Spatial attention in early vision for the perception of border ownership

Nobuhiko Wagatsuma
Laboratory for Computational Vision Science, Department of Computer Science, University of Tsukuba, Tsukuba, Japan

Ryohei Shimizu
Laboratory for Computational Vision Science, Department of Computer Science, University of Tsukuba, Tsukuba, Japan

Ko Sakai
Laboratory for Computational Vision Science, Department of Computer Science, University of Tsukuba, Tsukuba, Japan

Spatial attention alters contrast gain in early visual areas, which might affect the determination of border ownership (BO) that indicates the direction of figure with respect to the border. We investigated the role of spatial attention applied to early vision in the determination of BO with a computational model that consists of V1, V2, and posterior parietal (PP) modules. Attention alters contrast gain in the V1 module so that it enhances local contrast. The V2 module determines BO based on the surrounding contrast extracted by the V1 module. The simulation results showed that the attention significantly modulates BO; BO is even flipped in figures with ambiguous BO while BO is stable for unambiguous figures such as a simple square. To evaluate the model quantitatively, we carried out psychophysical experiments to measure the effects of attention in the perception of BO and compared the results with those from corresponding simulations. The model showed good agreement with human perception including the determination of BO for ambiguous random-block stimuli. These results indicate that the activity of BO-selective neurons could be modulated significantly by spatial attention that alters local contrast gain in V1, which may account in part for automatic, bi-stable perception in ambiguous figures.

Keywords: attention, border ownership, early vision, computational model, surrounding modulation


Introduction

Visual attention is a function that boosts our perception (Posner, 1980), enabling us to attend the most important information at the moment (e.g., Deco & Lee, 2004). Visual attention functions in two distinct modes: spatial attention and object-based attention. Both types of attention have been shown to enhance our perception from a number of aspects, such as spatial frequency and orientation discrimination, dominancy in binocular rivalry, and the contextual modulation (Ito, Westheimer, & Gillbert, 1998; Lee, Itti, Koch, & Braun, 1999; Mitchell, Stoner, & Reynolds, 2004). Although a number of studies have discussed the crucial role of higher-order cortical areas for attention modulation, recent studies have reported robust effects of attention on the primary visual cortex (e.g., Posner & Gilbert, 1999). Physiological studies have shown that spatial attention modulates V1 neurons to bias their activities in favor of a stimulus at the attended location (e.g., Luck, Chelazzi, Hillyard, & Desimone, 1997), with significant modulation in the presence of competing or contextual stimuli (Ito et al., 1998; Motter, 1993). Functional MRI studies have also shown attention modulation of V1 (e.g., Somers, Dale, Seiffert, & Tootell, 1999).

A number of psychophysical studies have reported the increase of apparent luminance contrast by spatial cueing in various paradigms (Bashinski & Bacharach, 1980; Hawkins et al., 1990; Palmer, Ames & Lindsey, 1993; Solomon, Lavie, & Morgan, 1997). Recent psychophysical studies have clarified that spatial attention alters apparent stimulus contrast even with a simple task (Carrasco, Ling, & Read, 2004; Pestilli & Carrasco, 2005). They have asked subjects to answer the orientation of a small Gabor patch if its contrast appeared to be higher than the other patch. When spatial attention was drawn to the location of a Gabor patch, subjects perceived higher contrast in that patch even though the physical contrasts were the identical. This experiment with such a simple paradigm showed that spatial attention alters the appearance of contrast without the influence of higher-level features, memory, or competition among multiple objects, suggesting a crucial role of early vision in attention...
modulation. Recent fMRI studies have also reported the increase of V1 activities in similar paradigms (Buracas & Boynton, 2007; Smith, Cotillon-Williams, & Williams, 2006). Several modeling studies have suggested that a change in the strength of divisive normalization in V1 leads to the modulation of apparent contrasts (e.g., Deco & Lee, 2004; Lee et al., 1999; Peters, Iyer, Itti, & Koch, 2005; Reynolds & Chelazzi, 2004). The major characteristics of the models showed good agreements with those of the attention-based modulation of human perception including contrast modulation.

Attention even alters the perception of an object, as is apparent in ambiguous figures such as Rubin’s vase (Hasson, Hendler, Bashat, & Malach, 2001; Pitts, Nerger, & Davis, 2007). Although spatial attention alters contrast sensitivity, it has not been clarified how, through contrast gain, attention accounts for more complex perception including figure-ground segregation, one of the earliest and most fundamental processes for perceptual organization and shape/object recognition. A majority of neurons in monkeys’ V2 and V4 show selectivity to border ownership (BO), their responses depending on which side of a border owns the contour (Zhou, Friedman, & von der Heydt, 2000). Computational studies have suggested that the cortical mechanisms underlying BO coding involve the surrounding modulation observed in early visual areas (Nishimura & Sakai, 2005; Sakai & Nishimura, 2006; Sugihara, Tsuji, & Sakai, 2007). Other recent studies are also consistent with involvement of early visual areas (Fowlkes, Martin, & Malik, 2007; Roelfsema, Lamme, Spekreijse, & Bosch, 2002; Sajda & Baek, 2004). BO could be altered by attention modulation of luminance contrast outside the classical receptive field (CRF).

These recent physiological and computational studies led us to propose the hypothesis that spatial attention alters contrast gain, which then modifies the activities of BO-selective neurons; these in turn give rise to the alternation of figural objects in ambiguous images. Although a number of studies have reported significant effects of attention in V2 and V4 (e.g., Reynolds, Chelazzi, & Desimone, 1999; Reynolds, Pasternak, & Desimone, 2000) including the direct influence to BO-selective cells in V2 (Qiu, Sugihara, & von der Heydt, 2007), few studies have investigated the perceptual role of attention modulation in the determination of Direction of Figure (DOF; Sajda & Finkel, 1995). We focus on the role of attention in early vision, specifically V1, to investigate bottom-up modulation that biases the determination of BO. The bottom-up modulation through V1 seems to be crucial for BO determination because apparent contrast in V1 is modulated by attention, the latency of the BO signal is short (Zhou et al., 2000), and the switch of figure is observed even for meaningless figures (Vecera, Flevaris, & Filapek, 2004).

