer memory) at a reduced scale (normally 29 point). Independent Gaussian noise was added to each pixel. Then the image was expanded by pixel replication to its final size—each pixel growing to become a square check—and copied to the screen. The expansion was normally a doubling of size, horizontally and vertically. This normal condition produced letters on-screen with a nominal point size of 58, and 2×2-pixel noise checks. One "typographer's point" corresponded to one pixel and 1/76 inch (slightly smaller than the standard 1/72 inch now used in the printing industry) at the display. The x-height of a typeface is the vertical extent of "a c e m n o r s u v w x z," the characters with no ascenders or descenders. The x-height of the Bookman font was exactly half the nominal size of the font. Thus the standard x-height of the Bookman font was 0.92 deg, i.e. about 1 deg. The average height (top to bottom) and width of the letters in each alphabet appear in Table A in Appendix A.

Most of the fonts were obtained commercially, but we created a few ourselves, using the font-drawing program Fontographer to create PostScript descriptions. (We are happy to share these homemade fonts—Checkers, Sloan, and Yung—with others for research purposes: http://psych.nyu.edu/pelli/software.html.) The bitmap "Arabic" font (Arabic and Farsi) and the PostScript "Devanagari" (Hindi-Sanskrit) and "Hebrew" (Hebraica) fonts were obtained from Linguist's Software http://www.linguistsoftware.com/. The PostScript "Armenian" font (Nork) was obtained from Lines, Fonts, and Circles (220 S. Kenwood, Suite 100, Glendale, CA, 91205). "Courier" is a TrueType font from Apple. "Bookman" (ITC Bookman Light), "bold Bookman" (ITC Bookman Demi), "Helvetica," and "Künstler" (KuenstlerScript TwoBold) are PostScript fonts from Adobe Systems http://adobe.com/. Our "Chinese" font (Yung) was created in Fontographer based on high-resolution scans of Yung Chih-sheng's beautiful calligraphy in a beginning Chinese primer (DeFrancis, 1976). The "Sloan" and "Checkers" fonts were created in Fontographer. The shapes of the 10 Sloan characters are specified by the NAS-NRC Committee on Vision (1980), based on Louise Sloan's design. The 2×3 and 4×4 Checkers fonts contain 26 random 2×3 and 4×4 checkerboards, each check being randomly "ink" or "paper." The selection of the 26 "letters" in each of the Checkers alphabets was done once, choosing checkerboards randomly, but excluding duplicates and simple translations.

2.5. Words

We constructed "alphabets" consisting of the 26 most common three- and five-letter words, from the Kučera and Francis (1967) count of printed English words. The most common three-letter word, "the," is 580 times as frequent as the 26-th most common, "man."

2.6. Noise

The noise was made up of square checks, each a luminance increment or decrement, sampled independently from a zero-mean Gaussian distribution truncated at ± 2 standard deviations. The power spectral density of a random checkerboard (with statistically independent check luminances) equals the product of the contrast power and the area of a noise check. The rms contrast of the noise was normally 0.25, so the power was $0.25^2 = 0.0625$. At the normal viewing distance of 60 cm, a side of a check (2 pixels) subtends 0.0633 deg, so the power spectral density N was normally $10^{-3.60}$ deg^2.

2.7. Threshold measurement

We measured contrast thresholds with the improved QUEST sequential estimation procedure (King-Smith, Grigsby, Vingrys, Benes, & Supowit, 1994; Watson & Pelli, 1983) for 40 trials. The procedure assumes that the observer's probability of correct identification P is a Weibull function of letter contrast c,

$$P(c) = \delta\gamma + (1-\delta)\left(1 - (1-\gamma)\exp - \left(\frac{c}{\alpha}\right)^{\beta}\right), \tag{1}$$

with all parameters provided by the experimenter, except the threshold contrast α, which is to be estimated. The program returns the threshold contrast α for a criterion proportion correct P. This estimate is the mean (not mode) of the posterior probability distribution function for the threshold parameter (King-Smith et al., 1994). The slope parameter β was set to 2.5. (Simulations indicate that mis-estimating β affects only the rate of convergence, i.e. it affects only the variance, not the mean, of the final threshold estimate.) For identification, the guessing rate γ is the reciprocal of the number of letters in the alphabet, since we presented all letters equally often. For detection, it is 0.5, since signal and blank were equally probable. For experienced observers the proportion correct at high contrast is nearly 1, so we set the lapsing (or "finger error") rate δ to 0.01. At the end of the 40-trial run we record the threshold contrast c for a criterion proportion correct $P = 0.64$ (for identification) or 0.82 (for detection). We report the average log contrast threshold estimate from several 40-trial runs. Threshold energy E for the alphabet is the average letter energy (across all the letters) at the threshold contrast for the alphabet.

2.8. Efficiency and the ideal observer

Tanner and Birdsall (1958) introduced the notion of comparing human and ideal thresholds to compute efficiency $\eta = E_{\text{ideal}}/E$. Here, E is the human observer's letter threshold energy (product of squared contrast and "ink" area) measured in the presence of display noise with power spectral density N. E_{ideal} is the threshold for the ideal observer. Pelli and Farell (1999) point out several advantages to instead computing *high-noise efficiency*[7]

$$\eta^+ = \frac{E_{\text{ideal}}}{E - E_0}. \tag{2}$$

E_0 is the letter threshold energy for the human observer, measured with zero display noise. η^+ counts only the extra energy needed to overcome the display noise, discounting the energy needed to see the signal on a blank screen. This has been called "central efficiency" (Barlow, 1978), "sampling efficiency" (Burgess et al., 1981), and "calculation efficiency" (Pelli, 1981, 1990). Pelli and Farell (1999) point out that empirical results invariably show η to be dependent on the noise level N, whereas η^+ is found to be independent of N, and thus a more stable property of the observer and task. The distinction between the two efficiencies, η and η^+, i.e. the correction for the zero-noise threshold E_0, becomes insignificant when the display noise is sufficiently strong to greatly elevate threshold, $E \gg E_0$. Since this was true for most of the efficiencies reported here, we will just say "efficiency," though it was always computed by Eq. (2).

The ideal observer performs the same task as the human—identifying letters in noise—and we measure its threshold in the same way: on each trial the ideal-observer computer program receives a noisy stimulus and returns an identification response, which is scored as right or wrong. The ideal observer has exact knowledge of the alphabet's letter forms, their relative frequencies, the letter contrast, and the statistics of the white Gaussian noise we added to the display.[8]

The mathematical description of the computation performed by the ideal observer is given by the theory of signal detectability for identifying one-of-many known signals in white noise (Van Trees, 1968). The ideal observer must decide from which of the 26 letters of the alphabet the letter-in-noise stimulus was most probably created, i.e. choose the hypothesis $\mathbf{H}_i$ (that the ith letter was presented) with highest posterior probability $P(\mathbf{H}_i|\mathbf{D})$ given data $\mathbf{D}$, the pixels of the noisy stimulus. Bayes's rule,

$$P(\mathbf{H}_i|\mathbf{D}) = \frac{P(\mathbf{H}_i, \mathbf{D})}{P(\mathbf{D})} = \frac{P(\mathbf{D}|\mathbf{H}_i)P(\mathbf{H}_i)}{P(\mathbf{D})}, \tag{3}$$

7 We follow the notation of Pelli and Farell (1999) with the minor enhancement of substituting "+" for * (as in η^+ and E^+) to help the reader remember that E^+ represents a threshold elevation $E - E_0$, and to avoid confusion with the well-established use of * to indicate correction for guessing.

8 In claiming that the ideal performs the same task as the human, we mean that the same prior knowledge and stimuli were provided, and that the same choice was demanded. The relevant prior knowledge is the letter shapes (provided on the response screen of every trial), relative letter frequency (always uniform), the statistics of the noise (always Gaussian, fixed variance), and the letter contrast. Letter contrast varied from trial to trial, as determined by the improved QUEST sequential estimation procedure (King-Smith et al., 1994), but the procedure homes in quickly, so that in the second half of the run there is little trial-to-trial variation in contrast, and it is these near-threshold trials that primarily determine the final threshold estimate. Also see *Footnote 12*.

computes the posterior probability from the likelihood $P(\mathbf{D}|\mathbf{H}_i)$ and the prior probabilities of the hypothesis and data. The ideal observer chooses the hypothesis that maximizes posterior probability, i.e., the right hand side of Eq. (3). However, the prior probability of the data, $P(\mathbf{D})$, in the denominator, is irrelevant to this choice because it is independent of the hypothesis, so we are simply maximizing the numerator $P(\mathbf{D}|\mathbf{H}_i)P(\mathbf{H}_i)$, the product of likelihood and prior probability of the hypothesis. $P(\mathbf{H}_i)$ is the frequency of the letter or word in the experiment. The likelihood $P(\mathbf{D}|\mathbf{H}_i)$ is the probability that $\mathbf{D}$ is observed given that $\mathbf{H}_i$ was presented. The ideal observer's maximum-posterior-probability classification reduces to a maximum likelihood classification when all m signals are equally probable, $P(\mathbf{H}_i) = 1/m$, as in the experiments reported here. This procedure compares the noisy stimulus with each of the 26 possible letter signals, and selects the hypothesis with highest likelihood. That is the signal with least mean-square difference with the stimulus (see Appendix A). The computation used by the ideal observer to identify images (letters) presented with equal probability in white Gaussian noise is derived in Section A.1 of Appendix A. Section A.2 shows that the ideal observer's probability of correct identification depends on the set of letters solely through the covariance between the letters, which is proportional to the amount of overlap when two letters are superimposed (i.e. their correlation). Finally, Section A.3 solves a special case to show explicitly how the ideal's threshold depends on the number and correlation of the letters.

The ideal observer is not intended as a model of the human observer. It merely provides a reference that allows us to place human performance on an absolute scale (Geisler, 1989). Human efficiency below 100% indicates a failure to fully utilize the available information. Finding a high human efficiency would rule out inefficient models. It's usually easy to impair an overly efficient model to match human efficiency, but difficult or impossible to salvage an inefficient model that has not been thus impaired.

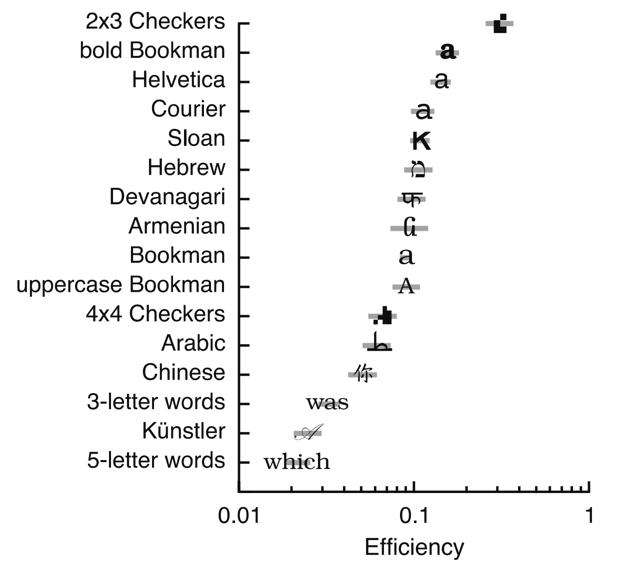

Fig. 3. Letter identification efficiency $\eta^+ = E_{\text{ideal}}/(E - E_0)$ for the fonts and alphabets of Fig. 1, sorted by efficiency. Traditional alphabets were tested with fluent readers; artificial alphabets were tested with trained observers (having completed at least 2000 identification trials). Note the cluster of traditional alphabets near 10% efficiency. Each point is the average of 5–10 thresholds for each of one or more observers, as specified in Table A. The horizontal bar through each point indicates its 68% confidence interval. Readers more accustomed to thinking about contrast should note that efficiency and energy are proportional to squared contrast.

2.9. Memory span

To read, people make about 4 fixations per second, over a wide range of conditions (Erdmann & Dodge, 1898; Legge, Mansfield, & Chung, 2001; Rayner & Pollatsek, 1989). Thus, reading rate is determined by the number of characters advanced between fixations. Provided there is no skipping (skimming), the person is reading that many characters per 1/4 s glimpse. Woodworth (1938) reviews the idea of span of apprehension, the number of characters that can be reported from a single glimpse. Woodworth notes that "span of apprehension" potentially confounds limitations in perceiving each letter with limitations in retaining and reporting a large number of letters. Legge et al. define "visual span" as a perceptual window through which readers recognize characters. Legge et al. ask the observer to report a briefly presented triplet of letters (to incorporate effects of acuity and crowding, but excluding any memory limitation) to measure the extent of this window for characters of a given size at various eccentricities, and use it to predict reading rate.

We measured memory span of novice, trained, and fluent observers using the "missing scan" paradigm (Buschke, 1963; Razel, 1974), which is based on Sperling's (1960) method of partial report. On each trial the observer was briefly shown a random string of one to eight letters.Wishing to emphasize the memory aspects of this task, and to minimize any limitation due to visibility of the letter, we showed the letters at high contrast, without noise, for 200 ms, long enough for an extended glimpse, yet too short for a useful eye movement. Under similar conditions, Sperling (1960) concluded that performance is limited by memory, not visibility. Trials were blocked by number of letters. Letters (0.7 deg x-height) were presented in a single row centered on fixation, with their centers 1.5 deg apart. After a 1-s delay, during which the screen was blank, the observer was shown the same string again, this time with one letter blanked out, and was asked to choose the missing letter from an adjacent display of the entire alphabet. The incomplete string and the alphabet remained visible until the observer responded. To analyze the results, we suppose that when the number of letters in the string is within the observer's "memory span" practically all responses will be correct, but that if the string's length exceeds the observer's memory span, then the proportion of correct responses will equal the fraction of the string remembered. *Memory span* is thus calculated as the proportion of correct responses multiplied by the number of letters in the string. Because lengthening the string can slightly reduce the memory span calculated this way, we used the string length producing the highest calculated memory span in our analysis.

In a control condition, the contribution of verbal rehearsal to memory span was assessed by a distractor method (Peterson & Peterson, 1959). The observer counted backwards, by 1's, aloud, at a brisk rate throughout the trial, from before the stimulus was presented until after the response was made.

2.10. Complexity

Perimetric complexity was defined above as inside-and-outside perimeter squared, divided by "ink" area. We wrote a computer program to measure the perimetric complexity of any bitmap in which 1's represent ink and 0's represent paper. Although the bitmap's pixels are nominally discrete squares, in fact they are not so well resolved by the video monitor and observer. Diagonal lines on the monitor appear smooth and straight, not jagged. Our algorithm[9] tries to take this into account in estimating the length of the perimeter.

[9] The ink area is the number of 1's. To measure the perimeter we first replace the image by its outline. (We OR the image with translations of the original, shifted by one pixel left; left and up; up; up and right; right; right and down; down; and down and left; and then bit clear with the original image. This leaves a one-pixel-thick outline.) It might seem enough to just count the 1's in this outline image, but the resulting "lengths" are non-Euclidean: diagonal lines have "lengths" equal to that of their base plus height. Instead we first thicken the outline. (We OR the outline image with translations of the original outline, shifted by one pixel left; up; right; and down.) This leaves a three-pixel-thick outline. We then count the number of 1's and divide by 3.

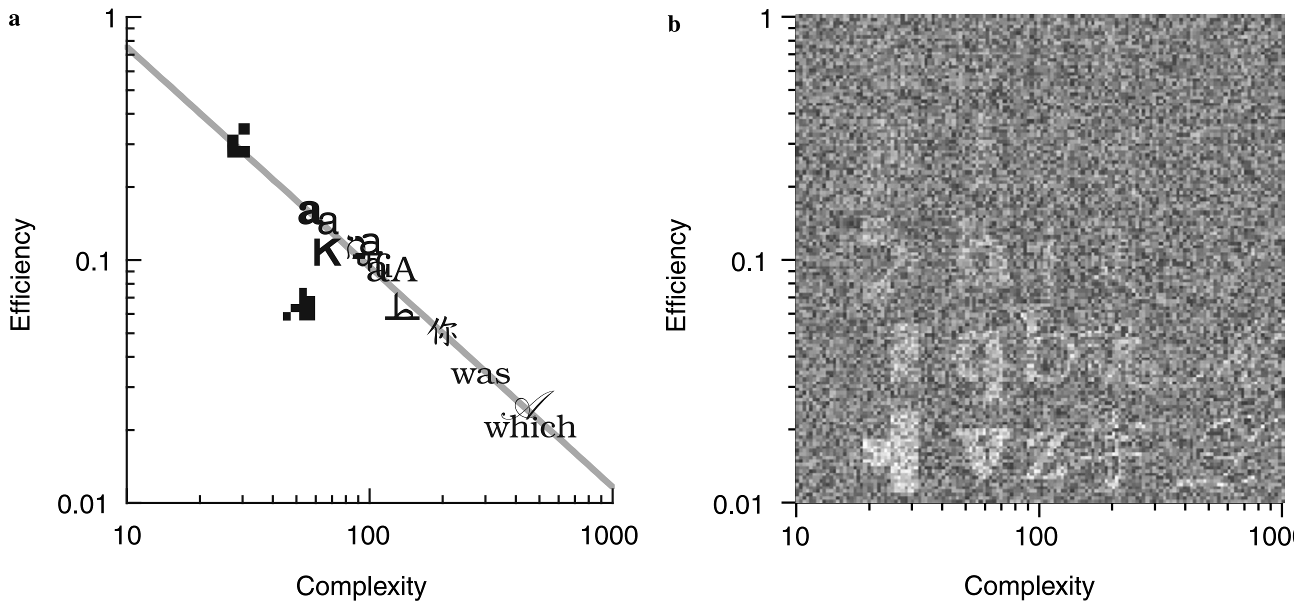

Fig. 4. (a) Identification efficiency, replotted from Fig. 3, as a function of perimetric complexity. (The letters clustered in the middle, from left to right, are: Hebrew, Courier, Devanagari, Bookman, Armenian. They have been dispersed slightly to lessen occlusion.) The complexity of a single letter is calculated by measuring the inside-and-outside perimeter, squaring, and dividing by the "ink" area (*Footnote 9*). The complexity of an alphabet is the average of the complexities of its letters. Log efficiency is highly correlated with log complexity. The best fitting line of slope −1, not shown, is an efficiency of 9 over complexity ($R^2 = 0.95$). The gray regression line has a slope of −0.91 and $R^2 = 0.96$. The fit excludes the 4 × 4 Checkers outlier, which is the only font here not based on a traditional alphabet. The regression line represents the best fit of the feature-detection model, Eq. (B.14), adjusting the proportionality constant ε_f and the psychometric function steepness β for best fit: $\eta = E_{\text{ideal}}/E \propto n^{-0.29}/n^{1-2/\beta} = n^{-0.91}$; thus $\beta = 5.3$. (Based on the evidence in Fig. 9b, discussed in Appendix B.5, we assume the number of detected features, k, is independent of complexity.) (b) This demonstration allows the reader to see what the letters look like at threshold. Letters from several alphabets (2 × 3 Checkers, bold Bookman, Bookman, Chinese, and Künstler) at several contrasts are displayed on a noise background. Each letter is displayed at the horizontal position corresponding to its alphabet's complexity and at the vertical position corresponding to efficiency if the reader can identify the letter 64% of the time. The faintest identifiable letters trace out the reader's efficiency as a function of complexity. Readers unable to identify Checkers and Chinese can still trace out the higher, but similar, line of just *detectable* letters. The unexpectedly fixed ratio of detection and identification thresholds is examined in Fig. 9b.

