The time dimension for neural computation

DeLiang Wang

Center for Cognitive Science and Department of Computer & Information Science
The Ohio State University
Columbus, OH 43210-1277, U.S.A.
Email: dwang@cis.ohio-state.edu

Abstract

The binding problem refers to how sensory elements organize into perceived objects. The issue of binding is hotly debated in recent years in neuroscience and related communities. Much of the debate, however, gives little attention to computational considerations – a rather curious status as the problem is originally formulated from the computational perspective. This article starts with two problems considered by Rosenblatt to be the most challenging to the development of perceptron theory 40 years ago, and argues that the main challenge is the figure-ground separation problem, which is intrinsically related to the binding problem. The central claim of the article is that introducing the time dimension is essential for systematically attacking Rosenblatt's challenge. The temporal correlation theory as well as its special form - oscillatory correlation theory, is discussed as an adequate representation theory to address the binding problem in neural computation. A computational mechanism for the oscillatory correlation theory - LEGION dynamics - provides a solution to the Minsky-Papert connectedness problem, which is an important example of the binding problem, and the mechanism is successfully applied to a variety of scene segmentation tasks. The plausibility and implication of the oscillatory correlation theory are discussed at the physiological, perceptual, and cognitive levels. A number of controversial issues regarding oscillatory correlation are considered and clarified. Finally, the time dimension is argued to be necessary for versatile computation.

Keywords: binding, desynchronization, figure-ground separation, LEGION, oscillatory correlation, segmentation, synchronization, temporal correlation, time dimension, versatility.

1. Rosenblatt's challenge

In his classic book forty years ago, "Principles of neurodynamics," Frank Rosenblatt (1962) summarized in the last chapter a list of problems facing the study of perceptron theory at the time. Two problems among the list "represent the most baffling impediments to the advance of perceptron theory" (p. 580). These are the problems of figure-ground separation and the recognition of topological relations. The development of the field of neural networks in the ensuing forty years has largely validated the foresight of Rosenblatt. In particular, major progress has been made in the understanding of error-correction procedures for training multi-layer and recurrent perceptrons (Rumelhart & McClelland 1986; McClelland & Rumelhart 1986; Bishop 1995; Arbib 2002). The field of neural networks has been firmly established, and has enjoyed tremendous success both as a field of theoretical study for brain function and as a technology for solving pattern recognition and related problems. On the other hand, progress has been extremely limited in addressing Rosenblatt's two chief problems in the framework of perceptrons.

The figure-ground separation problem concerns how to separate a figure from its background in a scene. Since natural scenes generally contain multiple objects, a closely related problem is the scene segmentation problem - the segmentation of a scene into its constituent objects. Literally speaking these two problems are not the same, but they are frequently treated as the same with different emphases: one on extracting target object and another on parsing the entire scene. As a result, the two terms, figure-ground separation and scene segmentation, are often used interchangeably; such is the case in Rosenblatt's book. The recognition of topological relations concerns how to compute various spatial relations between objects in a scene, e.g. whether object *A* is inside object *B* or whether *B* is to the left of object *C*. Solving this spatial recognition problem would require a solution to figure-ground separation, and in this sense, the latter is a more basic problem. Both problems can be treated as major aspects of the scene analysis problem.
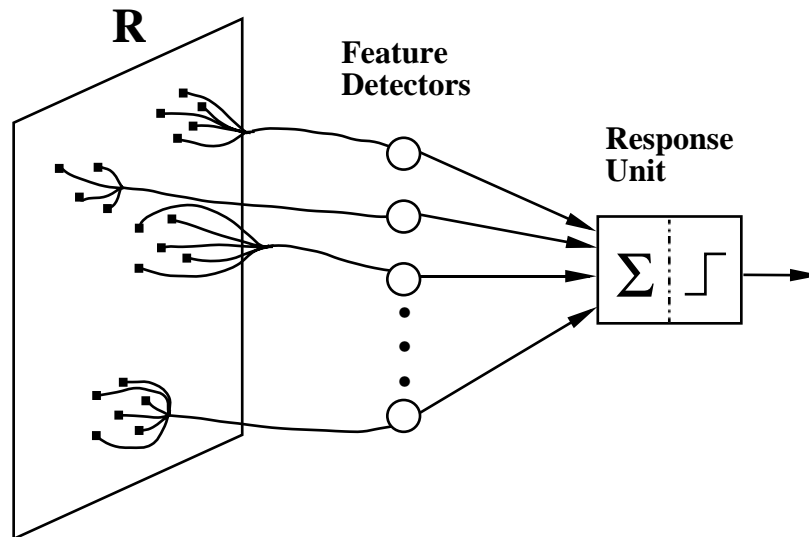
Figure 1. Network architecture of a perceptron. The input layer is denoted by *R*. Each feature detector receives input from a specific area of *R*. The response unit computes the weighted sum of all the detectors and checks whether the sum exceeds a certain threshold.

The deficiency of perceptrons for figure-ground separation was later hit hard in a landmark book by Minsky and Papert (1969). Through mathematical analysis, they pointed out that perceptrons are fundamentally limited in analyzing topological patterns. To better understand their mathematical results and perceptrons, let us give some details about perceptron theory. Perceptrons, introduced by Rosenblatt (1958; 1962), may be viewed as classification networks.

Figure 1 shows a typical perceptron that computes a predicate, which consists of a binary input layer R (symbolizing retina), a layer of binary feature-detecting units, and a response unit that represents the result of a binary classification. A feature detector senses a specific area of R, and it produces 1 if all of the pixels of the area are black and 0 otherwise. The response unit is a logic threshold unit whose input is a weighted sum of all the units in the detector layer and whose output is 1 if the sum exceeds the threshold and 0 otherwise. Minsky and Papert define the *order* of a predicate as the smallest number k for which one can compute the predicate with feature detectors that sense no more than k pixels of R. A predicate is topologically invariant, or is a topological predicate, if it is unchanged when the input figure undergoes distortion that does not alter the connectedness or inside-outside relationship among the parts of the figure.

Minsky and Papert (1969) proved that all except one topological predicate are of infinite order. That is, to compute such predicates requires feature detectors whose receptive fields are unbounded in size. One such predicate is connectedness: to tell whether an input pattern is connected or not. This predicate serves as the cornerstone for their analysis. The order of this predicate increases at least as fast as $|R|^{1/2}$. What are the implications of this result? For any fixed R, their theorem is not about *whether* a perceptron exists to solve the problem. Indeed, with a finite, discrete R, there are a finite number of connected patterns, and one can trivially find a perceptron whose feature detectors correspond to individual connected patterns and thus solve the problem. However, predicates that have unbounded order require large feature detectors relative to the size of R, and too many of them to be computationally feasible (Minsky & Papert 1969). For example, on a 2x2 R, the number of connected patterns is 13, and on a 3x3 R, it is 222. The number of connected patterns grows exponentially except for one-dimensional (1-D) R; see the Appendix for a proof. Clearly, this way of computing the connectedness predicate is computationally intractable for all but very small R's. In other words, their (negative) result is about the scalability or computational complexity of perceptrons.[1]

The discovery of the backpropagation algorithm for training multilayer perceptrons (Rumelhart et al. 1986)[2] led to remarkable resurgence of interest in neural networks, and has widely been heralded as having overcome the limitations of perceptrons uncovered by Minsky and Papert. Its ability to train multilayer networks demonstrates clearly that the backpropagation algorithm is more powerful than the perceptron learning rule that is applicable to only simple perceptrons with one layer of trainable weights (Figure 1 shows a simple perceptron since only the weights of the response unit are subject to training). However, although Minsky-Papert analysis is based on simple perceptrons their conclusions reveal a general problem in the perceptron framework for processing topological patterns, regardless of the learning rules employed. The general problem, according to their 1988 expanded edition (Minsky & Papert 1988), "had nothing to do with learning at all; it had to do with the relationships between the perceptron's architecture and the characters of the problems presented to it" (p. xii). They further stated that it is a representation problem, and "no machine can learn to recognize X unless it possesses, at least potentially, some scheme for representing X" (p. xiii).

Can a multilayer perceptron with backpropagation training recognize topological patterns? A direct attempt to address figure-ground separation was made by Sejnowski and Hinton (1987) based on a related learning algorithm - Boltzmann machine. They demonstrated its effectiveness on toy problems, and "its usefulness for large problems is still uncertain" (p. 703). A later study using an extended version of the backpropagation algorithm to perform a segmentation task was made by Mozer et al. (1992), but again, success on only toy problems was reported. Of course, analytical statements cannot be derived from these individual attempts, but as far as the connectedness predicate is concerned Minsky and Papert (1988) explicitly claimed that multilayer networks are no more powerful (p. 52). We find no report contrary to this claim since their 1988 edition. Further insight can be obtained by observing the architecture and training of a multilayer perceptron. The notion of an order may appear irrelevant for the multilayer perceptron given that

---

[1] Their book is often credited for the downfall of neural networks in the late 1960s.

[2] This article is the standard reference for the backpropagation algorithm though earlier versions had been discovered (see any textbook on neural networks).

hidden units typically receive input from all the input units. At issue, however, are how many hidden units are required and how long is the training process. Because of the infinite order of the predicate and the exponential increase of the number of connected patterns, for R not too small, the number of hidden units required to represent a solution would be prohibitively large. Also, a practical training process can use only a tiny fraction of all possible training samples, and it is hard to expect example-based learning procedures, such as backpropagation, to generalize to exponentially more unseen samples. Hence, the Minsky and Papert claim is a reasonable projection, and it is premature to regard that the limitations of simple perceptrons no longer exist for their multilayer offsprings.

To solve the figure-ground separation problem in the general form, the solution must be valid regardless of the shape, position, size, orientation, etc., of each object on the retina; that is, the solution must be invariant to topological transformations. In a sense, it presupposes a solution to the connectedness problem. As we shall see later, the ability to segment connected components of an image plus some counting mechanism yields a solution to detect the connectedness predicate. So, the limitations of perceptrons in dealing with topological patterns can be treated as a special case of Rosenblatt's challenge. In other words, besides the mathematical rigor of Minsky-Papert analysis, the major findings of the book should have been hardly surprising to Rosenblatt.

Modern neurocomputing research has made important advances in understanding generalization characteristics of multilayer networks and their statistical underpinnings, which in turn have led to more effective training algorithms and better generalization capability for certain classes of problems (Bishop 1995; Arbib 2002). However, mainstream neural networks are preoccupied with analyzing statistical properties of individual patterns, and have all but shunned away from Rosenblatt's challenge. But the challenge remains a considerable cloud. For example, Hinton and Sejnowski (1999) recently acknowledged that "a major challenge for unsupervised learning is to get a system of this general type to learn appropriate representations for images" (p. xiv).

The purpose of this article is to argue that introducing the dimension of time is essential for a systematic attack on Rosenblatt's challenge. I will first argue in Section 2 that Rosenblatt's challenge is intrinsically related to the binding problem, and a key representation to resolving the binding problem lies in the temporal correlation theory. In Section 3, I introduce a special form of the temporal correlation theory: oscillatory correlation, which is primarily motivated by computational/mathematical and biological considerations. In the framework of oscillatory correlation, I describe LEGION[3] dynamics which converts the temporal representation into a computational mechanism. As a rather straightforward application, LEGION yields a solution to the connectedness problem. In Section 4, I review a number of recent studies that apply LEGION networks to visual and auditory scene analysis tasks, and these results demonstrate the computational utility of the oscillatory correlation theory. Section 5 discusses the biological relevance and implications of the theory; both neurobiological and psychophysical studies will be discussed. Section 6 is devoted to a discussion of a number of issues, including some that are likely contentious, such as the role of attention in binding and internal time versus external time (or oscillator time versus physical time). Concluding remarks are given in Section 7, whose theme is that versatile neural computation requires the time dimension.
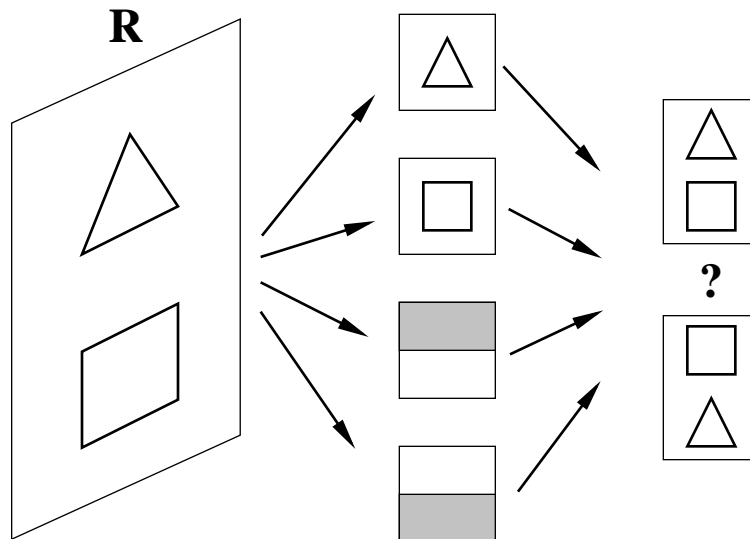
2. Binding problem

A fundamental attribute of perception is the ability to group elements of a perceived scene into coherent objects, which has been extensively studied in perceptual psychology under the title of perceptual organization or perceptual grouping (Palmer 1999). This ability is remarkable considering the fact that the input to perceptual organization is a retinal image and the organization takes place rapidly and effortlessly. Despite centuries of research from multiple disciplines, how perceptual organization is accomplished in the brain remains a mystery. Observing that visual

---

[3] LEGION stands for Locally Excitatory Globally Inhibitory Oscillator Networks (Wang and Terman 1995).

features, such as color, orientation, motion, and depth, are first extracted by feature-detecting neurons in different areas of the visual system, a related question is how these initial responses are bound together in the brain to form perceived objects? This is called the binding problem. At the heart of the problem is the fact that the sensory input contains multiple objects, which makes it necessary to resolve the issue of which features should bind with which others.   A related formulation is often illustrated using two objects of different shapes, say a triangle and a square, and different locations, say top and bottom, arranged in such a way that the triangle is at the top and the square is at the bottom (a layout discussed by Rosenblatt 1962;[4] see von der Malsburg 1999). This layout is shown in Figure 2. Given feature detectors for triangle, square, top, and bottom, how can the perceptual system bind locations and shapes in order to correctly perceive that the triangle is at the top (binding "top" and "triangle") and the square is at the bottom (binding "bottom" and "square"), rather than the other way around: the square is at the top and the triangle is at the bottom? What makes this formulation attractive is that people can make wrong feature conjunctions, producing "illusory conjunctions", when stimuli are presented very briefly (Treisman & Schmidt 1982; Treisman 1999). A potential confusion for this popular formulation (see Roskies 1999) is that object-level attributes such as shape and size are undefined before objects are perceived; that is, they are not defined before the more fundamental problem of figure-ground separation is solved.[5] For this reason, I will refer to the binding of local features to form perceived objects when talking about the binding problem.