We investigated the role of spatial attention applied to V1 with a computational model consisting of V1, V2, and posterior parietal (PP) modules. The PP module is designed to represent spatial attention that could be considered as a saliency map based on luminance contrast. In the model, spatial attention from the PP module alters contrast gain in the V1 module. The change in contrast signal then modifies the activity of BO-selective neurons in V2 because BO is determined solely by surrounding contrast. We excluded the feedback connections from PP to V2 and V2 to V1, as similar to a lesion study to examine the sole effect of V1 in which attention modulates contrast. The simulation result showed that the DOF could be flipped in ambiguous stimuli, depending on the location of attention. Although the activities of model BO cells were modulated, their DOF did not alter for unambiguous stimuli. To evaluate the model quantitatively, we carried out psychophysical experiments to measure the effects of attention on the perception of BO for ambiguous random-block stimuli and then compared the results with those of corresponding simulations. The model showed good agreements with human perception for modulation magnitude and its variance among stimuli. When the stimulus included a familiar shape, human determination of BO shifted in the direction of the familiar shape, whereas our model did not exhibit such a shift. However, there was no significant difference in the modulation magnitude between the model and psychophysical data. This disagreement in BO shifts, and agreement in modulation magnitude seems natural if spatial attention could work on early vision without the influence of familiarity or feature-based attention. Specifically, the stimulus invariance of modulation magnitude appears to indicate the effect of contrast modulation in V1 because this mechanism works independent of stimulus shape and familiarity. These results indicate that the activity of BO-selective neurons could be modulated significantly by spatial attention that alters local contrast gain in early vision.

**The model**

In our model, spatial attention enhances contrast gain in the early visual area. As a result, the activity of BO-selective neurons that determines BO from local contrast is modulated. Our model consists of three modules, V1, V2, and the posterior parietal (PP) module, as illustrated in Figure 1. V1 and PP are mutually connected for the application of spatial attention to V1, and for the modification of spatial attention in PP. Attention is applied only to modulate contrast gain in V1. The lack of connections between PP and V2 excludes the direct modulation of BO-selective cells by attention. It has been reported that spatial attention influences BO-selective cells in Macaque V2 (Qiu et al., 2007). The exclusion of this connection in the model corresponds to the dissection of the connections between PP and V2, which enables us to examine the exclusive role of attention.
applied to V1. The connections from V2 to V1 are also excluded for the same reason. Although a physiological study has reported that feedback from V2 to V1 did not show considerable influence on figure-ground segregation (Hupe´, James, Girard, & Bullier, 2001), it has not been clarified what is the role of the feedback for the establishment of V1 properties including surround modulation. A recent computational study has suggested that the feedback is crucial for surround modulation (Wielaard & Sajda, 2006). On the other hand, recent physiological studies have suggested that afferent connection from the lateral geniculate nucleus (LGN) evokes surround modulation (Naito, Sadakane, Okamoto, & Sato, 2007). We excluded the feedback from V2 to V1, and given a hard-wired surround modulation, in order to focus on the contrast modulation that is caused by attention applied to V1.

Each module consists of 100 x 100 model neurons positioned retinotopically. The model is designed so that 25 pixel corresponds to one degree (°) in visual angle. Although we do not focus on the temporal characteristics of the model, we introduce the dynamics into the model since V1 and PP are mutually connected including feedback. Therefore, activities of the model cells are represented by a partial differential equation with time (t) and space (x and y) variables. In the following equations, where focus is placed on dynamic change rather than spatial interaction, we omit space variables (x and y) or represent them as if they are constants. In the absence of external input, the activity of a neuron at time t, A(t), is given by

\[ \tau \frac{\partial A(t)}{\partial t} = -A(t) + \mu F(A(t)), \]  

where the first term on the right side is a decay, and the second term takes into account the excitatory, recurrent signal among the excitatory neurons. The non-linear function, \( F(A(t)) \), is given by

\[ F(A(t)) = \frac{1}{T_r - \tau \log(1 - (1/A(t)))}, \]

where \( \tau \) is a membrane time-constant (10.0 ms), and \( T_r \) is the absolute refractory time (0.5 ms). The dynamics of this equation as well as appropriate values for constants have been widely studied (e.g., Gerstner, 2000). Although the equation reproduces the temporal characteristics of neurons in vivo (Deco & Lee, 2004), we do not aim to study the dynamics of spatial attention in BO determination. We focus on the magnitude and its variance of BO-selective cells, which lead to the modulation in the perception of figure direction.

V1 module

Model V1 cells extract local, oriented contrast from input stimuli, with the convolution of the image with a set
of Gabor filters including four orientations. The response of the model cells is determined by convolution with the input, the previous response of the cell, and feedback signals from PP. The activity of a model V1 cell, $A_{\theta xy}^{V1}$, is given by

$$\frac{\partial A_{\theta xy}^{V1}(t)}{\partial t} = -A_{\theta xy}^{V1}(t) + \mu F(A_{\theta xy}^{V1}(t)) + J_{\theta xy}^{V1,excit}(t) + J_{\theta xy}^{V1}(t),$$

where $x$ and $y$ indicate spatial locations, and $\theta$ and $\omega$ are the preferred orientation and spatial frequency, respectively. $J_{\theta xy}^{V1}(t)$ represents random noise, and $\mu$ represents the scaling constant. The activities of model V1 cells are modulated by the signal from the PP module in an exponential fashion, as proposed by (Lee et al., 1999; Peters et al., 2005), which is represented by $J_{\theta xy}^{V1,excit}(t)$. This exponential modulation acts on divisive normalization in which a contrast signal at a location is divided by a spatial pool of neighborhood contrasts, as details described in Equation A2 in Appendix A. This normalization is an inhibitory mechanism that is crucial for the stability in recurrent computation. Note that the PP module is designed to represent spatial attention that could be considered as a saliency map based on luminance contrast. According to this mechanism in the model, spatial attention increases the local contrast gain so that contrast at the attended location is enhanced. The activity of this cell represents the response of a V1 neuron to the stimulus projected onto its CRF. A detailed mathematical description is given in Appendix A.