3. Results

3.1. Complexity

Fig. 3 shows the observers' letter identification efficiencies for 16 different typefaces and alphabets. The traditional alphabets were tested with fluent observers; the artificial alphabets were tested with trained observers. Ignore the 2 × 3 and 4 × 4 Checkers for a moment. Most of the traditional alphabets (Armenian, Devanagari, Hebrew, and most of the English typefaces) yield about 10% efficiency. Arabic and Chinese have lower efficiency, as does the ornate English display script, Künstler. The "alphabets" made up of the 26 most common three- and five-letter English words also yield surprisingly low efficiency, considering how frequently these words occur in everyday reading.[10] The more than 10-fold variation of efficiencies among alphabets in Fig. 3 led us to explore the visual characteristics of alphabets to discover how they determine efficiency.

It occurred to us that, through centuries of use, traditional alphabets might have been optimized to match the human visual system, making it difficult to design a new alphabet that would yield as high an efficiency. We also wondered whether Arabic and Chinese are inefficiently identified because their complexity exceeds the capacity of the visual memories that mediate letter identification (e.g., Attneave, 1955; Fehrer, 1935; Pelli & Farell, 1992; Weinstein, 1955). To test these ideas, we created two artificial alphabets, an (easy) 2 × 3 checkerboard and a (hard) 4 × 4 checkerboard. Each "Checkers" alphabet consists of 26 random checkerboards, from which repetitions and translations were excluded.

After 2000 trials of practice, we found that efficiency for 2 × 3 Checkers was 30%, three times that for the traditional alphabets, while efficiency for 4 × 4 Checkers was only 6%, about half that for traditional alphabets. The high efficiency for 2 × 3 Checkers is consistent with its low subjective complexity. Braille, which also uses a 2 × 3 grid to represent letters, yields higher acuity than English letters, by a similar amount in both visual and tactile tests (Loomis, 1982).

[10] In these alphabets each "letter" is a three- or five-letter word, treated as a single, visually complex pattern. This follows the spirit of past suggestions that very common words are recognized as whole visual patterns, like letters (e.g., Hadley & Healy, 1991).

In looking at the efficiencies in Fig. 3, it is apparent that the low efficiencies were produced by alphabets whose letters appear most complex (Arabic, 4×4 Checkers, Chinese, Künstler). A word is obviously more complex than a letter, and efficiency for identifying the 26 most common three-letter words is 3%, a third that for letters, and efficiency for five-letter words is a fifth that for letters. Pelli et al. (2003) consider the implications for word recognition of the reciprocity between efficiency and word length (also see Geisler & Murray, 2003). The relationship between human efficiency and perimetric complexity for all our fonts and alphabets is shown in Fig. 4a. The regression line in this figure has a log–log slope of -0.91, but a log–log slope of -1 (i.e. reciprocity) fits nearly as well. Contrary to our original expectation, we find that complexity is an excellent predictor of efficiency for identifying these familiar patterns. In fact, with the single exception of 4×4 Checkers, efficiency is about 9 divided by perimetric complexity, for all our alphabets. Since 2×3 Checkers is, in effect, Braille, it is a traditional alphabet. All the traditional alphabets are near the line. Only the artificial 4×4 Checkers alphabet lies off the line. Transformations that affect complexity, such as rendering letters bold or combining them into a word, also had a corresponding inverse effect on efficiency. These changes translate points along the line in Fig. 4a (e.g., from plain to bold Bookman or from single letters to five-letter words), not off it.[11] We will return to these complexity results in Section 4, showing that they are evidence that letter identification is mediated by feature detection.

The demonstration in Fig. 4b allows readers to see what these letters look like at threshold. Letters from several alphabets (2×3 Checkers, bold Bookman, Bookman, Chinese, and Künstler) at several contrasts are displayed on a noise background. Each letter is displayed at the horizontal position corresponding to its alphabet's complexity and at the vertical position corresponding to efficiency if the reader can identify the letter 64% of the time. The faintest identifiable letters trace out the reader's efficiency as a function of complexity. Few readers will know all these alphabets (Braille, Latin, and Chinese), so you may wish to trace out your *detection* threshold, since we find (Fig. 9b) that identification and detection thresholds track each other for these fonts.

Our 10% efficiency for common fonts is consistent with previous estimates of efficiency of identifying letters in static noise: Tjan, Braje, Legge, and Kersten (1995) report up to 16% and Gold et al. (1999a) report up to 8%. Parish and Sperling (1991) used a particularly bold font. From the aspect ratio of its stroke, we estimate the complexity of their font at 60 (about the same as bold Bookman), predicting an efficiency of $9/60 = 15\%$ (Fig. 4). By mistake, they plotted (in their Fig. 7) the square root of efficiency instead of efficiency (Sperling, personal communication). After correction, their peak efficiency of $0.42^2 = 16\%$ matches our prediction of 15%.[12]

3.2. Viewing conditions: Contrast, duration, size, and eccentricity

There are other simple ways of manipulating text that might affect the observer's efficiency even though they leave complexity unchanged. For any given alphabet, the ideal's performance depends only on signal-to-noise ratio, independent of overall contrast. So, in general, ideal performance depends only on the noise level, the covariance (or "template overlap") between the alphabet's letter patterns, and their frequencies of occurrence. Mathematically, these are the factors that determine the intrinsic difficulty of the identification task (see Appendix A). Though the ideal's threshold energy is independent of duration, size, and eccentricity, these parameters can affect human performance and thus efficiency.

Fig. 5 shows that efficiency η^{+} of letter identification is independent of noise level. Efficiency is the same whether the signal and noise are faint or strong, i.e. efficiency is independent of contrast. Pavel, Sperling, Riedl, and Vanderbeek (1987) reported a similar result for understanding of low-bandwidth video transmission of American Sign Language, finding a constant signal-to-noise ratio threshold, independent of contrast. Pelli (1981, 1990) showed the same contrast-independence of efficiency for pattern detection in noise, both for his threshold measurements and for all published visual detection thresholds in noise (see Pelli & Farell, 1999).

Fig. 6 shows that letter and word identification efficiencies are independent of duration for durations beyond 60 ms (letters) or 200 ms (words), out to at least 4000 ms. The slight drop in word efficiency at 60 ms cannot be simply a result of the words' higher complexity since Künstler letters have similar complexity, yet show no such drop.

One might suppose that we found no effect of duration because internal visual processing was allowed to continue indefinitely. By this reasoning, one might predict that a

[11] It would be interesting to extend this analysis to compare efficiency and complexity for individual letters, either by selective analysis of 26-way classifications, or by asking the observer to do 2-way classifications of letter pairs that have similar complexity.

[12] This agreement among studies is satisfying, but take it with a grain of salt, as there were nontrivial differences in the stimulus conditions and the precision of the ideal threshold estimates. First, Gold et al. used dynamic noise, and we used static noise. Efficiency in static noise is independent of signal duration (Fig. 6), but efficiency in dynamic noise is strongly dependent on signal duration (Raghavan, 1995). Second, Parish and Sperling bandpass filtered the letter and noise. We did not. Third, Parish and Sperling randomized letter contrast on each trial, whereas, in our experiments, the contrast uncertainty was negligible (Footnote 8). Contrast uncertainty raises the ideal observer's threshold, but does not affect human threshold for grating detection (Davis, Kramer, & Graham, 1983), so increasing uncertainty increases efficiency (for grating detection and probably for letter identification), since efficiency is the ratio of ideal and human thresholds. Finally, we implemented the ideal observer and measured its threshold precisely. Parish and Sperling established upper and lower bounds on the ideal energy threshold in the presence of contrast uncertainty. The upper bound is about 4 times the lower bound. Their efficiency estimate is based on the geometric average of the upper and lower bounds, so their efficiency estimate is within a factor of 2 of the true efficiency.

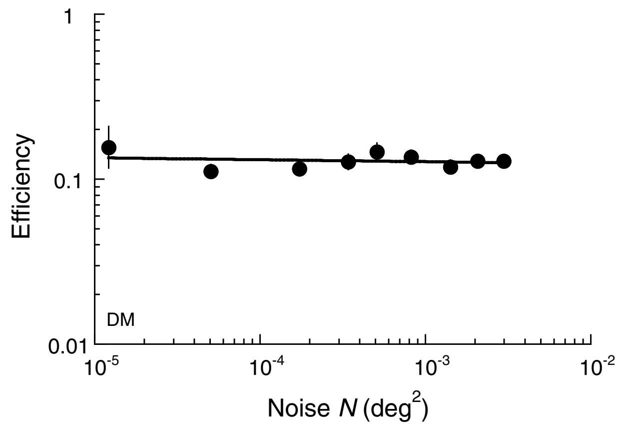

Fig. 5. Efficiency $\eta^+ = E_{\text{ideal}}/(E - E_0)$ as a function of noise power spectral density N. (We varied noise contrast; noise check size was fixed.) Each point is the average of five threshold estimates. The horizontal line represents constant efficiency, i.e. a fixed effective signal-to-noise ratio $D^+ = (E - E_0)/N$ at threshold. Unless otherwise indicated, the data reported in all the Figures are for observer DM identifying Bookman letters, 1 deg x-height, presented for 200 ms, with a background luminance of 50 cd/m^2. Thresholds are measured on backgrounds of no noise and a power spectral density N of $10^{-3.60}$ deg^2.

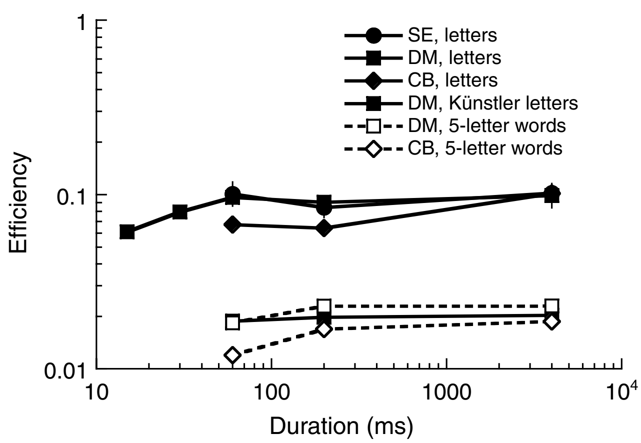

Fig. 6. Efficiency as a function of stimulus duration, for letters (Bookman and Künstler fonts) and words (just Bookman). The words were the 26 most common five-letter words (Kučera & Francis, 1967). Each point is the average of 4 threshold estimates from each of 2 observers. Künstler is about as complex as the five-letter words, so the reduced efficiency for words at 60 ms is not simply due to their perimetric complexity.

post-mask coming immediately after the signal would "erase" the target and terminate visual processing, producing very different results. However, it seems that post-masks are integrated; they do not "erase" the target (Eriksen, 1980; Schultz & Eriksen, 1977; Smithson & Mollon, 2006). We found that using a jumbled array of letters as a post-mask did reduce letter identification efficiency at signal durations of 60 ms and shorter, but had no effect at 200 ms and longer. This result is consistent with Raghavan's (1995) finding, using dynamic noise, that identification of small letters is mediated by a 100 ms integrator.

Fig. 7 shows that size *does* matter, somewhat. Efficiency peaks at 0.5 deg, dropping quickly for smaller letters—presumably due to acuity limitations. For larger letters, efficiency gradually falls 5-fold over a 120-fold increase in

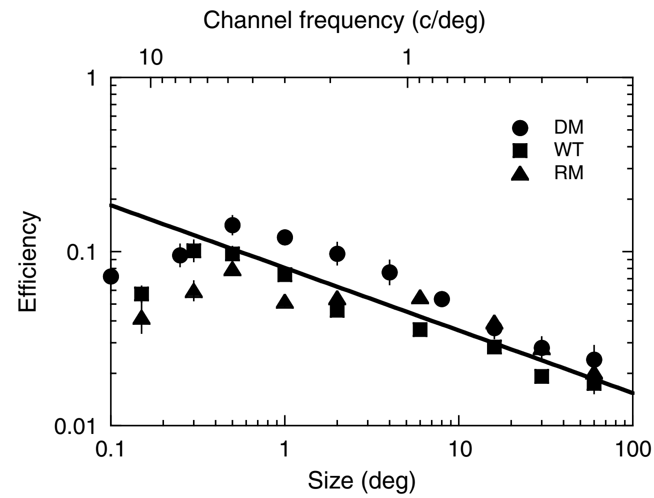

Fig. 7. Efficiency of letter identification as a function of size. The three observers show a similar dependence on size, with an average log–log slope of −0.35 over the range 0.5 to 60 deg. Each point is the average of 4 to 8 threshold estimates. The upper horizontal scale is the estimated channel frequency for identifying a Bookman letter at each size (Majaj et al., 2002).

size, i.e. a log–log slope of −0.35. This is a weak effect for such a powerful stimulus parameter; area is increasing more than 10,000-fold. Parish and Sperling (1991) found an even smaller effect of size for observers identifying band-pass filtered images of letters in noise. Their results, over a 32-fold range of size, suggest a log–log slope in the range 0 to −0.3. Presumably their smaller effect of size is a consequence of their bandpass filtering.[13] Our results demonstrate a consistent −0.35 log–log slope over a 120:1 range. As Parish and Sperling noted, the absence of a strong effect of size on efficiency for letter identification is consistent with the absence of a strong effect of size on reading rate (Legge, Pelli, Rubin, & Schleske, 1985).

The upper scale in Fig. 7 shows the "channel" frequency for each letter size, which might explain the effect of letter size on efficiency. The set of feature detectors used to do a task may be called a *channel*. Majaj et al. (2002) used critical band masking to determine the center frequency of the channel used to identify a letter, as a function of letter size. *Stroke frequency* is the number of strokes traversed by a horizontal cut through the letter, divided by the width of the letter. Testing a wide range of fonts and letter sizes, they found that channel frequency depends solely on stroke frequency of the letter, and that the dependence is a power law $f \propto f_{\text{stroke}}^{2/3}$, weaker than scale-invariance would predict. Majaj et al. note that efficiency for letter identification is highest for small letters, where channel frequency matches stroke frequency, and gradually falls for larger letters, as the channel:stroke frequency ratio rises. Thus the falling efficiency may be due to increasing mismatch between the letter and the channel (feature detectors).

[13] Majaj et al. (2002) found that identification of bandpass-filtered letters is scale invariant and that identification of unfiltered letters is not.

Tjan et al. (1995) report efficiencies for identifying simple three-dimensional blocks (wedge, cone, cylinder, and pyramid) seen from several viewpoints. As with letters, efficiency fell with increasing size, but their halving of efficiency with a tripling of size indicates a steeper log–log slope (-0.7) than the -0.35 that we find for letters.

Some studies have filtered the letter, keeping letter size fixed, finding a modest effect of bandpass center frequency on efficiency. Efficiency peaks at about 6 cycles/letter and falls off gradually at higher and lower frequencies (Gold et al., 1999a; Parish & Sperling, 1991; also see Majaj et al., 2002).

