Perception, 2006, volume 35, pages 541 – 559

DOI:10.1068/p5313

Font tuning associated with expertise in letter perception

Isabel Gauthier, Alan C-N Wong, William G Hayward¶, Olivia S Cheung
Department of Psychology, Vanderbilt University, Wilson Hall, Nashville, TN 37203, USA;
e-mail: isabel.gauthier@vanderbilt.edu; ¶ Department of Psychology, University of Hong Kong, Pokfulam Road, Hong Kong
Received 16 June 2004, in revised form 24 May 2005; published online 15 March 2006

Abstract. Font tuning (FT) occurs when observers recognize a sequence of letters presented in the same font faster than in different fonts (Sanocki 1987, 1988 *Journal of Experimental Psychology: Human Perception and Performance* **13** 267 – 278; **14** 472 – 480). Here, we test the hypothesis that FT is associated with expertise with a specific writing system. We developed a systematic search task allowing the measurement of FT over a large number of letters and generalized the finding of FT from English readers viewing Roman letters to Chinese readers viewing Chinese characters. Non-Chinese readers did not show evidence of FT for Chinese characters in this search task. We also used a simpler 3-letter identification task to directly compare novice and expert readers, and to explore FT for different aspects of font such as fill, slant, and aspect ratio. We found that experts tune to aspect ratio but not to the other font changes. These findings reveal that letters are not processed visually in the same manner as shapes, and suggest an explanation for the cortical specialization obtained in the visual system for letters.

1 Introduction

what is wrong with these words? They are written in a font called Hieronymous Boschian (The Chank Company, Minneapolis), with each letter created independently by a different artist during a fontmaking workshop at the University of Wisconsin-Stout in Menomonie, WI. The legibility of text printed in Hieronymous Boschian is reduced for at least two reasons. First, and perhaps most obvious, some of the letters are difficult to recognize (eg the w). A second reason lies in the lack of predictability from one letter to the next. The letters of 'good' fonts, those made by master calligraphers and used in books or that become standard on our computers, all share a common style. Indeed, it seems that one could predict the appearance of the subsequent letters after having seen the first 16 (Hofstadter and McGraw 1995). Here we investigate how such regularity in style for letters is used by expert readers, and whether it also influences the processing of letters of an unfamiliar alphabet.(1)

Previous research revealed that regularity in font information facilitates letter recognition in fluent readers. With a number of different tasks, Sanocki (1987a, 1987b, 1988, 1991) studied letter recognition in regular-font (all the letters in the same font) or mixed-font non-word strings (font alternating from one letter to the next). Letter recognition proceeded more efficiently in regular-font than in mixed-font strings, indicating that font information irrelevant to the task is nonetheless encoded by observers. We will call this phenomenon 'font tuning' (FT). Sanocki (1988) proposed that parameter values could be extracted for the font, which could facilitate letter perception because perceptual interpretations inconsistent with the expected font do not need to be considered. While this work reveals that FT can facilitate the recognition of letters, it is unknown whether it is one manifestation of a general phenomenon that applies to all shape perception, or whether it is specific to letters of a familiar alphabet;

(1) Note that we will use 'letter recognition' here to stand for the recognition of individual characters of a writing system, be it alphabetic or not.

more precisely, specific to a category where the regularities are available *and* the observers have enough experience to expect them.

We explore the possibility that FT is one hallmark of perceptual expertise with letters. Because print is almost always regular in size, orientation, and font, expert readers may have developed a recognition strategy that takes advantage of this situation. In other words, an important difference between letter and object recognition is that we rarely recognize a sequence of objects that share a common 'style'. For example, encoding the style of a chair is unlikely to help us to recognize the next object we happen to land our eyes on, be it a computer, a dog, or a plant. While it is possible to imagine grouping objects by style (clothes, furniture, art, breeds of dogs), it is hard to imagine that one could expect the same amount of regularity in these situations as with letters. Thus, regularities in style constrain letter perception in a specific way that may represent an important difference in our experience with letters versus object/shape processing.

Apart from Sanocki's work on this topic, the use of font information has been largely ignored by existing theories of letter and word recognition. Some models deal with font by using features specific to a given font (Keren and Baggen 1981; Townsend and Ashby 1982), limiting their generalization from one font to another. Other models assume that letters are recognized as sets of abstract features, those that are invariant across most fonts (eg '/', '-', and '\' = **A**) (Estes 1978; Gibson 1969; Gibson and Levin 1973; McClelland and Rumelhart 1981; Oden 1979; Schneider and Shiffrin 1977). In that case, the information specific to each font or typeface is simply noise to the process of letter recognition (McClelland and Elman 1986; Morton 1969; Oden and Massaro 1978; Studdert-Kennedy 1976). The idea that font is irrelevant to letter recognition is similar to abstractionist models of speech perception which assume that voice details are merely 'noise' in the process of spoken word recognition. However, several studies indicate slowed and less accurate performance at word identification for lists of words spoken by multiple speakers compared to a single one (Creelman 1957; Mullennix and Pisoni 1990; Mullennix et al 1989; Palmeri et al 1993). Similarly, words produced at varied speaking rates are identified and recalled more poorly than words produced at a single rate (Nygaard et al 1995; Sommers et al 1994). These are examples of how the perceptual system may incidentally use regularities in the flow of information in a domain of expertise.

Although FT has been demonstrated in experienced readers with letters of a familiar alphabet, its origins have not been discussed or investigated. We propose that FT is the result of expertise with print, because print is regular in font. Here we test the conjecture that FT is the hallmark of a type of processing that distinguishes letter recognition from the recognition of most other shapes and objects. This series of studies is designed to test the idea that FT is a phenomenon common to expert readers in different reading systems (Roman alphabet and Chinese characters), but one that is not obtained in novice observers performing the same task with the same characters.

In experiments 1 and 2, we introduce a guided-visual-search task to measure the time an observer needs to recognize 100 letters. We reason that, because letter perception is very fast, the cost in time produced by disrupting font regularity should be more sizeable if measured over the recognition of many letters. The effect of manipulating font regularity may also be larger in arrays of many letters, if FT is a cumulative process that becomes more pronounced as the number of letters in the same font increases. To preview the results, we found evidence that experts with English and Chinese writing systems demonstrate FT with Roman and Chinese characters. In contrast, novices in some conditions make more errors when font is mixed, but do not demonstrate FT in the same general and robust manner in this paradigm. In experiment 3, we generalize these findings with a simpler task involving the report of briefly flashed 3-letter displays.

We also manipulate more specific aspects of fonts and find that experts show FT specifically for aspect-ratio manipulations but not other changes such as slant or fill.

We should note that performance in the tasks we devised to study letter perception may not be directly relevant to the mechanisms implicated in normal reading, as these tasks are quite artificial and designed to reduce the impact of the knowledge of orthographic and phonological regularities that develop with reading experience. Interestingly, there remains a controversy in the literature regarding whether word reading is essentially wholistic (Johnston and McClelland 1974; Reicher 1969) or limited at an early stage by letter recognition (Pelli et al 2003). Here, we remain agnostic in this debate, and simply note that there is recent evidence for both specialization at the level of single letters and letterstrings in different parts of the left fusiform gyrus (James et al 2005). Thus, learning about how reading expertise influences single letter perception may be relevant in understanding the processes taking place in some but not all of these cortical areas.

2 Experiment 1

In this experiment we measured the time required to scan through a 10×10 letter matrix while searching for specific targets in a guided-search paradigm, and we compared performance in this task in three conditions: baseline, where all letters in a matrix were in the same font; regular, where five different fonts were used but all letters on a given row were in the same font; and mixed, where five fonts were used in a random order throughout the matrix. Our prediction was that FT should be reflected by faster performance in baseline than regular matrices, which in turn should be scanned faster than mixed font matrices. In addition, the number of targets to be found in a matrix was varied, mainly to ensure that participants could not easily tell when they had found the last target. Although we predicted a sizeable effect of the number of targets, we also expected the effects of FT should be independent of target number. This is because changes in font were expected to impact the recognition of each individual letter (and not other processes like that of selecting a target or deciding to encode the next target).

