

Using geometric moments to explain human letter recognition near the acuity limit

Lei Liu

School of Optometry, University of Alabama at Birmingham,
Birmingham, AL, USA



Stanley A. Klein

School of Optometry, University of California,
Berkeley, CA, USA



Feng Xue

EENT Hospital, Fudan University, Shanghai, China



Jun-Yun Zhang

State Key Laboratory of Cognitive Neuroscience and Learning,
Beijing Normal University, Beijing, China



Cong Yu

State Key Laboratory of Cognitive Neuroscience and Learning,
Beijing Normal University, Beijing, China



When the size of a letter stimulus is near the visual acuity limit of a human subject, details of the stimulus become unavailable due to ocular optical and neural filtering. In this study we tested the hypothesis that letter recognition near the acuity limit is dependent on more global features, which could be parsimoniously described by a few easy-to-visualize and perceptually meaningful low-order geometric moments (i.e., the ink area, variance, skewness, and kurtosis). We constructed confusion matrices from a large set of data (approximately 110,000 trials) for recognition of English letters and Chinese characters of various spatial complexities near their acuity limits. We found that a major portion of letter confusions reported by human subjects could be accounted for by a geometric moment model, in which letter confusions were quantified in a space defined by low-order geometric moments. This geometric moment model is universally applicable to recognition of visual patterns of various complexities near their acuity limits.

Keywords: visual acuity, computational modeling, object recognition

Citation: Liu, L., Klein, S. A., Xue, F., Zhang, J., & Yu, C. (2009). Using geometric moments to explain human letter recognition near the acuity limit. *Journal of Vision*, 9(1):26, 1–18, <http://journalofvision.org/9/1/26/>, doi:10.1167/9.1.26.

Introduction

Visual acuity, often defined as the minimal angle of resolution, is a most important visual function. While the primary visual acuity optotype is Landolt C (International Organization for Standardization (ISO), 1986; National Academy of Science National Research Council (NAS-NRC), 1980), visual acuity is measured almost exclusively with letters in literary adults. In the US, the most popular visual acuity assessment is the ETDRS visual acuity chart (Ferris, Kassoff, Bresnick, & Bailey, 1982), which is made of 10 uppercase English letters of a specific typeface (Sloan letters). Subjects have to read these letters of progressively smaller sizes until no reliable recognition is possible. Therefore, visual acuity assessment is a pattern recognition process under a condition where visual information is severely degraded by ocular optics. How does a human subject recognize these letters?

In vision science, studies of letter recognition have focused on spatial frequency channels. Recent studies of contrast thresholds for recognizing spatially filtered and/or

noise masked large letter stimuli indicated that a narrow band (1–2 octaves) centered at a relatively low frequency (1–2 cycles/letter) is critical for letter recognition (Ginsburg, 1977; Parish & Sperling, 1991; Solomon & Pelli, 1994). Bondarko and Danilova (1997) calculated the Fourier spectra of Landolt C and Snellen E, and concluded that the primary information that could indicate the orientations of these optotypes at the resolution limit also resides in a narrow band centered at 1.4–1.7 cycles/letter. The low frequency channels identified in these studies cannot transmit the visual information of fine details of the stimulus, and thus suggest an important role of global characteristics in letter recognition.

The identification of the critical band is a big step forward in our understanding of letter recognition near the acuity limit, but it falls short of explaining how information within the critical band is used to differentiate members of the stimulus set. For example, Bondarko and Danilova (1997) showed that spatial frequency amplitudes could be used to determine whether the gap of a Landolt C was in the horizontal direction, but amplitudes could not determine whether the gap was to the left or to the right.

Although phase information of Fourier components is in theory capable of differentiating mirror images, there is little empirical data or theoretical speculations on how human subjects utilize phase information in this situation.

As a working model, spatial template matching in various forms has been used to explain pattern recognition (Chung, Legge, & Tjan, 2002; Gold, Bennett, & Sekuler, 1999; Tjan, Braje, Legge, & Kersten, 1995). While such models have been powerful tools in testing hypotheses about visual channels, the notion that a set of fixed spatial templates exist somewhere in the visual system, ready to be matched with visual inputs, may not explain why human subjects can effortlessly recognize patterns they have never encountered, for example, a new font face.

Cognitive science has a tradition of characterizing letter stimuli as collections of features and explaining human letter recognition as a process of matching feature sets extracted from stimuli with those stored in memory (Neisser, 1967). A recent study argued that human letter recognition has to be performed by features, because a holistic letter channel or template matching would have resulted in much higher efficiency than observed (Pelli, Burns, Farell, & Moore-Page, 2006). At the basis of any feature analysis model is the set of perceivable attributes of letters that are crucial to the recognition task at hand. Feature sets of existing models for recognition of letters or letter-like stimuli consist mainly of natural graphic features, such as horizontal, vertical and oblique strokes, dots, and gaps or openings (Geyer & DeWald, 1973; Gibson, 1969; Laughery, 1969). While such models enjoy different degrees of successes in explaining human errors in recognizing briefly presented, large and high contrast letters, their applications to recognition of letters whose sizes are close to the acuity limit is questionable, because individual strokes of these letters are severely distorted, merge with each other, or simply vanish due to the low-pass filtering of the ocular optics.

It is generally agreed that feature analysis models, compared to template matching models, are more robust and have better potential to generalize. This is because in template matching, every physical pixel matters, regardless of its perceptual and cognitive significance. Therefore, template matching is vulnerable to translation, scaling, rotation, distortion, and noise. Feature analysis, on the other hand, extracts generic features with perceptual and cognitive significances, and the matching is done at the feature level. On the other hand, whether the advantage of a feature analysis model can materialize depends on the selection of a set of appropriate features. We propose a feature analysis model that uses moments of the stimulus image as its feature set.

Moments are commonly used in statistics and computer science to specify a random distribution in a hierarchical way. Lower order moments describe the global characteristics of the distribution, whereas higher order moments are associated with details. Unlike the feature sets of previous letter recognition models, moments decompose

2-D images into generic features that do not depend on a particular set of stimuli or an investigator's personal preference, and thus should be applicable to any letter/character set. Moment "shape descriptors" have been used as generic features for machine pattern recognition of letters and characters of all types of fonts, including handwritten ones (Mukundan & Ramakrishnan, 1998). Among various moments developed in computer science for image analysis, geometric moments (GMs) seem to be most relevant to human pattern recognition because, as the name indicated, they are directly related to geometric features of random patterns. For an N_x -by- N_y pixel spatial pattern $f(x, y)$, the double-sequence quantities defined in Equation 1 are the order $(p + q)$ GMs of the pattern

$$M_{p,q} = \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} x^p y^q f(x, y), \quad (1)$$

and $\{M_{p,q}\}$ ($p, q = 0, 1, 2, \dots$) are sufficient to uniquely specify a finite spatial function $f(x, y)$ (Hu, 1962). $M_{p,q}$ is the inner product of the geometric moment basis function $x^p y^q$ and the spatial pattern $f(x, y)$.

The direct relationship between low-order GMs and the geometric properties of 2-D images can be appreciated from the basis functions, $x^p y^q$. In GM analysis, a 2-D image is decomposed into basis functions $x^p y^q$, which are luminance distributions confined within the boundary of the image, and $M_{p,q}$ are scalars that represent the weights of the basis functions (the contributions of basis functions to the 2-D image). Figure 1b shows basis functions of some low-order GMs. Notice that low-order basis functions are simple luminance distributions or features, and the GMs are the weights or contributions of these features. Also notice that the basis functions of low-order GMs are low spatial frequency features in a sense that the luminance changes in these features are gradual. Figure 1b also shows a letter "E," and the result of reconstruction using GMs up to the 5th order. While this 5th-order reconstruction is missing many details, it contains adequate information for many purposes, for example, for identifying the opening of tumbling E's in an acuity test. When more moments are used in the reconstruction, details such as the center bar of the "E" will become evident. It is also worth mentioning that while pure GMs, such as $M_{0,q}$ and $M_{p,0}$, are 1-D basis functions, comparable to Gabors, mixed GMs, such as $M_{p,q}$ ($p \neq 0$ and $q \neq 0$), are true 2-D basis functions, comparable to plaids.

We hypothesize that human observers use global characteristics of stimulus images to recognize patterns whose sizes are close to the acuity limit and that these global characteristics can be adequately and parsimoniously described by low-order GMs. The low-order GMs signify and retain pertinent information of x - and y -distributions of ink on a rectangular domain that is too small to engage receptive fields with a large variety of envelopes. They are appropriate features to approximate