Figure 2. Binding problem in a perceptron with four detectors for triangle, square, top, and bottom. The response unit needs to tell whether the triangle is on top (and the square at bottom) or the square is on top (and the triangle at bottom).

How does the brain resolve the binding problem? Concerned with the difficulty for visual shape recognition that is created by multiple, simultaneously present objects, Milner (1974) suggested that different objects be separated in time, leading to synchronization of firing activity within the cells activated by the same object. Later, in one of most cited technical reports, von der Malsburg (1981) also suggested that the time structure of neural signals provide the neural basis for his correlation theory, which directly and systematically addresses the binding (integration) problem. In a subsequent study, von der Malsburg and Schneider (1986) gave a concrete demonstration of the temporal correlation theory for the task of segregating two auditory inputs based on their distinct onset times - an example of the cocktail-party problem (I will come back to

---

[4] Note that Rosenblatt himself did not formulate this as a binding problem.

[5] Treisman in formulating her Feature Integration Theory (Treisman 1986) also failed to make this distinction.

this problem in Section 4.) This is an important paper because it, for the first time, addresses a figure-ground separation task using neural oscillators as building blocks, and synchrony and desynchrony among neural oscillators to represent a solution to a binding problem.

We should note that the temporal correlation theory is a theory of representation, concerned with how neurons triggered by different objects are represented in a neural network, not a theory of computation. It does not by itself address *how* multiple patterns lead to multiple cell assemblies with different time structures. This is the key issue to be taken up in Section 3.

The main alternative to the temporal correlation theory is hierarchical coding. The central idea is to rely on individual neurons arranged in some cortical hierarchy to integrate information so that cells higher in the hierarchy respond to larger and more specialized parts of an object. Eventually, this leads to the scenario that individual objects are represented by individual neurons, and for this reason hierarchical coding is also known as the grandmother or cardinal cell representation[6] - derived from Cajal's neuron doctrine (Barlow 1972). In the following I will limit my discussion to computational considerations; see Gray (1999) for biological evidence for and against the hierarchical representation. There are several computational problems with this representation. First, to be able to bind features that belong to the same object, object representation must already exist in the brain and image analysis is thus limited to recognition of familiar objects in the image. In addition to the issue of how individual representations are created in the first place, hierarchical coding would not allow perceiving novel objects, an ability the perceptual system clearly possesses (Treisman 1999). Second, perceiving an object, with all its vivid details such as location, shape, color distribution, orientation, size, and many other dimensions, is different from simply identifying that the object is, say, a dog (Kahneman et al. 1992; Treisman 1999). This creates the following dilemma. If the representation explicitly encodes and prestores all such details in order to deal with all possible scenarios, a vast majority of which never occurs to an observer in a lifetime, it would require prohibitively many cells. If the representation stays above these details, the binding problem recurs when facing an image with multiple objects: how to make sure that only a relevant subset of image elements are fed to a recognizer? One possible way out of this dilemma is perform top-down search starting from stored templates, one by one. This could be computationally feasible if the number of patterns and their possible variations are limited (see Sect. 6.5 for further discussion), but it would considerably limit the scope of image analysis.

The binding problem in the visual domain is a subject of intense debate that has captured the interest of researchers from many disciplines. Recently journal *Neuron* published a special issue (vol. 24, No. 1, 1999) discussing the binding problem and it featured articles by leading researchers both for and against the two binding hypotheses discussed above.

The above discussions should make it clear that, from the standpoint of the temporal correlation theory, the figure-ground separation problem is basically the same as the binding problem and the theory is primarily motivated by the need to address the problem. On the other hand, we note that the hierarchical coding mechanism is not much different from the perceptron framework as discussed by Rosenblatt (1962), where a variety of architectures, including multilayer and recurrent ones, are studied. The challenge facing Rosenblatt, discussed in the previous section, should underscore limitations of hierarchical coding.

3. Oscillatory correlation theory and LEGION dynamics

To make progress one must study concrete mechanisms. We focus on a special form of temporal correlation, which we call *oscillatory correlation* (Terman & Wang 1995), whereby feature detectors are represented by oscillators and binding is represented by synchrony within an assembly of oscillators and desynchrony between different assemblies. This representation is illustrated in Figure 3. The oscillatory correlation proposal is motivated by three considerations.

---

[6] The term "grandmother cell" intuitively captures the claim implied by hierarchical coding that cells must exist that code one's grandmother.

First, the activity of an oscillator well describes that of a neuron or a local group of neurons. Second, oscillatory correlation is consistent with coherent oscillations in the brain. Third, the use of oscillators facilitates a good deal the computation of synchrony and desynchrony. Two remarks about the oscillatory correlation theory are in order. First, temporal correlation need not be oscillatory correlation and there can be other ways to encode different temporal structures. In practice, however, few other alternatives have been seriously investigated. Second, oscillators need not always produce periodic activity, but they do so only when they reach steady behavior.[7] Thus, if the external input to an oscillator changes with time, the oscillator activity may not exhibit periodic behavior although it is still mathematically an oscillator.

As we noted earlier, the oscillatory correlation theory is a representation, not a mechanism. How to compute required synchrony and desynchrony when facing an input scene is an issue largely separate from the representation, and it is obviously a critical one in order to solve the binding problem. This limited scope of the correlation theory for addressing the binding problem has been repeatedly criticized by its opponents (Shadlen & Movshon 1999; Ghose & Maunsell 1999). What they did not realize, however, is that the theory has already motivated major progress in addressing this very issue.

The discovery of coherent oscillations in the cat visual cortex in the late 1980s (Eckhorn et al. 1988; Gray et al. 1989) immediately triggered a lot of computational work aimed at either modeling the biological phenomenon or attacking the binding problem in neural computation. Most of the early models use long-range connections to achieve synchronization in an assembly of oscillators (for a comprehensive list of references see Terman & Wang 1995). However, long-range connections lead to the problem of indiscriminant segmentation: oscillators would synchronize no matter whether they are activated by the same object or different ones (Sporns et al. 1991; Wang 1993b). This is illustrated in Figure 4, where three objects comprise an input scene. It is immediately clear from the figure that the three objects are separated on the basis of connectedness, but globally connected networks cannot encode topology and thus fail to separate the different patterns in Figure 4. To accomplish this elementary task, one needs to use locally connected networks.

In general, to provide a computational mechanism for the oscillatory correlation theory, three key functions must be achieved: (1) The mechanism must be capable of synchronizing a locally coupled assembly of oscillators; (2) it must be capable of desynchronizing different assemblies of oscillators that are activated by different objects; (3) both synchrony and desynchrony must occur rapidly. These requirements became major stumbling blocks for many computational attempts subsequent to the experimental discovery of synchronous oscillations. A major reason is a analytical result by Mermin and Wagner (1966) from theoretical physics, which states that harmonic oscillators cannot synchronize with local connections (see Terman & Wang 1995, for more discussion). This result was not known to many in neural networks, and harmonic oscillators, due to their simplicity plus a prevailing belief that all oscillations could be somewhere reduced to harmonic oscillators, were widely used for achieving oscillatory correlation. This also explains why such models require all-to-all connectivity to achieve synchrony. Fortunately, different kinds of oscillators do yield qualitatively different behaviors in a network and the generality of harmonical oscillators is bounded (Somers & Kopell 1993; Wang 1993a). Further investigation led to the use of relaxation oscillators, which, unlike harmonic oscillators, exhibit two times scales.

Building on the prior work on coupled relaxation oscillators by Somers and Kopell (1993), Terman and Wang (1995; Wang & Terman 1995) proposed the LEGION architecture that achieves all of the three requirements.

---

[7] This description of oscillations excludes so-called strange oscillators, which are really chaotic units.
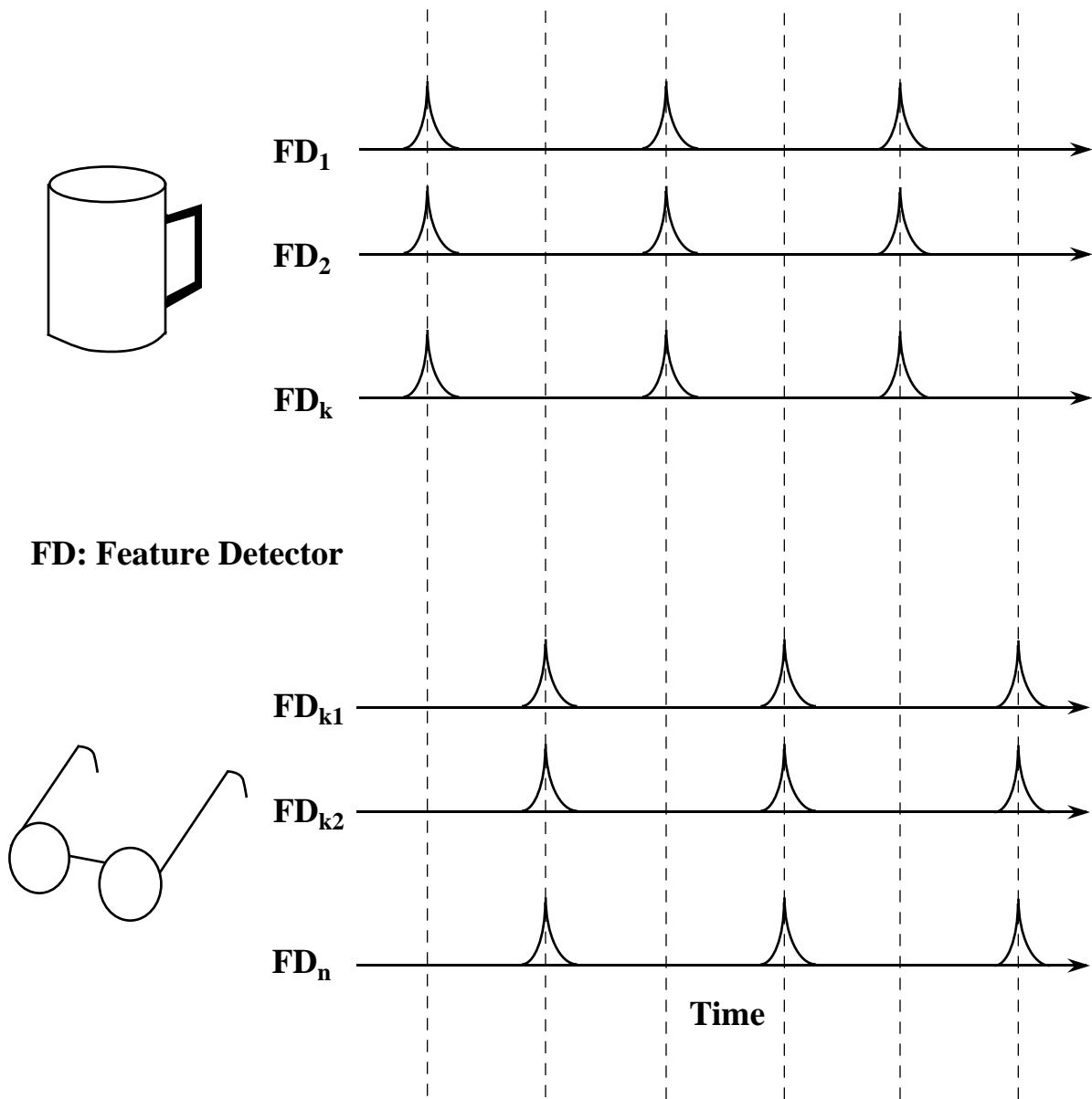
Figure 3. Oscillatory correlation representation. Each object (a cup or a pair of glasses) is represented by an assembly of feature detectors that are oscillators. Oscillators within an assembly synchronize their activity and different assemblies desynchronize.
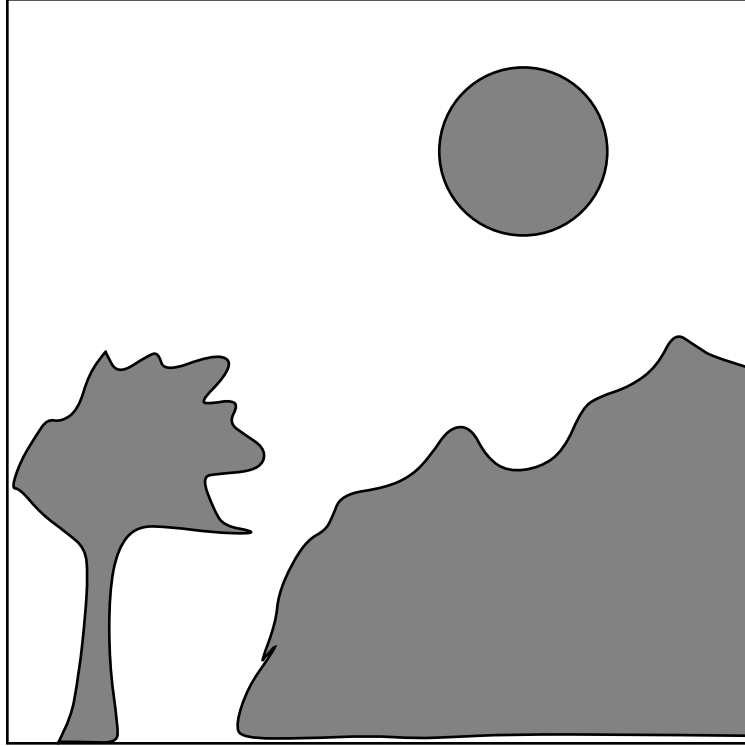
Figure 4. A visual scene with three caricature objects.

