Vision Research 62 (2012) 57–65

Contents lists available at SciVerse ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Seeing visual word forms: Spatial summation, eccentricity and spatial configuration

Chien-Hui Kao [1], Chien-Chung Chen *

Department of Psychology, National Taiwan University, Taipei 106, Taiwan

ARTICLE INFO

Article history:
Received 30 November 2011
Received in revised form 14 March 2012
Available online 4 April 2012

Keywords:
Text perception
Detection
Discrimination
Ideal observer
Chinese character

ABSTRACT

We investigated observers' performance in detecting and discriminating visual word forms as a function of target size and retinal eccentricity. The contrast threshold of visual words was measured with a spatial two-alternative forced-choice paradigm and a PSI adaptive method. The observers were to indicate which of two sides contained a stimulus in the detection task, and which contained a real character (as opposed to a pseudo- or non-character) in the discrimination task. When the target size was sufficiently small, the detection threshold of a character decreased as its size increased, with a slope of −1/2 on log–log coordinates, up to a critical size at all eccentricities and for all stimulus types. The discrimination threshold decreased with target size with a slope of −1 up to a critical size that was dependent on stimulus type and eccentricity. Beyond that size, the threshold decreased with a slope of −1/2 on log–log coordinates before leveling out. The data was well fit by a spatial summation model that contains local receptive fields (RFs) and a summation across these filters within an attention window. Our result implies that detection is mediated by local RFs smaller than any tested stimuli and thus detection performance is dominated by summation across receptive fields. On the other hand, discrimination is dominated by a summation within a local RF in the fovea but a cross RF summation in the periphery.

1. Introduction

Reading is one of the most common forms of communication in modern society. The first step in reading is to connect arbitrary visual symbols to their meanings and pronunciations. With training and practice, an adult can identify words effortlessly and efficiently. We are interested in the properties of the mechanism underlying visual word form perception.

Our approach combines two experimental paradigms. The first is the dual task paradigm developed by Thomas (1985a, 1985b). In a dual task experiment, there are two intervals in each trial and an observer has to make two responses in each trial. For instance, in a study by Thomas (1985a, 1985b), one interval contained a noise mask alone while the other contained the noise mask plus one of two possible targets. The observer had to first decide which interval contained the target (detection,) and then which of the two targets was shown (identification.) He found that the identification threshold was higher than the detection threshold. His result suggests that observers may rely on the response of one channel to detect a target and on comparison of different channels to identify it (Thomas, 1985b). That is, the difference in the two tasks reflects the properties of the mechanism that determines performance in these tasks. The second paradigm, the spatial summation paradigm, has been used to estimate the spatial extent of the receptive field of pattern detectors (Barlow, 1958; Tyler & Chen, 2006; Wilson & Wilkinson, 1998). In a typical spatial summation experiment, the task of an observer is to detect target stimuli of various sizes. In general, the target detection threshold decreases with the target size to a critical value. Beyond this critical size, there is little, if any, further threshold. Spatial summation has been used to test whether there is a specific visual mechanism for object processing: for example, for processing faces (Tyler & Chen, 2006). Analysis of spatial summation for visual words could help us to explore the factors that affect the perception of visual word forms.

We are interested in what information in a stimulus is used by the visual system to process visual word forms. It has been shown that recognition of a visual word is based on an analysis of its individual letters, which in turn relies on the analysis of local image features such as the length or width of the strokes (Martelli, Majaj, & Pelli, 2005; Pelli, Farell, & Moore, 2003; Pelli et al., 2006). However, there are also studies showing that analysis of local features alone does not provide enough information for visual word form analysis. Instead, these studies suggest that the spatial relationship between local features, or spatial configuration, is important for visual word form processing. (Kao, Chen, & Chen, 2010; Wong et al., 2011; Yeh et al., 2003). For example, Kao, Chen, and Chen (2010) found that observers have more difficulty identifying an inverted visual word form than an upright one. Since the local features

* Corresponding author.
E-mail address: c3chen@ntu.edu.tw (C.-C. Chen).
[1] Present address: Laboratory for Cognitive Brain Mapping, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.

0042-6989/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.visres.2012.03.015

are the same in both inverted and upright visual word forms, this result suggests that their spatial configuration is important. Wong et al. (2011) showed that matching the target parts of an English word could be interfered with by the other part of the word unless the two parts were not aligned. These studies show that spatial configuration plays an important role in visual word form perception.

To test the spatial configuration effect, we developed a way to manipulate the spatial configuration of components while keeping the local features of the components intact. We used Chinese characters or Hanzi (also called Kanji in Japanese), a character set used in several East Asian languages, as our stimuli. A character consists of several components and each component consists of several strokes. Hence, there are two levels of spatial configurations in a visual word: one is concerned with the spatial relationship among strokes in a character component, while the other, the spatial relationship among components in a character. In this study, we focused on the latter, or the global configuration. The spatial properties of a Hanzi character can be manipulated in a way that allows us to study how the visual system analyzes the spatial configuration of a visual word form.

We used four types of stimuli (real characters, non-characters, *Jiagu* characters, and scrambles) to study the effect of various types of visual information on text perception. All characters consisted of two components in a left–right configuration. In the real characters, the left component signaled the meaning and the right component signaled the pronunciation (see Fig. 1a for an example). Ninety percentage of all Hanzi have this configuration (DeFrancis, 1984). In addition to the real characters (Fig. 1a), non-characters were constructed by swapping the left and right components of a real character (Fig. 1b) to render it unpronounceable. As a control, we also used *Jiagu* characters, which are ancient characters discovered by archaeologists in the late 19th century. A *Jiagu* character has the same left–right configuration as a modern Hanzi character, but contains no components that would be familiar to modern readers (Fig. 1c). Hence, the non-characters kept the local components intact while destroying the rule-based spatial configuration between them, and the *Jiagu* contained no familiar components while the left–right spatial configuration remained intact. Scrambles (Fig. 1d) were also used to establish a baseline.

If there is a specific mechanism for the recognition of visual word forms, we would expect the thresholds for stimuli that are optimal for this mechanism to be lower than the thresholds for other types of stimulus. For instance, if the function of a mechanism is to analyze the spatial configuration of character components within a character, the contrast thresholds should be lower for a *Jiagu* character than for a set of scrambled lines, but on par with other character types. However, if the mechanism responds to the familiarity of components, the threshold for a *Jiagu* character would be higher than for a real character or a non-character, as a *Jiagu* character contains no familiar components.