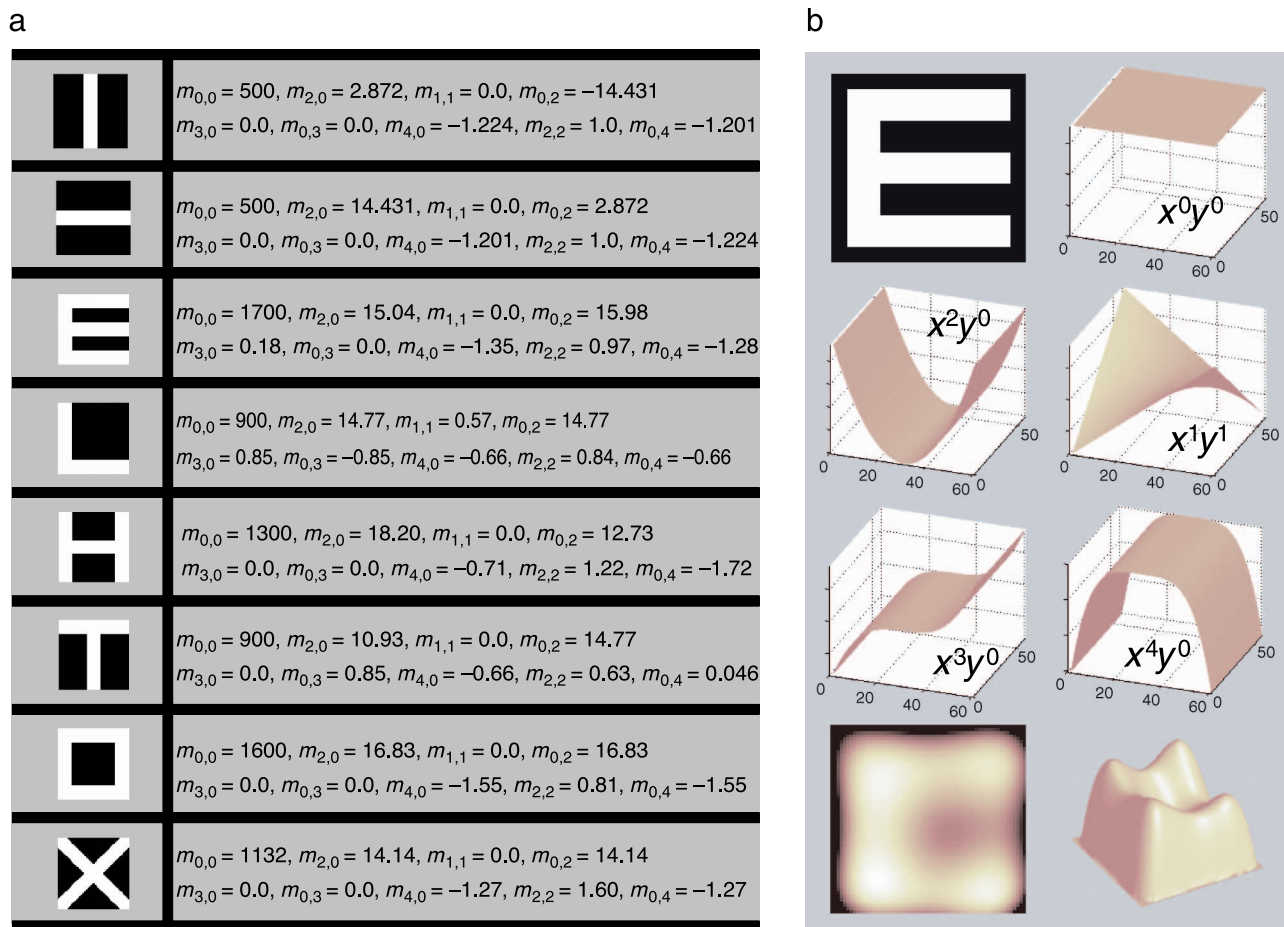


Figure 1. (a) Selected centralized GMs, $m_{p,q}$, of 50×50 pixels black and white images. The white pixels are considered “ink.” Moments $M_{p,q}$ defined in Equation 1 are related to $m_{p,q}$ through a series of affine transforms (Equations 2–5). (b) A stimulus “E”, the basis functions of selected GMs, and the result of reconstruction using GMs up to the 5th order. Basis functions x^0y^2 , x^0y^3 and x^0y^4 have the same shapes as x^2y^0 , x^3y^0 and x^4y^0 but are aligned along the y -axis. The result of reconstruction is shown both in gray map and in 3-D distribution.

global features of barely resolvable Sloan letters and Chinese characters because these stimuli have a fixed orientation and a predominantly orthogonal structure. For example, the observation that a human observer can determine the orientation of a Landolt C or a Snellen E by judging which side of the stimulus has less ink could be readily explained by the difference in the 3rd-order GMs (skewness) of the ink distribution. The contribution of height-to-width quotient to letter recognition demonstrated by Bouma (1971) could be associated with the ratio of 2nd-order GMs in y - and x -directions. Specifically, we propose a feature analysis model for letter recognition near the acuity limit, in which the feature set consists of global characteristics of stimulus patterns quantified by low-order GMs, and recognition is achieved by comparing GM compositions of letters. Because recognition confusions occur when crucial features are shared, we used the model to analyze the patterns of letter confusions obtained from identifying letters that were slightly above the acuity limit. In order to demonstrate that this model could be applied to a wide range of

over-learned patterns, we analyzed English letters and 6 groups of Chinese characters that spanned a wide range of spatial complexities. We were able to demonstrate that the Euclidean distances in a low-order GMs space could explain a large portion of human errors made in recognizing letters near the acuity limit.

Methods

Psychophysics and confusion matrix

Data were part of a large data set (approximately 110,000 trials) that was collected in a visual acuity study of English letters and Chinese characters (Zhang, Zhang, Xue, Liu, & Yu, 2007). Unlike the published visual acuity study that focused on the relationship between viewing conditions and correct responses, the current study focused on response errors or confusions.

		Acuity size (arcmin)	Stroke freq. (strokes/char)
Sloan	C D H K N O R S V Z	3.41	2.02
CC1	个 文 么 山 少 之 十 力 人 天	4.68	2.22
CC2	州 月 占 只 无 区 以 世 风 任	4.94	3.22
CC3	条 名 抓 电 各 好 取 医 非 多	5.44	3.80
CC4	益 售 很 资 建 做 学 深 张 事	5.69	4.40
CC5	势 数 想 续 验 率 常 群 策 领	5.97	4.73
CC6	需 感 题 路 德 增 解 就 赛 整	5.99	5.52

Figure 2. The seven groups of stimuli, their average acuity sizes, and average stroke frequencies. Acuity size was defined as the character size that resulted in a 66.9% correct recognition.

The test stimuli consisted of one group of Sloan letters and six groups of Chinese characters (CC1–CC6), 10 letters or characters in each group (Figure 2). The Chinese character groups differed by the number of strokes (2–4, 5–6, 8–9, 11–12, 13–15, and 16–18 strokes per character, respectively). They were selected from the 500 most frequently used Chinese characters according to an official character frequency table and were pre-screened for similarity in legibility based on the intermediate pair-wise Euclidean distances of their bitmaps. Different stimulus groups were tested separately. The observers' task was to identify the stimulus from a list of 10 letters or characters of the tested group. Figure 2 lists the acuity sizes and stroke frequency (Majaj, Pelli, Kurshan, & Palomares, 2002) for each stimulus group. Stimulus sizes for these stimulus groups were set at 0.1 log unit above the acuity sizes shown in Figure 2. More details regarding data collection can be found in Zhang et al. (2007).

One confusion matrix (CM) was constructed for each stimulus group based on six observers' pooled responses. Like most studies of human recognition errors (Geyer & DeWald, 1973; Townsend, 1971), only data from stimulus sizes that generated group average correct rates between 54% and 60% were used for CM construction. Since each experiment involves 10 letters, the CM was a 10-by-10 square matrix. The (i, j) cell contained the probability of the i th stimulus letter being reported as the j th letter, $c_{i,j}$. The diagonal line entries $c_{i,i}$ were probabilities of correct responses, and the off-diagonal line entries ($i \neq j$) were errors, or confusions. Because the i th column of a CM contained all the responses to the i th stimulus, the sum of the column was equal to 1.0.

A GM-based feature analysis model

Central geometric moments

GMs defined in Equation 1 are "raw" moments that are sensitive to pattern location and size. It is helpful to transform the variables x and y so that the moments are

location and size invariant (Alt, 1962). The coordinates of the center of gravity of the pattern are

$$\bar{x} = \frac{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} xf(x,y)/L}{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} f(x,y)/L} \quad \text{and} \quad (2)$$

$$\bar{y} = \frac{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} yf(x,y)/L}{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} f(x,y)/L},$$

where $L = \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} f(x,y)$ is the mean luminance. The variances of the pattern in x - and y -directions are defined as

$$\sigma_x = \sqrt{\frac{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} (x - \bar{x})^2 f(x,y)/L}{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} f(x,y)/L}} \quad \text{and} \quad (3)$$

$$\sigma_y = \sqrt{\frac{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} (y - \bar{y})^2 f(x,y)/L}{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} f(x,y)/L}},$$

and they are used to normalize the coordinates

$$x^* = (x - \bar{x})/\sigma_x \quad \text{and} \quad y^* = (y - \bar{y})/\sigma_y. \quad (4)$$

The new moments based on x^* and y^* were used in our study:

$$m_{p,q} = \frac{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} x^{*p} y^{*q} f(x^*, y^*)/L}{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} f(x^*, y^*)/L}. \quad (5)$$

The following adjustments were made. The 0th-order moments $m_{0,0}$ defined in Equation 5 is equal to 1.0. Because the mean luminance of the pattern might provide important information for pattern recognition, we set $m_{0,0}$ equal to L . The 1st-order moments, $m_{1,0}$ and $m_{0,1}$, are 0 for all patterns because the coordinate was shifted to the

centroid of the pattern. They were not used in the following simulation because the task was to recognize a single letter. Information about centroids of multiple letters is important in tasks such as word recognition or reading. The 2nd-order moments $m_{2,0}$ and $m_{0,2}$ defined in Equation 5 are equal to 1.0. Because the width/height ratio might be informative in letter recognition, we set $m_{2,0}$ and $m_{0,2}$ to σ_x and σ_y , defined in Equation 3. We subtracted 3 from the kurtosis in the x - and y -directions ($m_{4,0}$ and $m_{0,4}$) defined in Equation 5 to comply with the common practice that when the distribution along the x - or y -axis was a Gaussian, the kurtosis was zero.

The centralized GMs of some simple geometric shapes were shown in Figure 1a to illustrate that low-order GMs were directly associated with perceivable global properties of 2-D images. For binary images ($f(x, y) = 0$ or 1) in Figure 1a, the 0th-order GM is the number of ink pixels, which can be perceived as the general lightness or darkness of a letter. In Figure 1a, patterns with more strokes have a larger $m_{0,0}$ than those with fewer strokes and appear darker if individual strokes cannot be distinguished. Pure 2nd-order moments (moments that are 0-order in one direction) $m_{2,0}$ and $m_{0,2}$ are dispersion of ink distribution. For the vertical bar in Figure 1a (height/width = 5/1), $m_{2,0} < m_{0,2}$ (2.872 vs. 14.431). For the horizontal bar, $m_{2,0} > m_{0,2}$ (14.431 vs. 2.872). In fact, when only GMs up to the second order are considered, the original image is completely equivalent to a constant irradiance ellipse, whose size, orientation, aspect ratio, and center are completely specified by the GMs (Teague, 1980). Pure 3rd-order moments $m_{3,0}$ and $m_{0,3}$ represent the skewness of ink distributions on the x - and y -directions. For the letter “E,” $m_{3,0}$ has a positive value (0.18), but $m_{0,3}$ is zero. Perceptually, the letter appears darker on the left side (skewed to the left), but appears symmetric in the vertical direction. The distribution of letter “L” is heavily skewed to the left in the horizontal direction and to the bottom in the vertical direction ($m_{3,0} = 0.85$, $m_{0,3} = -0.85$). Notice that the skewness is both directional specific ($m_{2,0}$ vs. $m_{0,2}$) and positional specific (positive vs. negative). Pure 4th-order moments $m_{4,0}$ and $m_{0,4}$ specify whether ink distributions on the x - and y -directions are more peaked or flat-topped than a Gaussian. In letter “H,” the distribution along x -axis is the lowest in the middle, due to the two vertical strokes, and thus $m_{4,0}$ has a large negative number (-1.72). The distribution of letter “T,” on the other hand, has a strong peak in the middle, and $m_{4,0}$ has a positive number (0.046). Mixed moments indicate the clustering of ink pixels around oblique axes. For example, $m_{2,2}$ is large for letter “X,” but small for Chinese character □ (1.60 vs. 0.81).

Feature analysis models using GMs as feature sets

This model used GMs extracted from stimulus bitmaps as features. To simplify the notation, we used one index to denote moments involved in the model as $\mu_1, \mu_2, \dots, \mu_n$.

