# The neural computation of the aperture problem: an iterative process

Masato Okada,[1,2,CA] Shigeaki Nishina[3] and Mitsuo Kawato[1,3]

[1]Kawato Dynamic Brain Project, ERATO, JST and [3]ATR Computational Neuroscience Laboratories, Hikaridai 2-2-2, "Keihanna Science City" Kyoto 619-0288, Japan; [2]Laboratory for Mathematical Neuroscience, RIKEN Brain Science Institute, Hirosawa 2-1, Saitama 351-0198, Japan

[CA,2]Corresponding Author: okada@brain.riken.go.jp

Received 26 March 2003; accepted 22 May 2003

DOI: 10.1097/01.wnr.0000090586.35425.f4

The aperture problem is defined as one of integrating motion information from inside and outside of the aperture, and determination of the true direction of motion of a line. Much is known about it and many models have been proposed for its neural mechanisms. However, it is still a matter of debate whether the brain solves the problem by using only feed-forward neural connections, also known as the one-shot algorithm, or by using the iterative algorithm while utilizing feedback as well as horizontal neural connections. Here we show unequivocal evidence for the latter model. The model was tested using critically designed psychophysical experiments and the results were perfectly in line with the psychophysical performance of the observers. *NeuroReport* 14:1767–1771 © 2003 Lippincott Williams & Wilkins.

**Key words**: Aperture problem; Binding problem; Iterative algorithm; Motion perception; One-shot algorithm; Vision

## INTRODUCTION

Although the direction of movement of a line behind an aperture is ambiguous (Fig. 1a) [1], the brain resolves such ambiguity by integrating spatially distant motion information (Fig. 1b) in one of two ways. The one-shot algorithm extracts some feature points within an object and computes their unambiguous motion direction by utilizing only feed-forward neural connections between the hierarchical visual cortical areas [2–4]. The iterative algorithm, on the other hand, temporally modifies the distributed local representations of motion and binding properties for many small segments within the object by utilizing feedback and horizontal connections [5–7]. Whether the brain adopts the iterative algorithm has been a major topic of vision research, but the topic remains controversial.

Temporal changes of visual perception [8–11] or temporal changes in neural firings in visual areas [12,13,14] by themselves do not necessarily support iterative computation because they can be equally well explained within the one-shot framework by the delays and/or temporal dynamics of sensors and/or single neurons in the visual system. For example, Pack and Born [14] reported that MT cells encode motion perpendicular to the orientation bars at initial firings, but the responses of MT cells temporally and gradually change to encode true object motion. However, they concluded that their data could not discriminate the two possibilities. Here, we combine psychophysical experiments and computational studies to provide decisive evidence for the existence of iterative computation.

## MATERIALS AND METHODS

***Psychophysical experiments:*** Figure 1c shows the presentation sequences of three kinds of visual stimuli with a time abscissa for three different experiments. The target stimulus consists of three collinear line segments observed behind three apertures (shown in Fig. 1b) [15]. Here, the direction perpendicular to the line was defined as 45°, and the object motion direction was defined as 90°.

The target stimulus was followed by the standard stimulus of 300 ms, which consisted of random dots moving in the 75° direction within the central aperture. The subjects judged whether the perceived direction of motion of the central segment of the target stimulus was closer to 45 or 90° while comparing it with the standard stimulus (two-alternative forced choice: 2AFC). Although the object motion direction is variable depending on the quadrants where stimuli are presented, we call the object motion direction upward, hereafter, for simplicity.

