Review of Psychology
2001, Vol. 8, No. 1-2, 31-40 UDC 159.9

Visual feature integration with amplitude maximization

DRAŽEN DOMIJAN

A neural model of visual feature integration is proposed based on presynaptic inhibition of excitatory feedback connections. The same activity level or amplitude of corresponding nodes represents features that belong to the same object. This is achieved by spreading of activation from strongly activated nodes to weakly activated nodes but not in reverse. Spreading is controlled by presynaptic inhibition, which prevents unbounded activity growth. The model's representational capacity is far greater than in models based on temporal synchrony, it is equally applicable to the static and moving stimuli and it can represent hierarchical groupings. A network may operate as a short-term storage, which allows simultaneous feature integration over space and time.

Top-down signals from higher visual centers can influence the operation of the network either through direct excitatory input or indirectly through inhibitory interneurons. With direct input, model behavior is consistent with psychophysical data on object-based attentional selection and curve tracing task. Indirect influences are able to provide flexible task-dependent feature integration and prevent accidental bindings that are consequences of intrinsic properties of some scenes.

Anatomical and physiological investigations suggest that visual system analyzes stimuli in a parallel and distributed manner. Specialized areas exist for processing different visual attributes such as form, color, motion, etc. (DeYoe & van Essen, 1988; Livingstone & Hubel, 1988; Zeki, 1993). Distributed processing is efficient but it creates a problem for object recognition because it is not clear which features belong to the same object. It is assumed that a process of feature integration or feature binding operates on distributed cortical maps before they send signals to higher cortical areas responsible for object recognition (Muller, Elliott, Herrmann & Mecklinger, 2001; Sejnowski, 1986; Treisman & Gelade, 1980).

Recently, a large amount of experimental and theoretical work has been devoted to the search for a plausible mechanism for feature integration. One possibility that receives considerable attention is stimulus specific synchronization of oscillatory neural activity observed in cats and monkeys. Neurons that code the same object engage in synchronous oscillations, that is, their phase difference is reduced to zero. On the other hand, neurons that code different objects become desynchronized. Physiological measurements also reveal that the degree of synchronization among cortical neurons depends on their receptive field properties, where more similar receptive fields are more likely to induce synchronization (Eckhorn *et al.*, 1988; Gray, Konig, Engel & Singer, 1989; Gray & Singer, 1989; Singer & Gray, 1995; see also recent reviews in Eckhorn, 1999; Gray, 1999; and Salinas & Sejnowski, 2001). Based on this discovery, many computational mechanisms have been proposed (Hummel & Biederman, 1992; Mozer, Zemel, Behrmann & Williams, 1992; Wang & Terman, 1995; 1997). Although they differ in biophysical plausibility and mathematical sophistication, they all share the common assumption of phase locking of oscillatory activity in cells that represent the same object.

It is possible to achieve temporal synchrony without oscillatory units. Such an idea has been put forward by Fujii *et al.* (1996), who assumed that neurons function as coincidence detectors. The basic idea is that a neuron is sensitive to a temporal difference between spikes that arrive on them. When the temporal difference is small, a neuron will fire and signal that inputs are correlated and arrive from a single source, but if the temporal difference is large, a neuron will not fire because uncorellated input arrives from different objects. Recently, this hypothesis has been extended in a more complicated architecture with bidirectional functional connectivity (Watanabe, Aihara & Kondo, 1998; Watanabe, Nakanishi & Aihara, 2001).

However, synchronization as a mechanism for feature integration has been criticized on several grounds. First, psychophysical evidence concerning synchronization is

---

Dražen Domijan, Department of Psychology, Faculty of Philosophy, University of Rijeka, I. Klobučarića 1, 51000 Rijeka, Croatia. E-mail: ddomijan@human.pefri.hr (Correspondence concerning this article should be sent to this address).

not conclusive because some researchers fail to find support for it (Kandil & Fahle, 2001; Lamme & Spekreijse, 1998; Lehky, 2000), while others confirm correlation with perceptual performance (Alais, Blake & Lee, 1998; Usher & Donnelly, 1998). Furthermore, Lamme (1995) reported a difference in amplitude of neuronal responses in a cortex in relation to a figure-ground organization. Second, models of object recognition based on temporal synchrony such as Hummel and Biederman's (1992) JIM or Wang's (Wang & Terman, 1995, 1997) LEGION suffer from capacity limitations making them unsuitable for processing complex natural scenes, which contain large number of objects. Such a limitation is intrinsic to synchronization because the phase difference of cells that code different objects could not be arbitrarily small. Thirdly, synchronization models are not capable of representing hierarchical stimuli (Hummel & Holyoak, 1993). This refers to the fact that one grouping may become part of a larger group or it may contain a smaller group in itself. For instance, the human body may be perceived as a single entity but it contains parts such as legs and head, which may be treated as separate perceptual objects. Finally, proposed models of synchronization have not yet demonstrated how moving stimuli could be integrated in perceptual units.

Due to the above mentioned problems, it is reasonable to ask whether there is a mechanism for feature integration in a more traditional framework based on a neuron's average firing rate or its amplitude. Treisman and Gelade (1980) suggested that attention is responsible for feature integration. Recently, such an idea has been further developed by Roelfsema (1998; Roelfsema, Lamme & Spekreijse, 2000). The problem with this approach is that it has not been computationally developed yet. Wersing, Steil, & Ritter (2001) proposed a competitive-layer model that is based on amplitude representation. Their model is an additive recurrent network with linear threshold neurons. However, in order to achieve the desired behavior, it requires an annealing procedure, which is not a biophysically realistic mechanism. Besides, it is not clear whether their model could be used with hierarchical and moving stimuli and how attention influences its behavior.