For example, if a model used ink area, x - and y -direction skewness and x - and y -direction kurtosis, then these moments were denoted as $\mu_1, \mu_2, \mu_3, \mu_4$, and μ_5 . In a model involving n moments, each stimulus was represented by a vector $\{\mu_1, \mu_2, \dots, \mu_n\}$, in an n -dimensional moment space. The difference between the i th and j th stimuli, $d_{i,j}$, was measured by the distance between these stimuli in the n -dimensional moment space. To reflect different contributions of moments to the recognition of a set of stimuli, a weighting, w_k , was given to each moment dimension:

$$d_{i,j} = \sum_{k=1}^n \left[w_k (\mu_{ik} - \mu_{jk}) \right]^2. \quad (6)$$

This weighted Euclidean distance (Getty, Swets, Swets, & Green, 1979; Mukundan & Ramakrishnan, 1998) was a pure physical measure of pair-wise difference between stimuli. For a letter recognition experiment that involves k letters, $\{d_{i,j}\}$ forms a k -by- k symmetric matrix with zeros on the diagonal. To simulate human recognition performance, $d_{i,j}$ needed to be converted to a measure of perceptual similarity. There are both empirical evidence and theoretical justification (Shepard, 1987) that the conversion should be monotonic and should take an exponential shape:

$$s_{i,j} = e^{-\tau d_{i,j}} \quad (\tau > 0). \quad (7)$$

The quantity $s_{i,j}$ was called “measure of stimulus generalization” (Shepard, 1987) or “similarity scale” (Luce, 1963b) and was used in several previous studies of human letter recognition (Getty et al., 1979; Keren & Baggen, 1981; Loomis, 1990). For a set of k letters, $\{s_{i,j}\}$ was a k -by- k similarity symmetric matrix. Because $d_{i,i} = 0$, similarity between a stimulus and itself, $s_{i,i} = 1.0$, which sets the scale for similarity. The fact that all diagonal entries of similarity matrix $\{s_{i,j}\}$ equal 1.0 simply indicates that all stimuli had the same degree of similarity to themselves. The free parameter in Equation 7, τ , determined how fast similarity fell from perfection. Intuitively, the probability of confusion between the i th and j th letter should be proportional to perceptual similarity $s_{i,j}$, but $s_{i,j}$ was not the empirical probability of the i th letter being reported as the j th letter, because the column sum of matrix $\{s_{i,j}\}$ was not 1.0. A similarity matrix $\{s_{i,j}\}$ was converted into a theoretical CM through a column-wise normalization (Luce, 1963b):

$$c_{i,j} = s_{i,j} / \sum_{j=1}^k s_{i,j}. \quad (8)$$

The k -by- k CM $\{c_{i,j}\}$ differed from $\{s_{i,j}\}$ in several ways. First, each column of $\{c_{i,j}\}$ summed up to 1.0, indicating that it summarized all responses to a stimulus

letter. Second, the diagonal line entries were no longer equal, due to the column-wise scaling. They were now the rates of correct recognition. Finally, $\{c_{i,j}\}$ was no longer symmetric, again due to the scaling of Equation 8. This asymmetry, however, was not caused by response bias human subjects produced in letter recognition experiments, as we will discuss later. The outcome of a GM model was thus a theoretic CM $\{c_{i,j}\}$. For a model with n

GMs, there were n free parameters, including the τ for the similarity matrix and the $n - 1$ independent weightings of the moment space.

Implementation of models

To simulate letter recognition close to acuity limits, we filtered the letters with a published contrast sensitivity

Models	Key features/number of free parameters	Simulation details	Correlation with empirical						
			Sloan	CC1	CC2	CC3	CC4	CC5	CC6
CHOICE	Similarity matrix and response bias vector. 54 parameters	Interpreting empirical CM. Not involving stimulus.	0.921	0.949	0.960	0.958	0.918	0.954	0.910
GM13/LM13	All GMs or LMs up to 4th order, excluding $m_{1,0}$ and $m_{0,1}$ 14 parameters	GM No Bias Correction	0.768	0.896	0.642	0.779	0.628	0.747	0.617
		GM Bias Correction	0.829	0.895	0.685	0.868	0.684	0.789	0.633
		LM No Bias Correction	0.792	0.895	0.665	0.800	0.611	0.799	0.701
GM9/LM9	$m_{0,0}, m_{2,0}, m_{1,1}, m_{0,2}, m_{3,0}, m_{0,3}, m_{4,0}, m_{2,2}, m_{0,4}$ 10 parameters	GM No Bias Correction	0.698	0.861	0.547	0.723	0.477	0.672	0.517
		GM Bias Correction	0.713	0.858	0.578	0.793	0.554	0.665	0.528
		LM No Bias Correction	0.680	0.779	0.514	0.739	0.408	0.711	0.528
GM7/LM7	$m_{0,0}, m_{2,0}, m_{0,2}, m_{3,0}, m_{0,3}, m_{4,0}, m_{0,4}$ 8 parameters	GM No Bias Correction	0.694	0.770	0.392	0.708	0.366	0.673	0.392
		GM Bias Correction	0.706	0.768	0.483	0.765	0.410	0.671	0.395
		LM No Bias Correction	0.680	0.779	0.317	0.739	0.408	0.711	0.528
GM5 ₁ /LM5 ₁	$m_{0,0}, m_{2,0}, m_{0,2}, m_{3,0}, m_{0,3}$ 6 parameters	GM No Bias Correction	0.690	0.698	0.327	0.558	0.212	0.673	0.335
		GM Bias Correction	0.699	0.697	0.376	0.642	0.241	0.671	0.339
		LM No Bias Correction	0.670	0.697	0.374	0.609	0.175	0.711	0.402
GM5 ₂ /LM5 ₂	$m_{0,0}, m_{2,0}, m_{0,2}, m_{4,0}, m_{0,4}$ 6 parameters	GM No Bias Correction	0.578	0.676	0.290	0.603	0.247	0.568	0.280
		GM Bias Correction	0.565	0.668	0.188	0.687	0.247	0.571	0.277
		LM No Bias Correction	0.594	0.760	0.519	0.687	0.373	0.571	0.463
GM5 ₃ /LM5 ₃	$m_{0,0}, m_{3,0}, m_{0,3}, m_{4,0}, m_{0,4}$ 6 parameters	GM No Bias Correction	0.692	0.746	0.392	0.708	0.366	0.576	0.396
		GM Bias Correction	0.703	0.741	0.483	0.765	0.410	0.568	0.396
		LM No Bias Correction	0.680	0.734	0.519	0.686	0.370	0.568	0.402
GM3 ₁ /LM3 ₁	$m_{0,0}, m_{2,0}, m_{0,2}$ 4 parameters	GM No Bias Correction	0.577	0.556	0.142	0.427	0.160	0.578	0.109
		GM Bias Correction	0.565	0.508	0.111	0.489	0.221	0.574	0.084
		LM No Bias Correction	0.595	0.563	0.147	0.431	0.119	0.596	0.321
GM3 ₂ /LM3 ₂	$m_{0,0}, m_{3,0}, m_{0,3}$ 4 parameters	GM No Bias Correction	0.395	0.461	0.278	0.446	0.118	0.575	0.225
		GM Bias Correction	0.414	0.472	0.319	0.489	0.142	0.578	0.254
		LM No Bias Correction	0.351	0.423	0.238	0.492	0.068	0.592	0.200
GM3 ₃ /LM3 ₃	$m_{0,0}, m_{4,0}, m_{0,4}$ 4 parameters	GM No Bias Correction	0.600	0.633	0.290	0.603	0.292	0.531	0.172
		GM Bias Correction	0.571	0.614	0.283	0.687	0.187	0.534	0.089
		LM No Bias Correction	0.594	0.552	0.374	0.609	0.384	0.509	0.220
GM1/LM1	$m_{0,0}$ 2 parameters	GM No Bias Correction	0.476	0.311	0.213	0.354	0.037	0.595	0.072
		GM Bias Correction	0.319	0.178	0.136	0.347	-0.197	0.575	0.040
		LM No Bias Correction	0.476	0.311	0.213	0.354	0.037	0.595	0.072
CSFTM	Template matching 1 parameter	No Bias Correction	0.729	0.649	0.408	0.447	0.155	0.640	0.292
		Bias Correction	0.761	0.664	0.453	0.549	0.205	0.656	0.337
RANDOM	Random off-diagonal line elements	Average of 500 CMs	0.202	0.136	0.154	0.195	0.299	0.151	0.120

Table 1. Model summary. Key features of the models evaluated in this study are summarized. Correlation coefficients between model generated and empirical CMs are listed for each stimulus group. The three correlation coefficients in a cell were obtained from GM model without bias correction, GM model with bias correction, and LM (Legendre moments) without bias correction, respectively.

function (CSF) (see [Appendix A](#) for details). A total of 10 GM models were constructed using GMs of 4th order and below. Key features of these models were shown in [Table 1](#). Zero-order GM $m_{0,0}$ was used in all GM models, and the 1st-order GMs, $m_{1,0}$ and $m_{0,1}$, were never used. This is because when the center of the coordinate was moved to the center of gravity of stimulus image ([Equation 4](#)), $m_{1,0}$ and $m_{0,1}$ became 0 for all stimuli. Model GM1(lu) used $m_{0,0}$ (lu) only. Models GM3(lu,va), GM3(lu,sk), and GM3(lu,ku) used $m_{0,0}$ (lu) and one pair of pure GMs, $m_{2,0}$ and $m_{0,2}$ (va), $m_{3,0}$ and $m_{0,3}$ (sk), and $m_{4,0}$ and $m_{0,4}$ (ku). Models GM3(lu,va,sk), GM3(lu,va,ku), and GM3(lu,sk,ku) used $m_{0,0}$ and two pairs of pure GMs. We did not use the 2nd order GM $m_{1,1}$ to construct GM3 and GM5 models because it only had a non-zero value when there was a gross difference in ink distribution between opposite corners of a letter, as in a letter “L” ([Figure 1a](#)). Because most of the stimulus characters shown in [Figure 2](#) filled the corners of a square area rather symmetrically, values of $m_{1,1}$ are very close to zero. The average values of $m_{2,0}$ and $m_{0,2}$ of the 70 characters were 26.2 ± 0.51 and 25.7 ± 0.50 , respectively, but the average of the absolute values of $m_{1,1}$ was only 0.023 ± 0.003 . GM7 had all pure GMs up to the 4th order. GM9 added $m_{1,1}$ and $m_{2,2}$ to GM7. GM13 used all GMs up to the 4th order, excluding $m_{1,0}$ and $m_{0,1}$. GMs higher than the 4th order were not discussed because we found adding 5th-order GMs resulted in little change in fitting empirical CMs.

To quantitatively demonstrate the effectiveness of GM models in predicting human performance, we also implemented three other models: a template-matching model (CSFTM) derived from a CSF ideal-observer model (Chung et al., 2002), a parameter-heavy choice model (CHOICE; Townsend, 1971), and an average of 500 CMs that had the empirical CMs’ diagonal line entries but random off-diagonal line entries (RANDOM). Details of these models were presented in [Appendix A](#). Pearson correlation coefficients were used to quantify the agreement between vectorized empirical and theoretical CMs.