2.1 *Participants*

Seventeen undergraduates from Vanderbilt University participated for monetary reward or course credit. All participants had normal or corrected-to-normal visual acuity.

2.2 *Apparatus and stimuli*

20 letters (‘a’ to ‘t’) in five fonts (Georgia, Croissant, Larabi, Trebuchet, and Angelus) were used. Each matrix was composed of 10 rows of 10 letters. The letters ‘a’ to ‘t’ were repeated 5 times each in a random order, except for the target letters (see task description below). Each letter, about 0.7 cm high and 0.4 cm wide, was separated from neighboring letters by 1.3 cm (see figure 1). An entire matrix thus subtended a visual angle of 14.0 deg $\times$ 15.8 deg in height and width, respectively, at an approximate viewing distance of 60 cm (although viewing position was not fixed).

There were three font-regularity conditions (baseline, regular, and mixed), each consisting of thirty matrices of 100 letters, with each matrix containing a unique letter sequence. In the baseline condition, each matrix had all the letters in the same font. Six different matrices in each of the five fonts were constructed. In the regular condition, the letters on each row were shown in the same font, with two noncontiguous rows for each of the five fonts. In the mixed condition, each matrix was created by taking one regular matrix and reorganizing the letters randomly. Thus, crucially, the regular and mixed conditions included identical letter stimuli and the matrices differed only in their ordering of letters. For each font-regularity condition, there were ten matrices each with 2, 3, or 4 switches in target letters (see task description below).

Figure 1. Examples of matrices used in experiments 1 and 2. Top: two Roman matrices used in experiment 1, from the regular (left) and the mixed font conditions (right). Bottom: two Chinese matrices used in experiment 2, from the regular (left) and the mixed font conditions (right).

Because of experimental errors, two of the matrices had the wrong number of targets, leaving only nine matrices in the regular-4-target and baseline-4-target conditions, and eleven matrices in the regular-2-target and regular-3-target conditions. Presentation order of the matrices was randomized. The experiment was conducted on iMac computers (15 inch CRT), each with a monitor set to an 800×600 pixel resolution. Presentation of stimuli was controlled by RSVP software (Williams and Tarr, no date).

2.3 *Procedure*

Subjects were instructed as follows:

> "The first letter of the matrix is your target. Scan the matrix from left to right, top to bottom, until you find your target. The letter that immediately follows it is your new target. Keep scanning the matrix until you find this letter, and the letter that immediately follows it is your new target. Continue this process until you get to the end of the matrix, and then press the space bar as fast as you can. After pressing the space bar you will have to type in the final answer. The last target you were looking for and did not find is the final answer. Make sure to perform the entire task as fast and accurately as you can."

The experimenter explained the task using an example of a baseline matrix printed on paper, and gave the participant a second baseline matrix on paper and asked him/her to perform the task and report the answer. The participant was allowed to ask any question about the task at this point, as well as after performing 5 practice trials on the computer. In the test trials, accuracy of the final answer was recorded on each trial, as well as the time from matrix onset until the participant pressed the space bar and the matrix disappeared.

2.4 *Results*

Trials with response times below 2000 ms or more than 3 standard deviations above the mean for all participants were removed from the analysis (less than 1.7% of the trials). Figure 2 shows the response times for correct responses and mean accuracy in the three conditions. Participants were slowest in the mixed font condition, followed by the regular and the baseline conditions, with no sign of an interaction with target number. They also made more errors in the regular and mixed conditions than in the baseline condition, especially in the 4-target condition.

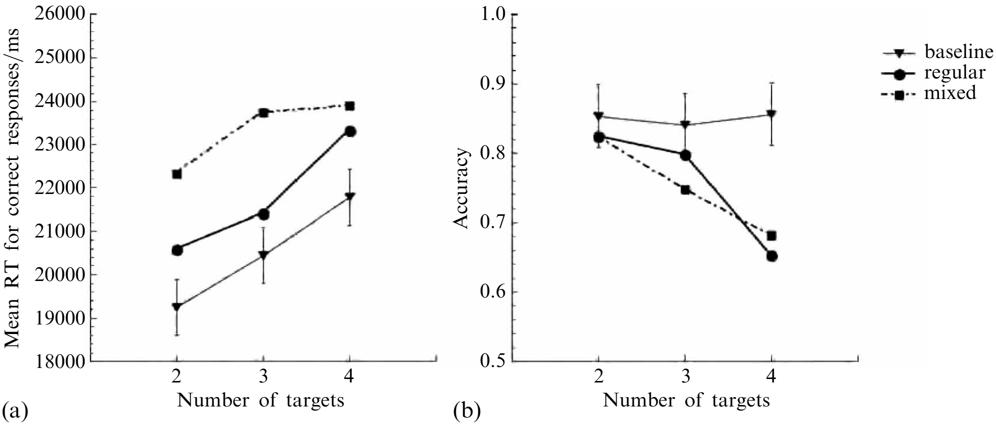


Figure 2. Accuracy for final target search and response times for correct trials in experiment 1 as a function of font condition (a) and target number (b). Error bars (only displayed in the baseline condition) show the pooled within-subjects SEM.

For each dependent variable, a 3×3 analysis of variance (ANOVA) was conducted with font regularity and number of targets as repeated factors. An ANOVA on mean response times for correct responses revealed a significant effect of number of targets ($F_{2,16} = 13.41$, $p < 0.0001$), with Scheffé tests ($p < 0.05$) indicating a significant difference between 3 and 4 targets, and a marginal difference ($p < 0.06$) between 2 and 3 targets. There was also a significant main effect of condition ($F_{2,16} = 32.50$, $p < 0.0001$), with all three conditions being significantly different from each other. The interaction between condition and number of targets was not significant ($F < 1$).

A similar ANOVA on accuracy showed a significant effect of number of targets ($F_{2,16} = 12.56$, $p < 0.0001$), with Scheffé tests revealing better performance in the conditions with 2 and 3 targets than in the condition with 4 targets. There was also a main effect of condition ($F_{2,16} = 7.25$, $p < 0.005$), with no difference between the mixed and regular conditions, but both being less accurate than the baseline condition. Finally, the interaction between target number and condition was significant ($F_{1,16} = 2.71$, $p < 0.05$). As can be appreciated from figure 2, the mixed and regular conditions both showed an effect of target number that was not obtained with the baseline condition, which was essentially at ceiling across all target number conditions.

2.5 *Discussion*

FT can be measured in a robust fashion with the matrix task. The same set of letters in the same fonts was used in the three conditions of this experiment. Nonetheless, participants were slower when several fonts were presented within the same matrix rather than only one, and slower still when font changed randomly within a matrix. One may suggest that the differences among the font-regularity conditions are related to the switch of target. For example, a reason for better performance in the baseline condition than the other two conditions might be that, in the baseline matrices, participants always found the matching letter in the same font that it had been presented as a target. However, this cannot account for the faster performance in the regular condition than in the mixed condition, because the mixed condition did not include more switches where the matching letter appeared in a different font from the target (mixed: 2.06 font switches on average; regular: 2.14 font switches). In addition, the condition effect occurred independently of target number, suggesting that the cost of font changes occurs on most letters of the matrix as they are rejected as non-targets, rather than in the recognition of task-relevant target letters.

These results extend the FT phenomenon with Roman letters to a new task. In experiment 2, we test whether FT would generalize to Chinese characters in Chinese readers.