### V2 module

Model V2 cells determine BO based solely on contrast signals surrounding the CRF that are extracted by the V1 module, as illustrated in (Figure 2A) (Sakai & Nishimura, 2006). Each model BO-selective cell has excitatory and inhibitory regions whose location, shape, and size give rise to the selectivity of the cell. The activity of a model V2 cell, $A_{\theta xy}^{V2,BO}(t)$, is given by

$$\frac{\partial A_{\theta xy}^{V2,BO}(t)}{\partial t} = -A_{\theta xy}^{V2,BO}(t) + \mu F(A_{\theta xy}^{V2,BO}(t)) - \gamma F(A_{\theta xy}^{V2,Inh}(t)) + I_{\theta xy}^{V2-V1,BO}(t) + J_{\theta xy}^{V2}(t),$$

where $N$ represents the type of BO-selective cell that is defined by the surround regions. An index BO represents BO selectivity in which, for the sake of simplicity, we limited ourselves to either the left or right in order to consider only vertical borders in the simulations. If the activities of model BO-left-selective cells are dominant over the activities of model BO-right-selective cells, the direction of figure is determined as left. $I_{\theta xy}^{V2-V1,BO} N(t)$ indicates random noise. Note that there is no direct attention effect for the BO determination in V2 module. DOF is determined solely by the bottom-up signals from V1 module, and attention modulation is applied only to V1 module. The exclusion of feedback from PP to V2 enables us to examine the effects of attention modulation in V1.

The activity of $I_{\theta xy}^{V2-V1,BO} N(t)$ is determined by retinotopical, afferent signals from V1 including surrounding contrast (Nishimura & Sakai, 2005; Sakai & Nishimura, 2006):

$$I_{\theta xy}^{V2-V1,BO} N(t) = O_{\theta xy}^{V1}(t)(O_{\theta xy}^{V1}(t) + O_{\theta xy}^{BO}(t)),$$

where

$$O_{\theta xy}^{BO} = Exc_{\theta xy}^{BO} - Inh_{\theta xy}^{BO}.$$ A detailed mathematical description is given in Appendix B. $O_{\theta xy}^{V1}$ is the feed-forward input from the V1 module, corresponding to the CRF responses. $O_{\theta xy}^{BO}$ is the contrast surrounding the CRF. $I_{\theta xy}^{V2-V1,BO}$ is determined from the summation of the CRF response, $O_{\theta xy}^{V1}$, and the surround response, $O_{\theta xy}^{BO}$, which represents surrounding modulation apparent in early visual areas (Jones, Grieve, Wang, & Sillito, 2001). The surround modulation in this module is given as a hard-wired circuit with a linear combination of the CRF signal and the surround contrast, although recent studies have suggested more detailed cortical mechanisms such as dynamics with feedback connections or pre-processing in LGN (Smith, Bair, & Movshon, 2006; Wieland & Sajda, 2006). Because, the exact mechanisms of surround modulation has not been clarified, and that mechanisms are out of our focus. The multiplication with $O_{\theta xy}^{V1}$ acts as a switch so that no response is observed when no stimulus is projected onto the CRF. Exc_{\theta xy}^{BO} and Inh_{\theta xy}^{BO} are excitatory and inhibitory surrounding contrast signals, respectively, which are determined by the spatial convolution of V1 responses with the corresponding surround regions with a Gaussian shape as illustrated in (Figure 2B). Although a wide diversity of BO selectivity has been reported in physiological experiments (Zhou et al., 2000), we selected 10 types of surround regions from a pool of Gaussians generated randomly (Nishimura, 2007). Although it may be intuitive to include this surrounding modulation in the V1 module, we included the process in the V2 module to simplify computation. It has been also reported that V2 neurons exhibit similar surrounding modulation (Ito & Komatsu, 2004). $A_{\theta xy}^{V2,inh}$ in Equation 4 shows the activity of an inhibitory neuron. We implemented a single inhibitory unit for each
of the V2 and PP modules in order to limit the activities of the module within a certain range. The activity of the model inhibitory cell for V2 is given by

$$\frac{\partial A_{\text{V2,inh}}(t)}{\partial t} = -A_{\text{V2,inh}}(t) + \lambda F\left(A_{\text{V2,inh}}(t)\right) + \kappa \sum_{x,y,N,BO} F\left(A_{N,BO}(t)\right),$$

where $\kappa$ and $\lambda$ are scaling constants. The inhibitory neuron receives inputs from excitatory neurons and inhibits all of them.

**PP module**

The PP module represents spatial attention that will facilitate the contrast processes within the attended
location in the V1 module. The PP module determines where attention from the activity of V1 module as well as the top-down spatial attention assigned for each condition should be applied. The activity of a model PP cell, $A_{xy}^{PP}$, is given by

$$\tau \frac{d A_{xy}^{PP}(t)}{dt} = -A_{xy}^{PP}(t) + \mu F(A_{xy}^{PP}(t)) - \gamma F(A_{xy}^{PP, inh}(t)) + I_{xy}^{PP-V1}(t) + I_{xy}^{PP, A}(t) + I_{xy}^{PP, noise}(t). \quad (8)$$

$I_{xy}^{PP, A}$ represents the bias of spatial attention given by a Gaussian in the simplified shape of a Mexican hat (Müller, Mollenhauer, Rösler, & Kleinschmidt, 2005) with standard deviation $\sigma_G = 1.0^\circ$. $I_{xy}^{PP-V1}$ represents afferent signals from V1 to PP. These two inputs determine the strength of attention. $A_{xy}^{PP, inh}$ represents an input from an inhibitory PP neuron whose activity is given similarly to Equation 7. $I_{xy}^{PP, noise}$ represents random noise. The distribution of activities within the PP module could be considered as a saliency map in the sense that the distribution determines the attending location, although we do not aim to construct a physiologically detailed saliency map.

Simulation results

We investigated the role of spatial attention applied to early vision in the determination of BO, with the computational model consisting of the V1, V2, and PP modules. Spatial attention from the PP module alters contrast gain in the V1 module so that it enhances local contrast. Because BO is determined from the surrounding contrast extracted by the V1 module, the enhanced contrast will modulate the BO signal. If BO is ambiguous or if its signal is relatively weak in nature, the perceived BO may flip to the other side, depending on the location of attention. On the other hand, if BO is apparent or if its signal is solid and stable, the perceived BO may not be changed even if the location of attention is altered.

