

## PAPER

# Perception and Saccades during Figure-Ground Segregation and Border-Ownership Discrimination in Natural Contours

Nobuhiko WAGATSUMA<sup>†a)</sup>, Mika URABE<sup>††</sup>, Nonmembers, and Ko SAKAI<sup>††b)</sup>, Fellow

**SUMMARY** Figure-ground (FG) segregation has been considered as a fundamental step towards object recognition. We explored plausible mechanisms that estimate global figure-ground segregation from local image features by investigating the human visual system. Physiological studies have reported border-ownership (BO) selective neurons in V2 which signal the local direction of figure (DOF) along a border; however, how local BO signals contribute to global FG segregation has not been clarified. The BO and FG processing could be independent, dependent on each other, or inseparable. The investigation on the differences and similarities between the BO and FG judgements is important for exploring plausible mechanisms that enable global FG estimation from local clues. We performed psychophysical experiments that included two different tasks each of which focused on the judgement of either BO or FG. The perceptual judgments showed consistency between the BO and FG determination while a longer distance in gaze movement was observed in FG segregation than BO discrimination. These results suggest the involvement of distinct neural mechanism for local BO determination and global FG segregation.

**key words:** eye movements, border-ownership, figure-ground segregation, grouping, visual perception

## 1. Introduction

When we view a scene and recognize an object within, we need to determine a region where the object occupies in the scene [1]. This problem is known as figure-ground segregation, which has been generally considered to require global scene understanding to solve the problem. From a view point of robot vision, the necessity of global scene understanding appears to be burden for the system. It appears to be beneficial if local image features can be used for the global figure-ground estimation. To explore plausible mechanisms that enable global figure-ground estimation from local image features, we investigated human strategies for global and local judgements in the segregation of an object from background.

Physiological studies have reported that the responses relevant to figure-ground arise from early- to intermediate-level visual areas [2]–[5]. A recent physiological study reported that figure-ground-dependent modulation in the primary visual cortex (V1) follows each fixation saccade, indicating autonomous bottom-up processing of figure-ground

segregation originating from an early stage [6]. von der Heydt and his colleagues reported that a majority of neurons in the monkey secondary visual area (V2) showed the selectivity to border-ownership (BO), their responses depending on which side of a border owns the contour [7]–[10]. BO is considered as a contour-based, *local* representation of the direction of figure (DOF). Note that we consider figure-ground (FG) as a *global* representation unless otherwise noted in the following sections. Carlson et al. [11] reported that the responses of the curvature-selective neurons in monkey visual area 4 (V4) showed the dependence on the FG organization of stimuli, *i.e.* the cells responded only when its preferred curvature was the contour of figure. It is expected that a surface-based local representation of FG is available in the intermediate-level visual areas.

Although BO signals could underlie the neural representation of FG, how the local BO signals contribute to global FG segregation has not been clarified. The extent of the receptive fields of the cells provides a crucial clue for understanding the mechanisms of BO and FG determination. Although the determination of FG needs a certain extent of view around an object, sometimes across a scene, the receptive fields of the BO-selective neurons reported in V2 extend only 4° in visual angle [7]. These neurons appear to code the DOF along a local contour with the influence from regions surrounding the receptive field [12]. The integration of the population responses of BO-selective cells might underlie FG segregation because the spatial pooling provides an access to global information which would make possible the cortical representation of a figural region [13]–[17]. However, little has been clarified on the link between BO and FG determination. An alternative could be possible that mechanisms independent of BO discrimination evoke FG segregation. It could also be possible that a mechanism takes place in addition to BO processing for the perception of FG. The perceptual and behavioral characteristics of BO and FG determination are expected to provide crucial information for the investigation of the neural and perceptual mechanisms underlying the figure-ground segregation. For instance, it would be crucial for prospective psychophysical experiments if the difference in the instruction for participants, whether to judge BO or FG, changes the behavior. It is important to clarify the differences and similarities in the determination of local BO and global FG.

The eye movement in the determination of BO and FG would provide important evidence in the investigation of FG segregation. Indices based on eye movements are used in a

Manuscript received January 19, 2019.

Manuscript revised September 8, 2019.

Manuscript publicized January 27, 2020.

<sup>†</sup>The author is with Toho University, Funabashi-shi, 274–8510 Japan.

<sup>††</sup>The authors are with University of Tsukuba, Tsukuba-shi, 305–8573 Japan.

a) E-mail: nwagatsuma@is.sci.toho-u.ac.jp

b) E-mail: sakai@cs.tsukuba.ac.jp

DOI: 10.1587/transinf.2019EDP7020

variety of fields [18], [19]. Eye movements and microsaccades reflect the strategy of visual processing [20]. The determination of BO is essentially local processing while that of FG could be global. The perception of local BO assignment often contradicts that of the global FG segregation, depending on instruction, window size, or direction of gaze [21], [22]. Other psychophysical studies have reported that FG segregation can occur without focal attention near the point of fixation [23], and that human development shows differences in detecting local and global configurations [24]. If the determination of BO is local processing while that of FG is global, the strategy of eye movement could be different. Because few studies have reported the eye movement in BO and FG determination, the analyses of the characteristics of the eye movement are expected to provide insightful evidence for the investigation of the mechanisms underlying FG segregation.

We investigated the perceptual responses and the spatiotemporal characteristics of eye movements for local BO discrimination and global FG segregation with filled natural-image patches. The patches were systematically selected to include a wide variety in contour shapes available in natural scenes [25]. First, we examined whether the perceptual DOF between the BO and FG tasks differ, and whether their difficulties differ. Specifically, we presented participants the patches of natural images in which figural regions were filled with either black or white. The participants judged either BO or FG in a trial and repeated the trials with different patches. The participants showed consistent judgement and similar difficulties between BO and FG determination for the most of stimuli, indicating similar perception between BO discrimination and FG segregation. Next, we examined the difference between the determination of BO and FG in eye movement including the distance of gaze and duration of fixation. Our result showed that gaze moves to farther locations in the FG segregation task compared to that in the BO discrimination task. This result appears natural because FG segregation requires the inspection of farther locations than local BO discrimination. The measured duration of the first fixation showed no significant difference between the BO and FG determination, but significantly longer duration while the fixation stayed on a figural region rather than a ground region. These results indicate the consistency in the perception between BO discrimination and FG segregation, but suggest the additional pooling strategies in saccades for FG segregation. Our analyses based on the perceptual responses and eye movements to a variety of the natural image patches would provide useful information in the investigation of the mechanisms underlying global FG segregation from local BO judgement.