The aim of the present paper is to present a biophysically plausible mechanism for feature integration that does not suffer from capacity limitations, allows attention or any other top-down signals to alter its operations, and is capable of representing hierarchical and moving stimuli.

Model description

In the context of the lightness perception, Grossberg and his colleagues developed an alternative approach to the visual feature integration (Cohen & Grossberg, 1984; Grossberg, 1987; Grossberg & Todorović, 1988). They introduced the nonlinear diffusion process as an implementation of a perceptual filling-in. Diffusion allows spreading of neural activity which integrates local contrast information in a global perception of object's lightness or color.

Pessoa & Neumann (1997) and Grossberg (1987) suggested that the same approach could be applied to the integration of form, texture and other visual attributes. However, this extrapolation is not straightforward when object recognition is considered. In lightness perception, two different surfaces with the same lightness should be represented with the same activity level in corresponding nodes. But in the task of visual object recognition, every object should be represented with a different activity level despite their similarity in size, form or color. Therefore, a certain type of transformation is needed in order to prevent wrong bindings. Consider the simple example with two horizontal lines as input

$$I = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{1}$$

where 0s correspond to a dark background and 1s to light points. A diffusion model would respond with the same activity for both bars, making them indistinguishable for an object recognition system. This may be circumvented with gating of input signals before they can reach a diffusion stage with a weight matrix that gives different emphasis to every location in the visual field such as

$$w = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \\ 21 & 22 & 23 & 24 & 25 \end{pmatrix} \tag{2}$$

or some random matrixes that have different value on every location. It should be noted that this is equivalent to the requirement for distinct starting phases of neural oscillations in the models of synchronization (Hummel & Biedermann, 1992). A diffusion model performs averaging of input values so final activity would approximately correspond to

$$x = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 8 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 18 & 18 & 18 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (3)$$

However, such different weighting is not enough as can be seen from the following input.

$$I = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (4)$$

Here, both lines will elicit same response in the diffusion network

$$x = \begin{pmatrix} 6 & 0 & 6 & 0 & 0 \\ 6 & 0 & 0 & 6 & 0 \\ 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (5)$$

Also, the diffusion model is not appropriate for binding moving stimuli because, as the object changes location, average activity will change due to the different weights at different locations.

Instead of averaging, here it is proposed that maximization of a node's amplitudes achieve binding that is free from accidental errors. Maximization means enhancing all cells that represent the same object to the same activity level as the strongest cell in that group. Therefore, the network assigned certain activity value to every object in a visual scene and object recognition system could easily discriminate between them. Representational capacity is not limited because every location in a network receives different emphasis from input gating so they could not be mixed if they do not correspond to the same object.

