MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

A model of visual working memory : Examining effect of intervening stimulus on network dynamics

Shigeru Kubota, Tsuyoshi Okamoto,
Kosuke Hamaguchi, and Kazuyuki Aihara

METR 2004–40 August 2004

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.i.u-tokyo.ac.jp/mi/mi-e.htm

A model of visual working memory : Examining effect of intervening stimulus on network dynamics

Shigeru Kubota[a,*], Tsuyoshi Okamoto[b],
Kosuke Hamaguchi[c], and Kazuyuki Aihara[a,b]

[a] Institute of Industrial Science, University of Tokyo,
4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan

[b] ERATO Aihara Complexity Modelling Project, JST,
45-18 Oyama, Shibuya-ku, Tokyo, 151-0065 Japan

[c]RIKEN Brain Science Institute,
2-1 Hirosawa, Wako-shi, Saitama, 351-0198 Japan

*kubota@sat.t.u-tokyo.ac.jp

August 2004

Abstract

In a delayed matching-to-sample task with visual test items, the stimulus-specific delay period activity can survive the intervening stimulus presentations in the prefrontal cortex, but not in the inferior temporal cortex. We construct a two-layer neural network model with interconnections to examine the effects of interaction in the cortical areas on the neural responses in the task. By analyzing the pattern dynamics of the network, we show that the interplay between excitation patterns in the cortices plays a key role in memory-related neuronal dynamics.

1 Introduction

In electrophysiological experiments using animals performing a working memory task, stimulus-specific persistent activity is observed in various brain areas such as the prefrontal cortex (PFC) [11, 13, 23], the inferior temporal cortex (ITC) [3, 22, 24, 25], and the posterior parietal cortex [5]. The observations that the PFC seems to contain especially many memory cells [28] and the results of brain imaging studies [7] indicate that PFC plays a preeminent role in holding information in memory. On the other hand, there exist also evidences that performance in working memory tasks depends on activity in cortical regions other than the PFC, such as the ITC [12] and the parietal cortex [26]. Thus, it is important to explore the effects of interaction between the PFC and the other regions that underlies memory-related neuronal dynamics as discussed in Wang [28].

A number of theoretical studies have shown that the neural activity can be sustained in the absence of external input stimuli by several possible ways such as the mutual excitation among neurons [1, 2, 4], bistable membrane dynamics of single neurons [14, 20], and successive feedfoward neural connections of 'synfire chain' type [8, 15, 16]. (See [10, 28] for a review.) The modeling studies by Camperi and Wang [2] and Compte and Brunel [4] have reproduced the visuospatial delayed response experiment [11] by the model network with spatially homogeneous interconnections. They have also examined the robustness of the memory of the cue stimulus against the intervening stimulus during a delay period (distractor). However, most of such modeling studies have been concerned with the activity in a localized network such as in the PFC, and the mnemonic neural dynamics induced by the interplay between the cortical areas has not been sufficiently understood.

Miller and colleagues recorded from PFC and ITC neurons of the monkeys during a delayed matching-to-sample (DMS) task [22–25]. In their experiments, a sequence of visual test stimuli were used in each trial. The first stimulus was the sample, which was followed by one or more test stimuli. The final test stimulus is the matching stimulus that is the same as the sample. The animals were required to ignore the intervening (nonmatching) stimuli between the first and final stimuli that differ from the sample and to respond only to the matching stimulus. The high levels of activation during delay period (delay activity) in PFC were compared with that in ITC [23,25]. The results demonstrate that the sample-selective delay activity in PFC is preserved throughout the trial but that in ITC is disrupted after the intervening stimulus presentations.

There exist several kinds of neurons in both PFC and ITC observed in the DMS task based on the classifications, such as whether the response during the stimulus presentation is stimulus-selective or the delay activity is sample-selective [23,25]. Miller et al. [23] investigated whether such response properties can occur together in the same neurons in PFC and found the tendency that cells with stimulus-selective responses also show sample-selective delay activity. The same tendency is also observed in ITC cells [3].

Miller et al. [25] analyzed data of the delay activity after the intervening stimulus in ITC neurons and pointed out the possibility that the information about the immediately preceding stimulus is included in the activity level of the ITC neurons, which implies that some ITC neurons can show delay activity correlated with the preceding intervening stimulus. However, the experiment of Chelazzi et al. [3] clearly shows that the ITC neurons with stimulus-selective responses do not have delay activity following the stimulus that has no behavioral relevance, such as the intervening stimulus in the DMS task. Although the behavior of the ITC cells in the delay period is not clear as discussed in Miller et al. [23], we hypothesize based on these observations that some proportion of ITC neurons with stimulus-selective responses can show sample-selective delay activity following the sample stimulus, but have no delay activity following the intervening stimuli.

This paper aims to present a two-layer neural network model consisting of higher and lower layers that model PFC and ITC cells and to examine the possible effects of interaction between the two cortical areas on the neuronal dynamics in the DMS task. We examine how the delay activity of neurons that have stimulus-selective responses is modulated by the intervening stimulus presentations. In particular, we show that the model network can qualitatively reproduce the following properties of such neurons. 1) In the PFC, sample-selective delay activity is preserved throughout the trial. 2) In the ITC, sample-selective delay activity is shown after the sample, but delay activity does not occur after the intervening stimuli. The mechanism of pattern dynamics underlying responses to the intervening stimulus presentations is explored by applying the results of the theoretical analysis [1, 18] to the model.

2 Neural network model

We consider the two groups of neurons with stimulus-selective responses that constitutes the PFC and ITC as shown in Fig. 1(a). We make the following two assumptions with respect to the neural connections among neurons in the two cortical areas:

Assumption I : Every two neurons constituting the same cortical area are connected via excitatory interconnections if the information encoded by firing of the neurons are sufficiently similar, and via inhibitory interconnections otherwise.

Assumption II : Every two neurons that belong to the different cortical areas are connected via excitatory interconnections only if information encoded by firing of the neurons are sufficiently similar.

Note that the assumptions of the existence of excitatory and inhibitory interconnections between neurons encoding similar and dissimilar information matches the Hebbian covariance rule about synaptic plasticity [6, 17, 27].

Now we construct a neural network model by describing the above two assumptions mathematically. As shown in Fig. 1(b), we rearrange each group of neurons in Fig. 1(a) in line such that the neurons encoding similar information are located at nearby positions. The x-axis is set in the direction parallel to the lines of neurons. We refer to the layers corresponding to the PFC and ITC neurons in Fig. 1(b) as layer H (the higher layer) and layer L (the lower layer), respectively.