In Experiment 1, the presentation time of the target stimulus (100, 200, 300, 400, 500, or 600 ms), the gap size (GS) and the aperture diameter (AD) were systematically varied, i.e. (AD,GS)=(1.1,0.1), (1.1,1.9), (1.1,1.7), (1.1,2.5), (1.9,1.7), (2.7,0.9), (3.5,0.1) degrees of visual angle. In Experiments 2 and 3, the duration of the target stimulus was fixed at 300 ms, and in addition to the target stimulus and the standard stimulus, preceding stimuli of various durations (0, 200, 400, or 600 ms) were given before the target stimulus (Fig. 1c). In Experiment 2, the center aperture and a translating line segment in it

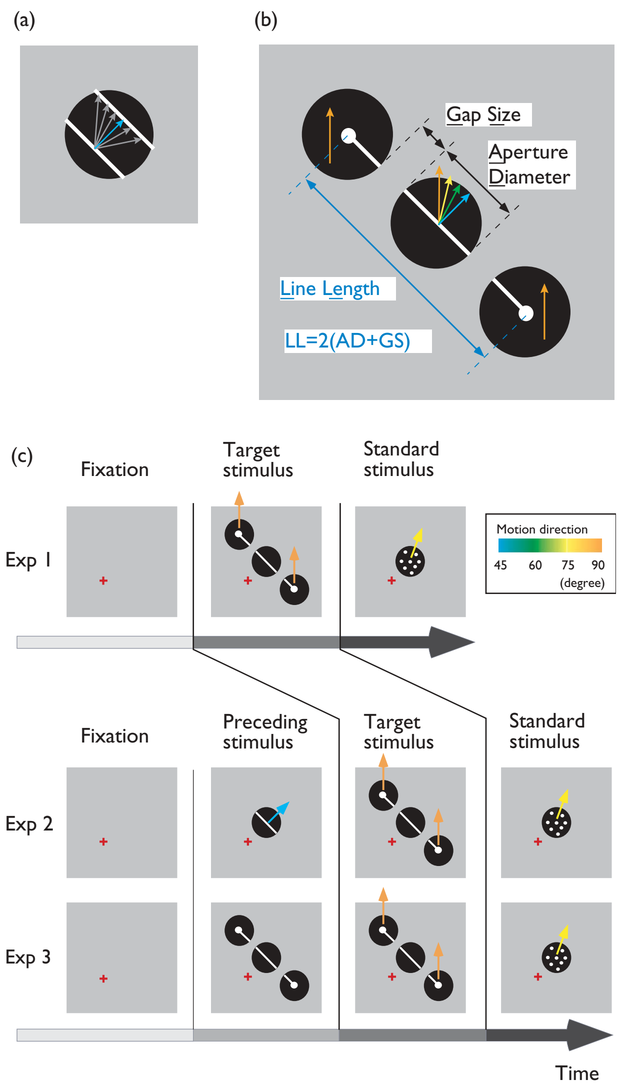

**Fig. 1.** (**a**) Illustration of the aperture problem. (**b**) Spatial configuration of the target stimulus. All three aperture diameters are the same, and the two endpoints pass through the centers of the two peripheral apertures. The two gap sizes of the occluded regions are the same. Therefore, the length of the whole line is twice the sum of the aperture diameter and the gap size (2 AD + 2 GS). The endpoints move upwards (90°) as denoted by the orange arrow. The green and yellow arrows represent 60 and 75°, respectively. (**c**) Display sequences in three experiments. First, a fixation point is presented alone at the center of the screen for 1500 ms. In Experiment 1, a series of target and standard stimuli follows the fixation. The stimulus in (b) is presented as the target. In Experiments 2 and 3, preceding stimuli are shown before the target stimuli.

were presented as the preceding stimuli (Fig. 1c). The line segment in the center aperture translated continuously and at a constant rate through the two periods. That is, the final location of the preceding stimulus was the same as the initial location of the target stimulus. In Experiment 3, the same three line segments as the target stimulus were statically presented as the preceding stimuli at the initial location of the target stimulus (Fig. 1c). The same 2AFC as Experiment 1 was required in Experiments 2 and 3.

