Multiple topological representation self-organized by spike-timing-dependent synaptic learning rule

Yutaka Sakai · Koji Wada

Received: 20 May 2008 / Revised: 11 November 2008 / Accepted: 11 November 2008 / Published online: 28 November 2008
© Springer Science+Business Media B.V. 2008

Abstract Position-and-scale-free representations of shapes are acquired by neurons in the inferior temporal (IT) cortex. So each neuron receives information from the whole visual field. Familiar shapes are extremely restricted from all the possible shapes on the whole visual field. So they must be clustered in the shape space to have mixed structure of continuity and discreteness. We demonstrate that multiple representation can be acquired in a spike-based model for topological maps based on the spike-timing-dependent synaptic plasticity (STDP), subjected to a set of inputs on multiple rings, which is a simple example of mixed structure. In this representation, the position on each ring is represented by a center of active neurons and the difference of rings is represented by a detailed pattern of active neurons. Neurons in the same region exhibit high activities for an input on the other ring. The result is consistent with the fact observed in IT cortex that neighboring neurons exhibit different preferences while the region of active neurons is continuously shifted for continuous changes of object.

Keywords Topological map · Inferior temporal cortex · Spike-timing-dependent plasticity

Introduction

It is known that spike frequency of a neuron in the primary visual cortex (V1) depends on the orientation of edge in a specific small region of visual field. In other words, a V1 neuron possesses its preferred position and orientation. The feature space composed of position and orientation is covered by population of V1 neurons. Thus, the features of orientation and position of local edge are extracted from a visual image and expressed as a pattern of spike frequencies of V1 neurons. These neurons are arranged on a cortex so that neurons preferring more close features might be more close on a cortex. Such an arrangement of neurons is called as a “topological map”. Topological maps are generally observed in sensory neural systems as well as V1. These maps are considered to be self-organized by changing of connection strength, and various models have been proposed (von der Malsburg 1973; Takeuchi and Amari 1979; Kohonen 1982).

On a higher stage of extracting visual features, a position-free representation is acquired. For example, in the inferior temporal (IT) cortex, a neuron exhibits high spike frequency selectively to a certain specific shape and independently from the visual position and size. It implies that an IT neuron receives information from the whole visual field. However, the number of IT neurons is too few to cover all shapes possible in high-dimensional visual space. So IT neurons should cover only shapes frequently seen, which are extremely restricted from the whole possible shapes. It is unlikely that all of the familiar shapes are continuously linked. On the other hand, there is a set of shapes that are linked continuously, for example a set of shapes obtained by rotation of a familiar object. Therefore, shapes of familiar objects are considered to be on separate clusters in the space of shapes. We can say that

Y. Sakai (✉)
Tamagawa University Brain Science Institute,
Tamagawa-gakuen 6-1-1, Machida-shi,
Tokyo 194-8610, Japan
e-mail: sakai@eng.tamagawa.ac.jp

K. Wada
Department of Electrical Engineering, Kochi National College
of Technology, Kochi, Japan

IT neurons receive inputs that have mixture of continuous and discrete structure.

In the case of such clustered inputs, the topological map models acquire patch-like structures dividing a cortex discretely. Continuous components in each cluster are represented by position in each patch. In fact, continuous changes of selectivity are observed in neighboring IT neurons (Fujita et al. 1992; Wang et al. 1998), and patch-like neurons selective to a shape are observed in IT cortices (Wang et al. 1998). These facts are consistent with the models for topological maps. However, an IT neuron generally exhibits high activity even for entirely different shapes from its preferred shape. The activity pattern for a given set of shapes differs among neighboring neurons (Tamura et al. 2005). This fact is contradictory to the traditional topological map models. It is found that IT cortex acquires multiple representation different from traditional topological maps.

In order to overcome this conflict, Wada et al. (2004) introduced random recurrent connections between local neurons, while traditional models for topological maps have symmetric and homogeneous connection patterns. They applied it to a set of inputs distributed on two parallel rings in 3-dimensional space. As a result, dual representation is acquired where the position on each ring is represented by the center position of active neurons localized in a region and the difference of rings is represented by the detailed activity pattern of neurons in the region. In this representation, neighboring neurons exhibit different activity patterns for the set of inputs, but they are likely to prefer similar inputs. It is consistent with the physiological results (Fujita et al. 1992; Wang et al. 1998; Tamura et al. 2005).

