Neuronal Mechanisms for Hierarchical Encoding in Inferior-Temporal Cortex

Narihisa Matsumoto [a,b,*] Masato Okada [b,c]

[a] *Graduate School of Science and Engineering, Saitama University, Saitama 338-8570, Japan*

[b] *RIKEN Brain Science Institute, Saitama 351-0198, Japan*

[c] *"Intelligent Cooperation and Control", PRESTO, JST, Saitama 351-0198, Japan*

Abstract

Sugase et al. found that global information is processed at initial transient firings of face-responsive neurons in the inferior-temporal (IT) cortex, and that finer information is processed at the subsequent sustained firings. An attractor network is employed to elucidate neuronal mechanisms producing this dynamics. The results of computer simulations show that the behavior of both neuronal population and a single neuron qualitatively coincides with Sugase's data. Therefore, the attractor network might be the neuronal mechanisms producing the dynamics of the face-responsive neurons in the IT cortex.

Key words: attractor network, face-responsive neurons, ultrametric patterns

* Corresponding author.
  Email address: `xmatumo@brain.riken.go.jp` (Narihisa Matsumoto).

1 Introduction

In the inferior-temporal (IT) cortex, which is regarded as a final stage of the visual cortex, some neurons respond to faces or complex objects [4,6]. Previous works on neuronal activity indicated that a pattern recognition is complete at initial transient firings of the neurons [10]. However, recent experimental results were shown, which suggest the possibility that not only the stationary state of the neuronal activity but its dynamics encodes some information [5,12]. Sugase et al. recorded the activity of the face-responsive neurons in the IT cortex by using single-unit electrodes, while presenting visual stimuli [12]. The visual stimuli in their experiment consisted of 12 human faces (3 models with 4 expressions), 16 monkey faces (4 models with 4 expressions), and 10 simple objects (2 models with 5 colors). The relationship among the stimuli was hierarchical because the stimuli were categorized into global categories, e.g., human, monkey, and object, and finer categories, e.g., individual and expression. The temporal change in neuron firing rates indicated that the single neuron encodes the information of the global categories at the initial transient firing and that of the finer categories at the subsequent sustained firing. This implies that the hierarchical relationship among the visual stimuli is extracted and it is coded in the transient and the sustained activity of the single neuron.

We employ an attractor network [2,3] to elucidate neuronal mechanisms producing the dynamics of the face-responsive neurons in the IT cortex. Amit et al. showed that the stationary state of the network might produce the sustained activity of the IT neurons by employing the attractor network [3]. It is easily anticipated that the convergence process to the attractor, i.e., attractor dynamics, might produce the dynamics of the face-responsive neurons.

Amari found that when the memory patterns with the hierarchical structure are stored in an associative memory model, not only a memory pattern but a mixed state, which is a nonlinear superposition of the memory patterns, becomes an attractor spontaneously [1]. The behavior of the face-responsive neurons as a population level indicated that the activity of the neuronal population is initially a superposition of patterns representing the different faces, and then it converges to a single pattern representing the specific face [8]. The mixed state represents the superposition of the different faces while each stored memory pattern represents each face. The results of computer simulations show that the macroscopic state of the network approaches the mixed state, and finally it converges to the memory pattern. The results qualitatively coincide with the population dynamics of the face-responsive neurons. The behavior of a single neuron in our model also coincides with their data.

2 Model

Our model consists of excitatory and inhibitory neurons [3]. The excitatory connections from the excitatory neurons to the inhibitory neurons are uniform. The inhibitory connections from the inhibitory neurons to the excitatory neurons are also uniform. There are no connections between each inhibitory neuron. Therefore, the population of the inhibitory neurons can be regarded as a single inhibitory neuron. The equations that describe the model are as follows: $\tau_{ex}\dot{I}_i^{ex} = -I_i^{ex} + \sum_{j\neq i}^N J_{ij}V_j^{ex} - T + H_i^{ext}, V_i^{ex} = \phi(I_i^{ex}), \tau_{in}\dot{I}^{in} = -I^{in} + \frac{1}{fN}\sum_{i=1}^N V_i^{ex}, T = k \times V^{in} = k \times \psi(I^{in})$. N is the number of excitatory neurons and J_{ij} is the synaptic weight from the j-th neuron to the i-th neuron. τ_{ex} and τ_{in} are time constants of the excitatory neurons and the inhibitory