Fig. 8a compares the effect of size for foveal and peripheral (5° right visual field) viewing. Surprisingly, efficiency is unaffected by the peripheral viewing for 1 deg and larger letters, i.e. for letters at least a factor of four above the local acuity limit. Efficiency falls for smaller letters in a similar way for foveal and peripheral viewing, but the fall-off occurs at a somewhat larger size in the periphery, corresponding to the lower acuity there. Parallel findings by Chung, Legge, and Tjan (2002) show contrast sensitivity for identifying filtered letters presented at eccentricities of 0°, 5°, and 10° peaking at similar spatial frequencies, with differences accountable by the shift in CSF with eccentricity. Many authors have suggested that equating the cortical size of peripheral and central targets will equate peripheral and central sensitivity (e.g., Rovamo, Virsu, & Näsänen, 1978). The dashed line in Fig. 8a shows the predicted central efficiency obtained by acuity scaling the peripheral efficiency, i.e. correcting for cortical magnification by shifting the data left by the ratio of acuities. (See Lennie, 1993; Toet & Levi, 1992, for discussion of the relation between acuity, retinal-ganglion-cell density, and cortical magnification.) Acuity scaling does align the results for acuity-size letters, but introduces a factor-of-two discrepancy in efficiency at letter sizes of 1 deg and larger. The cortical magnification hypothesis is that the effects of eccentricity on discrimination are cancelled by scaling that equates the size of the cortical representations. Contrary to this popular idea, our data indicate that peripheral and central viewing are most alike when targets on a noise background have the same retinal, not cortical, size.[14]

The success of acuity scaling for near-acuity-size letters demonstrates a correlation between efficiency and acuity-scaled size, but still does not explain why very small letters are identified less efficiently. We can rule out optics, partly because our efficiency measure Eq. (2) includes a correction for the zero-noise threshold, making it insensitive to optical attenuation (Pelli & Farell, 1999; Pelli & Hoepner, 1989), and partly because the optical quality of the retinal image hardly changes from 0° to 5° eccentricity, while acuity and retinal ganglion cell receptive field spacing change 5-fold (Lennie, 1993). Analogous graphs of efficiency for letter identification as a function of duration in dynamic noise show a region of uniform efficiency over the range of durations for which there are channels with matched integration times, and a sharp drop in efficiency outside that region (Raghavan,

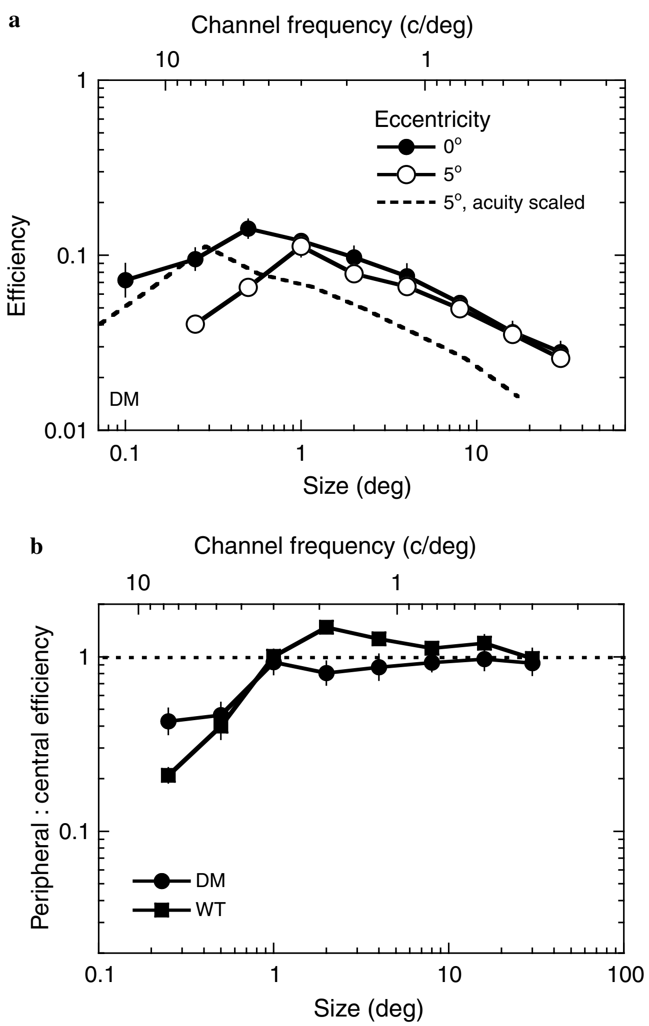

Fig. 8. (a) Efficiency as a function of size, for central and peripheral (5° right) viewing for observer DM. (Her central data also appear in Fig. 7.) The dashed line is a leftward translation of the peripheral efficiencies by a factor of 3.5, the ratio of her central and peripheral acuities. If peripheral and central viewing were most alike when stimuli were equated in cortical area, then the dashed line would superimpose the central efficiencies. The upper horizontal scale is the estimated channel frequency for identifying a Bookman letter at each size (Majaj et al., 2002). Note that their measurements were all central, but we would expect the same result with peripheral viewing (also see Pelli et al., 2004). (b) The ratio of peripheral and central efficiencies, for two observers, showing a ratio near 1 (dotted line) for all letters that are at least 4-fold bigger than the local acuity limit.

[14] A full discussion of this would exceed the scope of this paper. Briefly, threshold contrast depends on both the observer's equivalent noise level and efficiency (Pelli & Farell, 1999). We show in Fig. 8a that efficiency is independent of eccentricity (0° vs. 5°) for letters at least 4 times bigger than the acuity limit. We show elsewhere that a mosaic of neurons with equal responsivity and output noise but eccentricity-dependent spacing will have an equivalent input noise inversely proportional to density of the receptive fields (Raghavan, 1995; Raghavan & Pelli, in preparation). If acuity size is proportional to receptive field spacing, then acuity scaling of the signal increases signal energy in proportion to the equivalent noise power spectral density, conserving the signal-to-noise ratio and explaining why acuity scaling keeps contrast sensitivity approximately constant (Rovamo, Franssila, & Näsänen, 1992; Rovamo et al., 1978). Measuring efficiency, i.e. thresholds in high noise, reveals that the efficiency of the decision process itself is *independent* of eccentricity (Fig. 8).

1995). Likewise, perhaps efficiency falls near the acuity limit because the letters are smaller than the smallest available channel (but see Pelli, Levi, & Chung, 2004).

Fig. 8b emphasizes the eccentricity independence by plotting the ratio of peripheral and central efficiencies. Both observers show little or no effect of peripheral viewing for ⩾1 deg letters. The equal efficiency at all sizes (⩾1 deg) at two eccentricities (0° and 5°) complements the Rovamo et al. (1992) finding that efficiency for detecting a 3 c/deg grating (one size) is independent of eccentricity, from 0° to 8°.

The finding that contrast, duration, size, and eccentricity have so little effect on efficiency suggests that letters are identified similarly, if not identically, across this wide range of viewing conditions.

3.3. Detecting vs. identifying

Näsänen et al. (1993) reported a result closely related to Fig. 4, showing that efficiency for *detection* drops with increasing complexity. The important difference between our experiment and theirs, which included letters and gratings, was that their task was detection, whereas ours is identification. Empirically, the consequence is that, in their results, efficiency drops as the −0.5 power of complexity (weak summation), whereas we find that efficiency drops as the −0.91 power (almost no summation). In principle, the observer's task should affect the way information is integrated over space, since all of a letter's energy can contribute to detection, i.e. distinguishing it from a blank, whereas only the differences between letters are relevant to identification. The Näsänen et al. complexity measure (a space-bandwidth product) differs from ours, but we suspect that the two measures are proportional for unfiltered letters. Indeed, we measured human and ideal detection thresholds (for saying correctly, 82% of the time, "letter" or "blank," when a random letter was present or absent) for 2 × 3 Checkers, bold Bookman, Bookman, five-letter words, and Künstler and confirm their finding, but with our complexity measure, finding a log–log slope of −0.52 relating detection efficiency to complexity. This is shown in Fig. 9a, which plots detection and identification efficiency as a function of complexity for three observers, finding log–log slopes of −0.52 for detection (consistent with −0.5 in Näsänen et al.) and −0.76 for identification (consistent with −0.91 in our Fig. 4).

Since identification, unlike detection, depends on the differences between letters, we were not surprised that identification efficiency was (slightly) more affected by complexity,

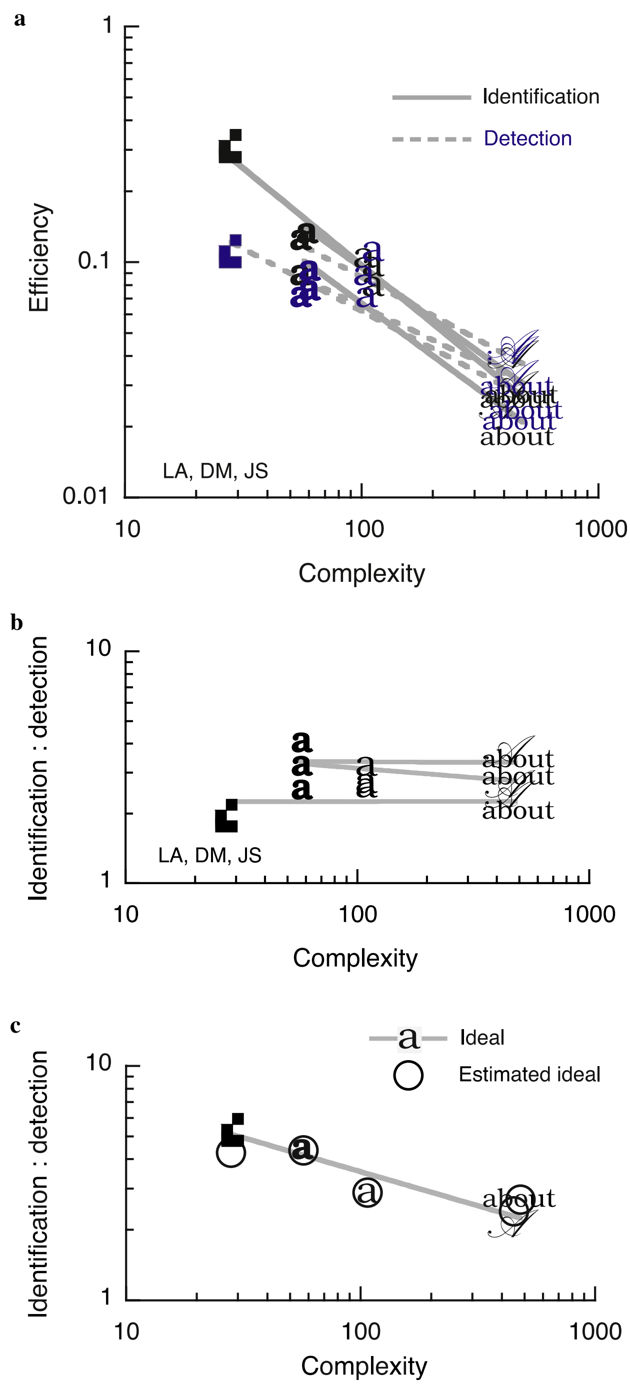

Fig. 9. (a) Efficiencies for letter identification (solid lines) and detection (dashed lines) for five "alphabets"—2 × 3 Checkers, bold Bookman, plain Bookman, Künstler, and five-letter words—plotted as a function of the alphabet's perimetric complexity, for three observers (Table B). The average log–log slope is −0.76 for identification and −0.52 for detection. (The slopes for DM, JS, and LA are −0.82, −0.72, and −0.74 for identification and −0.53, −0.54, and −0.48 for detection.) (b) The identification:detection threshold ratio E_{id}/E_{det} as a function of complexity for three observers (solid lines). The average log–log slope is −0.03, nearly zero. (The slopes for DM, JS, and LA are 0.00, −0.08, 0.00.) Collapsing across complexity and observers, the geometric average of the 13 ratios is $E_{id}/E_{det} \approx 2.8$, i.e. $\log E_{id}/E_{det} = 0.44 \pm 0.03 = \text{mean} \pm SE$. (c) The ideal observer's identification:detection threshold ratio E_{id}/E_{det} as a function of complexity. The circles are a check on the accuracy of the approximation derived in Appendix A.3, substituting the estimated ideal identification threshold, Eq. (A.24), for the true ideal, showing that the error is small. The ideal's threshold for detection is independent of complexity (zero slope), but its threshold for identification is lower for more complex alphabets because they tend to have lower correlation ρ (i.e. overlap), so the ratio drops, with a best-fitting log–log slope of −0.29 (gray line).

especially since we find that more complex alphabets tend to have less overlap (i.e. more difference) between letters. However, Fig. 9a hides the big surprise of these data. Turning to Fig. 9b, which replots the same data, we see that the ratio of human identification and detection thresholds is about 2.8, independent of complexity. (The energy ratio of 2.8 corresponds to a contrast ratio of $\sqrt{2.8} = 1.7$.) The near-zero slopes in Fig. 9b indicate that our observers defy the expectation that the different demands of detection and identification would result in different effects of complexity (and overlap) on threshold. The theoretical implications for feature detection are discussed in Section 4.3.

But, one asks, if detection and identification thresholds have the same dependence on complexity (Fig. 9b), how can it be that their efficiencies do not (Fig. 9a)? Efficiency depends on both human and ideal. We prove in Appendix A.2 that the ideal identification threshold depends only on the covariance between letters, not their complexity. Appendix A.3 shows that, for any given alphabet length and letter energy, we can capture most of the effect of the covariance matrix on the ideal threshold by considering just the average correlation (i.e. overlap) between letters. It turns out that our more complex alphabets have lower correlation ρ (i.e. overlap) between letters (see Table A). In $N = 10^{-3.60}$ deg^2 noise, the ideal detection threshold is about $E = 10^{-3.06}$ deg^2 for all these alphabets (Table B), but the ideal identification threshold is lower for the lower-correlation alphabets (as predicted by Eq. (A.24)), and the correlation (overlap) tends to be lower for more complex alphabets. This accounts for the -0.29 log–log regression line slope of the ideal's identification:detection threshold ratio in Fig. 9c. That is why human efficiencies for identification and detection have different slopes in Fig. 9a.

3.4. Learning

Obviously, efficiency should grow with experience identifying the letters of an alphabet, but we were surprised to find just how little experience is needed before the novice observer's performance reaches that of the fluent observer. Learning curves for five alphabets are shown in Fig. 10. All of the observers were native English speakers. The points represent efficiencies measured by successive 40-trial runs. All observers were adults, except that the English-letter data were provided by the 3-year-old daughter JF of one of the authors. She had only incidental experience with the English alphabet prior to the beginning of the experiment.

All the curves in Fig. 10a initially rise quickly and then continue rising ever more slowly. Efficiency reaches that of fluent readers (~10%; see Fig. 3) within 2000–4000 trials (about 4–8 h). The trials were spread out over several days for the adults, and over several weeks (15 min/day) for the 3-year-old. Shape differences in Fig. 10a correspond to log–log slope differences in Fig. 10b. There is a great deal of scatter, but the curves are remarkably similar to each other, and initial learning is fast, whether the alphabet is artificial or traditional, and whether the observer is an adult or a child. Fast identification learning has also been observed with patterns as diverse as faces and random textures (Gold, Bennett, & Sekuler, 1999b) and progresses similarly in the fovea and 10° eccentric (Chung, Levi, & Tjan, 2005).

Operationally, our letter-identification task requires only that the observer remember the noisy stimulus long enough to compare it with the 26 templates on the response screen, choosing the most similar. In principle, this procedure does not demand that the observer remember the alphabet at all, since it is provided on every trial. However, improvement in human performance seems to be accompanied by categorization of the noisy stimulus in advance of the response screen. Experienced observers immediately move the cursor to their choice as the response screen appears, rather than serially scrutinizing the many alternatives.

It occurred to us that the identification procedure might be learned more slowly than the alphabet and might thus determine the learning rate. However, observers AW and DM in Fig. 10 learned several new alphabets, one after another, and showed no obvious benefit of knowing the procedure. Conversely, we tested a fluent but unpracticed observer (RA) for 2000 trials of Bookman. His efficiency increased only slightly, from 6 to 7.3%, showing that the large rapid improvement in Fig. 10 represents learning of the letters, not the procedure.

Further practice, beyond 2000 trials, produces further improvement at an ever-decreasing rate. Our two most-experienced observers, DM and WT, after 50,000 trials (mostly Bookman), have reached 15% efficiency for the Bookman font—surpassing the 10% typical of untrained fluent observers. This is consistent with an extrapolation of the learning curves in Fig. 10, suggesting that trained and fluent observers improve at the same rate.

For another view of learning, Fig. 11 shows the effect of age, and thus reading experience, on English-letter identification efficiency. Our observers ranged in age from 6 to 68 years. There is some scatter in the data (and perhaps a hint that efficiency increases during the grammar-school years), but the best-fitting line has a slope insignificantly different from zero, rising only 0.03 log units over 50 years (from 8.5% at 10 years to 9% at 60 years). Thus, even decades of additional reading experience do not improve efficiency beyond what is learned in the first year of reading or in the first few thousand trials in the laboratory.

Most of the observers in Fig. 11 have read hundreds of millions of letters. Yet the 10% efficiency of these fluent readers is attained by trained observers in only a few hours' experience with a new alphabet.[15] But what about

[15] The relevant duration may be minutes, not hours. Suchow and Pelli (2005), testing with four letters, found that it took only a few trials for observers to adjust to working with a blank response screen, where each quadrant corresponded to one of the four letters. Testing observers with response screens with and without the letters, they found that, after the first few trials, learning proceeded at the same rate. This indicates that the observers learn letters primarily from the brief near-threshold exposure of the signal, not the static easily visible response screen. Here, a novice reaches the efficiency of a fluent reader in three thousand 200 ms presentations of the signal, a total exposure of 11 minutes.

