Neural mechanism of dynamic responses of neurons in inferior temporal cortex in face perception

Yuichiro Yamada · Yoshiki Kashimori

Received: 7 December 2011 / Revised: 30 May 2012 / Accepted: 6 July 2012 / Published online: 20 July 2012
© Springer Science+Business Media B.V. 2012

Abstract Understanding the neural mechanisms of object and face recognition is one of the fundamental challenges of visual neuroscience. The neurons in inferior temporal (IT) cortex have been reported to exhibit dynamic responses to face stimuli. However, little is known about how the dynamic properties of IT neurons emerge in the face information processing. To address this issue, we made a model of IT cortex, which performs face perception via an interaction between different IT networks. The model was based on the face information processed by three resolution maps in early visual areas. The network model of IT cortex consists of four kinds of networks, in which the information about a whole face is combined with the information about its face parts and their arrangements. We show here that the learning of face stimuli makes the functional connections between these IT networks, causing a high spike correlation of IT neuron pairs. A dynamic property of subthreshold membrane potential of IT neuron, produced by Hodgkin–Huxley model, enables the coordination of temporal information without changing the firing rate, providing the basis of the mechanism underlying face perception. We show also that the hierarchical processing of face information allows IT cortex to perform a “coarse-to-fine” processing of face information. The results presented here seem to be compatible with experimental data about dynamic properties of IT neurons.

Keywords Inferior temporal cortex · Dynamic response · Hierarchical processing · Face perception · Neural model

Introduction

In everyday life, we can recognize quickly and effortlessly complex visual scenes. To understand how we have such an amazing ability, it is important to understand the neural mechanism of object recognition. A number of studies have been made to investigate the neuronal mechanism of object recognition. Object shape information is processed in the ventral pathway, a hierarchical network of areas in occipital and temporal lobes (Ungerleider and Mishkin 1982). At early visual areas such as V1, V2, and V4, individual neurons signal simple properties of local edge fragments (e.g. position, orientation, curvature) (Gallant et al. 1993; Hubel and Wiesel 1962; Ito and Komatsu 2004; Pasupathy and Connor 1999). The neural mechanisms of information processing in these areas have been relatively well understood. On the other hand, at higher visual area such as inferior temporal (IT) cortex, objects are encoded by a sparse pattern of active neurons (Baker et al. 2002; Freedman et al. 2003; Fujita et al. 1992; Gross et al. 1972; Logothetis et al. 1995; Tsunoda et al. 2001; Young and Yamane 1992). However, how the visual information is represented in IT cortex remains unclear. How do the information about a whole object and its parts are represented in IT cortex? To address this issue, we focus on the mechanism of face perception. It is a good model system for studying the neural mechanism underlying the representation of object features, because face has a whole image and consists of well-defined parts such as eyes, nose, and mouth.

Recent experiments have reported dynamic responses of IT neurons in face perception. Hirabayashi and Miyashita

Y. Yamada · Y. Kashimori (✉)
Graduate School of Information Systems, University
of Electro-Communications, Chofu, Tokyo 182-8585, Japan
e-mail: kashi@pc.uec.ac.jp

Y. Kashimori
Department of Engineering Science, University
of Electro-Communications, Chofu, Tokyo 182-8585, Japan

(2005) demonstrated that the IT neurons exhibited different responses in spike correlations for face and non-face images. Face-like objects elicited neuronal activities that were more correlated than the corresponding non-face like objects, whereas the firing rates of the same neurons did not show a significant difference. Sugase et al. (1999) found that single IT neurons conveyed two different scales of facial information in their firing patterns, starting at different latencies. They showed from the statistical analysis for spikes of IT neurons that global information about categorization of visual stimuli was conveyed in the early part of the responses, and fine information about identity and expression was conveyed later. Furthermore, Kiani et al. (2005) reported that IT neurons responded to human and non-primate animal faces with comparable magnitude, but responded significantly more quickly to human faces than to non-primate animal faces.

The response property of IT neurons significantly depends on learning and experience, implying the importance of synaptic plasticity in producing the response property of IT neurons. There are much evidence for enhancement of IT neuronal responses by learning and experience. Neurons in monkey IT cortex showed enhanced sensitivity after training for novel objects (Kobatake et al. 1998; Logothetis et al. 1995; Rolls 1995), holistic multiple-parts configurations (Baker et al. 2002), and physically unrelated pairs of shapes (Messinger et al. 2005; Miyashita 2004). Furthermore, learning can shape the assignment of novel objects into classes (Rosenthal et al. 2001) by modulating the sensitivity of neurons in the IT and frontal cortex for features critical for the categorization processes (Freedman et al. 2003; Sigala and Logothetis 2002). Theoretical studies also showed that the enhanced sensitivity of IT neurons was caused by the learning based on the interaction between IT and prefrontal cortex (Kashimori et al. 2007; Soga and Kashimori 2009; Szabo et al. 2006). In these studies, the enhanced sensitivity of IT neurons was explained by an increase in the magnitude of IT neuronal responses induced by learning. However, it is not yet clear how the synaptic plasticity by learning and experience affects the spike dynamics of IT neurons.

In the present study, we present a model of IT cortex, which performs face perception via an interaction between three IT layers and face recognition layer. The three IT layers encode the information about face features processed by V4 maps with different spatial resolutions. We show using the computational model that the spike synchronicity between the IT layers, evoked by synaptic plasticity, is needed to make functional connections between these layers. After the learning, the information about a whole face and its parts are effectively combined by the synchronous firing of these layers, leading to a face perception. A nonlinear property of IT neuron, produced by

Hodgkin–Huxley model, enables the coordination of temporal information without changing the firing rate, providing the basis of the mechanism underlying face perception. The synaptic strength of the functional connections also affects the onset timing of the IT neuronal responses, associated with the familiarity of faces. Furthermore, the feedback from the face recognition layer to IT layers allows IT cortex to perform a “global-to-fine” perception of faces. The results presented here seem to be compatible with the experimental results on the temporal properties of IT neurons.

Model

Overview of our model

The network model of face perception is shown in Fig. 1. The model consists of three IT layers, face recognition layer (ITFR), and V4X ($X = B, M, F$). These IT layers and ITFR are mutually connected with excitatory synapses. ITX ($X = B, M, F$) receive the information about face image processed by the three V4 layers with different spatial resolutions. ‘B’, ‘M’, and ‘F’ indicate the broad, middle, and fine resolution, respectively. The ITB receives a face information encoded by V4B with a broad spatial resolution, and thereby encodes the coarse information of the whole image. The ITM encodes the information about the arrangement of face parts, represented by V4M with a middle resolution. The ITF binds the information of local curvatures encoded by V4F with a fine resolution, and thereby encodes the information about the fine features of face parts such as eyes, nose, and mouth. The ITFR combines the information encoded by ITB, ITM, and ITF layers, and represents a unique image of face stimulus to

Fig. 1 The neural network model of IT cortex. The ITX ($X = B, M$, and F) receive the output of V4X with different spatial resolutions. ‘B’, ‘M’, and ‘F’ mean the broad, middle, and fine spatial resolutions, respectively. The ITFR (face recognition; FR) has neurons responding to different faces. The unit of ITM, ITF, and ITFR contains a cluster of neurons, which is depicted by the network consisting of the small circles

perform the identification of faces. The ITB is a model of posterior IT, whereas the ITM, ITF, and ITFR layers correspond to areas involved in anterior IT.