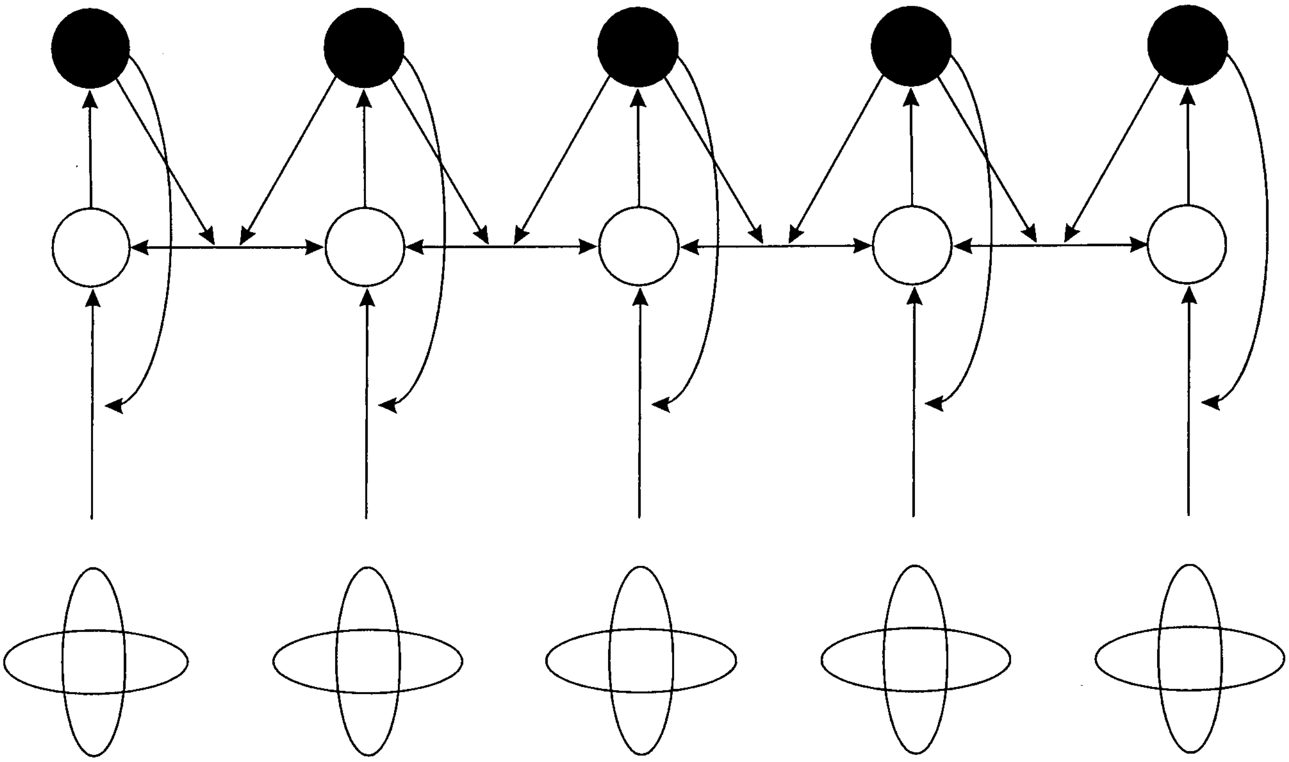

Figure 1. A model of visual feature integration based on amplitude maximization. Open circles are excitatory nodes and filled circles are inhibitory. Horizontal and vertical ellipses represent input from orientation selective cells in primary visual cortex.

Amplitude maximization is implemented as an excitatory feedback network with presynaptic inhibition. A model is illustrated in Figure 1. Formally, it is expressed as a set of differential equations

$$\frac{dx_i}{dt} = -Ax_i + Bf(I_i)\left(\left[w_iI_i - x_i\right]^+ + \left[x_{i+1} - x_i\right]^+ + \left[x_{i-1} - x_i\right]^+\right) \quad (6)$$

and

$$[y]^+ = \max(y,0) \quad (7)$$

where x_i denotes the neuron's activity at spatial position i; $-Ax_i$ is a passive decay that drives activity to the resting potential if no input is applied; B is a parameter that determine the strength of influence that input and neighboring cells may exert on a target cell; f( ) is a binary function with the value 1 if I_i0 and the value 0 otherwise. It is used because it restricts undesirable spreading of activity to inactive cells. Term I_i denotes input gated by synaptic weight w_i described by eq (2); x_{i+1} and x_{i-1} are nearest neighbors that deliver excitation. Input to the network are the activities from the distributed map of feature detectors such as simple or complex cells or directionally selective cells. If the object's form is used as a primary source of information for object recognition as in the model of Hummel and Biedermann (1992), then neighbors should be understood as spatially displaced cells with the same orientation as target cell. Therefore, for a complete representation of the object's form, separate feature binding networks with different orientations are needed.

Presynaptic inhibition is denoted with $-x_i$. Eq (7) describes rectification, which is necessary in a biologically plausible model, because it prevents excitatory connection from becoming inhibitory and vice versa.

The model operates in the following manner. If a cell receives the strongest input with respect to its neighbors, its inhibitory interneuron will prevent any excitatory influence that arrives on it through horizontal connections. Therefore, the cell will be protected or isolated from its neighbors and its activity will depend only on input value. On the other hand, if a cell does not receive the strongest input, its inhibitory interneuron will not be able to prevent excitatory influence from neighbors and the cell's activity will grow until it attains a value close to the activity level of its neighbors. In this way, activity spreads from strongly activated cells to the weakly activated cells but not in reverse. At the end of this process, all cells that represent the same object will attain the same activity value and that is the value of the strongest cell in the whole group for a particular object.

Computer simulations show that the proposed network converges to the equilibrium value that is close to the iteration of the following algebraic approximation

$$x_i = \max\left(w_iI_i, x_{i+1}, x_{i-1}\right) \quad (8)$$

at least for small values of A (.01) and large values of B (100). Therefore, the network response to the input matrix (1), is approximated by

$$x = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 9 & 9 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 19 & 19 & 19 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (9)$$

and to the input matrix (4) by

$$x = \begin{pmatrix} 11 & 0 & 9 & 0 & 0 \\ 11 & 0 & 0 & 9 & 0 \\ 11 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (10)$$

The network may also be used for surface representation. However, such representation requires a different network structure, because surfaces may overlap which will cause leaking of activity between cells that represent different objects. This is prevented by orientation selective cells which signal a boundary between surfaces. A variant of the model more suitable for surface representation is illustrated in Figure 2.

Mathematically, the network is described by

$$\frac{dx_i}{dt} = -Ax_i + B\left(\left[w_iI_i - x_i\right]^+ + \left[x_{i+1} - x_i - J_{i+.5}\right]^+ + \right.$$
$$\left. + \left[x_{i-1} - x_i - J_{i-.5}\right]^+\right) \quad (11)$$

where J denotes input from orientation selective cells which will prevent spreading between neighboring cells if they represent different surfaces. It is assumed that a signal from orientation selective cells is stronger than a signal transmitted through excitatory axons so there is no possibility for activity leaking. Spatial displacements of .5

