A Bayesian Framework for Crowding Effect

Zhenbo Cheng[1,3], Wenfeng Chen[2], Tian Ran[2], Zhidong Deng[1], Xiaolan Fu[2]

1. State Key Laboratory on Intelligent Technology and Systems, Tsinghua National La oratory for Information Science and Technology, Department of Computer Science, Tsinghua University, Beijing 100084, China
E-mail: michael@tsinghua.edu.cn

2. State Key Laboratory of Brain and Cognitive Science Institute of Psychology, Beijing 100101, China
E-mail: chenwf@psych.ac.cn

3. Information and Engineering College, Zhejiang University of Technology, Hangzhou,310014, China

Abstract: In crowding, neighboring distractors impair the visual perception of a presented ta get. We study influences by the configuration of distractors on the bias to perceive the orientation of a target. Our results show that: (a) when distractors are similar to each other but different from target, crowding is decreased; (b) when distractors form a subjective contour, crowding is also reduced. These results illustrate that crowding is weak whenever the target stands out from the context and strong when the target is grouped into the context as a part of a global percept. In addition, we show how a Bayesian model, based on the principle of spatial resolution of attention that is modulated by the large size of receptive fields, can account for the behavioral data.

Key Words: Crowding Effect, Bayesian Model, Generative Model, Inferential Model

1 INTRODUCTION

In crowding effect, surrounding distractors can impair the observer's ability to perceive the target stimulus[1]. It also impairs the discrimination of the target stimulus attributes such as orientation and contrast due to the influence of the distractors[2]. The perception of a target is largely influenced by contextual input that is made up of distractors. Previous studies have shown that crowding is reduced when the distractors group together[3, 4] or the target is different from the distractors[2]. Our purpose in the study is to test how distractors configuration affects crowding and to provide a computational model to account for the key results of the experiments.

In the experiments, the human observers were informed to identify the gap orientation of the central Varin-type figure [5] (target) by pressing four buttons (left/right/up/down)(see Figure 1a). Another four Varin-type figures (distractors) were arranged in a circle around the target. The observers were informed that they should make judgments according to the target alone regardless of the distractors. The mean percent correct responses in different experimental conditions were measured. Our results show that: (a) when distractors are similar to each other but different from target, crowding is decreased; (b) when distractors form a subjective contour, crowding is also reduced.

Bayesian methods have been proven successful in modeling tasks of human sensory processing[6, 7, 8]. These methods also have been used to build probabilistic models for selective attention[9, 10, 11]. In this study, we provide a computational model to show the close relationship between attention and crowding. In the framework of a model of the Bayesian process, our data suggest that the large size of receptive fields can modulate the spatial resolution of attention.

This work is supported by National Nature Science Foundation under Grant No. 90820305 and No. 60775040, and National Key Basic Research and Development Program of China under Grant No. 2006CB303101.

2 EXPERIMENTS: CROWDING EFFECTS

In the experiments, observers viewed the stimuli binocularly from a distance of 60 cm. The stimuli were a configuration of Varin-type figures (see Figure1b, c) that measured $1.2 * 1.2$ cm ($1.1 * 1.1^{\circ}$). They consisted of a central Varin-type target and four irrelevant Varin-type distractors surrounding the target. The gap orientation of the target was randomly chosen from one of the four directions: down, up, left, and right. The stimuli were presented randomly either to the left or right from the fixation point. The center to center distance of target and fixation (eccentricity) was about 5.5 cm that measured 5.8°. The distance between the distractors and the target (measured from centre to centre) was about 1.74° (0.3 relative to the target eccentricity).

There were three types of Varin-type shapes: four-line Varin arc shape, two-line Varin shape with arcs closely spaced, two-line Varin shape with arcs widely spaced. Stimulus presentation, response recording and experimental procedure were controlled by the E-Prime 1.2 [12]. The display device was a 17" ($32 * 24$cm) LENOVO monitor with $1024 * 768$ resolutions and the vertical refresh rate was 75 Hz. Twelve undergraduate/graduate students (age 21-26 years old) participated in this experiments for a small payment. All had normal or corrected-to-normal vision and normal color vision. The naive observers were used to minimize any systematic bias.