Models of ITX ($X = B, M, F, FR$)

The ITX ($X = M, F, FR$) contains M_X units, each of which is constructed with a cluster of N_X neurons, in which the neurons are weakly connected with each other, as shown in Fig. 1. The neurons within a unit respond to a face feature, with different sensitivities. The units make no synaptic connection within each IT layer, but they make synaptic connections across IT layers. The ITB has no cluster structure in contrast to other IT layer, because ITB only categorizes a coarse image of whole face. Experimental study (Hirabayashi and Miyashita 2005) has demonstrated that the temporal correlation of an IT neuron pair increases in response to face stimuli, while the firing rate remains unchanged. This suggests that the IT neurons may have a nonlinear property of neuronal responses. The usually used neuron models such as the leaky integrate-and-fire (LIF) and leaky integrator models are not suitable for IT neuron model, because an increase in input magnitude would result in both an increase in firing rate and temporal correlation, through an increase of synaptic connections elicited by learning. Hodgkin–Huxley model (Hodgkin and Huxley, 1952) exhibits a change of firing rate less than LIF, in response to an increase of input magnitude. This is due to a nonlinear dynamics of subthreshold membrane potential of Hodgkin–Huxley model, mainly produced by a partially activated K-channel. The weak inhibitory effect by the K-channel is responsible for an oscillation of subthreshold membrane potential, allowing a neuron to elicit a change of firing rate less sensitive to change of input magnitude. In contrast, LIF model has a firing rate sensitive to change of input magnitude. This is because subthreshold membrane potential of LIF model is determined by a linear equation, mimicking the effect of Na channels in the subthreshold region, and thereby it changes straightforwardly to change of input magnitude. Thus we expected that Hodgkin–Huxley model may enable only an increase in temporal correlation without increasing firing rate. Additionally, the ITX neurons have different sensitivities to a face feature, resulting from the difference in the conductance values of Na- and K-channels. Neurons have embryologically different numbers of Na- and K-channels. The different distributions of these channel conductances produce both a variability of spike timing and firing rate of IT neurons in response to visual stimuli. These conductance values were varied within $\pm 10\%$ of the standard values used in Hodgkin–Huxley model. The membrane

potentials of i th neuron within j th unit in ITX layer, $V_{ij,ITX}$ ($X = B, M, F, FR$), are determined by

$$\begin{aligned} C_m \frac{dV_{ij,ITM}}{dt} = & -g_{ij,ITM}^{Na} m^3 h(V_{ij,ITM} - V_{Na}) - g_{ij,ITM}^K n^4 \\ & \times (V_{ij,ITM} - V_K) - g_{ij,ITM}^L (V_{ij,ITM} - V_L) \\ & + \sum_{kl} w(ij, ITM; kl, ITF; t) S_{kl,ITF}(t) \\ & + \sum_{mn} w(ij, ITM; mn, ITB; t) S_{mn,ITB}(t) \\ & + I_{ij,V4M} + \sum_{pq} w(ij, ITM; pq, ITFR; t) S_{pq,ITFR}(t) \\ & + \sum_{i'} w(ij, ITM; i'j, ITM) S_{i'j,ITM}(t) \\ & + \xi_{ij,ITM}(t), \end{aligned} \quad (1)$$

$$\begin{aligned} C_m \frac{dV_{ij,ITF}}{dt} = & -g_{ij,ITF}^{Na} m^3 h(V_{ij,ITF} - V_{Na}) \\ & - g_{ij,ITF}^K n^4 (V_{ij,ITF} - V_K) - g_{ij,ITF}^L (V_{ij,ITF} - V_L) \\ & + \sum_{kl} w(ij, ITF; kl, ITM; t) S_{kl,ITM}(t) + I_{ij,V4F} \\ & + \sum_{pq} w(ij, ITF; pq, ITFR; t) S_{pq,ITFR}(t) \\ & + \sum_{i'} w(ij, ITF; i'j, ITF) S_{i'j,ITF}(t) + \xi_{ij,ITF}(t), \end{aligned} \quad (2)$$

$$\begin{aligned} C_m \frac{dV_{ij,ITFR}}{dt} = & -g_{ij,ITFR}^{Na} m^3 h(V_{ij,ITFR} - V_{Na}) - g_{ij,ITFR}^K n^4 \\ & \times (V_{ij,ITFR} - V_K) - g_{ij,ITFR}^L (V_{ij,ITFR} - V_L) \\ & + \sum_{kl} w(ij, ITFR; kl, ITF; t) S_{kl,ITF}(t) \\ & + \sum_{pq} w(ij, ITFR; pq, ITM; t) S_{pq,ITM}(t) \\ & + \sum_{rs} w(ij, ITFR; rs, ITB; t) S_{rs,ITB}(t) \\ & + \sum_m \sum_{n \neq j} w^I(ij, ITFR; mn, ITFR) S_{mn,ITFR}^I(t) \\ & + \sum_{i'} w(ij, ITFR; i'j, ITFR) S_{i'j,ITFR}(t) \\ & + \xi_{ij,ITFR}(t), \end{aligned} \quad (3)$$

$$\begin{aligned} C_m \frac{dV_{ij,ITB}}{dt} = & -g_{ij,ITB}^{Na} m^3 h(V_{ij,ITB} - V_{Na}) - g_{ij,ITB}^K n^4 \\ & \times (V_{ij,ITB} - V_K) - g_{ij,ITB}^L (V_{ij,ITB} - V_L) \\ & + \sum_{kl} w(ij, ITB; kl, ITM; t) S_{kl,ITM}(t) \\ & + I_{ij,V4B} + \sum_{pq} w(ij, ITB; pq, ITFR; t) S_{pq,ITFR}(t) \\ & + \xi_{ij,ITB}(t), \end{aligned} \quad (4)$$

where

$$S_{Y,ITX}(t) = \sum_{n=0}^M \frac{t - t_{Y,n}}{\tau_s} e^{-\frac{t-t_{Y,n}}{\tau_s}}, \quad (5)$$

$$(Y = i'j, kl, mn, pq, rs; X = B, M, F, FR),$$

$$S_{mn,ITFR}^I(t) = \sum_{l=0}^M \frac{t - t_{mn,l}}{\tau_s^I} e^{-\frac{t-t_{mn,l}}{\tau_s^I}}, \quad (6)$$

$$I_{ij,V4X} = \begin{cases} I_0, & \text{if } (i,j) = \arg \min_{ij} d_{ij} \\ 0 & \text{otherwise,} \end{cases} \quad (X=B, M, F), \quad (7)$$