As the first test of the model, we carried out simulations of the model with unambiguous and ambiguous figures in order to observe the model’s behavior. Specifically, we tested whether the model reproduces human perception: the switch of BO for an ambiguous figure depending on the location of spatial attention, and the unchanged BO for an unambiguous figure. An example of such an unambiguous figure is a single square. Several examples of ambiguous figures are shown in the following sections.

Figure 3. Simulation results for a single square. The bottom illustrations show the configuration of the stimulus (a white square) and the CRF (a small red circle) of the model cell being examined. Spatial attention of Gaussian shape with SD of $1^\circ$ (a shaded, reddish circle with the “x” icon at its center) is applied to either (1) the center of the square, (2) outside of the square, in the opposite direction from the square relative to the CRF, and (3) nowhere. Black and white bars indicate the ratio of the activities between BO-left (black) and BO-right (white) model cells computed as a percentage. Regardless of the type of spatial attention, the model determines the square as figure.
Throughout this paper, we have computed the summation of the total activity of model BO-right cells with 10 types of surrounding regions that realize a variety of BO selectivity. We then compared the sum with that obtained from model BO-left cells. The ratio of the two sums for BO-right and BO-left cells was computed, and the BO direction of the dominant group was considered to own the border. For the sake of simplicity, we took into account only horizontal BO directions, either to the left or right with respect to a vertical border.

A single square

We simulated the model with a single square as illustrated in Figure 3. Humans tend to perceive a white square as figure in this stimulus, regardless of the location of attention. As a control, we tested whether the model reproduces this perception. We carried out the simulations of the model with three conditions; spatial attention is applied to (1) the center of the square, (2) the outside of the square in a direction opposite to that of the square, and (3) nowhere. We expected that the model cells at the center (indicated by a small red circle) would show a BO-left response for all three conditions. The model cells determine the BO based on their surround modulation, the suppressive/facilitatory regions of which are localized and asymmetric with respect to the CRF. The simulation results show that the activities of the model BO-left cells are dominant over those of the model BO-right cells, regardless of the location of spatial attention. The model determined the square as figure, which was not affected by spatial attention, in agreement with human perception.

Ambiguous figure

If BO is ambiguous in nature, the computed direction of BO may flip to that side of the border toward which attention is directed. We tested the model with three ambiguous figures: an easy Necker cube (Figure 4A), white and black pillars (Figure 4B) (after Shepard, 1990), and arrows and men (Figure 4C) (Shepard, 1990). The easy Necker cube with two adjacent squares might be the simplest case to test, in which either the right or left square could be perceived as figure with respect to the center edge. This stimulus would also be a test to check the applicability of the model to a line drawing stimulus. When we view the white and black pillars, we tend to perceive either white or black pillars as figure. However, we tend not to perceive simultaneously both white and black pillars as figure. The arrows and men figure includes two objects to be perceived alternatively, walking men and arrows. We chose these three figures because they are considered to possess weak meaning in their shape or weak bias of meaning. For instance, in the case of the Rubin’s vase, the objects seen are human faces and a vase, the human face appearing to have a much stronger meaning for human observers. Therefore, feature-based attention may play a crucial role in such stimuli, an issue that will not be a focus of this study.

We investigated whether our model reproduces the switching of BO when the location of spatial attention is changed to the other side. The simulation results are shown in Figure 5. For the easy Necker cube and the pillars without attention, the ratios of BO on the right and the left were 50% because the contrasts on the both sides were identical or very similar. The model cannot determine BO if a stimulus is symmetric with respect to the CRF so that the contrasts on the both sides were identical. The model may not be capable of determining BO for other nearly symmetric stimuli such as spirals in (Baek & Sajda, 2005). If attention is applied to a single side of a symmetric stimulus, contrast on that side is enhanced so that the equal balance of contrast is disrupted, which leads the direction of BO toward that side. For all stimuli, a dominant
The model showed the alternation of BO direction according to an attentive location. We carried out psychophysical experiments to test whether the characteristics of the model agree with human perception of attention modulation in BO determination. We introduced random-block stimuli consisting of two adjacent, block stimuli, each of which is comprised of randomly chosen blocks with several constraints so that the BO direction is ambiguous at their border. A total of 40 types of stimuli were presented in the experiments: five base types and their mirror images with respect to the vertical and horizontal midlines for two contrast polarities. The test will be objective and reduce the effects of other possible cues for BO determination because these white-noise-type stimuli approximate all possible shapes (Nishimura, 2007).

First, we tested whether spatial attention actually yields an alternation of BO direction in the random-block stimuli. Further, we tested whether the magnitude of the modulation is invariant among the stimuli. Stimulus invariance has been reported in an electrophysiological study of motion perception (Treue & Maunsell, 1999) and an fMRI study for texture-defined stimuli (Appelbaum, Wade, Vildavski, Pettet, & Norcia, 2006). Stimulus invariance will also support the idea that other cues for BO, including feature-based attention derived from stimulus shape, are not significant among these stimuli. We carried out the simulations of the model with the same sets of stimuli and compared the results with those obtained from psychophysics. Quantitative agreement of modulation magnitudes between psychophysics and the corresponding simulation will provide a basis for quantitative predictions of the model for other stimuli.

Second, we separated the modulations of spatial attention and feature-based attention and examined the sole effects of spatial attention that are apparent in the presence of feature-based attention. An introduction of a familiar shape into a meaningless stimulus will shift human determination of BO toward the familiar shape (Vecera & O’Reilly, 1998), while the magnitude of the modulation due to spatial attention will be similar to that observed in the absence of the familiar shape. This prediction is derived from the fact that spatial attention in early vision does not have specificity for shape so that spatial attention and feature-based attention might operate independently and additively. We carried out psychophysical experiments to examine quantitatively these characteristics of attention modulation in BO determination and compared the results with those from model simulations. Our model should show no shift in BO preference in disagreement with psychophysics and should show that modulation magnitude is similar to that obtained without
any familiar shape, in agreement with psychophysics. Such agreement and disagreement between psychophysics and the corresponding simulations will support the validity of the model.