3.1 LEGION architecture

A LEGION network consists of three parts: (1) Its basic unit is a relaxation oscillator with two time scales; (2) oscillators are coupled with local excitation, which leads to rapid synchronization within a group corresponding to one pattern; (3) a global inhibitor desynchronizes different groups of oscillators.

Formally, a single oscillator i in LEGION is defined as a reciprocally connected pair of excitatory variable x_i and inhibitory variable y_i:

$$\dot{x}_i = 3x_i - x_i^3 + 2 - y_i + I_i + S_i + \rho \tag{1a}$$
$$\dot{y}_i = \varepsilon \left(\alpha \left(1 + \tanh(x_i / \beta) \right) - y_i \right) \tag{1b}$$

Here, I_i denotes external stimulation to the oscillator, S_i the overall coupling from the rest of the network, and ρ a noise term. The parameter ε is a small positive number. When coupling and noise are ignored and I is set to a constant, (1) defines a typical relaxation oscillator with two time scale induced by ε. The x-nullcline (i.e. $\dot{x}_i = 0$) is a cubic function and the y-nullcline is a sigmoid function.

If $I > 0$, the two nullclines intersect only at a point along the middle branch of the cubic, and in this case the oscillator produces a stable limit cycle, illustrated in Figure 5A. The oscillator is called *enabled*. The limit cycle alternates between a phase of relatively high x values and a phase of relatively low x values, called the *active* and *silent* phase respectively. Within each of the two phases the oscillator exhibits near steady-state behavior, and its trajectory in the silent phase corresponds to the left branch (LB) of the cubic and its trajectory in the active phase corresponds to the right branch (RB). In contrast to the behavior within each phase, the transition between the

8
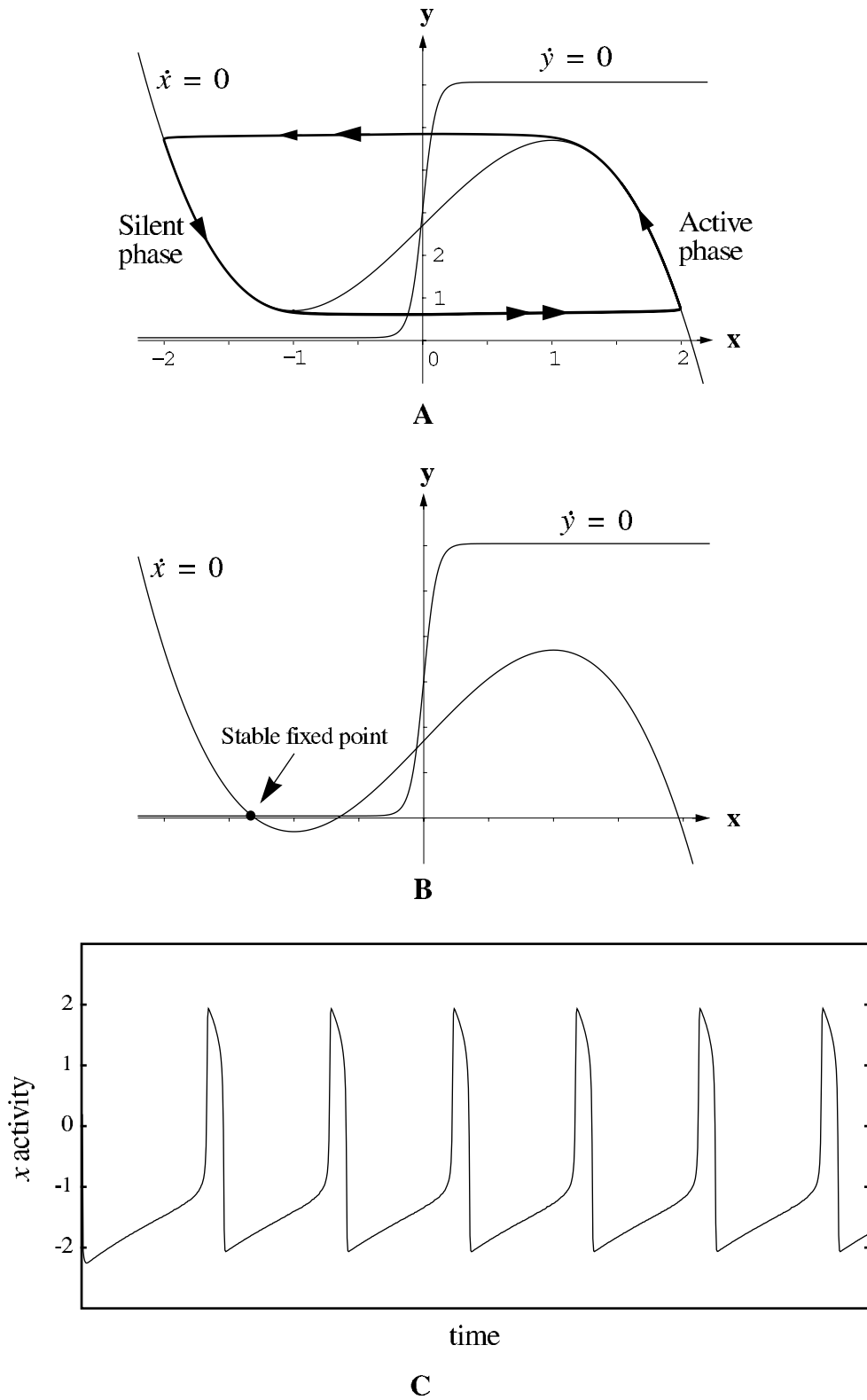
Figure 5. Single relaxation oscillator. **A**. The oscillator is enabled. In this case it produces a limit cycle, shown as the bold curve. The arrows indicate the direction of motion, and double arrows indicate jumping. **B**. The oscillator is excitable. In this case, it approaches the stable fixed point. **C**. The x activity of the oscillator with respect to time.

two phases takes place rapidly, and it is referred to as jumping. Such alternations between rapid change and slow change are characteristic of relaxation oscillations (van der Pol 1926). The parameter α determines the relative durations during which the limit cycle spends in the two phases - a larger α produces a relatively shorter active phase. If $I < 0$, the two nullclines of (1) intersect at a stable fixed point on LB of the cubic (see Figure 5B). In this case no oscillation occurs, and the oscillator is called *excitable*. Obviously, whether an oscillator is enabled or excitable depends on external stimulation. Hence, oscillations in (1) are stimulus-dependent.

For the oscillator defined in (1) may be interpreted either as a model of action potential generation, where x represents the membrane potential of a neuron and y represents the level of activation of ion channels, or oscillating bursts of neuronal spikes where x represents the envelope of the bursts. Figure 5C shows a typical trace of x activity, akin to a spike train. In fact, Equation (1) is dynamically very similar to standard neuronal models, including the FitzHugh-Nagumo equations (FitzHugh 1961; Nagumo et al. 1962) and Morris-Lecar model (Morris & Lecar 1981). All of these models can be viewed as simplifications of the classic Hodgkin-Huxley equations (Hodgkin & Huxley 1952).

For the simplest form of a 2-D LEGION network, an oscillator is excitatorily coupled with its four nearest-neighbors, and Fig. 6 shows the network architecture. The coupling term S_i in (1) is then given by

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

where H stands for the Heaviside step function, W_{ik} is the connection weight from oscillator k to i, and $N(i)$ is the set of four immediate neighbors of i. Both θ_x and θ_z are thresholds, and θ_x is chosen between LB and RB in the x dimension. Following Wang (1995), dynamic normalization is typically used to ensure that each oscillator has equal overall weights of dynamic connections, W_T, from its neighborhood.
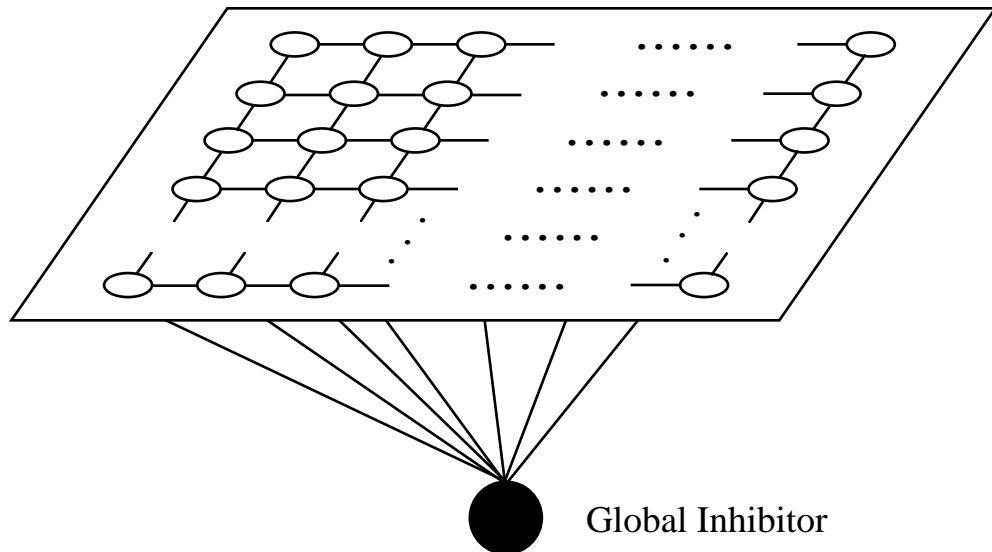
Figure 6. LEGION architecture. An oscillator is indicated by an open circle on the 2-D network which has four nearest-neighbor coupling. The global inhibitor receives input from and inhibits all the oscillators.

Finally, W_z in (2) is the weight of inhibition from the global inhibitor z, defined as

$$\dot{z} = \phi (\sigma_\infty - z) \tag{3}$$

Here, ϕ is a parameter, and $\sigma_\infty = 1$ if $x_i \geq \theta_z$ for at least one oscillator i and $\sigma_\infty = 0$ otherwise. If σ_∞ equals 1, $z \rightarrow 1$.

The system (1)-(3) has been extensively analyzed by Terman and Wang (1995). Let a pattern be a connected region. With ε sufficiently small, a LEGION network exhibits the mechanism of *selective gating*, whereby an enabled oscillator jumping up to the active phase rapidly recruits the oscillators stimulated by the same pattern, while preventing others from jumping up. They proved that, due to selective gating, the network rapidly achieves both synchronization within each oscillator assembly and desynchronization between different assemblies. Desynchronization between two assemblies means that they are never active (on RB) simultaneously. This dynamics will be illustrated in Section 3.3 when we discuss a solution to the connectedness problem. 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. See Wang (1999b) for a tutorial exposition of the selective gating mechanism and other related properties of the relaxation oscillators and their networks. In short, LEGION dynamics has met the three requirements for a computational mechanism of the oscillatory correlation theory.

3.2 Oscillation period and segmentation capacity

After a network of relaxation oscillators reaches stable limit cycles, it has the property that the oscillation period τ depends only on the parameters of a single oscillator. In the singular limit $\varepsilon \rightarrow 0$, τ has been calculated by Linsay and Wang (1998),

$$\tau = \tau_{LB} + \tau_{RB} = \ln(\frac{I_T + 4}{I}) + \ln(\frac{I - 2\alpha}{I_T - 2\alpha + 4}) \tag{4}$$

where τ_{LB} denotes the time spent on LB, and τ_{RB} on RB. I_T denotes the total input to an oscillator, and it equals $I + W_T - W_z$, where I is the level of external input to an enabled oscillator.