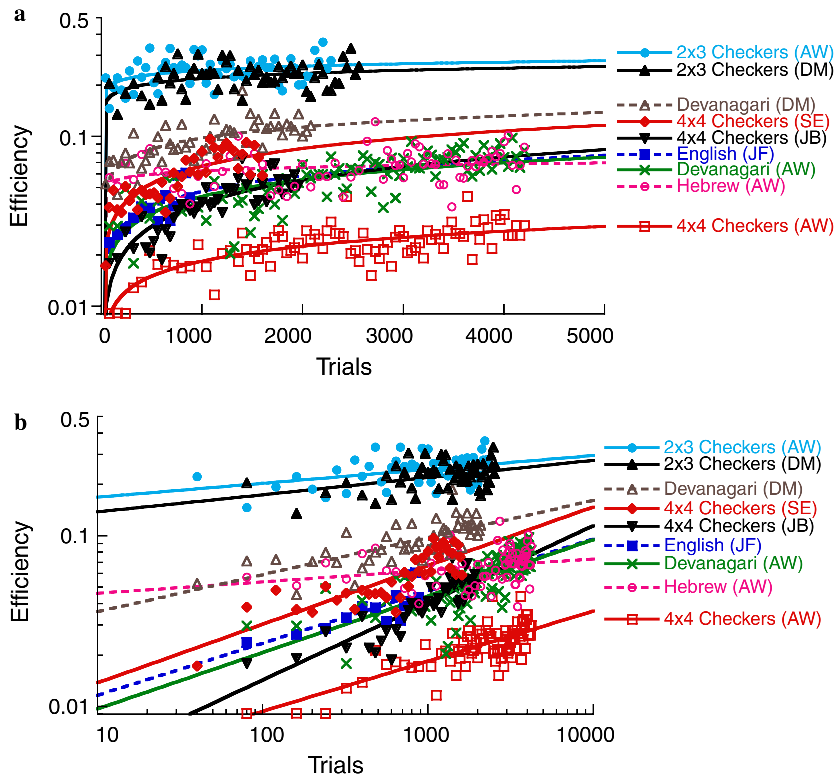

Fig. 10. Learning curves. Log efficiency is plotted as a function of: (a) trials and (b) log trials. Each point represents a single 40-trial threshold estimate. The lines represent linear regressions of log efficiency versus log trials. (The lines fit well, though they have no theoretical basis, and, in fact, make the nonsensical prediction of efficiency greater than 1 if extrapolated to millions of trials.) All observers were native-English-speaking Syracuse University students, with the exception of JF, the 3-year-old daughter of one of the authors. On each trial, the observer was briefly shown a letter from an initially unfamiliar alphabet, and then was asked to identify the presented letter on the response screen, which contained all the letters of the alphabet. It is possible that some of the learning (efficiency improvement) occurred through seeing the letters on the response screen. However, in similar experiments with just four letters, Suchow and Pelli (2005) found that, after the first few trials, learning proceeds at the same rate whether the four sections of the response screen display the letters they represent or remain blank. The negligible contribution of the letters on the response screen indicates that the observers learn the letter shapes (or whatever is improving efficiency) primarily through exposure to the near-threshold signal letter on the test screen, not the easily seen letters on the response screen. For some of these observers (AW, JF, and JB), high-contrast runs measuring percent correct were interleaved with efficiency measurements to provide an accurate estimate of δ to the improved QUEST sequential estimation procedure. For the other observers, δ was simply set to 0.01. The legend indicates the alphabet used in the nearest curve. Letters were presented for 200 ms to adults and 400 ms to young children under age 6 (see Section 2).

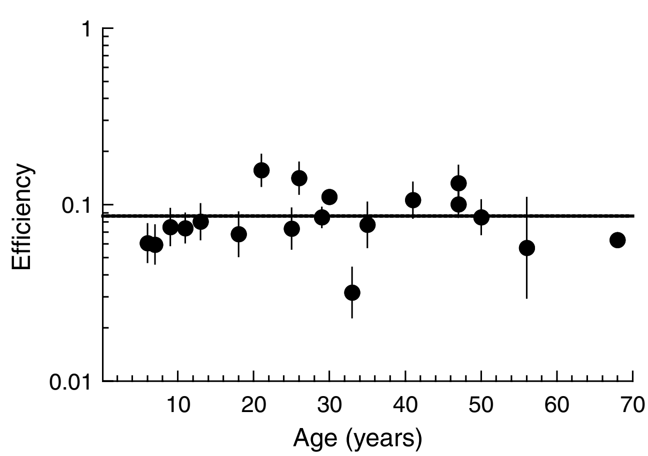

Fig. 11. Efficiency as a function of age. Each point is an average of 5 thresholds of a single observer, except the point for the 6-year-old child who quit after only 4 thresholds. The line is a linear regression of log efficiency versus age. Its slope is not statistically significantly different from zero. The graph includes three children, AR (age 6), HE (7), and AE (9), and two adolescents, KE (11) and RE (13).

cortical plasticity? Should not the years spent reading nearly a billion letters have produced a further improvement in efficiency? Each brain had years to refine its algorithm for letter recognition, but it did not improve upon the calculation developed during our trained observers' first three thousand trials. Those observers have acquired, perhaps permanently,[16] the ability to recognize the previously unfamiliar letters, but this plastic change took place within days, not years. (We will return to reading in Section 4.)

3.5. Memory span

Those years of reading did affect another visual capacity. Fig. 12 shows that fluent readers have a memory span of

[16] We tested one observer (AW) a year after she learned to identify Devanagari characters; her efficiency was undiminished, despite having had no further exposure to Devanagari.

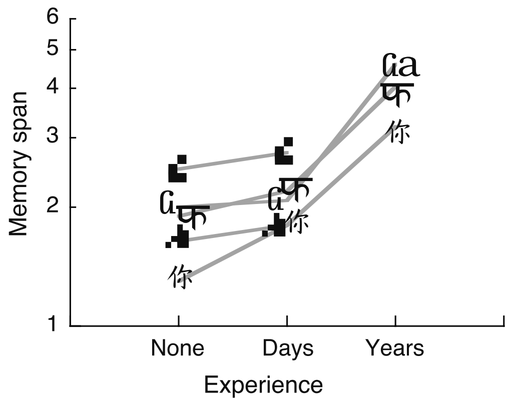

Fig. 12. Memory span as a function of experience. An experience of "none" refers to novice observers; "days" refers to trained observers; "years" refers to fluent observers. The letters, with 0.7 deg x-height, appeared in horizontal strings at high contrast, without noise, for 200 ms. One to four observers contributed to each data point, which gives their average memory span. The standard error of each point is less than 0.5.

four or five letters (5 for English and Armenian, and 4 for Devanagari) or three Chinese characters, whereas both novice and trained observers have a meager memory span of only two letters (or one Chinese character). Although our trained observers efficiently identify letters in their trained alphabets, they can only remember two at a time, not four or five like fluent readers. Our letter-identification training had no effect at all on memory span, even though it raised our novice observers' identification efficiency to that of fluent observers.

An alternative explanation that does not involve visual capacity might be that fluent observers have phonological labels for the letters, which they verbally rehearse while waiting for the response screen. To assess the contribution of verbal labeling and rehearsal, we measured memory span for one of our observers both with and without a verbal distractor secondary task. The observer was trained in 2×3 Checkers, and was fluent in English. The verbal task consisted of rapidly counting backwards aloud during the trial. If the trained-fluent difference in memory span were entirely due to rehearsal of verbal labels, we would expect the secondary verbal task to greatly reduce memory span for English, the alphabet in which the observer was fluent. In fact, the verbal interference lowered memory span for English by only 0.5 character (and raised memory span for the 2×3 Checkers alphabet by a mere 0.15 character). Though these results suggest that rehearsal of verbal labels may contribute slightly to measured memory span, the piddling effect of the secondary verbal task suggests that most of the difference between trained and fluent observers is due to differences in their visual memory (see Brooks, 1968). It is obvious that the effect of training in our observers is specific to the stimuli they encountered, raising the possibility that training with different alphabets might generate anatomically distinct neural changes. This would be consistent with evidence of segregated cortical activation during letter and digit recognition (Polk & Farah, 1999), presumably due to extensive exposure to within-category co-occurrences, at the same time and place.

4. Discussion

We have shown that perimetric complexity (perimeter squared over "ink" area) is an excellent predictor of human efficiency for identifying letters in traditional alphabets (including 2×3 Checkers as Braille). Efficiency increases quickly with experience. Three thousand trials with a given alphabet are enough for a novice observer of any age to become as efficient as a fluent reader, although the trained observer's memory span remains small. Efficiency is insensitive to viewing conditions, being independent of contrast, duration, and eccentricity, and only weakly dependent on size.

Despite an enormous variation in shape of the letters, we consistently found fast learning, efficiency inversely proportional to complexity, and a fixed identification:detection ratio. These results suggest to us a fundamental visual process rather than a language skill, as one might have imagined for this language-related task. Cats and monkeys—even pigeons (Morgan, Fitch, Holman, & Lea, 1976)—might learn to identify letters as efficiently as we do.

4.1. Complexity and Gestalt

Counterexamples to Fig. 4's reciprocal relation between complexity and efficiency are easily contrived. Starting with a simple character—a vertical bar, say—one could change its orientation in 5° steps to generate the 26 letters of a low-complexity alphabet. Despite the low complexity, human observers would be hard-pressed to accurately identify the letters of this alphabet and would have low efficiency, as is the case with many-category low-dimensional absolute judgments generally (Miller, 1956a). Alternatively, imagine forming a new alphabet by attaching the same Chinese character 謝 to the bottom of each of the 26 English letters. This "beard" would greatly increase the complexity of the alphabet, but would be ignored by both human and ideal observers. This bearded alphabet is complex, yet easy, because the complex part is irrelevant. The bar alphabet is simple, yet hard, because its variation is unidimensional. The success of complexity in accounting for efficiency in Fig. 4 suggests that such bad apples have already been culled from the set of alphabets that people actually use. Indeed, the outlier, 4×4 Checkers, is the only non-traditional alphabet (again counting 2×3 Checkers as Braille).

Our finding that low complexity facilitates perception of letter identity is to some extent an affirmation of the Gestalt approach to perception. The original Gestalt principles, set out around 1920, define various image properties, such as symmetry, that contribute to figural "goodness." Subsequent efforts, since 1950, have tried to replace the subjective principles by an objective measure of "goodness." Early Gestalt theory says that a "good" object is more readily perceived as whole (Wertheimer, 1923), and that visual memory tends to enhance an object's "goodness" (Wulf, 1922), much as we would say that low-complexity alphabets yield high efficiencies and that memories are simple, losing complexity (e.g., Attneave, 1955; Fehrer, 1935; Pelli & Farell, 1992; Weinstein, 1955). It is conceivable that prediction of efficiency for identification of arbitrary patterns, beyond letters, might be enhanced by measuring other image properties associated with Gestalt "goodness."[17]

Suggesting that "good patterns have few alternatives," Garner (1970) noted that the Gestalt criteria for "goodness" may implicitly determine the size of the set of similar images (i.e. images sharing the same symmetries, etc.). For letters, if one assumes that the spacing, along a stroke, between changes of stroke curvature is proportional to the stroke width, then the number of independent stroke segments will be proportional to the stroke's aspect ratio (our measure of complexity), and the set size (the number of possible letter forms) will be exponentially related to complexity. This is similar to Miller's (1956b) earlier observation that the appropriate computation in an identification or detection task depends crucially on what components (features) one assumes to be statistically independent.

Attneave and Arnoult (1956), in their thoughtful paper on how to physically describe visual shapes, introduced perimeter squared over area, and then said, "One further point should be made, however; neither [perimetric complexity] nor any other gestalt measure can possibly predict the relative identifiability of shapes except in a limited statistical way. The kinds and degrees of similarity which a shape bears to all the other shapes with which it might be confused [i.e. the rest of the alphabet] will clearly affect the difficulty with which it is identified..." Their disclaimer makes the strong correlation between perimetric complexity and efficiency (Fig. 4) all the more remarkable. But there is no contradiction; identifiability per se does depend on many other factors, e.g., similarity between letters. Happily, efficiency strips away these factors to expose a pure measure of human ability that is in fact predicted by Attneave and Arnoult's perimetric complexity.

Putting aside the complexity and learning results, our results are mostly negative. Efficiency is nearly independent of viewing conditions. That is quietly encouraging, suggesting that efficiency tells us something fundamental about the decision process, independent of how the stimulus is presented. But how do observers decide? And what are we to make of our positive result, the reciprocal relation between efficiency and complexity? Recall that part of our motivation in this study was to challenge the popular notion of feature detection ("probability summation"). Have we dented it? All these questions have one long answer, which, in brief, says that feature detection emerges unscathed, and, indeed, accounts for our complexity result. It is strongly indicated as a necessary first stage in all visual decisions. Let us now review the reasoning that led us to this conclusion, the opposite of what we originally set out to prove, as we try to understand why efficiency is inversely proportional to complexity. Thinking of this as a traditional summation experiment and comparing detection and identification will lead us to fit the standard feature-detection model to the results, showing that it accounts for our main findings.

4.2. Summation

Summation is a classic paradigm in vision research. Threshold is measured as a function of the signal extent in some dimension, e.g., duration or area. Typically the log–log graph is made up of several straight segments, with different slopes. Simple models are devised, each predicting one of those slopes and thus explaining a segment of the graph. Often, the location of each breakpoint in the graph (change of log–log slope, separating regions with different explanations) corresponds to a key parameter of a model, e.g., the time constant of temporal integration. Plotting efficiency instead of the raw threshold enhances the didactic value of this approach, burying the intricacies of the experimental measure in the computation of efficiency, and thereby simplifying the accounts of slope to generic explanations, as illustrated here (Fig. 4) and elsewhere (Pelli & Farell, 1999; Pelli et al., 2003). As signal extent increases from small to large, the efficiency of a detector that integrates over a fixed extent will rise with a log–log slope of +1 as long as the signal is smaller than the detector (because the signal extent is an increasing fraction of the extent over which the detector integrates noise). It will peak, with a log–log slope of zero, when the signal extent matches that of the detector's integrator. And it will fall, with a log–log slope of −1, once the signal becomes larger (because the detector collects a decreasing fraction of the signal energy). The peak efficiency may be less than 100% because the detector may be mismatched in other ways. Here we are interested in slopes 0 and −1. Complete summation corresponds to constant efficiency, a log–log slope of 0. No summation, failing to benefit from the increased signal extent, corresponds to a log–log slope of −1.

In this spirit, recall that perimetric complexity is additive for equal-area patterns (see Introduction), and may thus be understood as an extent. The −0.91 (nearly −1) slope in Fig. 4a over the entire measured range, 30 to 500, indicates that the complexity of our letters exceeds that of any of the observer's detectors. The peak presumably occurs off to the left, at a complexity under 30. Apparently the observer lacks detectors that would match patterns with complexities over 30. If we can model the detectors' receptive fields[18] (the

[17] As we will discuss below, the space-bandwidth products suggested by Näsänen et al. (1993, 1994) as measures of complexity yield similar correlation with efficiency as perimetric complexity does, and have the advantage of being applicable to arbitrary gray-scale images.

[18] A *receptive field* computes a weighted sum over the stimulus; the pattern of weights (an image) is the feature it "detects."

features) as binary images, then their perimetric complexity (defined only for binary images) must be less than 30. In other words, letter identification is mediated by detecting features no more complex than a rectangle  with an aspect ratio of 5:1. The number of such features, abutted end to end, that are required to form a letter is given, approximately, by the ratio of the complexities of the letter and the feature. E.g., forming a bold Bookman letter (complexity 57) would require two features of complexity 30, or more features if the features are less complex.

4.3. Feature detection (probability summation)

Appendix B takes the standard probability summation model for detection and supposes that its scope extends to include identification. The standard model tells us how many features are detected at any signal energy. We assume that the feature detection depends only on the stimulus (signal and noise) regardless of the task (detection or identification). Our finding that the ratio of identification and detection thresholds is independent of letter complexity (Fig. 9b) indicates that the same number of features, k, is detected at the identification threshold for all these alphabets. (Probability summation assumes that only one feature is detected, $k = 1$, at the letter's detection threshold.) This model predicts the reciprocal relation between efficiency and complexity (Figs. 4 and 8a). The precise log–log slope depends on β, which we took as a degree of freedom and found the fitted value to be in the typical range.The detection model and the measured identification:detection threshold ratio together indicate that identifying one of 26 letters (at 64% correct) is based on 7 ± 2 feature detections, for all alphabets tested, independent of the complexity of the alphabet (Appendix B).

Rather than start with a given list of features (e.g., Gibson, 1969), we left the features unspecified, making only the simplifying assumptions, in the tradition of probability summation, that they have equal energy and are detected independently.

We take the fixed identification:detection ratio of Fig. 9b as evidence that the same features mediate detection and identification of our letters. We would not expect this result to hold for degenerate alphabets like the bars and bearded alphabets considered at the beginning of this *Discussion*. Alphabets are designed for identification, tending to have low correlation. If letters had high correlation, so that most of their contrast energy were irrelevant to identification, then we would expect different features to mediate detection and identification, in which case the identification threshold could be arbitrarily higher than detection threshold. Indeed, Tjan et al. (1995) reported detection and identification thresholds for various renderings of a set of simple $2.8° \times 2.8°$ objects: wedge, cone, cylinder, and pyramid. The identification:detection ratio $E_{\mathrm{id}}/E_{\mathrm{det}}$ was 1.6 for line drawings of the objects, and 2.3 for silhouettes. Low-pass filtering the line drawings and silhouettes down to 0.5 c/object—reducing them all to nearly indistinguishable blurry blobs—increased the identification:detection ratio $E_{\mathrm{id}}/E_{\mathrm{det}}$ to 100 (Braje, Tjan, & Legge, 1995).