Procedure for the psychophysical experiments

We generated stimuli consisting of two adjacent random blocks so that their BO appears to be ambiguous, as illustrated in Figure 6. We tested whether covert spatial attention affects the perception of BO in these stimuli. A detailed description of the stimulus generation is given in Appendix C. Figure 7 shows an illustration of the experimental procedure. A bright red cross (7.5 cd/m², 0.5° side) was flashed for 400 ms to the left or right side at 2.5° in visual angle from the center, followed by a blank screen of 100 ms, which directed the attention toward the location of the dot. Subjects were instructed to fixate on a blue dot (fixation aid) that was presented at the center of the screen throughout the trial. A single random-block stimulus (8.8 or 11.4 cd/m²) was presented within a 4 × 4 region for 200 ms with its center aligned to the screen center. The apparent direction of BO at the screen center was measured using a two alternative forced-choice paradigm in which subjects were asked to indicate which side appears in front of the other (i.e., appears as figure) by pressing a mouse button. We did not ask participants to indicate the direction of BO to eliminate the response bias in which subjects are more likely to report toward the cued side than the other side (Driver & Baylis, 1996; Vecera et al., 2004). To further ensure that this paradigm does not produce the response bias toward the cued side, we carried out control trials in which we asked subjects to report which side appears more distant than the other (i.e., appears as ground). There was no significant difference in the results between the two tasks. Using a randomized order of appearance of stimuli, we carried out 80 trials for each condition, and repeated the task for three subjects including one of the authors (RS). As a control, we also measured the BO directions for the same stimuli without attracting attention by the flash of the red dot. Stimuli were generated on a Linux-based PC.

Figure 6. Random-block stimuli with ambiguous BO at the center of the stimulus. The top row on the left set shows five base-type blocks. Mirror images with respect to the vertical and horizontal midlines, and those with the opposite contrast yield a total of 40 stimuli.

Figure 7. The procedure of the psychophysical experiment. Subjects at the viewing distance of 115 cm were asked to decide which side (left or right) was perceived as figure using a 2AFC task. See text for details.
(Dell Precision 360) and displayed on a 21-in. CRT monitor (Dell P1130) with the refresh rate of 60 Hz. The viewing distance was 115 cm.

**Ambiguous block stimuli**

We examined whether spatial attention evokes an alternation of BO direction for random-block stimuli with ambiguous BO. The stimuli consisted of two adjacent, block stimuli with their vertical border positioned at the screen center where the fixation aid was located. The measured percentage of the apparent direction of BO is shown in (Figure 8A) with the stimuli tested indicated at the bottom. Note that data for all mirror images and contrast polarities for the three subjects were combined for each base-type stimulus. The apparent direction of BO was shifted toward the direction of attention for all stimuli presented, its modulation magnitude being indicated by arrows. Here, we focus on this modulation magnitude as the attention effect in the early visual area. To examine whether the magnitude of the modulation depends on the

![Figure 8. Results of the psychophysics and the corresponding simulations.](image)

(A) The result of the psychophysical experiment. Black bars show the apparent perception of BO left for each base-type stimulus indicated at the bottom. Three conditions for attending location are indicated by 0, 1, and 2, corresponding to nowhere, location 1, and location 2, respectively, as the locations indicated in each stimulus. Numbers at the bottom show the $m$ value that represents the magnitude of attention modulation (see Equation 9). Error bars indicate the standard error. The result shows the tendency for the attended block to be perceived as figure. Black arrows represent the magnitude of attention modulation. There is no significant difference in modulation magnitude among the stimuli. (B) The results of the corresponding simulations with the same stimulus sets. The icon at the top right corner shows the CRF of the cells being examined, and the size and location (1) of spatial attention. The model reproduces characteristics similar to those observed in psychophysics.
stimulus, we introduce a modulation index $m$, similar to that presented in Treue and Maunsell (1999),

$$m = \frac{P_b(\text{attn(black)}) - P_b(\text{attn(without)})}{P_b(\text{attn(black)}) + P_b(\text{attn(without)})}$$

$$+ \frac{P_w(\text{attn(white)}) - P_w(\text{attn(without)})}{P_w(\text{attn(white)}) + P_w(\text{attn(without)})},$$

where $P_b(\ )$ and $P_w(\ )$ indicate the proportion of trials on which the black and white shapes, respectively, are perceived as figure, and attn( ) represents the attention condition such as focusing attention on white and black shapes. In the following sections, we will refer to the $m$ index as the modulation magnitude. A one-way analysis of variance (ANOVA) of the $m$ values with the factor of stimulus types and the repetition of subjects showed no significant difference ($p = 0.940$). In the control experiments, we asked subjects to report which side appears more distant than the other (i.e., appears as ground). There was no significant difference in $m$ values between the test and control experiments (ANOVA, $p = 0.244$), indicating no significant bias toward the cued side. These results indicate that the magnitude of the modulation is independent of stimulus shape. The shape invariance of attention modulation supports that contrast modulation in early vision is crucial for the modulation of BO.

Figure 9. Random-block stimuli with a familiar shape. (A) Stimuli consisting of familiar and unfamiliar shapes. The top row on the left set shows three base-type blocks. Mirror images with respect to the vertical and horizontal midlines, and those with the opposite contrast yield a total of 24 stimuli. (B) Control stimuli consisting of two unfamiliar shapes.
We carried out simulations of the model with the same stimuli as those used in the psychophysical experiments to compare directly the model’s performance with human perception. Figure 8B shows the simulation results with the same conventions as those used in Figure 8A. The standard deviation of the modulation magnitude was 0.027, which is similar to that obtained from the psychophysics (0.048). We carried out one-way ANOVA of $m$ values for the psychophysics and the simulations with the repetition of stimulus types in order to compare their modulation magnitudes. The result showed no significant difference between the two ($p = 0.293$), indicating that the modulation of the model agrees quantitatively with human perception. These results indicate that the effects of spatial attention in early vision could be a crucial factor for the modulation of BO. Although all free parameters in the model are relatively insensitive to the BO determination, the results presented here do not prove the quantitative agreement of absolute values between the model and human responses in general. It should be noted, however, that since all parameters were fixed throughout, the validity of the model within the context of this paper is veridical.

