European Journal of Neuroscience, Vol. 20, pp. 1089–1100, 2004

The role of early visual cortex in visual integration: a neural model of recurrent interaction

Gustavo Deco[1] and Tai Sing Lee[2]
[1]Institute Catalana de Recerca i Estudis Avancats (ICREA) and Department de Tecnologia Universitat Pompeu Fabra Passeig de Circumval lacio 8, 09003, Barcelona, Spain
[2]Department of Computer Science & Center for the Neural Basis of Cognition, Room 115 Mellon Institute, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA

Keywords: biased competition, neural model, object recognition, primate, selective attention, visual cortex, visual search

Abstract

This paper presents a model on the potential functional roles of the early visual cortex in the primate visual system. Our hypothesis is that early visual areas, such as V1, are important for continual interaction among various higher order visual areas during visual processing. The interaction is mediated by recurrent connections between higher order visual areas and V1, manifested in the long-latency context-sensitive activities often observed in neurophysiological experiments, and is responsible for the re-integration of information analysed by the higher visual areas. Specifically, we considered the case of integrating 'what' and 'where' information from the ventral and dorsal streams. We found that such a cortical architecture provides simple solutions and fresh insights into the problems of attentional routing and visual search. The computational viability of this architecture was tested by simulating a large-scale neural dynamical network.

Introduction

Primary visual cortex has traditionally been considered as a processing module in a feedforward hierarchy, extracting local features, such as oriented edges or bars (Hubel & Wiesel, 1978) or performing a Gabor wavelet transform (Daugman, 1988; Lee, 1996), and then handing over the information to higher order visual areas for further processing. Significant progress has been made based on such feedforward schemes (Riesenhuber & Poggio, 2000). However, many recent neurophysiological experiments have demonstrated that the long-latency responses of V1 neurons reflects top-down attention (Motter, 1993; Hupe *et al.*, 1998; Roelfsema *et al.*, 1998; Ito & Gilbert, 1999) or other higher order perceptual context, such as 'figure-ground' (Lamme, 1995; Zipser *et al.*, 1996), sensitivity to subjective contour (Lee & Nguyen, 2001) and shape from shading (Lee *et al.*, 2002). These observed effects are presumably indicative of the influence of recurrent feedback from higher visual areas on V1. However, are these effects simply epiphenomena, a reflection of the heightened activities higher up, or do they serve a useful purpose? More generally, does V1 play a role beyond simple image processing?

In this work, we propose a neural architecture to explore these questions. Our basic thesis is that V1 might play a central role in integrating and coordinating computations amongst the higher visual areas utilizing the recurrent network connections in the visual system. Conceptually, similar ideas have been proposed earlier as the high-resolution buffer hypothesis (Mumford, 1996; Lee *et al.*, 1998). The rationale behind this hypothesis is that, because explicit and precise encoding of features and spatial information is only represented in V1 (and the LGN), higher level perceptual computations that involve high resolution details, fine geometry and spatial precision, such as the inference of abstract contour and shape, necessarily involve V1. Furthermore, as V1 is an area where all the information is implicitly available in retinotopic coordinates, it naturally provides a spatially registered common forum for all the higher order perceptual inferences to come back together. Thus, it could play the role of facilitating the integration of information from the different higher order modules.

Here, we investigated how V1 could serve to integrate and coordinate the computation of the identity (WHAT) and the location (WHERE) of objects in a visual scene, using the recurrent interaction between V1 and the dorsal and ventral streams. The system, as shown in Fig. 1, is minimal and serves primarily to illustrate the basic principle. It consists of three major modules, a ventral stream module (VM), a dorsal stream module (DM) and an early visual module (V1). Neurons in both the VM and DM are reciprocally connected to V1 neurons in many hypercolumns and thus have large receptive fields. Neurons in the DM compute and represent the positional information of an object, while neurons in the VM compute and represent the identity of objects. This decomposition of functions is consistent with the proposal of Ungerleider & Mishkin (1982).

In this model, for simplicity, computation within each module is mediated by winner-take-all competitive mechanisms using lateral inhibition. This, together with reciprocal connections between V1 and higher modules, implements the so-called biased competition mechanism as recently popularized by the work of Usher & Niebur (1996) and Reynolds *et al.* (1999) who used such a mechanism to model attention phenomena observed experimentally in V4 and inferotemporal cortex (IT). Our contribution is to extend this framework by bringing the dorsal stream and V1 into a recurrent interactive architecture for integrating 'what' and 'where' information.

Our simulations provide new insights into two long-standing problems: attentional routing and visual search. First, the dominant neural model for routing control is the shifter circuit (Olshausen *et al.*, 1993), which allows selection and channeling of information from

Correspondence: Dr Tai Sing Lee, as above.
E-mail: tai@cnbc.cmu.edu

Received 12 February 2004, revised 11 May 2004, accepted 25 May 2004

doi:10.1111/j.1460-9568.2004.03528.x

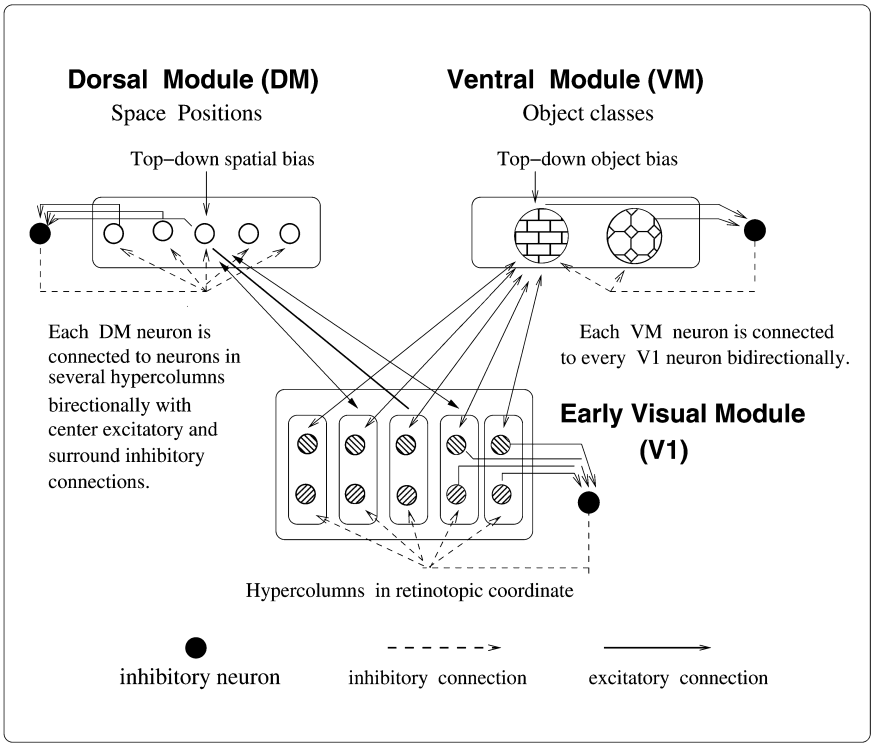

FIG. 1. The architecture of the proposed model depicting the three interacting modules of the system. The dorsal and ventral stream modules interact with V1 in this recurrent interactive architecture.

different retinotopic locations to the object recognition areas by dynamically modifying the synaptic weights of the feedforward connections. The shifter circuit provides an ingenious model for translation and scale invariant object recognition. While the shifter circuit is not impossible, it is rather complicated and primarily grounded on the idea of the router in electrical engineering. The model advanced here provides a much simpler and more neurally plausible alternative, in that no dynamic synaptic weight modification is needed in real time and routing is achieved by enhancement of V1 activities through recurrent interaction between V1 and the two streams.

A second insight provided by our model regards the possible mechanisms underlying visual search. Visual search has always been classified into serial search or parallel search in psychological literature as illustrated in Fig. 2 (Treisman & Gelade, 1980; Wolfe, 1998). When the target and distractors are different in their elementary features, the detection of the target is immediate and independent of the number of distractors. This suggests a parallel and 'preattentive' mechanism that can be implemented by the early retinotopic visual areas. On the other hand, when both target and distractors are composed of similar elementary features, the amount of time required to distinguish between them increases linearly with the number of distractors. This is said to suggest a serial attentional process. In our model, all of the bottom-up and top-down proposals about target are propagated and computed in parallel. Serial and parallel visual search phenomena could emerge from the same parallel mechanism without the need for two different and separate mechanisms.

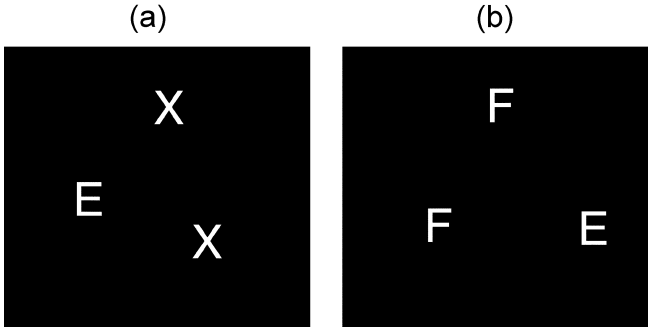

FIG. 2. Visual search can be classified into two types: (a) parallel search in which the time to find an E in a field of X's is constant regardless of the number of X's and (b) serial search in which the time to find an E in a field of F's increases linearly with the number of F's.