Let us assume that the neurons are densely distributed in each layer in Fig. 1(b). Then, by taking a continuum limit, the dynamics of cells in each layer can be described as follows:

$$\begin{aligned}\tau_{\mathsf{I}}\frac{\partial u_{\mathsf{I}}(x,t)}{\partial t} =& - u_{\mathsf{I}}(x,t) + \int_{-\mathsf{l}}^{\mathsf{l}} w_{\mathsf{I}}(x-y) f[u_{\mathsf{I}}(y,t)]dy \\ &+ \int_{-\mathsf{l}}^{\mathsf{l}} w_{\mathsf{IJ}}(x-y) f[u_{\mathsf{J}}(y,t)]dy + S_{\mathsf{I}}(x,t) - T_{\mathsf{I}}, \qquad (1)\\ &(I,J) = (H,L) \ \text{ or } \ (L,H), \qquad (2)\end{aligned}$$

where H and L denote the indexes for layer H and layer L; $u_{\mathsf{I}}(x,t)$ is the average membrane potential of neurons at position x in layer I at time t; $f(u)$ is the output function that determines the firing rate of neurons dependent on the membrane potential; $w_{\mathsf{I}}(x-y)$ is the function for the intralayer connectivity that represents the intensity of connections from neurons at position y in layer I to ones at position x in layer I; $w_{\mathsf{IJ}}(x-y)$ is the function for the interlayer connectivity that represents the intensity of connections from neurons at position y in layer J to ones at position x in layer I; τ_{I} is the time constant for neurons in layer I; $S_{\mathsf{I}}(x,t)$ is the external inputs applied to neurons at position x in layer I at time t; and $-T_{\mathsf{I}}$ $(T_{\mathsf{I}} > 0)$ is the resting potential of neurons in layer I. The second term in the right-hand side of Eq. (1) describes the synaptic inputs from neurons in the same layer and the third term represents those from neurons in the different layer. For the sake of simplicity, $f(u)$ is assumed to be the step-function satisfying $f(u) = 1$ for $u > 0$ and $f(u) = 0$ for $u \leq 0$. Since a neuron fires at a constant firing rate only when its membrane potential is above the threshold, we define $\{x|u_{\mathsf{I}}(x) > 0\}$ to be the excited region in layer I. We refer to the state where the excited region is a finite interval (x_1, x_2) as the local excitation with the length of $x_2 - x_1$.

Since the arrangement of neurons in each layer reflects the analogy of encoded information, Assumptions I and II can be simply described as the following equations by using the parameters for connectivity $K_{\mathsf{I}}^{\mathsf{exc}}$, $K_{\mathsf{I}}^{\mathsf{inh}}$ $(K_{\mathsf{I}}^{\mathsf{exc}} > K_{\mathsf{I}}^{\mathsf{inh}})$, K_{IJ}, σ_{I}, and σ_{IJ}:

$$w_{\mathsf{I}}(x) = K_{\mathsf{I}}^{\mathsf{exc}} \exp[-x^2/(2\sigma_{\mathsf{I}}^2)] - K_{\mathsf{I}}^{\mathsf{inh}}, \qquad (3)$$

$$w_{\mathsf{IJ}}(x) = K_{\mathsf{IJ}} \exp[-x^2/(2\sigma_{\mathsf{IJ}}^2)]. \qquad (4)$$

We can consider that the test stimuli used in the DMS task is applied to layer L corresponding to the ITC, the lower cortical area than the

PFC. Thus, this input stimuli is modeled as follows:

$$S_{\mathsf{L}}(x,t) = \begin{cases} A_{\mathsf{s}} \exp[-(x - x_{\mathsf{s}}^{\mathsf{k}})^2/(2\sigma_{\mathsf{s}}^2)], & \text{if } \ t \in [t_{\mathsf{s}}^{\mathsf{k}},\ t_{\mathsf{s}}^{\mathsf{k}} + \Delta T_{\mathsf{s}}]\ (k = 1, ..., N_{\mathsf{s}}), \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where N_{s} is the number of the test stimuli presented in a trial; A_{s} and σ_{s} are the intensity and the width of each input stimulus; $x_{\mathsf{s}}^{\mathsf{k}}$ is the center position of the kth stimulus; $t_{\mathsf{s}}^{\mathsf{k}}$ is the time of onset of the kth stimulus; and ΔT_{s} is the duration of each stimulus presentation. We assume $t_{\mathsf{s}}^{\mathsf{k}} \equiv (k-1)(\Delta T_{\mathsf{s}} + \Delta T_{\mathsf{d}})$ by using the time interval of the delay period ΔT_{d}. Note that, since A_{s} and σ_{s} take constant values independent of k, the intensity and the width of all input stimuli are the same throughout a trial.

Here let us consider the network function of memory erasure at the end of a trial. This function is linked to the high-level cognition process that controls the behavioral sequence in the task, so that we assume that the PFC plays a critical role in attaining the function. Therefore, to realize the function in the model network, we apply spatially uniform inhibitory inputs only to the neurons in layer H as follows:

$$S_{\mathsf{H}}(x,t) = \begin{cases} -A_{\mathsf{r}}, & \text{if } \ t \in [t_{\mathsf{r}}, t_{\mathsf{r}} + \Delta T_{\mathsf{r}}], \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where A_{r} denotes the intensity of the inhibitory inputs. t_{r} and ΔT_{r} are the start time and the duration of the inputs.

3 Simulation

By adjusting the parameter values for neural connections, we reproduced the neural dynamics in the DMS task (Fig.2). These parameter values are $K_{\mathsf{H}}^{\mathsf{exc}} = 9$, $K_{\mathsf{H}}^{\mathsf{inh}} = 3.6$, $K_{\mathsf{L}}^{\mathsf{exc}} = 4.5$, $K_{\mathsf{L}}^{\mathsf{inh}} = 1.8$, $K_{\mathsf{HL}} = 5$, $K_{\mathsf{LH}} = 1$, and $\sigma_{\mathsf{H}} = \sigma_{\mathsf{L}} = \sigma_{\mathsf{HL}} = \sigma_{\mathsf{LH}} = 2$, and the same values are used throughout the paper. This set of parameters implies that the intralayer connections are stronger in layer H than in layer L and that the top-down connections from layer H to layer L are weaker than the bottom-up connections from layer L to layer H (Fig. 1(b)).

The solid lines in Fig. 2(a) show the time course of the center positions of the four input stimuli applied to layer L. Their positions are $x_{\mathsf{s}}^1 = 0$, $x_{\mathsf{s}}^2 = 15$, $x_{\mathsf{s}}^3 = -10$, and $x_{\mathsf{s}}^4 = 0$. The first stimulus represents the sample, while the following two stimuli are the intervening (nonmatching) stimuli, so that their positions are different from the sample. The last stimulus is the matching one that is applied to the same position as the sample. The intensity and the width of the four input stimuli are $A_{\mathsf{s}} = 17$ and $\sigma_{\mathsf{s}} = 2$.

The duration of the stimulus presentation and the time interval of the delay are set as $\Delta T_{\mathsf{s}} = \Delta T_{\mathsf{d}} = 30$. Following the presentation of the four input stimuli to layer L, spatially uniform inhibitory inputs are applied to layer H. It is possible that animals erase memory of the sample so quickly after the presentation of the matching stimulus. But, to compare the delay period activity after the matching stimulus with that after the other stimuli, the time of onset of the inhibitory inputs is set as $t_{\mathsf{r}} = 240$, which is 30 time units after the termination of the matching stimulus. The intensity and the duration of the uniform inhibition to layer H are $A_{\mathsf{r}} = 15$ and $\Delta T_{\mathsf{r}} = 10$.