Results

Using GM models to predict human confusions

The seven empirical CMs are shown in [Figure 3](#). It was evident that empirical confusions we observed were not random events. This was reflected in the large differences in the values of the off-diagonal line elements (confusions). Some letter pairs seldom got confused (0.00%) and some letter pairs got confused quite frequently (10~20%, pink and blue numbers in [Figure 3](#)). As shown in CM_{Sloan}, the well-known confusions in English letters, such as “C”

vs. “O,” “D” vs. “O,” and “N” vs. “H”, were reproduced. These prominent confusions were most likely to be associated with perceptual similarities between characters, and thus could not be the result of random guessing. We constructed several versions of “random confusion” matrices, including Townsend’s (1971) “equiprobable” (every cell had the same value), an “equal legibility” (all diagonal line elements contained the mean legibility of the empirical CM, and all off-diagonal line elements contained mean confusion), and a CM that retained the empirical relative legibility while all incorrect reports were evenly distributed in the nine off-diagonal cells of a column. χ^2 -tests between the empirical CMs and all these random CMs were conducted, and the results showed that the probability that any of the empirical CMs were produced by random reporting was <0.0005 .

Correlations between whole empirical and theoretical CMs were high. The average correlation across the 7 stimulus groups was 0.970 ± 0.006 for the GM model using 13 GMs. In comparison, for those CMs that had empirical CMs’ diagonal lines but random off-diagonal line entries, the mean whole matrix correlation was 0.955 ± 0.012 . Evidently whole CM correlation provided little information about how well models could predict confusions made by human subjects. Therefore, in this section we ignore the diagonal information and focus on results obtained from fitting the 90 confusion entries.

[Figure 4a](#) shows empirical–theoretical correlation coefficients produced by GM13, GM9, GM7, the average of GM5(lu,va,sk), GM5(lu,va,ku) and GM5(lu,sk,ku), the average of GM3(lu,va), GM3(lu,sk) and GM3(lu,ku), and GM1(lu) (colored symbols). Also shown are correlations produced by the CHOICE model (black squares) and the RANDOM model (black diamonds). These correlation coefficients are shown as the first number in each cell in [Table 1](#). The CHOICE model, which produced a mean correlation of 0.938 with empirical CMs, served as practical upper boundaries of goodness of fit. The CMs without consistent confusion patterns (RANDOM model) produced a mean correlation of 0.180 with empirical CMs and served as the lower boundaries for goodness of fit. [Figure 4b](#) shows scatter plots of the 90 confusions of the CC1 empirical CM against the corresponding 90 theoretical confusions produced by the CHOICE, GM13, GM7, and GM1(lu) models, along with regression lines through the data points. If empirical and theoretical confusions matched perfectly, data points would all fall on a line with a slope of 1.0. Because most of the confusions in each CM had low probabilities except for a few prominent confusions, as shown in [Figure 3](#), the data were scaled and log transformed before regression analyses were performed and plots were made. Regression analyses showed that CHOICE, GM13, and GM7 models all had slopes close to 1.0 (0.898, 0.820, and 0.936), while the GM1(lu) model had a very shallow slope (0.137). In the plots, the data points were scattered widely for the GM1(lu) model but clustered more tightly around the

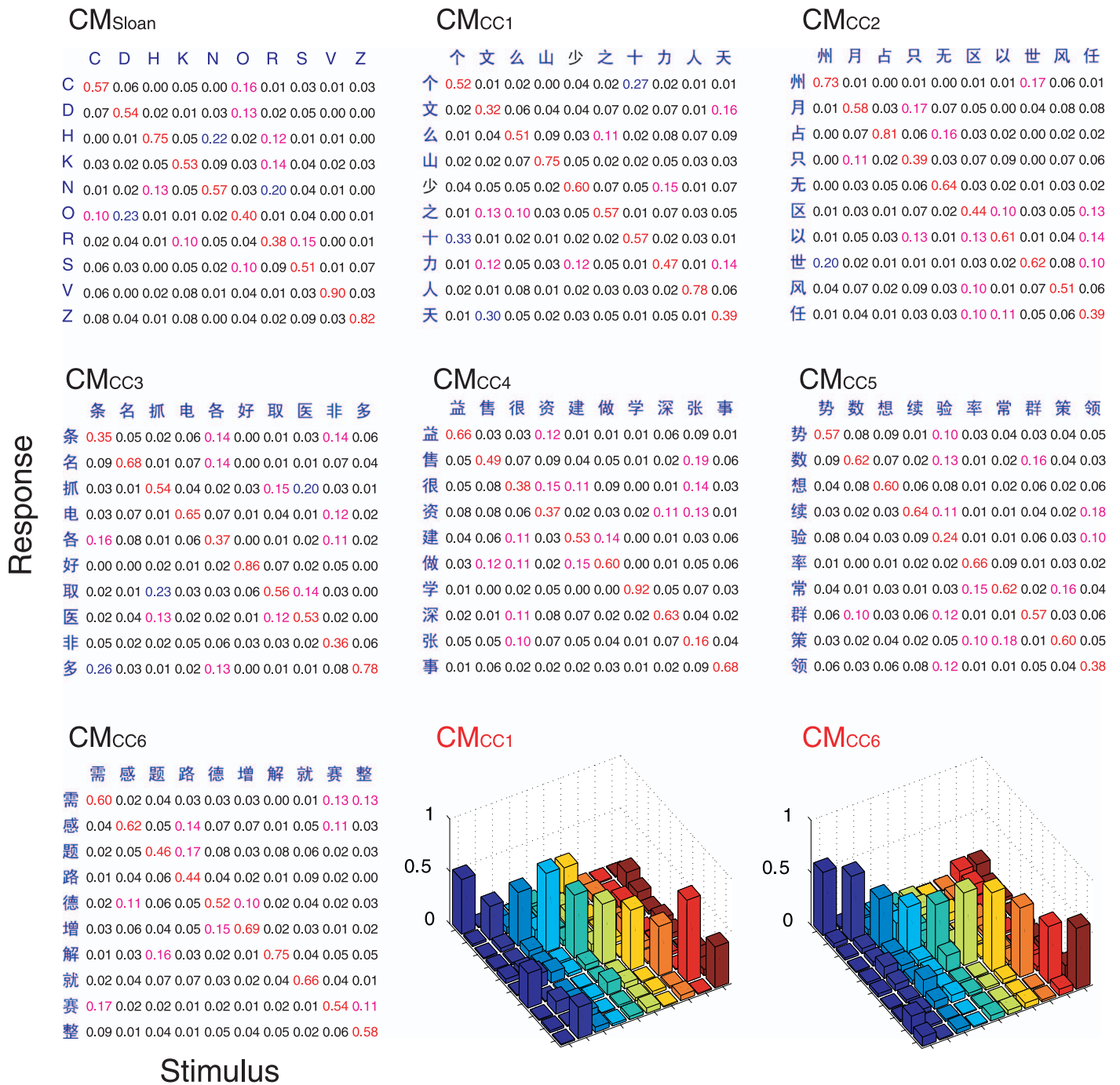


Figure 3. Empirical CMs obtained using Sloan, and CC1–CC6 stimulus groups. Off-diagonal line entries ≥ 0.1 are highlighted to indicate prominent confusions.

regression line for GM7, GM13, and CHOICE models. This was quantified by the adjusted R^2 of the linear regression model, which had values of 0.163, 0.484, 0.636, and 0.896 for GM1(lu), GM7, GM13, and CHOICE models, respectively.

It is obvious that the empirical–theoretical correlation increased with the number of GMs used in a GM model. However, just by having the mean luminance information of the stimulus ($m_{0,0}$), the correlations between GM1(lu) model CMs and empirical CMs were higher than the

lower boundary defined by RANDOM CM’s in 5 out of 7 stimulus groups. For example, the theoretical confusions obtained using the mean luminance values of Sloan letters and CC5 characters correlated with corresponding empirical confusions at 0.476 and 0.595, respectively, suggesting that even within these groups of stimuli of relatively uniform spatial complexity, letters of similar mean luminance were more likely to get confused when their sizes were close to acuity limits. With 7 GMs, correlation coefficients were 0.362–0.770. When all 13 GMs up to

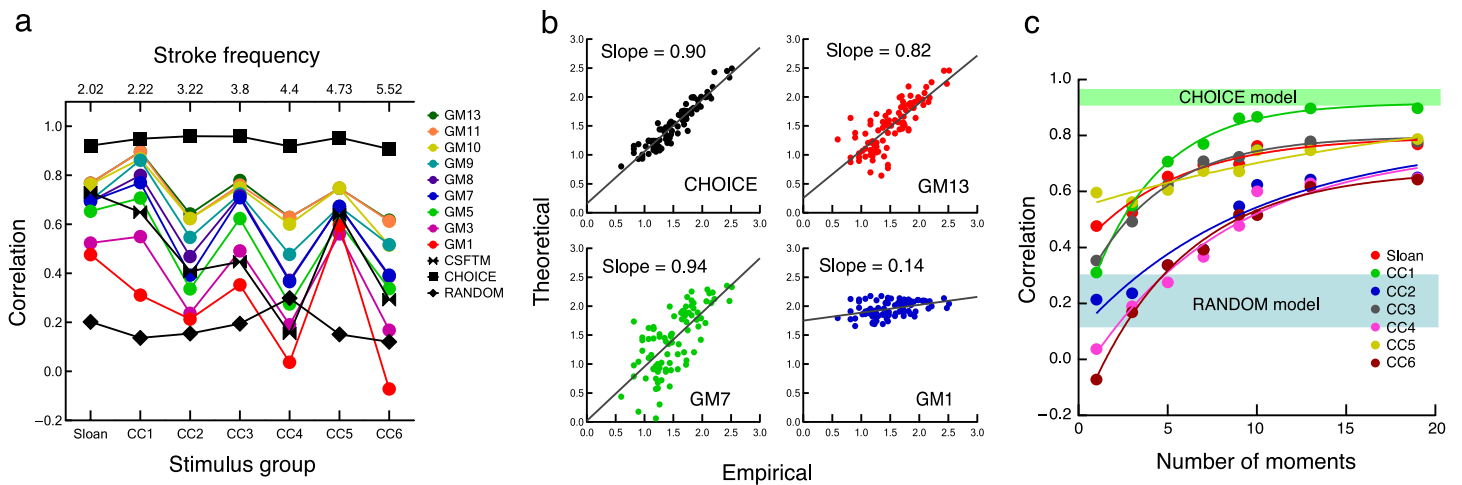


Figure 4. Fitting results. (a) Correlation coefficients between empirical and theoretical confusion entries of CMs. Theoretical CMs were produced by GM13, GM9, GM7, the average of GM5(lu,va,sk), GM5(lu,va,ku) and GM5(lu,sk,ku), the average of GM3(lu,va), GM3(lu,sk) and GM3(lu,ku), GM1(lu) and CSFTM models. Black squares and black diamonds are results from the CHOICE model and the RANDOM model, respectively. (b) Log transformed scatter plots of empirical confusions vs. theoretical confusions. Empirical confusions were from the CC1 stimulus group. Theoretical confusions were produced by the CHOICE, GM13, GM7, and GM1(lu) models. Best-fitting lines through the data points are drawn. The slopes of linear regression lines are also presented. (c) The relationship between the number of GMs used in GM models and model performance. The green and blue bands show the ranges of correlations obtained from the CHOICE and the RANDOM models.

4th order (excluding $m_{1,0}$ and $m_{0,1}$) were used, the correlation coefficients were at least 0.617. For Chinese characters in CC1, the correlation was about 0.896, approaching the level of performance of the 54-parameter CHOICE model.