In the following sections, we will first describe the basic architecture and operations of the model and then discuss how the model can address the issues of routing and visual search in a unified framework. We will then examine and interpret some of the neurophysiological findings in the early visual areas in the context of this framework and venture some experimental predictions and lastly, compare this architecture with the shifter circuit and other attentional and cortical models. A preliminary version of some of these ideas has been presented in Deco & Lee (2002).

The model

The model presented is a minimalist model designed to illustrate how interactions between early and higher visual areas can mediate routing and search in a recurrent interactive architecture. The model is composed of three modules (as shown in Fig. 1): an early visual module (or simply V1), which conceptually includes V1 and several other early visual areas such as the LGN and V2, a ventral stream module (VM) and a dorsal stream module (DM). These three modules are reciprocally connected in a parallel hierarchy in accord with anatomical data (Felleman & Van Essen, 1991).

Each unit in the modules represents a pool of spiking neurons with similar tuning properties. The activity of the unit is described by a dynamical equation derived from the mean-field approximation (Wilson & Cowan, 1972; Amit & Tsodyks, 1991). The mean-field approximation consists of replacing the temporally averaged discharged rate of a cell with the instantaneous ensemble average of the activity of the neuronal pool that corresponds to the assumption of ergodicity. According to this approximation, the activity of a cell assembly $A(t)$, without external input, is given by

$$\tau \frac{\partial A(t)}{\partial t} = -A(t) + \mu F(A(t)) \tag{1}$$

where the first term on the right hand side is a decay term and the second term takes into account the excitatory recurrent stimulation among the excitatory neurons within the pool.

$$F(x(t)) = \frac{1}{(T_r - \tau \log(1 - \frac{1}{\tau x(t)}))} \tag{2}$$

is a nonlinear input–output function for a spiking neuron with deterministic input $x(t)$, membrane time constant τ and absolute refractory time T_r. Equation 1 was derived by Gerstner (2000) assuming adiabatic conditions, i.e. that the activity changes slowly compared with the typical interval length.

Early visual module

The units in the V1 module are modelled as complex cells. The bottom-up input to a complex cell is given by the energy or the power modulus of the Gabor filters (Pollen *et al.*, 1989),

$$\begin{aligned} I^{E}_{mlpq} &= \sqrt{\left\| \langle G_{mlpq}, \Gamma \rangle \right\|^2} \\ &= \sqrt{\left\| \sum_{i=1}^{n} \sum_{j=1}^{n} G_{mlpq}(i,j) \Gamma(i - \frac{n}{2} + 2p, j - \frac{n}{2} + 2q) \right\|^2} \end{aligned} \tag{3}$$

where Γ is the input image and n is the number of pixels in the image covered by the receptive field of the cell along each dimension. G_{mlpq} is a family of Gabor wavelets (Lee, 1996) defined below

$$G_{mlpq} = a^{-m} \psi_{\theta_1}(a^{-m}(x - 2p) - a^{-m}(y - 2q)) \tag{4}$$

$$\psi_{\theta_1} = \psi(x \cos(l\theta_o) + y \sin(l\theta_o), -x \sin(l\theta_o) + y \cos(l\theta_o)) \tag{5}$$

and ψ, called the mother wavelet, is given by

$$\psi(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{8}(4x^2 + y^2)} \cdot \left[e^{ikx} - e^{-\frac{k^2}{2}} \right] \tag{6}$$

The whole family of 2D Gabor wavelets G_{mlpq} is generated by rotating and scaling the mother wavelet. In the above equations, $\theta_o = \pi L$ denotes the step size of each angular rotation where L is the number of orientations in the family of filters, l is the index of rotation, generating a preferred orientation $\theta = l\pi L$ and m denotes a scaling factor.

In our implementation, $a = 2$, $L = 8$ and $m = 1$, 2 and 4 are used, generating filters spanning three octaves with one octave step in scale (spatial wavelength, 2, 4 and 8 pixels) and eight orientations for each scale; $(x, y) = (2p, 2q)$ are the positions of the receptive field center. The input image Γ is a 66×66 pixel gray-level image with intensity values ranging from 0 to 255. The V1 module is composed of a lattice of 33×33 hypercolumns. Each hypercolumn contains 24 (eight orientations $\times$ three scales) excitatory units (pools). One inhibitory pool per scale is used to mediate competition among neurons (of different orientations and positions) within each scale. A total of three inhibitory pools are used.

The activity level of an excitatory pool unit in V1 is given by

$$\begin{aligned} \tau \frac{\partial A^{V1}_{mlpq}(t)}{\partial t} =& -A^{V1}_{mlpq}(t) + \mu F(A^{V1}_{mlpq}(t)) - \gamma F(A^{V1,I}_{mlpq}(t)) + I^{V1,E}_{mlpq}(t) \\ &+ I^{V1-DM}_{pq}(t) + I^{V1-VM}_{mlpq}(t) + I_o + v(t) \end{aligned} \tag{7}$$

where m, l, p and q are indices of scales, orientation selectivity and (p, q) are indices of spatial location in the retinotopic hypercolumn coordinate. The first two terms are decay and self-excitation as before and the third term is the competitive interaction ($\mu = 0.95$, $\gamma = 0.8$). The fourth term is the input current to the complex cell, the fifth term is the feedback from the DM to V1 and the sixth term is the feedback from the VM to V1. These feedback terms will be defined later when we describe the DM and VM. I_o is an additive Gaussian noise input to the unit, drawn from a normal distribution with zero mean and $\sigma = 0.02$. $I_o = 0.025$ is a bias current representing diffuse spontaneous background input to the unit. Note that when the same symbols are used in the subsequent equations, the parameter values are kept the same. The I or E in the exponents of the different terms indicates the activity of inhibitory or excitatory neurons, respectively.

The activity level of the inhibitory unit is given by

$$\tau_I \frac{\partial A^{V1,I}_{m}(t)}{\partial t} = -A^{V1,I}_{m} + \lambda F(A^{V1,I}_{m}(t)) + \kappa \sum_{l,p,q} F(A^{V1}_{mlpq}(t)) \tag{8}$$

where m is the scale index, $\kappa = 0.1$, $\lambda = 0.1$ and $\tau_I = 7$ ms. The first two terms are decay and self-excitation and the third is a function of the sum of the activities from all the excitatory pools at a particular scale within the whole module. This inhibitory pool receives input from all excitatory neurons from a particular scale and inhibits those neurons uniformly.

There are 26136 excitatory cell pools in the 33×33 hypercolumns. They are necessary to completely encode features and positions within an image (Lee, 1996). While inhibitory connections are more local and fine tuned for implementing precise computations in real V1 and V2 neuronal circuits, we found that one inhibitory neuron per scale is sufficient to mediate the basic competition between neurons in the same scale to produce the effect that we are seeking.

Dorsal visual module

The DM encodes spatial location and computes object location in the spatial domain. Spatial attentional selection is initiated by biasing a particular unit in the DM. The DM is a lattice of 66×66 units, each of which receives input from a number of the V1 hypercolumns. All the excitatory units in a spatial neighborhood (5×5 hypercolumns) in V1 are connected to a particular neuronal pool in the DM with a center-excitatory, surround-inhibitory weight profile. The connection weight between a DM unit at location (i, j) and a V1 unit at location (p, q) in the V1 lattice of a particular scale m and orientation l is defined by

$$W_{pqij} = C\mathrm{e}^{-\frac{(i-2p)^2+(j-2q)^2}{2\sigma_w^2}} - B \tag{9}$$

where $C = 1.5$ and $B = 0.5$. The weight is most positive when $i = p$, $j = q$ (i.e. the same retinotopic position) and decays in a Gaussian fashion spatially, with $\sigma_w = 2$, covering a spatial area of about five V1 hypercolumns along each dimension. As the largest receptive field size in each V1 hypercolumn extends to about 8×8 pixels, the diameter of the receptive field (defined as the 2 SD envelope of the Gaussian connections) of a DM neuron is about 17 pixel units.

The activity of an excitatory unit in the DM is given by

$$\begin{aligned}\tau\frac{\partial A_{ij}^{\mathrm{DM}}(t)}{\partial(t)} =& -A_{ij}^{\mathrm{DM}}(t) + \mu F(A_{ij}^{\mathrm{DM}}(t)) - \gamma F(A^{\mathrm{DM,I}}(t)) \\ &+ I_{ij}^{\mathrm{DM-V1}}(t) + I_{ij}^{\mathrm{DM,A}}(t) + I_{\mathrm{o}} + v(t)\end{aligned} \tag{10}$$

Most term definitions are similar to those of the V1 unit. $I_{ij}^{\mathrm{DM,A}}$ is the top-down external attentional bias that can be imposed on the unit and $I_{ij}^{\mathrm{DM-V1}}$ is the input from V1.