Figs. 2(b) and 2(c) show the spatio-temporal response patterns of the membrane potential (Fig. 2(b)) and the firing state (Fig. 2(c)) in layer H and layer L. Local excitations are elicited around $x = 0$ in the two layers in response to the sample stimulus. The excitations correlated with the sample are maintained in both layers during the delay following the sample. However, responses to the intervening stimuli are totally different in the two layers. In layer H, increased membrane potential by the intervening stimuli cannot cause the corresponding firing of neurons and the local excitation encoding the position of the sample stimulus is preserved. On the other hand, in layer L, the local excitation corresponding to the sample is disrupted by the intervening stimulus presentations, leading to the patterns in response to the two intervening stimuli, but the excitations disappear in the delay after the intervening stimuli. When the matching stimulus is applied, the excitation patterns occur again at the same position as the stimulus in both layers, which are maintained in the delay following the stimulus. Finally, by the uniform inhibitory inputs applied to layer H, the excitations in both layers disappear and the membrane potential of all neurons returns to the resting potential. Since the information about the sample stimulus is stored in the position of the local excitation in layer H until the matching stimulus appears, we can consider that the task succeeds from a viewpoint of information storage. The result shown in Fig. 2 matches the neuronal responses in the DMS task mentioned above.

Fig. 3 shows another simulation result corresponding to the case where the memory of the sample cannot be preserved and the task fails. The intensity of the input stimuli is $A_{\mathsf{S}} = 25$, which is larger than that used in Fig. 2, but all the other conditions are the same in the two simulations. Thus, the time course of the center positions of the input stimuli to layer L is represented by Fig. 2(a). Fig. 3(a) shows the spatio-temporal response patterns of the membrane potential in the two layers and Fig. 3(b) shows the corresponding spatio-temporal patterns of the firing state. In Fig. 3, the firing patterns change in response to the four input stimuli in both layers because of the large intensity of the inputs. Furthermore, the local excitations are maintained during the delay following each stimulus. The uniform inhibition to layer H erases the excitations in both layers just like Fig. 2.

Figs. 2 and 3 indicate that the effects of the intervening stimuli on the excitation pattern depends strongly on the stimulus intensity. Here, to examine the effects of intervening stimulus in detail, we perform simulations using only two input stimuli corresponding to the sample and the intervening stimuli (Fig. 4). The solid lines in Fig. 4(a) represent the time course of the center positions of the two inputs applied to layer L. We define Δx_{S} to be the distance between the center positions of the two inputs as shown in the figure. Fig. 4(b) shows the spatio-temporal firing pattern in layer I ($=H$ or L) in response to the input stimuli represented by Fig. 4(a). As shown in Fig. 4(b), Δx_{I} is defined as the distance between the position of the first stimulus and the center of the excited region in layer I at the end of the second stimulus presentation.

The relationship between Δx_{S} and Δx_{I} ($I = H$,L) was examined with changing the stimulus intensity A_{S}. Then, three types of response curves of Δx_{S} vs. Δx_{H} with Δx_{L} were found dependent on the stimulus intensity as shown in Figs. 4(c)-4(e). Note that we excluded the cases in which the stimulus intensity is too weak and delay activity cannot be found after the first stimulus. Figs. 4(c)-4(e) show typical examples for the three types of responses where the values of the stimulus intensity are sufficiently small ($A_{\mathsf{S}} = 10$), intermediate ($A_{\mathsf{S}} = 17$), and sufficiently large ($A_{\mathsf{S}} = 25$), respectively.

In cases where the input stimuli are sufficiently weak as shown in Fig. 4(c), Δx_{H} and Δx_{L} take almost the same values as that of Δx_{S} when the value of Δx_{S} is small. But the values of Δx_{H} and Δx_{L} converge to zero for sufficiently large values of Δx_{S}. These changes mean that the local excitations in the two layers move to the position of the intervening stimulus as far as the intervening stimulus is presented to the position near the sample, but that the excitation patterns are not affected by the intervening stimulus if the stimulus is applied to the position far from the sample. When the input stimuli are sufficiently strong as shown in Fig. 4(e), the excitation patterns change corresponding to the position of the intervening stimulus for all values of Δx_{S}. In all cases shown in Figs. 4(c) and 4(e), the local excitations in both layers are maintained during the delay period following the intervening stimulus at the same position as that at the end of the intervening stimulus presentation.

In cases where the stimulus intensity takes a intermediate value as shown in Fig. 4(d), the local excitations in the two layers move to the position of the intervening stimulus for the small values of Δx_{S}, just like Figs. 4(c) and 4(e), and the excitations in both layers are maintained through the delay following the intervening stimulus at the same position as that of the stimulus. On the other hand, for sufficiently large values of Δx_{S}, we can understand from Fig. 4(d) that the local excitation pattern in layer L changes in response to the intervening stimulus presentation, but that the excitation pattern in layer H is not affected by the stimulus. In this case, the excitation in layer H is maintained at the same position as that of the first stimulus during the delay after the intervening stimulus, whereas all the neurons in layer L become quiescent in the same delay, i.e., there is no delay activity after the intervening stimulus in layer L.

The pattern dynamics shown in Figs. 4(c)-4(e) can be summarized in the following three cases:

Case A : In cases where the intervening stimulus is applied to the position near the sample stimulus or the stimuli are sufficiently strong, the excitation patterns in the two layers change in response to the intervening stimulus and the excitations in both layers are maintained during the delay period after the intervening stimulus.

Case B : In cases where the sufficiently weak intervening stimulus is applied at the position far from the sample stimulus, the excitations in the two layers are not affected by the intervening stimulus presentation.

Case C : In cases where the intervening stimulus with intermediate intensity is applied to the position far from the sample stimulus, the excitation pattern in layer L changes corresponding to the intervening stimulus, but that in layer H is not affected. During the delay period after the intervening stimulus, the excitation in layer H is maintained, but that in layer L disappears.

The neural responses to the intervening stimulus shown in Fig. 2 corresponds to the dynamics in Case C and the responses in Fig. 3 matches Case A.

In the modeling studies of working memory for the localized network such as in the PFC [2, 4], the network behavior similar to Figs. 4(c) and 4(e) has been obtained. Thus, we can roughly understand that the patten dynamics in Case A and Case B can occur when the two layers with interconnections behave just like one layer. However, the pattern dynamics in Case C, which is obviously specific to the two-layer network, cannot be understood on the analogy of the one-layer network. All the three cases of pattern dynamics are explored in the following section by taking into account the effects of interaction between the two layers.

4 Mechanism of pattern dynamics

Here, we show an approximate but simple way to understand the mechanism underlying pattern dynamics in the model by applying the results of theoretical analysis for the network pattern formation [1, 18].

We consider the dynamics in the delay period for the cases where the centers of local excitations in the two layers lie on the same x-coordinate (Case I) and for the cases where the local excitations in the two layers are located at a great distance along the x-axis (Case II). The spatial configurations of the local excitations in the schematic diagrams shown in Figs. 5(a) and 5(c) agree with Case I and Case II, respectively.

