Contributed article

Object selection based on oscillatory correlation

D.L. Wang[*]

Department of Computer and Information Science and Center for Cognitive Science, The Ohio State University, 2015 Neil Avenue, Columbus, OH 43210, USA

Abstract

One of the classical topics in neural networks is winner-take-all (WTA), which has been widely used in unsupervised (competitive) learning, cortical processing, and attentional control. Owing to global connectivity, WTA networks, however, do not encode spatial relations in the input, and thus cannot support sensory and perceptual processing where spatial relations are important. We propose a new architecture that maintains spatial relations between input features. This selection network builds on Locally Excitatory Globally Inhibitory Oscillator Networks (LEGION) dynamics and slow inhibition. In an input scene with many objects (patterns), the network selects the largest object. This system can be easily adjusted to select several largest objects, which then alternate in time. We analyze the speed of selection, and further show that a two-stage selection network gains efficiency by combining selection with parallel removal of noisy regions. The network is applied to select the most salient object in gray-level images. As a special case, the selection network without local excitation gives rise to a new form of oscillatory WTA. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Object selection; Oscillatory correlation; Local cooperation; Global competition; Selective attention; LEGION; WTA

1. Introduction

Winner-take-all or maximum selector networks constitute a basic and large class of neural networks (see among others, Didday, 1970; Grossberg, 1976; Amari & Arbib, 1977; Koch & Ullman, 1985; Rumelhart & Zipser, 1986; Lippmann, 1987; Ermentrout, 1992; Yuille & Geiger, 1995). The WTA network in response to an input pattern produces only one active neuron with a highest input. The WTA dynamics is based on mutual inhibition, which can be implemented either by a global inhibitory unit or by mutual inhibitory connections. WTA networks have been applied to a variety of tasks, such as unsupervised (competitive) learning, pattern recognition, cortical processing, and selective visual attention.

The candidates in WTA competition are single neurons, corresponding to local representations. In contrast, the perceptual environment consists of patterned stimuli, which correspond to object structures, and objects seem to be the basic units of perception. Experimental data suggest that objects act as wholes in competition (Desimone & Duncan, 1995). Unless one adopts the view of local representations, whereby objects are represented by individual neurons (the *grandmother-cell* representation), existing

WTA networks are limited in addressing sensory and perceptual information processing. This is because, in distributed representations, each object is represented by an assembly of neurons encoding various features of the object, and such representations call for an assembly-level processing. One of the fundamental debates in neuroscience is whether objects are represented locally or distributedly. Here it suffices to point out that local representations face major theoretical and neurobiological problems (von der Malsburg, 1981; Abeles, 1991; Singer, 1993).

Take selective visual attention as an example. Roughly speaking, this refers to the phenomenon that attention is directed to one object (region) among many in a visual scene (for recent reviews see Desimone & Duncan, 1995; Neumann & Sanders, 1996; Egeth & Yantis, 1997). The so-called spotlight metaphor of attention speaks of a visual region under the attentional spotlight, the region that corresponds to the winning object. Because selective attention can be naturally viewed as a competitive process, WTA has played a major role in neural network theories of selective visual attention (see among others, Koch & Ullman, 1985; Niebur, Koch & Rosin, 1993; Niebur & Koch, 1996). Although these models exhibit selection due to WTA, they do not capture object-level selection required in selective attention as WTA provides only a winning *pixel* (or location), not a winning *object*. Also, it has been argued that selection mechanisms are indispensable for complex

* Tel.: + 1-614-292-6827; fax: + 1-614-292-2911.
E-mail address: dwang@cis.ohio-state.edu (D.L. Wang)
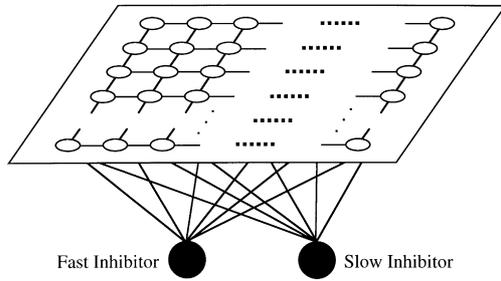
Fast Inhibitor   Slow Inhibitor

Fig. 1. Diagram of a 2-D selection network. An oscillator is indicated by an open circle, which connects with its four nearest neighbors with no wrap-around on the boundary. Two global inhibitors are indicated by the filled circles, and they connect mutually with each oscillator of the network.

visual computations such as image analysis (Tsotsos, 1995). Here again, selection must be at the object level in order to aid image analysis. In fact, it is generally assumed that the task of grouping pixels into regions or objects is performed before selective attention is involved (*preattentive*), by perceptual organization or Gestalt grouping principles (Palmer & Rock, 1994; Julesz, 1995; Desimone & Duncan, 1995). This view is consistent with recent claims that visual attention is surface (region) based (Nakayama, He & Shimojo, 1995; Mattingley, Davis & Driver, 1997).

It is conceivable that once a winning pixel is selected by WTA, a winning region may be extracted by some region-growing or filling-in process from that pixel. In this approach, a boundary detection mechanism needs to be in place so that the region growing is bounded to an appropriate region. The detection of closed boundaries in real images is a non-trivial computational problem. Even with successful closed boundary detection, objects, as recovered posteriorly, do not *participate* in the competition process, but rather pixels do and they would determine the outcome of object selection. One can imagine a hierarchical or multi-scale scheme where patches of pixels instead of individual pixels participate in the competition process so that selection takes place at a more abstract level.

Recently Terman and Wang proposed a neural architecture for dealing with the problem of perceptual organization and image segmentation (Wang & Terman, 1995; Terman & Wang, 1995). They studied a class of locally excitatory globally inhibitory oscillator networks (LEGION), whereby the building blocks are relaxation oscillators. LEGION is based on the representation of *oscillatory correlation*, whereby each object is represented by a synchronized oscillator group corresponding to the object and different objects in an image are represented by different oscillator groups which are desynchronized from each other. Oscillatory correlation is a special form of *temporal correlation* (Milner, 1974; von der Malsburg, 1981). Terman and Wang (1995) proved that global synchronization with local coupling is a robust property of relaxation oscillator networks and further such networks with a global inhibitory mechanism produce rapid desynchronization among different oscillator groups. More recently, the network has been

applied to segmenting gray-level images, giving rise to a number of major regions and a background comprising the rest of an image (Wang & Terman, 1997). To our knowledge, LEGION is the only oscillator network that has been rigorously shown to be capable of both rapid synchronization and desynchronization.

In this article, we study object-level selection using LEGION dynamics in object formation, i.e. the ability to group similar pixels into an object and segregate dissimilar regions into different objects. By introducing a new variable into each relaxation oscillator and a slow inhibitory mechanism, we show that the resulting network exhibits object selection. In response to a scene, this new selection network inhibits all the regions but the largest. If desired, the system can easily produce several largest objects by adjusting one parameter. We analyze the time the system takes to select the largest object, and based on this analysis we argue that the stage of perceptual organization, as performed by LEGION for example, is essential for efficient image analysis. A two-stage architecture is given that combines a LEGION layer and the selection layer, consistent with multistage theories of visual perception. Our integrated network is applied to identify the most salient regions in gray-level images. Finally, we give a new form of oscillatory WTA as a special case of the selection network.

2. Model description

We first describe the network architecture for object selection, which is called the *selection network*. Fig. 1 shows a typical 2-D selection network, with the simplest form of local connectivity; an oscillator is coupled only with its four immediate neighbors except on the boundaries where no wrap-around is used. In addition to the local connectivity, there are two global inhibitors, denoted by z_f and z_s for *fast* and *slow* inhibitors, respectively. Each inhibitor receives excitation from every oscillator of the 2-D grid and in turn inhibits each oscillator.