In addition to the issue of specific word form processing in the visual system, we are also interested in the cortical magnification factors for visual word forms. Eccentricity and visual word form size are critical factors in reading performance (Battista, Kalloniatis, & Metha, 2005; Chung, Mansfield, & Legge, 1998). The cortical magnification factor could be used to estimate the properties of the receptive fields at different eccentricities. In practice, it helps to determine the optimal display size at different eccentricities. Thus, in this study, we presented written words at different eccentricities and varied the size of the word form to compare the spatial summation of the visual word form mechanism in central and peripheral vision.

2. Methods

2.1. Observers

Three observers (KCH, LYL, and ST) participated in the detection task and three observers (KCH, ST, and LYY) were involved in the discrimination task. KCH is an author of this paper and the other two participants were paid observers and naïve as to the purpose of the study. All observers were native Mandarin speakers and had corrected to normal (20/20) visual acuity.

2.2. Stimuli

Fig. 1 shows examples of the stimuli used in the study. In the detection task, we used real characters and non-characters. The real characters were randomly selected from the 1500 most frequently used characters of the 5656 frequently used characters in the Academia Sinica Balanced Corpus (1998). The non-characters were selected from the norms prepared by Hue and Tzeng (2000). In the discrimination task, for a finer analysis, we also used pseudo-characters, which contain the same components as a real character and are organized with the same character construction conventions. We were interested in whether discrimination performance is affected by orthographic construction. The pseudo-characters, like the non-characters, were selected from the norm of Hue and Tzeng (2000). The average "character-likeness" was 2.4 for our pseudo-characters and 5.4 for the non-characters, on a 7-point Lickert scale (1: most like a real character; 7: most unlike a real character). The *Jiagu* characters were selected from the Digital Archive of the Oracle Bones Rubbing (Institute of History and Philology, Academia Sinica). The average number of strokes in real characters, pseudo-characters, non-characters, and *Jiagu* characters is 11.7, 11.4, 11.5, and 11.3, respectively. Since there is no significant difference in the number of strokes among these types of characters ($F(3, 152) = 0.46$, $p > .05$), these characters have a similar visual complexity. The scrambles were constructed by dividing the image of a character into 16 vertical strips. Then the positions of the 16 vertical strips were scrambled. There were 40 instances of

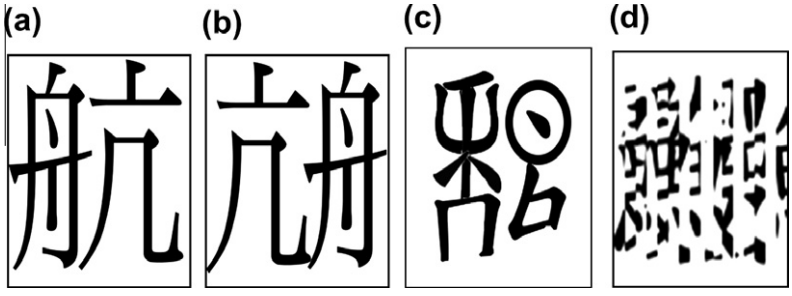

Fig. 1. Sample stimuli: (a) real character; (b) non-character; (c) *Jiagu* character; and (d) scramble.

each type of stimulus. The root mean squared error (RMSE) of five types of stimuli (real characters, pseudo-characters, non-characters, *Jiagu* characters, and scrambles is 49.58, 49.59, 50.29, 49.85, 49.59, respectively) showed there was no significant difference among the intensity of these images ($F(4, 190) = 0.48$, $p > .05$). The stimuli were presented at 1°, 2°, 4°, and 8° away from the central fixation along the horizontal meridian. They were presented at 10 visual angles, from 0.8° to 9.1° in half-octave increments.

2.3. Apparatus

The stimuli were presented on a HP P1130 (Trinitron 21 in. CRT) monitor controlled by a RADEON 9800 XT video card on a PC. The monitor resolution was 1280 (H) × 1024 (V) and the monitor input–output intensity function was measured with a light mouse photometer (Tyler & McBride, 1997). The mean luminance of the monitor was 30 cd/m^2. At a viewing distance of 54 cm, each pixel on the monitor equaled 0.03° in visual angle.

2.4. Procedure

We used a spatial two-alternative forced choice (2AFC) paradigm and the Ψ threshold-seeking algorithm (Kontsevich & Tyler, 1999) to measure the threshold at a 75% correct level. In the detection task, the experiment was composed of four blocks (real characters, non-characters, *Jiagu* characters, and scrambles.) In each trial, a stimulus was randomly presented on one side of the central fixation while the other side was blank. The duration of each presentation was 100 ms. Each trial had a stimulus presented at 1°, 2°, 4° or 8° from the central fixation. The observers were instructed to indicate which side contained a stimulus. The discrimination task was composed of four blocks (real characters vs. pseudo-characters, real characters vs. non-characters, real characters vs. *Jiagu* characters, and real characters vs. scrambles). The target was randomly presented on one side of the fixation, while the other side showed one of the other three types of stimuli. The two stimuli were presented simultaneously, at the same size, for 100 ms. Observers indicated which side contained a real character. The order of presentation within a block was randomized and the order of blocks was counterbalanced across observers. Feedback was provided via an auditory cue for both correct and incorrect trials. The observers were asked to focus on the center fixation point during the entire experiment. There were four threshold measurements and 40 trials for each threshold measurement. The contrast threshold values were the average of the four threshold measurements.

3. Results

Fig. 2 shows the contrast threshold for all types of stimulus at different eccentricities in the detection task. Each column of Fig. 2 denotes the data from one observer, and each row, an eccentricity. In each panel, the blue[2] circle, green square, pink triangle, and red inverted triangle denote real characters, non-characters, *Jiagu* characters, and scrambles respectively. The smooth curves are fits of the model discussed below.

For detection, the summation curves were similar for all eccentricities and stimulus types. When the target size was small, the detection threshold of a character decreased with the increase in its size, with a slope of −1/2 (indicated by the solid lines in Fig. 2) on log–log coordinates, up to a critical size. As character size increased beyond that critical size, there was little, if any, improvement in detectability. In general, character type had little effect on detection threshold ($F(3, 28)$ is between 0.28 to 0.75, $p > .05$ in eccentricity 1; $F(3, 36)$ is between 0.42 to 0.48, $p > .05$ in eccentricity 2; $F(3, 40)$ is between 0.18 to 0.47, $p > .05$ in eccentricity 4; $F(3, 40)$ is between 0.30 to 0.54, $p > .05$ in eccentricity 8). The apparent difference between the threshold for *Jiagu* characters and the thresholds for the other character types was within the noise range. The lack of effect of character type on detection suggests that it is a mediated mechanism whose response is not affected by familiarity or global features of a stimulus.