For a fixed set of parameters, both τ_{LB} and τ_{RB} are fixed; in particular, they do not vary as the number of objects on an input image increases. Given that each assembly stays in the active phase for the period of τ_{RB}, this property naturally leads to the fact that LEGION can segment only a limited number of patterns. This number is called the *segmentation capacity* (Wang & Terman 1997), and it corresponds to the ratio of τ to τ_{RB}. For typical parameter values, the capacity is about 5 to 7. Due to nonlinearity and intrinsic noise it becomes increasingly difficult to find parameter values that can robustly support a much larger capacity.

What happens if the number of patterns in an input image exceeds the segmentation capacity? The system then separates the entire image into as many segments as the capacity. In this case, each segment may either correspond to a single pattern or multiple ones.

3.3 A solution to the connectedness problem

As a concrete application of the LEGION dynamics described above, we now describe a solution to the connectedness problem (Wang 2000). Before explaining how to compute the predicate, we show the response of a two-dimensional LEGION network to two binary images: one connected and one disconnected. The size of the network is 30x30. The connected image is a
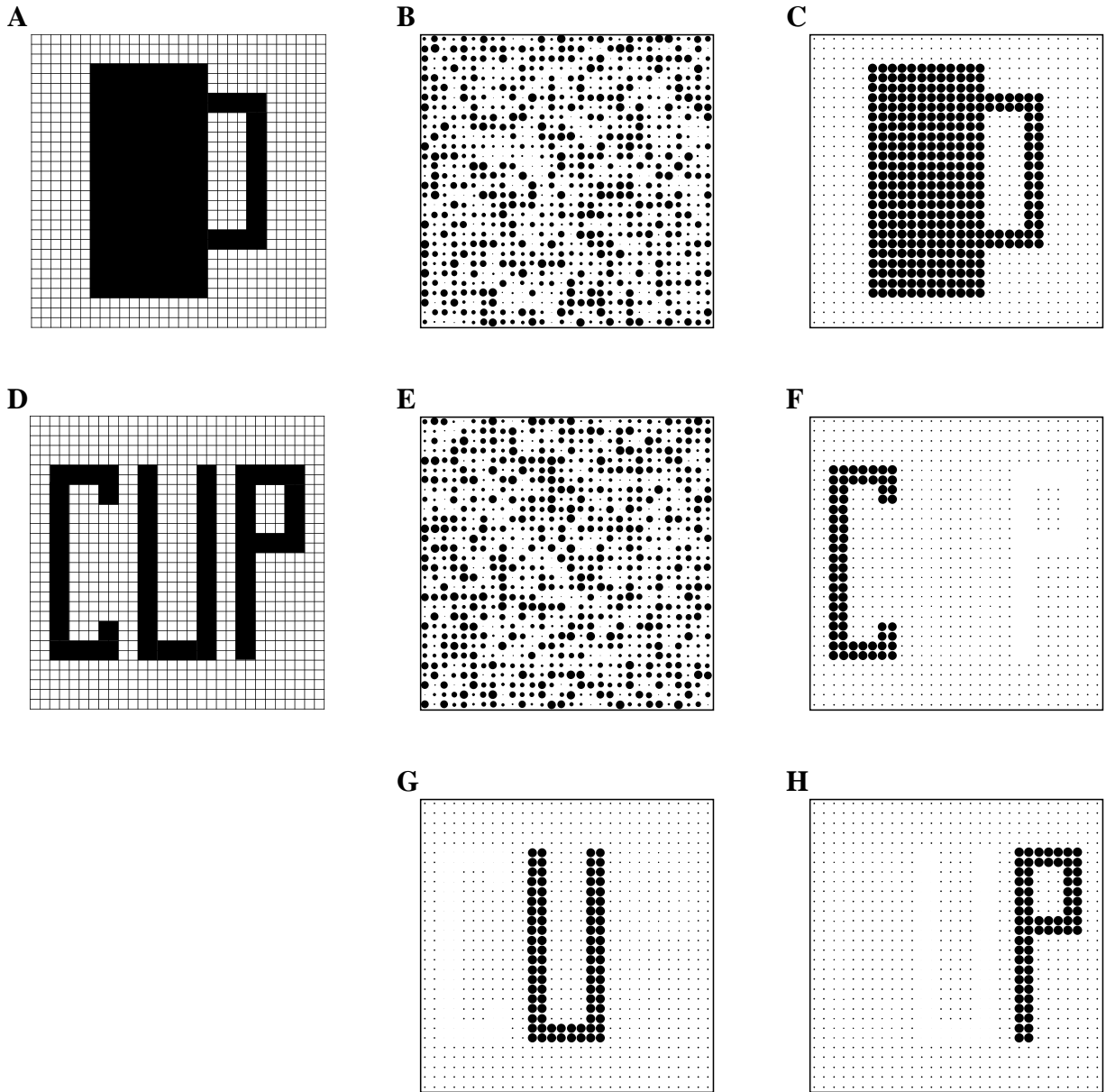
Figure 7. A. A connected cup image is presented to a 30x30 LEGION network. **B**. A snapshot at the beginning of system evolution. **C**. A subsequent snapshot taken shortly afterwards. **D**. A disconnected image with three patterns forming the word "CUP," is presented to the same network. **E**. A snapshot at the beginning of system evolution. **F.-H**. Subsequent snapshots taken shortly after the system starts. The parameter values are: ε = 0.02, α = 6.0, β = 0.1, ρ = 0.02, θ_x = -0.5, θ_z = 0.1, ϕ = 3.0, W_z = 1.0, and W_T = 8.0 (weights are identical before dynamic normalization). I = 0.2 for an enabled oscillator and I = -0.02 otherwise.

cup figure shown in Fig. 7A, while the disconnected one is the word "CUP" shown in Fig. 7D. The differential equations in (1)-(3) are solved using the fourth-order Runge-Kutta method. To indicate that the network has no binding preference at the beginning, we randomize the phase of each oscillator, as illustrated in Fig. 7B. In the figure, the diameter of a circle corresponds to the x activity of the respective oscillator. Fig 7C shows a snapshot of network activity shortly after the beginning. All the oscillators corresponding to the cup are synchronized, while the remaining ones are excitable. Figs. 7E-H shows the network response to "CUP", where Fig. 7E indicates the random initial conditions, and Figs. 7F-H show subsequent snapshots taken shortly afterwards. The effects of synchrony and desynchrony are clearly shown in the display, and the successive "popout" of the three segments continues until the input image is withdrawn.

To show the entire process of synchronization and desynchronization, Fig. 8A depicts the temporal activity of all the stimulated oscillators for the connected cup image, where unstimulated oscillators are omitted since they do not oscillate or are excitable. The oscillator activity corresponding to each connected pattern is combined in the display. Therefore it appears like a single oscillator when the assembly of oscillators representing the pattern are in synchrony. The upper panel shows the activity of the assembly representing the cup, and the middle one shows that of the global inhibitor. Synchrony occurs in the first oscillation period. The situation for the disconnected "CUP" is shown in Fig. 8B, where the upper three traces show the assembly activities corresponding to the three patterns. Fig. 8B shows that synchrony within each assembly and desynchrony between different ones are both achieved in the first two periods.

As illustrated in the above simulations, after a few oscillation cycles all the oscillators in one assembly, i.e. corresponding to one connected pattern, are synchronized whereas different assemblies are desynchronized. Furthermore, when an assembly jumps to the active phase the global inhibitor is triggered, and this happens as many times within an oscillation period as is the number of patterns in the input image. Thus, how many patterns are in the input image can be revealed by comparing the oscillation frequency of any enabled oscillator and the frequency of the global inhibitor. If they are the same this indicates that the input image contains one pattern, and thus the figure is connected. Otherwise, the input image contains more than one pattern and therefore the figure is disconnected. The accumulated activity of the global inhibitor over an oscillation period τ is $\int_{T-\tau}^{T} z \, dt$, where T denotes the current time. The corresponding average accumulated activity of all the enabled oscillators is given $\sum_i \int_{T-\tau}^{T} H(x_i - \theta_x) dt \Big/ \sum_i H(I_i)$, where the denominator indicates the number of the enabled oscillators. The connectedness predicate is then given by (Wang 2000)

$$\int_{T-\tau}^{T} z \, dt \Bigg/ \left(\frac{\sum_i \int_{T-\tau}^{T} H(x_i - \theta_x) dt}{\sum_i H(I_i)} \right) < \theta \tag{5}$$

The LHS (left-hand-side) of (5) gives the number of the patterns in the input image. Thus, $2 > \theta > 1$. In reality, with $\varepsilon > 0$ and system noise, synchrony within each pattern is not perfect (see Fig. 8); the active phase of an assembly, which directly triggers the global inhibitor, is slightly longer than that of a single oscillator within the assembly. Thus, θ should be chosen somewhat greater than 1, but certainly less than 2.

The LHS value of (5) for the two cases in Figure 8 is given in the two bottom traces of Fig. 8A and Fig. 8B, respectively. According to (4), $\tau \approx 5.27$ for the parameter values used in the simulations. A threshold $\theta = 1.6$ is used in Fig. 8. The figure shows that, beyond a short
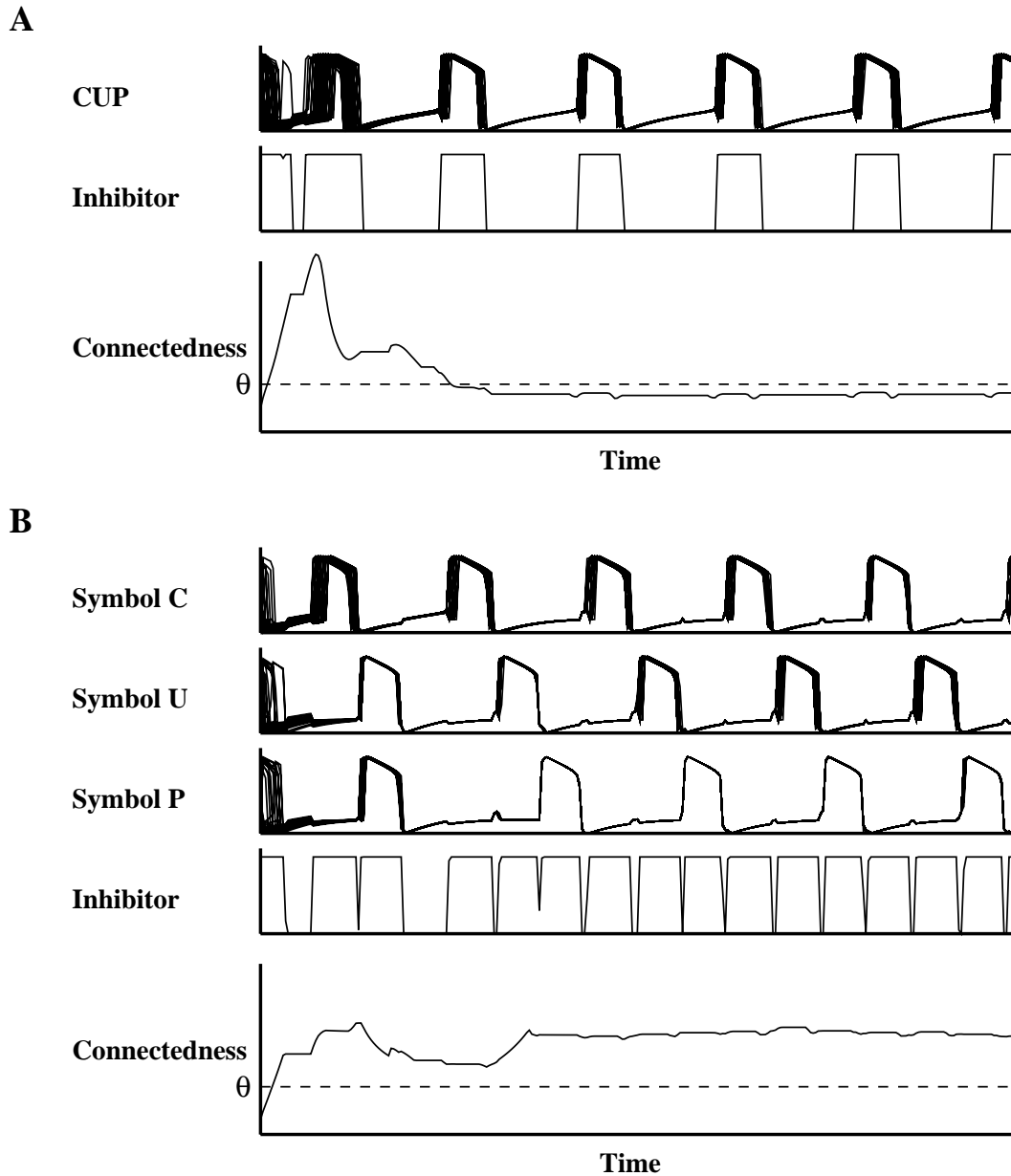
Figure 8. Temporal activity of every enabled oscillator (from Wang 2000). **A.** Result for the connected cup image. **B**. Result for the disconnected "CUP" image. In both **A** and **B**, the upper traces show the combined *x* activities of the oscillator assemblies indicated by their respective labels, the next-to-bottom trace the activity of the global inhibitor, and the bottom one the temporal activity of the RHS of (5) together with θ.

beginning duration that corresponds to the synchronization and desynchronization process, (5) correctly computes the connectedness predicate.