The building block of the selection network is a single oscillator i, which is defined as a feedback loop between an excitatory element x_i and an inhibitory element y_i:

$$\dot{x}_i = 3x_i - x_i^3 + 2 - y_i + I_i H(r_i - Cz_s) + S_i + \rho \quad (1a)$$

$$\dot{y}_i = \varepsilon(\alpha(1 + \tanh(x_i/\beta)) - y_i) \quad (1b)$$

where I_i represents external stimulation to the oscillator, and H stands for the Heaviside function, defined as $H(v) = 1$ if $v \geq 0$ and $H(v) = 0$ if $v < 0$. Within the Heaviside function, the variable r_i, to be defined later, indicates the size of the synchronized oscillator group that contains i at a particular time, and C is a system parameter. S_i represents the overall input from other oscillators in the network. The parameter ρ denotes the amplitude of Gaussian noise, which is introduced to test the robustness of the system and to assist in segregating different patterns. We note that the product of
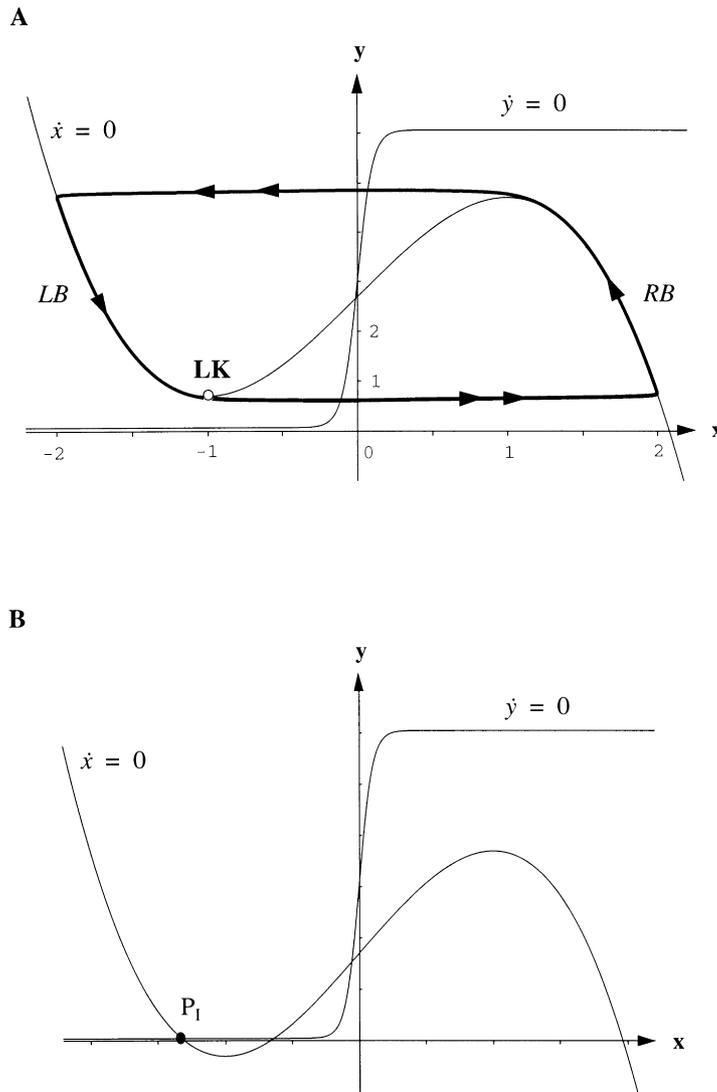
A

B

Fig. 2. Nullclines and trajectories of a single oscillator. (A) Trajectory of an enabled oscillator. The limit cycle trajectory is shown with a bold curve, and its direction of motion is indicated by arrowheads. LB and RB denote the left branch and the right branch of the cubic, respectively, and LK denotes the left knee of the cubic. (B) Trajectory of an excitable oscillator. The oscillator approaches the stable fixed point P_I.

external stimulation and the Heaviside function may be replaced by a sum, the details of which can be found in Wang and Terman, (1997, the technical report version in particular) who discussed a similar replacement in a different context.

The parameter ε is chosen to be a small positive number. Thus, if we ignore coupling and noise and set I to a constant, Eq. (1) defines a typical relaxation oscillator, similar to the van der Pol oscillator (van der Pol, 1926). The small parameter ε induces two time scales: a fast one dictated by the x variable and a slow one by the y variable. The x-nullcline of Eq. (1) is a cubic and the y-nullcline is a sigmoid. If $I > 0$ and $H = 1$, these nullclines intersect only at a point along the middle branch of the cubic with β small (see Fig. 2(A)). In this case, the oscillator produces stable oscillations for all sufficiently small values of ε, and it is

referred to as *enabled*. The periodic orbit alternates between *silent* and *active* phases of near steady-state behavior. As shown in Fig. 2(A), the silent and the active phases correspond to the left branch (LB) and the right branch (RB) of the cubic, respectively. The transitions between the two phases take place on the fast time scale, and they are thus referred to as *jumping*. The parameter α determines the ratio of the times the periodic solution spends in these two phases: a larger α results in a relatively shorter time in the active phase. In contrast, if $I \cdot H < 0$ the two nullclines of Eq. (1) intersect at a stable fixed point, P_I, along the left branch of the cubic (see Fig. 2(B)), and no oscillation occurs. In this case, the oscillator is referred to as *excitable*, meaning that it can be induced to oscillate by excitatory stimulation. We call an oscillator *stimulated* if $I > 0$, and *unstimulated* if $I < 0$. The dependency on I to

oscillate makes the oscillator *stimulus-dependent*. The parameter β controls the steepness of the sigmoid function, and is chosen to be small so that the sigmoid is close to a binary function (see Fig. 2).

Except for the introduction of the Heaviside function, Eq. (1) is the same as defined in the Terman and Wang oscillator model (Terman & Wang, 1995). It is this Heaviside term that allows the network to exhibit object selection. The coupling term S_i in Eq. (1a) is given by

$$S_i = \sum_{k \in N(i)} W_{ik} H(x_k - \theta_x) - W_z H(z_f - \theta_z) \qquad (2)$$

where W_{ik} is the connection weight from the oscillator k to oscillator i. $N(i)$, called the neighborhood of i, is the set of neighboring oscillators that connect to i. For the 2-D grid shown in Fig. 1, $N(i)$ are four nearest neighbors of i with no boundary wrap-around. The parameter θ_x is a threshold above which an oscillator can affect others in the network. Its value lies between LB and RB so that $x_i > \theta_x$ when oscillator i is on RB and $x_i < \theta_x$ when i is on LB. W_z is the weight of inhibition from the fast inhibitor z_f, whose activity is defined as

$$\dot{z}_f = \phi\left(\sum_k H(x_k - \theta_z) - z_f\right) \qquad (3)$$

where ϕ determines the rate at which the inhibitor reacts to stimulation, and it is chosen to be on the order of 1 resulting in fast dynamics[1]. θ_z is another threshold, and like θ_x it is chosen between LB and RB. We set $\theta_z > \theta_x$ and $1 > \theta_z > 0$ (see Terman and Wang, 1995, for the justification in choosing the threshold). It is clear from Eq. (3) that z_f approaches the number of oscillators in the active phase on the fast time scale.

In the following implementation connection weights to an oscillator are assumed to be normalized according to the mechanism of dynamic normalization (Wang, 1995). Dynamic normalization ensures that the weights of all connections to any stimulated oscillator quickly settle to a constant, W_T. Wang and Terman (1997) provided differential equations to realize such normalization. It is worth noting, however, that weight normalization is not a necessary condition for LEGION to work (Terman & Wang, 1995). Normalized weights serve the useful purpose of improving the quality of synchronization.

Coming back to the Heaviside term in Eq. (1a), r_i is defined as

$$\dot{r}_i = \lambda[z_f - r_i]^- H(x_i - \theta_x)\omega(\dot{z}_f, \delta) \qquad (4)$$

Here the function $[v]^- = v$ if $v < 0$ and $[v]^- = 0$ otherwise, and the window function $\omega(v, d) = 1$ if $|v| \leq d$ and 0 otherwise. The window function, within which δ is a small positive parameter, ensures that r_i does not change when z_f transiently increases or decreases. The parameter λ is

[1] Recall that slow dynamics is dictated by ε.

chosen on the order of 1. Notice that the rate of change in z_f is used in the definition of r_i dynamics. Such a rate change is easily detected in the nervous system by an onset or offset detector, and we directly use $\dot{z}_f$ in Eq. (4) for the sake of simplicity. It is easy to see from Eq. (4) that r_i changes its value only when both the Heaviside and the window functions are 1. These two conditions together assure that the oscillator group that i belongs to is in the active phase. More discussions are given in Section 3 when the network behavior is analyzed. When r_i is ready to change, it approaches, or reduces to, the value of the fast inhibitor on the fast time scale when $r_i > \dot{z}_f$, otherwise it remains unchanged. Thus, r_i records the number of oscillators that are in the active phase when i is in the active phase, and for this reason r_i is called the *residual potential*.