4.4. Are features revealed by letter confusions?

If an object is identified whenever enough of its features are detected, and these features are detected independently, then it might be possible to figure out what the features are by looking at observers' misidentifications. As we mentioned at the beginning, the psychology literature contains many attempts to account for people's confusions among faintly seen letters of a printed alphabet, typically uppercase Helvetica. The structure of the 26×26 confusion matrix has been carefully scrutinized, and checked for compatibility with various models of human letter recognition (e.g., Loomis, 1982, 1990; Townsend, 1971a, 1971b; Townsend & Ashby, 1982). This has revealed several useful facts. High-threshold theories, which posit that letters are seen veridically or not at all, are consistently rejected by the data: the confusion matrix is not merely a mixture of correct identifications and random guesses. Loomis (1982) showed that response bias plays at most a very small role in determining the structure of most confusion matrices. Luce's (1963) similarity-choice theory always provides an excellent fit, but does little to explain the recognition process. The success of similarity-choice theory, as with many other letter recognition models in the psychology literature, is clouded by its large number of degrees of freedom (more than half the number of cells in the 26×26 confusion matrix). It is therefore important to learn that physical measures of similarity between letters can account for much of the variance of the confusions or similarity parameters derived from the confusion matrix (Gervais et al., 1984; Loomis, 1982, 1990; Gibson, 1969; Townsend, 1971b). The physical measures include template overlap (i.e. the area of overlap when two characters are superimposed) and Euclidean distance between letter spectra filtered by the human visual contrast sensitivity function. However, attempts to use factor analysis or multidimensional scaling to identify physical descriptors ("features") mediating letter discrimination have had only limited success; global attributes of roundness and letter width are important, but the rest of the dimensions thus "revealed" have been uninterpretable (Cavanagh, personal communication; Townsend, 1971b).

4.5. Letters in words

Fig. 10 shows that our novice observers typically triple their efficiency over the course of 2000 trials, while Fig. 11 shows that additional decades of reading experience fail to improve letter identification efficiency. Apparently the algorithm for letter identification is optimized (for the particular letters) within hours, and further years of reading do not improve it. An alternative explanation for this non-effect of reading would be that reading does not expose the visual system to letters per se, but to words (Hadley & Healy, 1991). However, supposing a word superiority due to vastly more experience with words than letters makes it hard to understand why, in Fig. 4, "alphabets" made up of the 26 most common words produce results that fall on the same line as the letter results. The question of whether a word can be identified as a single visual element or only as a combination of letters has a long history, but, by an extension of the methods of this paper, we show elsewhere that words are identified as letter combinations after all (Pelli et al., 2003).

4.6. Clinical cases of letter-by-letter reading (alexia)

Our trained observers combine efficient letter identification with small memory span. This echoes clinical descriptions of Type 2 letter-by-letter readers, who identify letters fairly well, but find it difficult to recognize words even when they have successfully identified each of the letters (Patterson & Kay, 1982; Warrington & Shallice, 1980). "According to Kinsbourne and Warrington (1962, 1963, ...) the [disorder] arises from a difficulty in encoding many separate visual forms simultaneously, a disorder they call simultanagnosia. The main feature of this disorder is a reduced visual span; that is, the maximum number of items that may be reported from a briefly presented array of items is one or two, even though perceptual threshold for a single form seems to be within normal limits" (Arguin & Bub, 1994). Hanley and Kay (1992) "several times asked P.D. for his own impression of how he reads. ...He said that he first tries to form a mental picture of all the letters, which he then attempts to read as a word in his mind's eye. The problem that he apparently encounters when using this strategy is that he cannot 'see' all the letters simultaneously in his visual image." Similarly, after adult-age removal of congenital cataracts, Valvo's (1971) patient HS quickly learned to identify letters, but said, "it was impossible for me to make out whole words; I managed to do so only after weeks of exhausting attempts. In fact, it was impossible for me to remember all the letters together, after having read them one by one." Our trained observers share these characteristics; they identify letters as well as fluent readers do, but have a memory span of only one or two letters.

We were surprised that the two letter-related skills measured by identification efficiency and memory span are learned so independently. Letter identification for a foreign alphabet can be improved without improving memory span for that alphabet, as we have shown here, and preliminary data (not shown) suggest that practice at the memory task can greatly increase memory span with little effect on letter identification efficiency. These two skills may correspond to the two independent lesion-impaired "sub-systems" posited by Farah (1991, 1994) to account for the patterns of deficit in her review of 99 cases of associative agnosia (i.e., impaired visual recognition despite the ability to copy drawings).

4.7. Perceptual learning: Slow and fast

The dichotomy between identifying and remembering may help explain the very different learning rates found in related experiments. Kolers (1975) found that it took weeks of practice to learn to read mirror-inverted text at normal rates. Presumably, Kolers's observers spent most of their time modifying reading processes other than letter identification. Supposing that learning mirror-inverted text is like learning a foreign alphabet or a radically new font, we would expect identification efficiency of mirrored letters to reach fluent levels in a few hours, as in Fig. 10, but it seems very likely, especially considering the clinical evidence above, that fast reading demands a large memory span, which may develop slowly.

The learning curves in Fig. 10 show that the efficiency of trained letter identifiers reaches that of fluent readers in about three thousand trials. So-called "perceptual" learning of many simple discrimination tasks exhibits a similar time course, e.g., for learning vernier acuity (Fahle & Edelman, 1993), texture segregation (Karni & Sagi, 1991), and line orientation (Shiu & Pashler, 1991). Thus there is nothing in our data to support a distinction between learning letters and learning these other supposedly "low-level" perceptual discriminations.

4.8. Crowding

A letter in the peripheral visual field is much harder to identify if other letters are nearby. This is called "crowding" (for review see Pelli et al., 2004). Crowding may determine the visual span and thus reading rate. However, crowding did not affect any of the experiments reported here because we never presented more than one letter in the periphery.

4.9. Reading

One motive for this study was to gauge plastic cortical changes over the course of reading a billion letters. However, while the results clearly show a dramatic permanent change associated with learning an alphabet (Fig. 10), we found no correlation with number of years of reading (Fig. 11). Three thousand trials in the lab were enough to reach the efficiency of fluent readers. Clearly, if the goal is to improve efficiency of letter identification, ordinary reading is a poor substitute for identifying faint letters. While there are many differences between reading and our letter identification task, the pattern of our results—efficiency independent of contrast and dependent on complexity, identically for letters and words—suggests that the key difference is the effective signal-to-noise ratio,[19] which is high in books, and low in our task. In their 4-letter

[19] The effective signal-to-noise ratio is $D^+ = E/(N + N_{eq})$, where N_{eq} is the observer's equivalent input noise level (Pelli & Farell, 1999).

variant of our task, Suchow and Pelli (2005) found that the improved efficiency was due to seeing the near-threshold target letter, not the easily-visible letters on the response screen. Reading books neither demands nor encourages more than moderate efficiency. However, sometimes we do read at a low signal-to-noise ratio. Many signs, especially highway signs, are customarily read at the greatest possible distance, i.e. as soon as possible during approach. Good highway signs maximize sight distance by using low complexity—bold—lettering. Higher-efficiency observers can read signs from further away, i.e. sooner. Reading distant signs may be better practice for improving letter identification efficiency than ordinary reading. Conversely, extensive practice identifying letters in noise may improve acuity.

4.10. Children learning letters

Our finding that observers of all ages follow similar learning curves (Fig. 10) and have similar efficiency (Fig. 11) stands in contrast to reports of large improvement in shape identification during childhood. Gibson, Gibson, Pick, and Osser (1962) report a dramatic improvement between the ages of 4 and 8 in the ability to match letter-like shapes. Similarly, from age 2 to 7, Martelli et al. found a 100-fold improvement in efficiency for identifying one of three letters (Martelli et al., 2002; Martelli, Baweja, Mishra, Majaj, & Pelli, in preparation). However, there is no conflict among these results. The present experiment involved many more trials and thus provided training. We believe that our current study assessed capacity to learn, which apparently never changes, whereas the other studies assessed what is gained outside the lab as the child grows up. The Gibson et al. observers made just one match for each of twelve shapes, a total of 12 trials. Martelli et al. tested each observer for just 69 trials to measure a threshold (69 test cards at 23 different signal-to-noise ratios, most of which are well above or well below threshold). Our observers did hundreds (Fig. 11) or thousands (Fig. 10) of trials. We suspect that young children perform less well only in the first few trials. The steep rise in efficiency from age 2 to 7 (and much less steep from 7 to 10) found by Martelli et al. with their quick test (69 trials, mostly far from threshold) seems consistent with the nearly flat efficiency from age 6 to 10 found here (Fig. 11) with our extended test ($5 \times 40 = 200$ trials, mostly near threshold), which is enough to provide significant training.

Over the course of hundreds of trials, the learning curve for our youngest observer (JF, age 3, in Fig. 10) shows that she learned her ABC's just as quickly as our adult observers learned foreign alphabets. The poor performance of four-year-olds on the Gibson et al. task has been taken to indicate that, sometime between the ages of 4 and 8, children acquire the ability to discriminate among letter-like forms, a seemingly obvious prerequisite for learning to read (see Rayner & Pollatsek, 1989, for a discussion). However, we found that the ability to learn to identify letters was already mature in the 3-year-old we studied.

4.11. Are letters special?

We asked observers to identify letters because they do it so well. We wanted to show that people can recognize familiar objects much better than any model based on independent feature detection could. Letters are over-learned man-made elements of language, designed to be easily distinguished. Reading an hour a day, by age 50, you will process a billion letters and may devote certain areas of your brain to visual processing of letters and words (McCandliss, Cohen, & Dehaene, 2003; Polk et al., 2002; Steingrimsson, Majaj, & Pelli, 2003). When recognizing objects, people prefer categories that are "basic" over coarser and finer categories (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Like dog and chair, letters—a, b, c—are basic categories by Rosch's criteria: They have short names, they are learned first, and they are the most inclusive category that can still be represented by a concrete image of the category as a whole. Surely, we thought, if any stimulus set could help us prove that the human visual system is capable of doing something smarter than independent feature detection (probability summation), letters are it.

Instead, we find that after just three thousand trials, novice observers can identify the letters of a new alphabet as well as adult native speakers who have read fluently in this alphabet for all their lives. Probability summation, which we had sought to disprove, ultimately provides a parsimonious account of our main results, the reciprocal relation between efficiency and complexity and the fixed ratio of identification and detection thresholds, independent of complexity. The same visual channels that mediate the detection of gratings mediate the identification of letters (Solomon & Pelli, 1994) and reading (Majaj, Liang, Martelli, Berger, & Pelli, 2003). By this account, features are detected independently of each other, indifferent to whether the task is detection or identification and whether the stimulus is a letter or a leopard. In light of all this, and despite the unique role letters play in language, it seems that when it comes to identification and detection, letters are not special after all: Letter and word recognition are typical of everyday object recognition, and the underlying visual computations are the same as those used to recognize a spoon, a face, or a building.

4.12. The computation

We began by asking what computation underlies letter identification: how the remembered alphabet and the many pixels of the noisy stimulus letter are combined to reach a single decision. Recall that we originally sought to disprove feature detection, in the context of object recognition, by showing high efficiency for identification of complex objects. Like Wertheimer (1923) and Marr (1982), we were reluctant to think that the visual system is so crude as to reduce a rich visual stimulus to a mere list of assertions about elementary features. We imagined that, at least for familiar patterns, the visual system would linearly combine many simple receptive fields to synthesize complex receptive fields matched to the possible signals, e.g., the letters a–z. (Indeed, receptive fields of neurons in inferotemporal cortex seem about as complex as whole letters; Desimone, Albright, Gross, and Bruce, 1984; Tanaka, Saito, Fukada, & Moriya, 1991.) Making decisions about individual features, each with much less contrast energy than the whole signal, would greatly reduce our efficiency in detecting and identifying patterns in noise, since we would be limited by the thresholds of the feature detectors rather than that of a unit matched to the whole signal. However, our data reveal precisely this kind of inefficiency in visual identification: a reciprocal relation between efficiency and complexity (Fig. 4a). We also found a tight link between detection and identification (Fig. 9b). Both results are parsimoniously accounted for by feature detection: Perceptual decisions are based on independently detected features. A first stage of visual processing detects features independently, and only the detected features are available to subsequent stages of visual processing.

5. Conclusion

Visual object recognition can be studied by measuring threshold contrast for identifying letters in static noise. Efficiency—the ratio of ideal to human contrast energy at threshold—strips away the intrinsic difficulty of the task, leaving a pure measure of human ability that may be compared across a wide range of tasks. Efficiency of letter identification is independent of contrast, duration, and eccentricity, and only weakly dependent on size, indicating that observers are identifying letters by using a similar computation across this wide range of viewing conditions. Efficiency of identification for most traditional alphabets tested (Armenian, Devanagari, English, Hebrew) is about 10%. Across all tested typefaces, alphabets (except 4×4 Checkers), and even words, efficiency is approximately 9 divided by the perimetric complexity (perimeter squared over "ink" area). This, and the surprisingly fixed ratio of detection and identification thresholds, suggest that letter identification is mediated by detection of about 7 visual features.

Learning is fast—it takes only 2000–4000 trials for a novice observer to reach the same efficiency as a fluent reader—for all alphabets and ages (3–68). Extended reading, even reading a billion letters over forty years, does not improve efficiency of letter identification. Surprisingly, despite attaining the same efficiency as fluent readers, observers trained to identify new alphabets have the same meager memory span for random strings of these characters as observers seeing them for the first time, less than half the memory span of fluent readers. In this regard, our trained letter identifiers are strikingly like clinical cases of letter-by-letter readers. Apparently, learning to read fluently involves not only learning letters and words, but also increasing memory span from 1 or 2 letters, to 5 or so.

Our key result is the reciprocal relation between efficiency and complexity. This represents a severe early bottleneck in visual perception: simple forms are seen efficiently, complex forms inefficiently, as though they could only be seen by means of independent detection of multiple simple features.

Acknowledgments

Thanks to Chris Chiarello, Norma Graham, Gordon Legge, Jack Loomis, Marialuisa Martelli, Melanie Palomares, Jamie Radner, and Katharine Tillman for many helpful comments. Thanks to John Robson, who suggested comparing identification and detection thresholds (Fig. 9b), suggested estimating what effect intrinsic uncertainty would have (Appendix A.3), and helped us with typographic terminology. Thanks to George Sperling for dispelling the apparent inconsistency between our efficiencies and those of Parish and Sperling (1991). Thanks to Dennis Levi for discussions of the last channel, and him and an anonymous reviewer for helpful suggestions about eccentricity. Thanks to Bob Shapley for suggesting that Fig. 4a should include the probability summation prediction Eq. (B.14). Thanks to Stijn Oomes for catching a mistake in our discussion of additivity of complexity. Thanks to Mab Huang for lending us the Chinese primer, and to Steve Chamberlain for help in scanning characters from it. Thanks to Diana Balmori, Cesar Pelli, and Beau Watson for helping to make the Abstract clear. Thanks to Manoj Raghavan and Michael Landy for help with the appendices. Thanks to Jordan Suchow for help fine-tuning our conclusions on learning. Thanks to Doris Aaronson and Gay Snodgrass for telling us about and providing a copy of Micha Razel's dissertation. Thanks to Seema Khan for *Hopscotch*. And thanks to Misha Pavel for help with hačeks. Some of these results were presented at the Annual Meetings of the Association for Research in Vision and Ophthalmology in Sarasota, FL, May, 1994, 1995, and 1996. Most of the empirical results appeared in C.W. Burns's doctoral dissertation, Syracuse University, 1995. An earlier draft (8) was accepted by *Vision Research* in 1995, subject to minor revisions. Those, ultimately extensive, revisions resulted in the current draft (73), submitted anew to the journal. Intermediate drafts have been cited as "Pelli, D. G., Burns, C. W., Farell, B., & Moore, D. C. (accepted). Identifying letters. *Vision Research*." This work was supported by NIH grants EY04432 to Denis Pelli and EY09872 and EY012286 to Bart Farell.

Appendix A. The ideal observer

Signal detection theory showed that the detectability of an exactly known signal in white noise depends on only one number, E/N (Peterson, Birdsall, & Fox, 1954). That theoretical result gives energy a central role and is the basis for defining efficiency in terms of energy. The more general case of identification is similar, but not quite so simple.

Sections A.1 and A.2 derive the properties of the ideal identifier of letters in noise (for a fuller treatment see Duda & Hart, 1973; Van Trees, 1968). Section A.3 introduces simplifying assumptions (equal probability, energy, and cross-correlation) that allow us to solve for the ideal's threshold in our tasks, to compare with our assumption-free brute-force threshold measurements.

A.1. The Bayesian classifier

The task is to identify one of m possible signals $\mathbf{H}_1,\ldots,\mathbf{H}_m$ in the presence of added white noise with standard deviation σ. The ideal observer compares the noisy stimulus image with each of the noise-free signal images, one for each possible signal (e.g., a, ..., z), choosing the one with the maximum posterior probability. This maximizes the probability of correct identification. When the signals are words, the noisy stimulus image is compared with a template for each possible word. The received data $\mathbf{D}$ (i.e., the stimulus) and the m hypotheses $\mathbf{H}_1,\ldots,\mathbf{H}_m$ are all large vectors, each containing q pixels.