**Stimuli including a familiar shape**

The introduction of a familiar shape into a random-block stimulus, which appears to evoke feature-based attention in perception (Rotte, Heinze, & Smid, 1997;
Vecera & Farah, 1997), will shift the human determination of BO toward the familiar shape. However, the magnitude of the modulation due to spatial attention will be similar to that observed in the absence of the familiar shape. Since spatial attention in early vision does not have specificity for shape, spatial attention and feature-based attention might operate nearly independently and additively. We carried out psychophysical experiments to test this prediction together with the corresponding simulations of the model. We presented two adjacent random-block stimuli, similar to those used in the previous experiment but with one fixed to a square as a familiar shape as shown in Figure 9. We tested a total of 24 stimuli: three base patterns with their mirror images and two contrast polarities (familiar set). As a control, stimuli with the same blocks except for the square were also tested (unfamiliar set). We added a single rectangle to both stimulus sets to monitor whether subjects actually judge BO direction. The experimental procedures were identical to those described in the previous sections.

The measured percentage of the apparent direction of BO is shown in Figure 10, with the stimuli tested being shown at the bottom. From a comparison of the results for stimuli with and without the familiar shape, it is apparent that the direction of BO tends to agree with that of the familiar shape. There is a significant difference between stimulus sets with and without familiar shape in the determination of BO direction without attention (one-way

Figure 11. Simulation results corresponding to the psychophysics with a familiar shape, with the same conventions as those for Figure 8. (A) The results for a combination of familiar and unfamiliar shapes. The direction of BO is independent of that for the familiar shape, which disagrees with human perception. (B) The results for unfamiliar shapes. Black arrows show the modulation magnitude derived from spatial attention. There is no significant difference in modulation magnitude between the stimulus sets with and without a familiar shape, in agreement with human perception.
ANOVA with the repetition of stimulus types and subjects, $p = 0.021$). This tendency agrees with previous reports on BO determination with a familiar shape (Rotte et al., 1997). The apparent direction of BO shifts toward the direction of spatial attention for all stimuli including those with and without a familiar shape, there being no significant difference in modulation magnitude between the two stimulus sets (one-way ANOVA, $p = 0.620$). The modulation invariance between the conditions with and without a familiar shape could be attributed to the attention modulation in early vision. The values of the modulation magnitude are somewhat smaller than those in the previous section, which is caused by the smaller number of stimulus types. These results indicate that spatial attention is similarly effective regardless of the presence of a familiar shape. Modulation magnitude appears to be a reliable measure to indicate the strength of the effects of spatial attention. This result suggests further that the effect of feature-based attention derived from familiarity is stronger than that of spatial attention, so that the BO direction does not flip in stimuli containing a familiar shape.

To compare the model’s responses with human perception, we carried out simulations of the model using the same stimulus sets as were used in the psychophysical experiment. Figure 11 shows the simulation results. Although the psychophysics showed a tendency for a familiar shape to be perceived as figure, the model did not reproduce this property. There was no significant difference in the determination of BO direction between stimulus sets with and without a familiar shape (one-way ANOVA with the repetition of stimulus types, $p = 0.211$ for the “no attention” condition). This disagreement with psychophysics is natural because the model does not include any mechanism for feature-based attention. The model did not show a significant difference in modulation magnitude between the stimulus sets (one-way ANOVA, $p = 0.108$) in agreement with the psychophysics, thus indicating the quantitative validity of the model. These results suggest that spatial attention operates independently of feature-based attention in BO determination, and that the modulation of contrast gain in V1 underlies the modulation of BO direction based on spatial attention.

**Discussion**

We proposed a computational mechanism underlying attention modulation in BO perception, which modulates BO by control of contrast gain in the early visual area. Simulations of the model with an unambiguous figure showed stable activity of BO-selective model cells independently of the location of spatial attention, in agreement with human perception. In contrast, when ambiguous figures were presented, the computed BO direction was altered following the location of spatial attention. Our results suggest that ambiguous figures could be subjected to BO alternation that is due to a change in local contrast in early vision as a consequence of spatial attention. We carried out psychophysical experiments to test whether the characteristics of the model agree with human perception of attention modulation in BO determination. We introduced random-block stimuli that consisted of two adjacent, block stimuli that had an ambiguous BO at their border. Comparison of the modulation magnitude showed no significant difference between the model and psychophysics. When a stimulus included a familiar shape, human determination of BO was shifted in the direction of the familiar shape, while our model did not exhibit such a shift. However, there was no significant difference in modulation magnitude between the familiar and unfamiliar shapes for both the model and psychophysics. This similarity in the modulation values supports that the model provides a basis for modulation of BO based on spatial attention.

A crucial characteristic of the model is its short latency for attention modulation and BO determination. The model requires neither recurrent computation nor lateral connection. Physiological studies have shown that BO selectivity occurs about 50 ms after the response to the edge, and that the latency is independent of figure size (Zhou et al., 2000). These results suggest a role of fast feedback/feedforward channels in BO determination rather than intralayer lateral connections (Sugihara, Qiu, & von der Heydt, 2003). The coexistence of top-down attention and bottom-up attention in a single architecture is another crucial characteristic of the model. Physiological studies have shown both cases, namely that figure attracts attention and attention alters figure (Kimchi, Yeshurun, & Cohen-Savransky, 2007; Vecera et al., 2004). Both figure region and saliency map could be represented by populations of cells without a hypothetical homunculus that observes lower levels.

Physiological studies have shown attention modulation in a number of cortical areas including intermediate-level visual areas, such as V2 and V4, where a majority of neurons are BO selective. Although the present study has focused on investigating the indirect modulation of BO-selective cells through an afferent channel from V1, attention is likely to directly modulate BO-selective cells using a mechanism similar to the one we utilized for V1. The attention modulation in V1 alters the activities of BO-selective cells, and then the direct attention modulation in V2, including object-based attention, could further lead to complete figure/ground flipping (Wagatsuma & Sakai, 2006; Wagatsuma, Shimizu, & Sakai, 2008). However, the direct modulation of BO determination appears not to be straightforward because attention has to modulate selectively those cells with a particular BO preference to bias the BO direction. This could not be established easily by spatial attention with location specificity. If there is a group of cells in higher cortical areas that represents a figure or shape based on the activities of BO-selective
cells, feedback from these cells could realize such modulation in order to influence BO determination (Craft, Schütze, Niebur, & von der Heydt, 2007). However, such grouping cells are hypothetical and have not been reported physiologically. It may also be probable that feature- and object-based attention is influential for BO determination in intermediate-level visual areas. Although our simulation results exhibited the appropriate modulation of BO direction through gain control of V1, our study did not provide evidence to assure the magnitude of indirect modulation with respect to the direct modulation of intermediate-level areas. A further study of attention modulation in V2 and V4 is necessary to understand the modulation of BO determination.