The activity of the slow inhibitor is defined as

$$\dot{z}_s = \psi\left[\sum_k H(x_k - \theta_{zx}) - z_s\right]^+ - \mu\varepsilon z_s. \qquad (5)$$

Here the function $[v]^+ = v$ if $v \geq 0$ and $[v]^+ = 0$ otherwise. Both parameters ψ and μ are on the order of 1, which, due to ε, result in quick rise and slow decay for the slow inhibitor.

Let n be the total number of the oscillators in the network. It is easy to see from (5) that the maximum value of z_s is n. At the beginning we want all stimulated oscillators to be enabled. This is achieved by setting $r_i = C \cdot n$ so that initially the Heaviside in Eq. (1a) equals 1 for every oscillator.

3. Model behavior

With the Heaviside in Eq. (1a) set to 1, the system defined above simplifies to a standard LEGION network with the fast inhibitor only, which has been extensively analyzed by Terman and Wang (1995). Let a *pattern* be a connected region, and a *block* be a set of oscillators stimulated by a pattern. With ε sufficiently small, LEGION exhibits the mechanism of *selective gating*, whereby an enabled oscillator jumping up to the active phase rapidly recruits a block of oscillators stimulated by the same pattern, while preventing others from jumping up. The selective gating mechanism leads the network to rapidly achieve both synchronization within each block and desynchronization between different blocks. Desynchronization between two blocks means that they are never in the active phase simultaneously. In addition, the overall time the system takes to achieve both synchronization and desynchronization is no greater than m cycles of oscillations, where m is the number of patterns in the input image. The analytical results are established independent of the size and dimensionality of the network, and extend to lateral connections beyond nearest-neighbor coupling.

With the Heaviside in Eq. (1a) equal to 0 for an oscillator, the oscillator is excitable but can be recruited to oscillate by its neighbors through the S term in Eq. (1a). However, if the Heaviside is 0 for an entire block, then the block cannot
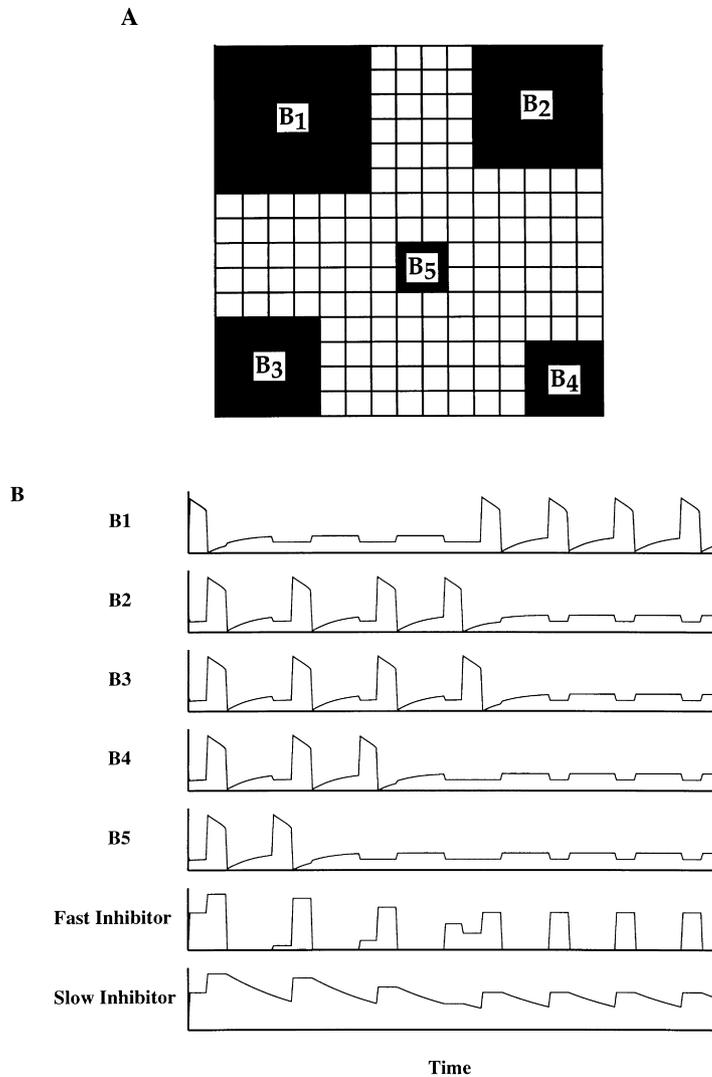
A

B

Fig. 3. (A) Input scene with five square patterns. The scene is mapped to a 15 × 15 selection network. Each pattern is labeled differently in (B). The upper five traces show the combined x activities of the five oscillator blocks indicated by their respective labels. These activities are normalized. The lower two traces show the normalized activities of the two global inhibitors. The simulation corresponds to the period from $t = 0$ to 45.

oscillate. This is the mechanism behind object (block) selection.

Let us assume that each block can synchronize in the first oscillation cycle. From the analysis in Terman and Wang (1995), the assumption is achieved by starting all the oscillators on LB only in the zone from the left knee (LK) up to the lowest LK of all the excited oscillators. In particular, the y value of LK is I (see Fig. 2(A)). Let $W_{min} = Min_{i,k \in N(i)} W_{ik}$. When a single neighbor of an oscillator jumps to RB, with $W_{min} > W_z$, the minimum positive shift of the nullcline for the oscillator is $W_{min} - W_z$. Thus, the zone $[I, I + W_{min} - W_z]$ satisfies the assumption.

In Eq. (4), the residual potential of oscillator i through z_f records the number of oscillators that synchronize with i. As z_f changes its activity quickly during the jumping process of an oscillator block, r_i may undesirably reduce its value in the transient jumping process. The ω function in Eq. (4) serves

to avoid this situation. As each block synchronizes in the first cycle, the residual potential is the same for all the oscillators in the same block, and r_i equals the size of the block if the block is segmented. The slow inhibitor through its quick rise and slow decay maintains a level of inhibition that must be overcome by a block if the block is to oscillate. The central idea for object selection is that the largest block sets the level of slow inhibition, which can then be overcome by that block only. The key is to choose an appropriate value for C in Eq. (1a).