3 Experiment 2

The goal of experiment 2 was to test whether FT can be obtained with characters of another writing system that differs in many important ways from the Roman alphabet. Chinese is written with characters called hànzi, each made out of 1 to 64 strokes and each associated with a syllable as well as meaning. The Chinese writing system is open-ended, although knowledge of about 4500 characters is sufficient to read Modern Standard Chinese. Each character takes about the same amount of space and those making multisyllable words are not physically grouped together like English words. Despite these differences from the Roman alphabet, we hypothesized that expertise with Chinese should lead to FT, because font is used in Chinese much in the same manner as it is in the Roman alphabet.

3.1 *Participants*

Seven undergraduate students from the Chinese University of Hong Kong participated for monetary reward or course credit. Chinese was the first language and English was the second language for all participants, and they had at least 15 years of experience reading each of them. All participants had normal or corrected-to-normal vision.

3.2 *Apparatus and stimuli*

8 Chinese characters were used, chosen to be simple (3 strokes) to moderately complex (9 strokes), in five different fonts. Only 8 characters were used to keep response making relatively simple, as the keys on the standard keyboard had to be labeled arbitrarily with new characters. The 8 keys from S to L on the keyboard were labeled with the 8 Chinese characters, in a different font from those used in the matrices. In a matrix, each letter, 0.8 cm high and 0.8 cm wide, was separated from neighboring letters by 0.4 cm. An entire matrix subtended 12.2 deg $\times$ 12.2 deg of visual angle with an approximate viewing distance of 60 cm.

Sixty matrices were constructed, thirty for the regular and thirty for the mixed condition, each with a unique letter sequence. Within each font regularity condition, there were ten matrices each with 3, 4, or 5 targets. Because of experimental error, one matrix had the wrong number of targets, leaving only nine matrices in the mixed-4-target condition and eleven matrices in the mixed-5-target condition. There was no baseline condition in this experiment because the crucial comparison is that between the mixed and regular matrices. Presentation order of the matrices was randomized. The experiment was conducted on iMac computers (15 inch CRT) each with a monitor set to an 800×600 pixel resolution. Presentation of stimuli was controlled by RSVP software (Williams and Tarr, no date). The rest of the procedure was the same as in experiment 1.

3.3 *Results*

Trials with response times below 2000 ms or more than 3 standard deviations above the mean for all participants were removed from the analysis (3.7% of the total trials). Figure 3 shows the response times for correct responses and mean accuracy in the two conditions. The response times show the same pattern as in experiment 1, with both an effect of target number and an effect of font regularity, with no apparent interaction between the two.

For each dependent variable, a 2×3 ANOVA with font regularity and number of targets as repeated factors was conducted. For mean response times in correct trials, the effect of target number was significant ($F_{2,12} = 15.32$, $p < 0.001$), with Scheffé tests ($p < 0.05$) showing a significant difference between 4 and 5 targets. The effect of condition was also significant ($F_{1,6} = 6.23$, $p < 0.05$), and there was no interaction between the two factors ($F < 1$). For accuracy, none of the effects was significant, with all Fs < 1 except for the effect of target number ($F_{2,12} = 2.89$, $p = 0.09$).

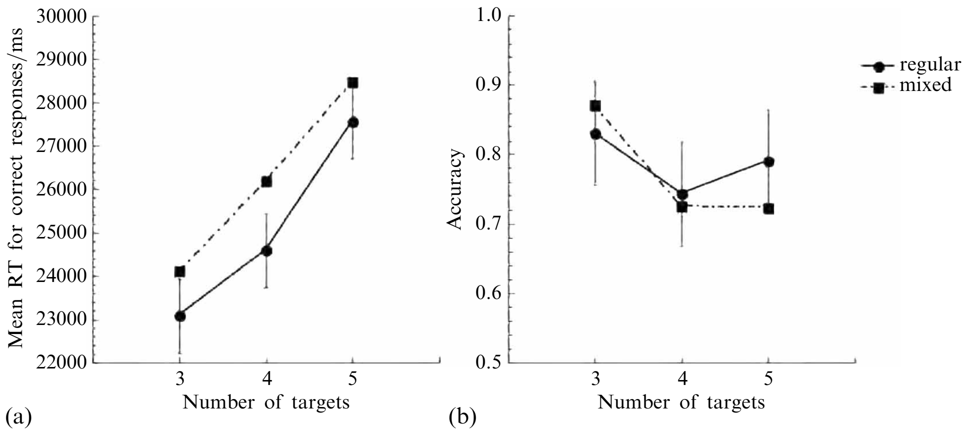


Figure 3. Accuracy of final target search and response times for correct trials in experiment 2, plotted as a function of font condition (a) and target number (b). Error bars (only displayed in the regular condition) show the pooled within-subjects SEM.

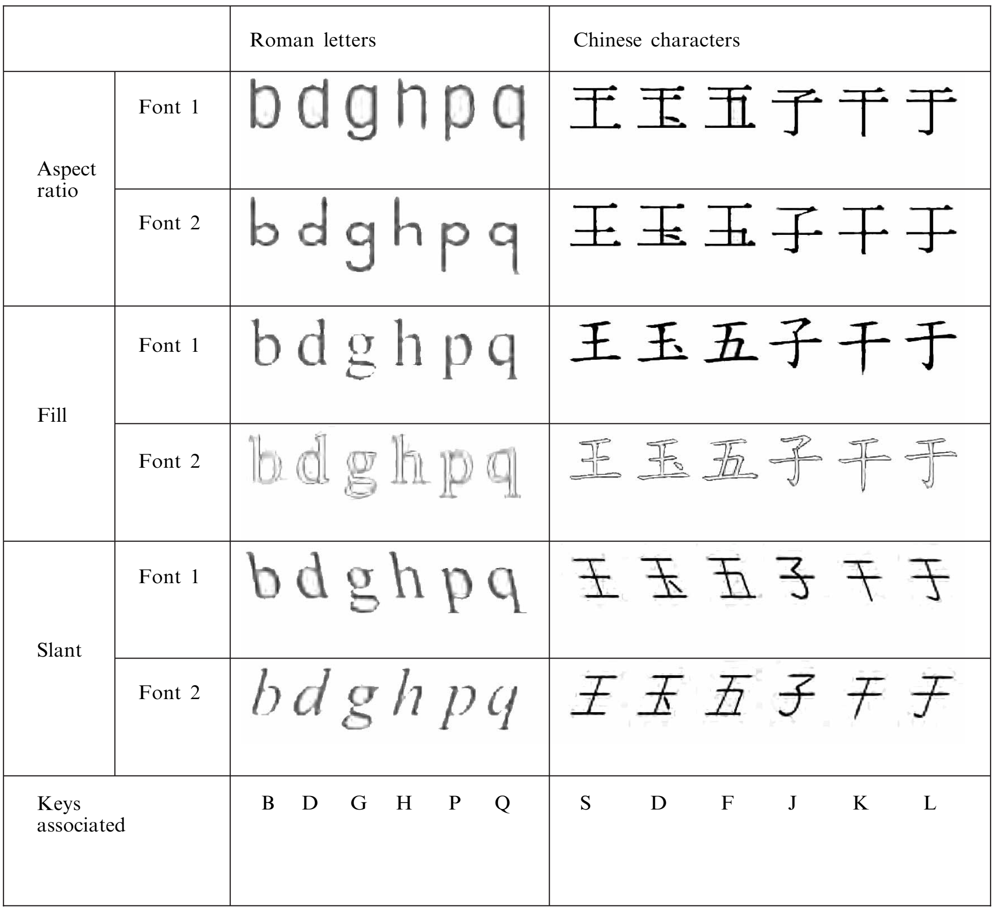
3.4 *Discussion*

The results of experiment 2 revealed that FT is not specific to the Roman alphabet and can be obtained with Chinese characters in experts (Chinese readers). One question however that is not answered by our results nor by prior work by Sanocki (1987a, 1987b, 1988) is whether tuning to shape regularities that are task-irrelevant is a general phenomenon that should be expected of any observer, or whether it depends to some extent on our experience with letters. In a pilot study, we replicated experiment 2 in a group of non-Chinese readers. The results were difficult to interpret, for a number of reasons. While these novice observers showed little evidence of FT, they also were much less accurate and took about twice as long to scan each matrix as experts in experiment 2. In addition, the novices reported often having to start again from the first letter of the matrix in some cases because they got confused as to what the current target was, which would contribute to much slower—and variable—responses. The matrix task appears too complex to afford a meaningful comparison of perceptual processing in novices and experts. The next experiment describes a simpler task that aims at isolating the perceptual aspects of FT, to test the hypothesis that this phenomenon occurs only in expert readers.