***Computational model:*** The proposed iterative model possesses two independent and local representations for the velocity and binding of many small segments on the line [16–20]. The coordinate $s$ is defined on the line length of $l$, and $s = 0$ and $s = l$ correspond to the endpoints. The velocity $\mathbf{V}(s)$ and the binding process $b(s)$ change so that the following cost function $E$ is minimized:

$$E(\{\mathbf{V}(s),\, b(s)\} = \int_{s\in \text{Aperture}} (\mathbf{V}(s)\cdot N(s) - \mathbf{V}^N(s))^2\, ds + \lambda_1 \int_{s\in \text{Line}} b(s)\left(\frac{\partial \mathbf{V}(s)}{\partial s}\right)^2 ds + \lambda_2 \int_{s\in \text{Line}} \left(\frac{\partial b(s)}{\partial s}\right)^2 ds, \qquad \text{eqn 1}$$

where $N(s)$ denotes the normal vector at point $s$ on the line. The first term represents the goodness of fit of the data. The second and third terms are spatial integrations of the squares of spatial derivatives of $\mathbf{V}(s)$ and $b(s)$ on the whole line. Here, $\lambda_1$ and $\lambda_2$ are regularization parameters. The velocity and the binding process are updated in the steepest descent direction of $E$, where $\tau_v$ and $\tau_b$ are defined as time constants for changes of $\mathbf{V}(s)$ and $b(s)$, respectively. The proportion of perceived upward movements was obtained from the mean motion direction $\theta$ of the velocity vector within the center aperture by the psychometric function ($P = 1/(1 + \exp(a_1\theta - a_2))$).

## RESULTS

***Psychophysical experiments:*** As an example of the results from Experiment 1, the proportion of upward perception of subject UN for (AD,GS) = (2.7,0.9) is plotted in the upper panel of Fig. 2a as a function of the presentation time of the targetstimulus. The perceived direction was perpendicular to the line (45°) at the beginning of the target-stimulus presentation period, and gradually changed to the motion of the line (upward, 90°). The lower three graphs of Fig. 2a plot the average proportions of upward perception for seven subjects as a function of the aperture diameter and the gap-size at the three presentation times of 100 ms, 200 ms, and 500 ms. First, the figure shows that with a longer presentation time, the data points move upward. An ANOVA revealed that the proportion of upward perception increased significantly ($p < 0.0001$) with the presentation time. Second, the three graphs show that the proportion of upward perception decreased either with the aperture diameter or the gap size. An ANOVA revealed that the proportion of upward perception decreased significantly ($p < 0.0001$) with increases in the line length for all seven subjects [15].

The slopes of solid lines that indicate the negative effect of the gap size was steeper than the slopes of the broken lines, which indicate the negative effect of the aperture diameter. Quantitatively, the occluded line length had, on average, 2.5 times the effect of the visible line length. The solid lines in Fig. 2b,c denote the proportion of upward perception averaged over six subjects as a function of the preceding stimulus duration in Experiments 2 and 3 for a specific aperture diameter of 1.9 and a gap size of 1.7, respectively.