$$d_{ij} = \sqrt{\sum_{mn} (Z_{mn,V4X} - w(ij, ITX; mn, V4X))^2}, \quad (8)$$

C_m is the membrane capacitance, $g_{ij,ITX}^z$ ($X = B, M, F, FR$; $z = Na, K, L$) are the maximum conductance of z channel, m, n , and h are the probabilities with which the channels are in the open state, and V_z ($z = Na, K, L$) are the equilibrium potential of ion z . The synaptic weight, $w(ij, ITM; kl, ITF; t)$, is the weight of synaptic connection from k th neuron within l th unit in ITF to i th neuron within j th unit in ITM, and $w(ij, ITM; mn, V4M)$ is that from (m, n) th neuron of V4M to i th neuron within j th unit in ITM. Other w 's also have the similar definitions. In Eq. (4), the value of i is 1 because the ITB has no cluster structure of neurons within unit. $I_{ij,V4X}$ ($X = B, M, F$) are the synaptic currents of ITX neuron evoked by the output of V4X neuron and $Z_{mn,V4X}$ is the output of (m, n) th neuron in V4X. The winner ITX neuron, determined by Eqs. (7) and (8), is activated by a constant current I_0 , which is disturbed by the noise inputs, $\xi_{ij,ITX}(t)$ that is given by the zero-mean Gaussian distributed with σ covariance functions. V4X is the input layer to ITX, whose output is simplified with the values of 0 and 1 because of stationary visual stimulus. $S_{Y,ITX}(t)$ ($Y = i'j, kl, mn, pq, rs$; $X = B, M, F, FR$) are the synaptic input evoked by the spike of the neuron located at the position Y in ITX layer, given by the sum of α -function described by Eq. (5), where $t_{Y,n}$ is the firing time of the latest spike of the neuron located at the position Y in ITX layer, and M is the number of spikes generated by the neuron until the latest spike firing. The inhibitory synaptic input of ITFR neuron, $S_{mn,ITFR}^I(t)$, is described by Eq. (6), similar to Eq. (5) except for the time constant τ_s^I . The second last terms of the right hand side of Eqs. (1–3), respectively, indicate the currents elicited by the firing of the intra-unit neurons. The third last term of the right hand side of Eq. (3) indicates the inhibitory current across the ITFR layer.

Representation of face information in V4X ($X = B, M, F$)

The V4B has a 2-dimensional array of neurons, as shown in Fig. 2a, and forms a retinotopic map. Each neuron takes 1

Fig. 2 Representation of face features in V4X ($X = B, M$, and F). **a** Face image on the retina (left), and the network model of V4B (right). The activated neurons are depicted by gray-colored cells. **b** Face image on the retina (left) and the network model of V4M (right). The small square in the face image indicates the point to which attention is directed. The V4M encodes the information about the arrangement of the face part around the attention point. **c** Image of face part (left) and the network model of V4F (right). The network has hypercolumns, which contains columns of neurons sensitive to different local curvatures. The V4F magnifies the local features of face part selected by attention

as its output when receiving the input evoked by a local curve of face, and does 0 in the absence of the input. The V4M and V4F respond to a face part selected by attention signal. As shown in Fig. 2b, the V4M has a retinotopic map, on which each neuron encodes the information about a location occupied by face part, that is, the information about the arrangement of face part in a whole face image. The V4F contains hypercolumns covering an image of face part, as shown in Fig. 2c. Each hypercolumn contains N_L columns of neurons responding to a local curvature. The V4F encodes the information about a local curvature of face part. The outputs of V4M and V4F neurons were determined by the similar way to the output of V4B neuron.

The weights of synaptic connections between ITX and V4X ($X = B, M, F$) were learned with Kohonen's self-organized maps (Kohonen 1995). The details for determining the synaptic weights are described in "Appendix". After the learning, the feature-detective neurons responding to coarse image of face were organized in ITB layer, the neurons responding to the arrangement of face parts were done in ITM layer, and those responding to the face

parts such as eyes, nose, and mouth were formed in ITF layer.

Learning in ITX ($X = B, M, F, FR$)

The neurons in the ITX ($X = B, M, F, FR$) are reciprocally connected with excitatory synapses, as shown in Fig. 1. The synaptic connections between ITX ($X = M, F, B, FR$) were updated based on Hebbian learning. Then the weight of synaptic connection from k th neuron within l th unit in ITY layer and i th neuron within j th unit in ITX layer, $w(ij, ITX; kl, ITY; t)$, is determined by

$$\tau_w \frac{dw(ij, ITX; kl, ITY; t)}{dt} = -w(ij, ITX; kl, ITY; t) + \eta \left(1 - \frac{w(ij, ITX; kl, ITY; t)}{w_{\max}} \right) S_{ij, ITX}(t) \cdot S_{kl, ITY}(t),$$

(9)

($X, Y = B, M, F, FR$),

where τ_w and $w_{\max}$ are the time constant and the maximum value of synaptic weight, respectively. The learning rate depends on the synaptic weight, so that the synaptic weight tends to the maximum value as the learning proceeds. η is the maximum learning rate.

Presentation of face stimuli during the learning

We used two face objects (FOs), two non-face objects (NFOs), and one animal face object (AFO), as shown in Fig. 3. The face objects have their parts at normal positions, while the non-face objects have a random arrangement of face parts. The animal face is a cat-like face shown in Fig. 3c. In the learning period of face, the face stimuli were frequently presented, while the non-face stimuli were rarely presented, because we encounter various faces every day but we do rarely faces with random arrangement of their parts. The animal face was presented less frequently than FOs. The presentation ratio of FOs, NFOs, and AFO was set at 7:1:2. The duration of stimulus presentation of one object was 400 ms, in which attention was changed at the cycle of 100 ms. The overall duration of the learning was 25 s.

Cross-correlation

The spike correlation was quantified as the cross-correlation,

$$C(\tau) = \frac{\sum_i (x_i - \bar{x})(y_{i+(\tau/\Delta t)} - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}, \quad (10)$$

of pair spikes. x_i and y_i are the data of the pair spikes averaged over intra-unit neurons, and were digitized with 0 and 1 based on the neuronal firing. Δt is the time step used in solving Eqs. (1) to (4), and $\Delta t = 0.05$ ms. $\bar{x}$ and $\bar{y}$ are the averaged values of x_i and y_i over 500 ms, respectively.

Parameters

The values of C_m , $g_{ij, ITX}^{Na}$, $g_{ij, ITX}^K$, $g_{ij, ITX}^L$ ($X = B, M, F, FR$), and the parameter values for determining m , h , and n were based on those used in Hodgkin–Huxley model (Hodgkin and Huxley, 1952). Other parameter values used in our model are described in Table 1. The values on the synaptic connections between IT neurons were determined so as to exhibit no synchronous firing of IT neurons in response to NFOs and synchronous firing for FOs.

Results

Emergence of spike synchronization between the IT layers involved in face perception

Figure 4 shows the temporal variations of spiking for a pair of ITM and ITF neurons after the learning. Each IT neuron shown in Fig. 4 fires at the same frequency in the time range of 60–190 ms. The firing frequency is nearly 80 Hz, which is in the range of gamma frequency. Face stimulus evoked the synchronous firing for every spike pair of ITM and ITF neuron, as shown in Fig. 4a. Figure 4b shows the spike patterns of the ITM and ITF neuron pair in response to non-face stimulus, with the same firing number as Fig. 4a. This indicates the spike patterns of an uncoupled ITM and ITF neuron pair, because non-face stimulus evokes little increase in synaptic connections between ITM and ITF neurons, as shown in Fig. 5 later. Of 9 spike pairs, only 2 spike pairs fire by chance at nearly the same timing. The spike synchronicity elicited by face stimulus gradually increased as the learning proceeded, whereas that elicited by non-face stimulus did not increase. The difference in the synchronicity comes from the

Fig. 3 The images used in the learning. **a** Face objects (FOs; FO1, FO2). **b** Non-face objects (NFOs; NFO1, NFO2). **c** Animal face object (AFO)