For Case I, we define a_H and a_L as the length of the local excitations in layer H and layer L. Without loss of generality, we assume that the center positions of the local excitations in the two layers are at $x = 0$. Then, the excited region of layer I $(= H, L)$ is represented by the interval $(-a_I/2,\ a_I/2)$. Consider that the network is at a steady state with the local excitation patterns and define $\bar{u}_I(x)$ as the steady membrane potential distribution in layer I. Let $S_{IJ}(x)$ be the synaptic inputs applied to neurons at position x in layer I from neurons in layer $J(\neq I)$. Then, we have

$$S_{IJ}(x) = \int_{-a_J/2}^{a_J/2} w_{IJ}(x-y)dy. \tag{7}$$

Now let us introduce the two characteristic functions $Z_I(x) \equiv S_{IJ}(x/2)$ and $Y_I(x) \equiv T_I - \int_0^X w_I(x')dx'$ that are defined for $x > 0$. By considering $S_{IJ}(x)$ as the external inputs to layer I, we apply the theoretical results in Kubota et al. [18] to the dynamics in layer I. Then, we can easily find that the value of a_I is equal to the x-coordinate of an intersection point of $Z_I(x)$ with $Y_I(x)$ and that the network is stable only when the gradients of these curves at the intersection satisfy $dY_I(a_I)/dx > dZ_I(a_I)/dx$. (Note that $Z_I(x)$ corresponds to the "$a - \hat{S}$ curve" in the discussion about graphic analysis in Kubota et al. [18].)

We have numerically calculated $\bar{u}_I(x)$, $Z_I(x)$, and $Y_I(x)$ for $I = H$,L with $S_{HL}(x)$ and $S_{LH}(x)$ as shown in Figs. 6(a)-6(d) where the neural connections in the network are the same as those used in the previous section. By comparing Fig. 6(a) with Fig. 6(c) and Fig. 6(b) with Fig. 6(d), we can find that the length of the excited region a_I for $I = H$,L is equal to the x-coordinate of the intersection of $Z_I(x)$ with $Y_I(x)$ for the stable solution as expected from the theoretical result.

We can understand that the functions $Z_I(x)$ and $Y_I(x)$ represent the effects of the interlayer synaptic inputs and the intralayer connections from their definitions. The fact that the steady local excitation solutions exist corresponding to the intersections of these functions means that the solution is determined by the two effects: interaction between the excitations in the different layers and the mutual excitation and inhibition among neurons in the same layer.

In Case II, neurons in the excited region in one layer have hardly any effect on excited neurons in the other layer, which is expected from the connectivity function in Eq. (4). Thus, the interaction between the excitations in the two layers are lost, so that we can assume approximately that the excitation patterns in both layers behave independently. Therefore, we apply the results of the theoretical studies [1, 18] to each layer under the condition of $Z_{\mathsf{I}}(x) \equiv 0$ $(I = H, L)$. (Note that, under this assumption, each layer can be considered to be one-layer network separated from the other.) Since we find two intersections of $Y_{\mathsf{H}}(x)$ with $Z_{\mathsf{H}}(x)(\equiv 0)$ from Fig. 6(c), there exist steady local excitation solutions in layer H that have the same length as the x-coordinates of the two intersections and the solution with greater length becomes stable. On the other hand, as shown in Fig. 6(d), there is no intersection of $Y_{\mathsf{L}}(x)$ with $Z_{\mathsf{L}}(x)(\equiv 0)$, so that any steady local excitation solution does not exist in layer L.

We can understand from the above discussion that, during a delay period, the local excitation in layer H can exist independently of whether it interacts with the excitation in layer L, but that the local excitation in layer L can exist only when the center of the excitations in both layers are located at the same x-coordinate and the two excitations interact with each other. Therefore, the pattern dynamics in Case A - Case C can be explained as follows.

1. As shown in Fig. 5(a), the sample stimulus elicits local excitations at the same position in the two layers, so that the two excitations are sustained through the delay period following the stimulus under the influence of the interaction between them.

2. When the intervening stimulus is applied to the position near the sample in Case A, the excited region in layer L moves to the position of the intervening stimulus according to the basic property of the network that the local excitation shifts in the direction such that the stimulus intensity increases [1, 19]. The changing pattern in layer L leads to

the changing inputs to layer H, so that the excitation in layer H also moves to the position of the intervening stimulus. In cases where the sufficiently strong intervening stimulus is applied to the position far from the sample in Case A, the intervening stimulus overcomes the inhibitory effects that the excited regions exert on distant neurons [1] and local excitations are elicited at the same position as that of the intervening stimulus in both layers. Thus, in Case A, presentation of the intervening stimulus always produces local excitations at the same position in layer H and layer L as shown in Fig. 5(b). Hence, during the delay period following the intervening stimulus, the excitations in both layers are maintained by the effect of interaction between them.

3. In Case B, the weak intervening stimulus applied to the position far from the sample stimulus cannot overcome the inhibitory effect caused by the excitation in layer L. Therefore, the excitation patterns elicited by the sample shown in Fig. 5(a) are preserved after the intervening stimulus presentation.

4. In Case C, the intervening stimulus with intermediate intensity is given to the position apart from the sample. Then, in layer L, the inhibitory effect caused by the excited region corresponding to the sample is overcome and the local excitation pattern is elicited in response to the intervening stimulus. However, in layer H, this inhibitory effect cannot be overcome so that the excitation pattern is not changed by the intervening stimulus. Hence, by the intervening stimulus presentation, the two local excitations are separated at a distance along the x-axis as shown in Fig. 5(c), leading to the loss of interaction between the two excitations. The excitation in layer H can persist during the delay after the intervening stimulus since it can exist even in the absence of the interaction with the excitation in layer L. But, the local excitation in layer L cannot be sustained without interacting with the excitation in layer H. Therefore, as shown in Fig. 5(d), only the local excitation in layer H can be maintained during the delay after the intervening stimulus.

The effect of the spatially uniform inhibitory inputs to layer H for the memory erasure, which has been used in Figs. 2 and 3, can also be explained like the above discussion. After the inhibitory inputs directly turn off the activation of neurons in layer H, neurons in layer L lose the interaction with those in layer H, so that the layer L cannot sustain the excitation pattern. As a result, activity of all the neurons returns to the resting state.

The conditions for the occurrence of the pattern dynamics in Case C, which matches the neural responses in a successful trial of the DMS task, can be summarized as follows:

Condition I : The intervening stimulus is applied to the position sufficiently far from the sample stimulus with intermediate intensity such that only the excitation pattern in layer L is affected.

Condition II : In a delay period, the local excitation in layer H can be maintained independently of the activity pattern in layer L, but the local excitation in layer L cannot persist without interacting with that in layer H.

By the intervening stimulus presentation satisfying Condition I, the positions of the excitations in both layers become apart from each other, which eliminates the interaction between them. Then, by Condition II, the excitation in layer H are sustained during the following delay period, while that in layer L disappears.

Condition I means that the ITC neurons receive the visual stimulus of an appropriate level of intensity from the lower visual areas and that the information of the intervening stimulus is sufficiently different from that of the sample stimulus.

In cases where Condition II holds, the following relationship is required as shown in Figs. 6(c) and 6(d): 1) there exists a solution satisfying $Y_{\mathsf{H}}(x) = 0$, and 2) there does not exist a solution satisfying $Y_{\mathsf{L}}(x) = 0$. Hence, if we assume that the resting potential of neurons in the PFC and ITC are the same, i.e., $T_{\mathsf{H}} = T_{\mathsf{L}}$, then the inequality of $\int_0^{x^c_{\mathsf{L}}} w_{\mathsf{L}}(x)dx < \int_0^{x^c_{\mathsf{H}}} w_{\mathsf{H}}(x)dx$ holds with a parameter $x^c_{\mathsf{I}}(> 0)$ satisfying $w_{\mathsf{I}}(x^c_{\mathsf{I}}) = 0$. Note that the relation implies that the excitatory connections among neurons in the PFC is stronger and/or more extensive than in the ITC.