The situation where the number of patterns on an image is greater than the segmentation capacity presents no difficulty. For any LEGION network with a capacity greater than 1, the predicate in (5) is not affected when numerous patterns appear in the input, because it is an assertion of whether the figure contains *just* one pattern or not. As discussed in Sect. 3.2, the LEGION network separates the image into as many segments as the capacity when it is exceeded by the number of patterns. This analysis, together with the result on the speed of LEGION segmentation (see Sect. 3.1) implies that the system takes at most as many cycles as the capacity to correctly detect connectedness, no matter how numerous input patterns are on an image.

The above solution to the connectedness problem is given in the most general form, regardless of the shape, size, orientation, position, etc., of each pattern, or the arrangement of various patterns in a picture; it works for objects with (see Fig. 7A) or without holes. The solution is analytically established, and the two key aspects contributing to the solution are recurrent LEGION architecture and the oscillatory correlation representation.

4. Towards a solution to the scene analysis problem

The oscillatory correlation theory and the LEGION mechanism together provide a general framework for addressing the scene analysis problem, which remains one of the most challenging problems in machine perception (Duda et al. 2001, p. 10). To deal with real-world scenes that are considerably more complex than binary images, connection weights between oscillators need to encode some measure of similarity between the corresponding scene elements. What determines the similarity between local sensory elements? In the visual domain, this has been systematically studied in Gestalt psychology (or Gestalt grouping, see Wertheimer 1923; Koffka 1935). The following is a list of major grouping principles that include both classical and new ones (Palmer 1999):

• *Proximity*. Elements that lie close in space tend to group.

• *Similarity*. Elements that have similar attributes, such as luminance, color, depth, or texture, tend to group.

• *Common fate*. Elements that move coherently (common motion) tend to group. We note that this may be regarded as an instance of similarity, and it is listed separately to emphasize visual dynamics.

• *Good continuity*. A set of elements that form smooth continuations of each other tend to group.

• *Connectedness* and *common region*. Connected elements tend to group. Similarly, elements that lie inside the same connected region tend to group.

• *Familiarity*. A set of elements that belong to the same familiar pattern tend to group.

Many of the principles are directly related to the emphasis of LEGION on local connectivity. In the case of real images, to apply grouping principles generally requires a separate process that extracts local features, which may simply be pixel values for intensity images or statistical features that characterize a textural pattern. In this section, we show that, in conjunction with feature extraction, LEGION networks can effectively perform challenging scene analysis tasks.

Before describing individual tasks, let us summarize the basic approach to scene segmentation. After a scene is presented, feature extraction first takes place, and extracted features form the basis for determining connection weights between oscillators. The oscillator network then evolves on its own. After a few oscillation cycles required for the synchronization and desynchronization process, assemblies that alternately jump to the active phase represent resulting segments. Different segments emerge from the network at different times, and it is *segmentation in time* that distinguishes the LEGION approach from others. In a broader context, this way of addressing the scene segmentation problem represents a concrete investigation of the dynamical approach to cognition (van Gelder & Port 1995; van Gelder 1998).

4.1 Image segmentation

Two issues immediately arise when handling real images: noise on an image and computing time required for integrating a large oscillator network. Noise can lead to many fragments which, in the presence of a limited segmentation capacity, can deteriorate the result of LEGION segmentation. To address this problem of fragmentation, Wang and Terman (1997) introduced a lateral potential for each oscillator, which allows the network to distinguish between major blocks and noisy fragments. The basic idea is that a major block must contain at least one oscillator, denoted as a leader, which lies at the center of a large homogeneous region. Such an oscillator receives large lateral excitation from its neighborhood, and thus its lateral potential is high. A noisy fragment does not contain a leader. The collection of all fragments is called the background. To alleviate the computational burden of integrating a large oscillator network, they abstracted an algorithm that follows major steps in the numerical simulation of LEGION dynamics, such as two time scales, jumping, and spread of activation. The resulting algorithm also removes the segmentation capacity. In their system for segmenting real images, each oscillator is connected to its 8-nearest neighbors, and the connection weight between two neighboring oscillators i and j is set proportional to $1/(1 + |I_i + I_j|)$, where I_i and I_j indicate the corresponding pixel values. The key parameter for segmentation is the level of global inhibition: W_z (see (2)). Larger values of W_z produce more and smaller segments. Figure 9B shows a typical result for an aerial image in Fig. 9A. The entire image is segmented into 23 regions, each of which corresponds to a different gray level in the figure, which indicates the phases of oscillators. In the simulation, different segments rapidly pop out from the image in time, as similarly shown in Figure 8. As can be seen from Figure 9B, most of the major regions are correctly segmented. The black scattered regions in the figure represent the background that remains inactive. Due to the use of lateral potentials, all these tiny regions are put to the background.

A   **B**
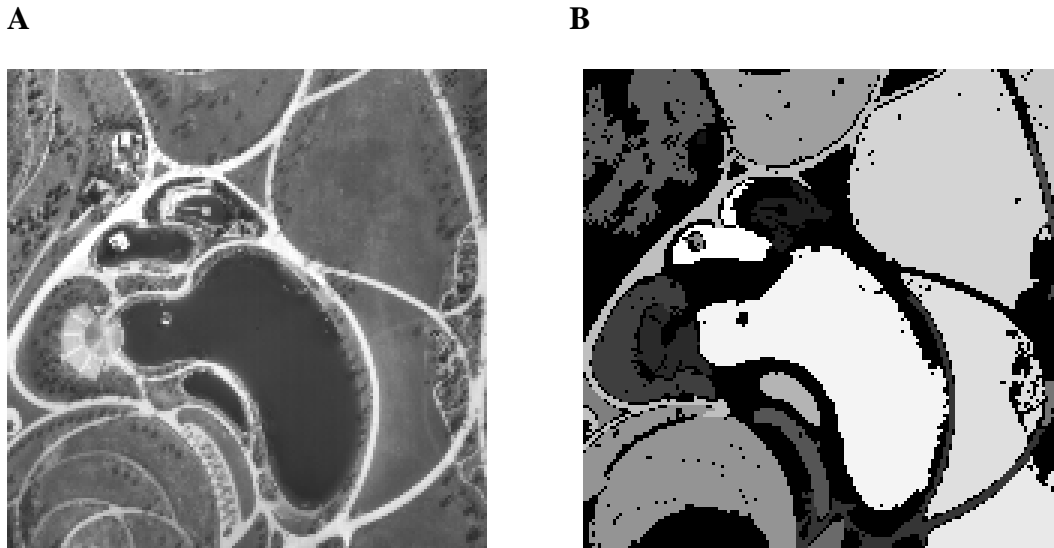
Figure 9. Intensity image segmentation (from Wang & Terman 1997). **A**. An intensity image with 160x160 pixels. **B**. Result of LEGION segmentation. Each segment corresponds to a distinct gray level, and the background corresponds to the black areas.

To further reduce sensitivity to noise on an image while preserving important features, Chen et al. (2000) proposed the idea of adapting dynamic weights to perform feature-preserving smoothing before LEGION segmentation. Their weight adaptation method is insensitive to termination times -

a common problem in various smoothing techniques in image processing. Moreover, they proposed to employ a logarithmic coupling term, which can be written as (cf. (2))

$$S_i = \frac{\sum\limits_{k \in N(i)} W_{ik} H(x_k - \theta_x)}{\ln(\sum\limits_{k \in N(i)} H(x_k - \theta_x) + 1)} - W_z H(z - \theta_z) \tag{6}$$

For a variety of large-scale aerial images from the United States Geological Survey (USGS), the resulting algorithm achieves very good segmentation results and performs better than other recent image processing algorithms, including nonlinear smoothing and multi-scale segmentation (Chen et al. 2000; Liu et al. 2001). Figure 10 gives two examples of extracting hydrographic objects from USGS satellite images. The original images containing water bodies are shown in the top row of Figure 10. The middle row shows the corresponding extraction results, where the water bodies are marked as white and superimposed on the original images. The bottom row provides the corresponding USGS 1:24,000 maps. A careful comparison between the extracted waterbodies and the maps indicate that the former portray the images even better, because stationary maps do not reflect well the changing nature of geography.

Common motion, or common fate, is another major grouping cue as we discussed earlier. Cesmeli and Wang (2000) applied LEGION to motion-based segmentation that considers motion as well as intensity for analyzing image sequences. In their system, two pathways perform an initial optic flow estimation and intensity-based segmentation in parallel. A subsequent network combines the two to refine local motion estimates. Motion analysis and intensity analysis complement each other since the former tends to be reliable in inhomogeneous, textured regions while the latter is most effective in homogeneous regions. The use of LEGION for segmentation allows for multiple motions at the same location, as in the case of motion transparency. The resulting system significantly reduces erroneous motion estimates and improves boundary localization. A typical example is given in Figure 11. A frame of a motion sequence is shown in Fig. 11A, where a motorcycle rider jumps to a dry canal with his motorcycle while the camera is tracking him. Due to the camera motion, the rider and his motorcycle have a downward motion with a small rightward component and the image background has an upright diagonal motion. Figure 11B shows the estimated optic flow after integrating motion and brightness analyses, and it is largely correct. The rider with his motorcycle is then segmented from the image background as depicted in Figure 11C. Their oscillator model has been favorably compared with a number of algorithms including the one by Black and Anandan (1996) based on robust statistics.

Other efforts include segmentation of range and texture images (Liu & Wang 1999; Cesmeli & Wang 2001), and contour extraction (Yen & Finkel 1998; Horn & Opher 1999). A recent study performs data clustering via synchrony and desynchrony in a network of integrate-and-fire oscillators (Rhouma & Frigui 2001).

4.2 Object selection

A classic topic in neural networks is neural competition. Winner-take-all (WTA) networks have been extensively studied (Didday 1970; Grossberg 1976; Amari & Arbib 1977; Rumelhart & Zipser 1986; Ermentrout 1992). WTA dynamics is based on global inhibition, either in the form of a global inhibitor or mutual inhibitory connections, and produces a winner that has the highest input. Such competitive dynamics has been applied to many tasks, and has played a major role in modeling selective visual attention (Koch & Ullman 1985; Niebur & Koch 1998). In WTA, individual neurons compete with each other, which corresponds to local representations. For perceptual processing, however, experimental data suggest that objects act as wholes in competition (Desimone & Duncan 1995; Nakayama et al. 1995; Driver & Baylis 1998; Wang et al. 2001). But, in order to capture object-level competition, one must address the binding issue.
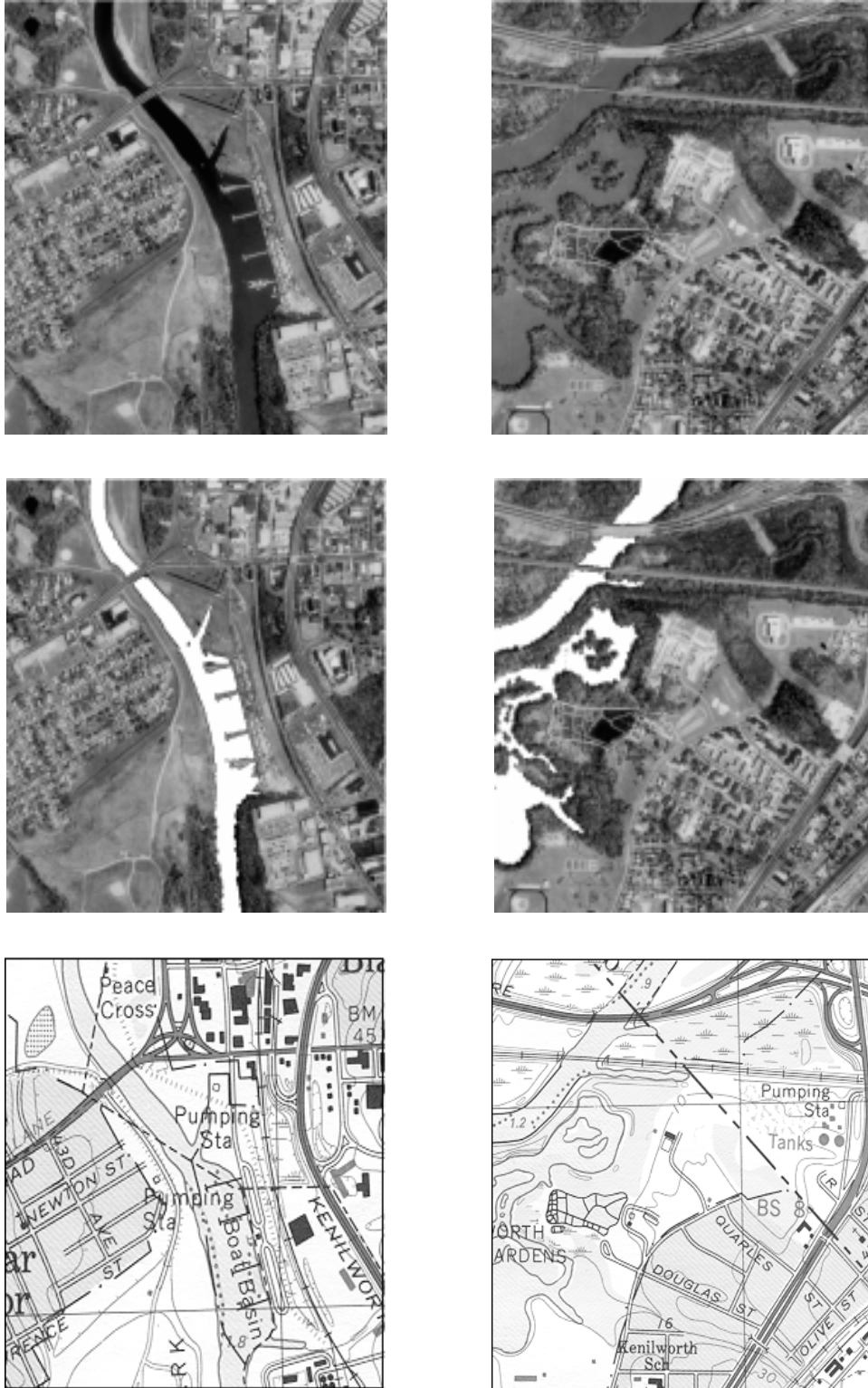
Figure 10. Extraction of hydrographic objects (from Chen et al. 2000). The top row shows two satellite images. The size of the left image is 670x606 and that of the right one is 640x606. The middle row shows the corresponding extraction results, where extracted objects are marked as white. The bottom row shows the corresponding topographic maps.