The similarity of shape between target and distractors and

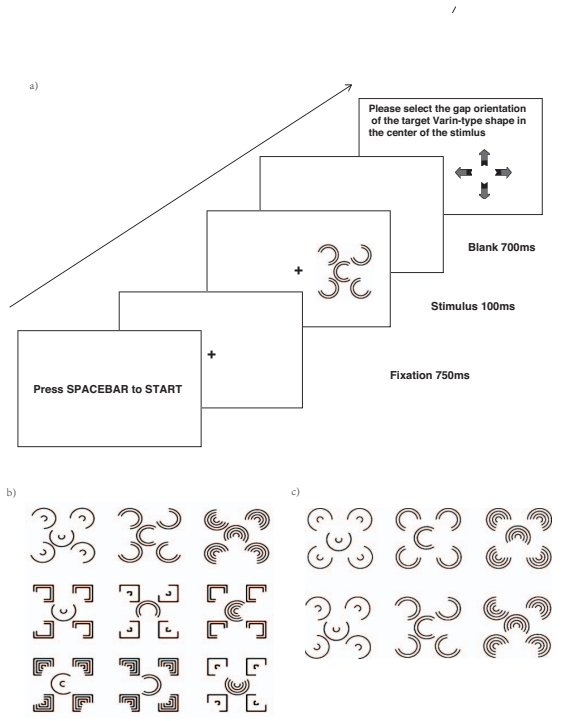

Figure 1: a)The observers initiated each trial by pressing SPACE BAR under the instruction on the screen. Then they fixated a small black cross, which was permanently presented on the screen during the trial. After a delay of 750 ms, a configuration of Varin-type figures was presented for 100 ms to either the left or right of the fixation. Following this, a blank screen was presented for 700ms. The observers had to identify the gap orientation of the central (target) Varin and click on the left/right/up/down arrow on the screen using the mouse to indicate their answer. There was no feedback about the accuracy of responses. b) Stimuli used in experiment 1. The shapes of target from left to right are: two-line Varin shape with arcs widely spaced, two-line Varin shape with arcs closely spaced, four-line Varin shape. The first row shows that the target and distractors have the same shape. The shape between target and distractors is different in the second and the third rows. c) Stimuli used in experiment 2. The upper row contains geometrically aligned and the subjective contour. The lower row does not contain the subjective contour.

the arrangement of the distractors made up of two experimental conditions: compatible trials (the same Varin-type shapes between target and distractors) and incompatible trials (the different Varin-type shapes between target and distractors or distractors were arranged to form the subjective contour). In the experiment 1, the orientation of target and distractors was unpredictably chosen on each compatible/incompatible trial. On each compatible trial of the experiment 2, the gap orientation of each distractors was also randomly chosen. On each incompatible trial of the experiment 2, the distractors were arranged to form a subjective contour.

In the experiments, the compatible and incompatible trials were run in two blocks of trials. There was also a baseline block in which the distractors was not presented. The order of blocks with different conditions was counterbalanced. Each block consisted of 120 experimental trials and 12 learning trials, where the three type of target Varin-type shapes appeared one third respectively. All blocks were finished within 30 minutes.

In the two experiments, data from 2.43% trials, where response times were deviated from the mean by three times the standard deviation or more (i.e., outlier), were excluded from the analysis. The mean percent correct responses are shown in Figure 2. A repeated-measures analysis of variance (ANOVA) showed significant main effects of similarity between target and distractors (Experiment 1 : $F(2, 22) = 62.12$, $p < .001$; Experiment 2 : $F(2, 22) = 30.677$, $p < .001$). Post-hoc comparison revealed that the accuracy of baseline conditions were higher than that of incompatible conditions (Experiment 1 : $p = .016$; Experiment 2 : $p = .0001$), and both the accuracy of incompatible and baseline conditions were higher than that of compatible conditions (Experiment 1 : $p < 0.001$; Experiment 2 : $p < 0.007$). There were no significant main effect of target type (Experiment 1 : $F(2, 22) = 1.399$, $p = 0.268$; Experiment 2 : $F(2, 24) = 0.506$, $p = 0.609$), and interaction (Experiment 1 : $F(4, 44) = 0.595$, $p = 0.668$; Experiment 2 : $F(4, 48) = 0.682$, $p = 0.608$). The values of F and p are the statistical outcomes when using ANOVA to analyze the experimental data.