In these models for topological maps (von der Malsburg 1973; Takeuchi and Amari 1979; Kohonen 1982; Wada et al. 2004), each neuron is assumed as an abstract unit that outputs value of spike frequency. The rule for updating connections also depends on spike frequency. On the other hand, a biological neuron outputs spikes. It is necessary to assume a unit as an ensemble of identical and independent neurons in order to reduce the realistic model to a frequency-based model. Every unit should be a neuronal ensemble. For example, in the simulation by Wada et al. (2004), we should assume that many neuronal ensembles on a cortex receive inputs commonly from three neuronal ensembles. There is not necessarily such a discrete grouping of neurons in a cortex. The present study attempts to examine reproducibility for multiple topological maps on a spike-based model that is free from the restriction of discrete groupings.

A spike-based model for topological maps has been proposed by Song and Abbott (2001), based on the physiological phenomena known as the spike-timing-dependent synaptic plasticity (STDP) (Bi and Poo 1998). In their

simulation, a topological map for inputs on one ring in high dimensional input space is organized on one-dimensional cortex. The present study examines what maps are organized in the equivalent model for a set of inputs on multiple rings.

Methods

Network structure

In the present study, we use a model equivalent to the model in Song and Abbott (2001). There are 1000 neurons on the input layer and 200 neurons arranged circularly on the output layer (Fig. 1). The output layer corresponds to a cortex. Each neuron on the output layer receives excitatory connections from randomly selected input neurons at 20% connecting rate, excitatory connections from 80 neighboring neurons on the output layer, and an inhibitory connection from an interneuron which excited by all of the output neurons. In addition, each neuron receives inhibitory inputs from 20% of 200 inhibitory neurons firing independently at rate 10 Hz (Poisson process). The strengths of inhibitory synapses and excitatory synapses on the interneuron are fixed as $w_{\text{inh}} = 0.2$ and $w_{\text{inter}} = 1.0$ respectively.

Input patterns

Each input neuron emits random spikes (Poisson process) at the current rate depending on its state: “up-state” with high spike rate λ^{up} or “down-state” with low spike rate λ^{down} . Up-down pattern is given as an input. Let up-state and down-state correspond to the values 1 and 0 respectively, and an input pattern is described by 1000-dimensional binary vector $\xi = (\xi(1), \xi(2), \dots, \xi(1000))$. We prepare k

Fig. 1 Model network structure. Lines show the excitatory (solid) and inhibitory (dashed) connections received by one (gray circle) of the output neurons. The other output neurons receive connections in the same way

Fig. 2 Topological structure of input pattern set in 1000-dimensional binary space in the case of two rings ($k = 2$)

sets of n random binary patterns $\{\xi_\kappa^\mu | \mu = 1, 2, \dots, n\}$ ($\kappa = 1, 2, \dots, k$) at a constant bit rate, i.e. the summation $\sum_i \xi_\kappa^\mu(i) = R$, which are arranged on respective rings at constant intervals measured by the overlap or Hamming distance,

$$\xi_\kappa^1 \cdot \xi_{\kappa'}^2 = \xi_\kappa^2 \cdot \xi_{\kappa'}^3 = \dots = \xi_\kappa^n \cdot \xi_{\kappa'}^1 = \begin{cases} Rm & (\kappa = \kappa') \\ RM & (\kappa \neq \kappa') \end{cases} \quad (1)$$

where m and M represent the intra-ring and inter-ring overlaps. Current input pattern switches randomly among the prepared set of input patterns $\{\xi_\kappa^\mu\}$ at a mean interval of 40 ms (Poisson process). Figure 2 schematically shows topological structure of the input patterns $\{\xi_\kappa^\mu\}$ in the case of 2 rings ($k = 2$). This set of inputs is topologically equivalent to the parallel two rings introduced by Wada et al. (2004). When the intra-ring overlap m is sufficiently larger than the inter-ring overlap M , the position on a ring described by μ is considered to be a relatively continuous feature to the discreteness of the difference of rings.

