From: AAAI-99 Proceedings. Copyright © 1999, AAAI (www.aaai.org). All rights reserved.

A Neural Network Model of Dynamically Fluctuating Perception of Necker Cube as well as Dot Patterns

Hiroaki Kudo†, Tsuyoshi Yamamura††, Noboru Ohnishi†, Shin Kobayashi‡, and Noboru Sugie‡

†Graduate School of Engineering, Nagoya University, Nagoya 464-8603, Japan
††Faculty of Information Science and Technology, Aichi Prefectural University, Aichi 480-1198, Japan
‡Faculty of Science and Technology, Meijo University, Nagoya 468-8502, Japan
kudo@nuie.nagoya-u.ac.jp, yamamura@ist.aichi-pu.ac.jp, sugie@meijo-u.ac.jp

Abstract

The mechanism underlying perceptual grouping of visual stimuli is not static, but dynamic. In this paper, the dynamical grouping process is implemented with a neural network model consisting of an array of (hyper)columns suggested by Hubel & Wiesel, where intracolumnar inhibition and intercolumnar facilitation are incorporated. The model was applied successfully to figures consisting of a set of dots yielding either of two ways of groupings from time to time due to neural fluctuations and fatigue. Then the model was extended to introduce dependency on fixation points as well as neural fluctuations and fatigue. Then, it was applied to the Necker Cube. The model output from time to time either of two ways of 3D interpretations depending on the fixation points.

Introduction

Perceptual grouping plays an essential role in segmenting objects in the scene and recognizing each of them. Gestalt psychologists have proposed that there are several factors underlying the grouping: they are factor of proximity, factor of similarity, factor of smooth continuation, and so on. Recently, computer implementations of the grouping processes have been reported (Stevens 1978; Hiratsuka, Ohnishi, and Sugie 1992). However, the mechanism underlying perceptual grouping of visual stimuli is not static; but dynamic as in Marroquin pattern (**Fig.1**) (Marr 1982). The dynamical aspect of grouping seems to reflect the flexible nature of human visual information processing to deal with ambiguous patterns. However, it has not been studied seriously. In this paper, the dynamical grouping process is implemented with a neural network model consisting of a 2D array of hypercolumns suggested by Hubel & Wiesel (1977), where intracolumnar inhibition and intercolumnar facilitation are incorporated. The model was applied successfully to figures which consist of a set of dots yielding either of two ways of groupings from time to time due to neural fluctuations and fatigue. Moreover, the model was able to interpret line drawings, a grouping at a higher level; it was applied to the Necker cube (**Fig.2**). It also exhibited dependence on fixation points about the Necker cube. The model output either of two 3D interpretations reflecting fixation point dependence as reported by Kawabata *et al.* (1978).

Copyright ©1999, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

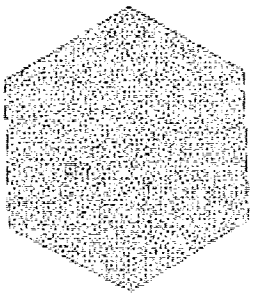


Figure 1: Marroquin pattern.

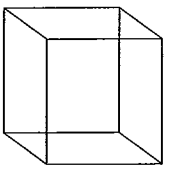


Figure 2: Necker Cube.

Neural Network Model

This model is based on the neural network model consisting of hypercolumnar structure. It has been used to explain the early visual process such as retinal rivalry (Sugie 1982). We extend it to deal with highly ambiguous figures of more than two interpretations.

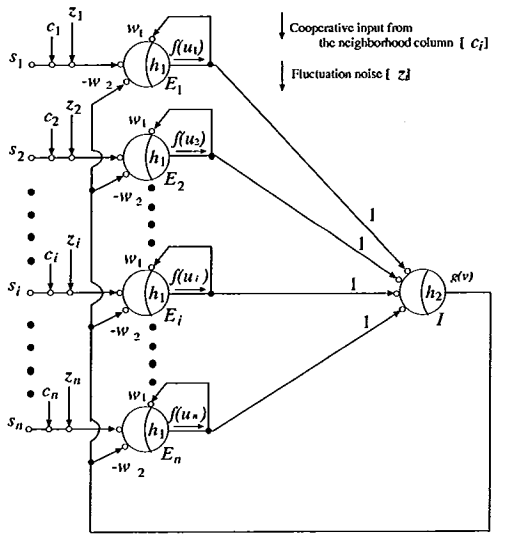


Figure 3: Generalized flip-flop.