Feedback from higher to lower cortical areas influences significantly responses of neurons and thus perception. In our model, the connection from V2 to V1 is excluded to examine the exclusive role of contrast modulation in V1 for BO determination. The effect of feedback from BO-selective cells has not been clarified for the modulation of contrast gain in V1 neurons. However, this feedback appears to be crucial and could alter the activities of V1 cells. For example, a modeling study (Hatori & Sakai, 2008) has suggested that the feedback from BO-selective cells is crucial for the formation of medial axis in V1 that represents a primary sketch of shape and surface (Lee, Mumford, Romero, & Lamme, 1998). The feedback from BO-selective cells to V1, as well as direct attention modulation of the BO-selective cells, appears to play a crucial role in figure-ground segregation and surface perception. Further studies on feedback modulations would be expected to clarify the perception of figure and surface.

Recent physiological and psychophysical studies have reported that object-based attention is observed in early visual areas (Roelfsema, Lamme, & Spekreijse, 1998; Shomstein & Behrmann, 2006). Deco and Rolls have proposed that both spatial- and object-based attention could work closely in a single architecture (Deco & Rolls, 2004; Rolls & Stringer, 2006). The incorporation of object- and feature-based attention into our model might allow a bias of BO direction with a familiar shape. Our model suggests that attention modulation of BO direction originates at least in part, from modulation of contrast sensitivity in the early visual area. Attention enhances the contrast that in turn changes the activity of BO-selective neurons. Our simulations show that enhanced contrast modulates or even flips the perception of figure direction, the modulation variance of which showed good agreement with human perception. These results provide important predictions for visual attention in figure-ground segregation and shape perception.

### Appendix A

A detailed mathematical description of the model V1 cell is given here. The input image, Input, is a $124 \times 124$ pixel, gray scale image with an intensity value ranging between zero and one. We designed the model so that 25 pixel corresponds to one degree ($^\circ$) in visual angle. Local contrast, $C_{\text{local}}$, is extracted by the convolution of the image with a Gabor filter, $G_{\omega}$:

$$C_{\text{local}} = \sum_{i=1}^{m} \sum_{j=1}^{n} G_{\omega}(i, j) \text{Input}(x - \frac{m}{2} + i, y - \frac{n}{2} + j),$$

(A1)

where index $x$ and $y$ indicate spatial location and $\omega$ indicates the spatial frequency. Because we have limited input stimuli, we chose a single frequency of 0.5$^\circ$ wavelength that is optimal for the extraction of contours from the stimuli. Orientation, $\theta$, is chosen from 0, $\pi/2$, $\pi$, and $3\pi/2$. $m$ represents the number of pixels in the Gabor filter $G_{\omega}$. An input image could be a line drawing, as long as the Gabor filter can effectively extract local contrast.

Spatial attention that is represented in PP modulates the contrast gain in V1 as in Equation 3 in the main text. Local contrast, $C_{\text{local}}$, is modulated by attention that is given by the feedback from PP to V1, $I_{xy}^{V1-PP}$. The modulated contrast, $I_{xy}^{V1-PP}$, is given by the following equations, as proposed by (Lee et al., 1999; Peters et al., 2005):

$$I_{V1\text{excit}}^{V1}(t) = \sum_{\theta} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} C_{\text{local}}(x - \frac{m}{2} + i, y - \frac{n}{2} + j) \cdot I_{xy}^{V1-PP}(t)$$

(A2)

$$F_{xy}^{PP}(t) = \sum_{i=1}^{k} \sum_{j=1}^{l} W_{ij} F(A_{xy}^{PP}(x - \frac{m}{2} + i, y - \frac{n}{2} + j)(t))$$

(A3)

$$W_{ij} = a \exp \left( -\frac{(i - \frac{m}{2})^2 + (j - \frac{n}{2})^2}{2\sigma^2} \right)$$

(A4)

where $W_{ij}$ represents connection weights of the Gaussian (Deco & Lee, 2004; Deco & Rolls, 2004) with the standard deviation of $\sigma_{ij}(0.53^\circ)$. $W_{ij}$ were chosen so that the total sum is one. $F(A(t))$ in Equation A3 is identical to that in Equation 2 in the main text. $A_{xy}^{PP}$ shows the activity of a model PP cell as shown in Equation 8 in the main text. $k$ and $l$ show the spatial extent of the feedback from PP cells to a single V1 cell. $a$, $\chi$, $\delta$, and $\gamma$ are constants. $S$ is a semi-saturation constant which prevents the denominator to be zero (Lee et al., 1999; Peters et al., 2005). In our simulation, we used $a = 0.25$, $\chi = 0.6$, $\delta = 0.1$, and $\gamma = 0.05$. The connection from V2 to V1 is excluded to examine the exclusive role of contrast modulation in V1 for BO determination. The effect of feedback from BO-selective cells has not been clarified for the modulation of contrast gain in V1 neurons. However, this feedback appears to be crucial and could alter the activities of V1 cells. For example, a modeling study (Hatori & Sakai, 2008) has suggested that the feedback from BO-selective cells is crucial for the formation of medial axis in V1 that represents a primary sketch of surface and shape (Lee, Mumford, Romero, & Lamme, 1998). The feedback from BO-selective cells to V1, as well as direct attention modulation of the BO-selective cells, appears to play a crucial role in figure-ground segregation and surface perception. Further studies on feedback modulations would be expected to clarify the perception of figure and surface.
Appendix B

A mathematical description of the surround modulation of a model BO-selective cell in V2 module is given here. The model determines BO based on surrounding contrast (Nishimura & Sakai, 2005; Sakai & Nishimura, 2006).