5 Discussion

In this study, we have constructed a two-layer neural network model that shows the neural responses in the PFC and ITC in the DMS task [23,25]. The mechanism underlying network pattern dynamics has been examined, and we have shown the conditions for the occurrence of the characteristic neural dynamics in the task.

In our model, the persistent delay activity in the PFC can exist independently of the activity of the ITC neurons, whereas the delay activity in the ITC neurons occur only in the presence of the feedback inputs from the firing neurons in the PFC that retain memory of the sample. This result adds a theoretical support to the idea that the PFC plays a central role in the working memory process [10]. The feedback inputs from the PFC to other brain structures are considered to work as bias signals that integrate the processing in multiple areas [21]. We have shown that, when the feedback input is terminated by applying inhibition to the PFC cells, the activation of the ITC cells is also turned off. Thus, it might be possible that, by cutting off the feedback inputs, the PFC resets the activity of other brain areas.

In the experiment with spatial version of the DMS task, the delay activity in the PFC is robust against presentation of the intervening stimulus [9], while the delay activity in the posterior parietal cortex is disturbed by the intervening stimulus [5]. These results are parallel to the observations of the delay activity in the PFC and ITC [23, 25], so that our model might be able to be applied to the dynamics of cells in the PFC and the posterior parietal cortex.

Our model shows the possibility that the responses of cells strongly depend on the characteristics of the intervening stimulus. As in Case A, if the intervening stimulus is very strong or the information of the intervening stimulus is similar to that of the sample, then the delay activity encoding the information of the intervening stimulus might occur in both PFC and ITC, which means the loss of memory of the sample. On the other hand, as in Case B, if the intervening stimulus is weak and dissimilar to the sample, then the delay activity correlated with the sample would be preserved in these cortical areas. These prospects match our experiences as follows. Suppose that, while you are memorizing some information (e.g., phone number), you receive another strong sensory input (e.g., an ambulance siren). Then, the memory might slip from your mind. If you catch some information similar to your holding item (e.g., when someone tells you phone number that differs from your memorizing number in only one digit), your memory would become confused.

The limitations in our model should be recognized. To perform the DMS task, multiple processes are required such as attending to the sample presentation, holding the memory against intervening stimuli, detecting the matching stimulus, and erasing the memory. The model has focused on the neural responses related to the memory storage and memory erasure in the task. But, the cells with various kinds of responses are observed in the task, e.g., cells with inhibitory responses to visual stimuli or fixation-related responses [23, 25], and it is possible that the different types of cells have different roles in the task and share such many functions. Thus, it should be examined in more comprehensive model how these different kinds of cells cooperate to attain the working memory function totally. Furthermore, there seem to exist several types of cells with different delay period activity in the ITC. As mentioned above, we consider that some ITC cells with stimulus-selective responses do not have persistent delay activity after the intervening stimulus presentation since the electrophysiological experiments indicate that the stimulus without behavioral relevance does not elicit delay activity in such ITC cells [3]. But, experimental data also show that some ITC cells might have delay activity correlated with the preceding intervening stimulus [25] and that other ITC cells have delay activity that predicts the coming stimulus [3]. Miller et al. [23] has observed “delay activity in IT cortex under all of the same conditions in which one would expect to find it in PF cortex - except after intervening stimuli”. This observation matches the dynamical property of ITC cells in our model very well, especially in the point that the delay activity in the ITC can exist only in the presence of the corresponding delay activity in the PFC. Although the exact behavior in the delay period of ITC cells and their roles are still controversial, we consider from all of these observations that our model network describes dynamics of some type of ITC cells that show stimulus-selective responses. More experimental data are required to elucidate the functional role of the complex delay period activity in ITC cells.

Actually, the delay activity observed in the experiment is often complicated and the firing rate can change during one delay period or between different delay intervals in one trial [23]. In this paper, the output function has been described as the step-function for the sake of mathematical simplicity. Although the simplification serves to elucidate the mechanism of the pattern dynamics by using the theoretical results of the network patten formation, the changing firing rates in real neurons cannot be expressed in this model. A more realistic neural element should be used in more advanced modeling studies, which makes it possible to compare the firing rate of the model neuron directly with the experimental data.

Acknowledgments

This study is partially supported by the Advanced and Innovational Research Program in Life Sciences from the Ministry of Education, Culture, Sports, Science, and Technology, the Japanese Government.

References

[1] S. Amari, Dynamics of pattern formation in lateral-inhibition type neural fields, Biol. Cybern. 27 (1977) 77-87.

[2] M. Camperi, X-J. Wang, A model of visuospatial working memory in prefrontal cortex: Recurrent network and cellular bistability, J. Comput. Neurosci. 5 (1998) 383-405.

[3] L. Chelazzi, J. Duncan, E. K. Miller, R. Desimone, Responses of neurons in inferior temporal cortex during memory-guided visual search, J. Neurophysiol. 80 (1998) 2918-2940.

[4] A. Compte, N. Brunel, P. S. Goldman-Rakic, X-J. Wang, Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model, Cerebral Cortex 10 (2000) 910-923.

[5] C. Constantinidis, M. A. Steinmetz, Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task, J. Neurophysiol. 76 (1996) 1352-1355.

[6] S. J. Cruikshank, N. M. Weinberger, Evidence for the hebbian hypothesis in experience-dependent physiological plasticity of neocortex: a critical review, Brain Res Brain Res Rev. 22 (1996) 191-228.

[7] M. D'Esposito, G. K. Aguirre, E. Zarahn, D. Ballard, R. K. Shin, J. Lease, Funtional MRI studies of spatial and nonspatial working memory, Cognit. Brain Res. 7 (1998) 1-13.

[8] M. Diesmann, M. O. Gewaltig, A. Aertsen, Stable propagation of synchronous spiking in cortical neural networks, Nature 402 (1999) 529-533.

[9] G. di Pellegrino, S. P. Wise, Visuospatial versus visuomotor activity in the premotor and prefrontal cortex of a primate, J. Neurosci. 13 (1993) 1227-1243.

[10] D. Durstewitz, J. K. Seamans, T. J. Sejnowski, Neurocomputational models of working memory, Nat. Neurosci. 3 (2000) 1184-1191.

[11] S. Funahashi, C. J. Bruce, P. S. Goldman-Rakic, Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex, J. Neurophysiol. 61 (1989) 331-349.

[12] D. Geffan, E. A. Murray, Monkeys (macaca fascicularis) with rhinal cortex ablations succeed in object discrimination learning despite 24-hr intertrial intervals and fail at matching to sample despite double sample presentations, Behav. Neurosci. 106 (1992) 30-38.

[13] P. S. Goldman-Rakic, Cellular basis of working memory, Neuron 14 (1995) 477-485.

[14] E. Guigon, B. Dorizzi, Y. Burnod, W. Schultz, Neural correlates of learning in the prefrontal cortex of the monkey: a predictive model. Cereb. Coretex 5 (1995) 135-147.