Basic Structure – hypercolumn –

In the human visual system, visual stimuli are received first by the retina. Then, outputs of retinal ganglion cells are sent to the V1 of the cerebral visual cortex via the lateral geniculate body. The V1 consists of a 2D array of hypercolumns, each of which corresponds to a specific local visual area preserving the topological relationship in the retina.

This columnar structure is modeled with network structure which has intracolumnar inhibitory as well as intercolumnar facilitator connection. **Fig.3** shows the network structure of one column. The units of $E_1, E_2, \cdots, E_i, \cdots, E_n$ correspond to neurons, each of which is selective to specific stimulus orientation of its own. Each neural output $f(u_i)$ is weighted (w_1) and feedback to itself, directly. It is also fed to an inhibitory neuron I with a unit weight, whereas the output of neuron I is fed to each of E_i's with weight w_2. These two inputs ensure that at the steady state only one of E_i's becomes activated or the winner depending on the inputs s_i's, c_i's, and z_i's (Amari 1978). Thus, we call the network shown in Fig.3 as generalized flip-flop. Now three kinds of inputs to each of E_i's are explained.

1. visual stimulus via the retina and the lateral geniculate body$\cdots$ $[s_i]$
2. an inhibitory input from I summing up all the outputs of E_i's $\cdots$ $[g(v)]$
3. a facilitator input from the orientation sensitive unit in the neighboring hypercolumn (intercolumnar facilitation)$\cdots$ $[c_i]$

This element of columnar structure (or, network) (Fig.3) is corresponding to the structure which locates at one point on visual cortex. As shown in **Fig.4**, a

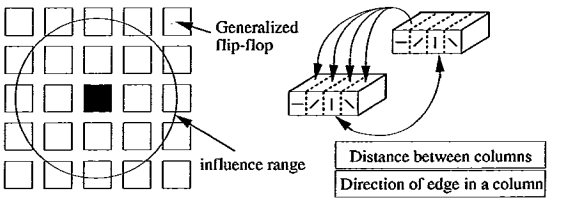


Figure 4: Intercolumnar facilitation network.

2D array of generalized flip-flops are arranged so that each of which corresponds to its own specific local visual field preserving the topological relation. Further, neighboring generalized flip-flops are facilitatively interconnected via c_i as described in 3 above. c_i is dependent on both the mutual distance and mutual difference in orientations between each pair of generalized flip-flops. In Fig.3, h_1 and h_2 are the thresholds of the corresponding units. u_i's and v are the inner potentials of the corresponding units. Here the inner potential means the total sum of weighted inputs and threshold. Thus u_i and v satisfy the following differential equations (1) and (2), respectively. The time constant for u_i is 1, and that for v is τ.

$$\dot{u}_i = -u_i + w_1 f(u_i) - w_2 g(v) - h_1 + s_i \quad (1)$$

$$\tau \dot{v} = -v + \sum_{i=1}^{n} f(u_i) - h_2 \quad (2)$$

As already mentioned, $f(u_i)$ and $g(v)$ are the excitatory and inhibitory outputs of E_i and I, respectively. Each E_i has its own excitatory feedback collateral with weight w_1, which plays the role of keeping its activity high once it is excited. The inhibitory unit I serves for keeping E_i's from saturation as well as for deciding the winner among E_i's. We define $f(u_i)$ and $g(v)$ as equations (3) and (4), respectively.

$$f(u) = \begin{cases} 1 & (u > 0) \\ 0 & (u \leq 0) \end{cases} \quad (3)$$

$$g(v) = \begin{cases} v & (v > 0) \\ 0 & (v \leq 0) \end{cases} \quad (4)$$

Further, we assume that $0 \leq s_i \leq s_{max}$.

As analyzed by Amari (1978), the generalized flip-flops behaves as follows:

1. When all the inputs s_i's do not attain the level of s_{min}, none of E_i's are excited.
2. When there are plural s_i's exceeding s_{min}, E_i corresponding to the maximal s_i becomes excited or the winner, while the others are suppressed and are not excited at the steady state.
3. Once one E_i becomes the winner, it remains excited even after the input s_i is set to 0. Only by resetting the whole system E_i is set to off.

