PSYCHOPHYSICAL ANALYSES OF THE SIZE EFFECTS OF SPATIAL ATTENTION ON FIGURE–GROUND ASSIGNMENT

Souta Hidaka[1], Shunji Satoh[2], and Jiro Gyoba[1]
[1]*Department of Psychology, Graduate School of Arts & Letters, Tohoku University*
[2]*Laboratory for Neuroinformatics, Riken BSI*
<hidaka@sal.tohoku.ac.jp, shun@brain.riken.jp, gyoba@sal.tohoku.ac.jp>

Abstract

The figure–ground assignment process is regarded as a preattentive process in hierarchal visual processing. However, some researchers have demonstrated that the spatial location of visual attention affects the figure–ground process. In contrast, the relationship between figure–ground assignment and the spatial size of attention has never been investigated. In this research, we psychophysically analyzed the effect of the spatial size of attention on the figure–ground assignment process. In our experiments, we measured the reaction time (RT) for identifying the edge shapes of the inner or outer regions contained in ambiguous figure–ground stimuli, using compounded letters to manipulate the size of the spatial attention. The results revealed that the RTs for inner edge shapes were faster than those for outer ones, followed by those for local letter discrimination. This indicates that local attention assigns the inner region as the figure, suggesting that the size of spatial attention affects figure–ground assignment.

The function of distinguishing an object with a shape (figure) from its background (ground) is called figure–ground assignment. In hierarchical visual processing models, figure‐ ground assignment is considered to be an autonomous or bottom-up process (Biederman, 1987; Julez, 1984; Kosslyn, 1987; Marr, 1982; Nakayama, He & Shimojo, 1995). This is because Gestalt psychologists have revealed some physical factors that are critical for this process (Graham, 1929; Koffka, 1935; Metzger, 1953; Rubin, 1915/1958). For example, relatively smaller or brighter regions are easily perceived as figures. In contrast, observers find it difficult to distinguish figures from the background when the figure–ground regions are equiluminant (Liebmann, 1927). Livingstone and Hubel (1987) demonstrated that figure–ground reversal frequently occurred when each region of the Rubin's face–vase figure was equiluminant. These dominant effects of physical information suggest that the figure‐ ground assignment process does not depend on form attention or object-level knowledge.

In contrast, it has been also pointed out that voluntary or spontaneous attention affects the figure–ground assignment process (Rubin, 1915/1958). For example, Baylis and Driver (1995; Driver & Baylis, 1996) investigated the effect of the location of spatial attention on this process. First, they demonstrated that the edge shape of a figure region is identified faster than that of a ground region in ambiguous figure–ground stimuli. Based on this, they investigated the manner in which a spatial precue affects the shape identification performance with respect to ambiguous stimuli. The results revealed that the cued region was identified faster than the uncued region when the precue validly predicted the direction of the test figures. However, the difference diminished when the direction of the test figures was independent of the precue or the precue was invalid. From these results, the researchers concluded that voluntary or endogenous spatial attention affects figure–ground assignment. Recently, it was demonstrated that automatic or exogenous spatial attention influences it, too.

Vecera, Flevaris, and Filapek (2004) pointed out two problematic areas in Baylis and Driver's experiment. First, the precue was presented outside of the figure–ground stimuli. Thus, it could be assumed that the precue had only an indirect effect on the figure–ground stimuli. The second problem pertained to the abrupt onset of the figure–ground stimuli, which could have cancelled the effect of the precue (Yantis & Jonides, 1984). Vecera et al. replicated Baylis and Driver's experiment by eliminating these problems, thereby confirming that automatic or exogenous spatial attention also affects the figure–ground assignment process, depending on the spatial location of visual attention.

As described above and in contrast with the traditional concepts, some researchers suggested that visual attention modulates the figure–ground assignment process. However, only the effect of the location of spatial attention has been reported so far, and several models have been suggested to illustrate the spatial components of visual attention. For example, the spatial shift of attention has been represented in the form of the spotlight model (Posner, Snyder & Davidson, 1980). Another metaphor called the zoom lens model suggests that spatial attention not only shifts in space but also changes its size (Eriksen & St. James, 1986). In addition, if we assume a limited resource of attention, the relatively larger size of spatial attention might cause global and coarse processing, while a local size of spatial attention would result in fine and detailed processing.