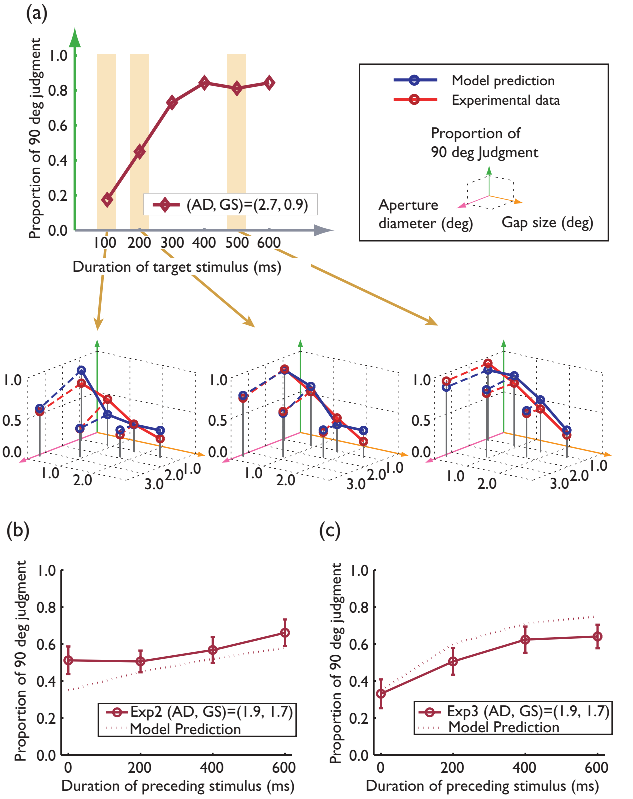

**Fig. 2.** (**a**) Experimental data and model reproduction. The lower figures show average values of proportions of upward perception for seven subjects (red) and corresponding model results (blue). These figures show how the motion perception changes for the presentation time of the target stimulus (100, 200, and 500 ms) and for different gap sizes and aperture diameters. (**b,c**) Examples of results of Experiments 2 and 3. Solid lines and dotted lines denote the experimental and theoretical results, respectively.

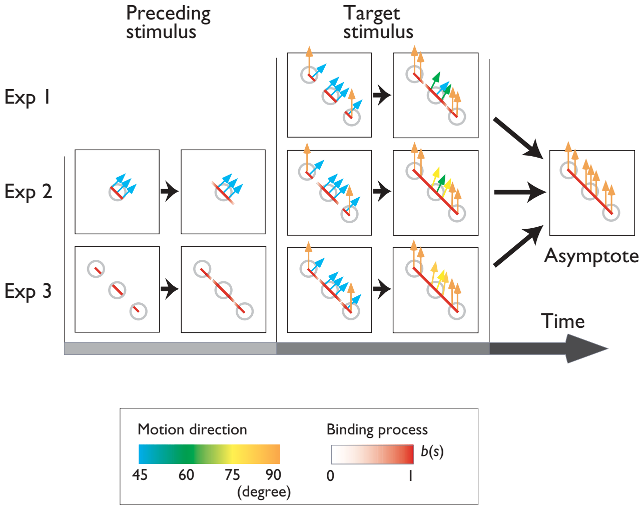

**Fig. 3.** Schematic explanation of the dynamics of the velocity vector and binding process. The abscissa denotes the time and the three rows correspond to the three different experiments. The corresponding motion vectors (blue to orange arrows) and the binding process (red lines) at the beginning and the end of the preceding stimulus presentation and the target stimulus presentation are shown. The velocity vector $\mathbf{V}(s)$ a small segment $s$ on the line is graphically represented by an arrow and its color. The binding process $0 \leq b(s) \leq 1$ is represented by the existence of the line and the red color. $b = 0$ indicates that the segment is not believed to belong to the line, and the color is white. $b = 1$ indicates that the segment is believed to belong to the line, and the color is red.

The proportion of upward perception increased significantly with increases in the preceding stimulus duration for all six subjects ($p < 0.02$ for Experiment 2 and $p < 0.005$ for Experiment 3). The averaged increase rate of this effect was 0.17 (proportion of upward perception)/sec in Experiment 2 and 0.26 (proportion of upward perception)/sec in Experiment 3.

***Computational model:*** For Experiment 1, at the beginning of the target stimulus presentation (the left rectangle in the first row of Fig. 3), the binding process is 1 for the visible part of the line and 0 for the occluded part. The velocity vector is upward for the two endpoints, and 45° for all of the other visible segments. As the presentation time elapses, the binding process diffuses into the occluded part across the T-junctions of the apertures, but not the endpoints (the third term in eqn 1). Thus the upward motion information is propagated gradually from the two endpoints into the interior points (the second term). If the presentation time of the target stimulus were to be infinitely long and, correspondingly, if the visual system had an infinitely long processing time, the asymptotic solution shown in the rectangle labeled 'Asymptote' would be obtained. Here the binding process would be 1 and the estimated velocity vector would be upward for all of the segments on the line. However, because of the finite presentation time, the obtained solution is on the way to this asymptotic solution. That is, the filling-in of the binding process is incomplete, and the motion direction of the central segment is $> 45^\circ$ but $< 90^\circ$ (the right rectangle in the first row of Fig. 3e). From this model, we can predict that the motion direction becomes closer to $90^\circ$ as the presentation time increases and as the line becomes shorter. We can further predict that the occluded line length has a stronger effect than the visible line length because both the binding process and the velocity information need to be propagated along occluded parts while only the velocity information needs to diffuse for visible parts, which is in agreement with the results of Experiment 1 as well as with those of Ben-Av and Shiffrar [15].