In the model shown in Fig.4, certain orientation selective units corresponding to the positions where visual stimuli do exist will be excited. Moreover due to intercolumnar facilitation, some other units may be excited as well even if there is not any corresponding visual stimuli. This facilitative effect may reflect the factor of smooth continuation complementing gaps along a smooth line. The winner-take-all nature of the model may correspond to only one interpretation of a stimulus figure at one time.

This kind of intercolumnar and intracolumnar interaction scheme has been proposed for the elucidation of self-organization mechanism of binocular stereopsis (Sawada and Sugie 1982). The generalized flip-flop is based on the system of winner-take-all. So, it is an appropriate model of binocular rivalry (Sugie 1982).

Extension to Deal with Dynamical Grouping

In the situation of dynamical grouping, humans do not always have a single stable percept, but have one of plural percepts competing one another from time to time. For example, when we look at **Fig.5**(a), the percepts alternate between (b) and (c). To deal with such phenomena, we introduce the following three factors into the model.

fluctuation of neural activities (z_i) As each neuron is under the influence of noises contained in the external stimulus as well as intrinsic fluctuations in cellular activities, the neural activities fluctuate from time to time. Therefore, only one neuron becomes the winner at the steady state, even if each input s_i to E_i is one and the same. z_i represents such noises.

neural fatigue Once a unit E_i continues to fire, the threshold h_1 of which becomes higher resulting in difficulty in firing on. This is the neural fatigue, which may cause the change in the winner unit. The detail of fatigue process will be described later.

fixation point When we look at stimulus figures, we usually change the fixation point from time to time, which causes the change in the retinal image.

Considering these factors, the processing in the model proceeds as shown in **Fig.6**.

First a fixation point is decided. Then the retinal image is formed accordingly. Next, at the stage of feature extraction, the position, and allowable orientations formed by grouping neighboring dots in the stimulus figure at each local region are extracted. Then at the stage of determining the optimal orientation, the inputs to each E_i is determined, which reflect the factors of perceptual grouping (proximity or mutual distance, similarity in shape or orientation , and smooth continuation).

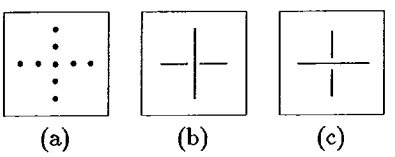


Figure 5: Example of dynamical grouping.

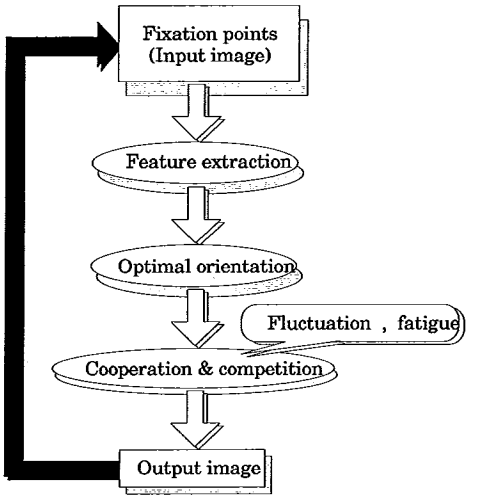


Figure 6: Model of dynamical grouping process.

At the stage of cooperation and competition, the proposed neural network decides the winner among orientation selective E_i's in each hypercolumn. In order to realize the neural fluctuations, we introduce one noise-generating neuron corresponding to each E_i. The output of the noise-generating neuron represents z_i. As for the neural fatigue, it is realized through the change in h_1 as already stated, the detail of which will be described in the next section.

At the stage of output image, the simulated percept at each instant is displayed, where E_i yielding the maximal output among other E_i's at each location at each instant is assumed to correspond to the perceived grouping (connection) between the dot of concern and one of the neighboring dots. Since the percepts may change from time to time except at the equilibrium state, the output images are generated and displayed at each instant.

As for the factor of change in fixation points, we assume the factor as restarting of the whole process. Thus the factor is introduced at the first of all the stages.

Simulation Studies

We implemented and simulated the proposed scheme. For simplicity, we assume the processes from the fixation point through the optimal orientation as a preprocessing prepared beforehand. The results of the preprocessing are given as inputs to the succeeding process of cooperation and competition shown in Figs.3

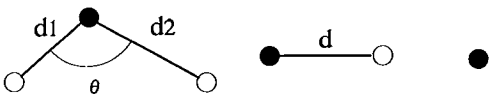


Figure 7: Variations of connection between elements.

and 4. One general flip-flop is assigned to each dot in the stimulus figure. Each of E_i's corresponds to a connection between the dot of concern and one of neighboring dots with the orientation specific to E_i. We assume the following three types of connections for each dot of concern as shown in **Fig.7**, where the filled circle represents the dot of concern and open circles represent neighboring dots.

