

Comparing the visual spans for faces and letters

Yingchen He

Department of Psychology, University of Minnesota,
Twin Cities, MN, USA



Jennifer M. Scholz

College of Optometry, The Ohio State University,
Columbus, OH, USA



Rachel Gage

Department of Psychology, University of Minnesota,
Twin Cities, MN, USA



Christopher S. Kallie

Department of Psychology, The Ohio State University,
Columbus, OH, USA



Tingting Liu

Department of Ophthalmology and Vision Science,
Eye and ENT Hospital of Fudan University,
Shanghai, China



Gordon E. Legge

Department of Psychology, University of Minnesota,
Twin Cities, MN, USA



The visual span—the number of adjacent text letters that can be reliably recognized on one fixation—has been proposed as a sensory bottleneck that limits reading speed (Legge, Mansfield, & Chung, 2001). Like reading, searching for a face is an important daily task that involves pattern recognition. Is there a similar limitation on the number of faces that can be recognized in a single fixation? Here we report on a study in which we measured and compared the visual-span profiles for letter and face recognition. A serial two-stage model for pattern recognition was developed to interpret the data. The first stage is characterized by factors limiting recognition of isolated letters or faces, and the second stage represents the interfering effect of nearby stimuli on recognition. Our findings show that the visual span for faces is smaller than that for letters. Surprisingly, however, when differences in first-stage processing for letters and faces are accounted for, the two visual spans become nearly identical. These results suggest that the concept of visual span may describe a common sensory bottleneck that underlies different types of pattern recognition.

Introduction

Only a few letters of text can be recognized accurately on each fixation in reading, a limitation known as the *visual span* (Tinker, 1955; O'Regan, Lévy-Schoen, & Jacobs, 1983; Legge, Ahn, Klitz, & Luebker, 1997; Legge, Mansfield, & Chung, 2001). The purpose of the present article is to determine whether a similar constraint applies to face recognition and, if so, how the size of the visual span for reading relates to the size of the visual span for face recognition.

In previous articles we defined the visual span for reading as the number of letters, arranged side by side as in text, that can be recognized accurately without moving the eyes. In normal central vision, the visual span is about 10 characters; approximately five characters can be recognized both left and right of fixation at or above 80% accuracy (cf. Legge et al., 2001). Recognition accuracy declines rapidly for letters farther from fixation. The visual span can be pictured as a small window for reliable letter recognition in the visual field. Reading involves moving this window through text, relying on either eye movements or automated text presentation such as rapid serial visual presentation.

Citation: He, Y., Scholz, J. M., Gage, R., Kallie, C. S., Liu, T., & Legge, G. E. (2015). Comparing the visual spans for faces and letters. *Journal of Vision*, 15(8):7, 1–13, doi:10.1167/15.8.7.

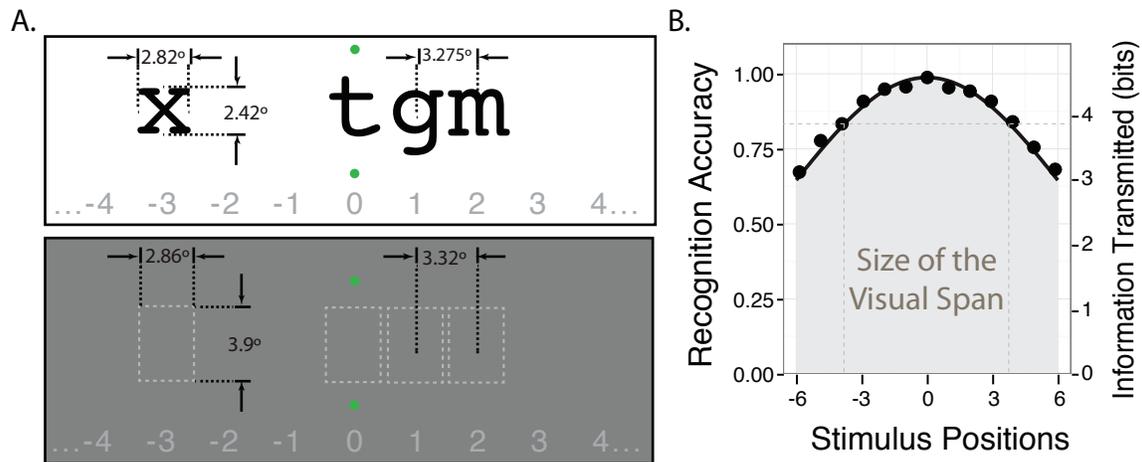


Figure 1. Sample visual span and schematic stimuli. (A) Schematic stimuli. The subjects needed to fixate between the two green dots and identify the faces or letters. A single letter and a single face are presented at slot -3 , and a trigram and a triface are centered at slot 1 . The light gray boxes indicate the bounding boxes for the faces (not visible to the subjects). The actual face stimuli are not shown here. The light gray numbers indicate the location of the slots and were not shown on the screen during the test. (B) Sample visual span. Dashed lines indicate the width of the visual span determined by an 80% recognition accuracy.

Several studies from our laboratory have shown a high correlation between the size of the visual span and reading speed when text parameters are varied, including character size and contrast (Legge et al., 2007), letter spacing (Yu, Cheung, Legge, & Chung, 2007), and retinal eccentricity (Legge et al., 2001). These results support the hypothesis that the size of the visual span is an important factor limiting reading speed.

In a typical visual-span measurement, subjects are required to recognize random strings of three letters called trigrams (Figure 1). We used strings of letters rather than isolated letters because of their closer approximation to fragments of English text. We measured recognition for trigrams at different horizontal locations, with position indicated by the number of letter slots left or right of the midline. In a trial, a trigram was presented very briefly, often 100 ms. Subjects were required to report all three letters in left-to-right order. Across a block of trials, percentage correct was accumulated for each letter slot. We refer to the resulting plot of letter accuracy versus letter position as a *visual-span profile*. We quantified the size of the visual span either as the breadth of the profile at a specified performance level (e.g., 80% correct) or by computing the area under the profile.

Legge (2007, chapter 3) identified three sensory factors that might limit the size of the visual span for reading: decreasing letter acuity away from fixation, crowding between adjacent letters, and errors in the encoding of the spatial order of letters (sometimes termed *mislocations*). Subsequent studies have shown that crowding is the major constraint (Pelli et al., 2007; He, Legge, & Yu, 2013; Wang, He, & Legge, 2014; Yu,

Legge, Wagoner, & Chung, 2014), with mislocations playing a lesser role and acuity playing a minimal role.

If the visual span limits the presymbolic sensory information for reading, we would expect that the factors determining its size would also affect patterns other than letters. A prediction follows: Similar visual-span constraints should produce a window within the visual field for recognizing faces and other types of patterns. Näsänen, Ojanpää, and Kojo (2001) and Ojanpää (2006) used this generalization in a study of search times and saccade sizes during eye-movement search through arrays of letters, words, faces, and other simple graphics. By studying search properties as a function of stimulus attributes such as contrast, they suggested that bottom-up constraints, summarized by a two-dimensional visual span, are important determinants of search behavior.

How might we expect the concept of visual span to generalize from alphabet letters to faces and other patterns? One possibility is that faces may have a smaller visual span than letters but be qualitatively similar. In a recent study, Wang et al. (2014) showed that Chinese characters have smaller visual spans than alphabet letters. They used a measure of pattern complexity (termed *perimetric complexity*, as defined by Attneave & Arnoult, 1956; Pelli, Burns, Farell, & Moore-Page, 2006) to categorize groups of Chinese characters. Using the trigram method for measuring visual-span profiles, they found that the profiles decreased in size as complexity increased. Their results may indicate that the size of the visual span is inversely related to pattern complexity.