3 THE BAYESIAN MODEL ON CROWDING

We define perception as a statistical problem in which the observer performs inference about the value of the target based on noisy sensory evidences created by a generative model. The generative model captures the general features of our task and the inference model infers the value of the target based on Bayesian methods. Both the generative and inference models are designed using probabilistic variables and the relationships that capture the uncertainties inherent from the generative process.

Figure 3 shows the structure of the generative and inference models. In the model, the three input units, representing the distractors and the target, are assumed to generate six input units which are belong to two groups, with small and large receptive fields. Based on the stimulus x, the generative model generates the noisy sensory measurements y. Given the measurements, the observer can infer the value of the target using the inference model based on the Bayesian methods.

3.1 The generative model

Let a stimulus pattern be made up of an array of three stimuli, x_1, x_2, x_3. The stimulus, x_2, is the target, and the other stimuli are the distractors. On each trial, the x_2 can be either $+1$ or -1. Following the setting of our experiment, we assume that the distractors are identical and they can be the same or different sign from the target, and the value of distractors can be either $+0.5$ or -0.5. In baseline conditions, the value of distractors is set to 0. In the compatible

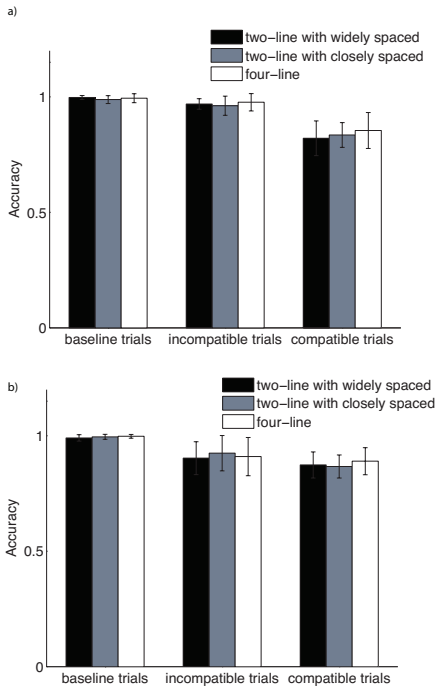

Figure 2: Results of the experiments. a) Experiment 1: the mean percent correct responses in three experimental conditions. b) Experiment 2: the mean percent correct responses in three experimental conditions. Error bars indicate standard errors of the means.

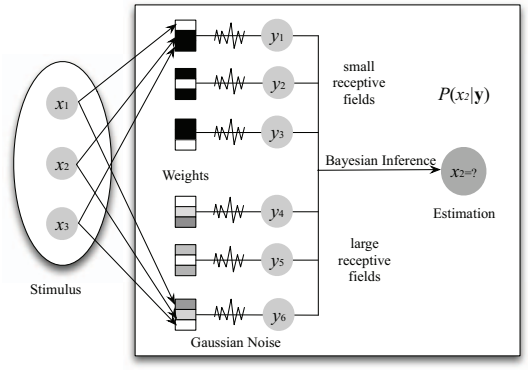

Figure 3: In the model, the three input units, representing the distractors and the target, are assumed to generate six input units which are belong to two groups, with small and large receptive fields. The weights are expressed by the Hinton diagrams to indicate how the receptive fields are represented. Based on the stimulus x, the observer generates the noisy sensory measurements y. Given the measurements, the observer infer the value of the target using the Bayesian methods.

conditions, the distractors has the same sign with the target, whereas the sign of target is different from that of the distractors on the incompatible trials.

To make simple, we assume that the three stimuli, $\mathbf{x}=[x_1, x_2, x_3]$, will evoke six populations of neurons with activities, $\mathbf{y}_t=[y_1(t), y_2(t), \ldots, y_6(t)]$. We consider that the three units $[y_1(t), y_2(t), y_3(t)]$ are with small receptive field, and another units $[y_4(t), y_5(t), y_6(t)]$ are with large receptive field. Since the small receptive field can be nicely confined to the visual stimulus, the units with small receptive field have only one inputs. On the other hand, the weight of each units with large receptive field comes from all three stimuli. All weights are positive, and equal or less than 0.3. The matrix of weights are given as

$$\mathbf{w} = \begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.3 \\ 0.3 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.2 \\ 0.1 & 0.2 & 0.3 \end{bmatrix}, \tag{1}$$

where each row of the weight matrix represents one of the six populations and each column of the weight matrix represents one of the stimuli.