Table 1 Parameter values used in the model of IT

Quantities	Equations	Values
(1) Time constant of synaptic currents		
EPSP (τ_s)	Eq. (5)	1.0 ms
IPSP (τ_s^i)	Eq.(6)	10.0 ms
Learning of synaptic weights		
Time constant (τ_w)	Eq. (9)	1000 s
$w_{\max}$	Eq. (9)	
ITB-ITM		3.0
ITM-ITF		3.0
ITX-ITFR (X = B, M, F) (feedforward)		5.0
ITFR-ITX(X = B, M, F) (feedback)		3.0
Maximum learning rate (η)	Eq. (9)	
ITB-ITM		0.01
ITM-ITF		0.03
ITX-ITFR (X = B, M, F) (feedforward)		0.5
ITFR-ITX (X = B, M, F) (feedback)		0.5
(3) The number of unit (M_X) in ITX		
(M_B, M_M, M_F, M_{FR})		(10, 16, 25, 5)
(4) The number of neurons in unit in ITX		
N_X (X = M, F, FR)		10
(5) Synaptic weights		
Connections between neurons within unit		
$w(ij, ITX; i'j', ITX)$ (X = B, M, F, FR)	Eqs. (1–3)	3.0
Inhibitory connections between ITFR neurons		
$w^I(ij, ITFR; mn, ITFR)$	Eq. (3)	−2.0
(6) Input current from V4X to ITX (X = B, M, F)		
I_0	Eq. (7)	10.0 $\mu\text{A}/\text{cm}^2$

difference in the frequency of stimulus presentation. The frequent presentation of face stimuli causes an increase in chance of coincident spiking of ITM and ITF neurons, and thereby facilitates the Hebbian learning of the synaptic connections between ITM and ITF neurons, as shown in Fig. 5. The spike timing of both neurons becomes closer as the number of coincident spikes are increased, resulting in the increase of the learning efficacy of synaptic weights. That is, the coincident spiking causes the increased efficacy of the learning, and the increased synaptic weight further facilitates the synchronous firing of ITM and ITF neurons. Thus, the

Fig. 4 Time courses of spiking of a pair of ITM and ITF neurons. Responses of these neurons to FO1 (a) and NFO1 (b). The upper and lower panels in a and b indicate the spike patterns of ITM and ITF neurons, respectively

synaptic changes induced by the coincident spiking and synchronous firing make ITM and ITF neurons more activate in the positive feedback loop. In contrast, infrequent presentation of non-face stimulus causes little chance of coincident spiking of ITM and ITF neuron, resulting in no significant facilitation of the synaptic learning, as shown in Fig. 5.

The synchronicity of the ITM and ITF neuronal firing was further facilitated by a top-down signal from ITFR, although it was partially organized by the connections between neurons in ITM and ITF. The synaptic weights of the connections conveying the top-down signal were increased during the learning. The effect of the top-down signal on the cross-correlation of neuronal firing of an ITM and ITF neuron pair is shown in Fig. 6. The top-down signal increased the peak value of the cross-correlation and enhanced the synchronicity of ITM and ITF neurons. The magnitude of the top-down signal from ITFR was about one third of the feedforward input to ITX neurons, which was set so that the firing of ITX neurons is not elicited by

Fig. 5 Temporal variations of the weights of synaptic connections from ITM to ITF neurons. The *solid* and *dashed* lines represent the synaptic weights involved in the representations of FO1 and NFO1, respectively

Fig. 6 Cross-correlations of spikes of an ITM and ITF neuron pair. The *solid* and *dashed* lines indicate the cross-correlations in the presence and in the absence of top-down signal from ITFR, respectively. The presence and absence of top-down signal are also depicted with “TDS(+)” and “TDS(-)” on the graph, respectively. The *vertical dashed line* indicates the *line of* $\tau = 0$, which represents the correlation of synchronicity

the top-down signal alone. The top-down signal only modulates the synchronicity of ITM and ITF neurons. The neuron pair between other IT layers also exhibited the synchronous firing similar to Fig. 4a (not shown here).

Taken together, the synchronicity in a pair neurons between two IT layers is organized due to the two effects, that is, the increased synaptic connections between IT layers and the top-down signal from ITFR layer.

Highly correlated response of ITF neurons in face perception

Figure 7 shows the firing properties for a pair of ITF neurons encoding the features of eye and nose after the learning. The

Fig. 7 Response properties of a pair of ITF neurons encoding the features of eye and nose. Cross-correlations of spikes (**a**) and firing rate (**b**) of these neurons in response to FO1 and NFO1. The *upper* and *lower panels* in (**b**) indicate the firing rates for the face and non-face stimuli, respectively

presentation of face stimulus caused a highly correlated firing between ITF neurons encoding face parts, as shown in Fig. 7a. Note that the correlation exhibited the time lag of 2 ms in contrast to the synchronicity of a pair of ITM and ITF neurons. This time lag is in agreement with the experimental result by Hirabayashi and Miyashita (2005). Although these ITF neurons are not directly connected with each other, the connections of the ITF neurons with the ITM and ITFR neurons allow the pair neurons to fire with the high correlation. The time lag of 2 ms comes from the weak coupling of the ITF neuron pair through the feedforward and feedback signals. In contrast, the presentation of non-face stimulus caused no highly correlated firing between these ITF neurons, as shown in Fig. 7a. This is due to the extremely weak connections of the ITF neurons with the ITM and ITFR neurons. In contrast to the spike correlation, the firing rates of these ITF neurons were almost the same in the responses to face and non-face stimuli, as shown in Fig. 7b. These results are in agreement with the experimental results by Hirabayashi and Miyashita (2005).

The pair of ITF neurons encoding the features of eye and nose changed only the temporal correlation of spiking, without affecting the firing rate of the ITF neurons in response to FOs and NFOs. This comes from a nonlinear

dynamics of subthreshold membrane potential of Hodgkin–Huxley model, mainly produced by a partially activated K-channel. The weak inhibition by the K-channel evokes an oscillation of subthreshold membrane potential in response to input, allowing the firing rate to be less sensitive to change of input magnitude. If the subthreshold membrane potential of neuron is determined by a linear equation as used in LIF model, the neuron responds straightforwardly to input and its firing rate becomes sensitive to input magnitude. Thus the linear property of neuron would result in both an increase in spike correlation and firing rate of neurons in response to an increase of input magnitude, with an increase in synaptic connections between ITX layers. This suggests that ITF neurons could have the nonlinearity in order to coordinate temporal information that is modulated by attention and weak signals through the interaction between IT and other brain areas.

Figure 8a shows the time course of neuronal correlations elicited by FO and NFO for an ITF neuron pair

Fig. 8 Temporal dynamics of spike correlation. **a** Time courses of spike correlations of a ITF neuron pair in response to FO1 and NFO1. The cross-correlation at the time 0 ms was calculated using spikes in 500 ms before the stimulus onset time, and the time course was calculated by the spikes in 500 ms window that was successively shifted in steps of 100 ms. **b** Time courses of firing rate of a pair of ITF neurons for FO1 and NFO1

encoding the features of eye and nose. The cross-correlations were calculated by the similar method to that used by Hirabayashi and Miyashita (2005), in which they were calculated using spikes in 500 ms window that was successively shifted by the steps of 100 ms. The neuronal correlations elicited by FO rapidly increased up to 500 ms after the stimulus onset, and then exhibited a slow fluctuation around a constant value, whereas those elicited by NFO gradually increased up to 600 ms, and then showed a fluctuation similar to that elicited by FO. The spike correlations elicited by FO were significantly stronger than those elicited by NFO. The firing rates of the ITF neuron pair for FO and NFO are shown in Fig. 8b. These firing rates rapidly increased up to 600 ms and then fluctuated around a constant value. In contrast to the spike correlations, there is little difference in the firing rates of this pair elicited by FO and NFO. The results exhibited the similar tendency in the case where the spike correlations and firing rates were averaged over all pairs of IT neurons encoding face parts of FOs and NFOs (not shown here). The results are consistent with the results by Hirabayashi and Miyashita (2005).