Both threshold and critical size depend on eccentricity. For a better illustration of the eccentricity effect, Fig. 3 replots the data in Fig. 2 from another viewpoint. Each column of Fig. 3 denotes the data from one observer, and each row a stimulus type. In each panel, the blue circle denotes the data collected at 1° of eccentricity; the green square, 2°; the pink triangle, 4°, and the red inverted triangle, 8°. Again, as the target size increased, the threshold decreased, up to a critical size, and then leveled out. For all character types, the threshold increased with eccentricity. For the small targets (i.e., smaller than the critical size), the threshold for stimuli at 8° of eccentricity was about 10.3–11.2 dB, or 3.3–3.6 times greater than at 1° of eccentricity. For large targets, the threshold for 8° peripheral stimuli was asymptotic at a level about 5–7 dB, or 1.6–2.3 times greater than the threshold for foveal stimuli. For all types of characters, the critical size increased with eccentricity. This suggests a cortical magnification factor for the receptive field of the detection mechanisms.

Fig. 4 shows the contrast discrimination threshold for all stimulus types at different eccentricities. As in Fig. 2, each column denotes the data from one observer, and each row an eccentricity. In each panel, the blue circle, green square, pink triangle, and red inverted triangle denote the compared targets: pseudo-characters, non-characters, *Jiagu* characters, and scrambles respectively. The smooth curves are fits of the model also discussed below.

The discrimination threshold decreased with target size with a slope of −1 (indicated by the solid lines in Fig. 4) up to a critical size that depended on stimulus type and eccentricity. At the same eccentricity and size, the contrast thresholds for discriminating between the real and pseudo-characters were the same as those for discriminating between the real and non-characters but were higher than for discriminating between real characters and *Jiagu* characters or scrambles. For smaller targets, the difference was about 7 dB, or a twofold increase in the discrimination threshold. The critical size for discriminating between real characters and pseudo- or non-characters was larger than for discriminating between real characters and *Jiagu* characters or scrambles. This suggests that the mechanism underlying discrimination is sensitive to character-like stimuli.

Fig. 5 replots the data from Fig. 4 to show the effect of eccentricity. Each column in Fig. 5 denotes the data from one observer, and each row a stimulus type. In each panel, the blue circle denotes the data collected at 1° of eccentricity; the green square, 2°; the pink triangle, 4°; and the red inverted triangle, 8°. As the target size increased, the threshold decreased to a critical size, and then leveled out. For all types of characters, the threshold increased with eccentricity. On the summation curve, for the same stimuli, the target threshold at 8° eccentricity was −5 dB greater than the threshold at the fovea for all types of targets. When the character size was larger than the critical size, the 8° peripheral threshold was asymptotic to a level about 8.58 dB (10 dB, 9.67 dB, 6.67 dB, and 8 dB for pseudo-characters, non-characters, *Jiagu* characters, and scrambles, respectively), or 2.69 times, greater than that of the foveal target. Again, there was a cortical magnification factor for the receptive field of the discrimination mechanisms.

At the same eccentricity, when stimuli were small, thresholds for the discrimination task had greater contrast, compared to those

[2] For interpretation of color in Figs. 2–5, the reader is referred to the web version of this article.

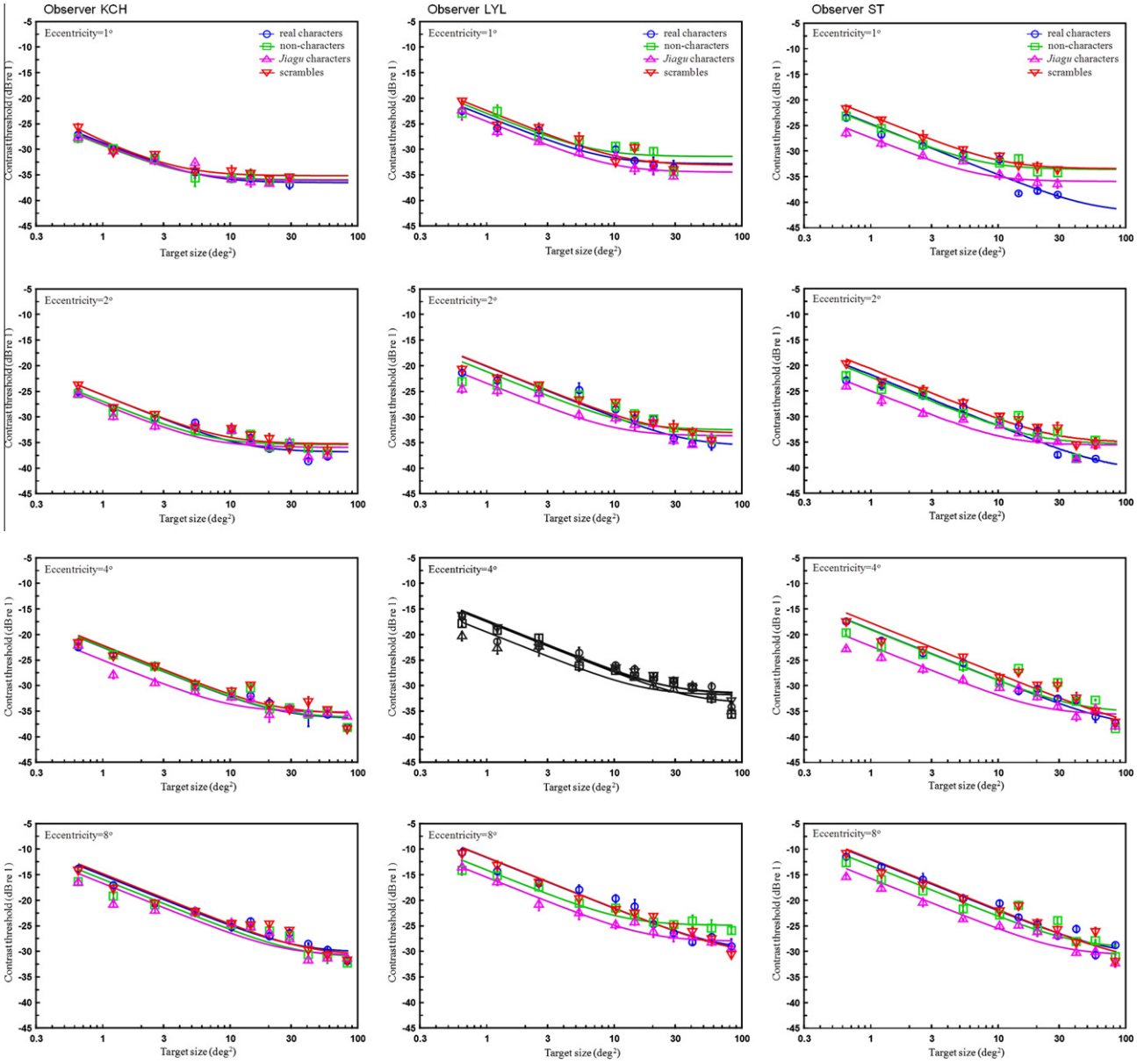

Fig. 2. The contrast threshold for all types of stimuli at different eccentricities in the detection task; detectability for different types of stimuli. Each panel represents different eccentricities in the detection task. Each row denotes one observer and each column denotes each eccentricity. The spatial summation curves are fits of the model defined in the text. The error bars are one standard error of the means.

in the detection task, for all character types. Beyond the critical size, there was no difference in threshold for the detection and discrimination task.