A   B   C
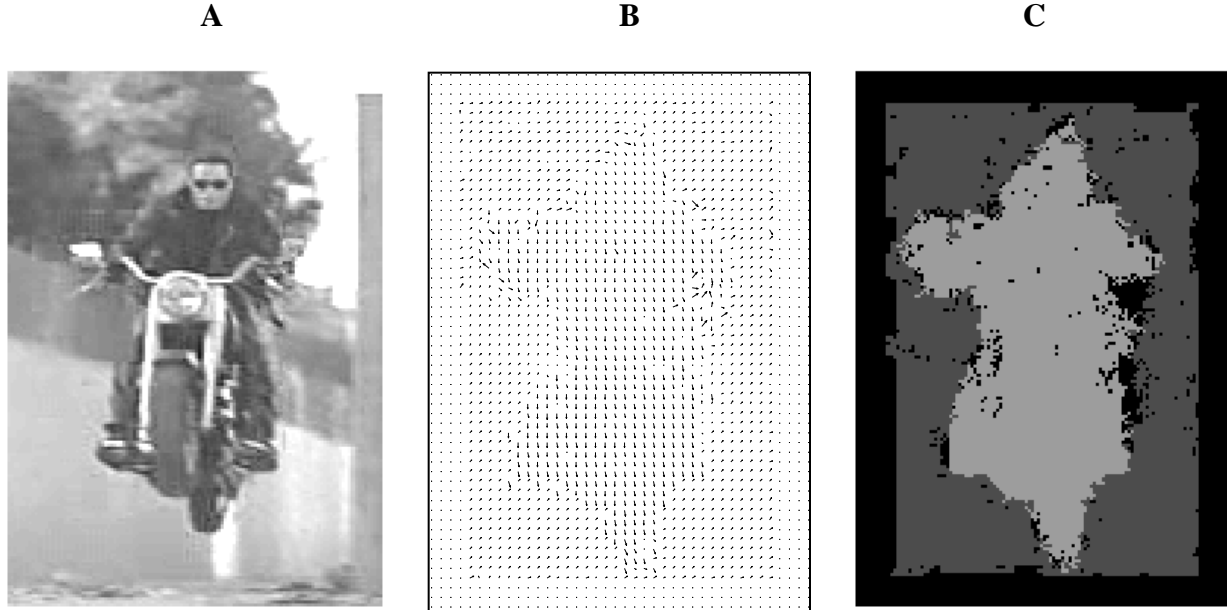
Figure 11. Motion segmentation (from Cesmeli & Wang 2000).   **A**. A frame of a motion sequence. **B**. Estimated optic flow. **C**. Result of segmentation.

Exploiting the LEGION mechanism for binding and slow inhibition, we have proposed an architecture for object selection (Wang 1999a). The selection network realizes a concrete form of object-level competition: size-based competition. In other words, in an input scene with many objects (patterns), the network attempts to select the largest one.   The basic idea is that after oscillating assemblies are formed, competition between assemblies takes place in time. When an assembly jumps to the active phase, it leaves an inhibitory trace via a slow inhibitor, which can be overcome only by larger assemblies. An analysis on the model shows that after a number of oscillation cycles, the largest assembly will be the only one that oscillates, while all the others are suppressed. The system can be adjusted to select several largest objects, which then alternate in time. Figure 12 shows the result of selecting the most salient object in an intensity image.   The original image (Fig. 12A) is first processed by a LEGION network, which yields a number of major segments (Fig. 12B). The selection network then extracts the largest segment – the cortex (Fig. 12C).

We stress that the saliency in our selection model is an object-level property, whereas it is a local, location-specific property when used in a saliency map (Koch & Ullman 1985; Niebur & Koch 1998). Our model is compatible with object-based theories of visual attention, while WTA models are compatible with location-based theories; see Pashler (1998) and Parasuraman (1998) for detailed description of these two contrasting theories of visual attention. We will come back to the issue of attention in Sect. 6.

4.3 Speech segregation

Similar to the visual domain, a listener in an auditory environment is exposed to acoustic energy from different sources. To understand the auditory environment, the listener must first disentangle the acoustic wave reaching the ears. This process is referred to as auditory scene analysis (Bregman 1990). According to Bregman (1990), auditory scene analysis takes place in two stages. In the first stage, the acoustic mixture reaching the ears is decomposed into a collection of sensory elements. Second, elements that are likely to have arisen from the same source are grouped to form a stream that is a perceptual representation of an auditory event. Important cues for auditory grouping include proximity in frequency and time, smooth temporal transition, onset and offset coincidence, and common location.

A

B

C
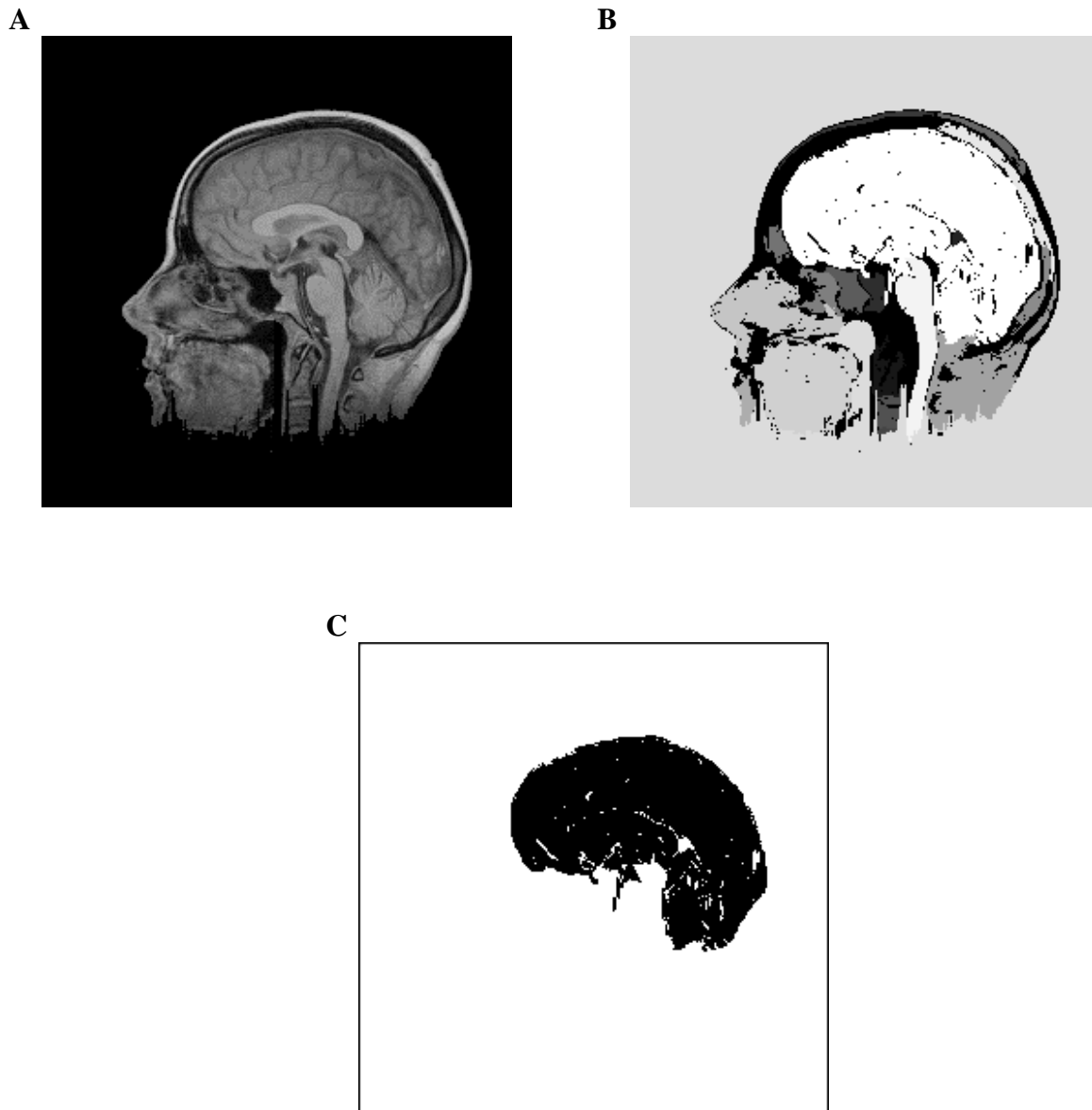
Figure 12. Object selection (from Wang 1999a). **A**. An MRI image with 257x257 pixels. **B**. Result of LEGION segmentation. **C**. Result of the selection network, which extracts the largest object.

Wang and Brown (1999) studied the speech segregation problem: separating target speech from its acoustic interference. Echoing Bregman's two-stage notion, their model consists of a stage of computing segments from an input scene, which is followed by a stage that groups segments into a target speech stream and a background. Segment formation is based on temporal continuity and cross-channel correlation between filter responses of adjacent auditory channels. Grouping is based the global pitch computed in each 20-ms time frame. Segment formation is performed by a LEGION network, and grouping is carried out by a laterally connected network of relaxation oscillators. A systematic evaluation shows that the system produces an improvement in signal-to-noise ratio (SNR) for every mixture. An example is given in Fig. 13, where the input scene is a mixture of male utterance and telephone ringing (see Fig. 13A). The segregated target is shown in Fig. 13B and the background in Figure 13C.

4.4 Integrated analysis

As discussed at the beginning of Sect. 4, the majority of the grouping principles do not involve either specific memory or recognition, and may be viewed as primitive or bottom-up. The familiarity principle, however, requires a memory recall and may be viewed as a top-down process. Memory-based segmentation has been previously studied (Wang et al. 1990; Horn & Usher 1991; Sompolinsky & Tsodyks 1994; Lourenco et al. 2000). But, without a primitive segmentation stage, the performance of these models is limited. A general solution to the scene analysis problem requires an approach that integrates bottom-up segmentation and top-down analysis.

Recently, we have investigated the integration of a primitive segmentation stage and associative memory on the basis of oscillatory correlation (Wang & Liu 2002). Our model consists of initial primitive segmentation, multi-module associative memory, and a short-term memory (STM) layer. Figure 14 shows the diagram of the model. Primitive segmentation is performed by LEGION, which separates an input scene into multiple segments. Each segment then activates the memory layer, and potentially multiple recalls interact in the STM layer, resulting in a common part. The pattern held in the STM layer projects to the LEGION network, and this top-down input performs memory-based grouping and segmentation. Memory-based grouping synchronizes multiple segments that belong to the same memory pattern. On the other hand, memory-based segmentation further separates a segment into multiple parts, correcting under-segmentation errors caused by primitive segmentation. It is worth emphasizing that the system achieves scene analysis entirely in phase space or time. This is consistent with physiological evidence that suggests an important role of synchronous oscillations in top-down processing (Engel et al. 2001).

The system has been evaluated on a set of 3-D line drawing objects, which are arranged in an arbitrary fashion to compose input scenes. Object occlusion arises due to 3-D arrangements between objects. One such scene is shown at the bottom of Figure 14. A systematic evaluation demonstrates that memory-based organization is responsible for a significant improvement in scene analysis performance (Wang & Liu 2002).