Figure 9 shows the scatter plots showing peak heights of cross-correlations (Z-scores) and mean firing rates for pairs of ITF neurons, elicited by FO and NFO. The period observed was divided into two periods (from 0 to 300 ms and from 300 to 600 ms, after stimulus onset). During the period from 0 to 300 ms, the peak height of the cross-correlation for FO was significantly higher than that for NFO, whereas the firing rates did not reveal significant dependence on the feature configuration. During the period from 300 to 600 ms, the spike correlations and firing rates for FO and NFO did not reveal a significant dependence of face configuration. Thus the spike correlation was modulated in feature configuration of a whole face during the period from 0 to 300 ms after stimulus onset, whereas the firing rates did not reveal configuration dependent bias within 300 ms after stimulus onset. The results are in agreement with the results by Hirabayashi and Miyashita (2005).

Neural mechanism for producing the time lag of ITFR neurons in response to human and animal faces

Figure 10 shows the response properties of ITFR neurons to FO and animal face object (AFO). The response of ITFR neurons elicited by AFO exhibited a time lag larger than that elicited by FO, as shown in Fig. 10a. The similar tendency appeared in the population of ITFR neurons (Fig. 11a). The points were largely distributed below the diagonal line. In contrast, the points for the firing rate of ITFR neurons were distributed along the diagonal line, as shown in Fig. 11b. This indicates that the firing rates

Fig. 9 Rapid emergence of feature configuration-dependent spike correlation. *Scatter plots* showing peak heights of cross-correlation in z-score (*left*) and firing rates for ITF neurons (*right*), elicited by FO1 and NFO1 in the periods of 0–300 ms (*top*) and 300–600 ms (*bottom*)

Fig. 10 Onset latency of ITFR neuronal firing for FO1 and animal face object (AFO). The firing rates were averaged over the neurons within units responding to these stimuli

exhibit no significant difference for FO and AFO stimuli. The difference in the time lags elicited by FO and AFO is due to the difference in development of synaptic connectivity between IT layers. Figure 12 shows the temporal variations of synaptic connection between the ITF neuron

encoding the feature of eye and that encoding a whole image, during the learning of FO and AFO. The synaptic connection elicited by FO became stronger than that elicited by AFO, because FO, presented more frequently than AFO, leads to a more efficient learning of the synaptic connections. The strong connection results in a spike initiation of ITFR neuron at an earlier time. This suggests that familiar faces, which we encounter frequently, strengthen the synaptic connections between ITFR and other IT layers and allow ITFR neurons to strongly combine the information about a whole face and its parts, leading to a quick response of IT neurons to familiar faces. In contrast, ITFR neurons in response to unfamiliar face such as AFO weakly binds these face information, causing an ITFR neuronal response with a larger time lag. The familiarity of face stimulus, however, does not affect the firing rate of ITFR neurons. Thus the experience-dependent learning well accounts for the time lags of ITFR neurons in response to different faces. The results shown in Figs. 10 and 11 are in agreement with the experimental results by Kiani et al. (2005).

Fig. 11 Scatter plots of onset latency and firing rate of ITFR neurons in response to FO1 and AFO

Functional roles of temporal aspects of IT neuronal responses in face perception

The temporal aspects of ITX neuronal responses ($X = B, M, F, FR$) were modulated by the learning, as discussed above. The learning of synaptic connections between ITX ($X = B, M, F$) layers and ITFR caused the synchronous firing of these IT layers, the highly correlated responses in ITF neurons, and the different time lags in ITFR neuronal responses to human and animal faces. The synchronous firing of ITM and ITF neurons activates effectively ITFR neurons, and the top-down signals emanating from these ITFR neurons then facilitate the synchronous firing of ITM and ITF neurons. Thus the synchronicity of neuronal activities in these IT layers is more stabilized by the positive loop of feedforward and feedback. The synchronous firing of IT neurons has a functional role in face perception. The synchronicity enables IT neurons to strongly combine the information about a whole face image and its parts, and the highly correlated responses of ITF neurons allow IT neurons to combine the features of face parts. As a result, these temporal properties of IT neurons are responsible for providing the visual stimulus with

Fig. 12 Temporal variations of the weights of synaptic connections from ITF to ITFR neurons during the learning. Only the two synaptic connections learned by FO1 and AFO are depicted

semantic information such as faces. Furthermore, the time lag of IT neuronal responses has another role in face perception, associated with the familiarity of face stimuli. The more familiar the visual stimulus is, the shorter time lag the IT neurons exhibit. This shorter time lag leads to a quick response to familiar faces. Thus the dynamic property of IT neurons may play crucial roles in face perception.

Role of ITB layer in the hierarchical processing of face information

Figure 13 shows the responses of ITFR neurons to FO1 stimulus. In the early period of face perception, the ITFR neurons sensitive to FO1 and those sensitive to FO2, respectively, exhibited the increased firing rates. These ITFR neurons are activated by the feedforward signals that convey the information about the coarse image of FO1 encoded by the ITB layer. After the early period up to 50 ms, the ITFR neurons sensitive to FO1 kept the firing rate, whereas those sensitive to FO2 rapidly decreased the firing rate, and then became silent, as shown in Fig. 13. This indicates that the IT layers convey the information about the coarse image of face earlier than that about the fine image. The feedforward signals from the ITB activate the ITFR neurons sensitive to FO1 and those to FO2, and then the feedforward signals from ITM and ITF neurons, conveying the information of the fine face image, enhance the activation of ITFR neurons sensitive to FO1, and suppress that of ITFR neurons sensitive to FO2. The result is consistent with the neurophysiological experimental result of coarse-to-fine processing by Sugase et al. (1999).

To examine the role of top-down signal from ITFR to ITM and ITF layers, we eliminated the connection between ITB and ITFR. The elimination prohibited the generation of top-down signal, resulting in a delay response of ITFR

Fig. 13 Responses of ITFR neurons to FO1. **a** Raster plot of ITFR neurons encoding FO1 and FO2. The label ‘i–j’ ($i = 1, 2; j = 1-5$) on the vertical axis means jth neuron within ith unit in ITFR layer. The ITFR neurons sensitive to FO1 and those sensitive to FO2 are indicated by “1–j” and “2–j” ($j = 1-5$), respectively. **b** Temporal variations of firing rates of the ITFR neurons corresponding to the raster plot

neurons to FO1 stimulus, as shown in Fig. 14. This indicates that the top–down signals from ITFR to ITM and ITF are required for the fast discrimination of face. To examine the neuronal mechanism underlying the fast discrimination of faces, we calculated the subthreshold membrane potentials of ITM and ITF neurons evoked by top–down signals alone. They exhibited a high temporal correlation, as shown in Fig. 15. This indicates that the top–down signal evokes a subthreshold membrane oscillation and brings the potential to beneath the firing threshold, enabling the IT neurons to have a quick response to face stimulus, as shown in Fig. 13. The subthreshold oscillation is also responsible for the enhancement of synchronous firing of ITM and ITF neurons shown in Fig. 6. The results shown in Figs. 14 and 15 suggest that the top–down signals from ITFR neurons could function as a prediction signal of the visual stimulus. The ITB sends rapidly coarse information about face stimulus to ITFR and then activates the ITFR neurons within a category of whole face. The top down signals from ITFR are sending back to the ITF and ITM and gives the information about possible faces, that is, a prediction signal of the face stimulus. Then the refinement of the relevant face is made by the combination of the top–down signal and the bottom-up signals received by the ITM and ITF neurons.