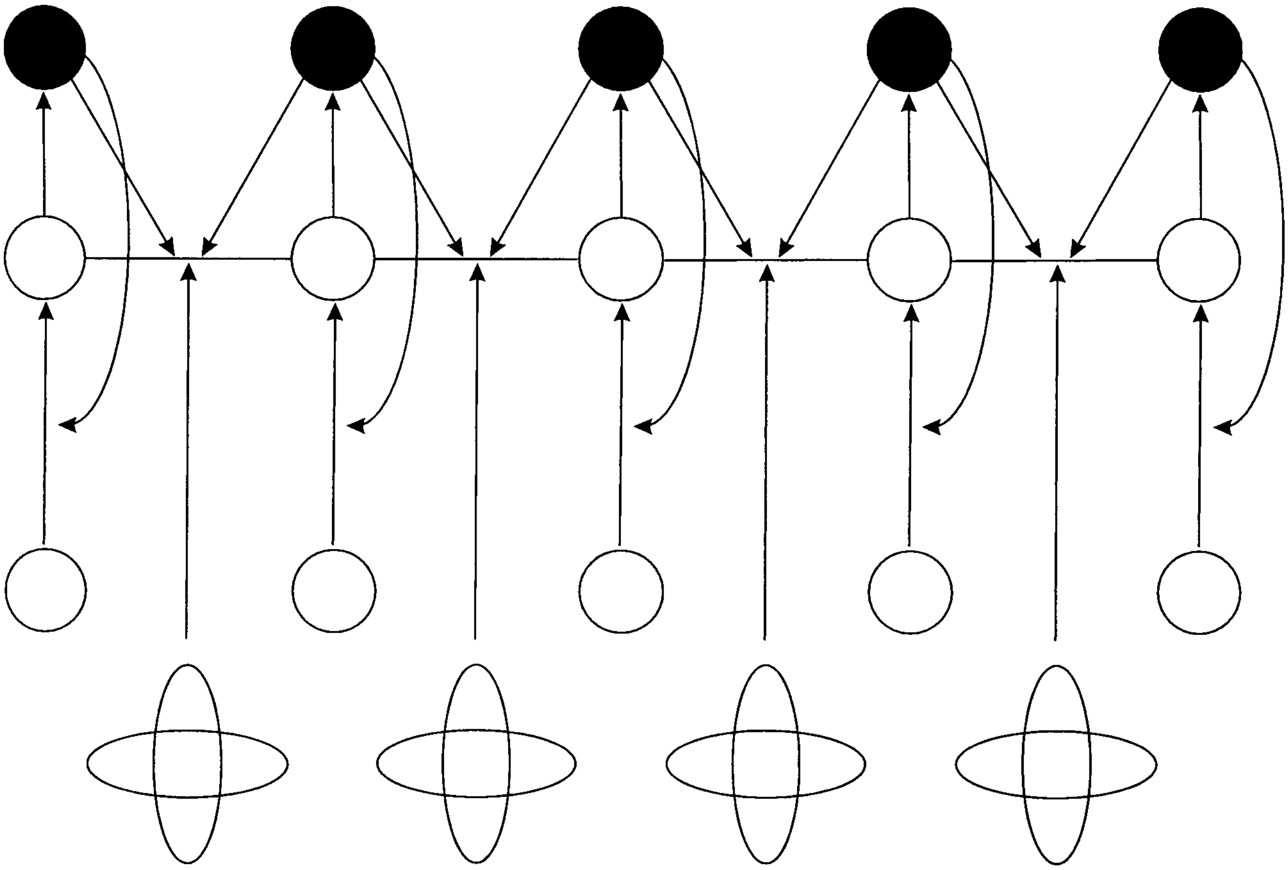

Figure 2. A variant of the proposed neural network model more suitable for surface representation. Orientation selective cells do not drive the network but provide restriction to activity spreading which allows the network to label different surfaces with different activity levels. Input to the network arrives from independent source that represents surfaces such as filling-in network in Grossberg's theory.

means that J does not influence network cells but influence axons between them. Therefore, i + .5 is a spatial position of an orientation selective cell that is placed between cells at spatial positions i and i + 1 in an activity spreading network which allows it to signal a boundary between them. Such a spatial coding achieves better surface representation. Term I does not represent direct input as in eq (6), rather it is an internally generated signal necessary for labeling different surfaces with different activity levels. Output of the networks for edge and surface labeling are illustrated in Figure 3a and 3b, respectively.

Humphreys, Cinel, Wolfe, Olson and Klempen (2000) provide neuropsychological evidence for distinct binding networks of form and surface properties. They presented data from a patient with bilateral parietal lesions, GK, who showed difficulty in integrating form and surface properties, as indicated by a large number of illusory conjunctions when asked to report shape and color of multiple objects in a visual field. On the other hand, a grouping has little effect on the number of illusory conjunctions indicating that integration of edge information into shape is left intact. Based on this finding, they conclude that binding is not a unitary process, but involves several networks that integrate different perceptual attributes. Damage to one such network may still leave other binding networks intact. This point is further developed by Humphreys (2001), who reviewed various neuropsychological data, and suggests that at least three different binding processes could be distinguished. The first is an integration of local edge signals into contours and integration of contours into coherent shapes. The second is an integration of shape and surface properties, and the third is a transient binding process that is sensitive to temporal cooccurrence of input.