1. the case with two connected neighbors, where the distances to them are $d1$ and $d2$ and the angle formed by two line segments are θ.
2. the case with one connected neighbor, the distance to which is d.
3. the case without any connection.

We set the initial value s_i for E_i considering perceptual grouping factors. That is, in the first case above, we assign the value is larger for smaller distance ($d1$ and $d2$) and for θ closer to 180 degrees. In the second case, the initial value is larger for smaller d, while in the third case the initial value is set to zero.

For those inputs described above, each generalized flip-flop outputs such a connection corresponding to E_i with the highest activity among others.

As for the neural fatigue, we changed the threshold h_1 of each E_i according to Eqs.(5) and (6) below. The changing profiles of h_1 with respect to time are shown in **Figs.8** and **9**. The former corresponds to the case where the inner potential is positive and the unit is firing, while the latter is applied in the case where the inner potential is negative and the unit not active.

$$h_1 = h_{1\_min} + \frac{1}{1+\exp\frac{-(t-h_\tau)}{T}} \cdot (h_{1\_max} - h_{1\_min}) \quad (5)$$

where t,h_τ,T,h_{1_min},h_{1_max} designate the time, the time constant, the increase rate of threshold, the minimum value of the threshold, and the maximum value of the threshold, respectively.

$$h_1 = h_{1\_min} + \exp(-at) \cdot (h_{1\_max} - h_{1\_min}) \quad (6)$$

where a means the time constant.

In **Fig10**, we show two sample visual stimuli for simulation studies. Since these stimuli are simple figures, we experimented without considering the factor of change in fixation points. As a measure of processing time unit, we introduce a prescribed time unit. It is the time to obtain one output image after giving the inputs to generalized flip-flops. We observed output

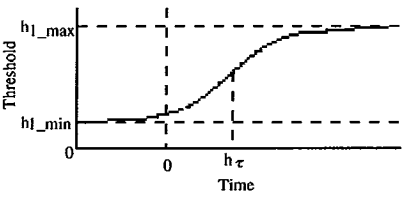


Figure 8: Change in threshold [1] due to fatigue.

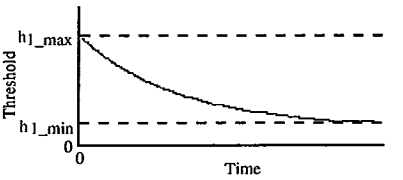


Figure 9: Recovery of threshold [2] from fatigue.

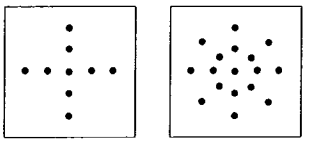
images during 6,000 time units (Fig.10(a)) and 3,000 time units (Fig.10(b)), respectively.

In **Fig.11**, we show the outputs of the neural network for the visual stimulus, Fig.10(a), in the course of the time after the stimulus presentation. We can see that at $t = 500$ a complete vertical line is perceived, while the horizontal line is interrupted in the middle. At $t = 3000$, however, the percept is just in the contrary. The horizontal line is complete, while the vertical line is interrupted in the middle. At $t = 1000, 2000, 3500$, and 5000, some of the dots are left alone without forming any connection with other dots. Such fluctuating percepts similar to human perception are caused primarily by the neural fluctuations and fatigue.

In **Fig.12**, we show the outputs of the neural network in the course of time after the presentation of the visual stimulus, Fig.10(b). When humans observe the stimulus, humans perceive in the course of the time either fragments of circles of various radii (arcs), or line segments of various orientations and lengths at various positions. The outputs shown in Fig.12 may be considered as simulating and the fluctuating human perception stated above. The simulated percepts in Fig.12 are mostly fragmentary circles or short lines. However, human percepts tend to prefer more complete circles or lines. This difference should be studied further by taking into consideration of more global measure of 'Gestaltian Praegnanz'.

Applying to Ambiguous Figures

When humans observe the figure shown in Fig.2 (Necker cube), human percepts alternate from time to time between either of two 3D interpretations shown in **Fig.13**(a) and (b), where hidden (only partly) lines are removed for convenience. Note that only one of the two interpretations are exclusively perceived at each in-



(a)Pattern 1 (b)Pattern 2

Figure 10: Sample stimulus 1,2.