In Experiment 2, during the preceding stimulus duration, the velocity vector remains at 45° for the visible part, but the binding process diffuses even into the occluded part outside the central aperture (second row of Fig. 3). Because this non-zero value of the binding process outside the aperture is used as the initial condition for the target stimulus presentation, the propagation of the object motion information from the two endpoints is accelerated compared with Experiment 1. Accordingly, the model predicts that the proportion of upward perception increases with increases in the duration of the preceding stimulus.

In Experiment 3, during the preceding stimulus duration, the velocity vector is zero for all three visible parts, but the binding process again diffuses into the occluded parts in the two gaps (third row of Fig. 3). We note that even if the binding process is fully activated along the line, the estimated motion direction of all visible segments of the line, except two endpoints, is perpendicular to the contour at the beginning of the stimulus motion onset. Nevertheless, because the binding process diffuses both from the central and peripheral apertures, they are larger than in Experiment 2. Therefore, the model predicts that the proportion of upward perception increases even more markedly with increases in the duration of the preceding stimulus than in Experiment 2. Our model not only qualitatively but also quantitatively reproduces the experimental results well. The three parameters (the diffusion constants for the velocity and the binding process ($\lambda_1$, $\lambda_2$), and the ratio of time constants for them ($\tau_b/\tau_v$) in our model were determined to best reproduce the results of Experiment 1. These were

$$(\tau_v, \tau_b, \lambda_1, \lambda_2, a_1, a_2) = (1, 15, 0.015, 0.0025, 0.024, 1.4)$$

and the variance accounted for (VAF) was 0.88. The model was robust against the parameter changes. The VAF of the parameter of 55% change was 0.87, and that for the 122% was 0.73. As shown by the blue lines in the lower graphs of Fig. 2a, the model quantitatively captured the effects of the presentation time, the aperture diameter, and the gap size well. While the best parameters for Experiment 1 were fixed, the model also quantitatively reproduced the results of Experiments 2 and 3 well (VAF = 0.87 and 0.67, respectively). As an example, dotted lines in Fig. 2b and 2c denote the model predictions regarding the proportion of upward perception in Experiments 2 and 3. That is, the generalization capability of the model was demonstrated. The predicted average increase rate of the upward perception in Experiment 2 was 0.13/sec (experimental data 0.17/s), and that in Experiment 3 was 0.30/sec (0.26/s).

## DISCUSSION

Many previous studies have provided circumstantial and/or indirect support to the existence of iterative computations in motion perception. For example, Lorenceau *et al.* [10] studied direction discrimination for lines moving obliquely relative to their orientation. They found that subjects initially perceived the direction perpendicular to the orientation of the lines. The perceived motion direction then gradually shifted to the actual direction. Their results can be equally well explained by their three-box model or the iterative model. Because Pack and Born [14] used the same aperture motion stimulus as Lorenceau *et al.* [10] in their MT recording studies, they admitted that the two computations still cannot be discriminated.

Here, we consider whether any one-shot algorithm can explain our experimental results. Within the one-shot algorithm framework, the current aperture problem amounts to extracting the two endpoints and computing their real motion vectors. In order to accommodate temporal perceptual changes, different time constants need to be assumed for motion detection in 45° and 90° directions. Furthermore, in order to reconcile the effects of line length and aperture size on perception, a variable combination of the outputs from these two kinds of motion detectors must be assumed. Consequently, any competent one-shot algorithm will consist of at least three black boxes [10,21,22]. The first box is the motion detector only in the direction perpendicular to the line, which always says the motion direction is 45°. The second box is the motion detector for the two endpoints, which always says the motion direction is 90°. The third box is the combiner, which mixes these two outputs with weights proportional to reliabilities in filter selection models [21,22]. It might thus be reasonable to assume that the second box is more time-consuming than the first box, and that its output gradually builds up over time while the output from the first box is immediate. If the combiner were to intelligently mix the two outputs by considering visible and occluded line lengths, three-box models could *qualitatively* explain the temporal increase of the upward perception, and its dependence on the gap size and aperture diameter in Experiment 1. However, the preceding stimulus in Experiment 2 excites only the first box. Furthermore, the preceding stimulus in Experiment 3 excites neither the first box nor the second box. Therefore, no three-box models, and accordingly no competent one-shot algorithm, can reproduce the results of Experiments 2 and 3.