When the largest block jumps to the active phase, z_s approaches the size of the block, say s_M, quickly. z_s decays when the block is on LB. Recently, Linsay and Wang calculated the time that a block spends on RB and LB in the singular limit $\varepsilon \to 0$ (Linsay & Wang, 1998). In particular, the time on LB is $\tau_L = \ln (I_T + 4)/I$, where $I_T = I + W_T - W_z$. According to Eq. (5), z_s decays to
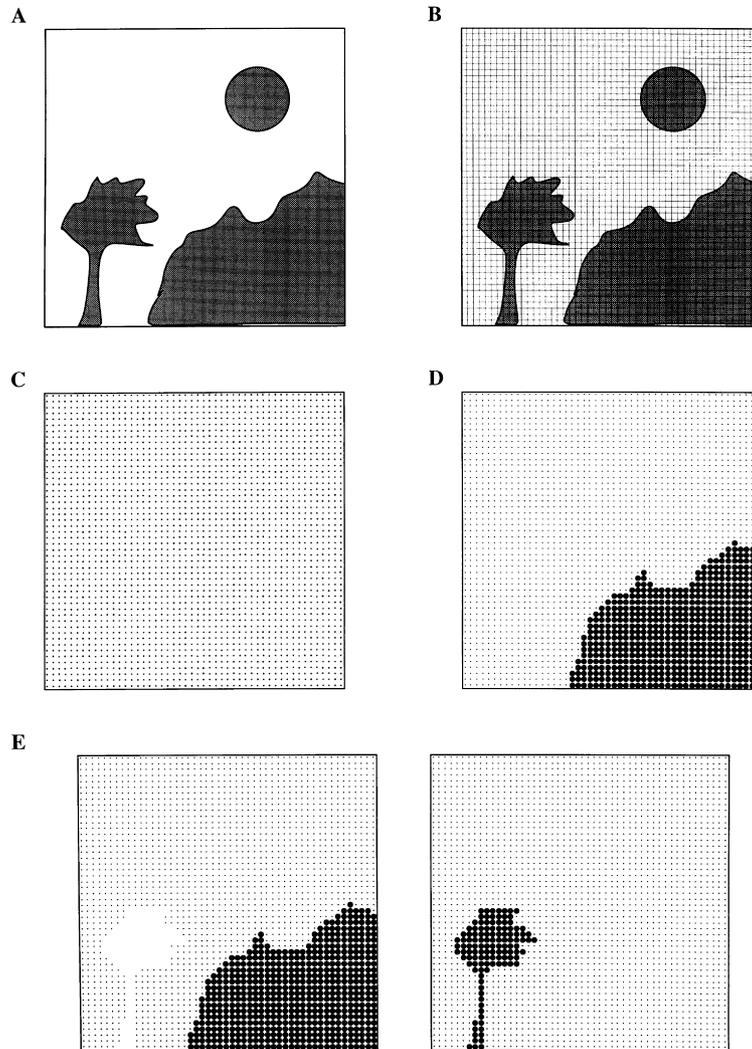
Fig. 4. (A) Input scene with three patterns. The same scene was used in Terman and Wang (1995). (B) The input scene of (A) is mapped to a 50×50 selection network. If and only if a square is entirely covered by the input is the corresponding oscillator considered externally stimulated. (C) The initial snapshot of the network. (D) A snapshot taken several cycles after the beginning. Only one pattern (*mountain*) is selected. (E) Two patterns (*mountain* and *tree*) are selected with a reduced C. The two subsequent snapshots are taken several cycles after the beginning.

$s_M\exp(-\mu\tau_L)$. Thus, if we set C to

$$C_M = \exp(\mu\tau_L) \qquad (6)$$

then this largest block is just able to jump after it spends τ_L on LB. Since during the time when the block is in the silent phase $z_s > s_M$, no other block can jump to the active phase. As soon as the block jumps to RB, the above dynamics repeats itself. Thus only the largest block can oscillate, and all other blocks are prevented from oscillating by the slow inhibitor. It is noted that C_M in Eq. (6) is independent of the specific size of the block. It is also noted that, though the external input I is factored in Eq. (6) through τ_L, C_M is independent of specific pixel intensities in an image for the following reasons. For a binary image, the external input is constant for all stimulated oscillators. For a gray-level image, as implemented in Wang and Terman (1997), the external input is set to a constant, and in this case intensity values are embodied in connection weights between

oscillators; the same is true in Section 6 when gray-level images are addressed. Thus, intensity values of a gray-level image do not enter Eq. (6). The above two notes together imply an important property: the parameter C_M does not need to be adjusted with specific input images.

The above straightforward analysis in terms of blocks does not apply in the context of LEGION, where different blocks may synchronize because of *rebound*, i.e., with the release of fast inhibition more than one block may jump to the active phase simultaneously. As described in Wang and Terman (1997), the LEGION network has a limited segmentation capacity. Taken together, given an arbitrary number of blocks the system in any cycle can produce at most as many segments as the capacity, where a segment is a synchronized oscillator population that desynchronizes from any other segment. Thus blocks may be arbitrarily grouped together, and r_i may have a value corresponding to a number of blocks. Wang and Terman (1997) refer to

A

Sun

Tree

Mountain

Fast Inhibitor

Slow Inhibitor

Time

B

Sun

Tree

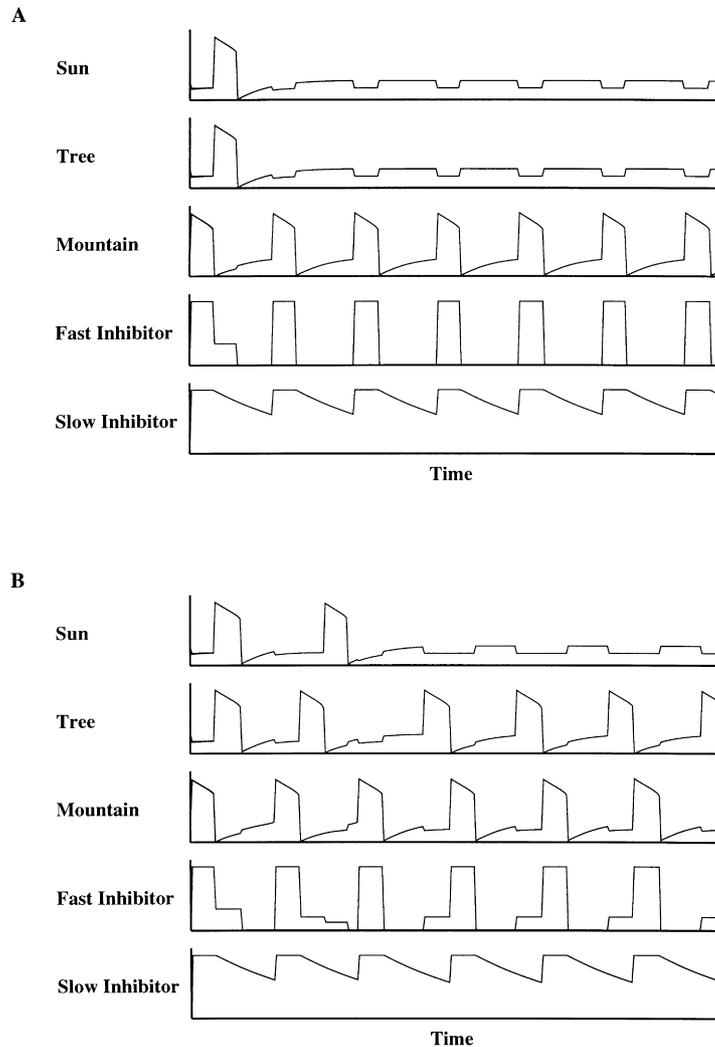Mountain

Fast Inhibitor

Slow Inhibitor

Time

Fig. 5. Temporal process of selection. (A) The upper three traces show the activities of the three oscillator blocks indicated by their respective labels. The lower two traces show the activities of the two global inhibitors. (B) The same display as in (A) but with a reduced C. See Fig. 3 legend for more explanations. Both simulations correspond to the same period from $t = 0$ to 36.

one-block segments as *simple segments* and others as *congregate segments*.

Fortunately, the analysis in terms of blocks applies similarly in terms of segments because each segment is synchronized and to the global inhibitors there is no difference between a block and a segment. Thus, after each cycle only the largest segment can oscillate. We claim, based on LEGION dynamics (Terman & Wang, 1995; Wang & Terman, 1997), that after each cycle at least one segment becomes a simple one. It is noted that the largest segment is at least as large as the largest block. Thus, if the simple segments separated in each cycle are not the largest block they will be prevented from oscillating in the future. If the largest block is separated as a simple segment, it may be prevented from oscillating temporarily, for it is possible that congregate segments are larger than the largest block. But at most by the mth cycle, the largest block will be the only block to oscillate. Thus object selection is guaranteed to be achieved.

The above process is illustrated in Fig. 3 with a scene having five square blocks of different sizes. The scene is mapped to a 15×15 selection network shown in Fig. 3(A). Each block has a distinct label. The details of the simulation will be discussed in Section 4; here it suffices to point out that the setup and the parameter values used in this simulation are exactly same as given in Section 4 when discussing Figs. 4 and 5. Fig. 3(B) shows one possible selection process, where the oscillators corresponding to the same block are combined in the display. The five upper traces show the activities of the five blocks, respectively, and the bottom two traces show the activities of the fast and the slow inhibitor, respectively. In this case, though the largest block, B_1, is separated in the first cycle, it is temporarily inhibited because the other segment separated in the first cycle—a congregate segment consisting of the remaining four blocks—is larger. The congregate segment formed in the first cycle is further segregated in the next several cycles, and by the fifth cycle no more active segment is larger than

B_1. Thus B_1 is activated in that cycle, and becomes the only winning block.