A network for surface integration presented in Figure 2 may be applied in image segmentation. It has an important advantage over previous algorithms such as FBF networks proposed by Grossberg and Wyse (1991). FBF networks require a different network layer for every surface in a vis-

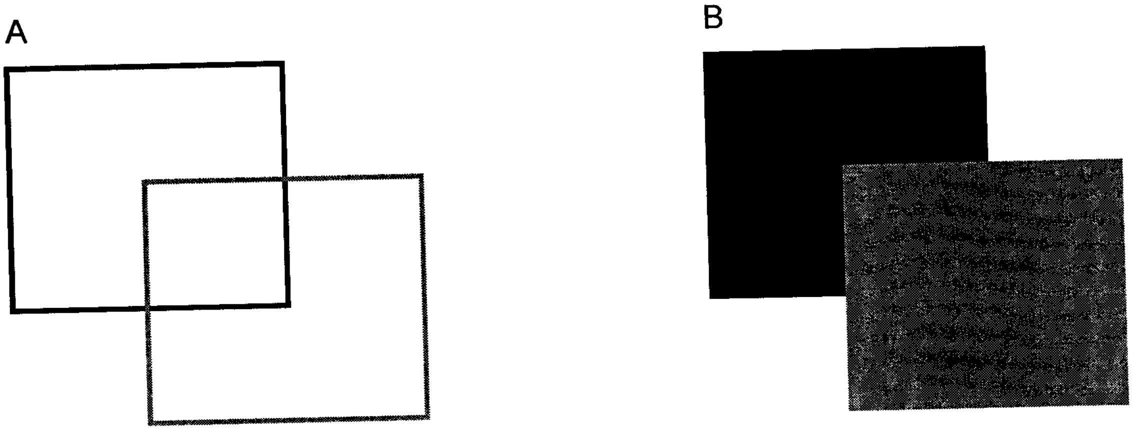

Figure 3. Output of the two variants of the network: A) when input came from orientation selective cells; B) when input came from surface representation.

ual field (i.e., segmentation is achieved when all surfaces are separated in their own network). However, the number of objects in a visual scene may vary considerably and it may be very large, so the number of layers in the network should be equal to the largest possible number of objects that may be encountered simultaneously, but this information is not known in advance. The present approach does not suffer from such a problem. One network layer is enough since every surface is labeled with a different activity level.

Spatiotemporal processing

The model proposed here may be used to bind features of moving objects. Observe that the maximum is left invariant if the stimulus is moving from a location that is more weighted to the location with a smaller weight. This is true if we assume that lateral excitatory signals travel faster than the stimulus moves. In that case, the strongest cell perturbs its neighbors before it is shut down due to the stimulus disappearance in its receptive field. Therefore, directional selective cells should be weighted in a way that gradient of weights follows preferred direction. For instance, cells selective for movement from left to right should have the largest weight at leftmost location and lowest at rightmost location. One potential problem with the network is its very slow decay rate, which is necessary for a smoother activity spreading. This implies that when an object moves, cells will be active even after a long period of time the object passes its locations, which will cause smearing of object representation (network will not represent objects but their trajectories). This problem could be overcome if we introduce a strong reset signal that selectively shuts down all cells, which has not been currently activated by the input. Therefore, when a stimulus moves, cells could track its current position while all cells that have been previously activated will be instantaneously deactivated.

Another important aspect of neural information processing is binding in a temporal domain. Different cells may process parts of the same object at different times due to the large amount of variability in their response latencies, which may cause a problem for object recognition system. Amplitude maximization solves this problem because it operates as a sort of short-term memory due to the slow decay rate. When a strong reset signal used for moving stimuli is removed from the model, the network will approximately retain its activity level despite the fact that input has ceased. Another possibility is that we assume that reset signal is activated with a certain delay (i.e., when input ceases, a reset signal waits for a while and then becomes activated). Cell activity is not reduced much during this delay period, and if another part of the object representation becomes active, it will be grouped with previously active parts. It is also possible that a reset signal may be dynamically manipulated from higher centers or from other parts of the object recognition network, where partial activation from a binding network allows recognition, but these centers may require further evidence from the input and therefore hold on already acquired data.

Hierarchical stimuli

An important advantage of the proposed neural network is its ability to represent hierarchical stimuli. This is achieved in a simple way by lifting cell activities by a certain amount, which allows an object recognition system to distinguish among different hierarchical levels. For instance, in the eq (1), cell activity may vary between 1 and 25 due to the input gating. However, if both objects are parts of the larger grouping, certain cells that detect higher order groupings may send strong feedback signal to the input network. Suppose that the strength of the feedback signal is 100 units, final activity in the feature integration network therefore will be 109 for an upper object and 119 for a lower object. In this way, the network successfully coded information that there is a single object coded by 100 units of activity composed of two parts coded by 9 and 19 units of activity. If there is even higher hierarchical level it may send 1000 units of activity to the input network. Also, if there is some other entity at the same hierarchical level it may send 2000 units and so on. In a similar vein, it is possible to represent hierarchical knowledge of an arbitrary complexity. An object recognition system may decode such hierarchical information by rounding off values that correspond to the hierarchical levels that are not interesting at that moment. For instance, if the lowest level is not interesting, the last two digits are rounded to the closest value so both parts signal 100 units of activity. Also, it is possible that different parts of the recognition system keep track of different hierarchical levels so it may simultaneously represent different hierarchical levels which is consistent with human capability.

Attention

Attentional influences may be incorporated in the model through inhibitory interneurons in at least two ways and through direct excitatory synapses from higher visual centers to the excitatory cells.