We fixed the following parameters throughout the simulations in this paper: $\lambda^{\text{up}} = 36.7$ Hz, $\lambda^{\text{down}} = 3.33$ Hz, $R = 200$, $n = 1,000$, and $m = 0.995$. The inter-ring overlap M is a control parameter in the simulations. At the bit rate $0.2 = R/1000$, the solution for Eq. 1 exists only when $k \leq 5$. Therefore, we simulated the cases that $k = 2, 3, 4, 5$.

Output neuron model

Each output neuron and the interneuron obeys the leaky integrate-and-fire model. We use the current-based synaptic inputs instead of conductance-based inputs. So we introduce a constant synaptic delay of 5 ms to include

effectively the open-and-closing time scale of conductance. Let $s_j(t)$ denote series of delta functions shifted 5 ms from the spike timings of the j -th neuron,

$$s_j(t) = \sum_k \delta(t - t_k^j - 5), \quad (2)$$

where $\{\dots, t_k^j, t_{k+1}^j, \dots\}$ are spike timings of the j -th neuron. The index j is assigned serially through every input and output neuron. The membrane potential of the i -th output neuron obeys the following equation,

$$\frac{dv_i}{dt} = -\frac{v_i}{\tau_v} + \sum_{j \in \text{exc.}} w_{ij} s_j(t) - w_{\text{inh}} \sum_{j \in \text{inh.}} s_j(t) \quad (3)$$

$$v_i > v_{\text{th}} \Rightarrow \text{spike} \ \& \ v_i = 0,$$

where the summation $\sum_{j \in \text{exc.}}$ is for all the excitatory connections received from the input and output neurons, and the summation $\sum_{j \in \text{inh.}}$ is for all the inhibitory inputs received from all the other output neurons and the source of random inhibitory inputs. The parameters are set as $\tau_v = 20\text{ms}$ and $v_{\text{th}} = 20$.

Learning rule

All the excitatory connections w_{ij} except for the connections on the interneuron are updated according to the STDP rule (Song and Abbott 2001), while the recurrent connections on the output layer are fixed in the frequency-based models (von der Malsburg 1973; Takeuchi and Amari 1979; Kohonen 1982; Wada et al. 2004). Let the time scales of timing window of spike-timing-dependent potentiation and depression be equal to each other, and the connections w_{ij} obey the following equations,

$$\frac{da_i}{dt} = -\frac{a_i}{\tau_a} + s_i(t), \quad (4)$$

$$\frac{dw_{ij}}{dt} = A^+ a_j \Theta[1 - w_{ij}] s_i(t) - A^- a_i \Theta[w_{ij}] s_j(t), \quad (5)$$

where the function $\Theta[\cdot]$ denotes the unit step function (Heaviside function), introduced in order to limit the strength of connection in the region $w \in [0, 1]$. The model parameters are fixed as $\tau_a = 20\text{ms}$, $A^+ = 0.005$ and $A^- = 1.05A^+$.

Data analysis

After approaching the steady state, we stop updating the connections (5) and examine activity pattern of the output neurons for each input pattern. The current input pattern is fixed to a certain pattern μ on a certain ring κ and spike frequencies of output neurons are estimated from 10 s spike series. In the same way, the frequency patterns or the activity patterns for all the input patterns are obtained,

$$\alpha_{\kappa}^{\mu} = (\alpha_{\kappa}^{\mu}(1), \alpha_{\kappa}^{\mu}(2), \dots, \alpha_{\kappa}^{\mu}(200)),$$

where $\alpha_{\kappa}^{\mu}(i)$ corresponds to the spike frequency of the i -th output neuron for input pattern μ on ring κ .