Fig. 14 Effect of top–down signal on the firing of ITFR neurons. The response of ITFR neurons to FO1 is shown under the elimination of the connection between ITB and ITFR layers. The raster plot and firing rate are depicted with the similar way to those of Fig. 13

Fig. 15 Effect of top–down signal on subthreshold membrane potentials of ITM and ITF neurons. Cross-correlation of subthreshold membrane potentials of a pair of ITM and ITF neurons, evoked by top–down signals from ITFR

Discussion

We have shown with the computational model of IT cortex that the information about a whole image and its parts is represented by a hierarchical processing in IT and the association between the whole image and its parts is effectively learned when the image has semantic information such as faces. The learning evokes the synchronous firing and highly correlated responses of IT neurons to face

stimuli, which is responsible for an efficient binding of the information about a whole face and its parts. Extent to which the association is learned depends on the frequency of experience. A nonlinear property of IT neurons, produced by Hodgkin–Huxley model, is needed to perform the temporal coordination of face information without increasing the firing rate. The learning by experience determines the familiarity of stimulus. The model also provides a plausible mechanism of coarse-to-fine processing of visual information in IT. The results shown by the model are consistent with the experimental results on the dynamic responses of IT neurons (Hirabayashi and Miyashita 2005; Kiani et al. 2005; Sugase et al. 1999).

Our model can explain systematically the temporal aspects of IT neuronal spiking evoked by faces and non-face objects, characterized by synchronous firing, highly correlated responses, short-term delay of spike onset, and temporal property of spikes in coarse-to-fine information processing. The neuronal mechanism underlying these temporal properties is based on experience-dependent learning in IT cortex. Many studies have demonstrated that experience-dependent learning modulates the IT neuronal responses (Baker et al. 2002; Freedman et al. 2006; Kobatake et al. 1998; Sigala and Logothetis 2002). Although these studies demonstrated the increased magnitude of IT neuronal responses elicited by learning, how experience-dependent learning affects the temporal properties of IT neuronal responses is poorly understood. We showed that increased synaptic connections between IT layers modulated the temporal properties of IT neuronal responses. This suggests that it is important for face and object recognition to what extent the functional connections between IT layers are flexibly modulated by learning. Our model provides a critical insight into understanding how experience can improve face recognition.

Our model has a long-term memory in IT, suggesting the involvement of glutamate receptors. Synchronization may also be enhanced by inhibitory interneurons through GABAergic receptors. Several studies indicate that different glutamate receptors and cholinergic and GABAergic systems contribute to object recognition of perirhinal cortex (Turchi et al. 2005; Wan et al. 2004; Winters and Bussey 2005a, b, c). This suggests that the similar neurotransmitters could be involved in object representation of IT. In addition, other transmitters such as serotonin (5-HT), norepinephrine (NE), and dopamine might be involved. The detailed modeling of synaptic connections in our model, with further pharmacological studies, might contribute to determination of neurotransmitters involved in learning of IT.

We developed the network model of IT that was constructed with the four IT layers. The ITB, ITM, and ITF layers represent the different face features based on the

local curvatures encoded by V4 layers with different spatial resolutions. There is anatomical and physiological evidence for supporting our model. In anatomical structure of IT cortex, ITB corresponds to a posterior region of IT (TEO), whereas ITM, ITF, and ITFR correspond to anterior regions of IT (AIT). For physiological data, it has been reported that IT neurons respond to more complex features of objects along the pathway from posterior to anterior region of IT (Kobatake and Tanaka 1994; Logothetis and Sheinberg 1996; Tanaka 1996). IT cortex contains the neurons sensitive to intermediately complex features of objects, corresponding to ITB in our model. Also, neurons in macaque IT represent both global and local shape of a hierarchical stimulus (Sripati and Olson 2009; Sugase et al. 1999). This implies that IT contains areas corresponding to ITF and ITFR. Furthermore, IT neurons have been shown to be remarkably sensitive to the configuration of objects rather than their fine features (Yamane et al. 2006), suggesting the existence of ITM. Our model is based on these experimental results of IT cortex.

In the present study, we have focused on the neural mechanism of face perception to investigate how the information about a holistic object and its parts is represented in IT. Single IT neurons respond selectively to highly complex visual stimuli such as faces (Desimone et al. 1984; Gross 2005; Perrett et al. 1982; Rolls 1984) and these neurons are found in relatively large clusters in IT (Perrett et al. 1992; Tsao et al. 2003; Zangenehour and Chaudhuri 2005). Recent experimental studies have reported that microstimulation of face-selective sites in IT strongly biases monkey's decision towards the face category (Afraz et al. 2006) and that these IT neurons are tuned to a fundamental dimension of face perception (Leopold et al. 2006). Thus IT is strongly associated with face perception. On the other hand, there is experimental evidence that face processing involves specific regions highly sensitive to faces. A region in ventral human cortex (fusiform face area, FFA) thought to be important for face perception responds strongly to faces and less strongly to non-face objects (Kanwisher et al. 1997). Furthermore, hierarchical processing of face information is found in a network of six interconnected face processing regions spanning the intermediate and high-level stages of the ventral pathway (Freiwald and Tsao 2010). How the face processing in IT is associated with that in these face-sensitive areas remains unclear. Further studies are required for understanding the neural mechanism of face perception.

For face representation in IT, it has been shown that faces are represented by a multidimensional 'face space', within which each point represents an individual face and the average of all faces resides at the center (Leopold et al. 2001, 2006; Valentine 1991). However, how each dimension and averaged face are organized remains unclear. Our

model may provide a clue for addressing this issue. The coarse information of faces encoded by ITB enables the ITFR to organize a face map, in which faces are categorized on the basis of the configuration of their parts. It may be possible to form the categorized face map using Kohonen's self-organized map (1995). Dominant region of ITFR contains the neurons sensitive to normal faces, because we frequently encounter normal faces. The region might correspond to the region sensitive to averaged faces. ITB activates ITFR neurons in a categorized region, and individual face is encoded by ITFR neurons within the categorized region by binding coarse information encoded by ITB with fine information encoded by ITM and ITF.

In this study, we focused on the relationship between spike dynamics and representation of face information in IT. However, object and face recognition are closely related with function of hippocampus and the adjacent perirhinal, entorhinal and parahippocampal cortices, involved in medial temporal lobe memory system. Recent study has shown that hippocampus plays a critical region in recognition and memory of novel objects (Bakker et al. 2008). It is of an important problem how the synchronous firing of IT neurons presented here is responsible for recollection of memory in hippocampus. Our model may provide a fundamental network model of IT to extend to a face recognition model including perception and memory system.