Thus far, the relationship between the spatial size of attention and figure–ground assignment has never been investigated, even though size is also a major characteristic of visual attention. The computational visual model proposed by Satoh (2006) already suggests that the size of spatial attention affects figure–ground assignment; in other words, the obtained figure shape in ambiguous figure–ground stimuli differs depending on the size of spatial attention. In order to achieve a further understanding of the above relationship, the current research employs psychophysical methods to investigate the manner in which the size of spatial attention influences figure–ground assignment.

Similar to the previous research, we measured reaction time (RT) in order to identify the shape of the jagged contour contained in ambiguous figure–ground stimuli (Baylis & Driver, 1995; Driver & Baylis, 1996; Vecera et al., 2004). We used square patterns consisting of inner and outer rectangles as figure–ground stimuli (Figure 1A). However, we already knew that a smaller region would be easily perceived as a figure. In order to suppress this kind of predominant figural perception for the inner rectangle, we set the luminance value of the outer rectangle to be brighter than that of the inner ones in Experiment 1A. In addition, both outer and inner regions were equiluminant in Experiment 2.

Moreover, we used compounded letters to operate the size of spatial attention (Navon, 1977; Figure 1B). Previous studies have proved that the discrimination of small letters requires local attention. In contrast, global processing becomes dominant while discriminating a large letter; therefore, the letter discrimination task could be used to manipulate the size of attention (Stoffer, 1993). We predicted that local attention would facilitate the figural perception of the inner rectangle, so the RTs for the inner edge shapes would be faster than those for the outer ones. In contrast, a reverse tendency would be observed when the size of spatial attention was global (Baylis & Driver, 1995).

We also investigated the temporal aspects of the effect of the size of spatial attention. Stoffer (1993) suggested that it takes 400–600 ms for spreading the size of spatial attention in the range of 1.9×2.1 deg. Based on this finding, it can be hypothesized that the effect of attention would become salient when the stimulus onset asynchrony (SOA) between the figure–ground stimuli and test figures would be less than 600 ms. In contrast, SOA exceeding 400 ms would reset the manipulation of the size of spatial attention, so the effect of attention would disappear in this condition. Based on these predictions, we set SOA as either 100 ms or 600 ms in the two experiments.

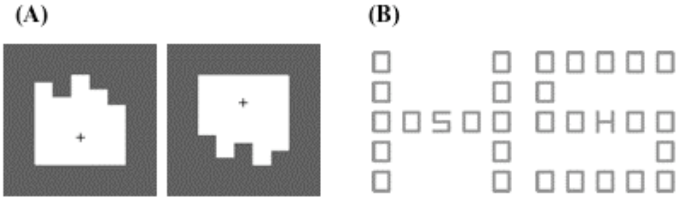

Fig. 1. Stimuli used in the experiments. (A) Ambiguous figure–ground squares. In the experiment, each region was colored either red or green. (B) Compounded letters.

Method

Participants and Apparatus

The participants comprised 16 (Experiment 1A) and 8 (Experiment 1B) undergraduate and graduate students of Tohoku University. All the participants had normal or corrected-to-normal vision, their average binocular visual acuity was 1.3, and all of them were right-handed. Except for one of the authors (S.H.), all the participants were naive with respect to the aims of this experiment. The stimuli were presented on a CRT display (Sony Trinitron GDM-FW900, 24 inch) with a resolution of 1600 × 1200 pixels and a refresh rate of 60 Hz. We used a customized PC (Dell-Dimension 8250) and the E-prime (Psychology Software Tools, Inc.) were used in order to control the experiment. The participants reported their response using a mouse and a numeric keypad, and their heads were fixed on a chin rest.