Given the stimuli $\mathbf{x}$ on each trial, the six populations of neurons $\mathbf{y}_t = [y_1(t), y_2(t), \ldots, y_6(t)]$will be evoked as a Gaussian fashion:

$$\begin{aligned} p(\mathbf{y}_t|\mathbf{x}) &= p(y_1(t)|\mathbf{x})p(y_2(t)|\mathbf{x})\ldots p(y_6(t)|\mathbf{x}) \\ &= N(\mu_1, \sigma_1^2)N(\mu_2, \sigma_2^2)\ldots N(\mu_6, \sigma_6^2)\,. \end{aligned} \tag{2}$$

$N(\mu, \sigma^2)$ denotes a Gaussian probability distribution with mean μ and standard deviation σ. Since each y_i depend on all the three stimuli, we calculate the mean of populations as $\mu = [\mu_1, \mu_2, \ldots, \mu_6] = \mathbf{wx}$.

When the distractors are consistent or easily grouped with the target (compatible case), the distractors and the target will be prone to build up a visual object of Gestalt (global percept) and the observer should make more effort to segment the target from the distractors. In this case, the distractors have the effect of adding so much extra noise to the units with large receptive fields compared with their signal about the target. This implies that the distractors will exert more impairment on the perception of the target. If the target can't be grouped with the distractors (incompatible case), the distractors and target will be difficult to make up of a visual object of Gestalt and the observer should segment the target from the distractors with less effort. In this case, only the small receptive fields will be useful, and so the distractor will exert little impairment.

We argue that, in the compatible case, the units with the large receptive fields are assumed to have high variance. This assumption captures the relatively uselessness of these large receptive fields in the compatible case. Conversely, in the incompatible case, the units with the large receptive fields are assumed to have low variance. In addition, since the distractor stimuli around the target don't prefer neither sign of the target, the mean values of the units with the large receptive fields are unchanged.

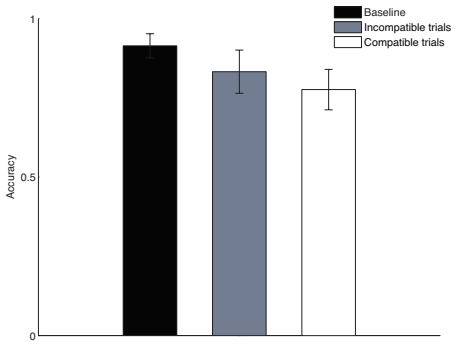

Figure 4: The mean percent correct responses in three experimental conditions. The bars show the mean percent correct responses as a function of different experimental conditions. Error bars indicate standard errors of the means.

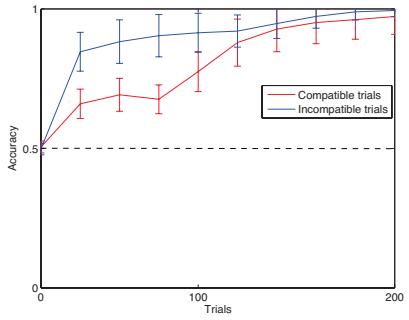

Figure 5: Accuracy versus trials (reaction time) in our model. The curves are the mean of accuracy over 50 calculations. Error bars are standard errors.

We also assume that successive observations $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t, \ldots$ are generated from $\mathbf{x}$ in an independent and identical case, that is

$$p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t|\mathbf{x}) = p(\mathbf{y}_1|\mathbf{x})p(\mathbf{y}_2|\mathbf{x}) \ldots p(\mathbf{y}_t|\mathbf{x}) \ , \quad (3)$$

where $\mathbf{y}_t = [y_1(t), y_2(t), \ldots, y_6(t)]$.