Although perimetric complexity is not a well-defined measure for the complexity of grayscale images such as faces (Watson & Ahumada, 2012), intuition suggests

that faces are more complex patterns than alphabet letters. If so, we would expect the visual span for face recognition to be narrower than the visual span for alphabet letters—that is, fewer adjacent faces should be recognizable without moving the eyes.

In support of the concept of greater pattern complexity for faces, we know that the minimum spatial-frequency requirements for recognition are greater for faces than for letters. Kwon and Legge (2011) used low-pass spatial-frequency filtering to find the minimum spatial-frequency cutoffs yielding 80% correct for sets of 26 letters and 26 familiar faces. They found cutoff spatial frequencies of 0.9 and 1.14 cycles/letter for lowercase and uppercase letters, respectively, in central vision and 2.59 and 4.23 cycles/face for faces with and without hair, respectively.

Findings on size scaling in peripheral vision also support the expectation of a smaller visual span for faces than letters. Strasburger, Rentschler, and Jüttner (2011) reviewed studies showing that performance on many perceptual tasks decreases with increased eccentricity in peripheral vision. For both letter and face recognition, this decreased performance can be compensated for by appropriately scaling the size and contrast of the stimuli according to the eccentricity, but the scaling factor is larger for faces than for letters (Melmoth, Kukkonen, Mäkelä, & Rovamo, 2000; Melmoth & Rovamo, 2003). From this, we might expect that visual-span profiles for faces would be narrower than profiles for letters—that is, recognition performance would decline more rapidly with increasing eccentricity.

Alternatively, since letters and faces are thought to rely on different mechanisms for pattern recognition, it may be hard to predict the size of the face visual span. Faces may be recognized holistically (e.g., Farah, 1991; Tanaka & Farah, 1993; Gauthier & Tarr, 2002), while letters may be recognized based on simple features (e.g., Pelli et al., 2006). This difference is also reflected in the influence of crowding on these two types of stimuli. Within a face, the facial features crowd each other, similar to the crowding between letters in a word (Martelli, Majaj, & Pelli, 2005). There appears to be an additional stage of crowding between the holistic representations of faces (Louie, Bressler, & Whitney, 2007; Farzin, Rivera, & Whitney, 2009). These differences in pattern recognition could lead to differences in the sizes of the visual spans for these two types of patterns or even to qualitative differences.

In order to compare visual spans for letters and faces, we adapted the trigram method to measure visual-span profiles for faces. Based on the narrower visual spans for Chinese characters (Wang et al., 2014), the presumptive greater complexity of faces than of letters, and the additional holistic crowding between faces, we expected to find narrower visual spans for faces.

Method

Subjects

Ten normally sighted college students (five males and five females) at the University of Minnesota participated in Experiment 1. Their binocular distance visual acuities (Lighthouse Distance Visual Acuity Chart) ranged from -0.1 to 0.06 logarithm of the minimum angle of resolution units (logMAR), mean = -0.052 . Five of these subjects (one male and four females) also participated in Experiment 2. Prior to the experiments, all subjects signed a consent form approved by the institutional review board (IRB) at the University of Minnesota.

Stimuli

The stimuli were generated and presented using MATLAB R2010a with Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and displayed on an NEC MultiSync cathode ray tube monitor (model FP2141SB-BK, NEC, Tokyo, Japan); refresh rate = 100 Hz; resolution = 1280×1024 controlled by a Mac Pro Quad-Core computer (model A1186, Apple Inc., Cupertino, CA). Viewing was binocular at a distance of 40 cm maintained with a chin rest.

Letter stimuli were single lowercase black letters or trigrams presented on a white background (luminance = 102 cd/m^2 ; Weber contrast = 98%) in the Courier font (Figure 1). The x -height of the letters was 2.42° and center-to-center spacing was 3.27° ($1.16 \times x$ -width), representing standard spacing for the Courier font.

The face stimuli were 26 grayscale images presented on a gray background (background luminance = 37.8 cd/m^2 ; RMS (root-mean-square) contrast range = 0.124 – 0.231 with a mean of 0.175). The selection and description of the faces are given in more detail by Kwon and Legge (2011). Briefly, they were images of 13 female and 13 male celebrities selected from the Google image database. The faces included hair and were all smiling and viewed from the front. Selection was made to exclude conspicuous external features such as glasses, beards, and hair accessories.

The face stimuli were presented for recognition either singly (one face/trial) or as a triface—that is, a horizontal row of three faces, analogous to the letter trigrams (Figure 1). All of the 26 faces were scaled in size to fit in a box of $2.86^\circ \times 3.9^\circ$ at the 40-cm viewing distance. The center-to-center spacing of the boxes was 3.32° , and the boundaries of the boxes were not visible to the subjects. Face size and spacing were chosen to be similar to the letter stimuli. All subjects were shown the set of 26 faces prior to the experiment to confirm

familiarity and to inform the subjects about the set of possible faces for recognition.

Visual-span measurement

Figure 1 illustrates the visual-span measurement. Subjects were asked to fixate between two vertically aligned green dots and recognize single stimuli (letters or faces) or triplets of horizontally adjacent stimuli (trigrams or trifaces). In each trial, a stimulus was presented for 250 ms. Single stimuli were presented at 17 evenly spaced slots along a horizontal line. The slot centered between the green dots was labeled position 0. The left slots were labeled with negative numbers (–8 to –1), and the right slots were labeled with positive numbers (1 to 8). Trigrams or trifaces were centered at slots –7 to 7, but the outermost stimuli would also fall on slots ± 8 . Since the innermost stimuli would never appear at slots ± 7 and ± 8 and the middle stimuli would never appear at slots ± 8 , we analyzed the data only from slots –6 to 6 to ensure an equal number of trials across positions.

For trials with single stimuli, the subjects responded by naming the letter or face and were required to give one of the 26 alternatives. For trials with triplet stimuli, we used two reporting methods. In the full report (used with letters only), the subjects named the three stimuli in left-to-right order. In the partial report, the subject was directed to name only one of the three stimuli—left, middle, or right. For face recognition tasks, the 26 celebrity names were displayed on the screen after each trial until the subject made a choice. The experimenter recorded the responses without providing feedback about correctness. A webcam was used to monitor the eye movement of the subjects except for the last three subjects (subjects 8, 9, and 10).

Experimental design

There were two experiments. In Experiment 1, the goal was to compare visual-span profiles for single letters, trigrams, single faces, and trifaces. For single letters and single faces, two blocks of visual-span measurements were conducted. In each block, there were 15 trials at each of the 17 slots (totaling 30 letters or faces/slot). For trigrams and trifaces, each block consisted of 150 trials, with the stimuli centered on each of the 15 slots 10 times. The trigram visual-span profiles were measured in one block using full report (totaling 30 letters/slot from slots –6 to 6), in keeping with previous measurements of visual spans for reading (Legge et al., 2001). The triface visual spans were measured with partial report. In pilot experiments, we attempted full report with the trifaces, but our subjects found it

difficult to keep their three responses in mind while consulting memory for the set of 26 alternative faces. Therefore, the subjects were tested with partial report in six blocks (totaling 60 faces/slot from slots –6 to 6) and were instructed to respond to only the left, middle, or right face in a given block. The sequence of reporting left, middle, and right face was counterbalanced.