4 Experiment 3

Experiment 3 had two goals. First, a simpler task was used to better tap into the perceptual aspects involved in FT by reducing the effect of a number of extraneous factors [this task was used by Jolicoeur (1990) to study orientation priming]. On each trial, we asked participants to identify 3 letters/characters presented briefly on the screen. Novice participants were trained with the verbal labels for the Chinese characters prior to the experiment, and both experts and novices were trained to associate the characters with the response keys. The purpose of this procedure was to minimize differences between novices and experts to the task, apart from their relative expertise with the characters.

Second, whereas in experiments 1 and 2 we used real fonts that differed in uncontrolled ways, in experiment 3 we explored whether FT depends on specific types of font change. In a pilot study aimed at exploring which font change would cause the biggest magnitude of FT, we tested English readers viewing Roman letters. Significant FT was found with aspect-ratio changes but not changes in line thickness, fill, slant, size, or presence of serifs. In this experiment we therefore used aspect-ratio changes, which we predicted would lead to robust FT, and we also used fill and slant manipulations, that were expected to have, at most, subtle effects (see figure 4a).



(a)

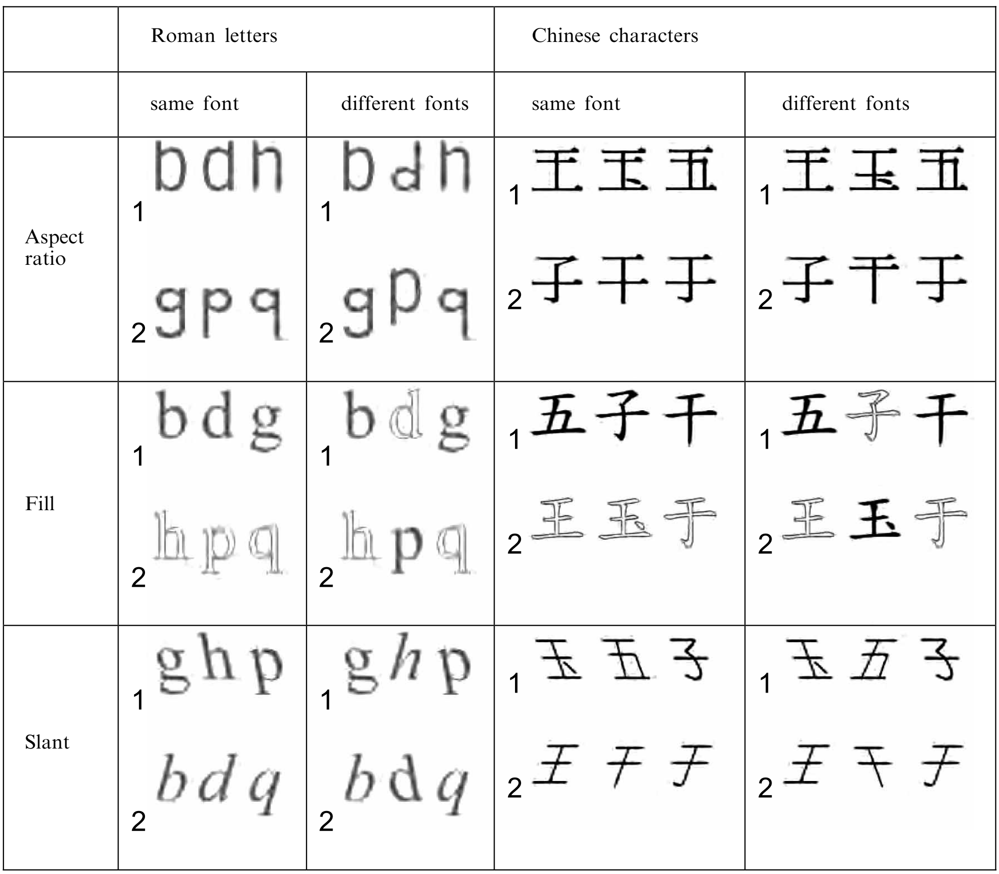
Figure 4. (a) Roman letters and Chinese characters and the response keys associated in experiment 3. (b) Examples of combinations in same-font and different-font conditions (two examples per cell).

4.1 *Participants*

Sixteen Chinese readers from the Chinese University of Hong Kong and eleven non-Chinese readers from Vanderbilt University participated for monetary reward. The Chinese readers had Chinese as their first language and English as their second language, and had at least 15 years of experience reading both Chinese and English. All non-Chinese readers either had English as their first language (or another language that uses Roman characters) or had at least 15 years of experience reading English texts. They reported no experience with Chinese characters. All participants had normal or corrected-to-normal visual acuity.

4.2 *Apparatus and stimuli*

6 lower-case Roman letters and 6 Chinese characters in various font types were used (figure 4a). For the aspect-ratio manipulation, the two Roman fonts differed in the vertical extent of the middle component (eg the loop in the letter ‘b’ in font 1 extends a vertical length 50% more than that in font 2). The size of the ascendants and descendents remained constant across fonts. The two Chinese fonts differed in the position of the central stroke, which is about 15 pixels higher in font 1 than in font 2. For the fill manipulation, the fonts were created by using 2-pixel wide outlines for the strokes for one font, and filling up all the space within the outlines for the other. For the slant manipulation, fonts were created by tilting the main vertical stroke by 72°



(b)

Figure 4 (continued)

counterclockwise (font 1) or clockwise (font 2). The contrast of the Roman letters was reduced by half for all three conditions because pilot studies indicated that some participants still identified the letters perfectly at the lowest presentation time.

A mask 1.5 cm high and 4 cm wide was created from fragments of the letters and characters. Each Roman letter was about 1 cm high and 0.8 cm wide and each Chinese character was about 1.2 cm high and wide. There were three font manipulations: aspect ratio, fill, and slant. For each font manipulation there were two sets of 6 letters in different fonts (font 1 and font 2 sets). On each trial, 3 letters were chosen from one or both font sets and presented as a trigram. There was no repetition of letters within each trigram. In the same-font condition, the 3 letters were in either a 'font 1 – font 1 – font 1' or 'font 2 – font 2 – font 2' combination (figure 4b). In the different-font condition, they were shown in a 'font 1 – font 2 – font 1' or 'font 2 – font 1 – font 2' order. For each font manipulation 48 same-font and 48 different-font trigrams were used (figure 4b). The different-font trigrams were generated by taking the 48 same-font trigrams and changing the font of the central letter (from font 1 to font 2, or vice versa).

Four iMac computers (15 inch CRT), each with a monitor set to a 1024×768 pixel resolution, were used. Presentation of stimuli was controlled by RSVP software (Williams and Tarr, no date).

4.3 *Procedure*

The experiment was divided into Roman letter and Chinese character sessions. We first describe the procedure used for Chinese – English bilinguals. All participants finished the Roman letter session first, followed by the Chinese character session. In each session there were three phases: key learning, calibration, and identification phases. The first two phases prepared the participant for the identification phase, where FT was measured.