The feedforward input $I_{ij}^{\mathrm{DM-V1}}$ from the V1 to a DM pool at location (i, j) is given by

$$I_{ij}^{\mathrm{DM-V1}}(t) = \sum_{m,l,p,q} W_{pqij} F(A_{mlpq}^{\mathrm{V1}}(t)) \tag{11}$$

The feedback from the DM to V1 is mediated by connections with weights specified by a Gaussian distribution as given below

$$I_{pq}^{\mathrm{V1-DM}}(t) = 0.6\sum_{i,j} W_{ijpq} F(A_{ij}^{\mathrm{DM}}(t)) \tag{12}$$

where

$$W_{ijpq} = A\mathrm{e}^{-\frac{(p-i/2)^2+(q-j/2)^2}{2\sigma_w^2}} - B \tag{13}$$

Note that the feedback to V1 is 0.6 of the feedforward connection in weight. The weaker feedback is important as it prevents V1 from being a mere slave to a higher order bias and allows lower level representations to change the 'opinion' of the higher order modules. Neurophysiologically, the efficacy of the feedback stimulation at evoking a postsynaptic response has also been found to be smaller than the efficacy of the feedforward connection (Salin & Bullier, 1995).

In this module, there is only one inhibitory cell pool that receives input from all excitatory pools and feeds back to inhibit every pool uniformly. This is sufficient for mediating the winner-take-all competition in the DM. Its activity is given by

$$\tau_I \frac{\partial A^{\mathrm{DM,I}}(t)}{\partial t} = -A^{\mathrm{DM,I}} + \lambda F(A^{\mathrm{DM,I}}(t)) + \kappa \sum_{ij} F(A_{ij}^{\mathrm{DM}}(t)) \tag{14}$$

Ventral visual module

The VM encodes object class or categorical information. Selection of a winner within the VM through competition corresponds to the identification of an object in the visual image. The VM contains a finite set of units, storing V1 response patterns (in the learned connection) on a number of objects. These include the tower and sculpture in the Paris scene (Fig. 3), the letters E, F, X, T and L and bars of different orientations. Each unit is connected to all neuronal units in V1.

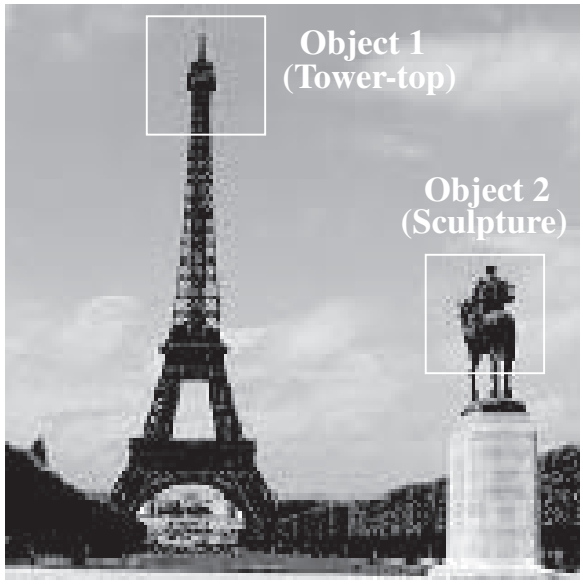

FIG. 3. A Paris scene, one of the test images used. The system has been trained by Hebbian learning to recognize the image of the tower top and sculpture in this scene.

The activity of an excitatory VM unit, coding object category c, is given by

$$\begin{aligned}\tau\frac{\partial A_c^{\mathrm{VM}}(t)}{\partial t} =& -A_c^{\mathrm{VM}}(t) + \mu F(A_c^{\mathrm{VM}}(t)) - \gamma F(A^{\mathrm{VM,I}}(t)) \\ &+ I_c^{\mathrm{VM-V1}}(t) + I_c^{\mathrm{VM,A}}(t) + I_{\mathrm{o}} + v(t)\end{aligned} \tag{15}$$

where $I_c^{\mathrm{DM,A}}$ is the top-down external attentional bias imposed on the VM pool for object class c and $I_c^{\mathrm{DM,A}}$ is the forward input from V1 to the VM unit c and is given by

$$I_c^{\mathrm{VM-V1}}(t) = \sum_{m,l,p,q} w_{cmlpq} F(A_{mlpq}^{\mathrm{V1}}(t)) \tag{16}$$

The feedback from the VM to V1 is also mediated by symmetrical but attenuated reciprocal connections

$$I_{mlpq}^{\mathrm{V1-VM}}(t) = 0.6\sum_{c=1}^{N_c} w_{cmlpq} F(A_c^{\mathrm{VM}}(t)) \tag{17}$$

The activity of the inhibitory VM unit is given by

$$\tau_{\mathrm{I}} \frac{\partial A^{\mathrm{VM,I}}(t)}{\partial(t)} = -A^{\mathrm{VM,I}} + \lambda F(A^{\mathrm{VM,I}}(t)) + \kappa \sum F(A_c^{\mathrm{VM}}(t)) \tag{18}$$

The memory of a particular object class c is encoded in the connection weight w_{cmlpq} between the VM unit (c) and the V1 units (m, l, p, q). The connection weights are trained by supervised Hebbian learning; the image containing the target is presented in V1 while a top-down bias is imposed on the VM unit coding for that object and a top-down bias is imposed on the DM unit coding for the spatial location where the object appears in the image. The active DM unit highlights the

corresponding hypercolumns in V1. The coactivation of the corresponding parts of V1 and the VM reinforces the association of the VM pool c and the appropriate V1 pools. The system is allowed to settle into a steady state for each presentation of the stimulus and the top-down bias signals. After convergence, all the relevant V1–VM connections are updated using the following Hebbian learning rule

$$\delta w_{cmlpq} = \eta F(A_c^{\mathrm{VM}}(t)) F(A_{mlpq}^{\mathrm{V1}}(t)) \quad (19)$$

where δw is the change in weight and η is the learning coefficient. A_c^{VM} is the activity of a neuronal unit for object c in the VM and A_{mlpq}^{V1} is the activity of a pool in V1 with a particular spatial frequency and spatial and orientation tuning. The image patch for each object is presented and learned at every possible position in the input coordinate. Thirty different presentations of each object per position are required to achieve convergence of the V1–VM connection weights.

A more complete hierarchy that resembles the ventral stream would consist of a cascade of areas each linked by local connections. Here we limit the VM–V1 interactions to a two-layer recurrent network because the Hebbian learning algorithm is better understood in this context. Note that most of the parameters are independent of our models, characterizing generic neuronal parameters such as membrane time constants. The parameters are chosen so that the system will arrive at a stable solution with a clear winner in both the VM and DM. Once chosen, the parameters are used throughout all simulations. The model does not critically depend on these parameters. This is the main justification for using the simplest model.

Results

The current system can operate in three modes: (i) spatial attention mode when a top-down bias is imposed on a DM unit; (ii) object attention mode when a top-down bias is imposed on a VM unit and (iii) preattentive mode in which no top-down bias is imposed. We will use the processing of the Paris scene (Fig. 3) to illustrate how the system works in these scenarios (Fig. 4). We will then compare the pool activities of the model with neural activities observed in V1 and V4 in electrophysiological experiments.

Spatial attention and routing

In the spatial attention mode (Fig. 4a), when a DM unit is excited by a top-down signal presumably coming from the executive control area in the prefrontal cortex, this positional bias signal will propagate down to activate and enhance the corresponding units in V1 as if there is a spatial attention beam. When an image (Fig. 3) is presented in a sustained fashion, V1 units will be excited by both the bottom-up input signals and the top-down DM signals. All the V1 neurons project their signals simultaneously and in parallel to all the units in the VM. Those that are biased positively by the DM units will provide a stronger input to the VM units, whose connection weights best match their activities pattern ultimately leading to the dominance of this VM unit over other VM units through the winner-take-all mechanism. As this winner unit is coding a particular object, the dominance of its response over other VM units corresponds to the system's recognition of the object at the location as highlighted by spatial attention.