The rhythmic synchrony in the gamma frequency range (30–100 Hz) plays critical roles in sensory encoding and feature binding (Gray et al. 1989; Frie et al. 2007; Fujita et al. 2007). The importance of feature binding has been proposed as the binding-by synchronization hypothesis (Singer 1999; Singer and Gray 1995), which states that neurons forming a functional assembly are bound together by synchronization of their action potentials, a mechanism that is probably based, in part, on experience-dependent refinement of intra- and inter-areal connections. In human EEG experiments, stronger gamma coherence was induced in response to upright Monney face, which was perceived as a face, than to the inverted one, which was not perceived as a face (Rodriguez et al. 1999). In the present study, FOs elicited the synchronous firing of neurons in ITB, ITM and ITF layers, and the spike coincidece evoked in these layers activated effectively the neurons in ITFR. The spike frequency shown in Fig. 4 was nearly 80 Hz within the range of gamma frequency. The top-down signals emanating from ITFR also facilitated the synchronization of ITX neurons ($X = B, M, F$). The synchronous firing evoked in the IT layers can strongly combine the information about a holistic face and its parts, and produce highly pairwise correlations of neuronal spiking within ITF layer. The results presented here suggest that the binding-by-synchronization mechanism could be responsible for face representation in IT.

The spike synchronicity of IT neurons in face recognition, caused by the increased synaptic connections, may play a role in clinical practice of patients who suffer recognition deficit. The synchronous activity would not be observed for subjects that have a deficit of feature binding of face. In that sense, the synchronicity of IT activity could be a measure of face recognition, that is, a measure of the extent to which subjects recognize a whole face and its parts. Therefore monitoring the synchronicity might provide a useful insight into development of rehabilitation training of recognition deficit. Furthermore, our model may be extended to a detailed model incorporating the mechanism of neurotransmitters. The extended model, as well as pharmacological studies, might contribute to identification of neurotransmitters involved in synaptic learning and pharmacological treatment of face recognition deficit.

In our model, FOs were presented more frequently than NFOs, because we encounter various FOs more frequently than NFOs in everyday life, as far as we see faces. Our model is based on the idea that IT cortex has a region devoted to face processing, independently of a region for processing object information. However, IT cortex may also employ a common region for processing face and object information. In this case, we are seeing various NFOs (objects except for faces) more frequently than FOs. Alternative factor causing synchronous firing of IT neurons is a significance of face information in communication of human society. The significance might affect synaptic efficacy of IT neurons through the interaction between IT and other brain areas involving the reward system such as basal ganglia. The significance of face stimuli may be able to strengthen synaptic connections between IT neurons, even if we see FOs and NFOs at any rates.

The result shown in Fig. 8 is in agreement with the experimental result by Hirabayashi and Miyashita (2005). The result averaged for all pairs of IT neurons also showed the similar tendency to their experimental result. However, there is a discrepancy between the two results for all IT neuron pairs. The time course of the calculated cross correlation of IT neuron pairs in response to FO was rapidly increased up to 500 ms, and then remained nearly constant, whereas the corresponding result by Hirabayashi and Miyashita exhibited an increased cross correlation up to 500–600 ms, and then a significant decreased one. The firing rates of IT neurons in response to FOs and NFOs also have the discrepancy similar to the spike correlation. The cross correlation and firing rates exhibited slight decreases for a single pair of IT neurons, but they did significant decreases for all pairs. These discrepancies might come from the average of neuronal responses elicited by various sensitivities of IT neurons or the modulation of attention and interaction between IT cortex and other cortical areas.

For coarse-to-fine processing, in addition to the study by Sugase et al. (1999), it has been also reported that representation of a global shape develops before than that of a local shape (Sripati and Olson 2009). These suggest that top-down signal elicited by global signal affects the processing of local signal. There is also evidence that the processing of stimuli is controlled by top-down influences that strongly shape the intrinsic dynamics of the thalamo-cortical networks and constantly create predictions about forthcoming sensory events (Engel et al. 2001). Our model showed that the top-down signal from ITFR weakly activated the neurons in ITM and ITF before arrival of feed-forward visual signals and could function as a prediction signal of forthcoming visual stimulus. Coherence among subthreshold membrane potentials in ITM and ITF, elicited by the top-down signals, could be exploited to express the functional relationship between a holistic face and its parts during prediction. The top-down signal plays an important role in coarse-to-fine information processing of face and binding of both information. The role of the prediction would be more important when the ITFR must process a large number of face objects.

Appendix: The synaptic weights between neurons in V4X and ITX (X = B, M, F)

The synaptic weights between neurons in V4X and ITX (X = B, M, F) were determined by the learning rule of Kohonen's self-organized map (Kohonen 1995). The V4B has a retinotopic map consisting of $N_{V4B} \times N_{V4B}$ neurons, in which each neuron has the output of 0 or 1, depending on whether it receives the input elicited by a face stimulus, as shown in Fig. 2a. The synaptic weight of the connection from (i, j)th neuron in V4B to the neuron within kth unit in ITB layer, $w_{V4B-ITB}(k, ij)$, is updated by $\Delta w_{V4B-ITB}(k, ij)$, given by

$$\Delta w_{V4B-ITB}(k, ij) = \lambda_1 (X_{ij, V4B} - w_{V4B-ITB}(k, ij)) \text{ for } k = k^*, \quad (11)$$

$$k^* = \arg \min_k d_k \quad (12)$$

$$d_k = \sqrt{\sum_{ij} (X_{ij, V4B} - w_{V4B-ITB}(k, ij))^2}, \quad (13)$$

where $X_{ij, V4B}$ is the output of (i, j)th V4B neuron and λ_1 is the learning rate. The parameter values used were: $N_{V4B} = 11$, $\lambda_1 = 0.01$.

The V4M has a retinotopic map consisting of $N_{V4M} \times N_{V4M}$ neurons, which has the same size as V4B, as shown in Fig. 2b. The V4M neurons encode the information about the relative position of the face parts chosen by attention. When attention is directed to the position (i_0, j_0) included

in a face part, the output of (i, j)th V4M neuron, $X_{ij, V4M}$, is determined by

$$X_{ij, V4M} = \begin{cases} 1, & |i - i_0| \leq d, \text{ or } |j - j_0| \leq d \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

Then the synaptic weight of the connection from (i, j)th neuron in V4M to kth neuron within lth unit in ITM layer is updated by $\Delta w_{V4M-ITM}(kl, ij)$, given by

$$\Delta w_{V4M-ITM}(kl, ij) = \lambda_2 (X_{ij, V4M} - w_{V4M-ITM}(kl, ij)) \text{ for } (k, l) = (k^*, l^*), \quad (15)$$

$$(k^*, l^*) = \arg \min_{kl} d_{kl}, \quad (16)$$

$$d_{kl} = \sqrt{\sum_{ij} (X_{ij, V4M} - w_{V4M-ITM}(kl, ij))^2}. \quad (17)$$

The synaptic weights of ITM neurons within unit were slightly modulated so as to have different sensitivities to the output of V4M. The parameter values were: $N_{V4M} = 11$, $d = 1$, and $\lambda_2 = 0.01$.

The V4F consists of $N_{V4F} \times N_{V4F}$ hypercolumns, each of which contains N_L columns, as shown in Fig. 2c. These neurons in the columns encode local curvatures of a face part chosen by attention. The V4F represents only local features, in contrast to the representations of global face by V4B and V4M. Thus the local face features encoded by V4F is a complement to the face features encoded by V4B and V4M. The synaptic weights between ITF and V4F were determined by the learning rules similar to Eqs. (15–17). The parameter values used were: $N_{V4F} = 3$, $N_L = 10$, and the learning rate was set at 0.01.