3

neuron, respectively. I_i^{ex} is the membrane potential of the i-th excitatory neu-
ron while I^{in} is the membrane potential of the inhibitory neuron. V_i^{ex} is the
output of the i-th excitatory neuron while V^{in} is the output of the inhibitory
neuron. $\phi(I)$ is a saturation output function of the excitatory neurons while
$\psi(I)$ is a threshold-linear output function of the inhibitory neuron. The out-
puts of these functions are mean firing rates of the neurons. f denotes the
mean firing rate of the memory pattern.

Since the visual stimuli in the Sugase's experiment had the hierarchical struc-
ture [12], ultrametric patterns $\boldsymbol{\xi}^{\mu,\nu}$ are used as the memory patterns. The
elements of the memory pattern are binary, $\{0, 1\}$. If $\xi_i^{\mu,\nu} = 1$, the mean
firing rate takes the maximal value of the i-th neuron, otherwise it takes 0.
The procedure for generating the ultrametric patterns is shown as follows. At
first, each element ξ_i^{μ} of a parent pattern $\boldsymbol{\xi}^{\mu}$ is generated independently by
$\text{Prob}[\xi_i^{\mu} = 1] = 1 - \text{Prob}[\xi_i^{\mu} = 0] = f$. Next, s memory patterns $\boldsymbol{\xi}^{\mu,\nu}$ are
generated, preserving a correlation with $\boldsymbol{\xi}^{\mu}$. μ denotes a group number while
ν denotes an element number. The memory patterns belonging to the same
group are correlated, while the memory patterns belonging to the different
groups are uncorrelated. Consequently, the memory patterns are structured
hierarchically. The parent pattern $\boldsymbol{\xi}^{\mu}$ is regarded as the representative pattern
of the group μ since $\boldsymbol{\xi}^{\mu}$ is considered to be average of the memory patterns $\boldsymbol{\xi}^{\mu,\nu}$
belonging to the same group μ. The synaptic weight J_{ij} from the j-th excita-
tory neuron to the i-th excitatory neuron is determined by the Hebb rule [7]:
$J_{ij} = \frac{1}{fN} \sum_{\mu=1}^{p} \sum_{\nu=1}^{s} \xi_i^{\mu,\nu} \xi_j^{\mu,\nu}$. Note that the representative patterns $\boldsymbol{\xi}^{\mu}$ are not
stored in the synaptic connections. External input $\boldsymbol{H}^{ext}$ is injected into only
the excitatory neurons. When the stimulus $\boldsymbol{\xi}^{\mu,\nu}$ is input, the external input
$\boldsymbol{H}^{ext}$ follows the memory pattern $\boldsymbol{\xi}^{\mu,\nu}$, i.e., $H_i^{ext} = H\xi_i^{\mu,\nu}$, where H denotes

4

the coefficient of the external input.

All initial values of $\boldsymbol{I}$ are set at 0. The values of the parameters are set as follows: $N = 3000$, $p = 20$, $s = 3$, $C = 0.3$, $\tau_{ex} = 10$ ms, $\tau_{in} = 2$ ms, $k = 8.0$, $f = 0.05$, $H = 0.055$.

3   Results

We introduce a measure to describe the macroscopic state of the network, i.e., behavior of the neuronal population. This is called overlap and $m^{\mu,\nu}$ represents a distance between the memory pattern $\boldsymbol{\xi}^{\mu,\nu}$ and the vector of the outputs $\boldsymbol{V}^{ex}$. This value is calculated by $m^{\mu,\nu}(t) = \frac{1}{fN} \sum_{i=1}^{N} \xi_i^{\mu,\nu} V_i^{ex}(t)$. As $m^{\mu,\nu}$ approaches 1, $\boldsymbol{V}^{ex}$ converges to $\boldsymbol{\xi}^{\mu,\nu}$. When $\boldsymbol{\xi}^{1,2}$ is input, the temporal change of the neuronal population obtained by computer simulations is drawn in Fig.1(a). The solid line shows a vector trajectory of $m^{1,2}$ between $\boldsymbol{\xi}^{1,2}$ and $\boldsymbol{V}^{ex}$. The abscissa axis is $m^{1,2}$, while the vertical axis is $m^{1,1}$ between $\boldsymbol{\xi}^{1,1}$ and $\boldsymbol{V}^{ex}$. (i) in Fig.1(a) shows the initial state of the network. Then, no neurons fire and $m^{1,1} = m^{1,2} = 0$. (ii) shows that the macroscopic state approaches the representative pattern $\boldsymbol{\xi}^{1}$ (t $\approx$ 33ms). (iii) shows that the macroscopic state converges to the memory pattern $\boldsymbol{\xi}^{1,2}$. This temporal behavior of the neuronal population is consistent with Sugase's data [8].