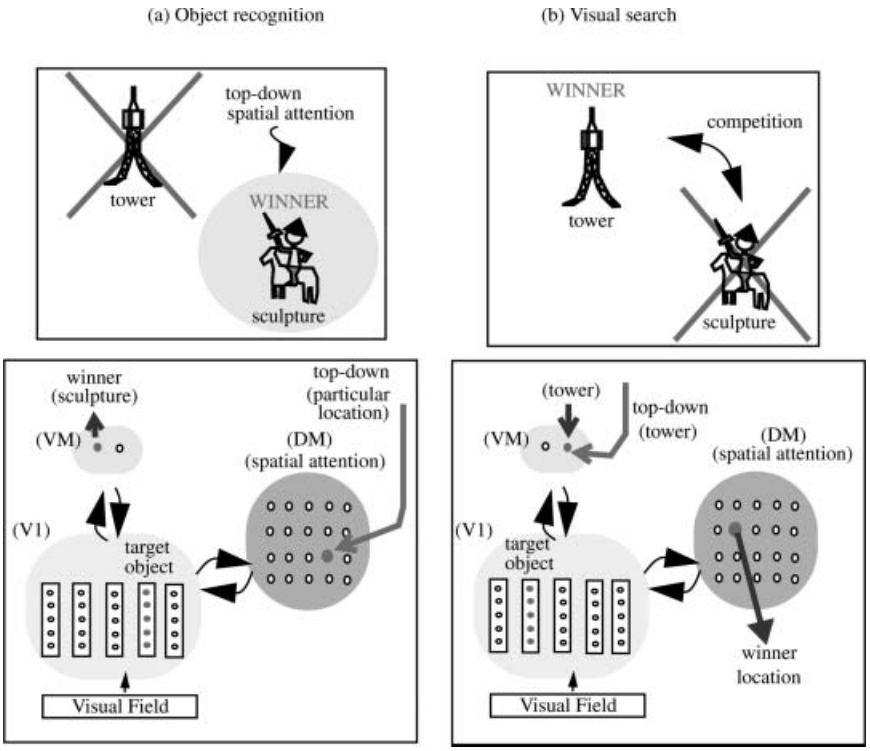

FIG. 4. (a) Recognizing the sculpture in the attended location. In the spatial attention mode, a top-down bias on a dorsal module unit covering the sculpture location can enhance the responses of the V1 units at the sculpture's location. The highlighted V1 (early module) hypercolumns suppress the other hypercolumns and provide stronger input to the sculpture neuron in the ventral module. Due to the stronger bottom-up bias, the sculpture unit in the ventral stream module (VM) becomes the winner through competition, corresponding to recognition of the object in the attended area. (b) Finding the tower in the scene. In the object attention mode, a top-down bias activates the tower unit in the ventral module. The tower unit in the VM back-projects the activity pattern associated with the tower image to all V1 units. The V1 hypercolumns that encode the tower image will be enhanced by the feedback from the VM. V1 units project their activities in parallel to the dorsal stream module (DM). The enhanced V1 units, however, exert a greater influence, resulting in a contraction of activity in the dorsal module's lattice to the location of the searched object, i.e. the tower.

Figure 5 shows the temporal evolution of averaged population activities in the form of a spatial map in the three different modules during spatial attention to the sculpture location in the Paris scene. We found that spatial attention bias can increase the spontaneous firing rate of the V1 units at the 'attended' location when it is applied 100 ms before stimulus onset, as observed by Luck *et al.* (1997). When the bottom-up input arrives at V1, the neuronal activity at the sculpture location is enhanced, providing a stronger drive to the VM sculpture unit and helping it to overcome the tower neuron and other object neurons in the VM. Conversely, when the spotlight is directed toward the tower location, the tower neuron eventually becomes the winner. These results demonstrate the effectiveness of the routing mechanism based on recurrent enhancement of V1 activities.

Our scheme of routing is distinct from the shifter circuit which mediates spatial attention by dynamically modifying synaptic connections along the ventral stream hierarchy with the pulvinar as the routing controller. In our scheme, no synaptic modification or delicate control is required; routing is accomplished simply by recurrent interaction between V1 and the DM and VM modules. Thus, this seems to furnish a simpler and more neurally plausible explanation. We will show later that this mechanism is sufficient to account for several spatial attentional phenomena observed in V4 and IT.

Object attention and visual search

In the object attention mode (Fig. 4b), a bias to a specific VM pool specifies the object to be searched in the visual scene. Let us illustrate how the system performs search with the following examples. To search for the sculpture in the Paris scene, a bias is imposed on the sculpture pool unit in the VM. As each VM unit is reciprocally connected to all units in V1, the activity of the biased VM unit effectively back-projects a response pattern that is associated with the image of sculpture simultaneously across all retinotopic locations in V1 in parallel. The V1 hypercolumns covering the sculpture location will resonate best with the feedback signals and their units' responses will be enhanced. The DM units at the sculpture location will, therefore, be activated more strongly because of the enhanced feedforward projection from V1. The winner-take-all mechanism in the DM will then be biased to choose the unit at the sculpture location to be the winner in the DM map (Fig. 6). When the VM tower neuron receives a top-down bias, the same mechanism will select the DM unit coding for the tower location to be the winner (Fig. 7). This demonstrates the visual search capability of the system.

Next, we consider the phenomena of serial and parallel search. Figure 8a contains an E in a field of X's. As the elementary features in E and X are quite distinct in orientation, V1 neurons of different orientation columns are activated. Hence, the difference between E and X is evident even at the level of V1 responses, allowing the E to pop out from the X's readily, independent of the number of X's in the image. This suggests that this computation is parallel, hence the name parallel search. On the other hand, Fig. 8b contains an E in a field of F's. These two characters share features of similar orientations and are detected by the same class of neurons in V1. Hence, E does not pop out readily from the F distractors. In human psychophysical experiments, the time required to localize E in a field of F's increases linearly with the number of F's. It has been suggested that this linear increase in time is a result of the engagement of a serial attentional search mechanism.

Interestingly, we found that both the serial and parallel search phenomena emerge from the same mechanism in this recurrent interactive architecture. We monitor the time required for the DM units at the E location to become the winner when the VM unit coding for E is biased. The time required for the difference between the maximum activities at the target location and the maximum activities in all the distractor locations (called polarization) to exceed a certain threshold is considered the search time, to be compared with the visual search time in psychophysical experiments. We found that the time for the system to search for an E in a field of X's is basically constant but the time to search for an E in a field of F's increases linearly at the rate of 25 ms per distractor (Fig. 8c) and to search for an E in a field of F's increases linearly at the rate of 25 ms per distractor (Fig. 8c and d). We also tested the system searching for L in a field of X's or in a field of T's. We found that the time required to find L is independent of the number of X's but increases linearly with the number of T's (Fig. 8e) for up to 16 T distractors.

Intuitively, E and F (or L and T) share very similar features and hence their differences cannot be detected by VM by one feed forward pass, as in the case of discriminating E or L from X's. However, the ambiguity between E and F (or between L and T) can be resolved by recurrent interaction between V1 and higher areas, resulting in additional time cost. The interaction between V1 and higher areas produces the emergent phenomenon of feature integration. The behavior of the model is consistent with the proposal of Duncan & Humphreys (1989) that serial visual search can potentially be solved by a parallel competitive mechanism. Exactly why the search time increases 'linearly' with the number of distractors in this system, however, is not understood and requires more investigation.

Comparisons with neurophysiological observations

Is this model relevant to our understanding of the visual system? By design, our model attempts to explain a potential functional role of the known feedback connections in the visual cortex but do the units in this system behave in a similar way as the neurons observed in electrophysiological experiments? We studied the temporal responses of the model's units and found that they are qualitatively similar to the observed temporal evolution of the responses of cortical neurons in several aspects.

The V1 units in our model exhibited a long-latency enhancement effect under top-down attentional modulation. Figure 9 shows the effect of top-down attention on V1 activities of neurons in both the object and spatial attention scenarios. Figure 9a compares the responses of V1 units with and without top-down object attention for the sculpture (Fig. 6). We noted that the initial (40–80 ms) responses of the V1 units are similar for the two cases but the units' responses are enhanced in the attention case at around 90 ms poststimulus onset. Note that the top-down attention signals and the bottom-up signals arrive at the V1 units at the same time. Nevertheless, it takes another 50 ms before the attentional effect becomes evident, suggesting that this contextual enhancement effect is not simply due to a top-down bias but rather involves continuous recurrent interaction of V1 with both the DM and VM modules in conjunction with competition within V1 itself.