Improvement of model performance with increasing number of GMs was more significant with lower order pure GMs. After all pure GMs were used (GM7), adding more GMs caused progressively less performance improvement. As shown in Figure 4c, the improvement saturated at about 9 to 11 GMs. Adding all six 5th-order GMs resulted in less than 0.3% of change in Sloan, CC1, CC2, and CC3 groups, and in 2%, 3%, and 5% improvement in CC4, CC5, and CC6. Therefore, for letter recognition near the acuity limit, pure GMs up to the 4th order explained a large portion of confusions made by human subjects, and adding higher order GMs did not seem to provide additional useful information.

Asymmetry of empirical CMs and effect of response biases

The empirical CMs we obtained were clearly asymmetric. For example, in Sloan letters, “D” was reported as “O” 23% of the time, while “O” was reported as “D” 13% of the time. Asymmetry can be appreciated by observing the row sums of CMs. The sum of the entries of a row in a CM indicates whether the report of a letter is more or less than its share. If subjects responded to all stimulus letters without bias, the row sums should all be close to 1.0. On

the other hand, if subjects have clear preference to report some letters, the row sums of these letters will be significantly greater than 1.0, and the row sums of other letters will become less than 1.0. In our Sloan letter CM, the row sums were 0.93, 0.88, 1.21, 0.97, 1.05, 0.83, 0.81, 0.93, 1.18, and 1.21. The standard deviation of the 10 row sums was 0.152. The subjects had a clear tendency to make more reports of “H” and “Z” (row sums 1.21) and not enough reports of “R” (row sum 0.81). In comparison, subjects made more uniform reports of the 10 characters in CC1. The row sums were 0.91, 0.81, 1.05, 1.05, 1.10, 1.05, 1.02, 1.02, 1.06, and 0.92 (standard deviation = 0.090).

Many researchers believe that perceptual similarity and confusion between two stimuli is determined only by stimulus properties and thus is inherently symmetric. Bias occurs because subjects have to make reports, and reports are influenced by subjects’ preference to stimuli. This view was clearly demonstrated in the CHOICE model, where a CM was a product of a symmetric similarity matrix and a bias vector (Equation A2). The CHOICE model provided explicit estimation of the bias $\{\beta_j\}$ vector (Equation A2) for each empirical CM. Because GM models produced symmetric similarity matrices, we could incorporate bias vectors estimated by the CHOICE model into these models. Specifically, a stimulus-driven similarity matrix $\{s_{i,j}\}$, obtained from a GM model was combined with the bias vector $\{\beta_j\}$ obtained from the CHOICE model to produce a biased theoretical CM. The optimization procedure was the same, because it was related only to the stimulus.

Without bias adjustments, the theoretical CMs were rather symmetric, with standard deviations of row sums around 0.04–0.08. The correlations between empirical and theoretical row sums were low, from -0.367 to 0.017 . Theoretical CMs with bias adjustments had row sums similar to those of the corresponding empirical CMs, and the correlations were high, from 0.717 to 0.907 ($p < 0.005$), indicating a faithful restoration of the bias observed in empirical CMs. Correlation coefficients between empirical CMs and biased theoretical CMs were shown as the second number in each cell in Table 1. It appeared that biased theoretical CMs generally correlated with empirical CMs better than unbiased, but the improvement was moderate at best.

Relative legibility

No model based on Euclidean perceptual distances, including the GM models and template matching models, can predict relative legibility directly, because the diagonal line entries of the model similarity matrix (Equation 7) are all 1.0. However, a model CM defined in Equation 8 did produce diagonal line entries of different values, due to the column-wise normalization. This was because each diagonal line entry was one minus the sum the nine off-diagonal line entries in the same column.

When the whole CM was optimized, the correlation between diagonal line entries of empirical and theoretical

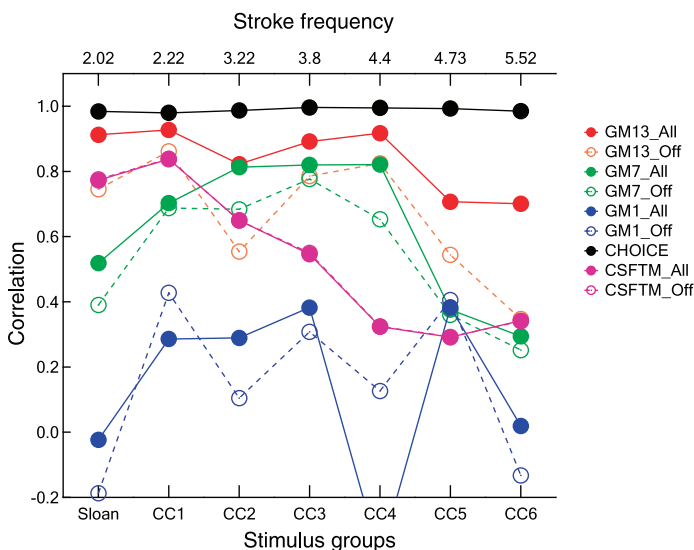


Figure 5. Correlation coefficients between diagonal line entries of empirical and theoretical CMs. Theoretical CMs were produced by GM13, GM7, GM1(lu), and CSFTM models. Solid symbols represent results obtained when the whole CMs were used in data fitting (CSFTM_All and CSFTM_Off were almost identical). Open symbols represent results of fitting only the confusions. Black circles are results from the CHOICE model.

CMs was higher than when only confusions were optimized. This was because whole CM optimization was mainly optimizing diagonal line entries. Figure 5 shows correlation coefficients between diagonal line entries of the 7 empirical CMs and theoretical CMs produced by GM13, GM7, GM1(lu), and CHOICE models. The CHOICE model produced the best correlation, 0.984 on average. The GM models each had two results, that from a whole CM fitting (solid symbol and solid line) and that from a confusion-only fitting (open symbol and dashed line). For Sloan letters, the empirical relative legibility was ROSKDCNHZV. The relative legibility produced by the whole CM fitting GM13 model was OSCRNDKHZV, which correlated with the empirical relative legibility at 0.912. GM models agreed with human subjects in that “R,” “O,” and “S” were more difficult to recognize and “H,” “Z,” and “V” were easier.

Discussion

GM as visual features

This study, as far as we know, is the first attempt to use low-order geometric moments as perceptual features to explain human recognition of characters near the acuity limit. Our study identified one set of perceptual features that might be used in recognizing these small patterns. Our study also showed how these features might be described, and how they might be used to choose one out of a set of 10 stimuli. GMs provide a systematic way to decompose 2-D patterns into visually perceivable features. A GM-based feature analysis model is thus universally applicable to any stimulus set.

Bouma (1971) measured confusions between 26 lower case letters near the acuity limit. While he did not propose an explicit model to fit his empirical CM, he did analyze perceptual distances between lower case letters and identified a set of 16 properties of lower case letters that served as cues to differentiate these letters through perceptual grouping. Many of these properties, or perceptual features, are global and could be related to lower order GMs. For example, the three “height-to-width quotients,” $H/W < 1.16$, $H/W > 1.16$, and $H/W > 1.22$, could be differentiated by $m_{2,0}/m_{0,2}$. The left, upper, right, and lower gaps could be perceived as skews of pattern luminance in different positions and thus can be differentiated by relative values and/or signs of $m_{0,3}$ and $m_{3,0}$. The rectangular envelope and circular envelope can be differentiated by the 4th-order GM $m_{2,2}$, which has a significantly higher value for a square than for a circle (1.0 vs. 0.664, regardless sizes). The fact that many perceptually important features identified by Bouma can be described by low-order GMs suggests that GMs provide a theoretical description of the basic features

human subjects use to recognize letters near the acuity limit.

One may expect that some kind of tendency in w_k values (Equation 6) may exist that reflect the intrinsic property of moment analysis, for example, the weights should be lower for higher order moments because they contribute less in low-pass-filtered stimuli. Our results, however, failed to reveal any consistent trend for w_k values. Only in CC5, w_k values decreased with GM order. In CC1, CC4, and CC6, w_k values first increased from 0th to 3rd order of GM, and then dropped at 4th-order GM. In Sloan and CC2, w_k values were the lowest at 2nd-order GM, went up at 3rd order, and then dropped at 4th order. In CC3, w_k values increased steadily from 2nd to 4th order. Our current understanding is that w_k values are very specific to the characteristics of the stimulus group. For example, if characters in a stimulus group all have similar ink pixels, then the 0th-order GM would contribute very little to differentiating characters in the group, and the weight for m_{00} will be small. This happened in CC6 where the variation of total number of pixels was reduced because a large number of strokes were packed in the same area. As a consequence, m_{00} did not contribute much to differentiating CC6 characters, and the weight for the 0th-order GM is very small. We have not yet found a satisfying way to explain the observed relationship between model weights and GM order.

Models with different number of parameters

The correlations produced by various models presented in the Results section only showed one aspect of model fitting. If we wanted to judge relative merits of models, the numbers of free parameters had to be considered, because models with more parameters tend to provide better fit to a set of data. There may not be one simple answer to this question, because the answer may depend on one's view of a balance between the models' abilities to fit data and its complexity, and there are more than one view on this matter. One method that determines a model's goodness of fit while taking its complexity into consideration is the Akaike's Information Criterion (AIC) (Akaike, 1974). The AIC is defined as $AIC = N - \ln(SS/N) + 2K$, where N is the number of data points to be fitted (90, if all confusions of a 10×10 CM are considered), SS is the residual sum of squares, and K is the number of free parameters. If A is a simpler model and B is a more complex one, then $\Delta AIC = AIC_B - AIC_A$ can help to determine the model that explains the data well with fewer free parameters. An evidence ratio, defined as $1/e^{-0.5\Delta AIC}$, is usually used to quantify how much more correct one model compared to the other. An alternative criterion, Bayesian Information Criterion (BIC), defined as $BIC = N \ln(SS/N) + K \ln(N)$, has a stiffer penalty for parameter usage (Schwarz, 1978). These criteria obviously will pass different judgments on relative merits of models with

different number of parameters. We used both AIC and BIC to give a more comprehensive view. Another advantage of AIC and BIC analyses is that they allow ranking multiple models according to their merits without having to adjust statistical criteria (α) like in multiple hypothesis tests.

Because AIC and BIC are related to a specific model fitting results of the specific experiment (N and K), it is difficult to judge whether an individual AIC or BIC value represents a good fit. For example, when the GM13 model was used to fit CC1 empirical CM, the BIC was -600.119 . It is difficult to say how good this fitting is, because there is not a universal best fitting criterion for BIC, like the 1.0 in correlation analysis. Only when two or more models are compared, does the importance of AIC and BIC become obvious. They help to select the more parsimonious, and thus the better model.