3.2 The inference model

Given the successive observations $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t$, the inference model will infer the value of the target x_2 from the marginal probability $P(x_2|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t)$. The marginal probability can be calculated from the posterior distribution as follows

$$P(x_2|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t) = \int_{x_1} \int_{x_3} P(\mathbf{x}|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t) dx_1 dx_3. \quad (4)$$

The posterior distribution $P(\mathbf{x}|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t)$ is the observer's belief about the value of stimuli after the variables $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t$ are observed. According to Bayes's rule, the posterior distribution is a function of the observer's belief at the previous time point, $P(\mathbf{x}|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{t-1})$, and the latest input $\mathbf{y}_t$, and is given as

$$P(\mathbf{x}|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t) = \frac{p(\mathbf{y}_t|\mathbf{x})P(\mathbf{x}|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{t-1})}{\sum_{\mathbf{x}'} p(\mathbf{y}_t|\mathbf{x}')P(\mathbf{x}'|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{t-1})}. \quad (5)$$

Given the initialized value, the posterior distribution can be calculated with a iterative process according to the Eq. (5). We assume that the prior probability of a trial being $x_2 = +1$ or $x_2 = -1$, before seeing any inputs should be equal to 0.5. The marginal probability $P(x_2 = -1|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t)$ is just $1 - P(x_2 = +1|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t)$. To make a perceptual decision, we can compare each of these two marginal probabilities against a decision threshold, p_{th}, and report that the target is $+1$ if $P(x_2 = +1|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t) > p_{th}$, report that the target is -1 if $P(x_2 = -1|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t) > p_{th}$, or continue observing otherwise.

The inference model is quite similar to a variant of the sequential probability ratio test (SPRT)[13]. The SPRT seems consistent with the performance of humans and animals in two-alternative decision tasks[14]. Furthermore, there are some computational neural models to implement the SPRT[15].

4 RESULTS

Figure 4 shows the results of inference based on the model. For each of the experimental conditions, figure 4 shows the the mean percent correct responses as a function of different experimental conditions. As in the data in our experiments, crowding is weak whenever the target stands out from the context (incompatible trials) and strong when the target is taken as the part of the context (compatible trials). Here, the threshold on the probability (p_{th}) is 0.9 and the variance of units with small receptive fields is 1. In compatible trials, the variance of units with large receptive fields is 8. The variance of units with large receptive fields is 2 in incompatible trials. The number of trials is 10 for each experimental condition.

Our model also provides a prediction that the accuracy will increase as the extension of the reaction time. The reaction time is the number of trials while the model finish a identifying task. The curves showing accuracy as a function of reaction time is plotted in Figure 5. They show that if the accuracy rates are the same, the model require fewer number of trials in the incompatible conditions than the compatible conditions.

5 DISCUSSION

In this paper, our experimental results have shown that the strength of crowding can be modulated by the shape or spatial configuration of distractors. When the shape of target is the same one of distractors, the crowding is increased, and when the spatial configuration of distractors make observer to form a subjective contour, the crowding is reduced. Strong crowding occurs when the target is grouped with the distractors so that they form a coherent pattern together.

We hypothesize that target-distractor segmentation in crowding is analogous to target-context segregation in Gestalt. When distractors can be grouped as a whole or when they are similar to each other but different from

the target, the target can be distinguished from distractors. However, grouping target and distractors together by Gestalt principles may interfere with target-distractor separation.

Then we have presented a computational model for crowding that captures the key results in our experiments. We assume that the perception is a statistical problem in which the observer performs inference about the value of the target based on noisy sensory evidences created by a generative model. We have suggested that the variance of the units with large receptive fields can be changed according to the arrangement or displacement of distractors. Our simulation results have shown that the change of the variance of the units with large receptive fields can result in the different identifying accuracy for the target stimulus. Note that the results happens automatically through inference, there is no explicit attentional control signal in our model, only inference and marginalization.

The assumptions of the units with the large receptive fields with high variance in the compatible case is rather simple. However, the assumptions conform to the hypothesis of Intriligator and Cavanagh about crowding[16]. They argued that crowding was an effect of insufficient spatial resolution of attention that was modulated by the large size of receptive fields. Further, the receptive fields of neurons in inferotemporal cortex are indeed large[17].

Our model is related to a previous Bayesian-inspired model of the attentional load theory[11]. This model also assumed that the receptive fields of units have a range of sizes. However, Dayan's model was to account for the reasons that inputs should automatically be attenuated to the extent that they do not bear on a task. Under our normative framework, it is appropriate to explain why grouping target and distractors together may interfere with target-distractor separation.