## 2. Method

### 2.1 Participants

We investigated the perception and the spatiotemporal characteristics of eye movements in the FG segregation and BO

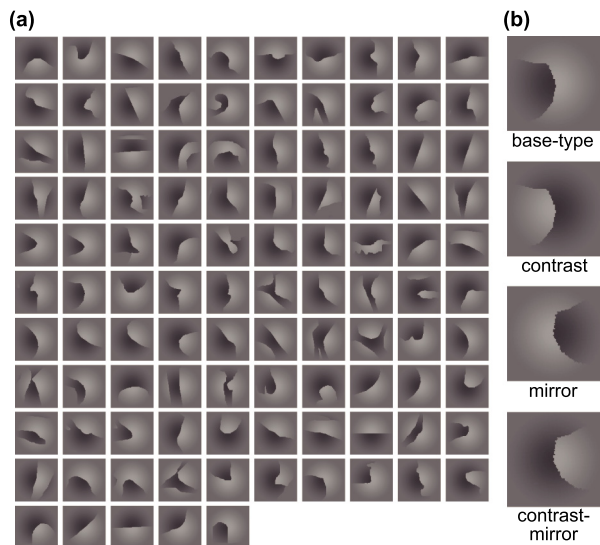
discrimination. Although the number of participant often vary across the visual psychophysical experiments that monitor eye movements, four to 35 observers participated in the previous reports that included the procedure similar to our experiments [18], [26]. In the present experiments, seven participants, five male and two female, in their twenties of age (mean age = 23, SD = 2) volunteered for the experiments. All participants had 20/20 visual acuity or better after correction. They were familiar with visual psychophysics but not aware of the purpose of the experiments. The present experiments were approved by the Research Ethics Committee. Prior to the experiment, all participants gave written informed consent approved by the Research Ethics Committee.

### 2.2 Apparatus

The stimuli were presented on a 24.1" LC monitor (ColorEdge CG242W; EIZO Corporation) that was placed in a dark room and refreshed at 60 Hz. The monitor was placed at a distance of 60 cm in front of the participants. Eye movements were monitored by the eyetracker machine (Tobii Eye Tracker X60; Tobii Technology AB) at a sampling rate of 60 Hz, controlled by MATLAB through Psychtoolbox [27], [28]. Given the 60 Hz sampling rate, our measurement reflects large shifts of eye-fixation. We obtained 51.4 points of eye position per second (SD = 13.1) via this machine. In order to compute and analyze characteristics of eye movements more precisely, we excluded the trials from the analysis that failed to provide enough points of eye position during the presentation of a test stimulus ( $n > 10$ ).

### 2.3 Stimuli

We presented black-and-white stimuli that were generated from the patches of natural images so that a variety of figural shapes were included. The generation of the stimuli was identical to Sakai et al. [25] except for filling regions. The patches of  $69 \times 69$  pixel were selected from a set of human-marked contours of natural scenes (Berkeley Figure/Ground Dataset, BFGD) [29] so as to assure a wide distribution of the contour features such as convexity, closure, and symmetry. Because of the wide distribution, the set of stimuli has a diversity in the contour features and their combinations. If we had chosen randomly patches from natural images, a number of the patches were similar to others according to the natural, non-uniform distribution of the features. A total of 105 patches were selected with the center of a patch always placed on the contour. We filled the one side of the contour with black and the other side with white. The luminance of the black and white were 0.54 and 82.6  $\text{cd/m}^2$ , respectively. We measured the luminance by using the luminance meter (LS-100, Minolta) at a distance of 60 cm from the display monitor. To smooth patch boundaries, we multiplied the patches with a Gaussian ( $\sigma = 17$  pixel). The generated stimuli are shown in Fig. 1 (a). To

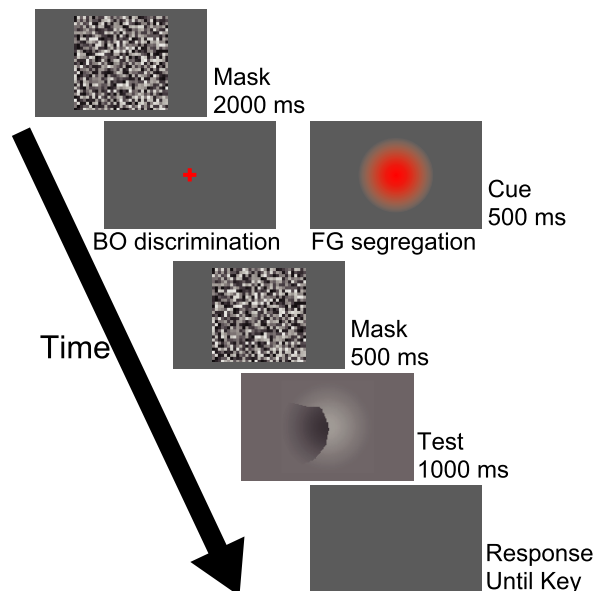


**Fig. 1** Black-and-white stimuli generated from the patches of natural images (BFGD [29]). (a) Base types of black-and-white stimuli. A total of 105 patches were selected for our experiment. The generation of these stimuli was identical to Sakai et al. [25] except for filling regions. (b) The variations of the stimuli. The stimulus with the opposite contrast (contrast), the mirror image with respect to the vertical midline (mirror) and that with both (contrast-mirror). These variations yield a total of 420 stimuli.

cancel out the effects of contrast and orientation, we included the patches with the opposite luminance and the mirror images with respect to the vertical midline. A total of 420 types of stimuli were presented in the experiments (105 base types  $\times$  2 contrast polarities  $\times$  2 mirror images; Fig. 1 (b)). The stimulus patches were systematically selected so that their figure-ground boundary (contour) included a wide variety of shapes available in natural scenes [25]. This wide variety of visual stimuli allowed us to investigate general characteristics of perception and eye movements for BO discrimination and FG segregation independent of boundary and object shapes.