In Experiment 2, the goal was to investigate the difference in trigram visual-span profiles for the usual full-report method and the partial-report method. We were interested in this issue because we wanted to determine if any of the difference in letter and face visual spans observed in Experiment 1 could be attributed to the difference in reporting method. In Experiment 2, five of the 10 subjects in Experiment 1 were tested on the trigram recognition task with partial report. The procedure was similar to the triface test in Experiment 1 (totaling 60 letters/slot from slots –6 to 6).

Data analysis

To quantify the size of the visual span for further comparison, we computed the area under the curve in terms of information transmitted (Figure 1B, right axis). The information transmitted in bits has a linear relationship with recognition accuracy, where 100% correct corresponds to 4.7 bits of information and chance-level performance corresponds to no information. The following formula was used to convert recognition accuracy to information transmitted:

Information transmitted in bits

$$= -0.036996 + 4.6761 \times \text{Recognition accuracy.} \quad (1)$$

This linear function was fitted from empirically measured confusion matrices for letter recognition (Beckmann, 1998; see Legge et al., 2001, for derivation). For the 26 faces, we applied the same formula as a linear approximation.

To evaluate the effect of experimental manipulations (e.g., faces vs. letters) on the size of the visual span (in number of bits), we used linear mixed effects (LME) models (Pinheiro & Bates, 2000) to fit our data. We modeled experimental manipulations as fixed effects and subject differences as random effects. For example, to compare the visual spans for single faces and single letters, a simple one-factor LME model would be described as follows:

$$y_{ij} = \beta_j + b_i + b_{ij} + \varepsilon_{ij}, i = 1, \dots, 10, \quad j = 1, 2,$$

$$b_i \sim N(0, \sigma_1^2), b_{ij} \sim N(0, \sigma_2^2), \varepsilon_{ij} \sim N(0, \sigma^2), \quad (2)$$

where y_{ij} is the size of the visual span (in bits) for the i th subject under the j th condition, β_j represents the mean size of the visual span for the j th type of stimulus ($j = 1$

for letters and $j = 2$ for faces), b_i represents the deviation from the group mean β_j for the i th subject, b_{ij} represents the effect of stimulus type within the i th subject, and ε reflects the noise in the measurement. In this case, the fixed effect for stimulus type would be an estimation of the group mean β_j , and the two levels of random effects would be b_j and b_{ij} , which are both assumed to follow independent normal distributions with a mean of 0.

In the Results section, if not otherwise specified, all the numerical results were extracted from LME models. We report the group mean visual spans estimated by the model for each condition as well as the difference between conditions and the 95% confidence interval (CI) for these differences.

In addition to the bits measure, we estimated the size of the visual span as the number of stimulus (letters or faces) positions between the 80% correct points on visual-span profiles (dashed lines in Figure 1B). To do this, we fit the group visual-span profiles with a split-Gaussian function (Legge et al., 2001) using nonlinear least squares regression:

$$y = \begin{cases} Ae^{-(x^2/2\sigma_L^2)}, & x < 0 \\ Ae^{-(x^2/2\sigma_R^2)}, & x \geq 0 \end{cases} \quad (3)$$

where y is the letter-recognition accuracy at position x , A characterizes the peak accuracy, and σ_L and σ_R are the standard deviations of Gaussians used to fit the data on the left and right sides of the peak, respectively. This measure allows us to express the size of visual spans as the number of adjacent letters or faces that can be recognized with high reliability.

Two-stage model

In addition to the LME analysis, we used a two-stage model to interpret our data comparing the visual spans for letters and faces. At the first stage, recognition is limited by factors that affect the processing of individual stimuli, whether presented in isolation or flanked by other stimuli. These factors may include visual acuity, contrast sensitivity, or within-stimulus crowding for faces. The recognition errors for isolated stimuli reflect the impact of these factors. For simplicity, we refer to this impact as an *acuity effect* but return to consider the likely underlying factors in the Discussion. The second stage represents the interfering effects of nearby stimuli on pattern recognition, which occurs only in the presence of flankers as in the trigram and triface stimuli. We analyzed our data in order to estimate the reliabilities of the two stages—that is, the probability that information for a correct response would be transmitted through the stage. The reliability of the two stages, R_1 and R_2 , can be computed if we assume that the two stages are serial and independent.

In that case, the measured probability correct for an isolated target would be equal to R_1 , and the probability correct for a flanked target would be $R_1 \times R_2$. The data can then be compared to yield separate estimates for R_1 and R_2 . For example, if the recognition accuracy is 90% for an isolated letter and 70% for a letter in a trigram, then

$$R_1 = 0.9 \quad (4)$$

and

$$R_1 \times R_2 = 0.7. \quad (5)$$

From here, R_2 can be computed as $0.7/0.9 = 0.78$. The value of R_2 can be interpreted as the letter recognition accuracy when first-stage processing is not a limiting factor (i.e., the reliability of the first stage is 100%). In this way we can remove the influence of the acuity effect from the visual span and estimate the influence of crowding.

Results

Experiment 1

Figure 2 shows the visual-span profiles for single letters, single faces, trigrams, and trifaces. The profile for single letters is almost at ceiling—that is, recognition accuracy is greater than 98% across the entire range of letter positions tested. The profile for single faces is narrower than that for single letters (Figure 2A). The recognition accuracy for faces decreases faster than that for letters for stimuli farther from the midline. For trigrams and trifaces, the profiles nearly overlap for positions near the midline but start to separate for more peripheral positions, from slot -5 on the left and slot 4 on the right (Figure 2B, gray lines). The profiles for most of the individuals resemble the average profile for the group. Although subject 10 appears to be an outlier, we retained the data in our analysis for two reasons. First, the results of our LME analyses were the same qualitatively whether or not we included this subject. Second, it is possible that this subject has exceptional peripheral vision.

According to the LME analysis, the group mean visual-span sizes in bits were as follows: single letters = 59.9, single faces = 52.9, trigrams = 51.4, and trifaces = 47.5. According to the LME model, visual spans for faces are smaller than visual spans for letters by 7.0 bits, 95% CI [4.1, 9.9], which is reflected in the comparison between single-letter and single-face spans. The effect of adding flanking stimuli (to produce trigrams and trifaces) is a reduction in visual-span size by 8.5 bits, 95% CI [5.8, 11.2], which is reflected in the comparison between single-letter and trigram spans. There is also an