4. Discussion

In this study, we found that detection thresholds were the same for all types of stimulus, given the same stimulus size and eccentricity. There was no significant difference in the detection task among different types of stimulus. This implies that there is no specific mechanism for visual word forms involved in the detection task. Detecting visual word forms relies on analysis of the local features of a stimulus. This was consistent with Pelli's finding that letter detection is limited only by local image features (Pelli, Farell, & Moore, 2003; Pelli et al., 2006). In the discrimination task, the threshold for discriminating between real and pseudo- or non-characters was higher than for discriminating between real characters and *Jiagu* characters or scrambles. This suggests a discrimination mechanism that is not sensitive to the difference between real, pseudo- and non-characters: i.e., it is insensitive to the arrangement of character components. It is easier to tell the difference between real and *Jiagu* characters, even though they share the same left–right spatial configuration between components. In addition, the threshold for discriminating real characters and scrambles was similar to that for discriminating real and *Jiagu* characters, even though the scrambles did not have the left–right configuration, indicating that similarity in spatial configuration has little effect on character discrimination. Notice that the detection threshold for *Jiagu* characters and scrambles was similar to that for real characters. Thus, the difference in statistical properties between real and *Jiagu* characters cannot explain why it is easier to discriminate between them. Instead, these results suggest that the visual word form discrimination mechanism is insensitive to the spatial configuration of the components.

Notice that our result does not suggest that spatial configuration plays *no* role in visual word form detection and discrimination.

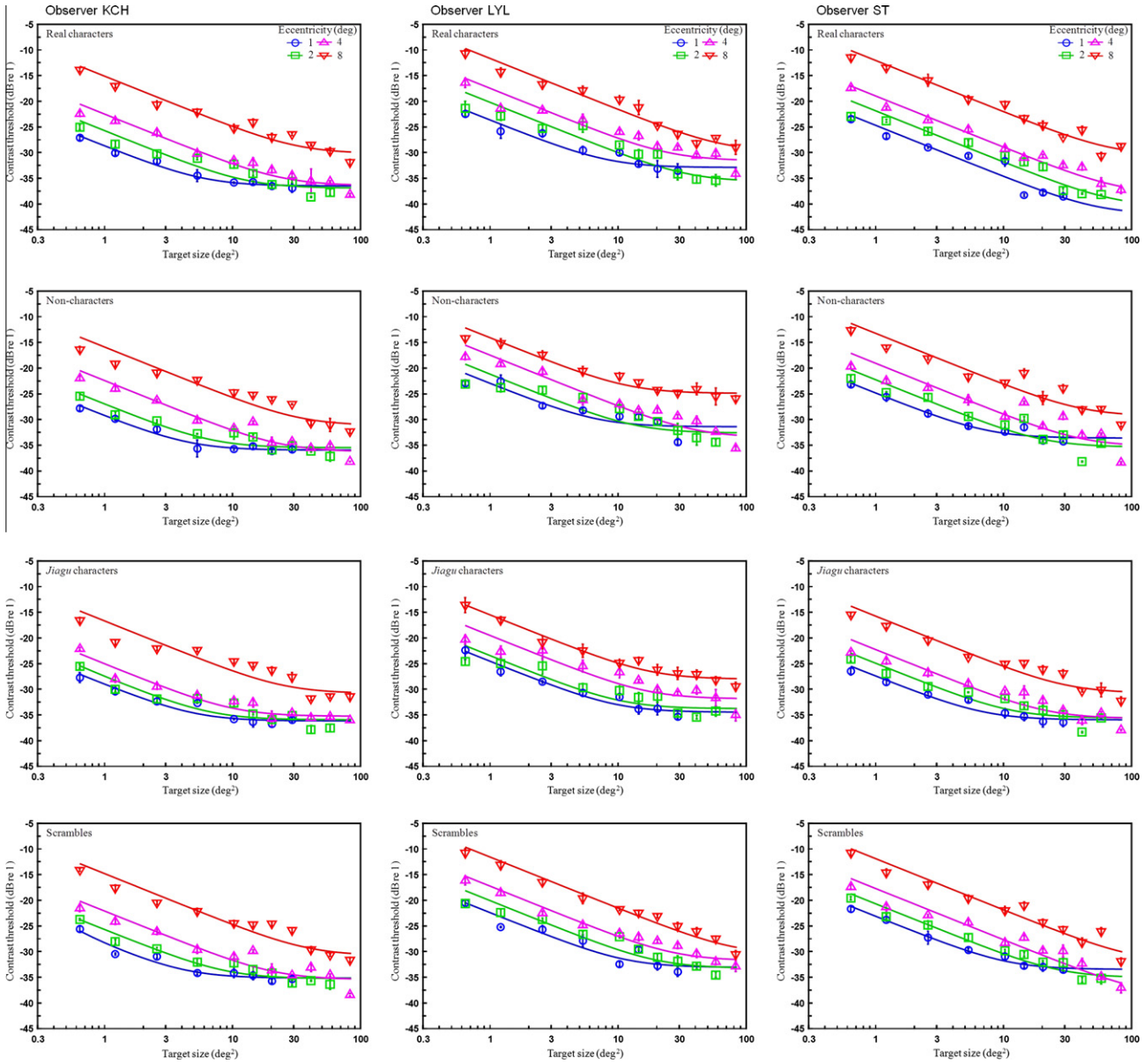

Fig. 3. Detectability for different eccentricities. Each row denotes one observer and each column denotes a stimulus type; from top to bottom: real characters, non-characters, *Jiagu* characters and scrambles, respectively. The spatial summation curves are fits of the model defined in the text. The error bars are one standard error of the means.

After all, we only manipulated the spatial configuration between components, not between strokes within a component. It is still possible that spatial configuration between strokes in a component is essential for visual word form detection and discrimination. That is, the spatial configuration effect reported in the literature (Kao, Chen, & Chen, 2010; Yeh & Li, 2002; Yeh et al., 2003) may result from a change of spatial configuration within a component. Therefore, it is possible that the mechanism for visual word form processing is to process a character component as a unit, rather than the whole character. This is consistent with studies that the recognition of a character relies on its components (Chen, Allport, & Marshall, 1996; Tsang & Chen, 2009).