If we suppose that the signal is $\mathbf{H}_i$, then we know the probabilities of the possible values for the pixel luminances in the noisy stimulus image. Let D_j be the luminance of the jth pixel in the image $\mathbf{D}$. It has a Gaussian probability density function

$$P(D_j|H_{ij}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\frac{-(D_j-H_{ij})^2}{2\sigma^2}, \tag{A.1}$$

with mean H_{ij}, the value of the jth pixel in signal $\mathbf{H}_i$, and standard deviation σ. Since the noise is independent from pixel to pixel, the overall probability of the image $\mathbf{D}$ is the product of the probabilities of all its pixels,

$$P(\mathbf{D}|\mathbf{H}_i) = \prod_{j=1}^{q} P(D_j|H_{ij}), \tag{A.2}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^q \exp\left(\frac{-1}{2\sigma^2}\sum_{j=1}^{q}(D_j-H_{ij})^2\right), \tag{A.3}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^q \exp\frac{-\|\mathbf{D}-\mathbf{H}_i\|^2}{2\sigma^2}. \tag{A.4}$$

If the prior probabilities are all equal then choosing i to maximize posterior probability, Eq. (3), reduces to simply maximizing the likelihood $P(\mathbf{D}|\mathbf{H}_i)$. By inspection of Eq. (A.4), we see that choosing the hypothesis $\mathbf{H}_i$ that maximizes the likelihood $P(\mathbf{D}|\mathbf{H}_i)$ is the same as minimizing the squared error $\|\mathbf{D}-\mathbf{H}_i\|^2$ between the hypothesis $\mathbf{H}_i$ and the data $\mathbf{D}$. That is what our ideal-observer computer program does. On each trial, it chooses the hypothesis closest to the data.

A.2. Determinants of ideal performance

All the ideal observer identification thresholds plotted in this paper were obtained by direct trial-by-trial implementation of the algorithm in the preceding section, using the same threshold estimation procedure as with our human observers. However, it is still helpful to know what aspects of the task and stimuli determine the ideal observer's level of performance. For the general case of identifying 1 of m equally probable signals in white noise, we show that the ideal's probability of correct identification depends only on the covariance of the signals (an $m\times m$ matrix), normalized by the noise level.

The probability P_{id} of a correct identification by the ideal observer is simply the probability that the correct hypothesis is nearest. Assume signal $\mathbf{H}_\kappa$. (The expected proportion correct will be the average over $\kappa = 1$ to m.)

$$P_{\text{id}} = P(\|\mathbf{D}-\mathbf{H}_\kappa\|^2 < \|\mathbf{D}-\mathbf{H}_i\|^2 \quad \text{for all } i\neq\kappa), \tag{A.5}$$

$$= P(\|\mathbf{X}\|^2 < \|\mathbf{X}-\mathbf{G}_{i\kappa}\|^2 \quad \text{for all } i\neq\kappa), \tag{A.6}$$

where $\mathbf{X} = \mathbf{D}-\mathbf{H}_\kappa$ is the white noise in the stimulus, and $\mathbf{G}_{i\kappa} = \mathbf{H}_i - \mathbf{H}_\kappa$ is the difference between each candidate signal and the correct signal.

$$=P(\|\mathbf{X}\|^2 < \|\mathbf{X}\|^2 - 2\mathbf{X}\cdot\mathbf{G}_{i\kappa} + \|\mathbf{G}_{i\kappa}\|^2 \quad \text{for all } i\neq\kappa), \tag{A.7}$$

$$=P\left(\frac{\mathbf{X}\cdot\mathbf{G}_{i\kappa}}{\|\mathbf{G}_{i\kappa}\|^2} < \frac{1}{2} \quad \text{for all } i\neq\kappa\right), \tag{A.8}$$

$$=P\left(Y_i < \frac{\|\mathbf{G}_{i\kappa}\|}{2\sigma} \quad \text{for all } i\neq\kappa\right), \tag{A.9}$$

where "$\cdot$" is the dot product and $Y_i = \mathbf{X}\cdot\mathbf{G}_{i\kappa}/\sigma\|\mathbf{G}_{i\kappa}\|$ is a normally distributed random variable (a scalar) with zero mean and unit variance. The probability density function (pdf) of the multivariate normal random vector $\mathbf{Y} = [Y_1,\ldots, Y_{\kappa\text{-}1}, Y_{\kappa+1},\ldots, Y_m]$ is determined by its covariance matrix

$$\langle Y_i Y_j\rangle_{\mathbf{X}} = \left\langle \frac{\mathbf{X}\cdot\mathbf{G}_{i\kappa}}{\sigma\|\mathbf{G}_{i\kappa}\|}\frac{\mathbf{X}\cdot\mathbf{G}_{j\kappa}}{\sigma\|\mathbf{G}_{j\kappa}\|}\right\rangle_{\mathbf{X}} = \frac{\mathbf{G}_{i\kappa}\cdot\mathbf{G}_{j\kappa}}{\|\mathbf{G}_{i\kappa}\|\|\mathbf{G}_{j\kappa}\|}, \tag{A.10}$$

where the angle brackets $\langle\ \rangle_{\mathbf{X}}$ indicate the average across all possible values of the noise $\mathbf{X}$. The probability P_{id}, in Eq. (A.9) above, of a correct identification by the ideal observer depends only on $\|\mathbf{G}_{i\kappa}\|/\sigma$ and the pdf of $\mathbf{Y}$, the latter of which depends only on $(\mathbf{G}_{i\kappa}\cdot\mathbf{G}_{j\kappa})/(\|\mathbf{G}_{i\kappa}\|\|\mathbf{G}_{j\kappa}\|)$. In terms of the original signals, we define the covariance $C_{ij} = \mathbf{H}_i\cdot\mathbf{H}_j$, so we can write

$$\frac{\|\mathbf{G}_{i\kappa}\|^2}{\sigma^2} = \frac{\|\mathbf{H}_i-\mathbf{H}_\kappa\|^2}{\sigma^2} = \frac{C_{ii}-2C_{i\kappa}+C_{\kappa\kappa}}{\sigma^2} \tag{A.11}$$

and

$$\frac{\mathbf{G}_{i\kappa}\cdot\mathbf{G}_{j\kappa}}{\|\mathbf{G}_{i\kappa}\|\|\mathbf{G}_{j\kappa}\|} = \frac{(\mathbf{H}_i-\mathbf{H}_\kappa)\cdot(\mathbf{H}_j-\mathbf{H}_\kappa)}{\sqrt{\|\mathbf{H}_i-\mathbf{H}_\kappa\|^2\|\mathbf{H}_j-\mathbf{H}_\kappa\|^2}}$$

$$= \frac{C_{ij}-C_{i\kappa}-C_{\kappa j}+C_{\kappa\kappa}}{\sqrt{(C_{ii}-2C_{i\kappa}+C_{\kappa\kappa})(C_{jj}-2C_{j\kappa}+C_{\kappa\kappa})}}. \tag{A.12}$$

Looking at Eqs. (A.11) and (A.12), we see that these expressions depend only on terms like C_{ij}/σ^2. Let $\mathbf{C}$ be the $m\times m$ covariance matrix of the m signals. Thus the ideal observer's level of performance (the expected value of P_{id}) depends only on $\mathbf{C}/\sigma^2$, i.e., the covariance of the signals, normalized by the noise variance.

A.3. The ideal observer's threshold

Although the preceding section showed that the ideal's level of performance depends on the signals only through their covariance matrix, that matrix is still quite big, with on the order of $m^2/2$ parameters. Reducing that matrix to just three parameters—energy, correlation, and number of signals—will help us understand how the correlation and number of letters in an alphabet affect the ideal's identification threshold.

We solve for the ideal's threshold in the special case of equally probable equal-energy signals with equal covariance between signals. Real alphabets have unequal energy and covariance, but we hope that the simpler case—which we can solve analytically—will provide insight into the general case.

The preceding sections of this appendix ignored the pixel area A, and terms like E and N that depend on pixel area, because the ideal observer's performance is independent of the displayed size of a pixel. However, using E and N will help us expose the connection between this derivation and the empirical work. The (average) signal energy is $E = A\langle C_{ii}\rangle_{i=1,m}$ and the noise power spectral density is $N = A\sigma^2$.

We assume that the elements in the covariance matrix have only two values. On the diagonal, all the elements have the value $C_{ii} = E/A$. Off the diagonal, all the elements have the value $C_{i\neq j} = \rho E/A$, where ρ is the cross correlation, $0 \leqslant \rho \leqslant 1$. Substituting into Eqs. (A.11) and (A.12) yields

$$\frac{\|\mathbf{G}_{i\kappa}\|^2}{\sigma^2} = 2\frac{(1-\rho)C_{ii}}{\sigma^2} = 2\frac{(1-\rho)E}{N} \quad \text{if } i \neq k, \tag{A.13}$$

and

$$\frac{\mathbf{G}_{i\kappa}\cdot\mathbf{G}_{j\kappa}}{\|\mathbf{G}_{i\kappa}\|\|\mathbf{G}_{j\kappa}\|} = \begin{cases} 1/2 & \text{if } i \neq j, \\ 1 & \text{if } i = j, \end{cases} \tag{A.14}$$

which reveals that performance P_{id} depends on ρ, E, and N solely through the signal-to-noise ratio $D = (1-\rho)E/N$. (The signal-to-noise ratio D is not to be confused with the data $\mathbf{D}$ or the pixel luminance D_j.) Thus we need only solve explicitly the case where all the signals are orthogonal, $\rho = 0$, because the effect of introducing nonzero (but equal) correlation is equivalent to reducing the energy from E to $(1-\rho)E$ (see Nuttall, 1962). ρ tells you what fraction of the energy is wasted.

With Eq. (A.13), the probability of correct identification Eq. (A.9) becomes

$$P_{\text{id}} = P\left(Y_i < \sqrt{\frac{E}{2N}} \quad \text{for all } i \neq \kappa\right), \tag{A.15}$$

where

$$Y_i = \frac{\mathbf{X}\cdot\mathbf{H}_i - \mathbf{X}\cdot\mathbf{H}_\kappa}{\sqrt{2NE/A^2}} = \frac{Z_i - Z_\kappa}{\sqrt{2}}. \tag{A.16}$$

$Z_i = \mathbf{X}\cdot\mathbf{H}_i/\sqrt{NE/A^2}$ is normally distributed with zero mean and unit variance. The Z's are statistically independent because the signals $\mathbf{H}_i$ are orthogonal and the noise $\mathbf{X}$ is white. Thus the probability of correct identification is

$$P_{\text{id}} = P\left(Z_i < Z_\kappa + \sqrt{\frac{E}{N}} \quad \text{for all } i \neq \kappa\right). \tag{A.17}$$

In Eq. (A.17), the comparisons indexed by i all share the same Z_κ, so their probabilities are not independent. The cumulative normal distribution is

$$P(Z_i < u) = \Phi(u) = \int_{-\infty}^{u} \mathrm{d}z\,\phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u} \mathrm{d}z \exp\frac{-z^2}{2}. \tag{A.18}$$

For $m-1$ independent identically distributed random variables,

$$P(Z_i < u \text{ for all } i \neq \kappa) = \prod_{i\neq\kappa} P(Z_i < u) = \Phi^{m-1}(u). \tag{A.19}$$

Thus Eq. (A.17) becomes

$$P_{\text{id}} = \int_{-\infty}^{\infty} \mathrm{d}z\ \phi(z)\ \Phi^{m-1}\left(z + \sqrt{\frac{E}{N}}\right). \tag{A.20}$$

This cannot be simplified further, but is easily integrated numerically. As we noted at the beginning, to deal with the more general case of nonzero (but equal) correlation between signals, we just replace E by $(1-\rho)E$,

$$P_{\text{id}} = \int_{-\infty}^{\infty} \mathrm{d}z\ \phi(z)\ \Phi^{m-1}\left(z + \sqrt{\frac{(1-\rho)E}{N}}\right). \tag{A.21}$$

At threshold (i.e. any fixed value of P_{id}) the dimensionless quantity $D = (1-\rho)E/N$ will depend only on m, the number of signals.

All that remains is to describe the dependence of D on m. Solving Eq. (A.21) numerically for D at threshold, $P_{\text{id}} = 0.64$, reveals that D is logarithmically related to the number of signals,

$$D = -1.189 + 4.757 \log m + \varepsilon, \tag{A.22}$$

with little error ($|\varepsilon| < 0.0372$) for any conceivable number of signals ($2 \leqslant m \leqslant 10^9$). After expanding D we can solve for the threshold energy,

$$E = \frac{N}{1-\rho}(-1.189 + 4.757 \log m + \varepsilon). \tag{A.23}$$

We use the ideal observer to characterize the intrinsic difficulty of each task. To aid our intuitions, we have solved for the ideal's threshold in a special case (equal probability, energy, and cross correlation) that may capture the most important differences between alphabets for identification by the ideal. Extrapolating this result to real alphabets—with *unequal* energies and cross correlations—we take E to be the average signal energy $E = A\langle C_{ii}\rangle_{i=1,m}$ and ρ to be the average cross correlation $\rho = \langle C_{ij}/\sqrt{C_{ii}C_{jj}}\rangle_{i\neq j}$. We designate the extrapolation, based on Eq. (A.23) with zero ε, as $\hat{E}$,

$$\hat{E} = \frac{N}{1-\rho}(-1.189 + 4.757 \log m). \tag{A.24}$$

Table A
Efficiencies for identifying letters in various alphabets by fluent readers of traditional alphabets and trained identifiers of artificial alphabets

Alphabet	Efficiency		Letters							Ideal threshold			Human observer thresholds			
	η^+	$\log \eta^+$	n	Complexity	Overlap	Height	Width	Area	$\log E_1$	$\log E$	$\log c$	$\log \hat{E}/E$	$\log E^+$	$\log c^+$	#	Fluent, *Trained*
Bookman	0.090	−1.05	26	107	0.34	1.1	0.95	0.38	−0.42	−2.60	−1.09	0.00	−1.55	−0.57	19	AR, HE, AE, KE, RE, KA, DM, RJ, KB, DA, JM, BD, MS, CC, BF, SC, GE, BS, RV
Uppercase Bookman	0.090	−1.05	26	139	0.36	1.3	1.27	0.52	−0.29	−2.63	−1.17	0.04	−1.58	−0.65	2	MP, BS
Bold Bookman	0.155	−0.81	26	57	0.53	1.2	1.02	0.66	−0.18	−2.44	−1.13	−0.01	−1.63	−0.73	3	DM, MP, BS
Courier	0.112	−0.95	26	100	0.45	1.0	0.93	0.35	−0.46	−2.56	−1.05	0.04	−1.61	−0.57	3	EG, MP, BS
Helvetica	0.142	−0.85	26	67	0.41	1.2	0.76	0.43	−0.37	−2.45	−1.04	−0.10	−1.60	−0.62	4	DM, CW, FA, WT
Künstler	0.025	−1.61	26	451	0.22	1.4	2.05	0.45	−0.35	−2.73	−1.19	0.06	−1.12	−0.39	2	WT, RM
Sloan	0.108	−0.97	10	65	0.58	1.9	1.78	1.87	0.27	−2.59	−1.43	−0.01	−1.62	−0.95	5	DM, MP, BS, FA, CW
Arabic	0.061	−1.22	28	137	0.21	1.0	1.01	0.35	−0.46	−2.50	−1.02	−0.17	−1.28	−0.41	2	TH, IM
Armenian	0.094	−1.03	37	106	0.33	1.1	0.82	0.29	−0.54	−2.54	−1.00	−0.01	−1.51	−0.49	1	VA
Chinese	0.051	−1.29	26	199	0.32	1.3	1.22	0.41	−0.39	−2.67	−1.14	0.06	−1.38	−0.49	2	KH, CH
Devanagari	0.097	−1.01	26	99	0.61	0.8	0.94	0.35	−0.46	−2.34	−0.94	−0.03	−1.32	−0.43	2	SA, *AW, DM*
Hebrew	0.106	−0.97	22	90	0.48	1.0	0.67	0.26	−0.59	−2.51	−0.96	−0.02	−1.53	−0.47	2	SS, *AW*
2×3 Checkers	0.308	−0.51	26	28	0.55	1.3	0.96	0.88	−0.06	−2.34	−1.14	−0.10	−1.83	−0.88	2	*DM, AW*
4×4 Checkers	0.066	−1.18	26	52	0.48	1.5	1.48	1.10	0.04	−2.50	−1.27	0.01	−1.32	−0.68	2	*SE, AW*
Three-letter words	0.033	−1.48	26	304	0.34	1.2	3.07	1.10	0.04	−2.64	−1.34	0.04	−1.16	−0.60	4	DM, MP, BS, CF
Five-letter words	0.022	−1.65	26	481	0.32	1.4	4.83	1.75	0.24	−2.62	−1.43	0.01	−0.97	−0.60	3	DM, MP, BS
Five-letter words	0.034	−1.47	2213	499	0.32	1.4	4.90	1.75	0.24	−2.28	−1.26	0.09	−0.81	−0.53	2	DM, CB

Across observers, the standard deviation of threshold $\log E^+$ (and thus efficiency $\log \eta^+$) is about 0.1. All sizes are in degrees. The nominal point size of all letters was 58 point. The viewing distance was 60 cm. Künstler, the uppercase decorative script, was initially unfamiliar to our observers, so we discarded each observer's first 400 trials. The words were all in Bookman. *Overlap* is the average area of intersection of two superimposed letters, divided by the average ink area of a letter. (We suppose that overlap is negligibly different, for our purposes, from average cross correlation ρ.) *Height* and *width* are the extents from the bottom of descender to top of ascender and left edge to right edge, respectively, averaged across all letters in the alphabet. *Area* is "ink" area, averaged across all letters in the alphabet, and equal to the energy at unit contrast, E_1. $E^+ = E - E_0$ is the energy threshold, corrected for the threshold E_0 with no display noise. c^+ is the contrast threshold, similarly corrected, $c^+ = (c^2 - c_0^2)^{0.5}$, where c_0 is the threshold with no display noise. The static noise rms contrast was 0.25, and the power spectral density N was $10^{-3.60}$ deg^2.