Template matching vs. Feature analysis

Template matching models in both spatial and frequency domains and with raw stimulus and filtered images have been built to account for English letter confusions (Blommaert, 1988; Gervais, Harvey, & Roberts, 1984; Loomis, 1990). The correlations between these model CMs and empirical CMs ranged from 0.24 and 0.70. Because some correlations were calculated using whole CMs, the goodness of fit of confusions could be much lower. To make direct comparison between the GM model and template matching models for letter confusions near the acuity limit, we derived a template matching (CSFTM) model from Chung's CSF ideal-observer model (Chung et al., 2002; see Appendix A). The performance of this model in predicting human letter confusions was shown in Figure 4a (black bow ties) and in Table 1. The correlations produced by the CSFTM model fitting was lower than the GM13 model in all stimulus groups, was comparable with that of GM9 in two stimulus groups, and was at the level of GM3 or GM5 in the remaining 4 stimulus groups. When used to predict human relative legibilities (Figure 5, purple circles), the CSFTM model relative legibilities correlated with empirical ones at 0.773 and 0.776 for whole matrix and confusion only optimizations. The GM13 model, when optimized for the whole CM, had higher diagonal line correlations than the CSFTM model at all stimulus groups, and when optimized for confusions only, had similar correlations as the CSFTM model at 4 stimulus groups and better correlations than the CSFTM model at the remaining 3 groups. The correlation of GM7 model fitting was lower than the CSFTM model at simpler stimulus groups (Sloan letter and CC1), and was equal or higher than the CSFTM model at more complex stimulus groups.

Using AIC, we found evidence that was overwhelmingly in favor of the 1-parameter CSFTM model than the 13-parameter GM13 model for the Sloan letters (AIC for

GM13 and CSFTM were -590.061 and -607.049 , respectively, and the evidence ratio was 4886.27). However, the evidence was overwhelmingly in favor of the GM13 model in 5 of the 6 CC stimulus groups (evidence ratios > 2280) with the exception of CC5 where the GM13 was still in favor, but the evidence was not very persuasive (2.05). Comparisons between CSFTM and GM9 produced the similar results. When the number of moments in GM models was further reduced, the CSFTM model was favored in more and more stimulus groups. It is also interesting to compare the CSFTM model with the 54-parameter CHOICE model. There was only weak evidence that CSFTM is better than CHOICE in Sloan letters (evidence ratio 3.84), but the evidence in favor of the CHOICE model was overwhelming in all CC stimulus groups (evidence ratio $> 2 \times 10^{11}$).

Using BIC, we found overwhelming evidence in favor of the CSFTM model over the GM13 model for the Sloan letters, CC2, CC5, and CC6 groups, and overwhelming evidence in favor of the GM13 model for CC1, CC3, and CC4 groups. When the CSFTM and CHOICE models were compared using BIC, the evidence was overwhelmingly in favor of the CSF model for all stimulus groups.

Therefore, under some situations, like the Sloan letters, the 1-parameter CSFTM model was always superior to other parameter-rich models. In other situations, such as CC1 and CC3, GM13 and GM9 were always superior to the CSFTM models, indicating that the improvement of model performance had outweighed the cost of using more parameters. In other situations, the relative merits of models depended on the criteria used.

The CSFTM model did not seem to be sensitive to the exact shape of the CSF. We simulated the CSFTM model using three front-end filters (see below). There were no significant differences between correlation coefficients produced by different CSFs among the 7 stimulus groups (repeated measures ANOVA, $F_{2,12} = 1.179$, $p = 0.341$).

The effect of the front-end filter

In the study, we used an empirical CSF as a front-end filter. GM features were extracted from filtered characters. To assess the influence of the front-end filter on the performance of GM models, we repeated the analyses using two additional filters, an empirical CSF obtained from one of our subjects (ZJY CSF, blue squares and curve in Figure A1a), and the 3.8-mm pupil point spread function of Campbell and Gubisch's (1966; CG PSF). The ZJY CSF had a slightly higher peak frequency than the CLT CSF, and a lower cut-off frequency. The CG PSF was fitted with the sum of two Gaussians (Geisler, 1984) and then convolved with bitmaps of the stimuli. The corresponding correlation coefficients produced by the CLT CSF, ZJY CSF, and CG PSF front-end filters agreed with each other very well. An ANOVA showed that

among the 70 sets of model simulations (7 groups \times 10 GM models) there were no significant filter effect ($F_{2,120} = 0.182$, $p = 0.893$) or filter*model interaction ($F_{18,120} = 0.419$, $p = 0.982$). These results indicated that the GM models are not very sensitive to the details of the front-end filtering. This should not be surprising, because the GM models used only the global features of the characters, which were attenuated to some degree but not eliminated or distorted by these linear filters. The insensitivity of GM model performance to the type of the front-end filter may also be explained by the feature matching process used by the model. We used the feature list extracted from a filtered stimulus character to match a set of stored feature lists that were extracted from identically filtered characters. This approach was appropriate in our study, where subjects had done thousands of trials of recognizing characters close to acuity thresholds. It would be interesting to investigate the effect of front-end filters on GM models when the stimulus feature set is extracted from a filtered character while the store feature set is extracted from unfiltered characters. This situation may occur when the subject has learned the stimuli at very large sizes and is first exposed the stimulus at sizes close to the acuity limits.

Relative legibility, response bias, and asymmetry of confusion matrices

While studies of letter recognition CM considered the observed asymmetry as the consequence of response bias (Townsend, 1971), which, strictly speaking, is the subject's tendency to report one letter more than others when the stimulus is not legible, relative legibility of the stimulus letters actually plays an important role in determining the magnitudes of pair-wise asymmetries. In our CM_{Sloan} , for example, "H" was much more legible than "R" (75% correct vs. 38%). This difference contributed to the large pair-wise asymmetry in which the probability of an "H" being reported as an "R" was 0.012 while an "R" being reported as an "H" was 0.123. Indeed, when we generate random CMs, we could produce substantial asymmetries among confusion entries if large variations in the diagonal line entries were introduced. Therefore, unless the stimuli are truly equally legible, asymmetry of an empirical CM is not a pure measure of subjects' response bias.

Inter-group variations

It is obvious from Figure 4a that GM models fitted Sloan letters, CC1, CC3, and CC5 better than CC2, CC4, and CC6. It seemed that this fluctuation did not happen only to GM modeling. Although the CHOICE model only concerned with empirical CMs and did not care what stimulus produced them, it echoed the fluctuation. The fluctuation was also in some degree mirrored by the

RANDOM model, which produced higher correlations with empirical CMs when other models produced lower correlations. Therefore, the fluctuation seemed to be related to properties of the empirical CMs. The mean correct rates of the empirical CMs (59.7%, 54.7%, 57.3%, 56.75, 54.4%, 55.0%, and 58.7% for Sloan and CC1–CC6) were similar and did not echo the pattern seen in Figure 4a. It was suggestive that when the RANDOM model performed best, the other models performed worst, for example, at CC4. Because the RANDOM model, after averaging many random confusion entries, had essentially uniform confusion entries, it seemed reasonable to suggest models may fit empirical CMs with more prominent confusion entries better than those with more uniform confusion entries. Indeed, model fitting appeared to be better when confusion entries had larger standard deviations, higher maximum values, and larger ranges. For example, the standard deviations for the confusions of the 7 groups were 0.0493, 0.0583, 0.0447, 0.0529, 0.0418, 0.0424, and 0.0392. Standard deviations, maximum values, and ranges correlated highly with GM13 model performance and were able to account for 78.2%, 76%, and 74.1% of the inter-group variations for this model.

As the complexity of the letter set increases, the standard deviation of the confusion entries decreases, suggesting that confusions became more uniformly distributed. As a consequence, performance of CSFTM and GM models declines. This can be seen in the goodness of fit of both the diagonal line entries and off-diagonal line entries of empirical CMs in Figures 4a and 5. Although GM fitting of human confusions obviously fluctuated across stimulus group, and seemed to have two tiers for goodness of fit, the change among each tier was small. In comparison, the CSFTM fitting, with the exception of CC5 in Figure 4a, decline steadily with stimulus complexity, quantified by the stroke frequency. The less rapid decline of the GM models may be explained by the larger numbers of parameters they used to fit the confusion entries.

Non-orthogonal GMs vs. orthogonal moments

Because the basis functions of GMs are not orthogonal, GM features have non-zero redundancy, meaning that GMs may not correspond to independent characteristics of the image, from a mathematical point of view. Various orthogonal moment descriptors have been proposed in computer image processing. Legendre moments (LM) defined below were orthogonal and real-valued:

$$L_{p,q} = \frac{(2p+1)(2q+1)}{(N-1)^2} \sum_{i=1}^N \sum_{j=1}^N P_p(x_i) P_q(y_j) f(i,j), \quad (9)$$

where $P_p()$ and $P_q()$ are Legendre polynomial of the p th and q th orders, and x_i and y_j are normalized pixel

coordinates in a unit square (Teague, 1980). Computational studies have demonstrated that orthogonal moments, such as $L_{p,q}$, are more efficient in reconstructing details of images. Ghorbel, Derrode, Dhahbi, and Mezhoud (2005) demonstrated that a 32×32 pixel “E” reconstructed from up to 50th-order GM produced 74 pixels that differed from the original. Only up to 8th-order LMs were needed to reach the same level of reconstruction accuracy. Human pattern recognition, on the other hand, typically does not require that much detail. Do LMs explain human recognition of small patterns better than GMs? Correlation coefficients between CMs produced by LM-based feature analysis models and empirical CMs were shown as the third number in cells in Table 1. LM models were slightly better than corresponding GM models in some cases, but the differences were generally small. For the 10 corresponding GM and LM models, the differences in empirical–theoretical correlations averaged across the 7 stimulus groups ranged from +0.019 (GM better than LM) to –0.083 (LM better than GM). Comparing to the dramatically increased efficiency of LM in reconstruction of images (8th order vs. 50th order), the improvement of LM over GM in interpreting human letter confusions near the acuity limit was mediocre. One possible reason is that the basis function of an LM of $(p+q)$ th order is a linear combination of GM basis functions up to the $(p+q)$ th order (Teague, 1980). It can be shown that at the lower orders that were most relevant to recognition of letters near the acuity limit, the basis functions of LM are similar to those of GM, and an LM was either the same order GM times a constant or a linear sum of two GMs of equal or lower order. For example, $L_{0,0} = m_{0,0}$, $L_{2,0} = (5/4)[(3/2)m_{2,0} - (1/2)m_{0,0}]$, $L_{1,1} = (9/4)m_{1,1}$, $L_{3,0} = (7/4)[(5/2)m_{3,0} - (1/2)m_{0,0}]$, and so on. Therefore, the difference between LM and GM compositions was relatively small when human recognition of small letters was concerned. It can be seen in Table 1 that GM1(lu) and LM1(lu) models produced identical simulation results. In other models, some LM space dimensions were linear combinations of two GM dimensions, and adjusting weightings on these dimensions might not produce the same results as in the GM space.

Visual acuity and beyond

Visual acuity is a unique condition where the subject is forced to use global structural properties of stimuli because of the optical and neural filtering. That is probably one of the reasons for the success of the low-order GM models. Recently, Watson and Ahumada (2008) used a CSF template matching ideal observer model to fit visual acuity data obtained under several aberrated conditions. The GM-based feature analysis model may also be used for this purpose. GM features can be extracted from aberrated images

degraded by internal noise and filtered by a neural transfer function. The feature sets can then be matched at different letter sizes in a Monte Carlo simulation to determine threshold letter sizes for different types and amounts of aberrations.