[15] K. Hamaguchi, K. Aihara, Quantitative information transfer through layers of spiking neurons connected by mexican-hat-type connectivity, Neurocomputing, 58-60 (2004) 85-90.

[16] K. Hamaguchi, M. Okada, M. Yamana, K. Aihara, Correlated firing in a feedforward network with mexican-hat type connectivity, submitted.

[17] D. O. Hebb, The organization of behavior, (Willey, New York, 1949).

[18] S. Kubota, K. Hamaguchi, K. Aihara, Theoretical analysis of local excitation pattern solutions in neural field model with general external inputs, submitted.

[19] C. R. Laing, A. Longtin, Noise-induced stabilization of bumps in systems with long-range spatial coupling, Physica D 160 (2001) 149-172.

[20] J. E. Lisman, M. A. P. Idiart, Storage of 7 ± 2 short-term memories in oscillatory subcycles, Science, 267 (1995) 1512-1515.

[21] E. K. Miller, J. D. Cohen, An integrative theory of prefrontal cortex function, Annu. Rev. Neurosci. 24 (2001) 167-202.

[22] E. K. Miller, R. Desimone, Parallel neuronal mechanisms for short-term memory, Science 263 (1994) 520-522.

[23] E. K. Miller, C. A. Erickson, R. Desimone, Neural mechanisms of visual working memory in prefrontal cortex of the macaque, J. Neurosci. 16 (1996) 5154-5167.

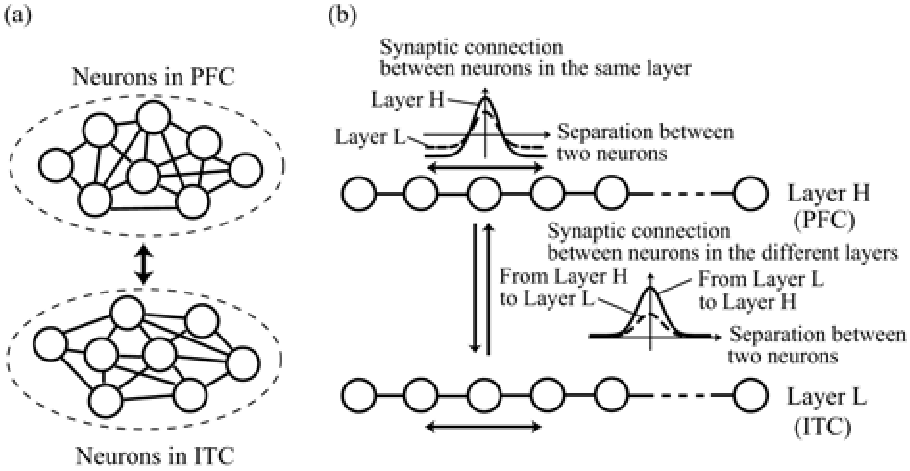
[24] E. K. Miller, L. Li, R. Desimone, A neural mechanism for working and recognition memory in inferior temporal cortex, Science 254 (1991) 1377-1379.

[25] E. K. Miller, L. Li, R. Desimone, Activity of neurons in anterior inferior temporal cortex during a short-term memory task, J. Neurosci. 13 (1993) 1460-1478.

[26] P. J. Olesen, H. Westerberg, T. Klingberg, Increased prefrotnal and parietal activity after training of working memory, Nat. Neurosci. 7 (2004) 75-79.

[27] P. K. Stanton, T. J. Sejnowski, Associative long-term depression in the hippocampus induced by hebbian covariance, Nature 339 (1989) 215-218.

[28] X-J. Wang, Synaptic reverberation underlying mnemonic persistent activity, Trends in Neurosci. 24 (2001) 455-463.



(a)
Neurons in PFC
Neurons in ITC
(b)
Synaptic connection between neurons in the same layer
Layer H
Layer L
Separation between two neurons
Layer H (PFC)
Synaptic connection between neurons in the different layers
From Layer H to Layer L
From Layer L to Layer H
Separation between two neurons
Layer L (ITC)

Fig. 1. Architecture of the two-layer neural network model. (a) The two groups of neurons corresponding to the PFC and ITC. (b) The model network where the neurons of each group in (a) have been rearranged in line such that the neurons encoding similar information are located at nearby positions.