In order to examine the properties of selectivities that the output neurons achieved through the learnings, we define the inter-ring correlation of activity patterns, C , and of smoothed activity patterns, $\tilde{C}$, which are the mean direction cosines between the activity vectors α_{κ}^{μ} and $\alpha_{\kappa'}^{\mu}$, and the smoothed activity vectors $\tilde{\alpha}_{\kappa}^{\mu}$ and $\tilde{\alpha}_{\kappa'}^{\mu}$, for $\kappa \neq \kappa'$,

$$C = \frac{2}{nk(k-1)} \sum_{\kappa \neq \kappa'} \sum_{\mu=1}^n \frac{\alpha_{\kappa}^{\mu} \cdot \alpha_{\kappa'}^{\mu}}{|\alpha_{\kappa}^{\mu}| |\alpha_{\kappa'}^{\mu}|},$$

$$\tilde{C} = \frac{2}{nk(k-1)} \sum_{\kappa \neq \kappa'} \sum_{\mu=1}^n \frac{\tilde{\alpha}_{\kappa}^{\mu} \cdot \tilde{\alpha}_{\kappa'}^{\mu}}{|\tilde{\alpha}_{\kappa}^{\mu}| |\tilde{\alpha}_{\kappa'}^{\mu}|},$$

$$\tilde{\alpha}_{\kappa}^{\mu}(i) = \sum_{j=-20}^{20} \alpha_{\kappa}^{\mu}(i+j),$$

where the index i for α satisfies periodic boundary conditions, for example $\alpha_{\kappa}^{\mu}(-j)$ means $\alpha_{\kappa}^{\mu}(200-j)$.

Results

Obeying the Eqs. 2–5, the excitatory connections $\{w_{ij}\}$ are updated depending on the timings of spikes caused by randomly switching input patterns of the prepared set $\{\xi_{\kappa}^{\mu}\}$. The initial values of feed-forward connections from the input layer to the output layer are uniformly distributed in $[0,1]$ and the initial values of recurrent connections on the output layer are set to zero. After a long enough time about 10,000 s, the distribution of $\{w_{ij}\}$ approaches a steady distribution.

Figure 3 shows the frequency patterns in the cases of two rings ($k=2$) with inter-ring overlap values $M=1, 0.5, 0.1$ as grey scale plots. The case that $M=1$ corresponds to the case that a ring completely coincides with the

Fig. 3 **a–c** Spike frequency patterns α_{κ}^{μ} (for ring 1) and $\alpha_{\kappa'}^{\mu}$ (for ring 2) of output neurons after sufficient self-organization (10,000 s) for the set of input patterns $\{\xi_{\kappa}^{\mu}\}$ in the case of two rings ($k=2$) with the

inter-ring overlap $M=1$ (**a**), 0.5 (**b**), 0.1 (**c**). **d**: Inter-ring correlations C (solid line marked by squares) and $\tilde{C}$ (dashed line marked by circles) are shown as a function of the inter-ring overlap M

Fig. 4 a–d Spike frequency patterns α_1^μ (for ring 1 of k rings) and α_2^μ (for ring 2 of k rings) of output neurons after sufficient self-organization (10,000 s) for the set of input patterns $\{\xi_k^\mu\}$ in the case of 2–5 rings ($k = 2, 3, 4, 5$) with the inter-ring overlap $M = 0.3$. **d** Inter-ring correlations C (solid line marked by squares) and $\tilde{C}$ (dashed line marked by circles) are shown as a function of the number of rings k

Fig. 5 The detailed activity patterns of neighboring neurons 1–3 after sufficient self-organization (10,000 s) for the set of input patterns $\{\xi_k^\mu\}$ in the case of 5 rings ($k = 5$) with the inter-ring overlap

$M = 0.3$. The activity patterns of neurons 1, 2 and 3 for input patterns $\mu = 400, 420, 430, \dots, 600$ on the individual rings are extracted to the bottom plots from the top plots

other, namely the case of one ring which is equivalent to the simulation by Song and Abbott (2001). In the case that $M = 1$ (Fig. 3a), localized neurons exhibit high activities for each input pattern and the center of active neurons is shifted as the position of the input pattern on the ring is shifted. It is confirmed that a topological map representing the position on the ring can be self-organized in the present spike-based model as well as the demonstration by Song and Abbott (2001).