References

- Afraz S-R, Kiani R, Esteky H (2006) Microstimulation of inferotemporal cortex influences face categorization. *Nature* 442:692–695
- Baker CI, Behrmann M, Olson C (2002) Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat Neurosci* 5:1210–1216
- Bakker A, Kirwan CB, Miller M, Stark CE (2008) Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science* 319:1640–1642
- Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4:2051–2062
- Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci* 2:704–716
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J Neurosci* 23:5235–5246
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2006) Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cereb Cortex* 16:1631–1644
- Freiwald WA, Tsao DY (2010) Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330:845–851

- Frie P, Nikolic D, Singer W (2007) The gamma cycle. *Trends Neurosci* 30:309–316
- Fujita I, Tanaka K, Ito M, Cheng K (1992) Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360:343–346
- Fujita K, Kashimori Y, Kambara T (2007) Spatiotemporal burst coding for extracting features of spatiotemporally varying stimuli. *Biol Cybern* 97:293–305
- Gallant JL, Braun J, Van Essen DC (1993) Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259:100–103
- Gray CM, König P, Engel AK, Singer W (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338:334–337
- Gross CG (2005) Processing the facial image: a brief history. *Am Psychol* 60:755–763
- Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the macaque. *J Neurophysiol* 35:96–111
- Hirabayashi T, Miyashita Y (2005) Dynamically modulated spike correlation in monkey inferior temporal cortex depending on the feature configuration within a whole object. *J Neurosci* 25:10299–10307
- Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* 117:500–544
- Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160:106–154
- Ito M, Komatsu H (2004) Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *J Neurosci* 24:3313–3324
- Kanwisher N, McDermott J, Chun M (1997) The fusiform face area: a module in human extrastriate cortex specialized for the perception of faces. *J Neurosci* 17:4302–4311
- Kashimori Y, Ichinose Y, Fujita K (2007) A functional role of interaction between IT cortex and PF cortex in visual categorization task. *Neurocomputing* 70:1813–1818
- Kiani R, Esteky H, Tanaka K (2005) Differences in onset latency of macaque inferotemporal neuronal responses to primate and non-primate faces. *J Neurophysiol* 94:1587–1596
- Kobatake E, Tanaka K (1994) Neural selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71:856–867
- Kobatake E, Wang G, Tanaka K (1998) Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J Neurophysiol* 80:324–330
- Kohonen T (1995) *Self-organized maps*. Springer, Berlin
- Leopold DA, O'Toole AJ, Vetter T, Branz N (2001) Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat Neurosci* 4:89–94
- Leopold DA, Bondar IV, Giese MA (2006) Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* 442:572–575
- Logothetis NK, Sheinberg DL (1996) Visual object recognition. *Annu Rev Neurosci* 19:577–621
- Logothetis NK, Pauls J, Poggio T (1995) Shape representation in the inferior temporal cortex in monkeys. *Curr Biol* 5:552–563
- Messinger A, Squire LR, Zola SM, Albright TD (2005) Neural representations of stimulus associations develop in the temporal lobe during learning. *Proc Natl Acad Sci USA* 98:12239–12244
- Miyashita Y (2004) Cognitive memory: cellular and network mechanisms and their top-down control. *Science* 306:435–440
- Pasupathy A, Connor CE (1999) Responses of contour features of neurons in cortical area MT in alert and anaesthetized macaque monkeys. *Nature* 414:905–908
- Perrett DI, Rolls ET, Caan W (1982) Visual neurons responsive to faces in the monkey temporal cortex. *Exp Brain Res* 47:329–342
- Perrett DI, Hietanen JK, Oram MW, Benson PJ, Rolls ET (1992) Organization and functions of cell responsive to faces in the temporal cortex. *Phil Trans R Soc B* 335:23–30
- Rodriguez E, George N, Lachaux J-P, Martinerie J, Renault B, Varela FJ (1999) Perception's shadow: long distance synchronization of human brain activity. *Nature* 397:430–433
- Rolls ET (1984) Neurons in the cortex of the temporal lobe and in the amygdale of the monkey with responses selective for faces. *Human Neurobiol* 3:209–222
- Rolls ET (1995) Learning mechanism in the temporal lobe visual cortex. *Behav Brain Res* 66:177–185
- Rosenthal O, Fusi S, Hochstein S (2001) Forming classes by stimulus frequency: behavior and theory. *Proc Natl Acad Sci USA* 98:4265–4270
- Sigala N, Logothetis NK (2002) Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415:318–320
- Singer W (1999) Neuronal synchrony: a versatile code for the definition of relation? *Neuron* 24:49–125
- Singer W, Gray CM (1995) Visual feature integration and temporal correlation hypothesis. *Annu Rev Neurosci* 18:555–586
- Soga M, Kashimori Y (2009) Functional connections between visual areas in extracting object features critical for a visual categorization task. *Vis Res* 49:337–347
- Sripati AP, Olson CR (2009) Representing the forest before the trees: a global advantage effect in monkey inferotemporal cortex. *J Neurosci* 29:7788–7796
- Sugase Y, Yamane S, Ueno S, Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400:869–873
- Szabo N, Stetter M, Deco G, Fusi S, Giudice PD, Mattia M (2006) Learning to attend: modeling the shaping of selectivity in inferotemporal cortex in a categorization task. *Biol Cybern* 94:351–365
- Tanaka K (1996) Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19:109–139
- Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque cerebral cortex. *Nat Neurosci* 6:989–995
- Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M (2001) Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat Neurosci* 4:832–838
- Turchi J, Saunders RC, Mishkin M (2005) Effects of cholinergic deafferentation of the rhinal cortex on visual recognition memory in monkeys. *Proc Natl Acad Sci USA* 102:2158–2161
- Ungerleider LG, Mishkin M (1982) Two cortical visual systems. In: Ingle DJ, Goodale MA, Mansfield RJW (eds) *Analysis of visual behavior*. MIT press, Cambridge, pp 549–586
- Valentine T (1991) A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quart J Exp Psychol A* 43:161–204
- Wan H, Warburton EC, Zhu XO, Koder TJ, Park Y, Aggleton JP, Cho K, Bashir ZI, Brown MW (2004) Benzodiazepine impairment of perirhinal cortical plasticity and recognition memory. *Eur J Neurosci* 20:2214–2224
- Winters BD, Bussey TJ (2005a) Transient inactivation of perirhinal cortex disrupts encoding, retrieval, and consolidation of object recognition memory. *J Neurosci* 25:52–61
- Winters BD, Bussey TJ (2005b) Glutamate receptors in perirhinal cortex mediate encoding, retrieval, and consolidation of object recognition memory. *J Neurosci* 25:4243–4251
- Winters BD, Bussey TJ (2005c) Removal of cholinergic input to perirhinal cortex disrupts object recognition but not spatial working memory in the rat. *Eur J Neurosci* 21:2263–2270

- Yamane Y, Tsunoda K, Matsumoto M, Phillips AN, Tanifuji M (2006) Representation of the spatial relationship among object parts by neuron in macaque inferotemporal cortex. *J Neurophysiol* 96:3147–3156
- Young MP, Yamane S (1992) Sparse population coding of face in the inferotemporal cortex. *Science* 256:1327–1331
- Zangenehour S, Chaudhuri A (2005) Patchy organization and asymmetric distribution of the neural correlates and face processing in monkey inferotemporal cortex. *Curr Biol* 15:993–1005