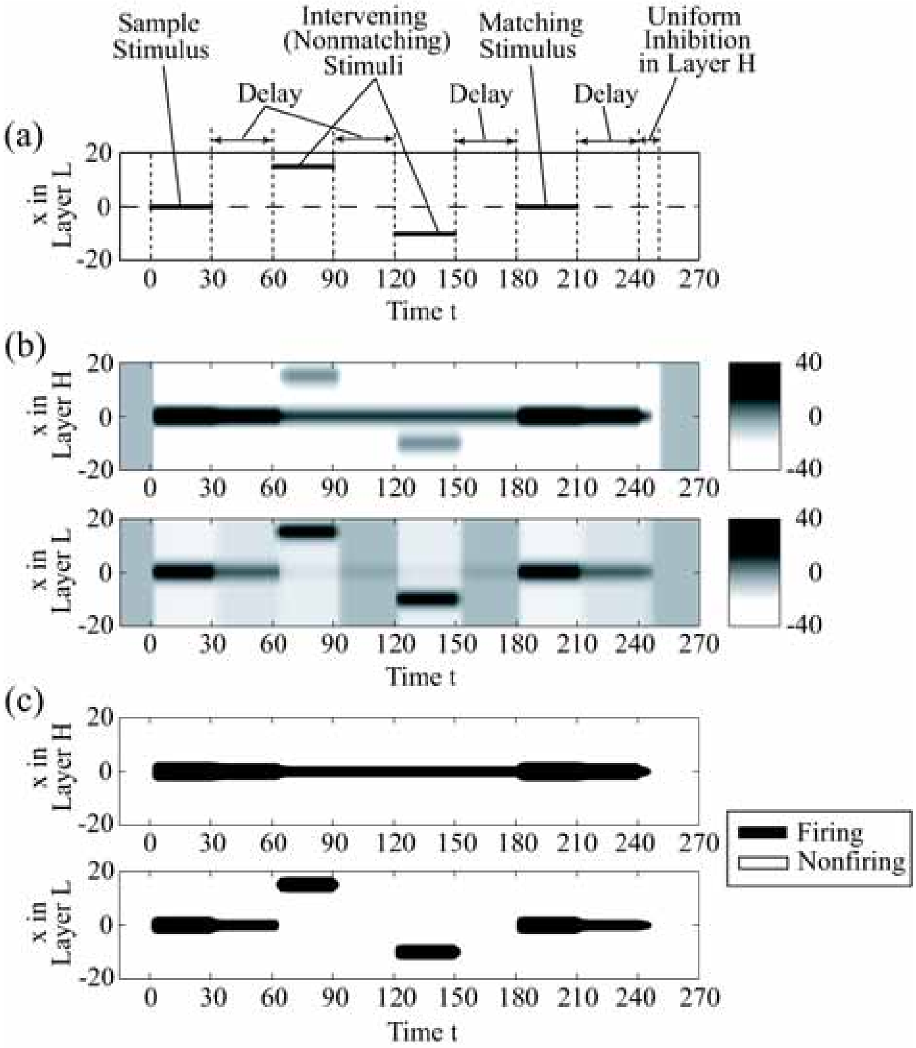


Fig. 2. A simulation result reproducing the neural responses in the DMS task. (a) The time course of the center positions of the input stimuli applied to layer L (solid lines) with the description of the time interval when spatially uniform inhibitory inputs are applied to layer H. (b) The spatio-temporal response patterns of the membrane potential in layer H and layer L. (c) The spatio-temporal response patterns of the firing state in the two layers. The parameter values are set as $\tau_{\mathsf{H}} = \tau_{\mathsf{L}} = 1$, $T_{\mathsf{H}} = T_{\mathsf{L}} = 7$, $l = 20$, $K_{\mathsf{H}}^{\mathsf{exc}} = 9$, $K_{\mathsf{H}}^{\mathsf{inh}} = 3.6$, $K_{\mathsf{L}}^{\mathsf{exc}} = 4.5$, $K_{\mathsf{L}}^{\mathsf{inh}} = 1.8$, $K_{\mathsf{HL}} = 5$, $K_{\mathsf{LH}} = 1$, $\sigma_{\mathsf{H}} = \sigma_{\mathsf{L}} = \sigma_{\mathsf{HL}} = \sigma_{\mathsf{LH}} = 2$, $A_{\mathsf{s}} = 17$, $\sigma_{\mathsf{s}} = 2$, $N_{\mathsf{s}} = 4$, $x_{\mathsf{s}}^1 = 0$, $x_{\mathsf{s}}^2 = 15$, $x_{\mathsf{s}}^3 = -10$, $x_{\mathsf{s}}^4 = 0$, $\Delta T_{\mathsf{s}} = \Delta T_{\mathsf{d}} = 30$, $A_{\mathsf{r}} = 15$, $t_{\mathsf{r}} = 240$, and $\Delta T_{\mathsf{r}} = 10$.

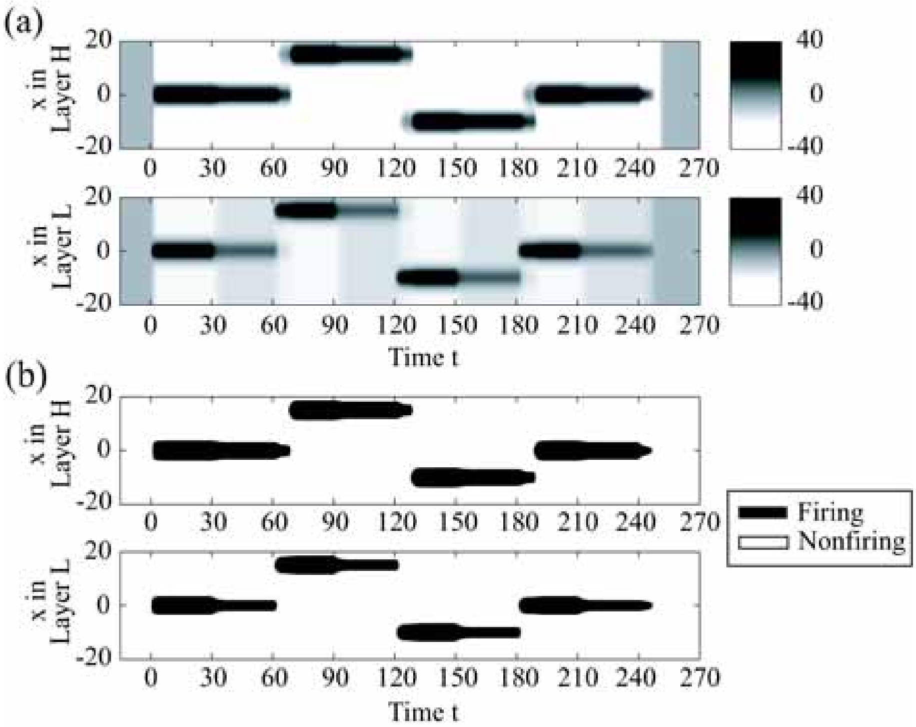


Fig. 3. A simulation result of the DMS task with stimulus intensity larger than that in Fig. 2. (a) The spatio-temporal response patterns of the membrane potential in layer H and layer L. (b) The spatio-temporal response patterns of the firing state in the two layers. The parameter values are the same as that used in Fig. 2 except for $A_{\mathrm{S}} = 25$.

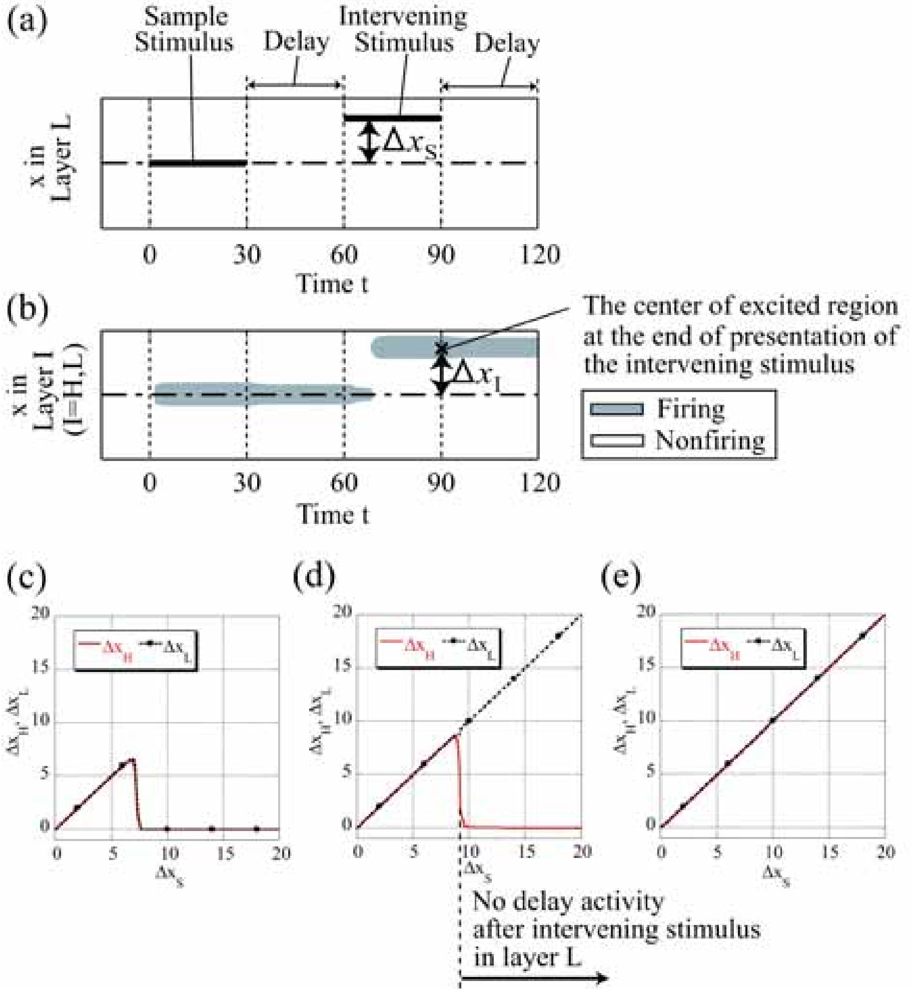


Fig. 4. The response characteristics of the network when the two inputs corresponding to the sample and the intervening stimuli are applied. (a) The time course of the center positions of the inputs applied to layer L and the definition of Δx_{S}. (b) The spatio-temporal firing pattern in layer I in response to the input stimuli shown in (a) and the definition of Δx_{I} $(I = H, L)$. (c)-(e) The relationship between Δx_{S} and Δx_{I} $(I = H, L)$ where the values of stimulus intensity are $A_{\mathsf{S}} = 10$ for (c), $A_{\mathsf{S}} = 17$ for (d), and $A_{\mathsf{S}} = 25$ for (e). The other parameter values different from those in Fig. 2 are $N_{\mathsf{S}} = 2$ and $A_{\mathsf{r}} = 0$. Δx_{S} corresponds to $x^2_{\mathsf{S}} - x^1_{\mathsf{S}}$.