First, V2 module pools the contrast signals that are modulated by spatial attention in V1 module over space and frequency,

\[
O_{\text{xy}}^1 = \sum_{i=1}^{m} \sum_{j=1}^{m} W_{\text{xy}ij} F \left( A_{\text{xy}}^1 \left( x - \frac{\sigma}{2} + i \right) \left( y - \frac{\sigma}{2} + j \right) \right), \quad (B1)
\]

\[
O_{\text{xy}}^{1,\text{cross}} = O_{\text{xy}}^1 (\theta + \frac{\pi}{2}) + O_{\text{xy}}^1 (\theta + \frac{\pi}{2}) \text{,} \quad (B2)
\]

\[
O_{\text{xy}}^{1,\text{iso}} = O_{\text{xy}}^1 (\theta) + O_{\text{xy}}^1 (\theta + \pi) \text{.} \quad (B3)
\]

\(x\) and \(y\) represent the location of the CRF of the model cell. \(m\) indicates the spatial extent of feed-forward from V1 module. \(O_{\text{xy}}^1\) shows the feed-forward input from V1. Index \text{cross} and \text{iso} represent contrasts orthogonal and parallel to the preferred orientation, \(\theta\), of the cell, respectively. \(W_{ij}\) represents the Gaussian function as shown in Equation A5. \(O_{\text{xy}}^{1,\text{cross}}\) and \(O_{\text{xy}}^{1,\text{iso}}\) show the modulated contrast of cross-orientations and iso-orientations, respectively.

Second, the surrounding signal, \(O_{\text{xy}}^{2,\text{BO}}\), is given by a linear combination of contrast signals from excitatory and inhibitory regions that are defined by Gaussian functions as illustrated in Figure 2B.

\[
O_{\text{xy}}^{2,\text{BO}} = \text{Exc}_{\text{xy}}^\text{BO} - \text{Inh}_{\text{xy}}^\text{BO}, \quad (B4)
\]

\[
\text{Exc}_{\text{xy}}^\text{BO} = c_a \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} E_N^\text{BO} (i, j) O_{\text{xy}}^{1,\text{cross}} \left( x - \frac{\sigma}{2} + i \right) \left( y - \frac{\sigma}{2} + j \right), \quad (B5)
\]

\[
\text{Inh}_{\text{xy}}^\text{BO} = c_b \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} E_N^\text{BO} (i, j) O_{\text{xy}}^{1,\text{iso}} \left( x - \frac{\sigma}{2} + i \right) \left( y - \frac{\sigma}{2} + j \right), \quad (B6)
\]

\(\text{Exc}_{\text{xy}}^\text{BO}\) and \(\text{Inh}_{\text{xy}}^\text{BO}\) represent the contrast signals within the facilitatory and suppressive regions, respectively. The index \(N\) represents the type of model BO cells that are distinguished by their surround regions. We implemented 10 types of surround regions from a pool of Gaussians generated randomly (Nishimura, 2007) to reproduce a diversity of BO selectivity (Zhou et al., 2000). \(E_N^\text{BO}\) and \(E_i^\text{BO}\) represent the facilitatory and suppressive regions of the model BO-selective cell. \(n_x\) and \(n_y\) indicate the spatial extent of excitatory and inhibitory regions. The combination of excitatory and inhibitory regions determines the property of BO selectivity. Such localized, asymmetric, and orientation dependent organization is observed in surrounding modulation in V1 neurons (Jones, Wang, & Sillito, 2002). \(c_a\) and \(c_b\) are connection strength. These constants for surrounding modulation (\(n_x\), \(n_y\), \(c_a\), and \(c_b\)) were determined following the references (Nishimura, 2007; Sakai & Nishimura, 2006). The balance of the facilitation and suppression determines modulation of the model BO-selective cell.

Third, the response of a model BO cell, \(I_{\text{xy}}^{\text{V2−V1,BO}}\), is determined based on a linear summation of the CRF signal, \(O_{\text{xy}}^1\), and the surround signal, \(O_{\text{xy}}^{2,\text{BO}}\) (Nishimura & Sakai, 2005; Sakai & Nishimura, 2006).

\[
\text{If } O_{\text{xy}}^1 (t) + O_{\text{xy}}^{2,\text{BO}} (t) > 0 \quad (B7)
\]

\[
I_{\text{xy}}^{\text{V2−V1,BO}} (t) = O_{\text{xy}}^1 (t) \times \left( O_{\text{xy}}^1 (t) + O_{\text{xy}}^{2,\text{BO}} (t) \right), \quad (B8)
\]

otherwise

\[
I_{\text{xy}}^{\text{V2−V1,BO}} (t) = 0. \quad (B9)
\]
For the determination of direction of figure, the activities of model BO-selective cells are pooled to represent the population activities. For the sake of simplicity, we took the summation of all activities of BO-right cells and that of BO-left cells. The dominant population was considered to own the boarder.

Appendix C

We carried out simulations with ambiguous random-block stimuli to investigate whether our model reproduces human perception. The random-block stimuli approximate all possible shapes under certain conditions. The method for the generation of random-block stimuli is given here.

A block stimulus consisted of \( n \) square blocks placed on \( m \times m \) grids, with center \( 2 \times 2 \) grids were fixed with the two blocks placed on one side of the center so that contrast within the CRF was kept identical for all stimuli. The other blocks \( (n-2) \) were placed randomly adjacent to the existing blocks except for the CRF region. We call this set of stimuli as \( n \)-block stimuli. Stimulus shape is more complex with a larger number of blocks and grids. In this paper, the number of grid was fixed to \( 4 \times 4 \) and that of blocks were 3, 4, 6, and 8.

We brought together two block stimuli to produce a stimulus with ambiguous BO at the center as illustrated in Figure 6. Ambiguous random-block stimuli were comprised of a combination of the block stimulus with BO left and that with right. All combinations of these stimuli were prepared, and if there was an overlap of blocks, we excluded such combination. Because this method generated a number of stimuli, we selected five stimuli that appears most ambiguous from visual inspection of several people who did not participate in the experiments.

Acknowledgments

This research was supported by JSPS and Ministry of ECSST of Japan (KAKENHI 19530648 & 19024011), The Okawa Foundation (06-23), and The Brain Science Foundation.

Commercial relationship: none.
Corresponding author: Nobuhiko Wagatsuma.
Email: wagatsuma@cvs.cs.tsukuba.ac.jp.
Address: Department of Computer Science, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8573, Japan.

Footnote

1 Pairwise \( t \) test showed a similar \( p \)-value. We described the result of ANOVA to enable a direct comparison with simulation results.

References


Roelfsema, P. R., Lamme, V. A., Spekreijse, H., & Bosch, H. (2002). Figure-ground segregation in a recurrent network architecture. Journal of Cognitive Neuroscience, 14, 525–537. [PubMed]