Stimuli

The ambiguous figure–ground squares consisted of outer and inner rectangles (Figure 1A). The visual angles of the rectangles were 3.3 × 3.3 deg and 2 × 2 deg, and the rectangles, divided by jagged contours, were colored red or green. The size of the jagged counters was, on an average, 0.35 deg. We made 35 types of jagged contours. From the pilot experiment, we consistently set the luminance value of the outer rectangles to be brighter than that of the inner ones in Experiment 1A; the luminance values of red and green were 21.53 cd/m^2 and 13.46 cd/m^2, respectively, when the outer rectangles were red, while the corresponding luminance values were 12.75 cd/m^2 and 20.78 cd/m^2 when the outer rectangles were green. In contrast, in Experiment 1B, based on the experimenters' observation, the luminance values were set at equiluminance using the flicker photometry method; the luminance value of red was 16.72 cd/m^2 and that of green was 16.96 cd/m^2.

The compounded letters were gray (average 7.42 cd/m^2) and 3.1 × 3.0 deg in size (Figure 1B), each letter being 0.4 deg in size. Only the center of the compounded letters had a letter shape, so the participants would focus their attention on the center of letter while attempting to identify the local letters. The global and local letters were either S or H; therefore, there were four types of compounded letters: Ss, Sh, Hs, and Hh. The letters were superimposed on the figure–ground stimuli in order to control the size of spatial attention over the figure–ground stimuli (Vecera et al., 2004; Figure 2). This also suppressed the occurrence of an abrupt onset (Yantis & Jonides, 1984). The probe objects presented as test figures for the edge shape identification task were rectangles of sizes 0.65 × 2 deg, with jagged contours. The probes were orange, with a luminance value of 15.98 cd/m^2, which was the average luminance value of the figure–ground stimuli. All of the stimuli were presented on a gray background (22.08 cd/m^2). The observation distance was 57.3 cm.

Procedure

Figure 2 illustrates the procedure. At the onset of each trial, the participants were instructed regarding the to-be-reported letters (global or local). When they pressed the Enter key on the numeric keypad, a white fixation cross appeared for 1000 ms. Following this, the compounded letters superimposed on the figure–ground stimulus were presented four times, and the four types of compounded letters were sequentially presented for 650 ms. The order of these letters was counterbalanced in each trial. The participants reported each letter shape (S or H) using the corresponding button on the mouse (left or right) within 650 ms. In the next display, the ambiguous figure–ground stimulus was presented. The SOA of this stimulus was 100 ms or 600 ms; this was randomly selected in the trials. Finally, two probes were presented side-by-side at the center of the display. The distance between them was 0.2 deg; their maximum duration was 2000 ms. One of the probes had the same contour shape as the preceding figure–ground stimulus. The probes were oriented either upward or downward, with each probe's shape identical to the inner or the outer edges. Another probe was randomly selected from the other 34 types of jagged contours. The participants reported which probe was similar to the figure–ground stimulus using the corresponding numeric keys (1 or 3).

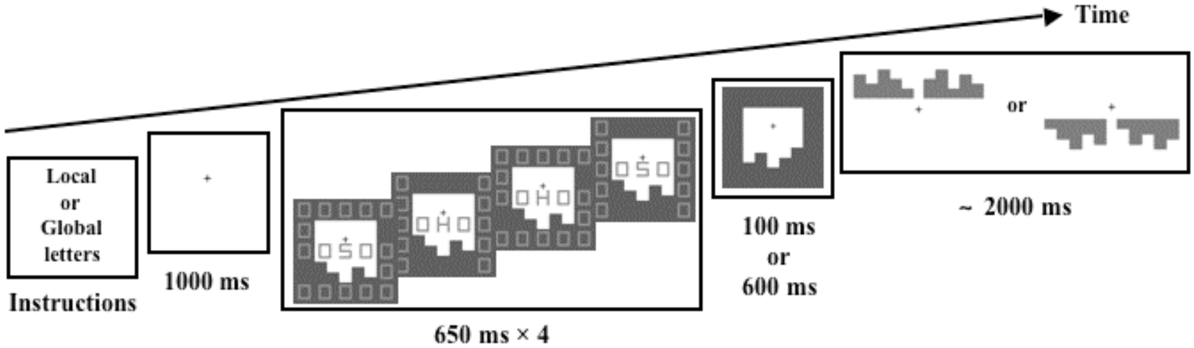

Fig. 2. Procedure of the experiments. In the experiments, the figure–ground stimuli were red and green, while the probe was orange against a gray background, with a white fixation point.