4.3.1 *Key learning phase.* Participants saw single letters on the screen and learnt the mapping between the letters and the keys (figure 4a). On each trial a letter was presented and the participant had to press its corresponding key as accurately and as fast as possible. A sheet with the characters and their names was provided during the first two blocks of training (48 trials per block). From the third block onwards a participant who reached accuracy of over 95% and an average reaction time of below 1000 ms in any block could proceed to the next phase.

4.3.2 *Calibration phase.* A presentation time of the letter trigrams was selected for each participant to prevent a ceiling effect in the identification phase. On each trial, participants saw a fixation cross for 500 ms, followed by the presentation of a trigram for a variable duration (see below) and then a mask until the first key response. The participant had to type the keys learnt in the previous phase for all 3 letters in an order from left to right. When each of 3 keys was typed, an underscore sign appeared under the position of its corresponding letter. The participant was given as much time as needed to respond as accurately as possible. A medium-pitch tone was given as feedback at the end of a trial when one or more of the 3 letters was not recognized. A staircase method (with steps of 50 ms; except when the presentation time went below 100 ms, when the steps were of 13 ms) was used to find the presentation time for which each participant performed correctly in 9 to 11 out of the 12 trials (75% to 91.67%) in two consecutive blocks. This presentation time was then used throughout the identification phase. The letters used here were the same ones used in the identification phase, albeit in a different font.

4.3.3 *Identification phase.* Participants went through the same task as the calibration phase. The only differences were that the presentation time was fixed and that each letter could appear in six different fonts (figure 4). There were 48 same-font and 48 different-font trigrams in each of the three font manipulations, forming a total of 288 trials in each of the Roman letter and Chinese character sessions. They were presented in random order in blocks each with 24 trials.

The non-Chinese readers went through the same experimental conditions, except that during the Chinese-character session they first learned the 'names' arbitrarily assigned to the Chinese characters in a 'naming' session. The names were 'som', 'dah', 'fep', 'jut', 'kax', and 'lic', with initials matching their corresponding keys ('S', 'D', 'F', 'J', 'K', and 'L'). In the naming session a character was presented together with four names under it. The participant had to press one of the keys 1 to 4 to determine which was the name of the character. When the participant reached accuracy of over 95% and an average reaction time for correct trials of below 1000 ms in any block of 48 trials, he/she could proceed to the calibration session.

4.4 *Results*

Data from three Chinese – English bilinguals and one non-Chinese reader were discarded, because their overall accuracy in the identification phase was over 2 standard deviations below the average of all participants.

For Chinese – English bilinguals, twelve observers were given a presentation time of 150 ms and one of 100 ms when viewing Chinese characters in the identification phase; eleven had a presentation time of 100 ms and two of 53 ms when viewing Roman letters. For non-Chinese readers, all had a presentation time of 250 ms when viewing Chinese characters; nine had a presentation time of 100 ms and one of 13 ms when viewing Roman letters.

Figure 5 shows the recognition accuracy in different conditions. Separate $2 \times 2 \times 2$ (group $\times$ letter type $\times$ regularity) analyses of variances (ANOVA) were conducted for each font manipulation. Within each font manipulation, group (Chinese – English bilinguals/English readers) was a between-subjects factor. Letter type (Roman letters/Chinese characters) and regularity (same font/different font) were within-subjects factors. The number of letters correct (maximum = 3) was the dependent variable.

4.4.1 *Aspect ratio (figure 5a).* Tuning to aspect ratio occurred only in expert situations. A significant three-way interaction between group, letter type, and regularity was found ($F_{1,21} = 8.86$, $p < 0.01$). Thus, separate analyses (letter type $\times$ regularity) were performed on the two groups of participants.(2)

Chinese – English bilinguals showed FT both with Roman letters and with Chinese characters. They were more accurate when the letters appeared in the same font than in different fonts. This was confirmed with a significant main effect of regularity ($F_{1,12} = 23.64$, $p < 0.001$). The bilinguals also performed better with Roman letters than Chinese characters, perhaps because the Roman letters are not as complex ($F_{1,12} = 19.84$, $p < 0.001$). The absence of regularity $\times$ letter type interaction ($F < 1$) showed that the magnitude of FT was similar for both types of characters.

Non-Chinese readers showed FT only with Roman letters but not with Chinese characters. There was a significant regularity $\times$ letter type interaction ($F_{1,9} = 11.9$, $p < 0.01$). Simple effect analyses showed that for Roman letters, accuracy was higher with same fonts than with different fonts ($F_{1,9} = 19.60$, $p < 0.002$), whereas for Chinese characters accuracy was not different between same and different fonts ($F_{1,9} = 1.48$, $p > 0.25$).

As shown in figure 5a, non-Chinese readers obviously identified Roman letters better than Chinese characters. On average they correctly named 2.66 Roman letters in each trial, but only 1.88 Chinese characters. It is possible that the absence of FT for Chinese characters in the non-Chinese readers is because on average they did not go

(2) All 2×2 ANOVAs were conducted separately for each group at a Bonferroni-corrected alpha level of 0.025. Subsequent simple effect analyses were performed with an alpha level of 0.013.

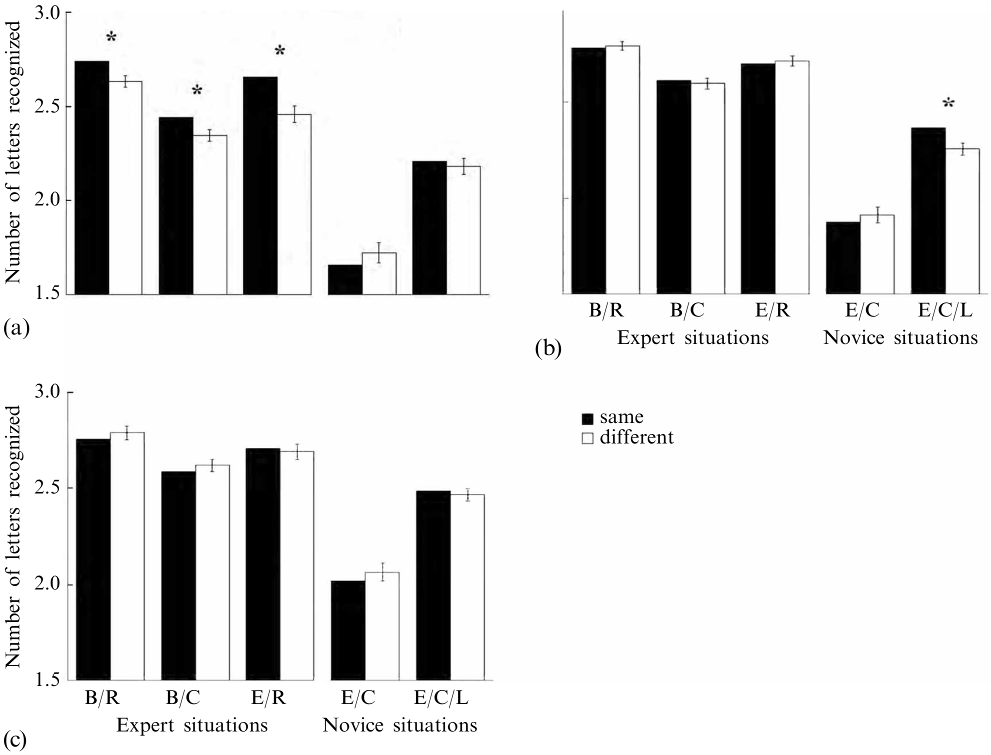


Figure 5. Number of letters recognized in experiment 3, in the same-font and different-font conditions for (a) the aspect-ratio, (b) fill, and (c) slant trials. B/R—bilinguals seeing Roman letters; B/C—bilinguals seeing Chinese characters; E/R—English readers seeing Roman letters; E/C—English readers seeing Chinese characters; E/C/L—English readers seeing Chinese characters with longer presentation times. Error bars show the standard errors of the differences between same-font and different-font conditions.

through more than 2 letters on each trial. To address this issue, we tested another group of twelve undergraduate English/non-Chinese readers with the same task with all training except that (i) they only went through the Chinese session; and (ii) the presentation time for the 3 characters was set at either 600 ms (seven participants) or 800 ms (remaining five). With longer presentation times, the performance level was more comparable for Chinese characters with that for the Roman letters (see the last pair of bars in figure 5a). Still, there was no difference in accuracy between the same-font and different-font conditions ($F < 1$).