4.1. Summation model

Spatial summation in visual target detection is a well-studied phenomenon. Principles governing spatial summation in different circumstances have been established in the literature. In the context of this study, two types of spatial summation are particularly relevant. The first is Ricco's law, or that the target threshold is inversely proportional to the target size (Barlow, 1958; Baumgardt, 1959). That is,

$$\theta = (s_1 \cdot A^{-1}) \tag{1}$$

where θ is the threshold, A is the target area and s_1 is a constant. This law applies when the target is sufficiently small and lies completely within the receptive field of a detector. As the target size increases, the overlap between the detector and the stimulus increases and in turn, generates a proportional increase in the response of the detector. The threshold decreases proportionally with the target size and thus has a slope of -1 in log–log coordinates.

The second is Piper's law, or that the target threshold is inversely proportional to the square root of target size. That is,

$$\theta = (s_2 \cdot A^{-1/2}) \tag{2}$$

where s_2 is a constant. This summation can occur when the target is large enough to cover the receptive field of several local detectors.

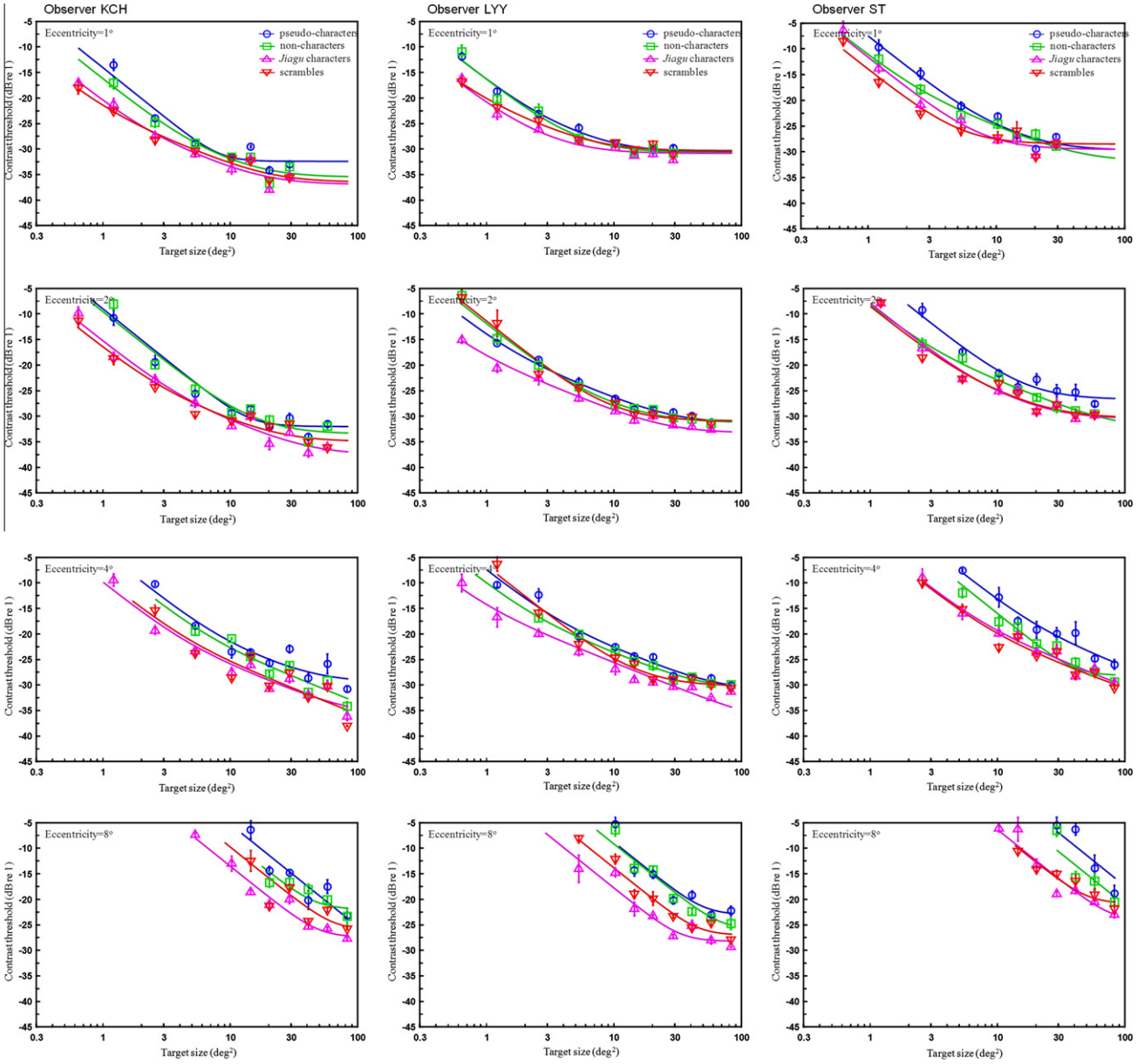

Fig. 4. The contrast threshold for all types of stimuli at different eccentricities in the discrimination task; the discriminability for different types of stimuli. Each column denotes one observer and each row denotes an eccentricity. The spatial summation curves are fits of the model defined in the text. The error bars are one standard error of the means.

Suppose that the threshold is determined by the combined response of the detectors whose receptive field overlaps with the target. As the target increases in size, it covers the receptive fields of more and more detectors. Assume that the noise in those detectors has the same distribution. The mean and the variance of the combined response increase with target size. Thus, the detectability of the target, usually defined as d' or the ratio between the mean and the standard deviation of the response (Green & Swets, 1966), increase with the square root of target size and in turn the target threshold decreases with target size with a slope of $-1/2$ (Tyler & Chen, 2000). An alternative to this scenario is that the threshold is determined by the detector with the greatest response to the target. In this case, the summation function should have a slope $-1/4$ (Tyler & Chen, 2000). However, there is no evidence of such summation in our data.

Piper's law has its own limitation. In any given trial, the visual system can only monitor a limited number of detectors (Pelli, 1985). Hence, if the target already covers these monitored detectors, or the attention window, a further increase of target size would not make a contribution to the response of the visual system. That is, when the target size is even larger than the attention window of the system, the threshold remains constant regardless of further increase in target size.