In the competition between blocks, it is possible that large blocks that are separated already may jump to the active phase together due to rebound. This intricate situation arises because z_s may continue decreasing when a small segment is in the active phase. During the time when the small segment is on RB, the decrease in z_s may be sufficient to allow more than one block to overcome slow inhibition. Thus one must be careful when analyzing the speed of selection. This consideration motivates us to define a non-increasing residual potential, because otherwise it can undo previous segmentation and cause a much slower selection process. With the non-increasing potential, temporary grouping between segmented blocks cause no harm except that it lengthens the oscillation cycle somewhat.

As a cycle in the selection process may be longer than that of a single oscillator, our conclusion is that in at most $O(m)$ cycles only the largest block is enabled. This is summarized in the following proposition.

Proposition 1. *The parameters in systems (1)– (5) can be chosen so that the system selects the largest block in* O(m) *cycles of oscillations, where* m *is the number of patterns in the input image.*

When $C > C_M$, the above behavior remains the same. But each cycle is longer, resulting in a slower selection process. This is because when the largest block arrives at LK, it cannot jump right away because the slow inhibition is still too high to overcome. On the contrary, when $C < C_M$, more than one block may survive the selection process and oscillate. This is because when the largest block reaches LK, only at which it can jump, z_s decays to a value smaller than s_M, which may be overcome by another block. Systematically reducing C can identify the second largest block. How small C should be to yield two winning objects depends on the ratio of the size of the second largest block to that of the largest one. As a special case, if there are two largest blocks of the same size, the network with a C value slightly smaller than C_M will select both blocks, which oscillate desynchronously.

The choice of μ in Eq. (5) also affects the speed of selection. If μ is very small, the system will take a long time to overcome the slow inhibition triggered by a large congregate segment. In contrast, if μ is very large, C_M must be large and magnify slow inhibition (see Eq. (1a)) by a large factor. This will lengthen oscillation cycles, because no block can jump up even when a small block jumps down.

4. Computer simulation

We illustrate how object selection is performed by simu-

lating a 50×50 selection network. We arbitrarily selected an input scene with three objects (designated by shape as the *sun*, a *tree*, and a *mountain*) as shown in Fig. 4(A). The scene is mapped to the network as shown in Fig. 4(B). In this simulation, $N(i)$ contains four nearest neighbors with no boundary wrap-around. For a stimulated oscillator, $I = 0.2$; otherwise, $I = -0.02$. Eqs. (1)–(5) were solved using the singular limit method (Linsay & Wang, 1998), which is a fast numerical method for integrating relaxation oscillator networks in the singular limit $\varepsilon \to 0$. When using the singular limit method, the window function in Eq. (4) is not needed, and neither are the detailed values for the following parameters: $\varepsilon, \rho, \beta, \phi, \lambda, \psi, \theta_x, \theta_z$. The following values were used for the remaining parameters: $\alpha = 6.5$, $W_z = 1.5$, $W_T = 8.0$, and $\mu = 0.125$. For the above parameters, $C_M = 1.6407$.[2] Thus, we choose $C = 1.64$, to be slightly less than C_M. The following simulation results have been confirmed by using a fourth-order Runge–Kutta method with $\varepsilon = 0.02$; the Runge–Kutta method is not used here because integration of large-scale systems as needed in Section 6 would be too intensive computationally.

Fig. 4(C) shows the initial instantaneous activity (snapshot) of the selection network. In the display, the diameter of each circle indicates the x activity of the corresponding oscillator. More specifically, the diameter is proportional to $(x - x_{min})/(x_{max} - x_{min})$, where x_{min} and x_{max} indicate the minimum and the maximum of x values of all oscillators. Fig. 4(D) shows a snapshot after the system reaches stable limit cycles, when the only active block is the one corresponding to the *mountain*—the largest object in the scene. The oscillators corresponding to the other two objects are prevented from oscillating by the slow inhibitor. To display the entire competition process, Fig. 5(A) shows the temporal evolution of every stimulated oscillator, where the oscillators belonging to the same block are combined in the display. The unstimulated oscillators remain excitable for the entire simulation, and are excluded from the display. The three upper traces represent the activities of the three oscillator blocks respectively, and the fourth and the fifth ones represent the activities of the fast and the slow inhibitors, respectively. It is noted that numerical integration using the singular limit method makes an oscillator block appear to be perfectly synchronized and look like a single oscillator in Fig. 5. The selection process is completed after just one cycle.

To illustrate the role of C, the same network is simulated with C reduced to 0.3. With the same random initial conditions shown in Fig. 4(C), the selection network produces two winning objects: the *mountain* and the *tree*. This is shown in Fig. 4(E), where two snapshots are given after the system reaches stable limit cycles. The two active blocks are separated from each other, and the *sun* block is inhibited.

---

[2] Notice that in calculating I_T, I should be 0 instead of 0.2 because once a block is in the active phase the Heaviside in Eq. (1a) is 0 for its oscillators.
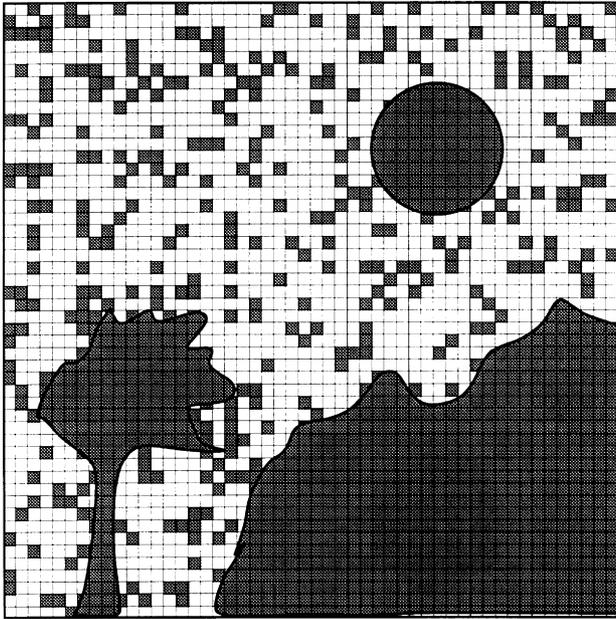
Fig. 6. The input scene of Fig. 4(B) with 20% noise added.

Fig. 5(B) gives the corresponding temporal activities of the network for the same duration as in Fig. 5(A).

5. LEGION as a preprocessing layer

As stated in Proposition 1, the time complexity of the selection process is $O(m)$. This complexity implies a very slow process if there are many objects in a scene. To be sure, a typical realistic scene does contain many objects due to the
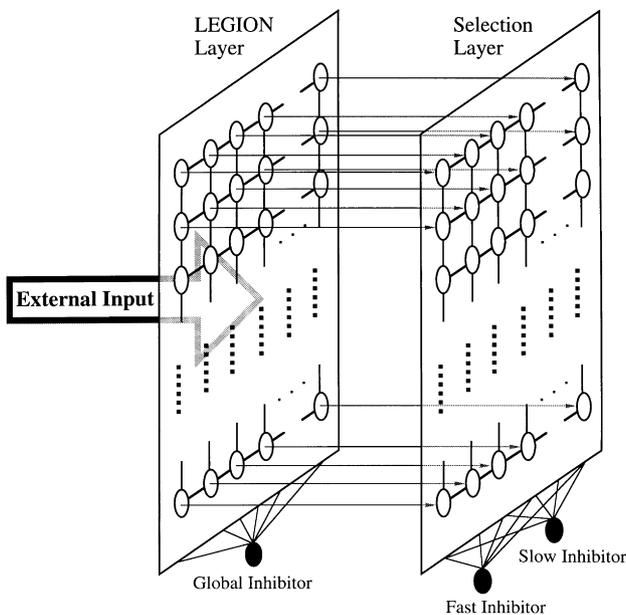

Fig. 7. Diagram of a two-stage selection network. External input is applied to the LEGION layer. The connections from the LEGION layer to the selection layer are one-to-one correspondence. See Fig. 2 legend for notation.

richness and noise of the real world. For example, Fig. 6 shows the same input as Fig. 4(B), with 20% noise added so that each uncovered box has a 20% chance of being covered (stimulated). With four nearest-neighbor coupling, Fig. 6 gives rise to hundreds of oscillator blocks. Thus, even though the selection network can eventually identify the largest block, it takes a very long time.