## 2.4 Procedure

In the present study, BO discrimination is defined as the determination of the DOF along a local contour at the screen center, the task of which may need no more than the inspection of a local area around the stimulus center. In contrast, FG segregation is defined as the perception of a figure, the task of which may need the inspection of the entire field of the stimulus. Figure 2 shows the procedure of the psychophysical experiments. The experiment started with the presentation of a mask at the center of the screen for 2000 ms. After the disappearance of the mask, in order to instruct the type of task, a cue stimulus was presented at the center of the screen for 500 ms. The BO discrimination and FG segregation tasks were instructed by the presentation of a red cross ( $2^\circ \times 2^\circ$ : left panel of the Cue on Fig. 2) and a red Gaussian circle ( $9^\circ \times 9^\circ$ : right panel of the Cue on Fig. 2), respectively. Note that the participants were instructed to freely view the stimuli. We used a larger cue ( $9^\circ \times 9^\circ$ ) for



**Fig. 2** Procedure for the psychophysical experiment. In the BO discrimination task, which was cued by the presentation of a red cross, the participants were instructed to judge the direction of figure along the contour of the filled natural-image stimulus at the screen center. In the FG segregation task, which was cued by the presentation of a red Gaussian, the participants were asked to report which of the black or white region was perceived in front of the other. Participants were instructed to respond to these tasks at the end of each trial (during the presentation of a blank screen). See text for details.

the FG task because some participants reported that a small cue ( $2^\circ \times 2^\circ$ ) was somewhat difficult for distinguishing tasks. Statistical analyses did not show the significant dependence of perceptual responses and eye movements on the dimension of the cue ( $9^\circ \times 9^\circ$  vs.  $2^\circ \times 2^\circ$ ). Subsequently, another mask was presented for 500 ms, and a stimulus ( $9^\circ \times 9^\circ$ ; Fig. 1) was presented at the center for 1000 ms (Test). At the end of each trial, a blank screen was presented until the detection of the participants' perceptual responses. All combinations (840) of the task (2), stimulus shape (105), and stimulus type (4) were tested with three repeats in a randomized order. This procedure was controlled by MATLAB through Psychtoolbox [27], [28].

## 2.5 Perceptual Responses

We recorded the perceptual responses during the presentation of a blank screen at the end of each trial (Fig. 2). In the BO task, the participants were asked to judge the local DOF, the right or left, along the contour placed at the center of the screen by pressing the arrow keys on the keyboard. In the FG task, the participants were asked to judge which region, the black or white, appeared in front of the other.

We analyzed the perception of DOF observed in the BO discrimination task and the FG segregation task. Note that DOF and a *figural region*, which are labeled by left-and-right and black-and-white, respectively, are used interchangeably in the following analyses. For instance, because the participants judged a figural region rather than DOF in

the FG task, we transformed the chosen region into the DOF at the center of stimulus.

First, to indicate which part in a stimulus is perceived as figure, and how consistent is the perception, we defined BO discrimination Index (BOI) and FG segregation Index (FGI) by combining the four stimulus variations, seven participants and three repeats. BOI was defined as

$$BOI(i) = \frac{B_{BO}(i) - W_{BO}(i)}{N_{BO}^{Par} \cdot N_{BO}^{Type} \cdot N_{BO}^{Trial}} \quad (1)$$

where  $i$  indicates stimulus shape ( $1 \leq i \leq 105$ ),  $B_{BO}$  and  $W_{BO}$  indicate the number of trials in which the perceived DOF was in the direction of black and white sides, respectively.  $N^{Par}$ ,  $N^{Type}$  and  $N^{Trial}$  represent the number of participants (7), stimulus variations (4) and repeats (3), respectively. The multiplication of the three results in the total number of trials. Note that for those stimuli with the contrast reversed (e.g., the second and fourth panels from the top (“contrast” and “contrast-mirror”) in Fig. 1 (b)), the original contrast was registered for counting  $B_{BO}$  and  $W_{BO}$ . FGI was defined similarly as

$$FGI(i) = \frac{B_{FG}(i) - W_{FG}(i)}{N_{FG}^{Par} \cdot N_{FG}^{Type} \cdot N_{FG}^{Trial}} \quad (2)$$

where  $B_{FG}$  and  $W_{FG}$  indicate the number of trials in which the perceived figural region was black and white regions, respectively. In these indices, the denominator represented the total number of trials for the stimulus shape,  $i$ . Therefore, BOI and FGI have a value ranging between  $-1.0$  and  $1.0$ . Positive BOI and FGI indicate that the black region (in the original patches) was perceived as figure more frequently than the white region, while negative values indicate that the white region tend to be perceived as figure. A unity of the indices indicates that the perception (the black region is figure) was perfectly consistent over the participants, variations in stimulus, and repeats, while zero indicates complete inconsistency, and a negative unity indicates the consistent perception of the other (white) region as figure.

Second, we defined the consistencies for BO discrimination (BOC) and FG segregation (FGC) across the participants, stimulus variation, and repeats. The sign of BOI and FGI indicates which region appeared figure (positive and negative for black and white regions, respectively). To cancel out the tendency that a particular region (black or white) is perceived as figure and clarify the consistency in the perception across the participants, stimulus variation, and repeats, we introduced indices that were absolute values of BOI and FGI. We defined the consistencies for BO discrimination (BOC) and FG segregation (FGC) as

$$BOC(i) = |BOI(i)| \quad (3)$$

$$FGC(i) = |FGI(i)| \quad (4)$$

where  $|\cdot|$  denotes absolute value. A unity of the indices indicates that the perception for a stimulus shape,  $i$ , was perfectly consistent over the participants, variations in stimulus and repeats, while zero indicates complete inconsistency.

## 2.6 Eye Movements

We recorded eye movements during the presentation of test stimuli. Prior to each session that takes about ten minutes, the eyetracker machine was calibrated to assure the accurate recording of eye positions. We reran the calibration until the machine succeeded to detect all 9 points on a grid within the spatial extent of stimulus presentation, unless the participant gave up the calibration. The overall accuracy (SD) of the gaze location was  $0.74^\circ$ .

We analyzed spatiotemporal characteristics of eye movements in the BO discrimination task and the FG segregation task. In this work, the distance in saccade and the location and duration of the first fixation were used as the indices for eye movements.