When attention enhances activity of the interneuron, it will strongly prevent any lateral excitatory influences despite the fact that neighbors may have stronger activity than the target cell. In other words, cells are isolated from neighbors and their activity will converge to the level defined solely by the input. Therefore, attention provides flexible bindings that are task dependent, that is, it may prevent some bindings if they are undesirable. Such behavior is consistent with Hummel and Biederman's idea that attention serves to prevent bindings. However, their hypothesis was motivated by the limitations of the mechanism of synchronization that they employ for feature binding. Their proposition is in contrast with Treismann's feature integration theory where attention plays a crucial role in binding distributed features. In the presented model, Treisman's position may be accommodated if we assume that, initially (before the start of the binding process), all interneurons have a strong activity level which prevents any activity spreading (another possibility is discussed below in relation to the curve-tracing task). When attention is employed to a particular location, excitation from attention network to the inhibitory interneuron is reduced for the location in a spotlight of attention, and binding is enabled. Therefore, both theoretical positions about the role of attention in feature binding are implemented in the same neural architecture. The only difference is the parametric variation, in the sense that attention may enhance or reduce internenuron activity.

Another possibility is that attention, or in more general terms, any top down signals from higher visual centers, directly influence excitatory cells through excitatory synapses. In that case, activity of a certain cell which lies at the location where attention is directed, will be enhanced but because of lateral connections this activity will spread to nearby locations that are connected with attended ones. Actually, spreading will continue until all cells that represent the same object as an attended cell converge to the activity value of the attended cell. Therefore, attention is spread to all locations occupied by the object, which is consistent with recent psychophysical investigations showing object benefits in feature detection task. When two overlapping objects are briefly presented, and the task is to report their two attributes, participants make fewer errors when both attributes belong to the same object (Duncan, 1984). Vecera and Farah (1994) extended his finding and showed that object benefit is equal regardless of whether objects are superimposed or not. They interpreted this as an evidence for a spatially invariant object processing and against grouped arrays of enhanced activity that correspond to objects in retinotopic map. However, Kramer, Weber and Watson (1997) challenged their conclusion on methodological grounds and provide evidence for a grouped array interpretation, which is in agreement with the mechanism proposed here. Recently, O'Grady and Muller (2000) have provided additional evidence for grouped array representation using a different paradigm.

In a curve-tracing task, participants are asked to judge whether two points lie on the same or different curves. The difficulty of the task may vary depending on the number of intersections among curves and the strength of their curvatures. An interesting finding is that the time to solve the task increases linearly with the distance between the points. When Euclidean distance is kept fixed while the length of the path between points is varied, the time to solve the task again increases with the increasing length (Roelfsema *et al.*, 2000). In a related neurophysiological study, Roelfsema *et al.* (1998) found that in an alert monkey's primary visual cortex, the neuron showed enhanced response if the traced curve is in its receptive field while the response is reduced if the neuron represents a distractor curve. In this way, attention labels target curve with increasing neural activity.

It should be noted that the proposed network does not automatically solve the curve tracing task because, ini-

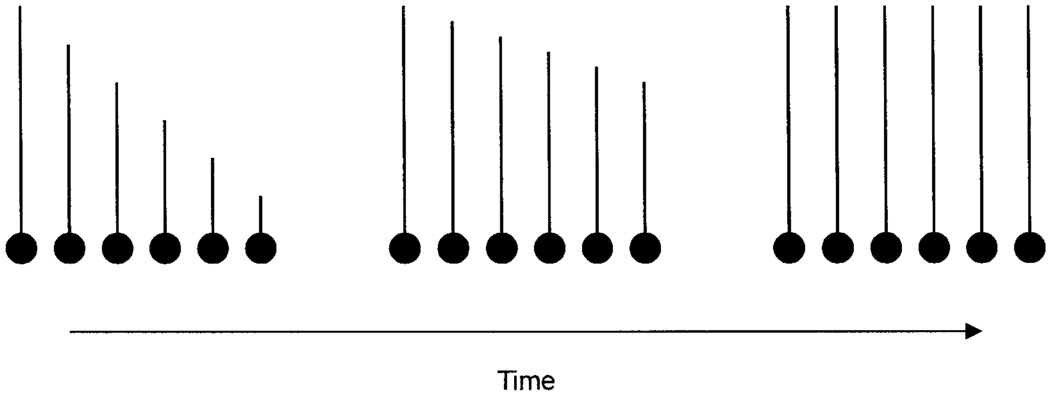

Figure 4. Time course of activity development. Height of vertical bars denotes activity level of corresponding cell. Left figure denotes start of the activity spreading and right figure shows final activity level.

tially, different segments of the curve may be labeled with different activity values since indicated points are treated as separated objects. Attention is required to give an additional excitatory signal which will be spread along the curve. Spreading is a process that requires some time, and if we assume that every synapse that should be passed along the object representation causes a small delay in signal transmission, the total time needed to complete the spreading will be linearly related to the size of the object or length of the path of the curve as shown by psychophysical investigations. Figure 4. illustrates temporal aspects of the network behavior.

An initial grouping achieved without attention could be interpreted as a base grouping in Roelfsema *et al.* (2000) theory while an activity after attention has been deployed is analogous to an incremental grouping. Recently, they showed using variant of the curve tracing task that there are indeed two components of the perceptual grouping; one that is achieved without attentional control and the second which requires attention (Scholte, Spekreijse, & Roelfsema, 2001). The mechanism presented here can support both types of grouping.