To perform scene segmentation effectively, both feature extraction and grouping are important. Feature extraction is modality- and cue-specific, whereas a binding mechanism may be generally applicable. Taken together, this body of work on LEGION-based segmentation gives strong indications that the LEGION networks provide a general and effective mechanism for scene segmentation. Although many other issues remain to be addressed in scene analysis, such as cue integration and multimodal integration, we believe that the LEGION mechanism has, for the main part, answered Rosenblatt's challenge regarding figure-ground separation. His challenge regarding spatial relations will be discussed in Section 6.
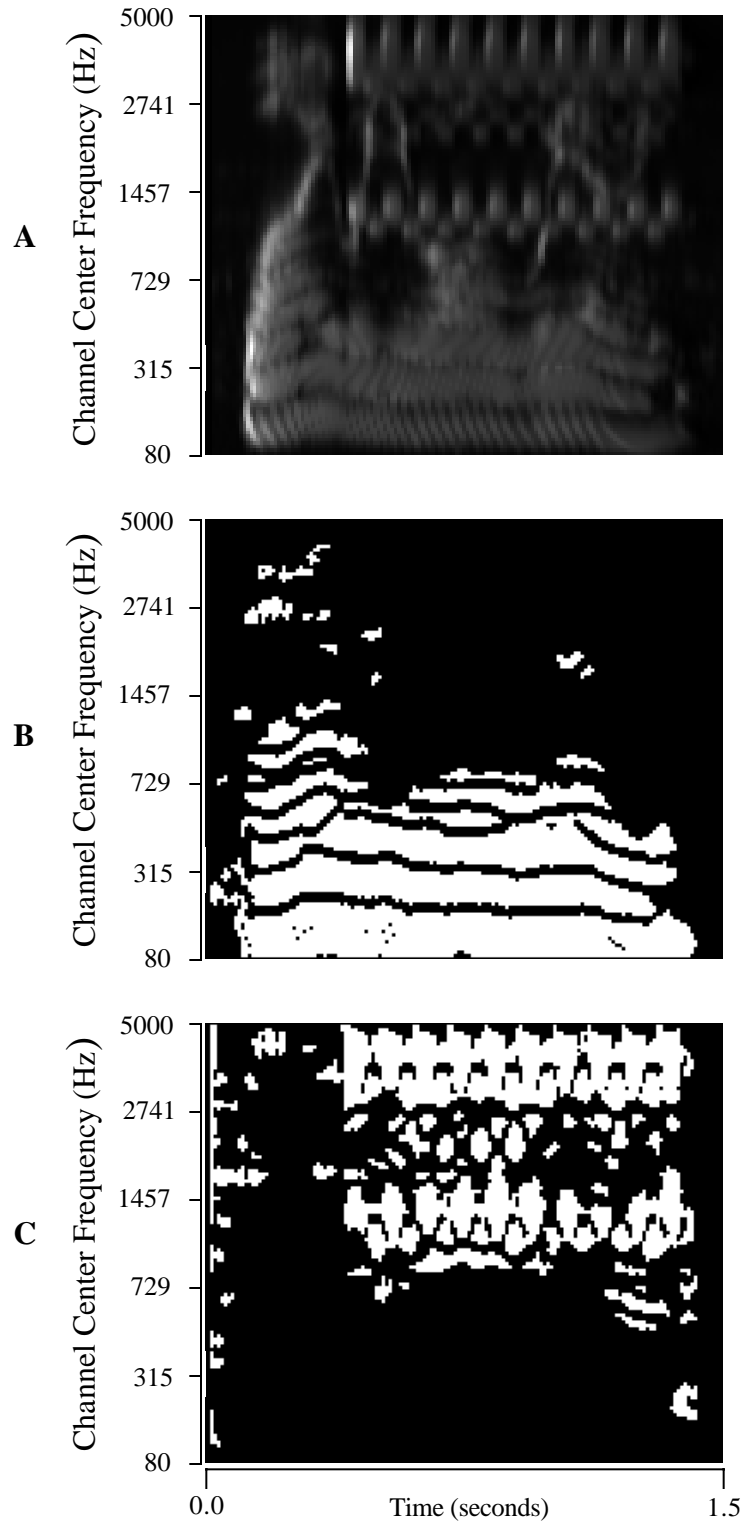
Figure 13. Speech segregation (from Wang & Brown 1999). **A**. Filter responses to a mixture of male utterance and telephone ringing. The response pattern is generated by 128 filter channels, whose center frequencies range from 80 Hz to 5 kHz, over 150 time frames. **B**. Segregated target speech, indicated by white pixels representing active oscillators at a time. **C**. Segregated background, indicated by white pixels representing active oscillators at a different time.
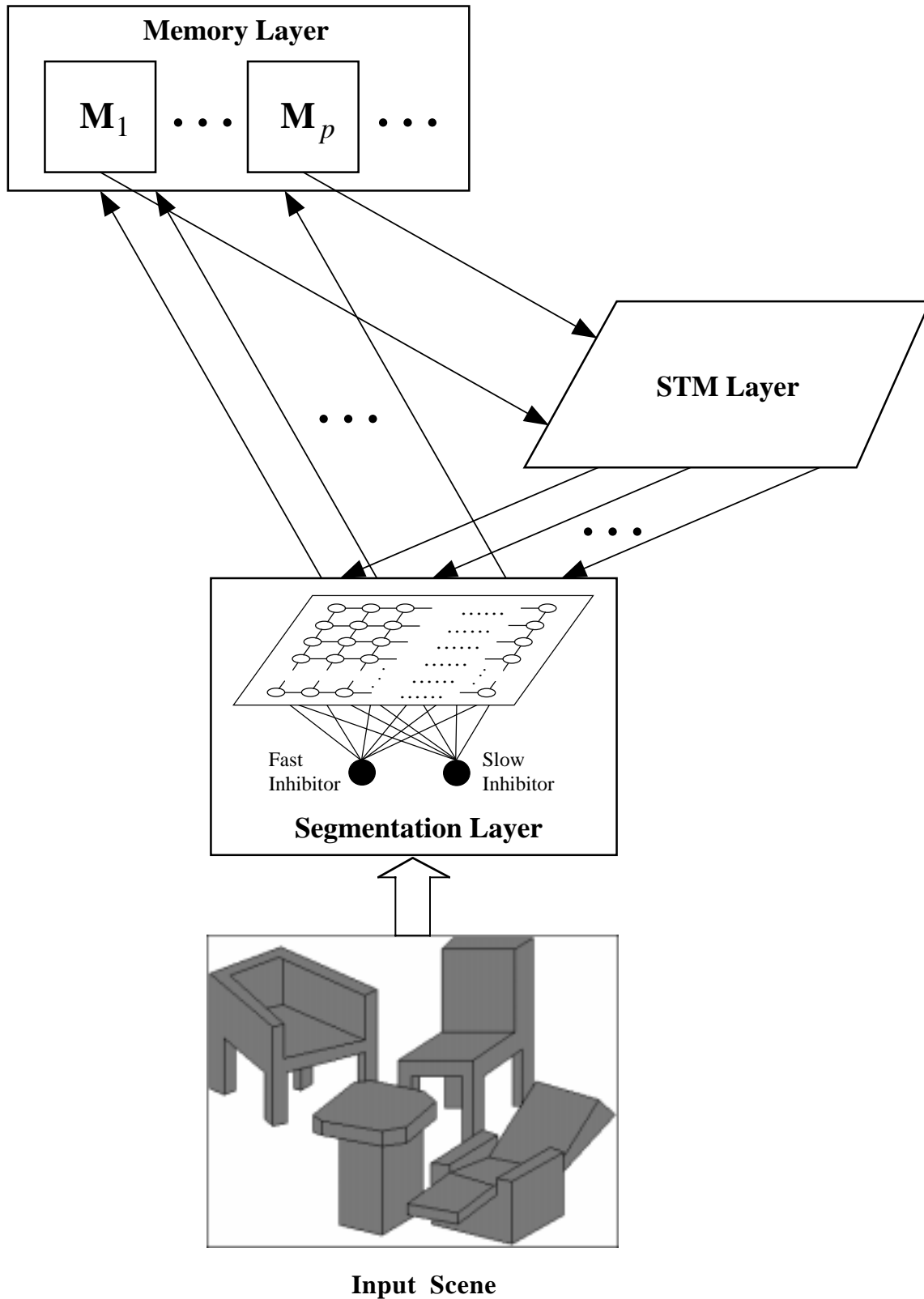
Figure 14. Integrated analysis network (from Wang & Liu 2002). The model consists of a LEGION segmentation layer, a multi-module memory layer, and an STM layer. Arrows indicate the directions of network connection.

5. Biological relevance and implications

5.1 Physiological considerations

There is an extensive and growing body of physiological evidence that supports the existence of coherent oscillations in various cortical regions as well as their potential role in feature binding. Experimental results come from different modalities and different animal species and humans, in both anesthetized and awake conditions. Early evidence of neural oscillations was obtained from sensory evoked potentials in the olfactory system (Freeman 1978) and the auditory system (Galambos et al. 1981). The accumulation of evidence has accelerated following the discovery of synchronous oscillations from cell recordings in the cat visual cortex (Eckhorn et al. 1988; Gray et al. 1989). Detailed reviews are given by Singer and Gray (1995), Usrey and Reid (1999), and Varela et al. (2001), and are not considered here. Two points are worth noting. First, theoretical investigation (Milner 1974; von der Malsburg 1981) on the binding problem predates and directly influences empirical work that uncovered visual coherent oscillations. Second, neural oscillations have been a controversial topic in neuroscience; see, for example, the *Neuron* issue mentioned in Section 2. It is fair to state that neural oscillations and synchrony are clearly present (Roskies 1999), and that the debate has largely been shifted from whether coherent oscillations exist to whether they play a major role in binding.

The following summarizes several important aspects of the experimental data on neural oscillations. Oscillation frequencies from different modalities and animal species generally range from 30 to 70 Hz, often referred to as 40 Hz oscillations. This frequency range is compatible with that of the EEG gamma rhythms, and hence such oscillations are also called gamma oscillations. Cortical oscillations depend on the presence of visual stimulation, but not on oscillating input. They are thus referred to as stimulus-dependent, not stimulus-driven. Synchrony in neural oscillations, i.e. phase locking with zero phase lag, occurs across a considerable extent of the cortex, beyond the distance of direct connections between cortical cells. Finally, the presence or absence of coherent oscillations correlate with perceptual organization for a broad range of perceptual stimuli.

In the auditory system, 40-Hz oscillations in localized brain regions have been recorded both at the cortical level and at the thalamic level, and these oscillations are synchronized over considerable cortical areas (Ribary et al. 1991; Llinás & Ribary 1993). Joliot et al. (1994) found evidence that directly ties coherent 40 Hz oscillations with perceptual grouping of clicks. Cell recordings in the auditory cortex show that neurons exhibit synchronous firing activity (Maldonado & Gerstein 1996; deCharms & Merzenich 1996; deCharms 1998). The study by Barth and MacDonald (1996) suggests that oscillations in the auditory cortex are originated within the cortex and synchrony is produced by intracortical interactions. The suggested anatomical substrate for coherent oscillations well agrees with that from the visual domain (Singer & Gray 1995). The Barth and MacDonald study further suggests that cortical oscillations can be modulated by the thalamus.

Concerning the LEGION architecture, local excitatory connections are broadly consistent with various lateral connections in the cortex. In the visual cortex, for example, horizontal connections (Gilbert & Wiesel 1989; Gilbert 1992) exist and they link pyramidal cells, which are known to be a chief type of excitatory neurons. With intracellular recordings and anatomic preparations, Gray and McCormick (1996) reported that pyramidal cells in the visual cortex may be responsible for generating synchronous cortical oscillations. The global inhibitor (see Fig. 6) serves to segment multiple patterns simultaneously present, thus exerting a global coordination. Crick (1984) has suggested that part of the thalamus, the thalamic reticular complex in particular, may be involved in the global control of selective attention. The thalamus is positioned at a key location in the brain: it receives input from and sends projections to almost the entire cortex. Thus, the global inhibitor could correspond to a neuronal group in the thalamus (Wang & Terman 1997); in this case, the activity of the inhibitor should be interpreted as the collective activity of the group.

Llinas and his colleagues (Ribary et al. 1991; Llinás & Ribary 1993), on the basis of their recordings from the auditory system, suggested that the thalamus plays the role of synchronizing

cortical oscillations through its mutual connections with the cortex rather than desynchronizing oscillator assemblies as suggested above. The question of whether cortical synchrony is produced by intracortical connections or thalamocortical connections may be answered by the following experiment. Let an auditory stimulus consist of two-tone interleaving sequences with high and low frequencies, respectively, so as to induce stream segregation (Bregman 1990). When streaming occurs, the LEGION model predicts that the global inhibitor oscillates with a frequency double that of cortical oscillations (Wang 1996). In contrast, the thalamocortical model for producing synchrony would predict the same frequency between cortical and thalamic oscillations. Note that the occurrence of stream segregation is a key condition, since, otherwise, the models do not yield contrasting predictions.