A comparison of distance in saccade is expected to provide meaningful evidence on the differentiation of processing between BO discrimination and FG segregation. To compare the distances of gaze transition in BO discrimination and FG segregation tasks, we defined the mean distances in the horizontal ( $d_x$ ) and vertical ( $d_y$ ) directions:

$$d_x = \frac{1}{N^{Trial}} \sum_{t=1}^{N^{Trial}} \left| \frac{\sum_{j=1}^{n_j} x_t^j}{n_t} - c_t^x \right| \quad (5)$$

$$d_y = \frac{1}{N^{Trial}} \sum_{t=1}^{N^{Trial}} \left| \frac{\sum_{j=1}^{n_j} y_t^j}{n_t} - c_t^y \right| \quad (6)$$

where  $x_t^j$  and  $y_t^j$  represent the  $j$ -th spatial locations of gaze during the trial  $t$  in the horizontal and vertical directions, respectively, and  $c_t^x$  and  $c_t^y$  represent the centers of gravity of the gaze locations during the trial.  $n_t$  indicates the total number of eye fixations for the trial  $t$ , and  $N^{Trial}$  represents the total number of trials.

In order to investigate quantitatively the spatiotemporal characteristics of eye movements, we measured the location (figure or ground) and duration of the first fixation after the stimulus onset. We defined saccade according to Engbert and Kligl [30] in which eye movements over a greater velocity (deg/sec) were considered as saccade while successive small movements were considered as fixation. The parameter of velocity threshold for the detection of saccades in Engbert and Kligl [30] was set to 6 (unitless).

For analyzing the characteristics of eye movements in more detail, we computed the location and duration of the first fixation during the BO and FG tasks. The analysis of the location and duration of the first fixation provides crucial information for understanding a variety of visual processing. For example, the location and duration of the first fixation after a stimulus onset have been reported to represent the characteristics of selective attention [26], [31], [32].



The location of the first fixation has been reported to agree with the region that pops out from the ground [17], [33]–[35]. The analysis of the location and duration of the first fixation, whether it was located in a figural or ground region and how long it stayed, would provide important evidence in the investigation of BO and FG determination.

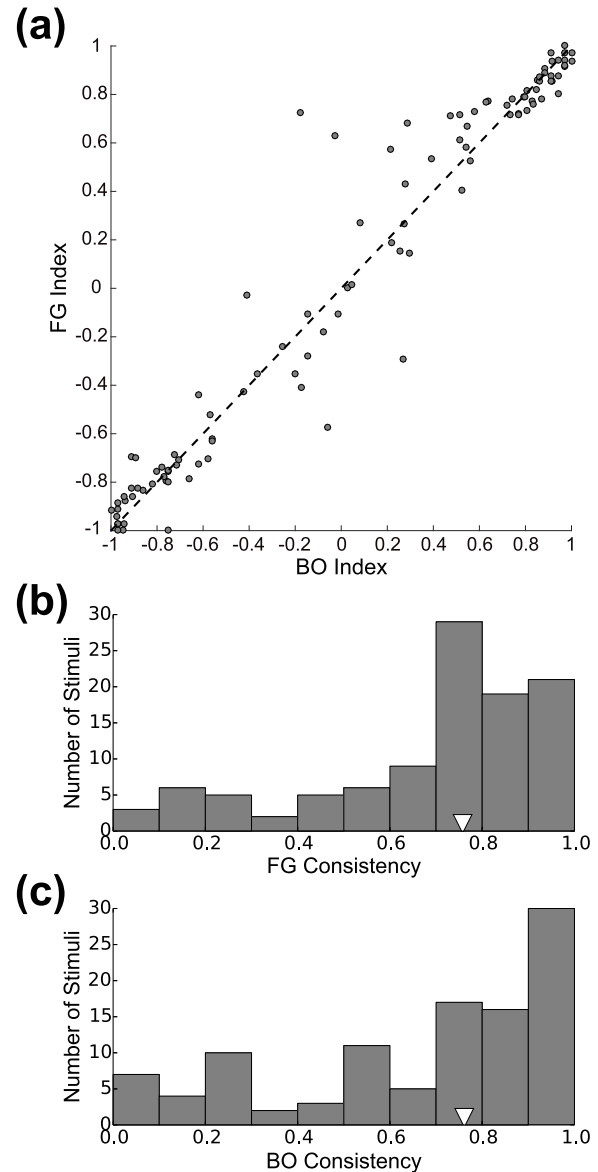
### 3. Results

We carried out the psychophysical experiments to compare the perceptual responses and the spatiotemporal characteristics of eye movements between BO discrimination and FG segregation. To test a variety of figural shapes, we generated a total of 420 stimuli from natural images in which figural regions are filled with black and white. The participants judged either BO or FG in a trial followed by the next trial with an either task (FG or BO) and one of the other stimuli. In the analysis of the perceptual responses, we compared the perceptual DOFs between the BO and FG tasks. In the analysis of eye movements, we examined the differences between the two tasks in the distance of gaze and duration of fixation. See also “Eye movements” in Method section for more detail.

#### 3.1 Similarity in the Perceptual Responses between the BO and FG Tasks

We investigated the perception of DOF observed in the BO discrimination task and the FG segregation task. We tested whether the figural regions perceived in the BO and FG tasks agree with each other. In this section, we compared the perceptual DOFs between the BO and FG tasks for each stimulus. To analyze quantitatively the perceptual DOFs between these tasks, we used BO discrimination Index (BOI) and FG segregation Index (FGI) (see Eqs. (1) and (2) in “Perceptual responses” in Method section). The distribution of the measured BOI and FGI with respect to each stimulus shape are shown in Fig. 3(a). A dot in the Fig. 3(a) represents the magnitude of the indices for a particular stimulus shape. A number of stimulus shapes showed large magnitudes close to one, and a fraction showed small magnitudes close to zero for both BOI and FGI, indicating confident perception of BO and FG. A number of dots are distributed along the diagonal (Pearson’s correlation coefficient = 0.88). To assure whether there is no significant difference between the BOI and FGI, we carried out a two-way ANOVA with factors of the task (2 levels; BO discrimination/FG segregation) and participant (7 levels) based on a factorial design including the interaction. The analysis showed no significance in the main factors ( $p = 0.171$  for task;  $p = 0.335$  for participant) and the interaction ( $p = 0.718$ ), indicating no statistical difference between the BOI and FGI. These results indicate a good agreement between BO and FG determinations.