To assess how accurately $\hat{E}$ predicts the ideal threshold energy for the alphabets used in this paper, we take the ratio $\hat{E}/E$ of estimated and actual thresholds for identification by the ideal. (We used overlap as our estimate of ρ.) Table A shows that $-0.17 \leqslant \log \hat{E}/E \leqslant 0.06$, indicating that $\hat{E}$ is a useful approximation, capturing the main effects of signal correlation and number (see Fig. 9c), but, in the general case, is no substitute for accurate determination of the ideal threshold (see Table A).

Appendix B. What probability summation tells us about letter identification

Visual science has explained how observers detect but has yet to explain how they identify. This paper strives to shed light on the latter. We lack a model for identification, but we have a good model for detection: probability summation (Brindley, 1954; Graham, 1977, 1980; Quick, 1974; Robson & Graham, 1981; Watson & Ahumada, 2005). Probability summation supposes that the observer detects the signal by detecting any one of its features, and that the features are detected independently. The model specifies the probability of feature detection at any contrast. To apply this model, we suppose that the number of features in a letter is proportional to its perimetric complexity and that the feature detections are autonomous, indifferent to whether the task is detection or identification. Then we ask what the detection model can tell us about what is happening at the measured threshold contrast for identification. Fitting this model to human performance indicates that identifying one of 26 letters (at 64% correct) is based on 7 ± 2 feature detections, for all alphabets tested, independent of the complexity of the alphabet.

This appendix defines feature detection and probability summation (Section B.1), calculates the number of features detected (Section B.3), applies this to letter detection and identification (Sections B.2 and B.4), and finally shows that the ratio of the model's identification and detection thresholds reveals the number of features detected at the identification threshold (Section B.5).

B.1. Feature detection and probability summation

One approach to modeling object recognition is to suppose that the visual scene is analyzed into components, which we will call *features*, and that objects are recognized as combinations of features (e.g., Graham, 1980; Townsend & Ashby, 1982). For the purposes of this appendix it does not matter what the features are, just that they are detected independently.

For computational convenience we assume that the probability P_{f} of asserting the presence of a particular feature is a Weibull function Eq. (1),

$$P_{\mathrm{f}}(\varepsilon) = 1 - (1 - \gamma_{\mathrm{f}})\mathrm{e}^{-(\varepsilon/\varepsilon_{\mathrm{f}})^{\beta/2}}, \tag{B.1}$$

where ε is the feature's energy in the signal, and γ_{f}, ε_{f}, and β are fixed parameters that we will set to fit the observer's performance.[20] We call ε_{f} the feature threshold. (Note that the exponent $-(\varepsilon/\varepsilon_{\mathrm{f}})^{\beta/2}$ is equivalent to the familiar $-(c/\alpha)^{\beta}$, where c is contrast, because energy is proportional to squared contrast.) The energy of the whole letter is E. For simplicity, we assume that the number of features, n, in a letter is proportional to its complexity, that each feature has energy $\varepsilon = E/n$, and that, for any given signal, the feature assertions are statistically independent. We suppose that the observer can access the stimulus solely through the list of feature assertions.

The formula for probability of feature assertion Eq. (B.1) can be re-written,

$$P_{\mathrm{f}}(\varepsilon) = 1 - (1 - P_{\mathrm{f}}(0))(1 - P^{*}_{\mathrm{f}}(\varepsilon)), \tag{B.2}$$

to show that it implies that a feature may be asserted for either (or both) of two statistically independent reasons, one *relevant* and one *irrelevant*.[21] The probability $P^{*}_{\mathrm{f}}(\varepsilon) = 1 - \mathrm{e}^{-(\varepsilon/\varepsilon_{\mathrm{f}})^{\beta/2}}$ of a relevant assertion depends on signal energy. The probability $P_{\mathrm{f}}(0) = \gamma_{\mathrm{f}}$ of an irrelevant (spontaneous) assertion is independent of signal energy. Discounting the irrelevant assertions, we say that a feature is *detected* if and only if it is asserted relevantly. Solving Eq. (B.2) for the probability of detection $P^{*}_{\mathrm{f}}(\varepsilon)$ is called "correcting for guessing,"

$$P^{*}_{\mathrm{f}}(\varepsilon) = 1 - \frac{1 - P_{\mathrm{f}}(\varepsilon)}{1 - P_{\mathrm{f}}(0)} = 1 - \mathrm{e}^{-(\varepsilon/\varepsilon_{\mathrm{f}})^{\beta/2}}. \tag{B.3}$$

In fact, our observers had low false alarm rates, so correcting for guessing makes little difference for our results.[22]

[20] Strictly speaking, assuming a Weibull function would imply a high threshold, i.e. that the assertions that occur at zero contrast arise through an entirely separate process from that which generates the contrast-dependent assertions at high contrast. However, for the purpose of calculating summation effects, Pelli (1985) has shown that a milder assumption (correcting for guessing) suffices to allow the use of fitted Weibull functions for accurate predictions of summation, without the high-threshold assumption.

[21] Hits and false alarms have been associated with similar activation of areas V1, V2, and V3 of visual cortex when the observer was trying to detect a low contrast target (Ress & Heeger, 2003).

[22] Incidentally, even though we assume that feature detections are all independent and obey this psychometric function, we are not ruling out the possibility of attentional and criterion effects on the feature detections, which could affect the parameters of the psychometric function. If correction for guessing seems incompatible with a subjective criterion, note that the high-threshold literature has sometimes confused two senses of "independence." It is not the same to ask whether hits and guesses occur statistically independently on each trial as to ask whether the mean rate of one can be changed without affecting the mean rate of the other. There is good evidence for the statistical independence of hits and guesses, which justifies correction for guessing. And there is good evidence that observers induced to change their criterion only manage to trace out a fixed ROC curve, rejecting the idea that observers can independently manipulate their hit and false alarm rates. So it is reasonable to correct for guessing and still expect the observer's criterion to affect both the hit and false alarm rates (see Pelli, 1985).

B.2. Detection threshold

The observer's detection threshold energy E_{det} is the letter energy required to achieve proportion correct P_{det}. For detection, the model says "letter" whenever any letter feature is asserted, and otherwise says "blank."[23] Thus the probability of saying "blank" is the probability that every feature will fail to be asserted.

$$P(\text{``blank''}|\text{letter}) = 1 - P(\text{``letter''}|\text{letter}) = [1 - P_{\text{f}}(E_{\text{det}}/n)]^n \tag{B.4}$$

$$P(\text{``blank''}|\text{blank}) = 1 - P(\text{``letter''}|\text{blank}) = [1 - P_{\text{f}}(0)]^n \tag{B.5}$$

(Note the use of P, not P^*, so guessing is still included.) Our detection experiments measured the proportion P_{det} of correct responses averaged across letter and blank trials, which were presented with equal probability,

$$P_{\text{det}} = 0.5P(\text{``letter''}|\text{letter}) + 0.5P(\text{``blank''}|\text{blank}), \tag{B.6}$$

$$= 0.5(1 - [1 - P_{\text{f}}(E_{\text{det}}/n)]^n) + 0.5[1 - P_{\text{f}}(0)]^n, \tag{B.7}$$

$$= 0.5\left(1 + (1-\gamma_{\text{f}})^n(1 - \mathrm{e}^{-n(E_{\text{det}}/n\varepsilon_{\text{f}})^{\beta/2}})\right). \tag{B.8}$$

Solving for threshold E_{det},

$$E_{\text{det}} = \varepsilon_{\text{f}} n^{1-2/\beta}\left(-\ln\left(1 - \frac{2P_{\text{det}}-1}{(1-\gamma_{\text{f}})^n}\right)\right)^{2/\beta}, \tag{B.9}$$

$$= a\varepsilon_{\text{f}} n^{1-2/\beta}, \tag{B.10}$$

where $a = (-\ln(1 - \frac{2P_{\text{det}}-1}{(1-\gamma_{\text{f}})^n}))^{2/\beta}$ is a constant.[24]

This is our old friend probability summation in a new guise. The familiar result is usually expressed in terms of contrast (e.g., Robson & Graham, 1981). Let c_{f} and c_{det} represent letter contrasts at feature and letter detection thresholds. The letter energy is $n\varepsilon_{\text{f}}$ at the feature threshold and is $E_{\text{det}} = a\varepsilon_{\text{f}} n^{1-2/\beta}$ at the letter threshold. Energy is proportional to squared contrast, so $c^2_{\text{det}}/c^2_{\text{f}} = E_{\text{det}}/n\varepsilon_{\text{f}} = an^{-2/\beta}$, which, after taking the square root, is the familiar probability summation prediction: $c_{\text{det}}/c_{\text{f}} = a^{0.5}n^{-1/\beta}$.

B.3. Number of features detected

The detection model can tell us how many features are detected, on average, at any energy E. Let k be the average number of letter features detected (i.e. after correction for guessing) at energy E. This equals the number n of features present times their probability of detection $P^*_{\text{f}}()$, which is a function of the feature energy $\varepsilon = E/n$.

$$k = nP^*_{\text{f}}\left(\frac{E}{n}\right) \tag{B.11}$$

Substituting for $P^*_{\text{f}}()$ from the definition in Eq. (B.3),

$$= n\left(1 - \mathrm{e}^{-\left(\frac{E}{n\varepsilon_{\text{f}}}\right)^{\beta/2}}\right). \tag{B.12}$$

To reach our final goal it will be helpful to solve for energy,

$$E = \varepsilon_{\text{f}} n\left(-\ln\left(1 - \frac{k}{n}\right)\right)^{2/\beta}. \tag{B.13}$$

That is exact. If few of the letter's features are detected, i.e. $k \ll n$, then

$$E \simeq \varepsilon_{\text{f}} n^{1-2/\beta} k^{2/\beta}, \tag{B.14}$$

because $\ln(1-x) \simeq -x$ for small x. As a check, note that the energy required, Eq. (B.14), to detect one feature ($k = 1$) corresponds to the detection threshold Eq. (B.10).

B.4. Identification threshold

Identification threshold energy E_{id} is the letter energy required to achieve proportion correct P_{id}. Up to here Eq. (B.14) our only assumption has been probability summation, a well-established model of detection by human observers. Now we boldly suppose that at least one aspect of the detection model is still relevant when the observer is looking at the same stimulus (a letter of energy E in noise N) but doing a different task, identification instead of detection. We assume that feature detection is independent of task, so Eqs. (B.13) and (B.14) remain valid. Our model's efficiency is the ratio of ideal E_{ideal} and model Eq. (B.14) thresholds. Thus we fit our model's efficiency to our measured efficiency of identification as a function of complexity (Figs. 4 and 8a). (The fitting uses degrees of freedom in the model; there are no degrees of freedom in the ideal.) The fit assumes that the number n of features is proportional to complexity and that the number k of features detected at the identification threshold is independent of complexity (for which we present evidence in Section B.5). This produces a line in a log–log plot of efficiency E_{ideal}/E as a function of n. The line has two degrees of freedom: β controls its slope and ε_{f} controls its vertical position. The ideal identification threshold E_{ideal} for these alphabets has a log–log slope of -0.29. Our model's identification threshold Eq. (B.14) has log–log slope $1 - 2/\beta$. Efficiency is the ratio of the ideal and model thresholds, with log–log slope $-0.29 - (1 - 2/\beta)$. This is -0.91 and -0.76 in Figs. 4 and 8a, so the fitted value of β is 5.3 and 3.8, respectively.

We did not measure the steepness β of the psychometric functions of our observers, but it is well-known that the steepness for detection (all that is relevant here) depends on the observer's subjective criterion, ranging from a β of 2 for a lax criterion to 8 for a strict criterion (Nachmias, 1981; Pelli, 1985), so the values of 5.3 and 3.8 required to fit our data are perfectly reasonable. It is hard to measure β precisely. Fitting his frequency of seeing data from a yes-no grating detection task, Watson (1979) reported 52 estimates of β, with a mean 4.65 and standard deviation 1.4.

[23] More precisely, we assume a maximum likelihood decision rule. The criterion for saying "letter" would be higher, e.g., two feature assertions, if the false alarm rate were high, but in our experiments the false alarm rate was very low, about 0.01.

[24] In our experiments we took $P_{\text{det}} = 0.82$ and estimated the false alarm rate to be $P(\text{``letter''}|\text{blank}) = 1 - (1-\gamma_{\text{f}})^n \approx 0.01$. Taking β to be in the range 2 to 8, a is in the range 1.04 to 1.01, about 1.

Table B
Identification (64% correct) and detection (82% correct) energy thresholds $\log E^+$ in noise $N = 10^{-3.60}$ $\deg^2$ for ideal and three human observers

	Complexity	Overlap	Identification threshold					Detection threshold			
			Est. ideal	Ideal	DM	JS	LA	Ideal	DM	JS	LA
2 × 3 Checkers	28	0.55	−2.44	−2.34	−1.83			−3.07	−2.11		
bold Bookman	57	0.53	−2.45	−2.44	−1.54	−1.55	−1.38	−3.09	−1.95	−2.06	−1.98
Bookman	107	0.34	−2.60	−2.60	−1.59	−1.61	−1.49	−3.06	−2.01	−2.11	−1.92
Künstler	451	0.22	−2.67	−2.73	−1.18	−1.29	−1.11	−3.05	−1.54	−1.71	−1.66
Five-letter words	481	0.32	−2.61	−2.62	−1.05	−1.03	−0.87	−3.04	−1.36	−1.50	−1.40

Each point is an average of 10–20 runs of 40 trials each, for the human observers, and 2 runs of 2000 trials for the ideal observer. The standard error of $\log E^+$ for each human observer is about 0.02 for the identification thresholds and between 0.05 and 0.10 for the detection thresholds. For both identification and detection, the standard deviation of threshold $\log E^+$ across observers is 0.1. (E_0 was measured with 2–3 runs of 40 trials each on two observers; its contribution to $E^+ = E - E_0$ is negligible for both identification and detection thresholds.) Those who have read earlier drafts of this paper may note that Table B and Fig. 9 have changed. The ideal thresholds for detection were formerly for 64% correct and are now for 82% correct. This change does not affect any of our conclusions.

B.5. Ratio of thresholds

The predicted ratio of identification and detection thresholds Eqs. (B.13) and (B.10) is

$$\frac{E_{\mathrm{id}}}{E_{\mathrm{det}}} = \frac{(-n\ln(1-k/n))^{2/\beta}}{a}. \tag{B.15}$$

If we assume, as we did above, that $k \ll n$, then n drops out, leaving

$$\frac{E_{\mathrm{id}}}{E_{\mathrm{det}}} \simeq a^{-1}k^{2/\beta}. \tag{B.16}$$

In fact, k/n need not be terribly small; Eq. (B.16) is a good approximation for Eq. (B.15) if fewer than half of the features are detected, $k/n < 0.5$. Finally, we solve Eq. (B.16) for k, the number of features detected at identification threshold,

$$k \simeq \left(\frac{aE_{\mathrm{id}}}{E_{\mathrm{det}}}\right)^{\beta/2}. \tag{B.17}$$

For human observers, the identification efficiency slope (Fig. 9a) indicates that $\beta = 3.8$, and we find (Fig. 9b) a constant ratio, $E_{\mathrm{id}}/E_{\mathrm{det}} = 2.8$, independent of complexity. For our conditions, $a \approx 1$ (*Footnote 24*). Plugging these values into Eq. (B.17), we learn that the measured identification:detection threshold ratio is consistent with detection of 7 ± 2 features at the identification threshold, independent of complexity, for all these alphabets.[25,26]

[25] Eq. (B.16) is a good approximation to Eq. (B.15) if fewer than half of the features are detected, $k/n < 0.5$. Since k/n cannot exceed 1 for any alphabet, e.g., 2 × 3 Checkers, assuming that k is fixed and that n grows in proportion to complexity implies that more complex alphabets have lower k/n, and that all alphabets with at least twice the complexity of 2 × 3 Checkers have $k/n < 0.5$, so Eq. (B.16) applies. Thus the Fig. 12b report of constancy of $E_{\mathrm{id}}/E_{\mathrm{det}}$ over a 16-fold range of complexity (see Table B), over most of which we are confident that Eq. (B.16) is valid, suggests that k is fixed.

[26] To estimate a confidence interval for k we rewrite Eq. (B.17) as $\log k = 0.5\,\beta \log aE_{\mathrm{id}}/E_{\mathrm{det}}$. We have $a \approx 1$ and $\beta = 3.8 \pm 0.5 = \text{mean} \pm SE$ and $\log E_{\mathrm{id}}/E_{\mathrm{det}} = 0.44 \pm 0.03$. For the latter two parameters the standard error (SE) is a small fraction of the mean so $\log k = 0.5 \times 3.8 \times 0.44 \pm ((0.5 \times 3.8 \times 0.03)^2 + (0.5 \times 0.5 \times 0.44)^2)^{0.5} = 0.84 \pm 0.12$. Thus k is about $10^{0.84} = 7$. The 68% confidence interval for $\log k$ is [mean − SE, mean + SE] = [0.84 − 0.12, 0.84 + 0.12] = [0.72, 0.96], so $5 \leqslant k \leqslant 9$, i.e. $k = 7 \pm 2$.

Thus, applying the standard probability summation model of detection to letter identification allowed us to measure how many features are detected at the threshold for identification (see Table B).

References

Alexander, K. R., Xie, W., & Derlacki, D. J. (1994). Spatial-frequency characteristics of letter identification. *Journal of the Optical Society of America A, 11*, 2375–2382.