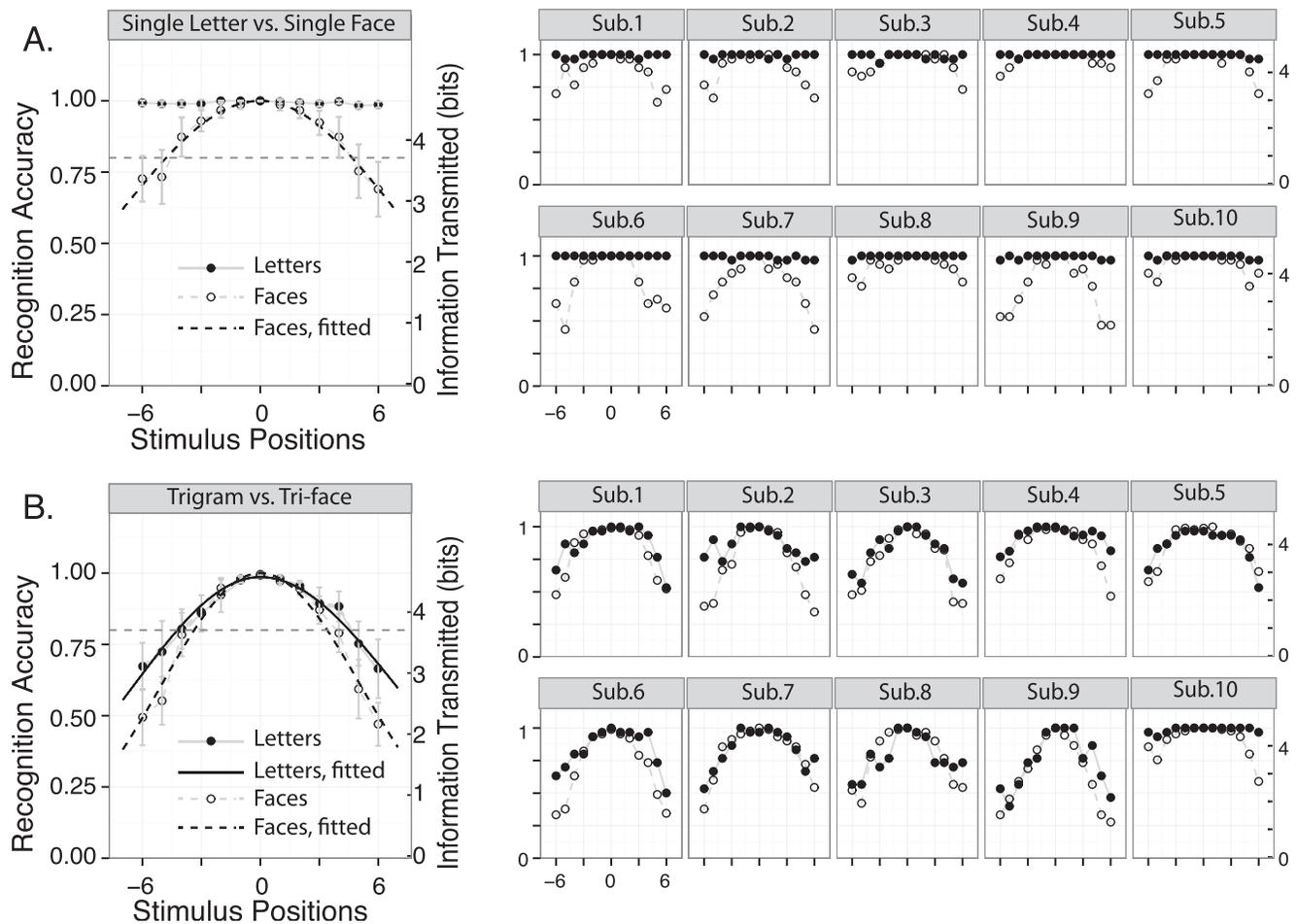


Figure 2. Results from Experiment 1. (A) Visual-span profiles for isolated letters and faces: group and individual data. (B) Visual-span profiles for trigrams and trifaces. Solid circles and solid gray line = letters (raw data); open circles and dashed gray line = faces (raw data); solid black line = fitted split-Gaussian curve for letters; dashed black line = fitted split-Gaussian curve for faces. Error bars indicate ± 1 SEM.

interaction between stimulus type and flanker type: The difference between letter and face spans is smaller when the flankers are present than when the flankers are absent. This interaction is 3.1 bits, 95% CI [0.1, 6.0], which explains why the difference between trigram and triface spans is only 3.9 rather than 7.0 bits. Another way to interpret the interaction is that adding flankers has a smaller influence on the size of the face-span size than the letter span.

From Figure 2A, the acuity effect (the decline in single-target performance away from the midline) is much stronger for faces than for letters. Does this difference in the acuity effect alone account for the difference between the trigram and triface visual-span profiles? To answer this question, we examined the second-stage reliabilities (R_2 ; see Method) associated with between-stimuli crowding for letters and faces. The corresponding derived profiles, depicting the reliabilities for the second stage, are shown in Figure 3. Due to the within-subject variability across measure-

ments, it could occasionally happen that the measured performance for a triplet of stimuli is better than that for isolated stimuli. In such cases, the derived R_2 value will be larger than the theoretical ceiling value of 1.0 (e.g., Figure 3, subject 1, open circle at position -4). But these variations are caused by pure chance and average out across subjects. Therefore, we retained these individual R_2 estimates rather than capping them at 1.0 when computing the group mean in order to not underestimate the true R_2 value.

Surprisingly, the second-stage reliability profile for the trifaces is higher than the trigram profile at most of the positions (Figure 3), implying less influence from crowding for faces than for letters. If we apply a similar linear conversion from reliability to information transmitted (Equation 1), the difference in the overall size is small (triface = 53.6 bits, trigram = 51.7 bits), but there is a trend for the face span to be larger than the letter span: estimated difference = 1.8 bits, 95% CI $[-1.6, 5.2]$.

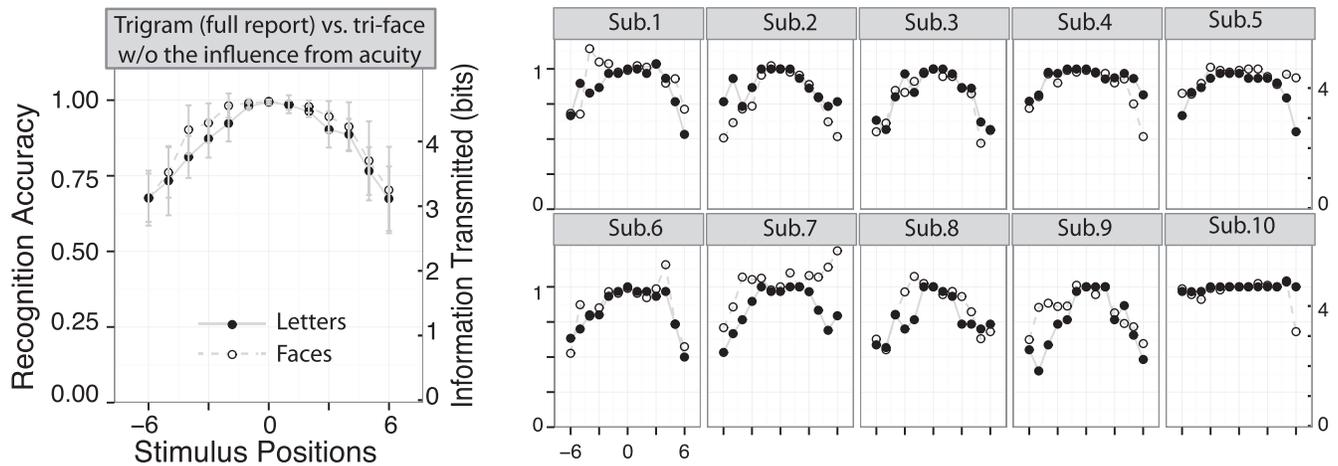


Figure 3. Profiles of second-stage reliability (R_2) for trigrams and trifaces. Solid circles and solid line = trifaces; open circles and dashed line = trigrams (full report).

Although the second-stage profiles in Figure 3 are assumed to be free of the first-stage acuity constraints, the comparison between the triface and trigram profiles might still be influenced by the different response requirements for the two types of stimuli: The subjects reported all three letters of the trigrams (full report) but only a single face in a designated position in the trifaces (partial report). To assess the effect of this difference, in Experiment 2 we compared trigram visual spans obtained with partial and full report.