4.4.2 *Fill (figure 5b).* There was no main effect or interaction involving regularity (all Fs < 1) for both Chinese and non-Chinese readers. A significant interaction between group and letter type was found ($F_{1,21} = 54.04$, $p < 0.001$), caused by the non-Chinese readers performing particularly poorly with Chinese characters.

As mentioned above, we tested an additional group of twelve non-Chinese readers on the Chinese characters with longer presentation times. With a longer presentation time, FT to fill was shown as the accuracy was higher when the characters were of the same font than of different fonts ($t_{11} = 3.558$, $p < 0.005$). Finding any kind of FT in novices was unexpected, but interestingly this result still reflects a difference between novice and expert letter perception. Experts do not show any FT for fill, and FT for aspect ratio was obtained for experts with much faster presentation times.

4.4.3 *Slant (figure 5c).* Neither group showed any tuning to slant. There was no main effect or interaction involving regularity (all Fs < 1). The interaction between group and letter type was significant ($F_{1,21} = 30.16$, $p < 0.0001$), a result of the worse performance of the non-Chinese readers viewing Chinese characters.

No evidence of FT for slant was shown for the additional group of non-Chinese readers tested with a longer presentation time. There was no difference between the same-font and different-font conditions ($F < 1$).

4.5 *Discussion*

The most interesting result in experiment 3 occurred with the aspect-ratio manipulation. FT to aspect ratio occurred only in experts, which included cases when English readers and Chinese – English bilinguals were seeing Roman letters, and when the bilinguals were seeing Chinese characters. In novice observers (ie non-Chinese readers seeing Chinese characters), no FT to aspect ratio was found, even when the overall performance for Chinese characters and Roman letters was matched. We found indications that novices, only for long presentation times, were sensitive to a 'fill' manipulation. The 'fill' manipulation is the most salient and is unlikely to go unnoticed even in a novice observer. One possibility is that the effect in novices reflects a disruption for different fill trials rather than a facilitation in the same fill trials, but without a neutral baseline this cannot be evaluated. It is unclear whether with a longer exposure duration this would also occur in expert situations, but would be difficult to test with the present task because experts would perform at ceiling. Therefore, this fill effect may either be specific to novices when accuracy is high enough for our measure to be sensitive, or it could be generally associated with long presentation times.

Our results suggest that expert and novice readers process font information in different ways: even when perceiving characters presented very briefly, experts are sensitive to subtle aspect-ratio manipulations that have no effect on novice performance under both long and short exposure durations. In contrast, when given long exposure durations, novices seem to be influenced by grosser aspects of font changes that do not appear to impact expert perception. Further research could focus on other types of font manipulations (eg serifs, size, curvature ...) to which experts may tune to, but the purpose of this study was to demonstrate that FT for some shape properties is associated with letter expertise.

A potential explanation for the difference between aspect-ratio and other font changes is that aspect-ratio changes, at least for Roman letters, can potentially result in changes in letter identity. For example, the letter 'd' could become an 'a' when the 'pole' becomes shorter and shorter with respect to the 'loop'. However, we believe it is unlikely to account for our findings of FT in this experiment, because, for the Chinese characters used, aspect-ratio changes could not lead to a change in character identity. Yet, we found evidence of FT in experts, for aspect-ratio manipulations performed on both types of characters.

5 Experiment 4

In experiment 3, aspect ratio was the only font manipulation that reliably produced FT in both our experts groups. Unfortunately, we did not plan a direct comparison between aspect ratio, fill, and slant in that experiment, and it is therefore difficult to know whether the pattern of results was caused by the relative strength of each of these manipulations. In experiment 4 we used the same stimuli as in experiment 3, but subjects were required to make matching judgments for the font manipulations rather than to ignore them. If aspect ratio leads to larger FT because the font manipulation is more salient than the other two manipulations, we reason that it should also be easier to judge that the letter triplets are presented in the same versus different fonts in this condition.

5.1 *Participants*
Fifteen Chinese–English bilinguals from the Chinese University of Hong Kong participated for course credit and fourteen non-Chinese readers from Vanderbilt University participated for monetary reward. The Chinese readers had Chinese as their first language and English as their second language, and had at least 15 years of experience reading both Chinese and English. All non-Chinese readers either had English as their first language (or another language that uses Roman characters) or had at least 15 years of experience reading English texts. They reported no experience with Chinese characters. All participants had normal or corrected-to-normal visual acuity.

5.2 *Apparatus and stimuli*
All materials were identical to those used in experiment 3.

5.3 *Procedure*
On each trial, participants first saw a fixation cross for 500 ms, followed by presentation of 3 characters. They had to decide whether the characters were presented in the same or in different fonts, and respond with the corresponding key responses ('1' for same and '2' for different). The characters stayed on the screen until a response was made, and a 200 ms blank was shown after response and before the next trial began. Participants were asked to perform the task as accurately as possible while being reasonably fast. After each incorrect trial, negative feedback was given in the form of a tone followed by an error message ("Sorry! Answer incorrect!") shown on the screen for 4 s.

As in experiment 3 there were 48 same-font and 48 different-font trigrams in each of the three font manipulations, forming a total of 288 trials in each of the Roman and Chinese character sessions. They were presented in random order in six blocks each with 48 trials. The experiment was divided into Roman letter and Chinese character blocks and all participants completed the Roman letter blocks first, as in experiment 3. Before each of the Roman and Chinese blocks, participants were shown two examples of each condition (eg same aspect ratio, different fill, etc) to study, and were given 24 practice trials.

5.4 *Results*
Data from one English reader were discarded, because accuracy was more than 2 standard deviations below the average of all participants. The resulting average accuracy was 96.4% (SD = 2.6%). Only correct trials with response time between 350 and 2650 ms were included in the analyses, resulting in only 1.05% of the trials being discarded.

Figure 6 shows the sensitivity and correct response times for each condition. For the two dependent variables, separate $2 \times 2 \times 3$ (group $\times$ character type $\times$ font manipulation) ANOVAs were conducted. In general, apart from the fact that bilingual subjects were faster, the pattern of results was similar for both groups. Performance (in terms of speed, which should be more sensitive on such an easy task) was the best for the fill manipulation, followed by slant and then aspect ratio. These differences were larger for Chinese characters than for Roman letters.

These observations are supported by the ANOVA on sensitivity, which revealed main effects of character type ($F_{1,27} = 10.38$, $p < 0.005$) and of font manipulation ($F_{2,54} = 5.05$, $p < 0.01$) as well as an interaction between these two factors ($F_{2,54} = 8.23$, $p < 0.001$). A posteriori tests (Scheffé) revealed that the fill and slant judgments did not differ between the two types of characters, whereas subjects (regardless of expertise) were more accurate on the Roman than Chinese characters for aspect-ratio judgments.