An alternative approach to the problem of interface between attention and perceptual integration has been developed by Grossberg and Raizada (2000; Raizada & Grossberg, 2001). Their network behaves in a similar manner as is described here, with activity spreading to the whole object representation. However, their model requires much more complicated neural architecture involving several processing stages and folded feedback from higher visual centers to the input stage with on-center off-surround network, which prevents unbounded growth of activity and at the same time mediates top-down attentional influences. Besides, such network does not have the ability to label different surfaces with different activity levels during initial preattentive grouping.

CONCLUSION

Presynaptic inhibition is a biophysically realistic mechanism confirmed by anatomical and physiological studies. It was discovered in a spinal cord (Eccles, Eccles, & Magni, 1961) and later in retina (Masland, Mills, & Cassidy, 1984) and hippocampus (Colmers, Lukowiak, & Pittman, 1987). An anatomical basis for this mechanism are axo-axonal synapses which have been discovered by electron microscopy in various locations in a mammalian central nervous system (Nicholls, Martin, & Wallace, 1991). In a neural network modeling, it was introduced by Yuille and Grzywacz (1989) in their model of winner-take-all behavior, where presynaptic inhibition provides ordering of the competitive interactions, in a way that the strongest cell receives the smallest amount of inhibition and therefore wins the competition. Here, it is shown that presynaptic inhibition may be useful in preventing unbounded exchange of activity in neural network with excitatory horizontal connections where cells reinforce each other and lose sensitivity to the input amplitude (Douglas, Koch, Mahowald, Martin, & Suarez, 1995; Grossberg, 1999). Due to the fact that inhibitory pathways do not interact with excitatory cells directly, presynaptic inhibition prevents excitatory reinforcement without causing any instability that would lead to the oscillations as in Li's (1998) model of contour integration in the visual cortex.

The model makes a testable neurophysiological prediction about the coding of locations in the cortex. Different locations are represented with different activity levels unless the same object occupies them. Such coding is expected to occur in posterior parietal cortex, which is known to be involved in spatial localization of objects in visual space (Ungerleider & Mishkin, 1982). Further research will elucidate whether the proposed mechanism may be applied to other aspects of cognitive functioning where binding plays an important role such as the modeling of language production and encoding in short-term memory.

In conclusion, visual feature binding may be achieved through amplitude maximization implemented as an excitatory feedback network with presynaptic inhibition. Presynaptic inhibition serves as a gate that allows strongly activated cells to influence weakly activated cells but not the opposite. Amplitude maximization is able to represent potentially infinitely many distinct objects in a visual field, it is free from accidental errors, it is equally applicable to static and moving stimuli and it is able to represent hierarchical stimuli. Moreover, the model allows higher visual centers to influence its operation in a way that is consistent with recent psychophysical research.

REFERENCES

ALAIS, D. BLAKE, R., & LEE, S.H. (1998). Visual features that vary together over time group together over space. *Nature Neuroscience, 1,* 160-164.

COHEN, M., & GROSSBERG, S. (1984). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Perception & Psychophysics, 36,* 428-456.

COLMERS, W.F., LUKOWIAK, K., & PITTMAN, Q.J. (1987). Presynaptic action of neuropeptide Y in area CA1 of the rat hippocampal slice. *Journal of Physiology, 383,* 285-299.

DEYOE, E.A., & VAN ESSEN, D.C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences, 11,* 219-226.

DOUGLAS, R.J., KOCH, CH., MAHOWALD, M., MARTIN, K.A.C., & SUAREZ, H.H. (1995). Recurrent excitation in neocortical circuits. *Science, 269,* 981-985.

ECCLES, J.C., ECCLES, R.M., & MAGNI, F. (1961). Central inhibitiory action attributable to presynaptic depolarization produced by muscle afferent volleys. *Journal of Physiology, 159,* 147-166.

ECKHORN, R. (1999). Neural mechanisms of visual feature binding investigated with microelectrodes and models. *Visual Cognition, 6,* 231-265.

ECKHORN, R., BAUER, R., JORDAN, W., BRISH, M., KRUSE, W., MUNK, M., & REITBOECK, H.J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? Muliple electrode and correlation analysis in the cat. *Biological Cybernetics, 60,* 121-130.

FUJII, H., ITO, H., AIHARA, K., ICHINOSE, N., & TSUKADA, M. (1996). Dynamical cell assembly hypothesis – Theoretical possibility of spatio-temporal coding in the cortex. *Neural Networks, 8,* 1303-1350.

GROSSBERG, S. (1987). Cortical dynamics of three-dimensional form, color, and brightness perception: I. Monocular theory. *Perception & Psychophysics, 41,* 87-116.

GROSSBERG, S. (1999). How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spatial Vision, 12,* 163-185.

GROSSBERG, S., & RAIZADA, R.D.S. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research, 40,* 1413-1432.

GROSSBERG, S., & TODOROVIĆ, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception & Psychophysics, 43,* 241-277.

GROSSBERG, S., & WYSE, L. (1991). Invariant recognition of cluttered scenes by a self-organizing ART architecture: Figure-ground separation. *Neural Networks, 4,* 723-742.

GRAY, C.M. (1999). The temporal correlation hypothesis of visual feature
integration: Still alive and well. *Neuron, 24,* 31-47.

GRAY, C.M., KONIG, P., ENGEL, A.E., & SINGER, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-column synchronization which reflect global stimulus properties. *Nature, 338,* 334-337.