After 40 practice trials, the participants took part in the two main sessions of the experiment. The session was divided based on the positions of the jagged counters (either top or bottom; Figure 1A). Each session consisted of 160 trials (to-be-reported letters (2; global, local) × SOA of the figure–ground stimuli (2; 100 ms or 600 ms) × probe directions (2; upward, downward) × positions of the correct probe (2; left, right) × repetitions (10)). The order of the sessions and the handedness assigned for pressing the mouse or numeric key were counterbalanced among the participants.

Results

We limited our analysis to the data in which the participants correctly discriminated between the last two compound letters. Thereafter, we analyzed only the median RTs for the correct probe identifications. In our analysis, we conducted a three-way repeated ANOVA with letters (global and local) × SOA (100 ms and 600 ms) × probes (inner and outer edges). In both experiments, there were main effects of probes. This showed that the inner edge shapes were identified faster than the outer ones (Experiment 1A: $F(1, 15) = 10.47$, $p < .005$; Experiment 1B: $F(1, 7) = 110.04$, $p < .001$).

Moreover, our analysis identified an interaction among letters, SOA, and probes in Experiment 1A ($F(1, 15) = 15.63$, $p < .001$). When SOA was 100 ms, we obtained an interaction between letters and probes ($F(1, 30) = 4.85$, $p < .03$; Figure 3A). Ryan's post hoc comparison revealed that the inner edge shapes were identified faster than the outer ones when the participants reported the local letters ($F(1, 60) = 14.91$, $p < .001$). In contrast, there was no RT difference when the to-be-reported letters were global ($F(1, 60) = 2.78$, $p > .01$). When SOA was 600 ms, there was no significant interaction between letters and probes ($F(1, 30) = 0.89$, $p > .35$). In Experiment 1B, there was an interaction between letters and probes ($F(1, 7) = 17.85$, $p < .004$; Figure 3B). Ryan's post hoc comparison revealed that the RT for identifying the inner edges was faster than that for the outer ones when the observers reported local letters ($p < .001$), while there was no significant difference for global letters ($p > .72$). Analyses for error rate confirmed that these results did not derive from a speed–accuracy trade-off.

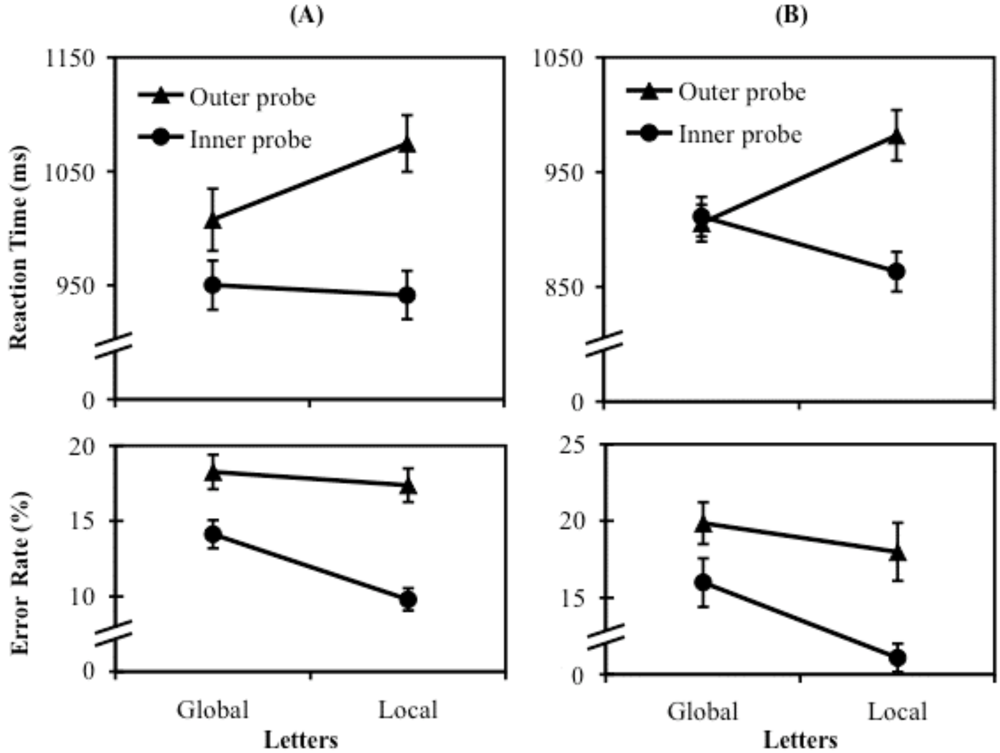

Fig. 3. Reaction time and percentage of errors for the letter–probe interactions in (A) Experiment 1A and (B) Experiment 1B. Error bar denotes the mean standard error.