Next, we examined the consistency for BO discrimination (BOC) and that for FG segregation (FGC) (see also Eqs. (3) and (4) in “Perceptual responses” in Method section). Specifically, we compared BOC and FGC, which



**Fig. 3** Agreement in the perceptual DOF between local BO discrimination and global FG segregation. (a) The distribution of the measured BOI and FGI. A dot represents the magnitude of the indices for a particular stimulus shape. A dashed line represents the diagonal. (b, c) The distributions of the consistency for FG segregation (FGC; (b)) and BO discrimination (BOC; (c)). Triangles represent the median. The distributions for the indices of the consistency were similar between FG and BO.

clarifies whether the difficulties of the judgements differ between the two tasks. The distributions of the indices are shown in Fig. 3 (b) and 3 (c). The 80% and 78% of stimulus shapes indicated  $FGC > 0.5$  and  $BOC > 0.5$ , respectively. Interestingly, these perceptual consistencies for BO and FG were similar to that predicted by the model based on local figure-ground cues [21]. We observe similar distributions in BOC and FGC. The medians for the indices were almost identical; 0.758 and 0.761 for FGC and BOC, respectively. There was no significant difference in the peak location between FGC and BOC (Mann-Whitney test,  $p = 0.488$ ).

These results indicate that BO discrimination and FG segregation show good agreements in the perceptual DOF and their consistency for a variety of shape in natural images.

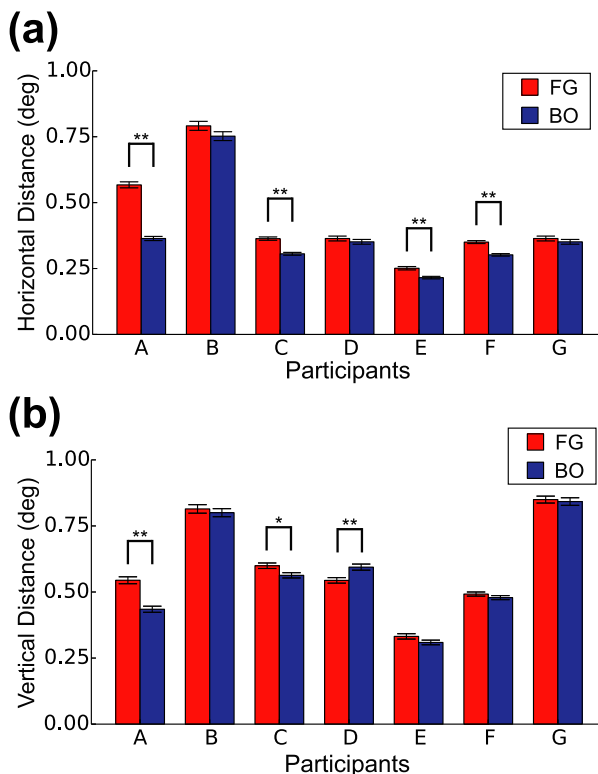
### 3.2 Longer Distances in Gaze Movements during FG Segregation

Characteristics of saccades have been frequently analyzed for exploring the mechanism of visual processing [36]–[38]. A comparison of the distance in saccade provides evidence whether the strategies for saccade are different between the BO discrimination and FG segregation. We computed the distance of gaze transition in the horizontal and vertical directions (see also Eqs. (5) and (6) in “Eye movements” in Method section).

The mean distances in the horizontal direction ( $d_x$ ) over trials for each of the seven participants are shown in Fig. 4 (a). To examine statistically the difference between the FG and BO tasks ( $d_x$ ), we performed a two-way ANOVA with factors of the task (2 levels) and participant (7 levels). The result showed significance in the main factors and the interaction ( $p < 0.001$ ). This result suggests the difference in the horizontal distance of gaze movement between the

FG and BO tasks. The significant difference in  $d_x$  among participants led us to examine the simple main effects of the task for each participant (pair-wise t-test). Four participants (A, C, E and F) showed the significantly greater  $d_x$  in the FG task than that in the BO task ( $p < 0.001$ ), and the other three participants did not exhibit significant difference ( $p = 0.109$  for B;  $p = 0.334$  for D;  $p = 0.300$  for G). Interestingly, all participants consistently showed longer mean  $d_x$  in the FG task compared to that in the BO task while some participants did not indicate significant difference in  $d_x$  between two tasks. The fact that FG segregation needs a longer movement compared to BO discrimination might be natural because the gaze might move to farther locations during FG segregation for the inspection of more global structure than BO discrimination. Although not all participants (four out of seven) reached the statistical significance, this result suggests that the strategies in visual processing appear to be different between FG segregation and BO discrimination.

Figure 4 (b) shows the mean distances in the vertical direction ( $d_y$ ) for the seven participants. We observe a longer distance for the FG task in six participants. A two-way ANOVA with factors of the task and participant showed significance in the main factors and the interaction ( $p < 0.001$ ). Further analyses with a factor of the task for each participant (pair-wise t-test) showed significance in three participants in which two (A and C) showed longer distances in the FG task ( $p < 0.001$  for A;  $p = 0.013$  for B) while the other showed that in the BO task ( $p < 0.001$  for D). These results suggest that eye movements in the vertical direction is not markedly different in performing the FG segregation and BO discrimination tasks. This result may be natural because the participants were instructed to judge BO in the horizontal direction.