Whether GM-based feature models are applicable to other experimental conditions, such as briefly flashed large and high-contrast letters, or large low contrast letters, remains to be seen. However, there are a few situations where an explicit model for utilizing global properties of visual stimuli is desirable. First, a window of 4 to 8 letters is necessary in order to reach maximum reading speed in normal subjects (Legge, Pelli, Rubin, & Schleske, 1985; Poulton, 1962), and 15 letters to the right of fixation could contribute to normal reading of English (Legge, Hooven, Klitz, Stephen Mansfield, & Tjan, 2002; Legge et al., 1985; Rayner & McConkie, 1976). When reading text whose size is suitable for foveal reading (≈ 3 – 4 times foveal acuity size), peripheral letters that contribute to reading are likely to be near the acuity size at their locations, because acuity drops sharply with retinal eccentricity. Second, patients who lose central vision due to diseases such as macular degeneration may be trained to read using intact peripheral vision. In such cases, a large text size has to be used, but the room for magnification is quite limited, because of the limited size of a magnifier aperture or a large print page. Therefore, reading is likely to be done near the acuity size of the intact part of the retina. Finally, there is a general consensus that visual patterns are recognized in a global-to-local manner under normal circumstances (Bouma, 1971; Eriksen & Schultz, 1978; Lupker, 1979; Sanocki, 1991; Townsend, Hu, & Kadlec, 1988). While there are debates whether global features can facilitate local feature detection (Lupker, 1979; Sanocki, 1991), there is little doubt that global features of a pattern are extracted and used first. The analog is a “focusing process” where a visual pattern appears as a “blob-like” form first, and more details emerge later (Bouma, 1971; Eriksen & Schultz, 1978; Lupker, 1979; Sanocki, 1991). A universal mathematical description of global characteristics of visual stimuli can be a powerful tool in study human performance under these situations.

Appendix A: Models and their implementations

Implementation of GM models

Because we were interested in recognition of letters near the acuity limit, we assumed that GM features were extracted from an internal representation of a stimulus

letter after it was degraded by ocular optics and early neural processing. In previous modeling efforts for human letter recognition, a low-pass filtering (Blommaert, 1988; Loomis, 1990) or a contrast sensitivity function (CSF) filtering (Chung et al., 2002) of the stimulus letters was proposed, the former representing the effects of the ocular optics and the latter representing the effects of early optical/neural processing. We chose to use CSF filtering before GM feature extraction. Specifically, we selected a normal adult CSF function published by Chung et al. (2002; Figure 2). A 3-parameter model (Mannos & Sakrison, 1974)

$$CSF(f) = afe^{-(bf)^c}, \quad (A1)$$

was used to fit the CSF. The best-fitting parameters were $a = 812.3$, $b = 1.071$, $c = 0.636$. The Chung, Legge and Tjan (CLT) CSF data are replotted in Figure A1a with the model fitting. A radial-symmetric 2-D filter in the frequency domain was then created, as shown in Figure A1b. The original stimuli were black-and-white bitmaps of Sloan letters and Chinese characters shown in Figure 2. All stimulus bitmaps were 50×50 pixels in size. Because the bitmaps were used as stimuli 0.1 log unit above the acuity sizes shown in Figure 2, the resolutions ranged 700 pixels/deg for the Sloan letters to 398 pixels/deg for CC6. The stimulus bitmaps were pasted on a large black background before being filtered. For each stimulus letter, the Fourier transform of its bitmap was multiplied with the CLT CSF filter, and then the inverse Fourier transform was taken. The result was a highly blurred version of the letter, shown in Figure A1c. The filtered stimulus was cropped to $1.2 \times$ of original stimulus size to save time in simulation. Cropped sizes of up to $3 \times$ of original stimulus size were tested, and no significant differences in simulation results were noticed. Procedures defined in Equations 2–5 were used to extract a set of n GMs, and a theoretic CM $\{c_{i,j}\}$ with $n + 1$ free parameters was generated according to Equations 6–8. An optimization routine was used to look for the best fitting parameters so that the sum of squared differences between a theoretic CM and the corresponding empirical CM was minimal. One of the debatable issues in fitting CMs is how to treat the diagonal line entries. In a typical letter recognition study, the average correct rate is set at 50%–75%, which means that the diagonal line entries are much larger than off-diagonal line entries. If the whole CM is fitted, the diagonal line entries are likely to dominate the fitting and may result in an over-optimistic goodness of fit, in which the fitting of the confusions can be quite poor (LeBlanc & Muise, 1985). In our study, we optimized both the whole CM and confusions entries.

The letter sizes used in the study were 0.1 log unit above acuity sizes shown in Figure 2.

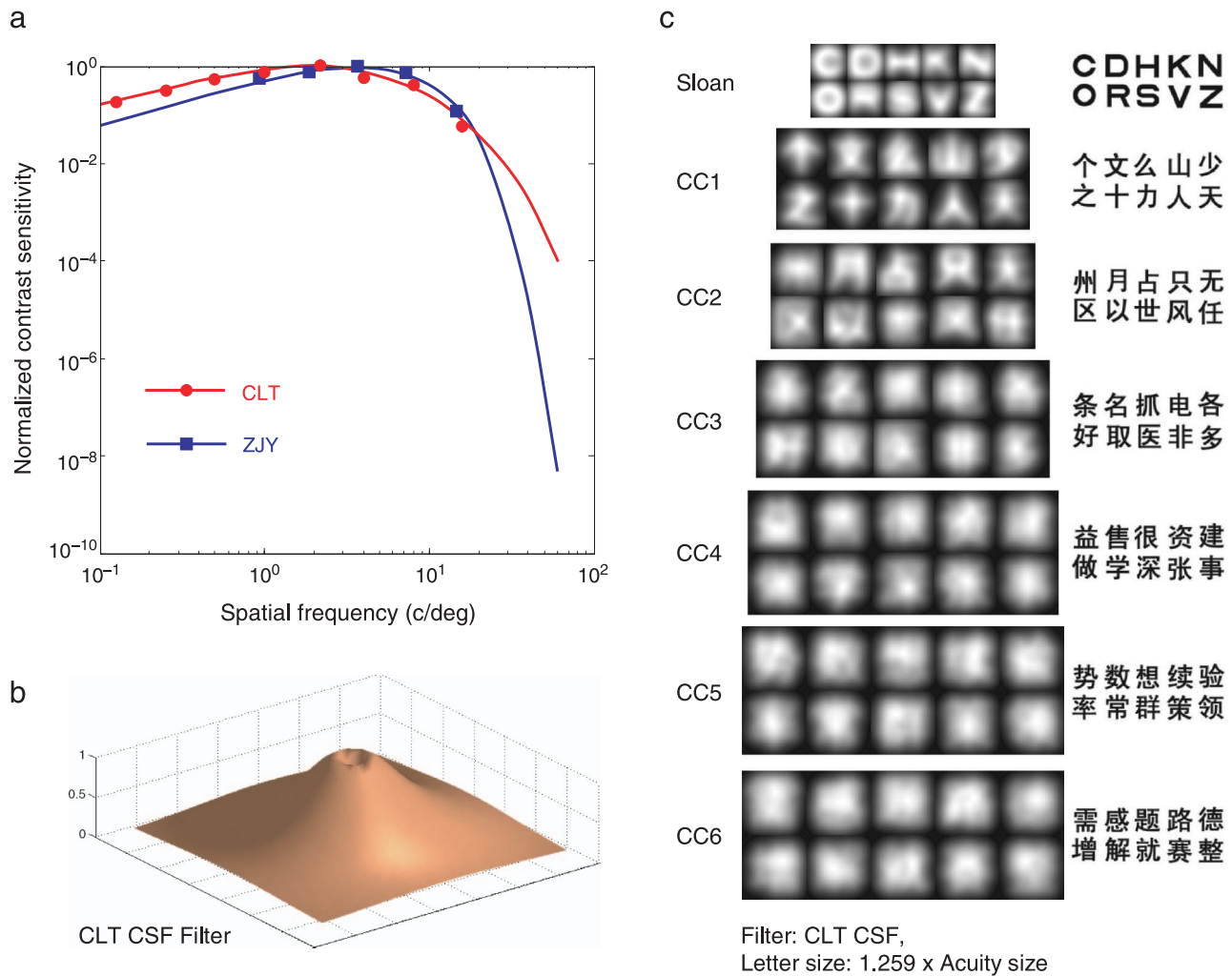


Figure A1. (a) An empirical CSF function from Chung et al. (2002). Data are fitted with a three parameter function. (b) CSF filter used in filtering the stimulus. (c) Filtered stimuli used in the model simulations.

CHOICE Model CMs and RANDOM Confusion CMs

To set up benchmarks for goodness of CM fitting and to gain insights into issues such as response bias, we also analyzed our empirical CMs with the choice model. Luce (1963a) first formulated a general model for choice experiments, which stipulated that the probability of the subject making a response of r given a stimulus s was determined by two scales, a similarity between r and s , and response bias. Townsend (1971) applied the choice model to recognition of uppercase English letters. The probability of the i th letter being reported as the j th letter was given by

$$c_{ij} = \eta_{ij}\beta_j / \sum_{k=1}^{26} \eta_{ik}\beta_k, \quad (\text{A2})$$

where $\{\eta_{ij}\}$ was a symmetric similarity matrix, and $\{\beta_k\}$ was the bias vector. The purpose of Townsend's choice model was to derive the unknown $\{\eta_{ij}\}$ and $\{\beta_k\}$ from a

known empirical CM. Therefore, for an experiment involving k letters, the model estimated $[k(k+1)/2] - 1$ parameters, $k(k-1)/2$ pair-wise similarities, and $(k-1)$ independent relative biases. For an experiment involving recognizing 10 letters, the number of parameters was 54. Because of the enormous number of parameters used, the choice model CM defined in Equation A2 usually provides an excellent fit of the empirical CM and thus can serve as a practical upper limit for goodness of fit. Explicit formulas provided by Townsend (1971) were used to calculate choice CM, similarity matrix, and response bias vector for each of our 7 empirical CMs.

A random CM did not contain a consistent confusion pattern. In each random CM, the diagonal line entries were the same as an empirical CM, and the off-diagonal line entries were randomly generated, satisfying only the restriction that each column sum was equal to 1.0. There was no free parameter in such a CM. An ensemble of such CMs represented hypothetical observer who made the same correct responses as human subjects but did not have a consistent confusion pattern. Each of our empirical CM

was correlated with 500 such random CMs. The average of the correlation coefficients served as a lower boundary for fitting empirical CMs. Because each random CM still had to satisfy the unit column sum requirement, the average correlations with empirical CMs had positive values (black diamonds in [Figure 4a](#)).