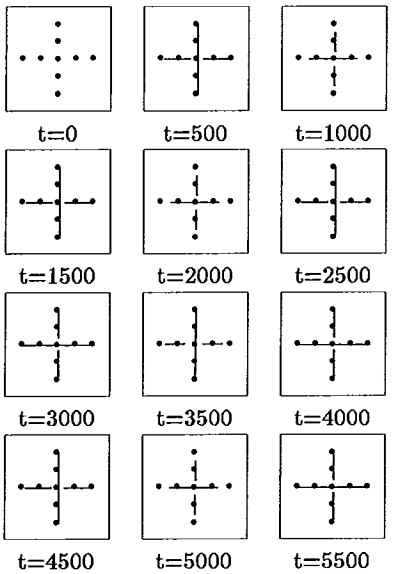


Figure 11: Time course of perception (pattern 1).

stant. We consider this kind of higher visual processes can be realized using the scheme of 2D array of generalized flip-flops shown in Figs.3 and 4. Each vertex of the visual stimulus in Fig.2 corresponds to one generalized flip-flop. Each E_i corresponds to one of the two interpretations at the assigned vertex. Facilitative intercolumnar interactions are introduced between a pair of E_i's with the same interpretation in generalized flip-flops for each pair of adjacent vertices.

According to Kawabata *et al.*, which of the two interpretations are preferred is remarkably dependent on which vertex the subjects look at (Kawabata, Yamagami, and Noaki 1978). It is reported that when the fixation point is around A in **Fig.14**, the subjects tend to perceive the interpretation 1 more often. When the fixation point is around A', however, the subjects tend to perceive the interpretation 2 more often. So we set s_i's dependent on the fixation point as follows, where let s_1 stands for the interpretation 1 and s_2 the interpretation 2. Let Pi denote a fixation point. Then s_1 is set to be proportional to the distance between A and A' divided by that between Pi and A plus α, where α is a positive constant to keep the value from diverging

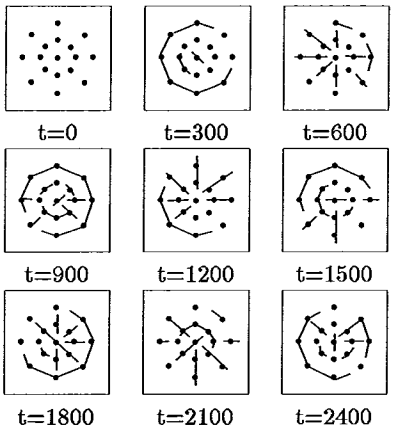


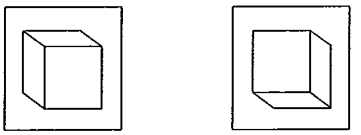
Figure 12: Time course of perception (pattern 2).

for Pi very close to A. Similar s_2 is set to be proportional to the distance between A and A' divided by that between Pi and A' plus α. Thus at $P1$, $P2$, $P3$, and $P4$, the ratios between s_1 and s_2 were set to 5:1, 2:1, 1:2, and 1:5, respectively.

The simulation studies were conducted while the fixation points were shifted from $P1$ through $P4$ successively. At each Pi, the duration of fixation was 1,500 time units. As an example of the inner potential change, the result on the vertex A is shown in **Fig.15**. In **Fig.16** are shown some of the snapshot percepts of the model. The duration ratios between the interpretation 1 and 2 are summarized in **Table 1**. It is obvious that the percepts fluctuate from time to time. We can see that at $P1$ the interpretation 1 is overwhelmingly dominant. As the fixation points were shifted towards $P4$, the interpretation 2 becomes dominant gradually. However, even at $P4$, the dominance of interpretation 2 over 1 is not so overwhelming as that of 1 over 2 at $P1$. Thus as a whole the duration ratio of interpretation 1 and 2 during 0 — 6,000 time units is 53.4 : 46.6 preferring the interpretation 1. To see the hysteresis effects due to the shifts in fixation points, we carried out simulation studies for the cases of no shifts in fixation points. Each fixation started from the same initial condition. The result is shown in **Table 2**. The change in interpretations ($P1$ versus $P4$, and $P2$ versus $P3$) is almost symmetrical with respect to two interpretations. Thus the results in Table 1 can be understood as reflecting the hysteresis effects. These behaviors of the model coincides well with the findings by Kawabata *et al.*

Concluding Remarks

Grouping process is dynamic in nature. However, in most cases only one kind of grouping is possible. Therefore the perception is stable and grouping seems static only apparently. In some pathological cases, the



(a)Interpretation 1 (b)Interpretation 2

Figure 13: 3D interpretation of Necker cube.