The ANOVA on response times revealed a significant main effect of group ($F_{1,27} = 9.76$, $p < 0.005$), with bilingual subjects responding faster than the non-Chinese readers. There was also a main effect of font manipulation ($F_{2,54} = 35.07$, $p < 0.0001$), but this factor also interacted with the type of character ($F_{2,54} = 17.09$, $p < 0.0001$). Again, a posteriori

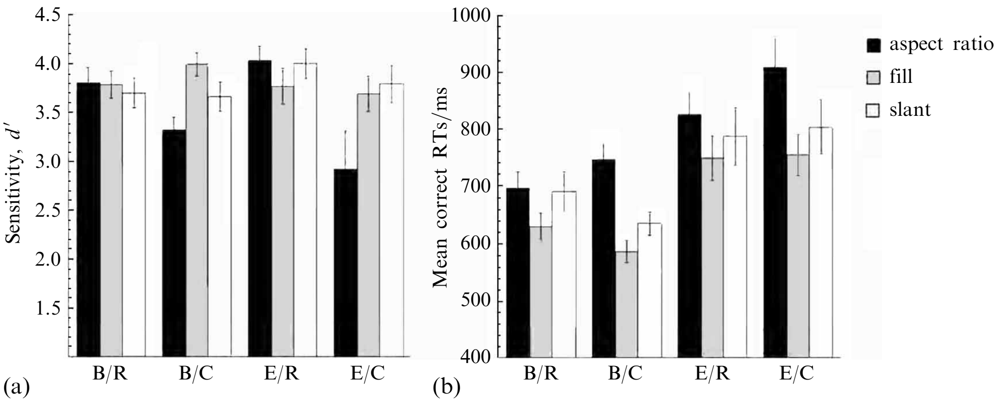


Figure 6. (a) Sensitivity and (b) mean response times (RTs) for correct responses in experiment 4. For notation see figure 5. Error bars show the SEM.

tests (Scheffé) revealed that, as for sensitivity, performance was slower for Chinese than for Roman characters only for aspect-ratio judgments.

5.5 *Discussion*

The results of experiment 4 clearly suggest that the aspect-ratio manipulation was not the most salient of all three font manipulations, for either novice or expert observers. Likewise, the fill manipulation showed no indication of being relatively more salient for novices than for experts. Although we found that our aspect-ratio manipulation was stronger for Roman than for Chinese characters, this has no bearing on the expertise effects obtained in experiment 3.

These results suggest that the salience of the font manipulations, at least when this dimension is task-relevant, cannot explain the results obtained in experiment 3. This is consistent with the possibility that the FT results in experiment 3, especially for experts in the aspect-ratio condition, arise from expert letter-recognition mechanisms that may apply normalization based on font information but that are different from those used when attending to font information (which is likely not a task with which expert readers have much experience).

6 General discussion

The most interesting implication of our findings is that letters are not perceived in the same manner as shapes(3) by expert readers: expert letter-recognition takes advantage of font regularity and this may be part of what supports the extremely fast performance of experts compared to novices.

The use of regularity of style by expert readers may be an important factor that distinguishes visual letter perception from the recognition of other shapes and objects. Regularity of font or style is an important property of print that is generally not present in other cases of object recognition. The analogy for the 'font' of an object would be its subordinate level (or simply its style in the case of man-made objects). For instance the 'font' of a chair may be 'mission'. Even in cases like furniture, where different basic-level objects (chair, table, dresser) can be seen together in the same style, we rarely recognize many of these objects in rapid succession, without other intervening objects (such as a plant, a person, a dog). The high occurrence of such style regularity for letters seen in rapid succession could lead readers to expect and use this information when recognizing letters.

(3) Here we assume that Chinese characters are processed like any other type of simple 2-D shape by non-Chinese readers.

It is also interesting to compare letter expertise to other cases of expertise. One of the most studied is face expertise (Diamond and Carey 1986; Gauthier et al 2000; Tanaka 2001). Faces typically need to be recognized at the subordinate level, whereas letters do not (Wong and Gauthier, submitted). Therefore, in one sense letter recognition is more similar to object recognition than to face recognition. With letters, as with objects, we generally need to distinguish between shapes that have different structural descriptions (Biederman 1987). In contrast, faces are often recognized at the individual level, and since they all share the same structural description (two eyes side-by-side above a nose, etc), more metric variations become crucial. In addition, sequences of different faces in the same 'font' are relatively rare (eg perhaps different individuals in the same family?).[(4)] In contrast to such differences with face recognition, letter recognition is in other ways more similar to face recognition than to object recognition. That is, for letters the subordinate-level information (eg appearance of a letter in a certain font type) appears to be automatically encoded, whereas it is not typically the case with common objects (Jolicoeur et al 1984; Tanaka and Taylor 1991). In the case of faces, subordinate-level information is generally thought to be encoded automatically because individuation is such an important (and frequent) aspect of processing faces. In the case of letters, we propose that it may be encoded because of the ubiquitous regularities of style present in print.

Thus, we suggest that it is the omnipresence of font regularity in printed text, combined with the benefits that using such regularity may confer, which leads to FT in expert readers. Accordingly, we propose that other important aspects of reading expertise (eg mapping orthography onto phonology, semantics) are not relevant in producing FT. A corollary prediction is that experience in the rapid recognition of any shapes that are consistently presented in style-regular contexts should lead observers to develop FT.

It is interesting to speculate whether or not FT may be responsible for the cost involved in reading mixed-case stimuli, which is larger for non-words than for words (Besner and Johnston 1989). Mayall et al (1997) investigated several possible causes for the effect of case mixing, and found evidence for the effect of at least two factors in this phenomenon. At least some of the disruption in mixed-case stimuli appears to be due to inappropriate groupings of letters of the same case, and this effect is reduced when the spacing between letters is increased. However, another factor seems to have an influence, as suggested by the fact that strings with letters of different case but equated in size, or with letters of the same case but of different size, both produced a cost in reading time but were not impacted by increasing letter spacing. Mayall et al (1997) proposed that the loss of transletter features, which are larger than single letters but smaller than whole words, may cause these disruptions. However, this rests on the untested assumption that such transletter features are not influenced by increasing space between each letter. One possibility based on our results is that some of the slowing down caused by mixing case is due to disruptions in the use of regularities in font (with the additional factor that different cases of the same letter can be composed of different parts). Indeed, we found evidence for FT under both small (experiment 3) and large letter spacing (experiment 1). In the future the effects of spacing letters in combination with font manipulations, as well as the letter-presentation asynchrony, could be investigated to study the spatial and temporal characteristics of FT.

Our results suggest that rapid FT is a hallmark of our visual expertise with letters, and as such it is a candidate as a mediator for specialization for letters in the visual cortex.

[(4)] An interesting possibility is that there may be contextual effects in face recognition that may be analogous to those described here with letters, for instance for a face seen in the context of other faces of the same race, or of different races. However, this situation is likely to be different from letter recognition, as faces rarely require rapid successive recognition.

In other words, letters may come to recruit parts of the visual system that are the best suited for the computations supporting fast FT. There is evidence for specialization in the left extrastriate cortex for letters (Gauthier et al 2000; Hasson et al 2002; Polk et al 2002; Puce et al 1996). For example, an area in the occipito-temporal region responds more to Roman letters than to digits or Chinese characters in participants who do not read Chinese (James et al 2005), while the same area responds to Roman letters and Chinese characters in individuals who have experience with both (Wong et al 2005). Just as it has been suggested that faces engage a specific part of the visual system because they are processed automatically by experts at the subordinate level using a configural strategy (Gauthier and Tarr 2002), some of the visual areas specialized for letter perception may be related to the specific computations supporting fast FT. This hypothesis receives some support from an fMRI study in which a letter-selective area in occipito-temporal cortex showed habituation to the same letter in the same but not different fonts (Gauthier et al 2000). Future work will need to investigate whether this habituation occurs across different letters of the same font and whether it is associated with expertise.

In sum, our results reveal that the phenomenon of FT described by Sanocki (1987a, 1987b, 1988) generalizes across writing systems and that expert readers use task-irrelevant variations in shape differently from novices. This finding reveals how expert letter processing differs from both the recognition of common objects and from expertise with faces. Future studies could investigate the mechanisms that mediate FT, comparing normalization (eg Martin et al 1989) to exemplar models (Goldinger 1998; Sanocki 1992), as well as explore the neural bases of this tuning process (eg Wong et al 2004).