Next, we compare the temporal behavior of a single neuron in our model with in Sugase et at.'s data. The neurons in our model are classified into $8(= 2^3)$ classes according to the stored patterns $(\xi_i^{1,1}, \xi_i^{1,2}, \xi_i^{1,3})$. These 8 classes can be effectively classified into three groups: $(0,1,0)$, $(1,1,1)$ and $(1,0,1)$. Hereafter, we call the i-th neuron with $(\xi_i^{1,1}, \xi_i^{1,2}, \xi_i^{1,3}) = (0,1,0)$ the "(0,1,0) neuron" and

so on. If the first group ($\mu = 1$) is regarded as the human category, $\boldsymbol{\xi}^{1,1}$, $\boldsymbol{\xi}^{1,2}$, and $\boldsymbol{\xi}^{1,3}$ correspond to the first individual, the second one and the third one, respectively. The $(0, 1, 0)$ neuron has the finer information to discriminate the second human individual from other individuals since it fires when $\boldsymbol{\xi}^{1,2}$ is input and does not fire when either $\boldsymbol{\xi}^{1,1}$ or $\boldsymbol{\xi}^{1,3}$ is input. Figure 1 (b) shows the temporal change of the $(0, 1, 0)$ neuron obtained by the computer simulations. The lines of $\boldsymbol{\xi}^{1,1}$, $\boldsymbol{\xi}^{1,2}$ and $\boldsymbol{\xi}^{1,3}$ are the firing patterns for the first individual, the second one, or the third one, respectively. The abscissa axis is time (ms), while the vertical axis is firing rate normalized by the maximal firing rate of the neuron. The neuron responds to $\boldsymbol{\xi}^{1,2}$ persistently, while it responds to $\boldsymbol{\xi}^{1,1}$ and $\boldsymbol{\xi}^{1,3}$ transiently. This behavior of the $(0, 1, 0)$ neuron coincides with that of a single neuron in Sugase et al's data [11]. The $(1, 1, 1)$ neuron codes all of $\boldsymbol{\xi}^{1,1}$, $\boldsymbol{\xi}^{1,2}$ and $\boldsymbol{\xi}^{1,3}$ as "1" while it codes all of the memory patterns belonging to the different groups as "0" under a sparse coding scheme, i.e., $f \rightarrow 0$. This neuron has the global information to discriminate the memory patterns in the first group from in the other groups. On the other hand, it does not have the finer information to discriminate the memory pattern in the first group from the other memory patterns in the first group. Computer simulations show that the neuron responds to $\boldsymbol{\xi}^{1,1}$, $\boldsymbol{\xi}^{1,2}$ and $\boldsymbol{\xi}^{1,3}$ persistently. The $(1, 0, 1)$ neuron codes $\boldsymbol{\xi}^{1,1}$ and $\boldsymbol{\xi}^{1,3}$ as "1", while it codes $\boldsymbol{\xi}^{1,2}$ as "0". This neuron has smaller amount of the finer information to discriminate $\boldsymbol{\xi}^{1,2}$ from $\boldsymbol{\xi}^{1,1}$ and $\boldsymbol{\xi}^{1,3}$ than the $(0, 1, 0)$ neuron since the probability of "0" is higher than that of "1" under the sparse coding scheme. Computer simulations show that the neuron responds to $\boldsymbol{\xi}^{1,2}$ transiently, while it responds to $\boldsymbol{\xi}^{1,1}$ and $\boldsymbol{\xi}^{1,3}$ persistently.