As described in Wang and Terman (1997), a LEGION network with a lateral potential for each oscillator has the ability to remove small noisy regions. This consideration and psychophysical theories of preattentive processing lead us to suggest a two-stage process for object selection. The first stage corresponds to a LEGION layer, which organizes the input scene into a number of major regions and a background that is the rest of the scene. The second stage corresponds to the selection layer as described in Section 2 (see Fig. 1). Fig. 7 shows the two layer network. The connections from the LEGION layer to the selection layer are in a one-to-one correspondence.

More specifically, S_i in the selection layer includes a new term $W \cdot H(\hat{x}_i - \theta_x)$, where W indicates a uniform connection strength and $\hat{x}_i$ indicates the x value of oscillator i in the first layer. In addition, I_i is set to be high if and only if the corresponding oscillator in the first layer is enabled. Thus, the oscillators in the selection layer corresponding to the background of the input scene will have low I values. With the coupling term $W \cdot H(\hat{x}_i - \theta_x)$, the second layer can synchronize with the first layer. Thus, an oscillator block formed in the first layer transfers to the second one. As stated in Section 3, lateral connections in the selection layer have normalized dynamic weights, which are formed on the basis of synchronization between neighboring oscillators. Given transferred block structures it is straightforward to form such dynamic weights based on Hebbian learning. See Wang and Terman (1997) for a specific way of implementing dynamic normalization. Thus, after a beginning period, non-zero dynamic connections are established only between the oscillators of the same block.

The coupling between the two layers is designed so that W switches to 0 after the LEGION layer completes its segmentation process. With W equal to 0, the selection layer behaves on its own except for the I term that is stable by then. Thus, the selection process takes place. Given the fact that the blocks in the selection layer are already separated, one more cycle is sufficient to suppress all except the largest block. The LEGION layer completes its segmentation in no greater than $M + 1$ cycles, where M is the number of major image regions (Wang & Terman, 1997).[3] Thus, in at most $M + 2$ cycles, there is only one active block in the selection layer—the one corresponding to the largest object.

A question arises from the above scheme of coupling between the two layers: how does the selection layer know when the LEGION layer has completed its segmentation? As analyzed in Wang and Terman (1997), out of $M +$

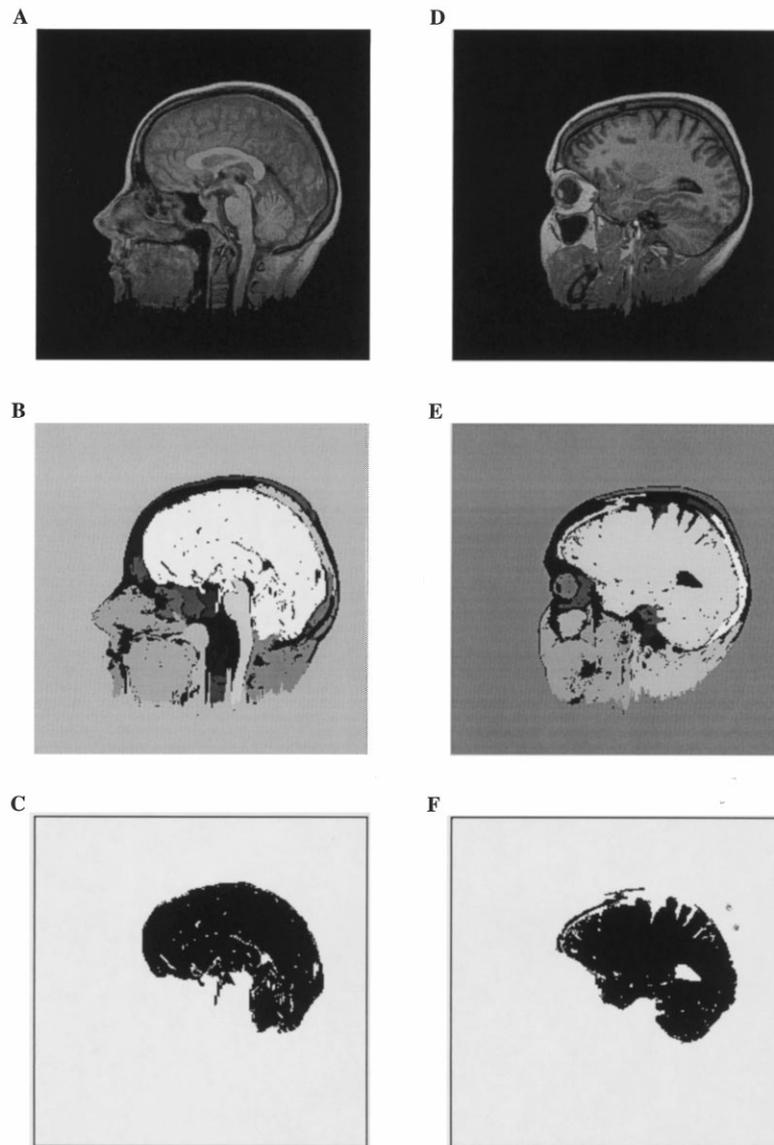[3] The technical report in particular contains the details of analysis.

Fig. 8. (A) An MRI image consisting of 257 × 257 gray-level pixels (courtesy of N. Shareef). (B) A gray map showing the segmentation result of the image in (A) by a 257 × 257 LEGION layer. (C) A snapshot of the selection layer showing the result of selection. Only one region (upper part of the brain) is selected by the network. (D) Another MRI image consisting of 257 × 257 pixels (courtesy of N. Shareef). (E) A gray map showing the segmentation result of the image in (D) by a 257 × 257 LEGION layer. (F) A snapshot of the selection layer showing the result of selection. Only one region (the brain) is selected by the network.

1 cycles for full segmentation, small noisy regions can be eliminated by LEGION within the first two cycles, and the remaining cycles (if any) are for segmenting major regions. The task of segmentation, however, can be performed by the selection layer equally well. This suggests an alternative scheme for coordinating the coupling; switch off W after the first two cycles. The above conclusion on the overall speed of selection still holds by similar analysis. That is, in at most $M + 2$ cycles the largest block is the only one active block in the selection layer.

For input scenes with many insignificant regions, the two-layer network represents remarkable speedup for object selection. The speedup comes from the fact that noisy regions are removed *in parallel* by the LEGION layer.

Coming back to the scene in Fig. 6, the LEGION layer segments it into three major regions, corresponding to the *sun*, the *tree*, and the *mountain*, respectively, as well as the background corresponding to the remaining stimulated oscillators (see Wang & Terman, 1997). After that, the selection layer takes one more cycle to select the mountain region, giving a similar result as Fig. 4(D). The entire process takes five cycles, four of which are used for segmentation (Wang & Terman, 1997).

6. Gray-level images

When the two-layer selection network is applied to a

External Input

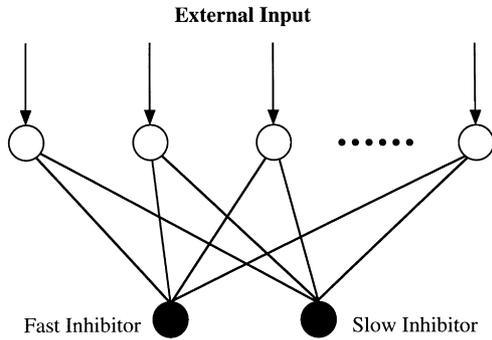Fast Inhibitor   Slow Inhibitor

Fig. 9. Diagram of an oscillatory WTA network. Each oscillator receives an external input, and has mutual connections with two global inhibitors.

gray-level image, the LEGION layer segments the image into a number of major regions, each region triggering an oscillator block in the selection layer. The blocks thus formed in the second layer compete with each other, and the largest one will win the competition and become the only block selected by the overall network. It is noticed that the selection layer does not deal with gray-level images directly, but with oscillatory activity from the LEGION layer.