Figure 9b shows that a similar delay in attentional modulation can also be seen in the case of spatial attention. Note that the attentional effect occurs earlier in V1 in Fig. 5 because the top-down attention was applied 100 ms prior to stimulus onset. For stimulus-triggered spatial attention, as in this case, we have assumed that the top-down signal arrives at V1 at the same time as the bottom-up signal. This delay in attentional enhancement effect resembles the long-latency contextual modulation effects observed in the early visual cortex (Lamme, 1995; Zipser *et al.*, 1996; Lee *et al.*, 1998, 2002). In those experiments, it was found that the responses of V1 neurons were significantly enhanced when

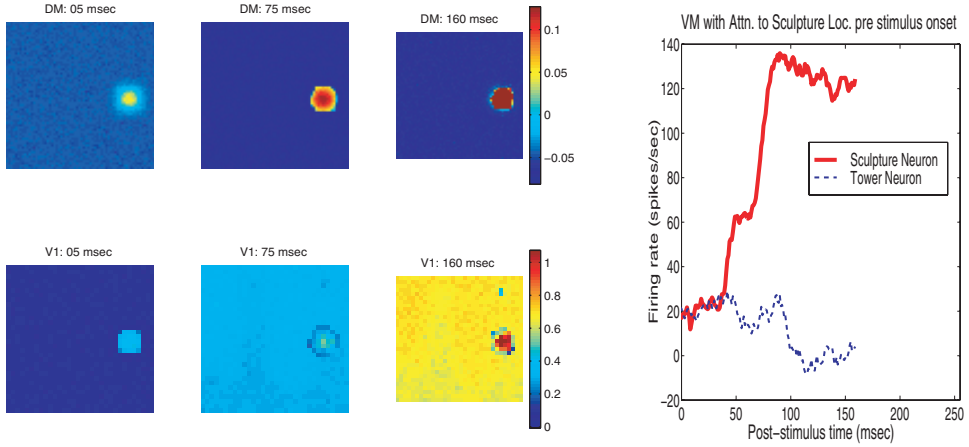

FIG. 5. Effect of spatial attention at the sculpture location. The average population activities at retinotopic maps of V1 and the dorsal module at three time points poststimulus onset show that top-down bias imposed on the dorsal stream module (DM) unit at the sculpture location enhances V1 activities at that location, which leads to an increase in activity of the sculpture unit in the ventral stream module (VM), ultimately making it the winner. The signal at each point in the V1 map is the sum of all the neural activities within the corresponding hypercolumn.

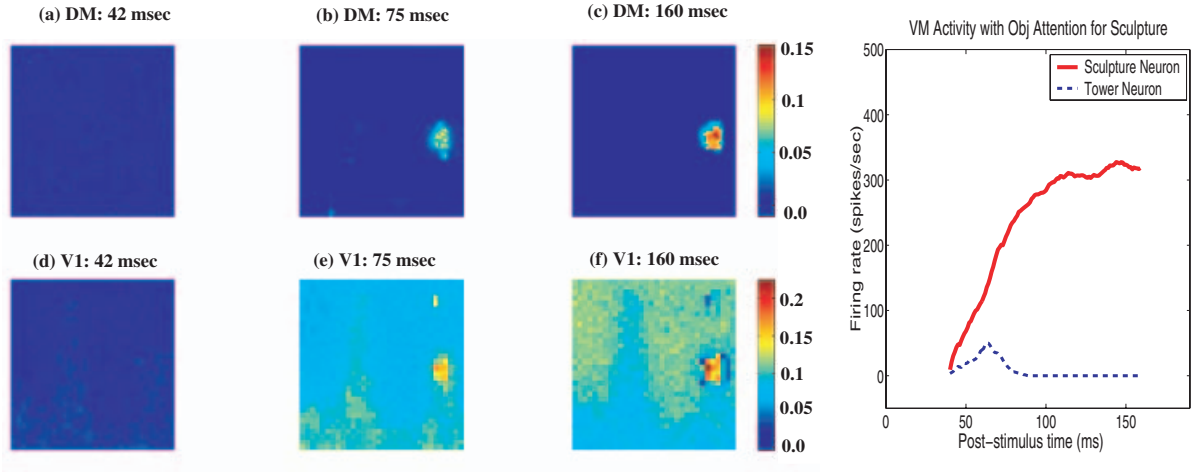

FIG. 6. Effect of object attention bias for the sculpture. The average population activities at retinotopic maps of V1 and the dorsal module at three time points poststimulus onset show that top-down bias imposed on the ventral stream module (VM) sculpture unit enhances V1 activities at the sculpture location by resonance which, in turn, allows the dorsal stream module (DM) unit at the sculpture location to overcome the other locations, corresponding to the localization of the sculpture in the scene.

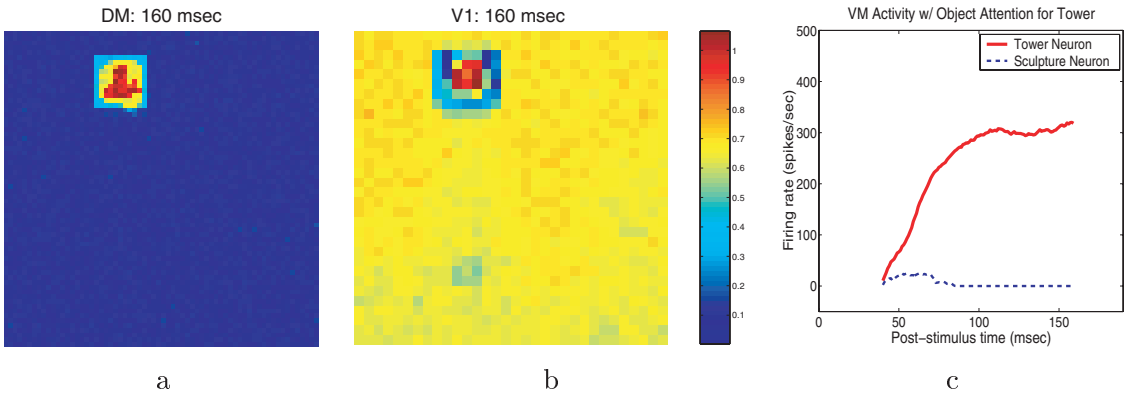

FIG. 7. Effect of object attention bias for the tower. The steady state responses of the three modules are shown. DM, dorsal stream module; VM, ventral stream module.

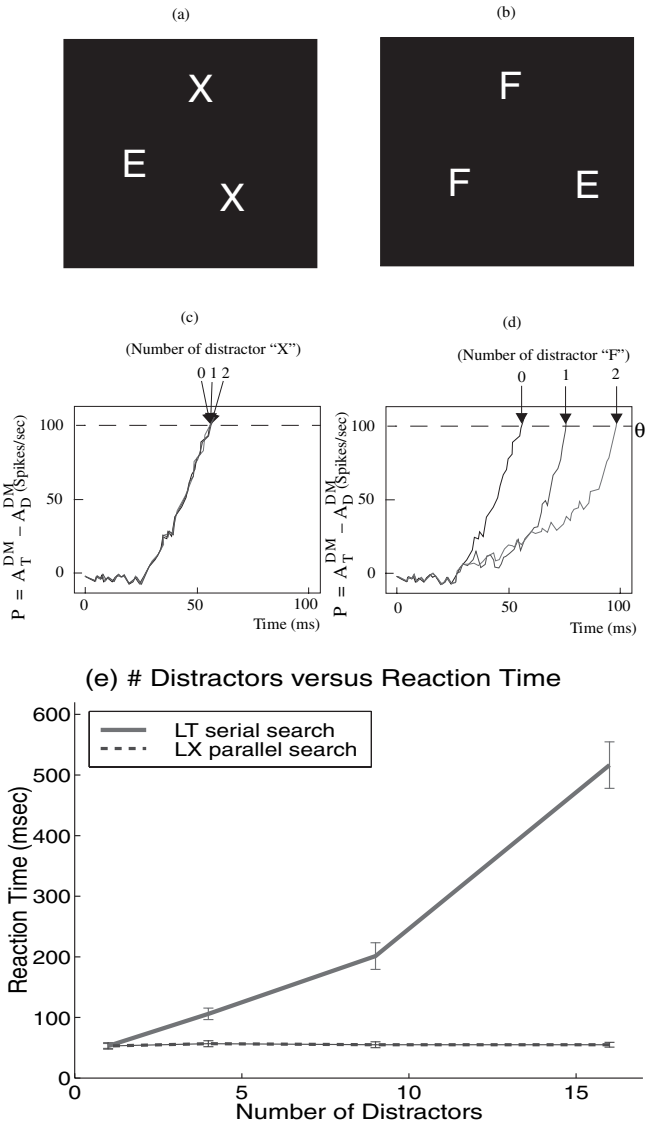

FIG. 8. Serial and parallel search emerge from the same mechanism. (a and b) Parallel and serial search stimuli, respectively. (c and d) Difference in the activation at the target location relative to the distractor locations as time evolves. The time required for this signal to cross threshold is assumed to be related to the time required by humans to find the target. The time required to find the E in X's is constant while the time required to find the E in F's increases linearly with the number of F's. (e) The visual search experiment is repeated for other stimuli with a large number of distractors. Finding the L in X's requires constant time (parallel), while finding the L in T's increases linearly with the number of T's (serial). DM, dorsal stream module.

their receptive fields were situated within a figure in a visual scene compared with when they were located in the background (Lamme, 1995; Zipser *et al.*, 1996; Lee *et al.*, 1998) or as a part of the distractors (Lee *et al.*, 2002). Lee *et al.* (1998) have suggested that this figure-ground effect, as reproduced in Fig. 9c, might be a highlighting signal that colors the figure to enhance its saliency. Our results suggest that the mechanisms underlying this long-latency contextual enhancement effect might be related to spatial and/or object attention feedback mechanisms. Further, this evidence suggests that such an enhancement effect might not simply be reflecting perceptual saliency *per se* (Lee *et al.*, 2002) but might be serving a deeper purpose of allowing the different higher order visual areas to communicate and coordinate their computations through recurrent interactions. Interestingly, in our simulation, we found that the observed spatial or object attentional enhancement is stronger for weaker stimuli; this surprising result has been recently confirmed experimentally by the work of Reynolds *et al.* (1999).