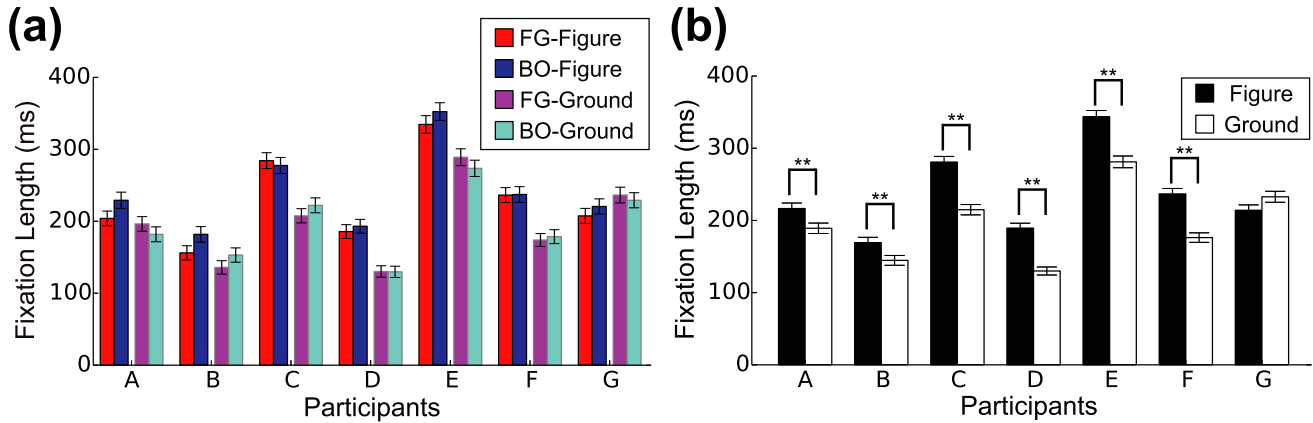


**Fig. 4** The distances of the eye movements during the BO discrimination and FG segregation tasks. (a) The mean distances among stimuli in the horizontal direction ( $d_x$ ) for seven participants. Blue and red bars represent  $d_x$  for the BO and FG tasks, respectively. Error bars are the standard error of all trials. Asterisks indicate significant differences between the two tasks (\*\*  $p < 0.01$ ; \*  $p < 0.05$ ). (b) The mean distances in the vertical direction ( $d_y$ ) for the two tasks. Conventions are the same as those in (a). The mean distance of the eye movements was greater for the FG task in the horizontal direction.

### 3.3 Longer Durations on Figural Regions in Early Fixations

For investigating the spatiotemporal characteristics of eye movements in more detail, we computed the location and duration of the first fixation for the seven participants during the BO and FG tasks. Figure 5 (a) summarizes the results in which the durations of the first fixation are categorized by its location (figure and ground) and the task.

First, we analyzed the duration of the fixations to *figure* regions (red and blue bars in Fig. 5 (a)). To examine this tendency, we carried out a two-way ANOVA with factors of the task (2 levels) and participant (7 levels) for the duration to figure regions. We found significance in the main factors and the interaction ( $p < 0.05$ ). Further analyses with a factor of the task for each participant showed no significance for all participants ( $p = 0.053$  for A;  $p = 0.072$  for B;  $p = 0.644$  for C;  $p = 0.573$  for D;  $p = 0.277$  for E;  $p = 0.954$  for F;  $p = 0.307$  for G), suggesting no difference between FG segregation and BO discrimination in the duration of the fixation to figural regions. However, almost (six out of seven) participants show longer duration of the fixations to figure regions in the BO task than that in the FG task.



**Fig. 5** The duration and location of the first fixations during the BO discrimination and FG segregation tasks. (a) Blue and red bars indicate the mean duration of the first fixation to the figural region in the BO and FG tasks, respectively. Cyan and magenta represent those to the ground in the BO and FG tasks, respectively. (b) The merged duration of the first fixation to the figure and ground regions for seven participants. Black and white bars indicate the durations of the first fixation to the figural and ground regions, respectively. Asterisks indicate significant differences between the two regions ( $t$ -test: \*\*  $p < 0.01$ ; \*  $p < 0.05$ ). The first fixation stayed longer on figural regions.

Second, to quantify the duration of the fixation to *ground* regions (magenta and cyan bars in Fig. 5 (a)), we performed similarly a two-way ANOVA with factors of the task and participant. We found the significance in the main factor of participants ( $p < 0.001$ ) but not in the task or interaction. Further analyses for each participant showed no significance in the task for all participants ( $p = 0.249$  for A;  $p = 0.182$  for B;  $p = 0.291$  for C;  $p = 0.964$  for D;  $p = 0.321$  for E;  $p = 0.707$  for F;  $p = 0.584$  for G), suggesting no difference between FG segregation and BO discrimination in the duration of the fixation to ground regions.

Finally, in order to explore whether the first fixation stays longer on a figural region in the two tasks, we performed a three-way ANOVA with factors of BO discrimination/FG segregation task (2 levels), participant (7 levels) and figure/ground location (2 levels) for the duration of first fixation. The result showed significance in the location, participant ( $p < 0.001$ ) and interaction between the location and participant. Since the participant was significant, we carried out for each participant a two-way ANOVA with factors of the task and location for the duration of first fixation. Six out of seven participants showed significantly longer durations on figure regions than that on ground ( $p < 0.05$  for A, B, C, D, E, F;  $p = 0.080$  for G), and no significance for the main factor of the task and the interaction. For clarifying the significance, the duration of first fixation for two tasks were merged with respect to figure and ground location for each participant. Figure 5 (b) summarizes the merged duration of the fixation to figure and ground regions for each participant. We observe longer durations of the fixation on figural regions. This result indicates that the first fixation stays longer in a figural region independent of the task. It has been reported that the first fixation often stays on a region that pops out based on bottom-up processing [17], [34], [35]. Although the underlying

cortical mechanisms have not been clarified, our result suggests that early-level bottom-up processing such as pop-out plays an important role in both BO discrimination and FG segregation.

#### 4. Discussion

To examine the perceptual responses and the spatiotemporal characteristics of eye fixations in local BO discrimination and global FG segregation, we carried out psychophysical experiments with black-and-white filled stimuli generated from natural-image patches. BO discrimination was defined as the determination of DOF along a local contour. This task needs no more than the inspection of a local area around the stimulus center. In contrast, FG segregation was defined as the global perception of a figure, the task of which needs the inspection of the entire field of the stimulus. The participants showed consistency in the perception between the BO and FG judgements for the most of the stimuli. We analyzed the spatiotemporal characteristics of eye movements to investigate the link between local BO discrimination and global FG segregation. The gaze moved to farther locations under the FG segregation task compared to that under the BO discrimination task. The duration of the first fixation did not show significant difference between the BO and FG determination but the duration of the fixation to the figural region was significantly longer than that to the ground region. These results indicate the similarities in the perception and the most of tendencies of saccade between the tasks. Only the difference we found was the longer distance in gaze movement during FG segregation. These results suggested the distinct neural mechanism for perception of local BO determination and global FG segregation.