4   Summary

We employed the attractor network to elucidate the neuronal mechanisms producing the dynamics of the face-responsive neurons in the IT cortex. Our model stored the memory patterns with the hierarchical structure corresponding to the hierarchical relationship among the visual stimuli. The results of the computer simulations showed that the macroscopic state of the network approaches the mean state of similar patterns transiently, and finally it converges to the memory pattern. This implies that the hierarchical structure of the memory patterns is extracted, which qualitatively coincides with the population dynamics of the face-responsive neurons [8]. The behavior of the $(0, 1, 0)$ neuron also coincides with the dynamics of the single neuron [11]. Therefore, the attractor network might be the neuronal mechanisms producing the dynamics of the face-responsive neurons in the IT cortex.

Acknowledgements

References

[1]   S. Amari, Neural theory of association and concept-formation, Biol. Cybern. 26 (1977) 175–185.

[2]   D.J. Amit, Modeling brain function: the world of attractor neural networks,

(Cambridge Univ. press, Cambridge, 1989).

[3] D.J. Amit, N. Brunel, and M.V. Tsodyks, Correlations of cortical hebbian reverberations: theory versus experiment, J. Neurosci. 14 (1994) 6435–6445.

[4] C. Bruce, R. Desimone, and C.G. Gross, Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque, J. Neurophysiol. 46 (1981) 369–384.

[5] R.W. Friedrich and G. Laurent, Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity, Science 291 (2001) 889–894.

[6] I. Fujita, K. Tanaka, M. Ito, and K. Cheng, Columns for visual features of objects in monkey inferotemporal cortex, Nature 360 (1992) 343–346.

[7] D.O. Hebb, The Organization of Behavior: A Neuropsychological Theory, (Wiley, New York, 1949).

[8] N. Matsumoto, M. Okada, K. Doya, Y. Sugase, and S. Yamane, Dynamics of face responsive neurons in the temporal cortex, in: Soc. Neurosci. Abstracts, Vol. 27, 2001, pp. 1048.

[9] S. Panzeri, E.T. Rolls, F. Battaglia, and R. Lavis, Speed of feedforward and recurrent processing in multilayer networks of integrate-and-fire neurons, Network 12 (2001) 423–440.

[10] E.T. Rolls and M.J. Tovee, Processing speed in the cerebral cortex and the neurophysiology of visual masking, Proc. R. Soc. Lond. B Biol. Sci. 257 (1994) 9–15.

[11] Y. Sugase, Dynamic encoding of facial information represented by neuronal responses in the primate temporal cortex, PhD thesis, University of Tokyo, 1999.

[12] Y. Sugase, S. Yamane, S. Ueno, and K. Kawano, Global and fine information coded by single neurons in the temporal visual cortex, Nature 400 (1999) 869–873.
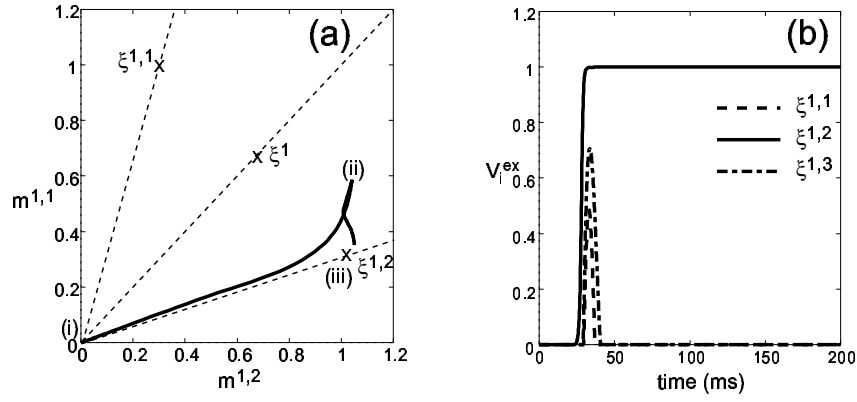
Fig. 1. (a): temporal change of the neuronal population when $\boldsymbol{\xi}^{1,2}$ is input. The solid line shows a vector trajectory of $m^{1,2}$. (b): temporal change of $(0,1,0)$ neuron.