A summation curve then contains three segments: when the target size is smaller than the receptive field of the smallest visual word form detector, the threshold decreases proportionally with the target size and thus the summation curve would have a slope of -1 in log–log coordinates (Eq. (1)). When the target size is within the attention window of the system, the threshold decreases with the increase in target size, with a slope of $-1/2$ on log–log coordinates (Eq. (2)). Finally, when the target size is larger than the attention window, the threshold remains constant regardless of the target size. That is

$$\theta = s_3, \tag{3}$$

where s_3 is a constant.

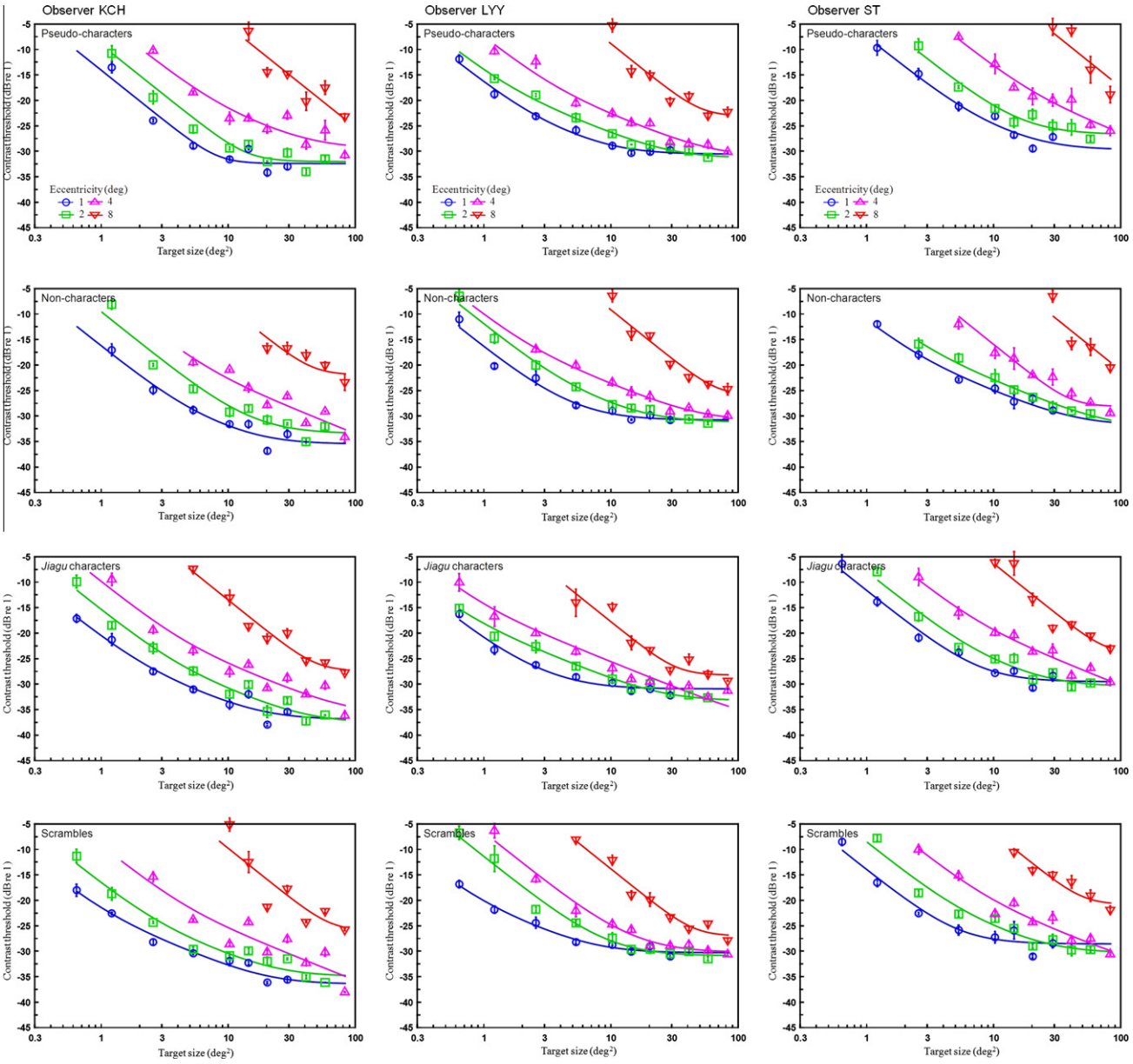

Fig. 5. The contrast threshold for all types of stimuli at different eccentricities in the discrimination task. The discriminability for eccentricities. Each row denotes one observer and each column denotes a stimulus type. The spatial summation curves are fits of the model defined in the text. The error bars are one standard error of the means.

We combined the three segments with a fourth-power Minkowski summation that provides for smooth transitions between the segments of the curve (Tyler & Chen, 2006) to fit our data. That is, the thresholds θ for a character stimulus were fit with an equation:

$$\theta_i = \left((s_{1i} \cdot A^{-1})^4 + (s_{2i} \cdot A^{-1/2})^4 + s_{3i}^4\right)^{1/4} \quad (4)$$

where the subscript i denotes the character type, and s_1, s_2, and s_3 are relative weights of the three summation segments. The area of A was computed as ω^2, where ω is the width of the character stimulus. The smooth curves in Figs. 2–5 are the fits of this model. Table 1 lists the fit parameters of s_1, s_2, and s_3. The RMSE of the model fit was between 1.29 to 1.53 dB in the detection task and 1.12 to 1.37 dB in the discrimination task for all stimulus types. This is only slightly larger than the mean standard error of measurement, which ranged from 0.53 to 0.68 dB in the detection task and 0.68 to 0.81 dB in the discrimination task. This model accounted for 98.6% of all variation in the data. That is, this model accounts for virtually all the systematic variations in the data.

If we consider the detection data only, the model can be simplified further. The summation curve at the smaller target size, as shown in Figs. 2 and 3, actually had a slope of −1/2, as indicated by the solid curves in the figures. Hence, we tested whether a reduced model with $s_1 = 0$ could fit the data. That is, the thresholds θ_i was fitted with

$$\theta_i = \left((s_{2i} \cdot A^{-1/2})^4 + s_{3i}^4\right)^{1/4} \quad (5)$$

The goodness of fit for the reduced model (Eq. (5)) is similar to the full model (Eq. (4)). The sum of square error (SSE) pooled across observers increases from 75.33 for the full model to 75.36 for the reduced model. This difference is not significant ($F(1, 22) < 0.01$, $p > .05$). The implication is that even the smallest character used in our experiment is larger than the receptive size of the detectors.

Table 1
Parameters s_1, s_2, and s_3 in the detection and discrimination task.