Our analyses of the spatiotemporal characteristics of eye movements during the BO discrimination and the FG

segregation suggest the involvement of distinct mechanisms for the determination of BO and FG. A possible mechanism might be a sequential processing from the BO determination to FG segregation: local BO is computed mainly from the examination of a figural region, and the computed BOs are gathered over a longer distance for the determination of global FG. Interestingly, a recent study indicated the critical roles of neurons in V4 for the FG segregation [39]. This evidence led us to consider a mechanism that the afferent transmission from V2 gradually establish the neural responses in V4 for the perception of FG segregation. The neural responses in V4 for the FG segregation might be determined based on the integration of the responses of BO-selective cells in V2.

A variety of computational models have been proposed for the investigation of the neural mechanisms underlying the determination of BO and FG. Interestingly, some computational models implied that the responses of BO-selective neurons in V2 underlay the responses of V4 for figure regions. Craft et al. [14] and Mihalas et al. [40] hypothesized grouping neurons in V4. In their models, the grouping neurons not only modulated the responses of neurons with BO selectivity in V2 via feed-back projections but also coded the rough sketch of object regions by integrating the local DOF given by BO-selective neurons. These models suggest that the responses of BO-selective neurons in V2 play important roles for the determination of global FG segregation. These previous modeling studies appeared to support the hypothesis of the sequential processing from the BO determination to the FG segregation.

We found the significance in the interaction between figure/ground location and participant for the duration of first fixation. Previous study based on spatiotemporal analysis of eye fixations and computational model implied that people with autism spectrum disorder (ASD) tended to have atypical preferences for particular objects in natural images [41]. This report and our results imply the possibility that the duration of fixation might be determined depending on the characteristics of neuronal network for each participant.

The investigations of the interaction between BO and FG appear to be important in order to further understand neural mechanisms underlying the perception of BO and FG. A possible method for investigating the interaction is the utilization of stimuli in which local BO and global FG are contradictory. The perception of local BO assignment sometimes contradicts to that of global FG segregation, depending on a variety of factors [21], [22]. For the present experiment, we used the filled natural-image patches from BFGD as the stimuli [29] in which BO and FG were consistent by nature. Experiments with the contradictory stimuli would provide the characteristics of the interaction between BO and FG. For instance, if the processing of BO and FG is sequential and the BO determination takes place prior to that of FG segregation, saccades during FG determination may be modulated by the contradiction while saccades during BO determination may not be. It is expected that

distances in saccades and gaze movements for the contradictory stimuli might markedly reflect the interaction between local BO and global FG. Investigations on the interaction between BO and FG in the contradictory stimuli would provide useful evidence for understanding the mechanisms of FG segregation.

## Acknowledgments

This work was supported by Grants-in-aid from JSPS (KAKENHI 26280047 and 17H01754), Research Institute of Electrical Communication, Tohoku University (RIEC) (H28/A13), and the Mitsubishi Foundation (30317).

## References

- [1] N. Rubin, "Figure and ground in the brain," *Nat. Neurosci.*, vol.4, pp.857–858, 2001.
- [2] V.A. Lamme, "The neurophysiology of figure-ground segregation in primary visual cortex," *J. Neurosci.*, vol.15, no.2, pp.1605–1615, 1995.
- [3] V.A. Lamme, K. Zipser, and H. Spekreijse, "Masking interrupts figure-ground signals in V1," *J. Cognit. Neurosci.*, vol.14, no.7, pp.1044–1053, 2002.
- [4] C.F. Altmann, H.H. Bühlhoff, and Z. Kourtzi, "Perceptual organization of local elements into global shapes in the human visual cortex," *Curr. Biol.*, vol.13, no.4, pp.342–349, 2003.
- [5] C.W. Tyler, H.A. Baseler, and L.L. Kontsevich, "Predominantly extra-retinotopic cortical response to pattern symmetry," *Neuroimage*, vol.24, pp.306–314, 2005.
- [6] A. Gilad, Y. Pesoa, I. Ayzenshtat, and H. Slovov, "Figure-ground processing during fixation saccades in V1: indication for high-order stability," *J. Neurosci.*, vol.34, no.9, pp.3247–3252, 2014.
- [7] H. Zhou, H.H. Friedman, and R. von der Heydt, "Coding of border ownership in monkey visual cortex," *J. Neurosci.*, vol.20, no.17, pp.6594–6611, 2000.
- [8] F. Qiu, T. Sugihara, and R. von der Heydt, "Figure-ground mechanisms provide structure for selective attention," *Nat. Neurosci.*, vol.10, pp.1492–1499, 2007.
- [9] N. Zhang and R. von der Heydt, "Analysis of the context integration mechanisms underlying figure-ground organization in the visual cortex," *J. Neurosci.*, vol.30, no.19, pp.6482–6496, 2010.
- [10] J.R. Williford and R. von der Heydt, "Figure-ground organization in visual cortex for natural scenes," *eNeuro*, vol.3, no.6, 2016. DOI:10.1523/ENEURO.0127-16.2016.
- [11] E. Carlson, R. Rasquinha, K. Zhang, and C.E. Connor, "A sparse object coding scheme in area V4," *Curr. Biol.*, vol.21, no.4, pp.288–293, 2011.
- [12] K. Sakai and H. Nishimura, "Surrounding suppression and facilitation in the determination of border ownership," *J. Cognit. Neurosci.*, vol.18, pp.562–572, 2006.
- [13] P.R. Roelfsema, L.A. Lamme, H. Spekreijse, and H. Bosch, "Figure-ground segregation in a recurrent network architecture," *J. Cognit. Neurosci.*, vol.14, pp.1044–1053, 2002.
- [14] E. Craft, H. Schütze, E. Niebur, and R. von der Heydt, "A neural model of figure-ground organization," *J. Neurophysiol.*, vol.97, no.6, pp.4310–4326, 2007.
- [15] Z. Kourtzi and C.E. Connor, "Neural representation for object perception: structure, category, and adaptive coding," *Ann. Rev. Neurosci.*, vol.34, pp.45–67, 2011.
- [16] K. Sakai, H. Nishimura, R. Shimizu, and K. Kondo, "Consistent and robust determination of border ownership based on asymmetric surrounding contrast," *Neural Netw.*, vol.33, pp.257–274, 2012.
- [17] A.F. Russell, S. Mihalas, R. von der Heydt, E. Niebur, and R.