Experiment 2

Figure 4A compares the trigram visual-span profiles for partial and full report. The full-report profile lies below the partial-report profile, especially left of fixation. For the full-report profile, the left side is narrower than the right side, but this asymmetry disappears when using partial report. This pattern holds for three of the five subjects. The difference in size between profiles measured with full and partial report is small but notable: full report = 51.2 bits, partial report = 53.6 bits; estimated difference (partial – full) = 2.4 bits; 95% CI [–1.7, 6.5]. When we used the partial-report trigram data to examine the effects of stimulus type and flanker presence as in Experiment 1, the interaction was 0.5 bits, 95% CI [–3.4, 4.4] bits, suggesting that when reporting methods were equalized, the presence of flankers imposed similar effects on the size of letter and face spans.

Using the method described earlier, we derived the second-stage reliability profiles for the trigram partial-report data and compared them with the second-stage profiles for the trifaces (Figure 4B). The resulting second-stage profiles for trigrams and trifaces overlap each other at most of the stimulus positions except for the tails on the right. The overall sizes of the second-

stage profiles were very similar: triface = 52.1 bits, trigram = 54.0 bits; estimated difference = 1.9 bits; 95% CI [–4.4, 5.5] bits. The confidence interval of the difference is nearly symmetrical around 0, suggesting that the two visual spans have nearly identical sizes.

To summarize the results from the two experiments, (a) recognition of single faces declined more rapidly than recognition of single letters in peripheral vision, (b) partial-report measurement yielded larger visual spans than did full report, (c) the difference between triface and trigram visual spans appeared to be primarily due to first-stage constraints, and (d) derived profiles for a presumed second-stage limitation due to between-stimuli crowding yielded nearly identical effects for trifaces and trigrams.

Size of visual span in number of letters or faces

The LME analysis provided estimates of visual-span size in terms of the area under the visual-span profiles (transformed to bits of information transmitted). It is also interesting to compute the width of the profiles in terms of the number of letters or faces exceeding an accuracy criterion (e.g., 80% correct). We used the split-Gaussian fits (see Method; Equation 3) to compute these widths. Table 1 lists the three parameters of the split-Gaussian fits: peak amplitude A and standard deviations σ_L and σ_R (to the left and right of the peak, respectively). The table also shows the number of stimuli left (VS_L) and right (VS_R) of the fixated stimulus that exceeded the 80% criterion. Finally, the table shows the total width of the visual span (VS_T), which is the number of stimuli to the left plus the number of stimuli to the right plus the single stimulus at fixation. For example, at the 80% criterion, the group trigram profile in Experiment 1 (Figure 2B, solid black curve) has 4.2 letters to the left, 4.5 letters to the

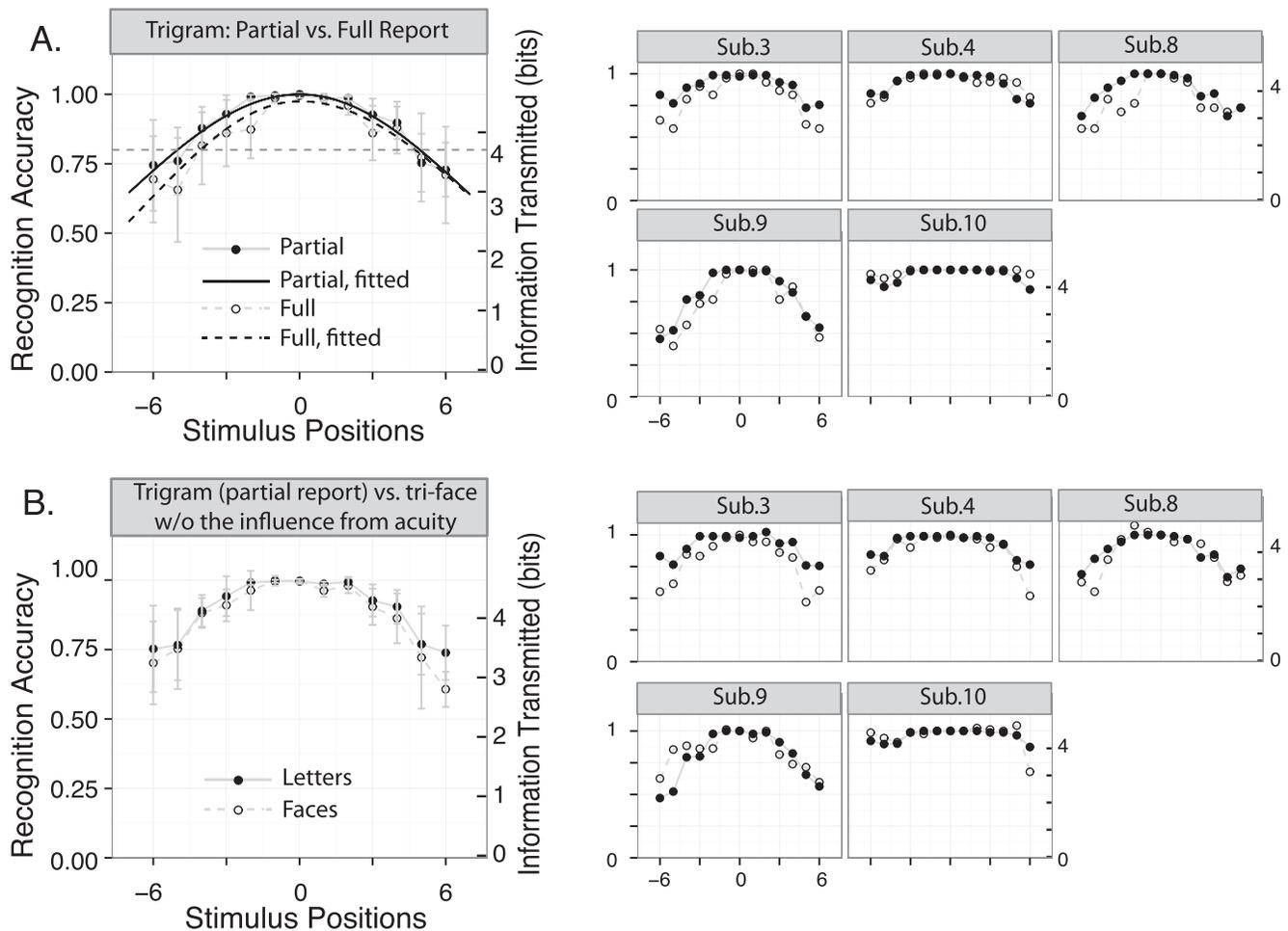


Figure 4. Results from Experiment 2. (A) Trigram visual-span profiles using partial versus full report. Solid line = partial report; dashed line = full report. Solid circles and solid gray line = partial report (raw data); open circles and dashed gray line = full report (raw data); solid black line = fitted split-Gaussian curve for partial report; dashed black line = fitted split-Gaussian curve for full report. (B) Profiles of second-stage reliability (R_2) for trigrams and triface, both using partial report. Solid circles and solid line = triface; open circles and dashed line = trigrams.

right, and one letter in the middle for a visual-span size of 9.7 letters. The triface profile in Experiment 1 (Figure 2B, dashed black curve) has a corresponding visual-span width of 7.8 faces.