Figure 10a and b reveals similarities between the responses of the VM units and activities of V4 or IT neurons observed in electrophysiological experiments on spatial attention. Desimone and colleagues (Moran & Desimone, 1985; Chelazzi *et al.*, 1993; Reynolds *et al.*, 1999) had observed the following phenomena in V2, V4 and IT neurons. Consider a neuron that prefers a vertical over a horizontal bar in its receptive field in Fig. 10a. When stimulated by a vertical bar alone, the neuron responds well. When stimulated by a horizontal bar alone, the neuron responds poorly. When both the vertical and the horizontal bar are present inside the receptive field, the neuron responds moderately. This suggests that the presence of the horizontal bar exerts an inhibitory effect. It is believed that a horizontal neuron at the same spatial location competes with the vertical neuron, thus suppressing its activity. When the animal is cued to pay attention to the vertical bar or the location of the vertical bar, the neuron's response is restored to its earlier rigor as if the horizontal bar does not exist. This has been attributed to a top-down bias imposing on the vertical neuron, compensating for the inhibition from the horizontal neuron. The behavior of the vertical VM neuron in our model exhibits a qualitatively similar phenomenon when it is stimulated by the same set of stimuli (Fig. 10b). It is worth noting that spatial attention shrinks the effective receptive field of the VM unit. These similarities to neurophysiological findings add to the plausibility of our model as a reasonable approximation to the biological system, suggesting that the attentional effect observed in V1 and higher ventral visual areas might, in part, be coordinated by the dorsal stream via V1.

Discussion

The main hypothesis that this paper explores computationally is that early visual cortex can act as a high-resolution buffer for the dorsal and ventral streams to integrate 'higher level' spatial and object information. Our work demonstrates that interaction of 'what' and 'where' information can occur early in the visual system and that recurrent interaction between higher order areas and the early visual areas, such as V1 and V2, may play an important role in mediating visual search and attentional routing.

On the issue of visual search, we demonstrate that a parallel recurrent interactive mechanism is sufficient to produce the so-called serial and parallel effects in visual search. When the component features of compound objects are similar, more time is apparently needed for recurrent interaction with the higher area to disambiguate the different objects in the visual scene. In this context, feature integration (Treisman & Gelade, 1980) can be thought of as a process emerging from recurrent interaction between early visual cortex and the various extrastriate areas, rather than a process in some intermediate stages in the visual hierarchy. However, this parallel dynamic is limited to processing within each fixation and should thus be limited in spatial scope. The visual system clearly moves the eyes to scan and search for objects in a visual scene in a serial fashion.

On the issue of routing, this model provides a simpler alternative to the shifter circuit (Olshausen *et al.*, 1993). The shifter circuit achieves routing by dynamically selectively modifying the feed-forward synaptic connections from V1 to IT. We show that a recurrent interactive architecture can produce this dynamic routing simply by enhancing and suppressing neural responses in the early visual cortex, such as V1. Our

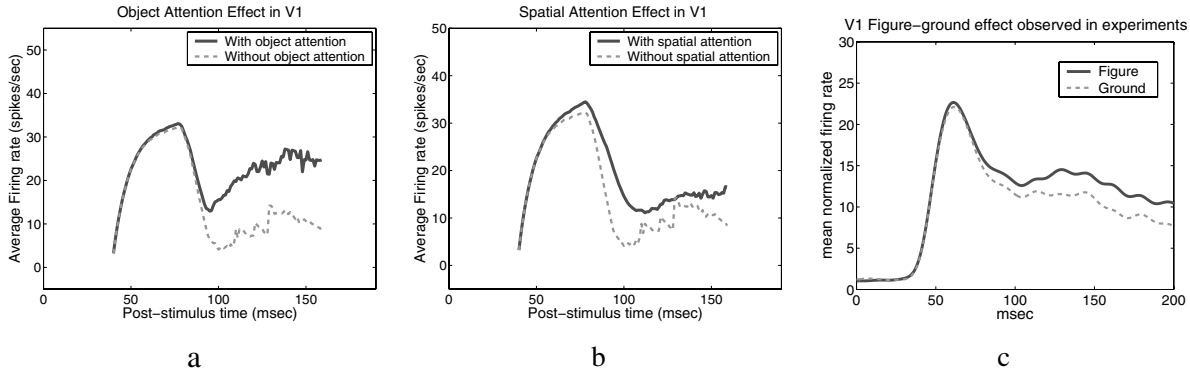

FIG. 9. Delayed attentional effect in V1. (a) Temporal evolution of averaged responses of V1 units at the location of the sculpture (4 × 4 hypercolumns) with and without object attention bias at the ventral stream module sculpture neuron. Attentional bias and bottom-up input arrive at the same time to V1 units but it is not until 90 ms after stimulus onset that the attentional enhancement becomes significant in V1 units. (b) Temporal evolution of averaged responses of V1 units at the location of the sculpture with and without spatial attention bias to the corresponding dorsal stream module unit. Attentional bias and bottom-up input arrive at V1 units at the same time (40 ms) but it is not until 90 ms after stimulus onset that the attentional enhancement becomes significant in V1 units. Both spatial and object attention take time for the effect to become evident in V1. (c) Temporal evolution of averaged responses of V1 units when their receptive fields were inside a texture figure (Figure) vs. when their receptive fields were in a texture background (Ground) as observed in a single-unit experiment in macaque V1. The difference in the responses to the figure and ground emerged at about 80 ms poststimulus onset. Stimulus-triggered spatial and/or object attention is a sufficient, but not necessarily the only one, explanation for the figure-ground phenomenon. Adapted from Fig. 8 in Lee *et al.* (1998).

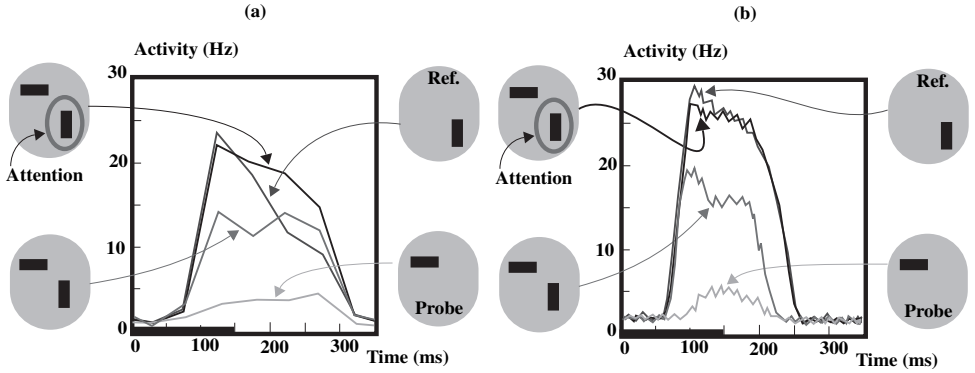

FIG. 10. Biased competition effect in the ventral module. (a) A sketch of experimental findings from Reynolds *et al.* (1999) showing that spatial attention can eliminate the competition effect introduced by the nonpreferred stimulus. (b) Response of the vertical unit in the ventral stream module module is similar to the experimental observations. The stimulus is assumed to be presented for 150 ms (thick horizontal bar). The time constant τ is increased from 7 to 15 ms to make the temporal response profiles more similar to the experimental data. The effect is, however, qualitatively the same for the shorter time constant.

model is unique in three aspects: (i) the dorsal stream and the early visual cortex are involved in our model for mediating spatial attention; (ii) the computation in the ventral stream is completely parallel and (iii) top-down position signals and top-down object pattern signals interact through V1. The combination of these ideas makes routing possible without dynamic modification of the synapses.