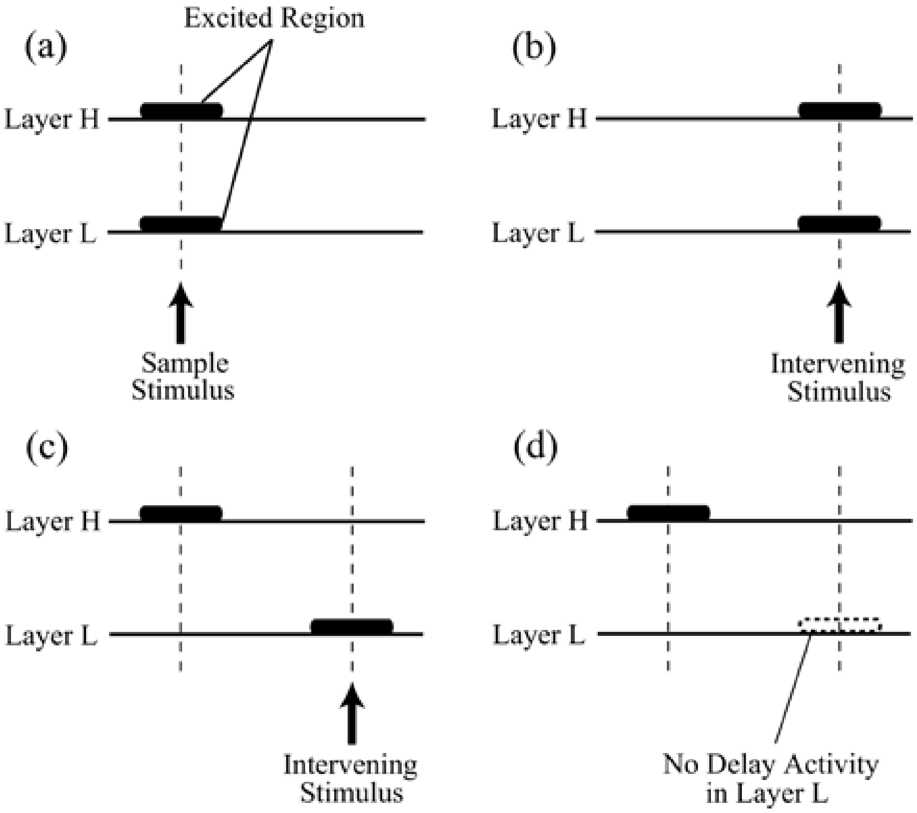


Fig. 5. Schematic diagrams of the pattern dynamics in the two-layer network. (a) The network state where the local excitations are elicited at the same position in the two layers by the sample presentation. (b) The state where the local excitations are produced at the position of the intervening stimulus in the two layers (Case A). (c) The state where the excitation in layer L has moved to the position of the intervening stimulus (Case C). (d) The delay period activity where the excitation in layer L has disappeared after the intervening stimulus presentation (Case C).

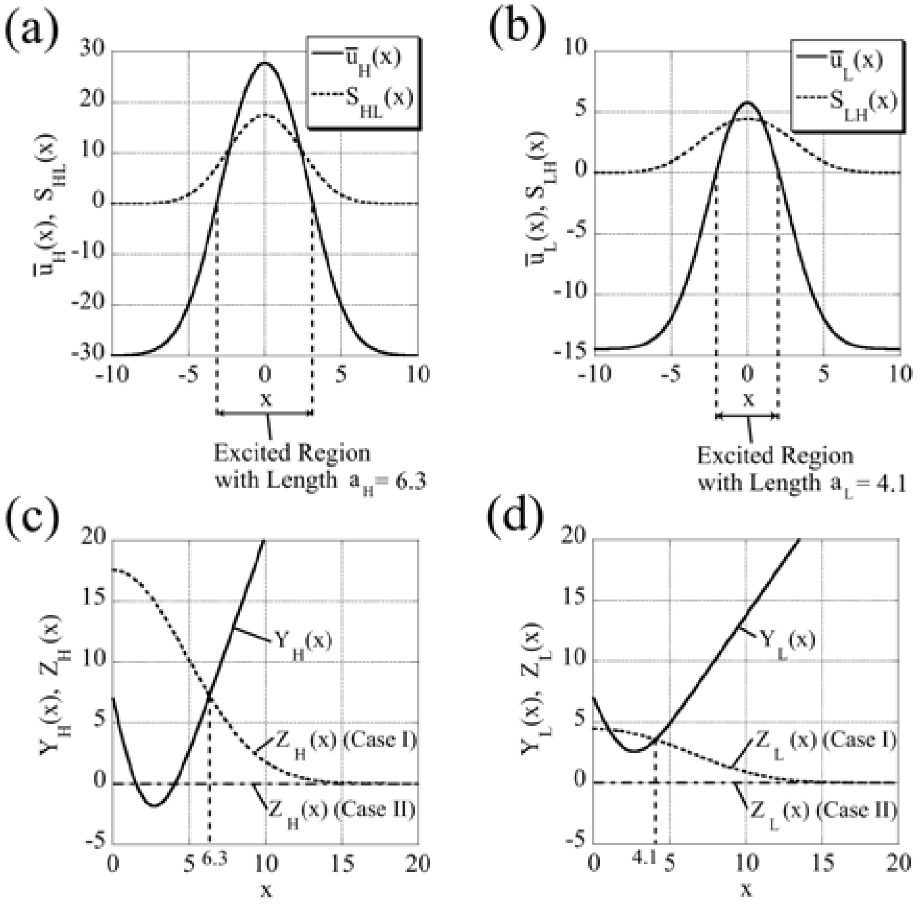


Fig. 6. Applying the results of theoretical studies [1, 18] to the pattern dynamics in Case I and Case II. (a) The plot of $\bar{u}_{\mathsf{H}}(x)$ and $S_{\mathsf{H\,L}}(x)$ in Case I. (b) The plot of $\bar{u}_{\mathsf{L}}(x)$ and $S_{\mathsf{L\,H}}(x)$ in Case I. (c) The relationship between $Y_{\mathsf{H}}(x)$ and $Z_{\mathsf{H}}(x)$ for Cases I and II. (d) The relationship between $Y_{\mathsf{L}}(x)$ and $Z_{\mathsf{L}}(x)$ for Cases I and II.