Eccentricity	s_1	s_2	s_3
Detection task			
1	0.01	0.06	0.03
2	0.01	0.08	0.03
4	0.01	0.12	0.06
8	0.30	0.28	0.30
Discrimination task			
1	0.2	0.06	0.4
2	0.4	0.08	0.4
4	0.95	0.12	0.4
8	4.8	0.06	0.1

Detection performance is dominated by a summation of localized detectors within an attention window.

All terms in the full model (Eq. (4)) were necessary to fit the discrimination data. The mean square error was 1.28 and was close to the standard error of measurement. Notice that, a segment of −0.5 slope was necessary for the model to fit the data. If we set $s_2 = 0$ the goodness of fit became significantly worse ($F(1, 22) = 4.32$, $p < .05$). The model fits confirmed that the summation curve at the smaller target size had a slope of −1, as indicated by the solid curves in Figs. 4 and 5, and then turned into a section of slope −1/2 before leveling out.

Some may question the validity of applying the summation principles originally used for detection to discrimination. However, in the signal detection theory (Green & Swets, 1966), the threshold is determined by d', or the difference between the mean internal responses to the signal and noise scaled by the standard deviation of the noise distribution. In the 2AFC paradigm we used, the task of the observer in the detection experiment was to distinguish the target in one interval from a blank screen in the other, while in the discrimination task, he/she had to distinguish the target from a comparison stimulus. The observer's performance in both tasks was based on the difference between two intervals. Hence, mathematically, both the detection and discrimination thresholds are governed by the same d' computation. Therefore, we can use the same model to fit both detection and discrimination data.

4.2. The size of the visual word form detectors and the eccentricity effect

As discussed above, the segment of the summation curve with slope −1 accounts for the summation within the receptive field of the local detector while the segment with slope −1/2 accounts for the summation across local detectors within the attention window. Hence, the intersection between the segments of the summation curves with slopes of −1 and −1/2 is where the summation within the local detector ends and the summation between detectors begins. That is, this intersection denotes the size of the receptive field of the local detectors, ω_{T1}. One can calculate this size by combining Eqs. (1) and (2) and thus

$$\omega_{T1} = s_1/s_2 \tag{6}$$

Similarly, the intersection of the slope −1/2 and 0 segments denotes the size of the attention window, ω_{T1} that can be calculated by combining Eqs. (2) and (3). That is,

$$\omega_{T2} = s_2/s_3 \tag{7}$$

In the detection task, since there is no −1 slope segment in our summation curve, we can conclude that all our stimuli were larger than the receptive field size of the visual word form detector. Nevertheless, we can still estimate the width of the attention window, ω_{T2}.

Fig. 6 shows ω_{T2} derived from detection data averaged across three observers for different character types and eccentricities. As the eccentricity increased, the size of the attention window also increased ($F(3, 32) = 137.7$, $p < .05$), regardless of stimulus types (i.e., no significant interaction between eccentricity and stimulus type: $F(9, 32) = 2.32$, $p > .05$). The stimulus type had no statistically significant effect on attentional window size ($F(3, 32) = 4.24$, $p > .05$). The filter size was twice as large in the periphery as in the fovea.

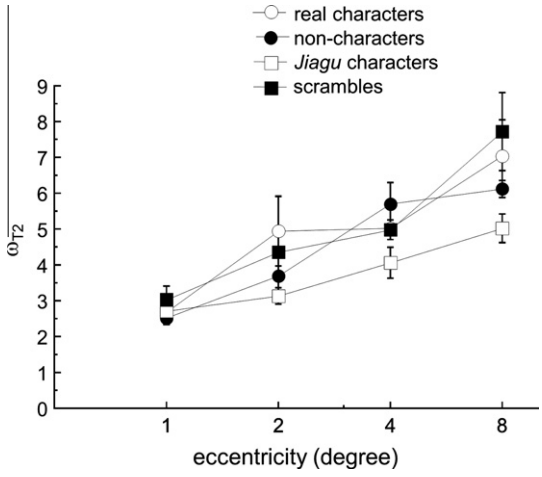

Fig. 6. The width of the attention window, ω_{T2}, for the detection task averaged across three observers. The error bar represents one standard error of measurement.

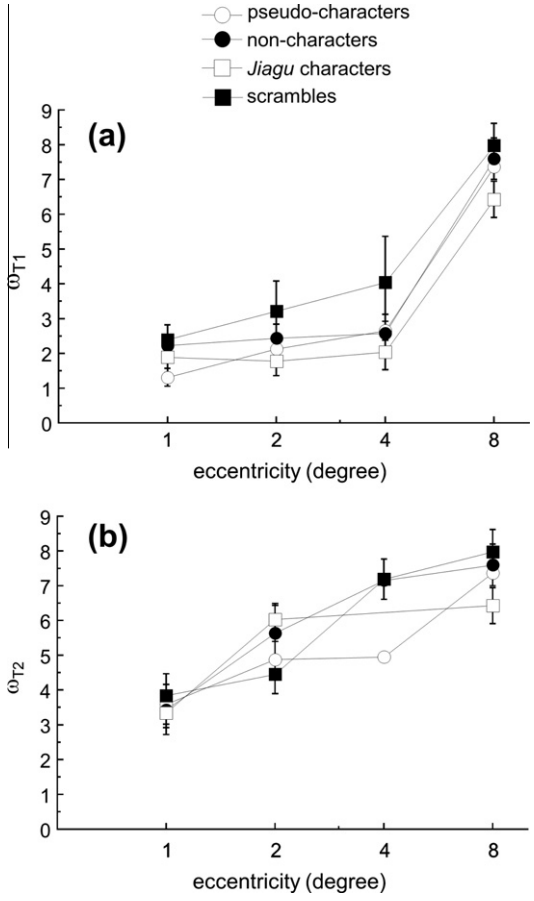

Fig. 7. The width of the local receptive field, ω_{T1}, and the attention window, ω_{T2}, for the discrimination task averaged across three observers. The error bar represents one standard error of measurement.

The summation curves for the discrimination task contain all three segments. Thus, we can estimate the sizes of both the local receptive field, ω_{T1}, and the attention window, ω_{T2}. Fig. 7 shows ω_{T1} (Panel A) and ω_{T2} (Panel B) as a function of eccentricity in the discrimination task. The estimated ω_{T2} of *Jiagu* characters at 4° of eccentricity was beyond the measurable range of the current experiment, and was excluded from further analysis. Both ω_{T1} and ω_{T2} increased with eccentricity ($F(3, 32) = 110.158$, $p < .05$ and $F(2, 24) = 88.15$, $p < .05$ respectively) regardless of stimulus type ($F(9, 32) = 0.57$, $p > .05$ and $F(6, 24) = 1.6$, $p > .05$). The stimulus type had no statistically significant effect on both ω_{T1} and ω_{T2} ($F(3, 24) = 0.2$, $p > .05$ and $F(2, 24) = 88.15$, $p < .05$ respectively). For all stimulus types, the parameter ω_{T2} was larger than ω_{T1} in the central viewing conditions ($t(11) = 4.69–5.68$, $p < .05$ for 1° and 2° eccentricity), while the parameter ω_{T1} was the same as ω_{T2} at the peripheral beyond 8° of eccentricity ($t(11) = 0.53$, $p > .05$). This implies that the observers' performance in discriminating visual word forms is dominated by local filters in the fovea a summation across local filters within the attention window in the periphery.