## CONCLUSION

It has been extremely difficult to demonstrate the iterative process in vision with only one approach among psychology, physiology, and modeling. In the present study, we combined modeling with psychophysical experiments, and provided more decisive evidence of iterative computation. Our iterative model may be neurally implemented in the following way. The same neuron encodes both motion information and binding information using different information carriers. For example, according to Singer and Gray [23], spike synchrony might be used to encode binding information, whereas the firing rate can be used for motion encoding. While motion information propagates through bidirectional horizontal connections, spike synchrony carrying binding information also propagates using the same horizontal connections. However, the bidirectional gating mechanism for firing rate and spike synchrony is still unknown. Recordings of V1 and/or MT neural responses to our stimuli may reveal these possible neural implementations.

## REFERENCES

1. Allach H. *Psychol Forsch* **20**, 280–325 (1935). (English translation in *Perception* **25**, 1317–3167 (1996)).
2. Rolls ET and Tovee MJ. *Proc R Soc Lond B Biol Sci* **257**, 9–15 (1994).
3. Thorpe S, Fize D and Marlot C. *Nature* **381**, 520–522 (1996).
4. Fellemann DJ and Van Essen DC. *Cerebr Cortex* **1**, 1–47 (1991).
5. Mumford D. *Biol Cybern* **66**, 241–251 (1992).
6. Kawato M, Hayakawa H and Inui T. *Network Comput Neural Syst* **4**, 415–422 (1993).
7. Gilbert CD and Wiesel TN. *J Neurosci* **3**, 1116–1133 (1983).
8. Nakayama K and Silverman GH. *Vision Res* **28**, 739–746 (1988).
9. Yo C and Wilson HR. *Vision Res* **32**, 135–147 (1992).
10. Lorenceau J, Shiffrar M and Wells N. *Vision Res* **33**, 1207–1217 (1993).
11. Watanabe T and Cole R. *Vision Res* **35**, 2853–2861 (1995).
12. Lamme VAF, Supèr H and Spekreijse H. *Curr Opin Neurobiol* **8**, 529–535 (1998).
13. Optican LM and Richmond BJ. *J Neurophysiol* **57**, 162–178 (1987).
14. Pack CC and Born RT. *Nature* **409**, 1040–1042 (2001).
15. Ben-Av M and Shiffrar M. *Vision Res* **35**, 2889–2895 (1995).
16. Marr D. *Vision*. San Francisco: Freeman; 1980.

17. Hildreth H. *The Measurement of Visual Motion*. Cambridge, MA: MIT Press; 1984.
18. Poggio T, Torre V and Koch C. *Nature* **317**, 314–319 (1985).
19. Grossberg S and Mingolla E. *Psychol Rev* **92**, 173–211 (1985).
20. Koechlin E, Anton JL and Burnod Y. *Biol Cybern* **80**, 25–44 (1999).
21. Nowlan SJ and Sejnowski TJ. *J Opt Soc Am Ser A* **11**, 3177–3200 (1994).
22. Snowden RJ and Verstraten FAJ. *Trends Cogn Sci* **3**, 369–377 (1999).
23. Singer W and Gray CM. *Annu Rev Neurosci* **18**, 555–586 (1995).

Acknowledgements: We wish to thank S. Nishida, F. E. Pollick, S. Shimojo and F. A. J. Verstraten for their comments on the manuscript. This work was supported by the Human Frontier Science Projects, and the Special Coordination Fund to Brain Science.