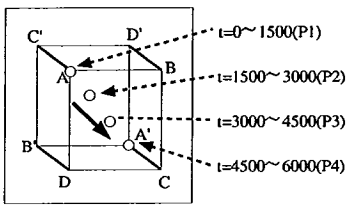


Figure 14: Movement of fixation points.

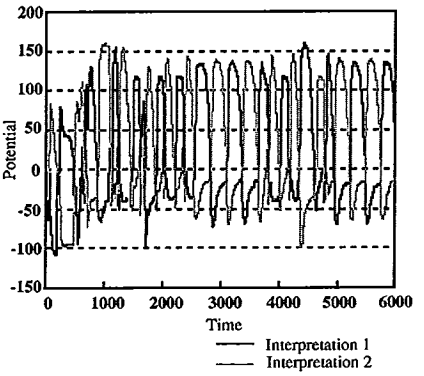


Figure 15: Potential change on vertex A.

dynamical aspect of grouping shows up. The present article may be the first serious attempt to simulate some of the typical phenomena, related to figures which consist of dots and the Necker cube. The Gestalt concept of similarity should be extended to include 3D interpretations as in the Necker cube.

Acknowledgements

The research was supported in part by the Grant-in-aid of the Ministry of Education for "Quantum Information Theoretical Approach to Life Science".

References

Stevens, K. A. 1978. Computation of Locally Parallel Structure. *Biol. Cybernetics* 29:19–28.

Hiratsuka, S., Ohnishi, N. and Sugie, N. 1992. Extracting Global Structures Using Perceptual Grouping.*IEICE Trans. D-II* J76-D-II:74–83. (in Japanese)

Marr, D. 1982. *Vision.* New York.: W.H.Freeman

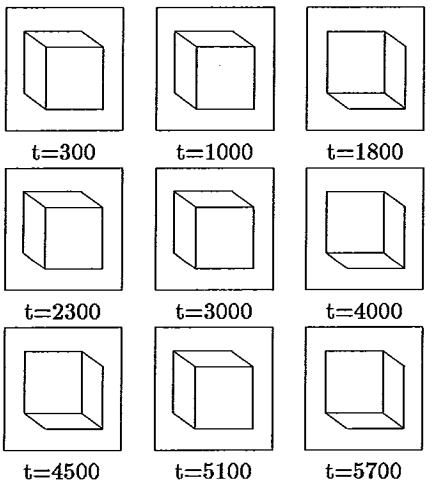


Figure 16: Time course of 3D interpretation.

Table 1: Duration ratio of two 3D interpretations, where movements of fixation points exist.

t [time unit]	Interpretation 1 [%]	Interpretation 2 [%]
0~1500	97.8	2.1
1500~3000	42.3	57.7
3000~4500	51.6	48.4
4500~6000	33.9	66.1
0~6000	53.4	46.6

Table 2: Duration ratio of two 3D interpretations, where there was no movements of fixation points.

	Duration ratio [%]			
fixation point	P1	P2	P3	P4
interpretation 1	98.7	51.7	44.9	5.4
interpretation 2	1.3	48.3	55.1	94.6

Hubel, D. H., Wiesel, T. H. 1977, Functional Architecture of Macaque Monkey Visual Cortex, Proc. Roy. Soc. Lond. B. 198. 1–59

Kawabata, N.; Yamagami, K.; and Noaki, M. 1978. Visual Fixation Points and Depth Perception. *Vision Research* 18.:853–854

Sugie, N. 1982. Neural Model of Brightness Perception and Retinal Rivalry in Binocular Vision. *Biol. Cybernetics* 43.:13–21

Amari, S. 1978. *Mathematical Principles of Neural Networks.* Tokyo. :Sangyo Tosho (in Japanese)

Sawada, R. and Sugie, N. (Amari, S., and Arbib, M. A. eds.) 1982. Self-organization of Neural Nets with Competitive and Cooperative Interaction. *Lecture Notes in Biomathematics 45, Competition and Cooperation in Neural Nets.* :238–247