CSF Template-Matching Model (CSFTM)

To compare GM models with alternative models for letter recognition, we derived a CSF template-matching (CSFTM) model from a CSF ideal-observer model (Chung et al., 2002) and applied it to our 7 groups of stimuli. The CSF ideal-observer model was essentially a template-matching model that minimized the sum of squared differences between a stimulus and a template (Tjan et al., 1995). Specifically, the model calculated all sums of squared differences between a CSF filtered stimulus letter and all similarly filtered template letters, and selected as the response the template that had the smallest sum of squared differences from the stimulus. For a CSF filtered stimulus letter S and a filtered template letter T , the sum of squared difference is

$$D_{s,t} = \sum_x \sum_y (S(x,y) - T(x,y))^2. \quad (\text{A3})$$

For a k -letter recognition experiment, $\{D_{s,t}\}$ is a k -by- k matrix with zeros on the diagonal, similar to $\{d_{i,j}\}$ in [Equation 6](#). We used procedures similar to [Equations 7](#) and [8](#) to create a CSF template-matching similarity matrix and a confusion matrix. The CSFTM model has only one free parameter, τ in [Equation 7](#).

The implementation of a CSFTM model was similar to that for GM models. The same CSF-filtered letters were used as inputs, and the same optimization procedures were used. Because template matching was sensitive to relative positions between the stimulus and template, we displaced the stimulus relative to the template in 1-pixel step for 5 pixels in both horizontal and vertical directions in calculating $D_{s,t}$. The best result among the 25 relative positions was taken as the $D_{s,t}$. Seven and 9-pixel relative shifts were tested and no significant differences in simulation results were noticed. Loomis (1990) used similar method to study CMs of 26 English letters. Instead of a CSF, Loomis used a low-pass filter to simulate the effect of ocular optics.

In our modeling effort, we computed a theoretical CM for each set of free parameters, calculated the correlation between the theoretical CM and the corresponding empirical CM, and looked for a set of parameters that maximized theoretical/empirical CM correlation. An alternative is a Monte Carlo simulation of a psychophysical experiment, where an idea observer equipped with an optimal decision making strategy tries to recognize

luminance noise degraded characters (Pelli et al., 2006; Tjan & Legge, 1998; Watson & Ahumada, 2008). The outcome of this simulation is also a theoretical CM, which can then be used to compare with an empirical CM. Tjan et al. (1995) gave mathematical proof that the calculation of ideal observer recognition performance of Gaussian noise degraded objects is equivalent to the calculation of template matching. It would be interesting to investigate how the ideal observer decision making would affect GM model performance.

Reconstruction of a 2-D image with geometric moments

Because a finite spatial function $f(x, y)$ can be completely specified by geometric moments $\{M_{p,q}\}$ ($p, q = 0, 1, 2, \dots$) (Hu, 1962), one should also be able to reconstruct $f(x, y)$ from $\{M_{p,q}\}$. However, reconstruction with GM is not as straightforward as an inverse Fourier transform (Ghorbel et al., 2005), because the basis functions of GM are not orthogonal. Teague (1980) proposed a moment matching method to reconstruct any function $f(x, y)$ from GMs up to a given order N_{\max} . The idea is to obtain a continuous function $g(x, y) = g_{00} + g_{10}x + g_{01}y + g_{20}x^2 + g_{11}xy + g_{02}y^2 + \dots$, whose GM exactly match those of $f(x, y)$ up to the order N_{\max} . The constant coefficients g_{pq} of $g(x, y)$ are determined by

$$\int_{-1}^{+1} dx \int_{-1}^{+1} dy (g_{00} + g_{10}x + g_{01}y + g_{20}x^2 + g_{11}xy + g_{02}y^2 + \dots) x^p y^q \equiv M_{p,q}. \quad (\text{A4})$$

[Equation A4](#) results in a set of linear equations. Solving these equations, which is equivalent to calculating an inversed matrix of known $M_{p,q}$, uniquely determines all the coefficients of $g(x, y)$. In theory, exact reconstruction of an $N_x \times N_y$ image can be made by using moments $M_{p,q}$, where $p = 1, 2, \dots, N_x$, and $q = 1, 2, \dots, N_y$. We used our implementation of this method in Matlab to create the reconstruction shown in [Figure 1b](#). One inconvenience of moment matching reconstruction is that the coefficients g_{jk} depend on the order of the moments used in reconstruction. For example, g_{22} for the same image has different values for $N_{\max} = 4$ and $N_{\max} = 5$. This is because each coefficient g_{jk} is a linear combination of $M_{p,q}$ up to the N_{\max} order.

Acknowledgments

This research was supported by a Natural Science Foundation of China grant NSFC-30725018 and a Chang-Jiang Scholar professorship (CY), and by the Beijing Normal University Project 111 (LL and CY).

Commercial relationships: none.

Corresponding authors: Lei Liu; Cong Yu.

Emails: liul7788@uab.edu; yucong@bnu.edu.cn.

Addresses: School of Optometry, University of Alabama at Birmingham, Birmingham, AL 35294, USA; Institute of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.
- Alt, F. L. (1962). Digital pattern recognition by moments. *Journal of Association Computing Machinery*, *9*, 240–258.
- Blommaert, F. J. (1988). Early-visual factors in letter confusions. *Spatial Vision*, *3*, 199–224. [PubMed]
- Bondarko, V. M., & Danilova, M. V. (1997). What spatial frequency do we use to detect the orientation of a Landolt C? *Vision Research*, *37*, 2153–2156. [PubMed]
- Bouma, H. (1971). Visual recognition of isolated lower-case letters. *Vision Research*, *11*, 459–474. [PubMed]
- Campbell, F. W., & Gubisch, R. W. (1966). Optical quality of the human eye. *The Journal of Physiology*, *186*, 558–578. [PubMed] [Article]
- Chung, S. T., Legge, G. E., & Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Research*, *42*, 2137–2152. [PubMed]
- Eriksen, C. W., & Schultz, D. W. (1978). Temporal factors in visual information processing. In J. Requin (Ed.), *Attention and performance* (vol. VII). New York: Academic Press.
- Ferris, F. L., 3rd, Kassoff, A., Bresnick, G. H., & Bailey, I. (1982). New visual acuity charts for clinical research. *American Journal of Ophthalmology*, *94*, 91–96. [PubMed]
- Geisler, W. (1984). Physical limits of acuity and hyper-acuity. *Journal of the Optical Society of America A, Optics and Image Science*, *1*, 775–782. [PubMed]
- Gervais, M. J., Harvey, L. O., Jr., & Roberts, J. O. (1984). Identification confusions among letters of the alphabet. *Journal Experimental Psychology: Human Perception and Performance*, *10*, 655–666. [PubMed]
- Getty, D. J., Swets, J. A., Swets, J. B., & Green, D. M. (1979). On the prediction of confusion matrices from similarity judgments. *Perception & Psychophysics*, *26*, 1–19.
- Geyer, L. H., & DeWald, C. G. (1973). Feature lists and confusion matrices. *Perception & Psychophysics*, *14*, 471–482.
- Ghorbel, F., Derrode, S., Dhahbi, S., & Mezhoud, R. (2005). *Reconstructing with geometric moments*. Paper presented at the International Conference on Machine Intelligence (ACIDCA-ICMI'05), Tozeur, Tunisia.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.
- Ginsburg, A. P. (1977). *Visual information processing based on spatial filters constrained by biological data*. Unpublished Ph.D., University of Cambridge, Cambridge, England.
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, *39*, 3537–3560. [PubMed]
- Hu, M. K. (1962). Visual pattern recognition by moment invariant. *IRE Transactions on Information Theory, IT-8*, 179–187.
- International Organization for Standardization (ISO) (1986). Visual acuity testing: Standard optotype and its presentation (No. ISO 8596). International Organization for Standardization.
- Keren, G., & Baggen, S. (1981). Recognition models of alphanumeric characters. *Perception & Psychophysics*, *29*, 234–246. [PubMed]
- Laughery, K. R. (1969). Computer simulation of short-term memory: A component decay model. In G. T. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (vol. VI). New York: Academic Press.
- LeBlanc, R. S., & Muise, J. G. (1985). Alphabetic confusion: A clarification. *Perception & Psychophysics*, *37*, 588–591. [PubMed]
- Legge, G. E., Hooven, T. A., Klitz, T. S., Stephen Mansfield, J. S., & Tjan, B. S. (2002). Mr. Chips 2002: New insights from an ideal-observer model of reading. *Vision Research*, *42*, 2219–2234. [PubMed]
- Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading—I. Normal vision. *Vision Research*, *25*, 239–252. [PubMed]
- Loomis, J. M. (1990). A model of character recognition and legibility. *Journal Experimental Psychology: Human Perception and Performance*, *16*, 106–120. [PubMed]
- Luce, D. R. (1963a). Detection and recognition. In D. R. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (vol. I, pp. 103–188). New York: John Wiley and Sons.
- Luce, D. R. (1963b). Psychophysical scaling. In D. R. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (vol. I, pp. 245–307). New York: John Wiley and Sons.

- Lupker, S. J. (1979). On the nature of perceptual information during letter perception. *Perception & Psychophysics*, *25*, 303–312. [PubMed]
- Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research*, *42*, 1165–1184. [PubMed]
- Mannos, J. L., & Sakrison, D. J. (1974). The effects of a visual fidelity criterion on the encoding of images. *IEEE Transactions on Information Theory*, *20*, 525–535.
- Mukundan, R., & Ramakrishnan, K. R. (1998). *Moment functions in image analysis: Theory and applications*. Singapore: World Scientific.
- National Academy of Science National Research Council (NAS-NRC) (1980). Recommended standard procedures for the clinical measurement and specification of visual acuity. Report of working group 39. Committee on vision. *Advances in Ophthalmology*, *41*, 103–148. [PubMed]
- Neisser, U. (1967). *Cognitive psychology*. New York: Meredith.
- Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research*, *31*, 1399–1415. [PubMed]
- Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, *46*, 4646–4674. [PubMed]
- Poulton, E. C. (1962). Peripheral vision, refractoriness and eye movements in fast oral reading. *British Journal of Psychology*, *53*, 409–419. [PubMed]
- Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research*, *16*, 829–837. [PubMed]
- Sanocki, T. (1991). Effects of early common features on form perception. *Perception & Psychophysics*, *50*, 490–497. [PubMed]
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323. [PubMed]
- Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, *369*, 395–397. [PubMed]
- Teague, M. (1980). Image analysis via the general theory of moments. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, *70*, 920–930.
- Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research*, *35*, 3053–3069. [PubMed]
- Tjan, B. S., & Legge, G. E. (1998). The viewpoint complexity of an object-recognition task. *Vision Research*, *38*, 2335–2350. [PubMed]
- Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, *9*, 40–50.
- Townsend, J. T., Hu, G. G., & Kadlec, H. (1988). Feature sensitivity, bias, and interdependencies as a function of energy and payoffs. *Perception & Psychophysics*, *43*, 575–591. [PubMed]
- Watson, A. B., & Ahumada, A. J., Jr. (2008). Predicting visual acuity from wavefront aberrations. *Journal of Vision*, *8*(4):17, 1–19, <http://journalofvision.org/8/4/17/>, doi:10.1167/8.4.17. [PubMed] [Article]
- Zhang, J. Y., Zhang, T., Xue, F., Liu, L., & Yu, C. (2007). Legibility variations of Chinese characters and implications for visual acuity measurement in Chinese reading population. *Investigative Ophthalmology & Visual Science*, *48*, 2383–2390. [PubMed] [Article]