Arguin, M., & Bub, D. N. (1994). Functional mechanisms in pure alexia: evidence from letter processing. In M. J. Farah & G. Ratcliff (Eds.), *The neuropsychology of high-level vision*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Arnoult, M. D. (1960). Prediction of perceptual responses from structural characteristics of the stimulus. *Perceptual and Motor Skills, 11*, 261–268.

Attneave, F. (1955). Symmetry, information, and memory for patterns. *American Journal of Psychology, 68*, 183–193.

Attneave, F. (1957). Physical determinants of the judged complexity of shapes. *Journal of Experimental Psychology, 53*, 221–227.

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin, 53*, 452–471.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Barlow, H. B. (1978). The efficiency of detecting changes of density in random dot patterns. *Vision Research, 18*, 637–650.

Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(4), 509–522 [PDF].

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

Bouma, H. (1971). Visual recognition of isolated lower-case letters. *Vision Research, 11*, 450–474.

Braje, W. L., Tjan, B. S., & Legge, G. E. (1995). Human efficiency for recognizing and detecting low-pass filtered objects. *Vision Research, 35*, 2955–2966.

Briggs, R., & Hocevar, D. J. (1975). A new distinctive feature theory for upper case letters. *Journal of General Psychology, 93*, 87–93.

Brindley, G. S. (1954). The order of coincidence required for visual thresholds. *Proceedings of the Physical Society of London Series B, 67*, 673–676.

Brooks, L. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology, 22*, 349–368.

Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science, 214*, 93–94.

Burns, C. W. (1995). Identifying letters. Ph.D. thesis. Special Report ISR-S-33. Institute for Sensory Research, Syracuse University, Syracuse, NY 13244, USA.

Buschke, H. (1963). Retention in immediate memory span estimated without retrieval. *Science, 140*, 56–57.

Campbell, F. W., & Robson, J. G. (1968). The application of Fourier analysis to the visibility of gratings. *Journal of Physiology, 197*, 551–566.

Cattell, J. M. (1896). The time taken up by cerebral operations. Part 3. *Mind, 11*, 377–392.

Cavanagh, P. Correlational and structural models of alphabetic confusion. Unpublished manuscript.

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

Chung, S. T. L., Levi, D. M., & Tjan, B. S. (2005). Learning letter identification in peripheral vision. *Vision Research, 45*(11), 1399–1412.

Cortázar, J. (1966). *Hopscotch*. New York: Pantheon Books, p. 36.

Davis, E. T., Kramer, P., & Graham, N. (1983). Uncertainty about spatial frequency, spatial position, or contrast of visual patterns. *Perception & Psychophysics, 33*, 20–28.

DeFrancis, J. (1976). *Character text for beginning Chinese* (2nd ed.). New Haven, CT: Yale University Press.

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

Edelman, S., Flash, T., & Ullman, S. (1990). Reading cursive handwriting by alignment of letter prototypes. *International Journal of Computer Vision, 5*(3), 303–331.

Ellis, W. D. (1938). *A source book of Gestalt psychology*. New York: Harcourt, Brace, and Co.

Erdmann, B., & Dodge, R. (1898). *Psychologische Untersuchungen über das Lesen auf experimenteller Grundlage*. Halle: Niemeyer.

Eriksen, C. W. (1980). The use of a visual mask may seriously confound your experiment. *Perception & Psychophysics, 28*(1), 89–92.

Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: effects of stimulus orientation, range and of feedback. *Vision Research, 33*, 397–412.

Fairbank, A. (1952). *A book of scripts*. Hammondsworth, Middlesex, England: Penguin Books Ltd.

Farah, M. J. (1991). Patterns of co-occurrence among the associative agnosias: implications for visual object representation. *Cognitive Neuropsychology, 8*, 1–19.

Farah, M. J. (1994). Specialization within visual object recognition: clues from prosopagnosia and alexia. In M. J. Farah & G. Ratcliff (Eds.), *The neuropsychology of high-level vision* (pp. 133–146). Hillsdale, NJ: Lawrence Erlbaum Associates.

Farell, B., & Pelli, D. G. (1993). Can we attend to large and small at the same time? *Vision Research, 33*, 2757–2772.

Fehrer, E. V. (1935). An investigation of the learning of visually perceived forms. *American Journal of Psychology, 47*, 187–221.

Garner, W. R. (1970). Good patterns have few alternatives. *American Scientist, 58*(1), 34–42.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review, 96*(2), 267–314.

Geisler, W., & Murray, R. (2003). Cognitive neuroscience: Practice doesn't make perfect. *Nature, 423*(6941), 696–697.

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.

Geyer, L. H. (1977). Recognition and confusion of the lowercase alphabet. *Perception & Psychophysics, 22*, 487–490.

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

Gibson, E. J., Gibson, J. J., Pick, A. D., & Osser, H. (1962). A developmental study of the discrimination of letter-like forms. *Journal of Comparative and Physiological Psychology, 55*, 897–906.

Gibson, E. J., Osser, H., Schiff, W., & Smith, J. (1963). An analysis of critical features of letters tested by a confusion matrix. In *A basic research program on reading*. Cooperative Research Project No. 639, Cornell University, Ithaca, NY.

Gilbert, C. D., & Wiesel, T. N. (1992). Receptive field dynamics in adult primary visual cortex. *Nature, 356*, 150–152.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999a). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research, 39*(21), 3537–3560.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999b). Signal but not noise changes with perceptual learning. *Nature, 402*(6758), 176–178.

Graham, N. (1977). Visual detection of aperiodic spatial stimuli by probability summation among narrowband channels. *Vision Research, 17*, 637–652.

Graham, N. (1980). Spatial-frequency channels in human vision: detecting edges without edge detectors. In C. Harris (Ed.), *Visual coding and adaptability* (pp. 215–252). Hillsdale, NJ: Erlbaum.

Graham, N. V. S. (1989). *Visual pattern analyzers*. Oxford: Oxford University Press.

Hadley, J. A., & Healy, A. F. (1991). When are reading units larger than the letter? Refinement of the unitization reading model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 1062–1073.

Hanley, J. R., & Kay, J. (1992). Does letter-by-letter reading involve the spelling system? *Neuropsychologia, 30*(3), 237–256.

Höffding, H. (1891). *Outlines of psychology*. New York: Macmillan.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences of USA, 88*(11), 4966–4970.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Research, 34*, 885–912.

Kinsbourne, M., & Warrington, E. K. (1962). A disorder of simultaneous form perception. *Brain, 85*, 461–486.

Kinsbourne, M., & Warrington, E. K. (1963). The localizing significance of limited simultaneous visual form perception. *Brain, 86*, 697–702.

Kolers, P. A. (1975). Memorial consequences of automatized encoding. *Journal of Experimental Psychology–Human Learning and Memory, 1*, 689–701.

Krauskopf, J., Duryea, R. A., & Bitterman, M. E. (1954). Thresholds for visual form: further experiments. *American Journal of Psychology, 67*, 427–440.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Legge, G. E., Mansfield, J. S., & Chung, S. T. (2001). Psychophysics of reading. XX. Linking letter recognition to reading speed in central and peripheral vision. *Vision Research, 41*(6), 725–743.

Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading–I. Normal vision. *Vision Research, 25*, 239–252.

Lennie, P. (1993). Roles of M and P pathways. In R. Shapley & D. M.-K. Lam (Eds.), *Contrast sensitivity* (pp. 201–213). Cambridge, MA: MIT Press.

Loomis, J. M. (1982). Analysis of tactile and visual confusion matrices. *Perception & Psychophysics, 31*, 41–52.

Loomis, J. M. (1990). A model of character recognition and legibility. *Journal of Experimental Psychology—Human Perception and Performance, 16*(1), 106–120.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. B. Bush, & E. Galanter (Eds.). *Handbook of mathematical psychology* (Vol. 1). New York: Wiley.

Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research, 42*(9), 1165–1184.

Majaj, N. J., Liang, Y. X., Martelli, M., Berger, T. D., & Pelli, D. G. (2003). The channel for reading. *Journal of Vision, 3*(9), 813, Available from http://www.journalofvision.org/3/9/813/.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman & Company.

Martelli, M., Baweja, G., Mishra, A., Chen, I., Fox, J., Majaj, N. J., et al. (2002). How efficiency for identifying objects improves with age. *Perception*, ECVP Supplement. Available from: <http://www.perceptionweb.com/perception/ecvp02/0370.html>.

Martelli, M., Baweja, G., Mishra, A., Majaj, N. J., & Pelli, D. G. (in preparation). Learning to recognize objects: efficiency for faces, letters, and pictograms. *Vision Research*.

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences, 7*, 293–299.

Miller, G. A. (1956a). The magic number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Miller, G. A. (1956b). Human memory and the storage of information. *IRE Transactions on Information Theory, IT-2*(3), 129–137.

Morgan, M. J., Fitch, M. D., Holman, J. D., & Lea, S. E. G. (1976). Pigeons learn the concept of an 'A'. *Perception, 5*, 57–66.

Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision Research, 21*, 215–223.

Nakayama, K. (1990). The iconic bottleneck and the tenuous link between early visual processing and perception. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 411–422). Cambridge: Cambridge University Press.

National Academy of Sciences-National Research Committee on Vision (1980). Report of working group 39: recommended standard procedures for the clinical measurement and specification of visual acuity. *Advances in Ophthalmology*, *41*, 103–148.

Näsänen, R., Kukkonen, H., & Rovamo, J. (1993). Spatial integration of band-pass filtered patterns in noise. *Vision Research, 33*, 903–911.

Näsänen, R., Kukkonen, H., & Rovamo, J. (1994). Relationship between spatial integration and spatial spread of contrast energy in detection. *Vision Research, 34*, 949–954.

Näsänen, R. E., Kukkonen, H. T., & Rovamo, J. M. (1995). A window model for spatial integration in human pattern discrimination. *Investigative Ophthalmology & Visual Science, 36*, 1855–1862.

Nuttall, A. H. (1962). Error probabilities for equicorrelated M-ary signals under phase coherent and phase incoherent reception. *IRE Transactions, IT-8*(4), 305–314.

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience, 13*, 4700–4719.

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

Patterson, K., & Kay, J. (1982). Letter-by-letter reading: psychological descriptions of a neurological syndrome. *Quarterly Journal of Experimental Psychology, 34A*, 411–441.

Pavel, M., Sperling, G., Riedl, T., & Vanderbeek, A. (1987). Limits of visual communication: the effect of signal-to-noise ratio on the intelligibility of American Sign Language. *Journal of the Optical Society of America (A), 4*(12), 2355–2365.

Pelli, D. G. (1981). Effects of visual noise. Ph.D. Thesis. Cambridge University, Cambridge, England. Available from http://psych.nyu.edu/pelli/pub/pelli1981thesis.pdf [PDF].

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: Cambridge University Press.

Pelli, D. G., & Farell, B. (1992). Visual memory. *Perception, 21*(Supplement 2), 71.

Pelli, D. G., & Farell, B. (1999). Why use notice? *Journal of the Optical Society of America A-Optics Image Science and Vision, 16*(3), 647–653.

Pelli, D. G., & Hoepner, J. A. (1989). Letters in noise: a visual test chart that ''bypasses'' the optics. *Noninvasive assessment of the visual system, 1989 technical digest series* (Vol. 7, pp. 103–106). Washington, DC: Optical Society of America.

Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature, 423*(6941), 752–756.

Pelli, D. G., Levi, D. M., & Chung, S. T. (2004). Using visual noise to characterize amblyopic letter identification. *Journal of Vision, 4*(10), 904–920, Available from http://journalofvision.org/4/10/6/.

Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research, 31*, 1337–1350.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). Theory of signal detectability. *Transactions of the IRE PGIT, 4*, 171–212.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193–198.

Polk, T. A., & Farah, M. J. (1999). The neural development and organization of letter recognition: evidence from functional neuroimaging, computational modeling, and behavioral studies. *Proceedings of the National Academy of Science, 95*, 847–852.

Polk, A. T., Stallcup, M., Aquirre, G. K., Alsop, D. C., D'Esposito, M., Detre, J. A., et al. (2002). Neural specialization for letter recognition. *Journal of Neuroscience, 14*, 145–159.

Pons, T. P., Garraghty, P. E., & Mishkin, M. (1988). Lesion-induced plasticity in the second somatosensory cortex of adult macaques. *Proceedings of the National Academy of Science, 85*, 5281–5297.

Quick, R. F. Jr., (1974). A vector-magnitude model of contrast detection. *Kybernetik, 16*(2), 65–67.

Raghavan, M. (1995). Sources of visual noise. Ph.D. Thesis. Syracuse University, Syracuse, NY.

Raghavan, M., & Pelli, D. G. (in preparation). Photon and cortical noise limit what we see. *Nature*.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.

Razel, M. (1974). The path of visual scanning and the order of search in short-term memory. Unpublished Ph.D. dissertation, Psychology Dept., New York University, New York, NY.

Ress, D., & Heeger, D. J. (2003). Neuronal correlates of perception in early visual cortex. *Nature Neuroscience, 6*(4), 414–420.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3*(Suppl.), 1199–1204.

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research, 21*(3), 409–418.

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382–439.

Rovamo, J., Franssila, R., & Näsänen, R. (1992). Contrast sensitivity as a function of spatial frequency, viewing distance and eccentricity with and without spatial noise. *Vision Research, 32*, 631–637.

Rovamo, J., Virsu, V., & Näsänen, R. (1978). Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature, 271*, 54–56.

Schultz, D. W., & Eriksen, C. W. (1977). Do noise masks terminate target processing? *Memory & Cognition, 5*(1), 90–96.

Shiu, L.-P., & Pashler, H. (1991). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Investigative Ophthalmology & Visual Science, 32*, 1041.

Simard, P. Y., LeCun, Y. A., Denker, J. S., & Victorri, B. (2001). Transformation invariance in pattern recognition—tangent distance and tangent propagation. *International Journal of Imaging Systems and Technology, 11*(3), 181–194.

Smithson, H., & Mollon, J. (2006). Do masks terminate the icon? *Quarterly Journal of Experimental Psychology, 59*(1), 150–160.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature, 369*, 395–397.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*(11, Whole No. 498), 1–29.

Steingrimsson, R., Majaj, N. J., & Pelli, D. G. (2003). Where are letters processed and learned? Neural specialization for letter processing under different learning conditions. *Perception*, *32* supplement. Available from http://psych.nyu.edu/pelli/posters.html#2003.

Suchow, J. W., & Pelli, D. G. (2005). Learning to identify letters: generalization in high-level perceptual learning. *Journal of Vision 5*(8), 712a. Available from http://www.journalofvision.org/5/8/712/ [PDF].

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology, 66*, 170–189.

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

Thomas, J. P. (1985). Detection and identification: how are they related? *Journal of the Optical Society of America (A), 2*, 1457–1467.

Thompson, D'A. W. (1917). *On growth and form*. Cambridge University Press.

Thompson, R. F., Berger, T. W., & Madden, J. (1983). Cellular processes of learning and memory in the mammalian CNS. *Annual Review of Neuroscience, 6*, 447–491.

Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. J. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research, 35*, 3053–3070.

Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research, 32*, 1349–1357.

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

Townsend, J. T. (1971b). Alphabetic confusion: a test of models for individuals. *Perception & Psychophysics, 9*, 449–454.

Townsend, J. T., & Ashby, F. G. (1982). Experimental test of contemporary mathematical models of visual letter recognition. *Journal of Experimental Psychology—Human Perception and Performance, 8*, 834–854.

Townsend, J. T., Hu, G. G., & Evans, R. J. (1984). Modeling feature perception in brief displays with evidence for positive interdependencies. *Perception & Psychophysics, 36*, 35–49.

Ullman, S. (1989). Aligning pictorial descriptions: an approach to object recognition. *Cognition, 32*, 193–254.

Ullman, S. (1995). The visual analysis of shape and form. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 339–350). Cambridge, MA: MIT Press.

Valvo, A. (1971). *Sight restoration after long-term blindness: The problems and behavior patterns of visual rehabilitation*. New York: American Foundation for the Blind [The HS quote also appears in Wandell (1995,p. 390).

Van Trees, H. L. (1968). *Detection, estimation, and modulation theory*. New York: Wiley.

Verghese, P., & Pelli, D. G. (1992). The information capacity of visual attention. *Vision Research, 32*, 983–995.

Wandell, B. A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer Associates.

Warrington, E. K., & Shallice, T. (1980). Word-form dyslexia. *Brain, 103*, 99–112.

Watson, A. B. (1979). Probability summation over time. *Vision Research, 19*, 515–522.

Watson, A. B., & Ahumada, A. J. Jr., (2005). A standard model for foveal detection of spatial contrast. *Journal of Vision, 5*(9), 717–740.

Watson, A. B., & Fitzhugh, A. E. (1989). Modelling character legibility. *Society for Information Display Digest of Technical Papers, 20*, 360–363.

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics, 33*, 113–120.

Watson, A. B., & Robson, J. G. (1981). Discrimination at threshold: labelled detectors in human vision. *Vision Research, 21*, 1115–1122.

Watson, A. B., Thompson, P. G., Murphy, B. J., & Nachmias, J. (1980). Summation and discrimination of gratings moving in opposite directions. *Vision Research, 20*, 341–347.

Weinstein, M. (1955). Stimulus complexity and the recognition of visual patterns. Doctoral dissertation. Ohio State University.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II. [Laws of organization in perceptual forms, translated in Ellis (1938)]. *Psychologische Forschung, 4*, 301–350.

Woodworth, R. S. (1938). *Experimental psychology*. New York: Holt.

Wulf, F. (1922). Über die Veränderung von Vorstellungen (Gedächtnis und Gestalt). *Psychologische Forschung, 1*, 333–373 [Tendencies in figural variation, translated in Ellis (1938)].