Consistent with previous findings (Legge et al., 2001), the table also reveals a small left–right asymmetry in the shape of the trigram visual span. The profile is slightly wider to the right than to the left. The group profile for faces in Experiment 1 does not exhibit this asymmetry. The small asymmetry in the trigram visual span observed with full report disappeared in partial report (Table 1, Experiment 2).

Discussion

In this study, we generalized the concept of visual span to the recognition of faces. We measured the

pattern recognition performance for letters and faces and compared the resulting visual-span profiles. We found a smaller visual span for faces than for letters. Using our two-stage model, the difference between faces and letters resides mainly in the first stage, which reflects the processing of individual stimuli. When differences in their first-stage processing are taken into account and the reporting methods are equated (partial report for both types of stimuli), the visual-span profiles for letters and faces are surprisingly similar.

Why does face recognition decline more quickly with eccentricity than letter recognition?

We have found that the major difference in visual spans for letters and faces resides at the first stage in our model—that is, in the recognition of individual stimuli. As shown in Figure 2A, letter recognition

Condition	Fitted parameter (<i>SE</i>)			Visual span size		
	<i>A</i>	σ_L	σ_R	VS_L	VS_R	VS_T
Experiment 1 (10 subjects)						
Single face	1.00 (0.013)	7.15 (0.38)	6.99 (0.36)	4.8	4.7	10.4
Triface, partial report	1.00 (0.015)	5.04 (0.18)	5.10 (0.18)	3.4	3.4	7.8
Trigram, full report	0.99 (0.014)	6.53 (0.032)	6.98 (0.38)	4.2	4.5	9.7
Experiment 2 (five subjects)						
Single face	1.00 (0.020)	7.94 (0.73)	8.01 (0.74)	5.3	5.3	11.6
Triface, partial report	1.00 (0.023)	5.28 (0.31)	4.99 (0.27)	3.5	3.3	7.8
Trigram, full report	0.97 (0.025)	6.46 (0.56)	7.54 (0.83)	4.1	4.7	9.8
Trigram, partial report	1.00 (0.016)	7.49 (0.52)	7.39 (0.50)	5.0	4.9	10.9

Table 1. Curve fitting results using nonlinear least squares regression and visual-span sizes represented in number of symbols. The fitted parameters for the split-Gaussian curves are peak amplitude A and standard deviations σ_L and σ_R (to the left and right of the peak, respectively). Values in parentheses are estimated standard errors of the fitted parameters. VS_L (VS_R) = number of symbols that can be recognized to the left (right) of the midline (not including the symbol on the midline); VS_T = total number of recognizable symbols (including the symbol on the midline). The fittings for single letters are not included here because performance was always near ceiling and the fitting would be unreliable.

remains close to the ceiling value of 100% correct for the range of letter positions tested in this study. Across a similar range of eccentricities, recognition accuracy for single faces declined from near 100% to about 70% at six letter positions left and right of the midline. In the following paragraphs, we briefly consider acuity, within-face crowding, and memory as potential contributors to this difference.

Acuity could account for the steeper decline for faces if face recognition is limited by the acuity for individual parts of faces (e.g., eyes, nose, mouth) or the ability to discriminate the separations between parts of faces. In our experiments, the faces were approximately matched in size with the letters. As a result, the face parts were smaller than the letters and would reach their acuity limit at a smaller eccentricity. We can make this argument more concrete. Letter acuity increases linearly with eccentricity up to 30° (Anstis, 1974), and the acuity size S at eccentricity E can be described by

$$S = S_0(1 + E/E_2), \quad (6)$$

where S_0 is the acuity size at the fovea and E_2 is the eccentricity at which the acuity size doubles compared with the fovea (Levi, Klein, & Aitsebaomo, 1984). For high-contrast letters, typically $S_0 = 0.083^\circ$ and $E_2 = 1.5^\circ$ (Legge, 2007). Applying this formula to letters of an x -height of 2.42° used in the present study, the acuity limit would not be encountered until such letters reached an eccentricity well beyond 30° —much farther into the periphery than our letter stimuli. Accordingly, it is not surprising that our subjects performed close to 100% for isolated letter recognition at all tested locations. Now suppose that Equation 6 applies to parts of faces, such as the separation between the eyes. The average interpupil distance of our face stimuli is 1° . According to Equation 6, letter stimuli subtending 1° would reach the acuity limit at about 16.6° from the

fovea. In our experiment, this eccentricity corresponds to the fifth stimulus position left or right of fixation. This analysis indicates that if face recognition is limited by acuity for resolving eye separation or other similar configural features of faces and if this form of acuity follows a rule similar to Equation 6, it is plausible that face recognition would show a steeper decline with eccentricity than letters when faces and letters are matched for angular size.

A second factor limiting recognition of individual faces in peripheral vision is crowding between the parts of faces (Martelli et al., 2005). This within-face crowding may account for the steeper decline we observed for faces than for letters, but we think this is unlikely to be the main reason. The distance between facial features (e.g., interpupil distance) is smaller than the distance between letters. If intraface crowding limits recognition for single faces and interletter crowding limits letter recognition for trigrams, we would expect the visual span for single faces (Figure 2A, open circles) to be narrower than the visual span for letter trigrams (Figure 2B, solid circles). But the opposite is true: The visual span for single faces is wider than the visual span for trigrams (Table 1). This comparison may indicate that crowding is not the primary factor accounting for the steeper decline for recognizing single faces in peripheral vision.

A third factor potentially limiting peripheral recognition of faces is memory load for the set of stimulus alternatives. For both letters and faces, increased set size impairs recognition performance when eccentricity is large, but faces are influenced more than letters (Melmoth et al., 2000; Melmoth & Rovamo, 2003). We tried to minimize the memory load by familiarizing the subjects with the face set prior to the experiment and displaying the celebrity names after each presentation, but it is still possible that letters have an advantage over

faces due to their greater familiarity. To briefly summarize, while acuity, crowding, and memory load may all contribute to the steeper decline of peripheral face recognition compared with letter recognition, we suggest that acuity is the dominant factor.

Similar reliability of second-stage processing

The processing of the second stage, which reflects the influence of nearby stimuli, does not differ between faces and letters under our experimental conditions. This similarity is surprising for at least two reasons: Faces seem to be more complex stimuli than letters, and faces may experience feature-level crowding as well as structure-level crowding. We briefly discuss these two issues in the following paragraphs.

Complexity

Intuition suggests that faces are more complex stimuli than letters, and there is evidence for an effect of target complexity on crowding for Chinese characters, alphabet letters, and some simple nonletter symbols (Zhang, Zhang, Xue, Liu, & Yu, 2009; Bernard & Chung, 2011; Wang et al., 2014). Wang et al. (2014) measured the visual spans of sets of Chinese characters across a range of complexity levels using the perimetric complexity definition (Attneave & Arnoult, 1956; Pelli et al., 2006) as well as alphabetic letters. All of the Chinese character sets were more complex than the set of alphabet letters (lowercase Courier). Unlike our results for letters and faces, they found that the visual spans for their character sets decreased in size with increasing complexity. The differences could not be attributed to acuity. If more complex targets have narrower visual spans, why are the (second-stage) profiles for faces so similar to those for letters?