Acknowledgments. Research was supported by NEI Grant R01-EY13441, a grant from the James S McDonnell Foundation to IG, and a grant from the Research Grants Council of the Hong Kong SAR (Project No. HKU 4260/03H).

References

Besner D, Johnston J C, 1989 "Reading and the mental lexicon: On the interaction of visual, orthographic, phonological and lexical information", in *Lexical Processes and Representation* Ed. W Marslen-Wilson (Cambridge, MA: MIT Press) pp 291 – 316

Biederman I, 1987 "Recognition-by-components: A theory of human image understanding" *Psychological Review* **94** 115 – 147

Creelman C D, 1957 "Case of the unknown talker" *Journal of the Acoustical Society of America* **29** 655

Diamond R, Carey S, 1986 "Why faces are and are not special: An effect of expertise" *Journal of Experimental Psychology: General* **115** 107 – 117

Estes W K, 1978 "Perceptual processing in letter recognition and reading", in *Perceptual Processing* Eds E C Carterette, M P Friedman (New York: Academic Press) pp 163 – 220

Gauthier I, Skudlarski P, Gore J C, Anderson A W, 2000 "Expertise for cars and birds recruits brain areas involved in face recognition" *Nature Neuroscience* **3** 191 – 197

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

Gauthier I, Tarr M J, Moylan I, Anderson A W, Skudlarski P, Gore J C, 2000 "The fusiform 'face area' is part of a network that processes faces at the individual level" *Journal of Cognitive Neuroscience* **12** 495 – 504

Gibson E J, 1969 *Principles of Perceptual Learning and Development* (New York: Appleton-Century-Croft)

Gibson E J, Levin H, 1973 *The Psychology of Reading* (Cambridge, MA: MIT Press)

Goldinger S D, 1998 "Echoes of echoes? An episodic theory of lexical access" *Psychological Review* **105** 251 – 279

Hasson U, Levy I, Behrmann M, Hendler T, Malach R, 2002 "Eccentricity bias as an organizing principle for human high-order object areas" *Neuron* **34** 479 – 490

Hofstadter D R, McGraw G E Jr, 1995 "Letter spirit: Esthetic perception and creative play in the rich microcosm of the roman alphabet", in *Fluid Concepts and Creative Analogies* Ed. D R Hofstadter (New York; Basic Books) pp 407 – 466

James K H, James T W, Jobard G, Wong A C-N, Gauthier I, 2005 "Letter processing in the visual system: different activation patterns for single letters and strings" *Cognitive and Affective Behavioral Neuroscience* **5** 452 – 466

Johnston J C, McClelland J L, 1974 "Perception of letters in words: seek not and ye shall find" *Science* **184** 1192 – 1194

Jolicoeur P, 1990 "Orientation congruency effects on the identification of disoriented shapes" *Journal of Experimental Psychology: Human Perception and Performance* **16** 351 – 364

Jolicoeur P, Gluck M, Kosslyn S M, 1984 "Pictures and names: Making the connection" *Cognitive Psychology* **16** 243 – 275

Keren G, Baggen S, 1981 "Recognition models of alphanumeric characters" *Perception & Psychophysics* **29** 234 – 246

McClelland J, Elman J, 1986 "The TRACE model of speech perception" *Cognitive Psychology* **18** 1 – 86

McClelland J L, Rumelhart D L, 1981 "An interactive model of context effects in letter perception. Part 1. An account of basic findings" *Psychological Review* **88** 375 – 407

Martin C S, Mullenix J W, Pisoni D B, Summers W V, 1989 "Effects of talker variability on recall of spoken word lists" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15** 676 – 684

Mayall K, Humphreys G W, Olson A, 1997 "Disruption to word or letter processing? The origins of case-mixing effects" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **23** 1275 – 1286

Morton J, 1969 "Interaction of information in word recognition" *Psychological Review* **76** 165 – 178

Mullennix J W, Pisoni D B, 1990 "Stimulus variability and processing dependencies in speech perception" *Perception & Psychophysics* **47** 379 – 390

Mullennix J W, Pisoni D B, Martin C S, 1989 "Some effects of talker variability on spoken word recognition" *Journal of the Acoustical Society of America* **85** 365 – 378

Nygaard L C, Sommers M S, Pisoni D B, 1995 "Effects of stimulus variability on perception and representation of spoken words in memory" *Psychological Science* **57** 989 – 1001

Oden G C, 1979 "A fuzzy logical model of letter identification" *Journal of Experimental Psychology: Human Perception and Performance* **5** 336 – 352

Oden G, Massaro D, 1978 "Integration of featural information in speech perception" *Psychological Review* **85** 172 – 191

Palmeri T J, Goldinger S D, Pisoni D B, 1993 "Episodic encoding of voice attributes and recognition memory for spoken words" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **19** 309 – 328

Pelli D G, Farell B, Moore D C, 2003 "The remarkable inefficiency of word recognition" *Nature* **423** 752 – 756

Polk T A, Stallcup M, Aguirre G K, Alsop D C, D'Esposito M, Detre J A, Farah M J, 2002 "Neural specialization for letter recognition" *Journal of Cognitive Neuroscience* **14** 145 – 159

Puce A, Allison T, Asgari M, Gore J C, McCarthy G, 1996 "Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study" *Journal of Neuroscience* **16** 5205 – 5215

Reicher G M, 1969 "Perceptual recognition as a function of the meaningfulness of stimulus material" *Journal of Experimental Psychology* **81** 275 – 280

Sanocki T, 1987a "Visual knowledge underlying letter perception: A font-specific approach" *Dissertation Abstracts International* **47**(7-B) 3147

Sanocki T, 1987b "Visual knowledge underlying letter perception: Font-specific, schematic tuning" *Journal of Experimental Psychology: Human Perception and Performance* **13** 267 – 278

Sanocki T, 1988 "Font regularity constraints on the process of letter recognition" *Journal of Experimental Psychology: Human Perception and Performance* **14** 472 – 480

Sanocki T, 1991 "Looking for a structural network: effects of changing size and style on letter recognition" *Perception* **20** 529 – 541

Sanocki T, 1992 "Effects of font- and letter-specific experience on the perceptual processing of letters" *American Journal of Psychology* **105** 435 – 458

Schneider W, Shiffrin R M, 1977 "Controlled and automatic human information processing: I. Detection, search, and attention" *Psychological Review* **84** 1 – 66

Sommers M S, Nygaard L C, Pisoni D B, 1994 "Stimulus variability and spoken word recognition: I. Effects of variability in speaking rate and overall amplitude" *Journal of the Acoustical Society of America* **96** 1314 – 1324

Studdert-Kennedy M, 1976 "Speech perception", in *Contemporary Issues in Experimental Phonetics* Ed. N J Lass (New York: Academic Press) pp 243 – 293

Tanaka J W, 2001 "The entry point of face recognition: evidence for face expertise" *Journal of Experimental Psychology: General* **130** 534 – 543

Tanaka J W, Taylor M, 1991 "Object categories and expertise: Is the basic level in the eye of the beholder?" *Cognitive Psychology* **23** 457 – 482

Townsend J T, Ashby F G, 1982 "Experimental test of contemporary mathematical models of visual letter recognition" *Journal of Experimental Psychology: Human Perception and Performance* **8** 834 – 864

Wong A C-N, Gauthier I, submitted, "The basic level as the entry point of expert letter perception" *Visual Cognition*

Wong A C-N, Jobard G, James T W, James K H, Gauthier I, 2005 "Experience-associated neural selectivity to single characters", paper presented at the Human Brain Mapping Conference, June, Toronto, Canada

Wong P C, Nusbaum H C, Small S L, 2004 "Neural bases of talker normalization" *Journal of Cognitive Neuroscience* **16** 1173 – 1184

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION

VOLUME 35 2006

www.perceptionweb.com