In the case of the more separate rings, $M = 0.5$ (Fig. 3b), we can see the same type of correspondence between the input position on a ring and the center of active neurons. In addition, neurons in the same region exhibit high activity even for the input pattern on the corresponding position of the other ring. So the center of active neurons represents the input position on the rings regardless of the difference of rings. Is there any difference in the activity patterns between the two rings?

Figure 3d shows the inter-ring correlation of activity patterns, C (solid line marked with squares), and of smoothed activity patterns, $\tilde{C}$ (dashed line marked with circles). We can see that the inter-ring correlation of activity patterns decreases as the input rings separate more, while that of smoothed activity patterns stays near 1. So in the middle range of inter-ring overlap, $M = 0.3$ – 0.5 , the position on each ring is represented by the center of active neurons on the output layer, and the difference of rings is represented by the detailed activity patterns of the localized active neurons.

In the case of the still more separate rings, $M = 0.1$ (Fig. 3c), topological maps for the position on each ring is still observed, but the correspondence between the two rings is broken. The inter-ring correlation of the smoothed activity patterns also vanishes (dashed line marked by circle in Fig. 3d).

Multiple representations observed in the middle range of inter-ring overlap do not vanish up to 5 rings. Figure 4a–d shows the activity patterns for two of k rings after sufficient self-organization for 2–5 rings input patterns $\{\xi_k^\mu\}$ ($k = 2, 3, 4, 5$) respectively, in the case that $M = 0.3$. Figure 4e shows their inter-ring correlations C and $\tilde{C}$. The inter-ring correlations change little up to 5 rings, while topological representation becomes slightly noisy.

In the case of 5 rings, we show the selectivities of neighboring neurons in Fig. 5. Neurons 1–3 are all selective to around pattern $\mu \sim 500$ for all 5 rings (top row). However, each of the three neurons is selective to different rings. The detailed selectivity patterns are different.

Discussion

We found that the present spike-based model acquires multiple representation in which the position on each ring is represented by the center of active neurons on the output layer, and the identity of ring is represented by the detailed activity patterns of the localized active neurons. In addition, in the middle range of inter-ring distance, positions on the individual rings represented by a center of active neurons correspond. This type of multiple representation is consistent with the results in Wada et al. (2004). The duality

commonly originates in random recurrent connections on the output layer, which are also self-organized by the STDP rule in the present model, while they are prepared in advance in the model by Wada et al. (2004).

The present result implies the possibility that multiple topological maps for different features are acquired in a sheet of neural field and the different maps are distinguished with the detailed pattern of active neurons. The results are consistent with the properties of IT neurons. Note that single topological representation like V1 and multiple one like IT can be self-organized by the identical model for different topological structures of input patterns. This implies the possibility that representation in V1 and IT may be self-organized by a common computational principle.

Acknowledgment This work is supported in part by Grants-in-Aid for Scientific Research on Priority Areas, No.18019034.

References

- Anderson J, Rosenfeld E (eds) (1988) Neurocomputing: foundations of research. MIT Press, Cambridge
- Bi G, Poo M (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J Neurosci* 18:10464–10472
- Fujita I, Tanaka K, Ito M, Cheng K (1992) Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360:343–346
- Kohonen T (1982) Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43. Reprinted in (Anderson and Rosenfeld, 1988)
- Song S, Abbott LF (2001) Cortical development and remapping through spike timing-dependent plasticity. *Neuron* 32:339–350
- Takeuchi A, Amari S (1979) Formation of topographic maps and columnar microstructures in nerve fields. *Biol Cybern* 35:63–72
- Tamura H, Kaneko H, Fujita I (2005) Quantitative analysis of functional clustering of neurons in the macaque inferior temporal cortex. *Neurosci Res* 52:311–322
- von der Malsburg C (1973) Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik* 14. Reprinted in (Anderson and Rosenfeld, 1988)
- Wada K, Kurata K, Okada M (2004) Self-organization of globally continuous and locally distributed information representation. *Neural Netw* 17:1039–1049
- Wang G, Tanifuji M, Tanaka K (1998) Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. *Neurosci Res* 32:33–46