4.3. Cortical magnification factors

Eccentricity and visual word form size are critical factors in reading performance (Battista, Kalloniatis, & Metha, 2005; Chung, Mansfield, & Legge, 1998). To reveal the relationship between eccentricity and visual word form size, we measured the cortical magnification factors for visual word form processing with Eq. (8). The relationship between the critical size ω_T and eccentricity E can be described by

$$\omega_T = \omega_{T0} * (1 + E/E_2) \quad (8)$$

where E is the eccentricity of the stimulus, ω_{T0} is the estimated critical size at the fovea and E_2 controls the cortical magnification factor. ω_T and E_2 were free parameters to be estimated. The E_2 parameter is used to represent the rate of change of the variable of interest as a function of eccentricity (Chung, Mansfield, & Legge, 1998; Levi, Klein, & Aitsebaomo, 1985; Toet & Levi, 1992). We found the cortical magnification, E_2, at 0.82° visual angle, fitted detectability for visual word forms. From fovea to periphery, the critical size increased 4.23-fold. These results were consistent with the finding of Levi, Klein, and Aitsebaomo (1985), in which the E_2 was 0.68 for identifying the acuity of alphabetic words. In addition, critical size increases with eccentricity with E_2 0.22 in the discrimination task. The target size in the periphery was 5.32 times greater than in the fovea. The E_2 value was lower in the discrimination task than in the detection task, implying that the variable changes more quickly with eccentricity for discriminating visual word forms than for detecting visual word forms (Chung, Mansfield, & Legge, 1998).

5. Conclusion

This study demonstrated the spatial extent of the receptive fields of the visual word form mechanism for detection and discrimination tasks. The summation curves for detecting different types of visual word forms were similar. This implies that the response of the detection mechanisms depends on the local features of a visual word form, such as line segments. It is easier to discriminate real characters from pseudo- or non-characters than real characters from *Jiagu* characters or scrambles. This implies that the discrimination mechanism is sensitive to the familiarity of the components of a character rather than spatial configuration per se. We fit a model containing the local receptive fields and a summation across these filters within an attention window to the data. From the model fits, we further confirmed that detection performance for visual word forms is mediated by a summation of localized detectors within an attention window while discrimination performance is mediated, in the fovea, by a summation of receptive fields and by a summation across local filters within the attention window in the periphery. In addition, the cortical magnification factor is greater for the detection mechanism than for the discrimination mechanism. This implies that the discrimination mechanism focuses more on foveal stimuli.

Acknowledgment

This study is supported by National Science Council (Taiwan), NSC 96-2752-H-002-007-PAE and NSC 99-2410-H-002-081-MY3 to C.C.C.

References

Academia Sinica Balanced Corpus (1998). (Version 3) [CDROM] from Academia Sinica.
Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. *Journal of Physiology, 141*(2), 337–350.
Battista, J., Kalloniatis, M., & Metha, A. (2005). Visual function: The problem with eccentricity. *Clinical and Experimental Optometry, 88*(5), 313–321.
Baumgardt, E. (1959). Visual spatial and temporal summation. *Nature, 184*(Suppl. 25), 1951–1952.
Chen, Y. P., Allport, D. A., & Marshall, J. C. (1996). What are the functional orthographic units in Chinese word recognition: The stroke or the stroke pattern? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49*, 1024–1043.
Chung, S. T., Mansfield, J. S., & Legge, G. E. (1998). Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision. *Vision Research, 38*(19), 2949–2962.
DeFrancis, J. (1984). *The Chinese language: Fact and fantasy*. Honolulu: University of Hawaii Press.
Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
Hue, C. W., & Tzeng, A. (2000). Lexical recognition task: A new method for the study of Chinese character recognition. *Acta Psychologia Sinica, 32*, 60–65.
Kao, C. H., Chen, D. Y., & Chen, C. C. (2010). The inversion effect in visual word form processing. *Cortex, 46*(2), 217–230.
Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research, 39*(16), 2729–2737.
Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research, 25*(7), 963–977.
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), 58–70.
Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A: Optics and Image Science, 2*(9), 1508–1532.
Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46*(28), 4646–4674.
Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature, 423*(6941), 752–756.
Thomas, J. P. (1985a). Detection and identification: How are they related? *Journal of the Optical Society of America A: Optics and Image Science, 2*(9), 1457–1467.
Thomas, J. P. (1985b). Effect of static-noise and grating masks on detection and identification of grating targets. *Journal of the Optical Society of America A: Optics and Image Science, 2*(9), 1586–1592.
Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research, 32*(7), 1349–1357.
Tsang, Y. K., & Chen, H. C. (2009). Do position-general radicals have a role to play in processing Chinese characters? *Language and Cognitive Processes, 24*(7/8), 947–966.
Tyler, C. W., & Chen, C. C. (2000). Signal detection theory in the 2AFC paradigm: Attention, channel uncertainty and probability summation. *Vision Research, 40*(22), 3121–3144.
Tyler, C. W., & Chen, C. C. (2006). Spatial summation of face information. *Journal of Vision, 6*(10), 1117–1125.
Tyler, C. W., & McBride, B. (1997). The Morphonome image psychophysics software and a calibrator for Macintosh systems. *Spatial Vision, 10*, 479–484.
Wilson, H. R., & Wilkinson, F. (1998). Detection of global structure in Glass patterns: Implications for form vision. *Vision Research, 38*(19), 2933–2947.
Wong, A. C., Bukach, C. M., Yuen, C., Yang, L., Leung, S., & Greenspon, E. (2011). Holistic processing of words modulated by reading experience. *PLoS ONE, 6*(6), e20753.
Yeh, S. L., & Li, J. L. (2002). Role of structure and component in judgments of visual similarity of Chinese characters. *Journal of Experimental Psychology: Human Perception and Performance, 28*(4), 933–947.
Yeh, S. L., Li, J. L., Takeuchi, T., Sun, V., & Liu, W. R. (2003). The role of learning experience on the perceptual organization of Chinese characters. *Visual Cognition, 10*(6), 729–764.