A shortcoming of our current model is that it requires the presentation of all the training examples in all spatial locations during the training process. This makes the training process unnecessarily long and redundant as all the V1–IT connections are essentially learning the same weights. A simple weight replication algorithm that generalizes the connections learned at one location to all locations can solve this problem, although how this replication is carried out by the visual system is an open question. Another shortcoming of our model is that, for simplicity, we have modelled the VM pathway as a two-layer network rather than as a hierarchy. Thus, the number of objects that the system can be trained to recognize is severely limited. There is ample evidence suggesting that the visual system uses a hierarchy to solve this problem (Logothetis & Sheinberg, 1996; Pasupathy & Connor, 2002) and such ideas have already been explored by some existing neural models (Riesenhuber & Poggio, 2000). The shifter circuit of Olshausen *et al.* (1993) decomposes the task of scale and translation invariance from the task of object recognition which, in their model, is deferred primarily to IT. The forward connections between cortical areas along the visual hierarchy in their model serves primarily for routing. Our model emphasizes the fact that the forward connections are used primarily for encoding objects and that routing emerges as a result of recurrent interaction of the 'what' and 'where' pathways in the early visual cortex. This view is more consistent with neurophysiological findings as well as existing neural models. Many aspects of vision have not been addressed in our model, such as contrast invariance and scale invariance. Contrast normalization can be achieved by simple normalization of responses within each hypercolumn (Heeger, 1992) and scale invariance can potentially be solved by training the network with input of a same object in multiple scales and in the framework of hierarchy. An interesting question is whether a top-down bias, such as object attention bias, can propagate down a hierarchy of visual areas without losing its effectiveness. A generalization of Hebbian learning to multiple layers is not trivial. Recent integration of our model with the VisNet model of Wallis & Rolls (1997) has shown that it is possible for the top-down signals to propagate down the hierarchy (Rolls & Deco, 2002).

Recently, a more powerful 'shifter circuit', called a map-seeking circuit, has been proposed by Arathorn (2002) which also bears a strong resemblance to the proposals of Mumford (1992) and Ullman (1995). His map-seeking circuit is similar to our model in that it allows V1 to make parallel proposals to the ventral module and uses the similarity between the feedback synthesis and the bottom-up proposals to assess the fitness of the proposals. However, his circuit still requires the dynamic adjustment of gain or weight of the feedforward pathways rather than using a dorsal module or utilizing V1 activity modulation as in our model. While his circuit is more powerful at this stage as it can handle translation, scale and rotation invariance, our scheme is more neurally plausible. It would be interesting to investigate whether his map-seeking circuit can be implemented in the recurrent interactive scheme proposed here.

The behavior of the units in our model has some qualitative similarities to the behaviors of V1, V4 and IT neurons observed in neurophysiological experiments (see Figs 9 and 10). Specifically, we show that both stimulus-triggered spatial and object attention effects can take 50 ms to develop in V1. This suggests that stimulus-evoked attention might underlie some of the observed long-latency contextual modulation in V1 and V2 neurons (Lamme, 1995; Zipser *et al.*, 1996; Lee *et al.*, 2002). Further, our model also exhibits the biased-competition attentional effect and receptive field-shrinking effect observed in V4 and IT (Moran & Desimone, 1985; Reynolds *et al.*, 1999). Top-down feedback provides a mechanism for contextual effect in the early visual cortex, resulting in the sensitivity of the neurons to the stimuli outside their classical receptive fields, particularly in the later part of the neurons' responses (Lamme, 1995; Lee *et al.*, 2002). Although, in some cases, top-down spatial attention is shown to increase the spontaneous activities of early visual neurons (Luck *et al.*, 1997), most of the contextual modulation observed requires visual stimulus within the classical receptive field for the cells to spike. In our model, even though top-down bias can increase the early cortical neurons' activities, without the bottom-up signals, the relatively diffuse and weaker top-down signals might not be strong enough to drive the recurrent circuit to form a stable representation in the early visual cortex. Such qualitative consistency suggests that the basic principles advocated here might be relevant to understanding the visual system. In this paper, we have stressed the role of top-down attentional effects in early visual processing. A complete model of attentional control should include a more sophisticated component for computing bottom-up saliency beyond the 'winner-take-all' lateral inhibition mechanism in this paper (see Lee *et al*., 1999).

It should be noted that the model advanced here is a minimalistic model. Parameters are chosen to drive the system to fixed points corresponding to zero activity for the losers and large activation for the winners (see Usher & Niebur 1996 for a description of the fixed point attractor of the types of equations that we used). While the VM and DM neurons in our model exhibit winner-take-all phenomena, the responses of actual visual neurons in both the dorsal and ventral streams are known to be more graded. Furthermore, the attentional enhancement that we saw in model V1 units is also larger than that commonly observed *in vivo*. We studied this model for its simplicity and the parameters are generic so that their particular values are not critical to the normal performance. Recent experimental evidence suggests that selective attention might not need to be implemented by enhanced firing rates but could be implemented by enhanced oscillation or synchrony in neuronal ensembles, particularly in the early visual areas (Murthy & Fetz, 1996; Fries *et al.*, 2001; Steinmetz *et al.*, 2000). We have not explored this dimension in this study and assume that activity in a neuronal pool indicates averaged firing rate. However, synchrony could be another measure of activity level, as it is known that synchrony can provide a stronger input to the postsynaptic pool. Future research is needed to generalize this model to incorporate the idea of synchrony as a measure of input current and pool activity in the model (see also Niebur & Koch, 1994). Nevertheless, we show that this very simple system is already able to demonstrate some basic and interesting results qualitatively. Further research is necessary to develop a quantitatively accurate model.

This paper makes three principal conceptual conjectures: (i) the early visual cortex, including V1 and V2, plays a very central role in mediating spatial attention/routing and in mediating visual search; (ii) the dorsal stream is critical for mediating spatial attention in the ventral stream and that, conversely, object attention can influence the parietal area and (iii) the bottom-up and top-down computations in the ventral stream are parallel and feature integration is a parallel rather than a serial process. These conjectures yield some interesting experimental predictions at a conceptual level. First, if V1 indeed plays a critical role coordinating routing and search, disrupting the activities in V1 up to 100 ms poststimulus onset should severely undermine the visual search performance and spatial attention effect in humans or monkeys by interrupting recurrent interaction in the system. Such disruption can potentially be induced by trans-cranial magnetic stimulation as in the work of Kamitani & Shimojo (1999). Second, if the dorsal stream is responsible for the spatial attentional effect observed in the ventral stream, deactivating the parietal cortex pharmacologically should eliminate the biased-competition spatial attentional effect in V4 in primate electrophysiological experiments.

The model also suggests that significant spatial attentional effects should progress from the parietal cortex to V1 and then to the ventral areas, while object attention effects should progress from ventral areas to V1 and then to the parietal cortex. Finally, one might test whether serial search is mediated by a parallel or serial mechanism by examining the response dynamics of the visual neurons in V1 and V2 when the monkey engages in a conjunctive visual search task, using either optical imaging or by simultaneously recording from multiple neurons in different hypercolumns. For example, suppose a monkey has to search for a shape with conjunctive features among similar distractors in a delayed-match-to-sample paradigm. If the search mechanism is serial, one might see the enhancement effect move from one visual item to another in the course of a trial and the locus of movement may vary randomly from trial to trial. However, if the search mechanism is parallel, the enhancement effect should emerge in multiple locations simultaneously, contract to the location of the correct target and then be relatively constant across trials. Obviously, clever experimental designs and rigorous statistical methods are needed to test this hypothesis.

The prediction that attentional effects can be seen to propagate through V1 between the dorsal and ventral stream might depend on the resolution of analysis. For integration of information of fine details, V1 needs to be involved. For integration of information of a larger scale, V2 or V4 can serve as the integration buffer. Recent functional imaging data (Martinez *et al.*, 1999) seem to suggest that spatial attention effects are first evident in the extrastriate cortex (V4 and IT) and later in V1, rather than the other way around, suggesting that the simplistic scheme of our model is not complete. These data might suggest that, along the dorsal and ventral hierarchy, there could be many layers of cross-talk but the cross-talk at the higher levels is more coarse and more global. The interaction can first occur in the higher areas and then penetrate back to V1, resulting in coarse-to-fine processing. This scenario is very probable and can be seen as a generalization of the current model. The general notion of V1 serving as high-resolution buffer could still be correct but might only come into play when precise spatial localization and precise feature

discrimination are required, as suggested also by the mental imagery experiments of Kosslyn *et al.* (1995).

This model is a synthesis of many existing ideas and thus shares many features with existing cortical models that emphasize the importance of parallel recurrent feedback in sorting out disambiguities. These include the adaptive resonance of Grossberg (1987), interactive activation of McClelland & Rumelhart (1981), pattern-theoretic feedback proposals of Mumford (1992), counter-streams model of Ullman (1995), Kalman filter model of Rao & Ballard (1999) and map-seeking circuit of Arathorn (2002). This model also utilizes mechanisms common to many existing biologically motivated attentional models (Buracas *et al.*, 1996; Usher & Niebur, 1996; Braun *et al.*, 2001; Horwitz *et al.*, 1999; Lee *et al.*, 1999; Reynolds *et al.*, 1999; Rolls & Milward, 2000; Deco & Zihl, 2001; Rolls & Deco, 2002). The similarities among many of these ideas reflect a convergence of thinking on the mechanisms of cortical computation. A unique feature of our model is its suggestion that integration of 'what' and 'where' information can happen much earlier than previously thought and that V1 might play a central role in the registration and integration of various kinds of abstract higher level information into a coherent percept. In the network model proposed here, many apparently contradictory phenomena and processes, such as spatial and object attention, serial and parallel search, selective routing and top-down feedback, can be understood as different aspects or manifestations of a single unified system.