One issue in comparing the complexity of letters and faces is definitional. Perimetric complexity is well defined for letters or Chinese characters but not for grayscale images such as faces (Watson & Ahumada, 2012). Nevertheless, empirical findings imply that faces, like Chinese characters, are more complex patterns than alphabet letters. For instance, size scaling of stimuli in peripheral vision can equate recognition performance with central vision. The corresponding scaling factors increase with complexity for Chinese characters (Zhang et al., 2009) and are larger for faces than for letters (Melmoth et al., 2000; Melmoth & Rovamo, 2003). It has also been observed that the minimum spatial-frequency requirement for more complex Chinese characters is higher than that for simpler characters (Wang, He, & Legge, 2013) and higher for faces than for letters (Kwon & Legge, 2011). Consistent with these findings, we found that the effect

of acuity in first-stage processing is larger for faces than for letters (Figure 2A).

If it is granted that faces are more complex stimuli than letters, perhaps some other aspect of face images offsets the expected effect of complexity on crowding. As one possibility, Wallace and Tjan (2011) found that images of real objects are less influenced by crowding than letters. They proposed that in grayscale images, informative features are defined by local variations of contrast, which better survive crowding. Perhaps the grayscale nature of our face images made them less vulnerable to crowding than letters despite their presumptive greater complexity.

Structure-level crowding

More generally, face recognition may differ from the recognition of other categories of objects. In addition to feature-level crowding, there is evidence for another type of crowding between higher level, structural information in faces. For upright faces (both real images and Mooney faces), crowding by flanking faces is stronger when the flankers are upright rather than inverted (Louie et al., 2007; Farzin et al., 2009). On the other hand, houses or inverted faces do not show this inversion effect in crowding (Louie et al., 2007). It is unknown whether letters will suffer from this type of high-level crowding, but letters are typically believed to be recognized based on simple features rather than on more holistic properties (e.g., Pelli et al., 2006). In fact, a closer analogy to faces might be words, where the crowding between letters in a word resembles the crowding between parts in a face (Martelli et al., 2005). But even for words, the between-stimuli crowding is determined by low-level features but not by high-level holistic representations (Yu, Akau, & Chung, 2012). Surprisingly, this additional level of crowding did not make the second-stage profile for faces smaller than that for letters. As discussed previously, the reason might be that the grayscale nature of face images reduces the impact from crowding (Wallace & Tjan, 2011).

Equating visual-span profiles for faces and letters?

We have interpreted the difference in visual-span profiles for faces and letters as being primarily due to a difference in the processing of individual patterns (letters or faces) in peripheral vision. It might be possible to modify stimulus conditions so that the face and letter visual-span profiles are more nearly matched in shape—for instance, by reducing the contrast of the letters. If so, we would expect the peripheral decline in single-letter recognition to match the decline in single-face recognition. Legge et al. (2007) measured visual spans for low-contrast letters. At 3% to 4% contrast

(about twice the threshold value), the visual-span profiles for letters narrowed, resembling the face visual spans in the current study. Legge et al. (2007) did not measure the peripheral decline in recognition performance for single letters at this low contrast. But Strasburger et al. (2011) presented an empirically based formula describing the relationship between acuity size, contrast, and retinal eccentricity for digit characters in peripheral vision (their figure 17). Using their formula, we calculated that digit characters with 3% contrast and subtending 2.42° would reach acuity threshold at about 12.5° retinal eccentricity. In our study this eccentricity would correspond to four character positions left or right of the midline, about equal to the position at which face recognition begins to decline. This analysis shows that, consistent with our proposal, both visual-span profiles and acuity limitations in peripheral vision would be very similar for faces and low-contrast letters.

We acknowledge that our two-stage model may be oversimplified in its assumptions of serial and independent processing. Nevertheless, our results suggest that pattern recognition, regardless of the type of stimulus, has a common sensory bottleneck that severely limits the number of simultaneously recognizable targets.

Partial versus full report

In Experiment 2, we studied the influence of report type (partial vs. full) on the shape of the visual span for letters. Partial report yielded a larger visual span than full report in the trigram task (Figure 4A). The benefit associated with partial report may be related to changes in high-level processing, such as more focused visual attention or reduced load on working memory. Notably, this benefit becomes larger with increased recognition difficulty, a pattern also observed by Wang et al. (2014). Therefore, while the visual span depicts a limitation at the sensory level, its size may be influenced by the interaction between lower and higher level processes.

Real-life implications

It has been proposed that the size of the visual span for letter recognition imposes a limitation on reading speed in both normal vision (Legge et al., 2001, 2007; Pelli et al., 2007) and low vision (Cheong, Legge, Lawrence, Cheung, & Ruff, 2008). Intuitively, as the visual span gets smaller, fewer letters are recognized on each fixation, requiring shorter saccades and more fixations to step through text.

Similarly, the visual span for faces might be a sensory bottleneck limiting face-related tasks such as searching for a face in a crowd. Imagine, for example, that you are searching through a class photo to find a particular classmate. Several fixations, each revealing the identity of a few faces, might be required. A smaller visual span, due to viewing conditions or vision impairment, would likely result in a longer search.

Both reading and visual search typically involve eye movements. Oculomotor control in such tasks will typically involve factors in addition to visual span. McConkie and Rayner (1975) defined the perceptual span as the region around fixation in which printed information influences reading behavior. Operationally, it refers to the region of visual field that influences eye movements and fixation times in reading. For recent reviews and extension of the concept to tasks other than reading, see Ojanpää (2006) and Rayner (2009). For reading English text, the perceptual span extends 15 characters to the right of fixation and four characters to the left of fixation (McConkie & Rayner, 1975; Rayner, Well, & Pollatsek, 1980). For visual search tasks using two-dimensional arrays of symbols, the perceptual span was between 5×5 and 7×7 items for characters (Näsänen et al., 2001) but no larger than 3×3 for faces (Näsänen & Ojanpää, 2004). These results are qualitatively consistent with our finding that the visual span for faces is smaller than that for letters.

An important difference between our experimental procedure and most real-world visual-recognition behavior is the involvement of eye movements. Whereas our visual-span measurements require stable fixation, reading and finding a face in a crowd require frequent saccades. Recent studies suggest that making an eye movement toward a crowded target may enhance the recognition of the target (Harrison, Mattingly, & Remington, 2013; Lin et al., 2014). It may prove important to take the effects of eye movements into account in order to capture the dynamic changes of visual span in real-world tasks.

Keywords: face recognition, letter recognition, visual span, crowding, visual acuity

Acknowledgments

A preliminary report was given at the annual meeting of the Association for Research in Vision and Ophthalmology (Scholz, Legge, Liu, & Kallie, 2012). The authors thank MiYoung Kwon for the development of the set of face stimuli and the Statistical Clinic at the University of Minnesota for their help with the mixed effects modeling. This research was supported by NIH Grant EY002934 to Gordon Legge and a grant

from the University of Minnesota Undergraduate Research Opportunities Program to Jennifer Scholz.

Commercial relationships: none.

Corresponding author: Yingchen He.

Email: hexxx340@umn.edu.

Address: Department of Psychology, University of Minnesota, Twin Cities, MN, USA.