Discussion

The results of both experiments consistently showed that local letter discrimination promoted the identification of the inner edge shapes, as compared to the outer edge shapes, strongly suggesting that the size of spatial attention affects the figure–ground assignment process. However, we did not observe the effect of global attention on this process. The main effect of probe illustrated that the RTs for identifying the inner probes were faster than those for the outer ones. Thus, it could be inferred that the effect of a small region of focus on figure–ground perception might reduce the effect of global attention. In addition, we found the predicted temporal aspects of the effect of spatial attention in Experiment 1A, but not in Experiment 1B. In Experiment 1A, the luminance bias would clearly separate two regions of the figure–ground stimuli. In contrast, due to the Liebmann effect, a very frequent figure–ground reversal was induced in Experiment 1B. This difference of stimulus information might affect whether the figure–ground assignment process relies on the stable physical properties or the manipulated spatial attention when SOA is 600 ms.

Although figure–ground assignment is regarded as a preattentive process, some researchers have proved that it is influenced by the location of spatial attention. The present research strongly suggests that the process is also modulated by the size of spatial attention.

References

Baylis, B.C., & Driver, J. (1995). One–sided edge assignment in vision: 1. Figure–ground segmentation and attention to objects. *Current Directions in Psychological Science*, *4*, pp. 140-146.

Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, *94*, pp. 115-147.

Driver, J., & Baylis, G.C. (1996). Edge-assignment and figure–ground segmentation in short-term visual matching. *Cognitive Psychology*, *31*, pp. 248-306.

Eriksen, C.W., & St. James, J.D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*, pp. 225-240.

Graham, C.H. (1929). Area, color, and brightness difference in a reversible configuration. *Journal of General Psychology*, *2*, pp. 470-481.

Julez, B. (1984). A brief outline of the text on theory of human vision. *Trends in Neurosciences*, *6*, pp. 41-45.

Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt Brace.

Kosslyn, S.M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review*, *94*, pp. 148-175.

Liebmann, S. (1927). Über das Verhalten farbiger Formen bei Helligkeitsgleichhe von Figur und Grund, *Psychological Research*, *9*, pp. 300-353.

Livingstone, M.S., & Hubel, D.H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience*, *7*, pp. 3416-3468.

Marr, D. (1982). *Vision*: *A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: W.H. Freeman.

Metzger, W. (1953). *Gesetze des Sehens*. Frankfurt, Germany: Waldemar Kramer.

Nakayama, K., He, Z.J. & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S.M. Kosslyn, & D.N. Osherson (Eds.), *Vision.* In *Invitation to Cognitive Science*. Cambridge: M.I.T. Press; pp. 1-70.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*, pp. 353-383.

Posner, M.L., Snyder, C.R.R., & Davidson, B.J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*, pp. 160-174.

Rubin, E. (1915). *Visuell Wahrgenommene Figuren*. Copenhagen: Glydendalske.

Rubin, E. (1958). Figure and ground. In D.C. Beardslee & M. Wertheimer (Eds.). *Reading in perception*. Princeton, NJ: Van Nostrand; pp. 194-203.

Satoh, S. (2006). A visual model for object detection based on active contours and level-set method. *Biological Cybernetics*, *95*, pp. 259-270.

Stoffer, T.H. (1993). The time course of attentional zooming: A comparison of voluntary and involuntary allocation of attention to the levels of compound stimuli. *Psychological Research*, *56*, pp. 14-25.

Vecera, S.P., Flevaris, A.V., & Filapek, J.C. (2004) Exogeneous spatial attention influences figure–ground assignment. *Psychological Science*, *15*, pp. 20-26.

Yantis, S., & Jonides J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, pp. 601-621.