Acknowledgements

G.D. was supported by Siemens, AG. Corporate Technology during this research and T.S.L. is supported by NSF 9984706 and NIH MH64445. We thank C. Yang and colleagues at the CNBC for helpful comments on the manuscript.

Abbreviations

IT, inferotemporal cortex; DM, dorsal stream module; VM, ventral stream module.

References

Amit, D. & Tsodyks, M. (1991) Quantitative study of attractor neural network retrieving at low spike rates. I. Substrate spikes, rates and neuronal gain. *Network*, **2**, 259–273.

Arathorn, D.W. (2002) *Map-Seeking Circuits in Visual Cognition: A Computational Mechanism for Biological and Machine Vision*. Stanford University Press, Palo Alto.

Braun, J., Koch, C. & Davis, J. (2001) *Visual Attention and Cortical Circuits*. MIT Press, Cambridge, MA.

Buracas, G.T., Albright, T.D. & Sejnowski, T.J. (1996) Varieties of attention: a model of visual search. Proceedings of the 3rd joint symposium on neural computation. *Inst. Neur. Comput.*, **6**, 11–25.

Chelazzi, L., Miller, E., Duncan, J. & Desimone, R. (1993) A neural basis for visual search in inferior temporal cortex. *Nature*, **363**, 345–347.

Daugman, J. (1988) Complete discrete 2D-Gabor transforms by neural networks for image analysis and compression. *IEEE Trans. Acoust., Speech, Sign. Process.*, **36**, 1169–1179.

Deco, G. & Lee, T.S. (2002) A unified model of spatial and object attention based on inter-cortical biased competition. *Neurocomputing*, **44–46**, 769–774.

Deco, G. & Zihl, J. (2001) Top-down selective visual attention: a neurodynamical approach. *Vis. Cogn.*, **8**, 119–140.

Duncan, J. & Humphreys, G. (1989) Visual search and stimulus similarity. *Psychol. Rev.*, **96**, 433–458.

Felleman, D.J. & Van Essen, D.C. (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex*, **1**, 1–47.

Fries, P., Reynolds, J.H., Rorie, A.E. & Desimone, R. (2001) Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, **291**, 1560–1563.

Gerstner, W. (2000) Population dynamics of spiking neurons: Fast transients, asynchronous states, and locking. *Neur. Comput.*, **12**, 43–89.

Grossberg, S. (1987) Competitive learning: from interactive activation to adaptive resonance. *Cogn. Sci.*, **11**, 23–63.

Heeger, D.J. (1992) Normalization of cell responses in cat striate cortex. *Vis. Neurosci.*, **9**, 181–198.

Horwitz, B., Tagamets, M.-A. & McIntosh, A.R. (1999) Neural modeling, functional brain imaging and cognition. *Trends Cogn. Sci.*, **3**, 91–98.

Hubel, D.H. & Wiesel, T.N. (1978) Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. B (Lond.)*, **198**, 1–59.

Hupe, J.M., James, A.C., Payne, B.R., Lomber, S.G., Girard, P. & Bullier, J. (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, **394**, 784–787.

Ito, M. & Gilbert, C.D. (1999) Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron*, **22**, 593–604.

Kamitani, Y. & Shimojo, S. (1999) Manifestation of scotomas by transcranial magnetic stimulation of human visual cortex. *Nat. Neurosci.*, **2**, 767–771.

Kosslyn, S., Thompson, W.L., Kim, I.J. & Alpert, N.M. (1995) Topographical representations of mental images in primary visual cortex. *Nature*, **378**, 496–498.

Lamme, V.A.F. (1995) The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.*, **15**, 1605–1615.

Lee, D.K., Itti, L., Koch, C. & Braun, J. (1999) Attention activates winner-take-all competition among visual filters. *Nat. Neurosci.*, **2**, 375–381.

Lee, T.S. (1996) Image representation using 2D Gabor wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.*, **18**, 959–971.

Lee, T.S., Mumford, D., Romero, R.D. & Lamme, V.A.F. (1998) The role of the primary visual cortex in higher level vision. *Vision Res.*, **38**, 2429–2454.

Lee, T.S., & Nguyen, M. (2001) Dynamics of subjective contour formation in early visual cortex. *Proc. Natl Acad. Sci. USA*, **98**, 1907–1911.

Lee, T.S., Yang, C.F., Romero, R.D. & Mumford, D. (2002) Neural activity in early visual cortex reflects behavioral experience and higher order perceptual saliency. *Nat. Neurosci.*, **5**, 589–597.

Logothetis, N.K. & Sheinberg, D.L. (1996) Visual object recognition. *Annu. Rev. Neurosci.*, **19**, 577–621.

Luck, S., Chelazzi, L., Hillyard, S. & Desimone, R. (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.*, **77**, 24–42.

Martinez, A., Anllo-Vento, L., Sereno, M.I., Frank, L.R., Buxton, R.B., Dubowitz, D.J., Wong, E.C., Hinrichs, H., Heinze, H.J. & Hillyard, S.A. (1999) Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat. Neurosci.*, **2**, 364–369.

McClelland, J.L. & Rumelhart, D.E. (1981) An interactive activation model of context effects in letter perception. Part I: an account of basic findings. *Psychol. Rev.*, **88**, 375–407.

Moran, J. & Desimone, R. (1985) Selective attention gates visual processing in the extrastriate cortex. *Science*, **229**, 782–784.

Motter, B. (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.*, **70**, 909–919.

Mumford, D. (1996) Commentary on 'Banishing the homunculus' by H. Barlow. In Knill, D.C. & Richards, W. (Eds), *Perception as Bayesian.* Cambridge University Press, UK, 501–504.

Murthy, V.N. & Fetz, E.E. (1996) Oscillatory activity in sensorimotor cortex of awake monkeys: synchronization of local field potentials and relation to behavior. *J. Neurophysiol.*, **76**, 3949–3967.

Niebur, E. & Koch, C. (1994) A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *J. Comput. Neurosci.*, **1**, 141–158.

Olshausen, B., Andersen, C. & Van Essen, D. (1993) A neural model for visual attention and invariant pattern recognition. *J. Neurosci.*, **13**, 4700–4719.

Pasupathy, A. & Connor, C.E. (2002) Population coding of shape in area V4. *Nat. Neurosci.*, **5**, 1332–1338.

Pollen, D.A., Gaska, J.P. & Jacobson, L.D. (1989) Physiological constraints on models of visual cortical function. In Cotterill, R.M.J. (Ed.), *Models of Brain Function*. Cambridge University Press, UK, pp. 115–135.

Rao, R. & Ballard, D.H. (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.*, **2**, 79–87.

Reynolds, J., Chelazzi, L. & Desimone, R. (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.*, **19**, 1736–1753.

Riesenhuber, M. & Poggio, T. (2000) Models of object recognition. *Nat. Neurosci.*, **3** (Suppl.), 1199–1204.

Roelfsema, P.R., Lamme, V.A. & Spekreijse, H. (1998) Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, **395**, 376–381.

Rolls, E.T. & Deco, G. (2002) *Computational Neuroscience of Vision*. Oxford University Press, Oxford.

Rolls, E.T. & Milward, T. (2000) A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition and information-based performance measures. *Neur. Comput.*, **12**, 2547–2572.

Salin, P.A. & Bullier, J. (1995) Corticocortical connections in the visual system: structure and function. *Physiol. Rev.*, **75**, 107–154.

Steinmetz, P.N., Roy, A., Fitzgerald, P.J., Hsiao, S.S., Johnson, K.O. & Niebur, E. (2000) Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature*, **404**, 187–190.

Treisman, A. & Gelade, G. (1980) A feature-integration theory of attention. *Cogn. Psychol.*, **12**, 97–136.

Ullman, S. (1995) Sequence seeking and counter streams: a computational model for bi-directional information flow in the visual cortex. *Cereb. Cortex*, **5**, 1–11.

Ungerleider, L.G. & Mishkin, M. (1982) Two cortical visual systems. In Ingle, D.J. (Ed.), *Analysis of Visual Behavior*. MIT Press, Cambridge, MA, pp. 549–586.

Usher, M. & Niebur, E. (1996) Modeling the temporal dynamics of IT neurons in visual search: a mechanism for top-down selective attention. *J. Cogn. Neurosci.*, **8**, 311–327.

Wallis, G. & Rolls, E.T. (1997) Invariant face and object recognition in the visual system. *Prog. Neurobiol.*, **51**, 167–194.

Wilson, H. & Cowan, J. (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biol. Cybernet.*, **12**, 1–24.

Wolfe, J.M. (1998) Visual search: A review. In Pashler, H. (Ed.), *Attention.* University College London Press, London, pp. 13–77.

Zipser, K., Lamme, V.A.F. & Schiller, P.H. (1996) Contextual modulation in primary visual cortex. *J. Neurosci.*, **16**, 7376–7389.