References

- Anstis, S. M. (1974). A chart demonstrating variations in acuity with retinal position. *Vision Research*, *14*, 389–392.
- Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin*, *53*, 452–471.
- Beckmann, P. J. (1998). *Preneural factors limiting letter identification in central and peripheral vision* (Doctoral dissertation). University of Minnesota, Minneapolis.
- Bernard, J.-B., & Chung, S. T. L. (2011). The dependence of crowding on flanker complexity and target-flanker similarity. *Journal of Vision*, *11*(8):1, 1–16, doi:10.1167/11.8.1. [PubMed] [Article]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.
- Cheong, A. M. Y., Legge, G. E., Lawrence, M. G., Cheung, S.-H., & Ruff, M. A. (2008). Relationship between visual span and reading performance in age-related macular degeneration. *Vision Research*, *48*, 577–588, doi:10.1016/j.visres.2007.11.022.
- Farah, M. J. (1991). Patterns of co-occurrence among the associative agnosias: Implications for visual object representation. *Cognitive Neuropsychology*, *8*(1), 1–19.
- Farzin, F., Rivera, S. M., & Whitney, D. (2009). Holistic crowding of Mooney faces. *Journal of Vision*, *9*(6):18, 1–15, doi:10.1167/9.6.18. [PubMed] [Article]
- Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 431–446, doi:10.1037//0096-1523.28.2.431.
- Harrison, W. J., Mattingley, J. B., & Remington, R. W. (2013). Eye movement targets are released from visual crowding. *The Journal of Neuroscience*, *33*, 2927–2933, doi:10.1523/JNEUROSCI.4172-12.2013.
- He, Y., Legge, G. E., & Yu, D. (2013). Sensory and cognitive influences on the training-related improvement of reading speed in peripheral vision. *Journal of Vision*, *13*(7):14, 1–14, doi:10.1167/13.7.14. [PubMed] [Article]
- Kwon, M., & Legge, G. E. (2011). Spatial-frequency cutoff requirements for pattern recognition in central and peripheral vision. *Vision Research*, *51*, 1995–2007, doi:10.1016/j.visres.2011.06.020.
- Legge, G. E. (2007). *Psychophysics of reading in normal and low vision*. Mahwah, NJ: Erlbaum.
- Legge, G. E., Ahn, S. J., Klitz, T. S., & Luebker, A. (1997). Psychophysics of reading—XVI. The visual span in normal and low vision. *Vision Research*, *37*, 1999–2010, doi:10.1016/S0042-6989(97)00017-5.
- Legge, G. E., Cheung, S.-H., Yu, D., Chung, S. T. L., Lee, H.-W., & Owens, D. P. (2007). The case for the visual span as a sensory bottleneck in reading. *Journal of Vision*, *7*(2):9, 1–15, doi:10.1167/7.2.9. [PubMed] [Article]
- Legge, G. E., Mansfield, J. S., & Chung, S. T. L. (2001). Psychophysics of reading. XX. Linking letter recognition to reading speed in central and peripheral vision. *Vision Research*, *41*, 725–743.
- Levi, D. M., Klein, S. A., & Aitsebaomo, P. (1984). Detection and discrimination of the direction of motion in central and peripheral vision of normal and amblyopic observers. *Vision Research*, *24*, 789–800, doi:10.1016/0042-6989(84)90150-0.
- Lin, H., Rizak, J. D., Ma, Y., Yang, S., Chen, L., & Hu, X. (2014). Face recognition increases during saccade preparation. *PLoS One*, *9*(3), e93112, doi:10.1371/journal.pone.0093112.
- Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: Selective interference between configural representations of faces in crowded scenes. *Journal of Vision*, *7*(2):24, 1–11, doi:10.1167/7.2.24. [PubMed] [Article]
- Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision*, *5*(1):6, 58–70, doi:10.1167/5.1.6. [PubMed] [Article]
- McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during fixations in reading. *Perception & Psychophysics*, *17*, 578–586.
- Melmoth, D. R., Kukkonen, H. T., Mäkelä, P. K., & Rovamo, J. M. (2000). The effect of contrast and size scaling on face perception in foveal and extrafoveal vision. *Investigative Ophthalmology and Visual Science*, *41*, 2811–2819. [PubMed] [Article]
- Melmoth, D. R., & Rovamo, J. M. (2003). Scaling of letter size and contrast equalises perception across

- eccentricities and set sizes. *Vision Research*, *43*, 769–777, doi:10.1016/S0042-6989(02)00685-5.
- Näsänen, R., & Ojanpää, H. (2004). How many faces can be processed during a single eye fixation? *Perception*, *33*, 67–77, doi:10.1068/p3417.
- Näsänen, R., Ojanpää, H., & Kojo, I. (2001). Effect of stimulus contrast on performance and eye movements in visual search. *Vision Research*, *41*, 1817–1824.
- O'Regan, J. K., Lévy-Schoen, A., & Jacobs, A. M. (1983). The effect of visibility on eye-movement parameters in reading. *Perception & Psychophysics*, *34*, 457–464.
- Ojanpää, H. (2006). *Visual search and eye movements: Studies of perceptual span* (Doctoral dissertation). University of Helsinki, Finland.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442, doi:10.1163/156856897X00366.
- Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, *46*, 4646–4674, doi:10.1016/j.visres.2006.04.023.
- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, *7*(2):20, 1–36, doi:10.1167/7.2.20. [PubMed] [Article]
- Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-PLUS (Statistics and computing)*. New York, NY: Springer.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*, 1457–1506, doi:10.1080/17470210902816461.
- Rayner, K., Well, A. D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics*, *27*, 537–544.
- Scholz, J. M., Legge, G. E., Liu, T., & Kallie, C. S. (2012, May). *Comparing the visual span for letters and faces*. Presentation at the Annual Meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL, USA.
- Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5):13, 1–82, doi:10.1167/11.5.13. [PubMed] [Article]
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*, 225–245, doi:10.1080/14640749308401045.
- Tinker, M. A. (1955). Perceptual and oculomotor efficiency in reading materials in vertical and horizontal arrangements. *The American Journal of Psychology*, *68*, 444–449.
- Wallace, J. M., & Tjan, B. S. (2011). Object crowding. *Journal of Vision*, *11*(6):19, 1–17, doi:10.1167/11.6.19. [PubMed] [Article]
- Wang, H., He, X., & Legge, G. E. (2013, Jul). Spatial frequency bandwidth requirements for recognizing Alphabetic and Chinese characters. Presentation at the 9th Asia-Pacific Conference on Vision, Suzhou, China.
- Wang, H., He, X., & Legge, G. E. (2014). Effect of pattern complexity on the visual span for Chinese and alphabet characters. *Journal of Vision*, *14*(8):6, 1–17, doi:10.1167/14.8.6. [PubMed] [Article]
- Watson, A. B., & Ahumada, A. J. (2012). Modeling acuity for optotypes varying in complexity. *Journal of Vision*, *12*(10):19, 1–19, doi:10.1167/12.10.19. [PubMed] [Article]
- Yu, D., Akau, M. M. U., & Chung, S. T. L. (2012). The mechanism of word crowding. *Vision Research*, *52*, 61–69, doi:10.1016/j.visres.2011.10.015.
- Yu, D., Cheung, S.-H., Legge, G. E., & Chung, S. T. L. (2007). Effect of letter spacing on visual span and reading speed. *Journal of Vision*, *7*(2):2, 1–10, doi:10.1167/7.2.2. [PubMed] [Article]
- Yu, D., Legge, G. E., Wagoner, G., & Chung, S. T. L. (2014). Sensory factors limiting horizontal and vertical visual span for letter recognition. *Journal of Vision*, *14*(6):3, 1–17, doi:10.1167/14.6.3. [PubMed] [Article]
- Zhang, J.-Y., Zhang, T., Xue, F., Liu, L., & Yu, C. (2009). Legibility of Chinese characters in peripheral vision and the top-down influences on crowding. *Vision Research*, *49*, 44–53, doi:10.1016/j.visres.2008.09.021.