Although we demonstrated in this work how our model accounts for a core set of experimental data on the Crowding task, we leave for future work a rich set of additional findings, such as how the brain implement the Bayesian computations, and how the brain change the variance of units with large receptive fields. There are two main computational principles for these problems. One is that attentional selection consists of bottom-up sensory inputs, which can modulate the variance of units with large receptive fields. The subjective contour may be encoded in primary visual cortex[18]. Another possible hypothesis is that attentional selection comes from top-down information. It would be interesting to design neural model to verify these hypotheses.

6 CONCLUSIONS

In this study, we not only provide a novel explanation for crowding, but also suggest a computational model for the explanation. Our hypothesis is that crowding is weak whenever the target stands out from the context and strong when the target is grouped into the context as a part of a global percept. We have suggested that the attentional control signal of the target can be modulated by the units with large receptive fields. Based on some simple assumptions, the attentional control signal of the target in our model is the results of statistical inference problem based on Bayesian methods.

REFERENCES

[1] H. Bouma, Interaction effects in parafoveal letter recognition, Nature, Vol.226, No.5241, 177-178, 1970.

[2] T. P. Saarela, B. Sayim, G. Westheimer, and M. H. Herzog, Global stimulus conguration modulates crowding. J Vis, Vol.9, No.2, 5.1-11, 2009.

[3] T. Livne and D. Sagi, Conguration influence on crowding, Journal of Vision, Vol.7, No.2, 2007.

[4] L. Renninger and P. Verghese, Orientation discrimination in the periphery depends on the context, J. Vis., Vol.7, No.9, 585-585, 2007

[5] D. Varin, Fenomeni di contrasto e diffusion cromatica nell organizzazione spaziale del campo percettivo, Rivista di Psicologia, Vol.65, 101-128, 1971.

[6] M. O. Ernst and M. S. Banks, Humans integrate visual and haptic information in a statistically optimal fashion, Nature, Vol.415, No.6870, 429-433, 2002.

[7] Y. Weiss, E. P. Simoncelli, and E. H. Adelson, Motion illusions as optimal percepts, Nat Neurosci, Vol. 5, No. 6, 598-604, 2002.

[8] A. A. Stocker and E. P. Simoncelli, Noise characteristics and prior expectations in human visual speed perception, Nat Neurosci, Vol.9, No.4, 578-585, 2006.

[9] P. Dayan, R. Zemel, and L. GCNU, Statistical models and sensory attention, in Articial Neural Networks, 1999. ICANN 99. Ninth International Conference on (Conf. Publ. No. 470), Vol.2, 1999.

[10] A. J. Yu, P. Dayan, and J. D. Cohen, Dynamics of attentional selection under conflict: toward a rational bayesian account, J Exp Psychol Hum Percept Perform, Vol.35, No.3, 700-717, 2009.

[11] P. Dayan, Load and attentional bayes, in Advances in Neural Information Processing Systems 21, D. Koller and D. Schuurmans and Y. Bengio and L. Bottou, 369-376, 2009.

[12] W. Schneider, A. Eschman, and A. Zuccolotto, E-prime users guide, Pittsburgh, PA: Psychology Software Tools, 2002.

[13] A. Wald, Sequential analysis. New York: Wiley, 1947

[14] R. Ratcliff, P. L. Smith, A comparison of sequential sampling models for two-choice reaction time. Psychological Review, Vol.111, 333-346, 2004

[15] J. I. Gold, M. N. Shadlen, Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. Neuron, Vol.36, 299-308, 2002

[16] J. Intriligator and P. Cavanagh, The spatial resolution of visual attention, Cognitive Psychology, Vol.43, No.3, 171-216, 2001.

[17] C. G. Gross, D. B. Bender, and C. E. Rocha-Miranda, Visual receptive fields of neurons in inferotemporal cortex of the monkey, Vol.166, No.3910, 1303-1306, 1969

[18] T. S. Lee and M. Nguyen, Dynamics of subjective contour formation in the early visual cortex, Proc. Natl. Acad. Sci., Vol. 98, No. 4, 1907-1911, 2001