To relate the previous study on LEGION (Wang & Terman, 1997), we use two images that have been processed by LEGION to illustrate object selection. These are two MRI (magnetic resonance imaging) images of the human head. The first one, shown in Fig. 8(A), is a midsagittal section with 257×257 gray-level pixels. When presented to the LEGION layer with a certain parameter setting, the image is segmented to 21 regions plus a background as shown in Fig. 8(B) (see Wang & Terman, 1997). In the figure, we use gray levels to indicate the phases of the oscillator blocks, resulting in a *gray map*. The background is indicated by scattered black areas. The salient regions segmented include two parts of the brain: the upper part and the brainstem with a part of the spinal cord, parts of the extracranial tissue, parts of the bone marrow, the neck part, the chin part, the nose part, and the vertebral segment. Like the LEGION layer, the selection layer has eight nearest-neighbor coupling within itself. The layer is integrated using the singular limit method with the following parameter values: $\gamma = 6.5$, $W_z = 0.7$, $W_T = 8.0$, and $\mu = 0.125$. For these parameter values, $C_M = 1.6558$. Thus, we choose $C = 1.65$. In the selection, we ignore the homogeneous background region, which is the largest (see Fig. 8(B)). Fig. 8(C) shows the output of the selection layer. The only selected region is the upper part of the brain, which is the largest segment from the LEGION layer.

The second MRI image is shown in Fig. 8(D), which is a sagittal section through one eye. The image has the same size as the one in Fig. 8(A). Again, the image is first segmented by the LEGION layer, yielding 17 regions plus a background. Fig. 8(E) shows the gray map of the segmented regions and the background is indicated by scattered black regions. The salient regions segmented include the

entire brain, the eye, the sinus, parts of the bone marrow, and parts of the extracranial tissue. The selection layer has the same configuration and the same parameter values as for processing the first MRI image. Again, we ignore the homogeneous background region. Fig. 8(F) shows the result of selection. The brain region, which is the largest segment from the first layer, is selected by the two-layer network.

As discussed in Wang and Terman (1997), different parameter settings for the LEGION layer produce different segmentation results. For example, when the level of global inhibition is increased, the LEGION layer can further segment the upper part of the brain (Fig. 8(B)) into the cerebral cortex, the cerebellum, the callosum/fornix regions and its surrounding septum (Wang & Terman, 1997). Regardless of parameter settings in the first layer, the selection layer works in the same way; it always selects the largest segment generated by the first layer.

7. WTA as a special case

The selection model described above readily yields a new WTA network based on oscillatory dynamics. To turn the selection network of Fig. 1 to a WTA network, the following modifications are sufficient. First, the lateral connections between the oscillators are removed, resulting in a group of oscillators communicating through the two global inhibitors only. Thus, $S_i = -W_z H(z_f - \theta_{xz})$. This architecture is shown in Fig. 9, and it is a familiar one for WTA (see among others, Didday, 1970; Amari & Arbib, 1977; Ermentrout, 1992). Second, we let the external input to an oscillator, I_i, gate the input of the oscillator to the two inhibitors. More specifically, we define

$$\dot{z}_f = \phi\left(\sum_k I_k H(x_k - \theta_{zx}) - z_f\right) \tag{7}$$

$$\dot{z}_s = \psi\left[\sum_k I_k H(x_k - \theta_{zx}) - z_s\right]^+ - \mu\varepsilon z_s \tag{8}$$

Finally, the I term in Eq. (1a) is set to a constant value.

With the above modifications, each oscillator forms its own block, the size (or saliency) of which corresponds to the input level to the oscillator. Thus, the analysis in Section 3 applies. In particular, the dynamics for object selection guarantees that the oscillator with the largest input will win the competition and become the only active oscillator. Further, Proposition 1 implies that the time complexity for the WTA network is $O(m)$, where m is the number of inputs or oscillators. It is noted that as each oscillator forms a unique block, the window function in Eq. (4) is not needed.

8. Discussion

Our model can easily be extended to incorporate a decay mechanism so that the selection process shifts from one

object to another. The idea of using a decay mechanism was proposed by Koch and Ullman (1985) for producing attentional shifting. With such mechanism in place, our network can readily select the second largest object after the largest one is selected, and the third largest, etc. Thus, it implements a form of shifting in object selection. This occurs because slow inhibition decays if the slow inhibitor is not stimulated by any active block, and eventually the next largest block can overcome slow inhibition and oscillate.

Object selection studied so far is based on the size of an object. Obviously size is the only one aspect of object saliency (Treisman & Gormican, 1988; Desimone & Duncan, 1995), and it is our intention to use size as a concrete way of embodying the saliency, In Section 7, stimulus strength is used as the saliency measure. In general, saliency includes other measures such as brightness, color, motion, texture, context, and familiarity. The selection framework described here may be extended to include other measures. For example, if motion should be an important measure, we can assign moving objects higher magnitudes in competition much like what we did for stimulus input in Section 7. This way, a smaller moving object can win over a large stationary object.

The architecture of the selection network and the oscillations used are very similar to the LEGION network, whose biological plausibility has been discussed at length elsewhere (Terman & Wang, 1995; Wang & Terman, 1997). We shall not repeat these arguments, but merely point out that both the neural oscillations and the basic network architecture are consistent with the neurobiological data. In addition, slow inhibition, which plays an important role in this work, is frequently used in biologically realistic models (Wang & Rinzel, 1992; LoFaro, Kopell, Marder & Hooper, 1994; Terman & Lee, 1997).

A WTA network incorporating oscillatory dynamics has been studied by Ermentrout (1992). In his model, a basic unit of the network is not intrinsically oscillatory, but can generate oscillations together with a global inhibitor that couples with every unit. He observes stationary WTA behavior with fast global inhibition and network oscillations as inhibition slows down. Unlike his model, our basic unit is an intrinsic oscillator, which results in very different network dynamics. In Niebur and Koch (1996), an architecture similar to that described in Section 7 is used for producing WTA behavior. In their model, each basic unit is an integrate-and-fire neuron which resembles an oscillator. WTA is achieved through strong global inhibition, resulting in that the neuron with the highest input is the only neuron that generates spikes. No analysis regarding the rate of competition is given for either model. The main difference between our selection network and these two models is that the selection network incorporates local coupling between the oscillators, which then leads to organization of the input image and selection at the object level. To our knowledge, the same distinction holds with regard to all other WTA models.

Conceptually, besides competition as used in all forms of

WTA dynamics, our selection network includes cooperation between neighboring oscillators in the network. The cooperation between elements that likely form the same object gives rise to object selection, as opposed to pixel selection. In other words, the selection network incorporates *both* competition through global inhibition and cooperation through local excitation.

Oscillatory correlation is usually used as a mechanism of binding features into objects in the presence of multiple objects. Given that object selection usually selects only one item, one might ask the question, why bother with oscillatory correlation in the first place? It is true that oscillatory correlation as a representation is not needed for representing the *result* of selection, if only one object is selected. However, the dynamics of our model shows that the *process* of generating the selection is intimately related to oscillatory correlation. In the process, multiple organizations participate in competition, and each organization corresponds to a synchronized segment. Moreover, the slow inhibitor, which plays a critical role in object selection, depends on oscillatory recharge from the winning object in order to prevent other organizations from being activated. As noted above, other WTA models have not demonstrated selection at the object level. Also, as far as we know, no computational models have addressed object-based selection systematically. In addition, the flexibility of our model in selecting more than one object may prove to be a very useful property, given that it is not a clear-cut phenomenon that attentional mechanisms select just one organization (Neumann & Sanders, 1996).