- Etienne-Cummings, "A model of proto-object based saliency," *Vis. Res.*, vol.94, pp.1–15, 2014.
- [18] Y. Chujo, Y. Jono, K. Tani, Y. Nomura, and K. Hiraoka, "Corticospinal excitability in the hand muscles is decreased during eye movement with visual occlusion," *Percept. Mot. Skills*, vol.122, no.1, pp.238–255, 2016.
- [19] E. Zimmermann, M.C. Morrone, and P. Binda, "Perception during double-step saccades," *Sci. Rep.*, vol.8, pp.1–9, 2018. DOI:10.1038/s41598-017-18554-w.
- [20] L. Itti and C. Koch, "Computational modelling of visual attention," *Nat. Rev. Neurosci.*, vol.2, pp.194–203, 2001.
- [21] C.C. Fowlkes, D.R. Martin, and J. Malik, "Local figure-ground cues are valid for natural images," *J. Vis.*, vol.7, no.8, pp.1–9. DOI: 10.1167/7.8.2.
- [22] S.-H. Kim and J. Feldman, "Globally inconsistent figure/ground relations induced by a negative part," *J. Vis.*, vol.9, no.10, pp.1–13, 2009. DOI: 10.1167/9.10.8.
- [23] R. Kimchi and M.A. Peterson, "Figure-ground segmentation can occur without attention," *Psychol. Sci.*, vol.19, no.7, pp.660–668, 2008.
- [24] R. Kimchi, B. Hadad, M. Behrmann, and S.E. Palmer, "Microgenesis and ontogenesis of perceptual organization: evidence from global and local processing of hierarchical patterns," *Psychol. Sci.*, vol.16, no.4, pp.282–290, 2005.
- [25] K. Sakai, S. Matsuoka, K. Kurematsu, and Y. Hatori, "Perceptual representation and effectiveness of local figure-ground cues in natural contours," *Front. Psychol.*, vol.6, 2015. DOI: 10.3389/fpsyg.2015.01685.
- [26] H. Sheridan and E.M. Reingold, "Chess player's eye movements reveal rapid recognition of complex visual patterns: Evidence from a chess-related visual search task," *J. Vis.*, vol.17, no.3, pp.1–12, 2017. DOI: 10.1167/17.3.4.
- [27] D.H. Brainard, "The psychophysics toolbox," *Spat. Vis.*, vol.10, no.4, pp.433–436, 1997: 10.
- [28] D.G. Pelli, "The VideoToolbox software for visual psychophysics: Transforming numbers into movies," *Spat. Vis.*, vol.10, no.4, pp.437–442, 1997.
- [29] D. Martin, C.C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *Proc. IEEE ICCV*, vol.2, pp.416–423, 2001.
- [30] R. Engbert and R. Kliegl, "Microsaccades uncover the orientation of covert attention," *Vis. Res.*, vol.43, no.9, pp.1035–1045, 2003.
- [31] K. Rayner, "Eye movements in reading and information processing: 20 years of research," *Psychol. Bull.*, vol.124, no.3, pp.372–422, 1998.
- [32] J.M. Henderson, "Eye movement control during visual object processing: Effects of initial fixation position and semantic constraint," *Can. J. Exp. Psychol.*, vol.47, no.1, pp.79–98, 1993.
- [33] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vis. Res.*, vol.40, no.10–12, pp.1489–1506, 2000.
- [34] S. Straube, C. Grimsen, and M. Fahle, "Electrophysiological correlates of figure-ground segregation directly reflect perceptual saliency," *Vis. Res.*, vol.50, no.5, pp.509–521, 2010.
- [35] B. Hu, R. Kane-Jackson, and E. Niebur, "A proto-object based saliency model in three-dimensional space," *Vis. Res.*, vol.119, pp.42–49, 2016.
- [36] B. Lia and J.F. Olavarria, "The distribution of corticotectal projection neurons correlates with the interblob compartment in macaque striate cortex," *Vis. Neurosci.*, vol.13, no.3, pp.461–466, 1996.
- [37] P.L. Abel, B.J. O'Brien, B. Lia, and J.F. Olavarria, "Distribution of neurons projecting to the superior colliculus correlates with thick cytochrome oxidase stripes in macaque visual area V2," *J. Comp. Neurol.*, vol.377, no.3, pp.313–323, 1997.
- [38] K. Inagaki, Y. Hirata, and S. Usui, "A model-based theory on the signal transformation for microsaccade generation," *Neural Netw.*, vol.24, no.9, pp.990–997, 2011.
- [39] H. Hasuike, Y. Yamane, H. Tamura, and K. Sakai, "Representation of local figure-ground by group of V4 cells," *Lecture Notes in Computer Science*, vol.9947, pp.131–137, 2016.
- [40] S. Mihalas, Y. Dong, R. von der Heydt, and E. Niebur, "Mechanisms of perceptual organization provide auto-zoom and auto-localization for attention to objects," *Proc. Natl. Acad. Sci. USA*, vol.108, pp.7583–7588, 2011.
- [41] S. Wang, M. Jiang, X.M. Duchesne, E.A. Laugeson, D.P. Kenndy, R. Adolphs, and Q. Zhao, "Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking," *Neuron*, vol.88, no.3, pp.604–616, 2015.



**Nobuhiko Wagatsuma** received the B.S., M.S. and Ph.D degrees from University of Tsukuba in 2004, 2006 and 2009, respectively. He is currently Lecturer at Faculty of Science, Department of Information Science, Toho University. His research interests include visual attention and computational neuroscience.

**Mika Urabe** received the B.S. and M.S. degrees from University of Tsukuba in 2014 and 2016, respectively.



**Ko Sakai** received B.S. and M.S. from University of Electro-communications, Tokyo, and M.S. and Ph.D from University of Pennsylvania, Philadelphia, PA, in 1983, 1985, 1991, and 1995, respectively. He is currently Professor at Dept. of Computer Science, University of Tsukuba. His research interests include vision science and computational neuroscience.