GRAY, C.M., & SINGER, W. (1989). Stimulus specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences, 86,* 1698-1702.

HUMMEL, J.E., & BIEDERMAN, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99,* 480-517.

HUMMEL, J.E., & HOLYOAK, K.J. (1993). Distributing structure over time. *Behavioral and Brain Sciences, 16,* 464.

HUMPHREYS, G.W. (2001). A multi-stage account of binding in vision: Neurpsychological evidence, *Visual Cognition, 8,* 381-410.

HUMPHREYS, G.W., CINEL, C., WOLFE, J., OLSON, A., & KLEMPEN, N. (2000). Fractionating the binding process: neuropsychological evidence distingushing binding of form from binding of surface features, *Vision Research, 40,* 1569-1596.

KANDIL, F.I., & FAHLE, M. (2001). Purely temporal figure-ground segregation. *European Journal of Neuroscience, 13,* 2004-2008.

LAMME, V.A.F. (1995). The neurophysiology of the figure-ground segregation in primary visual cortex. *Journal of Neuroscience, 15,* 1605-1615.

LAMME, V.A.F., & SPEKREIJSE, H. (1998). Neuronal synchrony does not represent texture segregation. *Nature, 396,* 362-366.

LI, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Compuation, 10,* 903-940.

LIVINGSTONE, M.S., & HUBEL, D.H. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science, 240,* 740-749.

MASLAND, R.H., MILLS, J.W., & CASSIDY, C. (1984). The functions of acetylcholine in the rabbit retina. *Proceedings of the Royal Society London, B, 223,* 121-139.

MOZER, M.C., ZEMEL, R.S., BEHRMANN, M., & WILLIAMS, C.K.I. (1992). Learning to segment images using dynamic feature binding. *Neural Computation, 4,* 650-665.

MULLER, H.J., ELLIOTT, M.A., HERRMANN, C.S., & MECKLINGER, A. (2001). Neural binding of space and time: An introduction. *Visual Cognition, 8,* 273-285.

NICHOLLS, J.G., MARTIN, A.R., & WALLACE, B.G. (1992). *From neuron to brain.* 3rd ed. Sunderland, MA: Sinauer.

O'GRADY, R.B., & MULLER, H.J. (2000). Object-based selection operates on a grouped array of locations. *Perception & Psychophysics, 62,* 1655-1667.

PESSOA, L., & NEUMANN, H. (1997). Why does the brain fill in? *Trends in Cognitive Sciences, 2,* 422-424.

RAIZADA, R.D.S., & GROSSBERG, S. (2001). Context-sensitive binding by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition, 8,* 431-466.

ROELFSEMA, P.R. (1998). Solutions for the binding problem. *Journal of Biosciences. 53,* 691-715.

ROELFSEMA, P.R., LAMME, V.A.F., & SPEKREIJSE, H. (1998). Object-based attention in the primary visual cortex of macaque monkey. *Nature, 395,* 376-381.

ROELFSEMA, P.R., LAMME, V.A.F., & SPEKREIJSE, H. (2000). The implementation of visual routines. *Vision Research, 40,* 1385-1411.

SALINAS, E. & SEJNOWSKI, T.J. (2001). Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience, 2,* 539-550.

SCHOLTE, H.S., SPEKREIJSE, H., & ROELFSEMA, P.R. (2001). The spatial profile of visual attention in mental curve tracing. *Vision Research, 41,* 2569-2580.

SEJNOWSKI, T.J. (1986). Open questions about computation in cerebral cortex. In J.L. McClelland & D.E. Rumelhart (Eds.), *Parallel distributed processing, Vol. 2.* (pp. 372-389). Cambridge, MA: MIT Press.

SINGER, W., & GRAY, C.M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience, 18,* 555-586.

TREISMAN, A., & GELADE, G. (1980). A feature-integration theory of attention. *Cognitive Neuropsychology, 12,* 97-136.

UNGERLEIDER, L.G., & MISHKIN, M. (1982). Two cortical visual systems. In D.J. Ingle, M.A. Goodale, R.W.J. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549-586). Cambridge, MA: MIT Press.

WANG, D.L., & TERMAN, D. (1995). Locally excitatory globally inhibitory oscillator networks. *IEEE Transactions on Neural Networks, 6,* 283-286.

WANG, D.L., & TERMAN, D. (1997). Image segmentation based on oscillatory correlation. *Neural Computation, 9,* 805-836.

WATANABE, M., AIHARA, K., & KONDO, S. (1998). A dynamical neural network with temporal coding and functional connectivity. *Biological Cybernetics, 78,* 87-93.

WATANABE, M., NAKANISHI, K., & AIHARA, K. (2001). Solving the binding problem of the brain with bi-directional functional connectivity. *Neural Networks, 14,* 395-406.

WERSING, H., STEIL, JJ., & RITTER, H. (2001). A competitive-layer model for feature binding and sensory segmentation. *Neural Computation, 13,* 357-387.

YUILLE, A.L., & GRZYWACZ, N.M. (1989). A winner-take-all mechanism based on presynaptic inhibition feedback. *Neural Computation, 1,* 334-347.

ZEKI, S. (1993). *A vision of the brain.* Oxford, UK: Blackwell.

Received: October 2001
Accepted: December 2001