The two-layer selection network (Fig. 7) may provide a neurocomputational framework for addressing the issues of selective attention and perceptual organization. The LEGION layer may be viewed to correspond to the stage of perceptual organization or preattentive processing. The psychophysical studies suggest that processing in this stage is parallel and yields multiple organizations that feed to the subsequent attentive stage (Mack, Tang, Tuma, Kahn & Rock, 1992; Trick & Pylyshyn, 1994; Julesz, 1995). Because of the parallel nature of LEGION, regions or organizations appear to pop out instantaneously. The properties of the LEGION layer are consistent with the stage of perceptual organization. In particular, the emphasis of LEGION on emergent synchrony based on local connectivity, which reflects both connectedness and proximity, is strikingly similar to a recent theoretical account of perceptual organization by Palmer and Rock (1994). Their theory attaches special importance to the so-called uniform connectedness, and regards the formation of uniform and connected regions as fundamental.

Computationally, the first layer in the selection network performs parallel removal of small noisy regions in an image. Given that the selection layer does not require oscillatory input, the LEGION layer may be replaced by alternative neural network models for image segmentation (for a review see Wang & Terman, 1997), such as the model of

Grossberg and Wyse which combines a boundary contour module and a feature contour module (Grossberg & Wyse, 1991; Grossberg & Wyse, 1992. For more recent developments of this approach, including how to treat object occlusion, the reader is referred to Grossberg (1994; 1997) and Grossberg and McLoughlin (1997). LEGION is chosen so that our entire selection network is based on the oscillatory correlation framework.

The selection layer may be viewed to correspond to the stage of selective attention. It acts on the results of the first stage, and selects one region based on its saliency from candidate regions provided by the first stage. That the selection layer acts directly on regions (surfaces) is well consistent with psychophysical and neuropsychological evidence suggesting that visual attention is surface-based (Nakayama et al., 1995; Mattingley et al., 1997). The processing time in the selection layer is a linear relation to the number of objects, reminiscent of sequential search. All these features of the selection network are consistent with the processing of selective attention, which is thought to be based on segment-by-segment scrutiny and serial shifting (Treisman, 1988; Julesz, 1995; Neumann and Sanders, 1996). Section 5 demonstrates the utility of a parallel process preceding the selection in terms of computational complexity. We consider the general agreement between the model and the perceptual phenomena very encouraging. However, an in-depth assessment of the selection network as a model of selective attention is a topic of future research.

Acknowledgements

References

Abeles, M. (1991). *Corticonics: neural circuits of the cerebral cortex*. New York: Cambridge University Press.

Amari, S., & Arbib, M. A. (1977). Competition and cooperation in neural nets. In J. Metzler, *Systems neuroscience* (pp. 119–165). New York: Academic Press.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

Didday, R. L., (1970). The simulation and modeling of distributed information processing in the frog visual system. PhD dissertation, Stanford University.

Egeth, H. E., & Yantis, S. (1997). Visual attention: control, representation, and time course. *Annual Review of Psychology*, *48*, 269–297.

Ermentrout, G. B. (1992). Complex dynamics in winner-take-all neural nets with slow inhibition. *Neural Networks*, *5*, 415–431.

Grossberg, S. (1976). Adaptive pattern classification and universal recording: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics*, *23*, 121–134.

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics*, *55*, 48–120.

Grossberg, S. (1997). Cortical dynamics of three-dimensional figure-ground perception of two-dimensional pictures. *Psychological Review*, *104*, 618–658.

Grossberg, S., & McLoughlin, N. P. (1997). Cortical dynamics of three-dimensional surface perception: Binocular and half-occluded scenic images. *Neural Networks*, *10*, 1583–1605.

Grossberg, S., & Wyse, L. (1991). A neural network architecture for figure-ground separation of connected scenic figures. *Neural Networks*, *4*, 723–742.

Grossberg, S., & Wyse, L. (1992). Figure-ground separation of connected scenic figures: Boundaries, filling-in, and opponent processing. In G. A. Carpenter & S. Grossberg, *Neural networks for vision and image processing* (161–194). Cambridge, MA: MIT Press.

Julesz, B. (1995). Dialogues on perception. Cambridge, MA: MIT Press.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227.

Linsay, P. S., & Wang, D. L. (1998). Fast numerical integration of relaxation oscillator networks based on singular limit solutions. *IEEE Transactions on Neural Networks*, *9*, 523–532.

Lippmann, R. P. (1987). An introduction to computing with neural nets. *IEEE Acoustics, Speech and Signal Processing Magazine*, pp. 4–22.

LoFaro, T., Kopell, N., Marder, E., & Hooper, S. L. (1994). Subharmonic coordination in networks of neurons with slow conductances. *Neural Computation*, *6*, 69–84.

Mack, A., Tang, B., Tuma, R., Kahn, S., & Rock, I. (1992). Perceptual organization and attention. *Cognitive Psychology*, *24*, 475–501.

Mattingley, J. B., Davis, G., & Driver, J. (1997). Preattentive filling-in of visual surfaces in parietal extinction. *Science*, *275*, 671–674.

von der Malsburg, C. (1981). The correlation theory of brain function. Internal Report 81-2, Max-Planck-Institute for Biophysical Chemistry.

Milner, P. M. (1974). A model for visual shape recognition. *Psychological Review*, *81* (6), 521–535.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S. M. Kosslyn & D. N. Osherson, *An invitation to cognitive science*, (pp. 1–70). Cambridge, MA: MIT Press.

Neumann, O. & Sanders, A. F. (Eds.). (1996). *Handbook of perception and action: Attention*. San Diego: Academic Press.

Niebur, E., & Koch, C. (1996). Control of selective visual attention. In D. Touretzky & M. C. Mozer & M. E. Hasselmo, *Advances in neural information processing systems* (pp. 802–808). Cambridge, MA: MIT Press.

Niebur, E., Koch, C., & Rosin, C. (1993). An oscillation-based model for the neuronal basis of attention. *Vision Research*, *33*, 2789–2802.

Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: the role of uniform connectedness. *Psychological Bulletin & Review*, *1*, 29–55.

van der Pol, B. (1926). On 'relaxation oscillations'. *Philosophical Magazine*, *2* (11), 978–992.

Rumelhart, D. E., & Zipser, D. (1986). Feature discovery by competitive learning. In D. E. Rumelhart & J. L. McClelland, (pp. 151–193). *Parallel distributed processing*, 1. Cambridge, MA: MIT Press.

Singer, W. (1993). Synchronization of cortical activity and its putative role in information processing and learning. *Annual Review of Physiology*, *55*, 349–374.

Terman, D., & Lee, E. (1997). Partial synchronization in a network of neural oscillators. *SIAM Journal on Applied Mathematics*, *57*, 252–293.

Terman, D., & Wang, D. L. (1995). Global competition and local cooperation in a network of neural oscillators. *Physica D*, *81*, 148–176.

Treisman, A. (1988). Features and objects: The fourteenth Barlett memorial lecture. *Journal of Experimental Psychology*, *140A*, 201–237.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*, 15–48.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers

enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*, 80–102.

Tsotsos, J. K. (1995). Toward a computational model of visual attention. In T. V. Papathomas, C. Chubb, A. Gorea & E. Kowler, *Early vision and beyond* (pp. 207–226). Cambridge, MA: MIT Press.

Wang, D. L. (1995). Emergent synchrony in locally coupled neural oscillators. *IEEE Transactions on Neural Networks*, *6* (4), 941–948.

Wang, D. L., & Terman, D. (1995). Locally excitatory globally inhibitory oscillator networks. *IEEE Transactions on Neural Networks*, *6* (1), 283–286.

Wang, D. L., & Terman, D. (1997). Image segmentation based on oscillatory correlation. *Neural Computation*, *9*, 805–836 (for errata see *Neural Computation*, 9 (1997) 1623-1626); see also Technical Report 19, The Ohio State University Center for Cognitive Science, Columbus, OH, 1996.

Wang, X. J., & Rinzel, J. (1992). Alternating and synchronous rhythms in reciprocally inhibitory model neurons. *Neural Computation*, *4*, 84–97.

Yuille, A. L., & Geiger, D. (1995). Winner-take-all mechanisms. In M. A. Arbib, *Handbook of brain theory and neural networks* (pp. 1056–1060). Cambridge, MA: MIT Press.