In Section 3.1 we mentioned that our single oscillator model is dynamically very similar to other neuronal models for generating membrane potentials or oscillating bursts of neuronal spikes. From the modeling perspective, relaxation oscillations best match oscillating envelopes of bursting activity. Figure 15 shows such an oscillating burst recorded from a single pyramidal neuron in the visual cortex (Gray & McCormick 1996). It is easy to see that the envelope of the burst would naturally be described as a relaxation oscillation.[8] Oscillating bursts have been argued to be more effective for synaptic transmission, thus are better candidates for binding, than single spikes (Gray & McCormick 1996). It is worth noting that the choice of relaxation oscillators (Terman & Wang 1995), motivated purely by computational considerations, is consistent with subsequent experimental data.
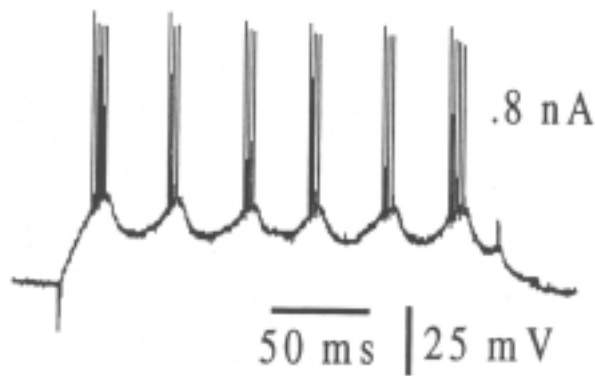
Figure 15. Membrane potential of a single neuron recorded from the cat striate cortex (from Gray & McCormick 1996).

5.2 Perceptual considerations

Based on a series of psychophysical experiments, Chen (1982; 1990) observed that human perception is sensitive to topological properties of stimuli; in particular, humans are more accurate in discriminating rapidly presented visual stimuli that have distinct topologies (number of holes). According to him, topological perception constitutes a basic and early part of perceptual organization. One can view that a hole inside a connected pattern is a distinct pattern, and as a result, a LEGION network will produce distinct responses to patterns with different numbers of holes. The difference in the number of segmented patterns emerging from LEGION provides an explanation for topological perception (Wang 2000). The fact that LEGION exhibits a fixed segmentation capacity generates the following prediction: topology-based discrimination occurs only up to a certain number of holes. On the other hand, a capacity limitation is not predicted by Chen's account based on mathematical topology (Chen 1990).

---

[8] See Wang et al. (1990) for a model that produces oscillating bursts, not just envelopes.

The Cesmeli and Wang model for motion analysis, described in Sect. 4.1, exhibits a number of important properties in human motion perception; these include motion transparency and a solution to the so-called blank wall problem - how to perceive a moving surface when no local motion signal can be detected in the interior of the surface. Subsequently, Cesmeli et al. (in press) showed that the model can account for the intriguing barber pole illusion (Wallach 1935), in which the perceived direction of motion of a grating changes merely as a result of changing the shape of an aperture. The model can also simulate a set of quantitative data from human perception of symmetrical and asymmetrical plaids (Stoner et al. 1990; Lindsey & Todd 1996), created by two superimposed gratings. Furthermore, the model is supported by a recent physiological study using moving plaids; Castelo-Branco et al. (2000) reported that neurons in two visual cortical areas synchronize their responses when the two gratings form a single moving surface, but the synchrony disappears when the two gratings form separate moving surfaces.

By extending LEGION to the auditory domain, Wang (1996) proposed an oscillator network to address stream segregation. The basic architecture is a 2-D LEGION network: one dimension represents time and another one represents frequency. The network demonstrates a set of psychophysical phenomena (Bregman 1990), including dependency on spectral and temporal proximity, sequential capturing, and competition among different perceptual organizations. Also, it is well known that the ability of listeners to identify two simultaneously presented vowels, or double vowels, can be improved by introducing a difference in fundamental frequency between the vowels. Brown and Wang (1997) proposed an oscillatory correlation model to explain this phenomenon, which represents the perceptual grouping of auditory frequency channels as synchronized oscillations.

5.3 Cognitive considerations

A basic implication of the oscillatory correlation theory in general, and the LEGION mechanism in particular, is a capacity limitation on segmentation and binding (see Section 3.2). The notion of a limited capacity naturally arises from relaxation oscillations, which have a non-instantaneous active phase, and a LEGION network precisely characterizes the capacity. This property of relaxation oscillators is not shared by spiking neurons (Campbell et al. 1999) or chaotic maps (Zhao & Macau 2001). Though the existence of such a capacity is sometimes viewed as a computational weakness (Wersing et al. 2001; Zhao & Macau 2001), we point out that capacity limitation is a fundamental property of cognitive processing. Capacity limits arise from a variety of information-processing tasks, including memory retrieval, attention, mental operations (e.g. addition), enumeration (subitizing), multi-object tracking, etc. (for reviews see Pashler 1998; Cowan 2001). Arguments have been made that limited capacity is a strength rather than weakness for information processing (e.g. MacGregor 1987; Kareev 1995).

When studying memory-based segmentation in an oscillator network, Wang et al. (1990) explicitly linked the model capacity with the magic number (7±2) of human STM capacity (Miller 1956). Subsequently, Lisman and Idiart (Lisman & Idiart 1995) developed a more detailed, oscillation-based STM model, where the 7±2 capacity results from the interaction between the gamma oscillation and a slower rhythm in the theta-alpha range (5 to 12 Hz). The magic number 7 symbolizes the existence of a limited capacity, but should not be taken literally. A recent, comprehensive examination concludes that the capacity is actually about 4 (Cowan 2001).

It is well documented that both STM and attention exhibit a limited capacity. What is the relation between them? Though often discussed in the literature as related, few studies directly address the question. The clearest answer is attempted by Cowan (1995; 2001), who provides a theoretical framework that ties a large number of studies from a variety of empirical paradigms. According to him, the focus of attention has a capacity about 4, and this is the only source of capacity limitations in cognitive processing; in other words, capacity limits exhibited from STM and other tasks result from capacity-limited attention. Attention provides a "global workplace" for mental operations (Baars 1988; Cowan 2001).

A typical situation to demonstrate capacity limits is reaction time (RT) in enumeration. Many experiments have consistently shown that the time people take to count small objects, say marbles,

increases very slowly when the number of objects goes up from 1 to 4, but rises at a much faster pace after that (Jevons 1871; Kaufman et al. 1949; Mandler & Shebo 1982). Increases in RT in both conditions follow a linear trend. Put it differently, the RT slope is small from 1 to 4 items, and becomes much larger when the number of items is greater than 4. From the viewpoint of the LEGION dynamics, assuming a segmentation capacity of 4, up to 4 items can be segregated and fit into a single oscillation period (see Fig. 8). Beyond this capacity, some segments contain multiple items and need to be further segregated to support correct counting. But the oscillators corresponding to different items within a segment are already synchronized, and, as shown by Wang (1999a) in the context of object selection, further segmentation can produce only one new item per oscillation period because of the way the selective gating mechanism works. Thus, two distinct slopes would result. Our explanation differs from that given by Pylyshyn (1994). He assumes that there is a preattentive stage to select a limited number of individual items, which is a stage intermediate between parallel processing and serial attention. Unlike our explanation, the capacity limit is an assumption in Pylyshyn's theory.

When studying stream segregation, Wang (1996) suggested a shifting synchronization theory to explain the loss of temporal order when streaming occurs (Bregman 1990; see Sect. 6.2). The main point of the theory is that attention rapidly alternates between multiple streams. A later study (Wang 1999a) shows how to selectively focus on one or a small number of visual patterns. Though closely related to attention, neither study purported to be an explicit theory of attention. Wrigley and Brown (2001) recently proposed a two-layer oscillator model of auditory attention. In their model, the first layer is a LEGION array that performs stream segregation and the second layer performs attentional selection. Motivated by these studies and Cowan's analysis, I suggest that oscillatory correlation, originally proposed to address the binding problem, may also be viewed as a theory of attention with the following additional claims and qualifications. First, attention holds all the organizations that correspond to enabled and separated oscillator assemblies. The term "organization" is neutral to individual modalities, and can mean a visual object, an auditory stream, a chunk, and so on, in a specific situation. This claim implies that attention is paid to more than one organization simultaneously. By "simultaneity" I refer to a psychological time scale, or a psychological moment (Pöppel & Logothetis 1986), approximately in the range 10-50 ms (Wang et al. 1990). The psychological moment is the finest time scale for conscious awareness, and it roughly corresponds to the periods of gamma oscillations. Note that this account differs from the Wrigley and Brown model that does not allow attention to be shared by more than one stream. Also, by virtue of the LEGION mechanism, our claim implies a limited capacity of attention, which is the same as the segmentation capacity. Second, multiple organizations within the focus of attention oscillate on a physiological time scale (up to 10 ms), which has a finer time resolution than the psychological moment and is thus too fine to enter conscious experience. The phases of oscillator assemblies give distinct identities for the organizations attended to at a time.

I realize that this is a potentially provocative suggestion, stated here without a systematic development. Nonetheless, the theory immediately leads to two very broad implications: (1) Perceptual organization, or feature binding, is the same process as attending; (2) attention is capacity-limited but can be directed to more than one object (see Cowan 2001, for an extensive argument). That perceptual organization requires attention sharply contrasts with the popular view that there is a preattentive process that operates on the sensory input in parallel without the involvement of attention. Pashler (1998) examined this view in detail and concluded "only a very small amount of evidence even bears on it, and these data are somewhat equivocal" (p. 235). Indeed, data from several experiments specifically designed to address this issue suggest that attention is needed for typical "preattentive" tasks. These include a feature-based visual search task (Joseph et al. 1997) and an auditory streaming task (Carlyon et al. 2001). Perceptual organization would require a form of divided attention (Pashler 1998). Then, how to reconcile between a capacity limit of 4 (Cowan 2001) and a phenomenological observation that one can focus on only one thing at a time? A capacity limit represents an upper bound on the number of items held by attention, and it does not necessarily mean that the attention span is constantly full. It may be possible, for instance, for a subject to selectively attend to one thing or two in order to extract more detailed information from the attended items. Even in the case of selective attention, unselected

27

items still receive some analysis. In the classic experiment of Cherry (1953), for example, listeners can detect the change of the speaker gender (and a tone) from the "unattended" ear. Furthermore, when a subject is asked to perform a demanding task, such as repeating spoken phrases in Cherry's experiment, it may be that the task itself occupies several "slots" in the attention span, making it difficult to attend to other items at the same time.

6. Discussion issues

6.1 External time vs. internal time

In the oscillatory correlation theory, time plays the role of binding: different segments unfold in time. Let us refer to this putative role of time as internal time. Time is also a dimension in the physical world, and indeed a defining dimension for auditory and other temporal patterns. To complicate the matter further, time, in the form of common onsets and offsets, is also a grouping principle (Bregman 1990; Leonards et al. 1996; Lee & Blake 1999). Let us refer to this as external time: time that is external to the organism. A potential difficulty for the double use of time has been raised (Brown 2002), and used as an argument against the temporal correlation theory (Shadlen & Movshon 1999).

Onset/offset detectors are identified in the auditory system (Popper & Fay 1992), and have been used in auditory models. With such detectors, grouping based common onsets/offsets is not unlike that based on other cues. A similar idea can be extended to the visual domain. How to distinguish internal time from external time as used in temporal patterns depends on how time is represented in temporal patterns. In neural network modeling of temporal patterns, time is usually coded by delay lines, decay traces, or exponential kernels (Wang 2002). Delay lines convert time to space, and hold the most recent patterns for a certain period of time. This way of representing time has also been used in the context of auditory segregation (Wang 1996; Wang & Brown 1999), and as demonstrated in this case the potential conflict of the double use is not present. Decay traces encode time implicitly and compactly, but have limited discriminative power; for instance, it is unclear how they could underlie a variety of auditory functions, such as pitch and rhythm perception. The use of exponential kernels strikes a reasonable compromise between these two cases, and it converts time into a logarithmic axis of space so that more recent traces are represented with higher temporal resolutions. The distinction between internal time and external time can be made similarly as in the case of time delays. This way of coding time bears resemblance to how space is coded on the retina: higher resolution for image parts nearer to the fovea. Both delay lines and exponential kernels form a shifting representation (Wang 1996). Given the high resolution of temporal processing in the auditory system (Moore 1997), it is likely that internal time needs to be preserved during the shifting process. This raises the interesting issue of how internal time can be maintained during neural transmission. Though one can imagine ways of dealing with this issue, it has not been systematically addressed.

6.2 Spatial relations

In terms of addressing Rosenblatt's challenge, our discussion so far is exclusively on the figure-ground separation problem. The other problem is how to compute topological or spatial relations among objects. We asserted earlier that one needs to first solve the separation problem in order to compute geometrical relations. Given the challenge of solving the figure-ground separation, very little research has been conducted to address the relation problem.

Building on the LEGION ability to perform figure-ground separation, Chen and Wang (2001) recently addressed one particular question: how to tell whether a dot belongs area A or area B, as illustrated in Figure 16. This can be phrased as how to compute the inside/outside relation (Ullman 1984). The solution by Chen and Wang is to first separate the two areas apart using a LEGION network, and then decide whether the oscillator corresponding to the dot is in the assembly

representing *A* or that representing *B*. It is interesting to contrast this solution with that offered by Ullman (1984; 1996) in his framework of visual routines. His suggested solution to the inside/outside problem is a visual routine called coloring, which spreads activity from the dot location until the boundary of an area is reached. Although LEGION dynamics for synchronization has some resemblance to a coloring process, there are several differences between the two solutions. First, the coloring routine is described as a serial algorithm, while for LEGION synchronization emerges from a network of interacting oscillators. Second, the time course of synchronization enables Chen and Wang to explicitly distinguish between effortless perception with simple boundaries (see the upper frame of Fig. 16) and effortful perception with convoluted boundaries (see the lower frame of Fig. 16). It is not clear how a qualitative distinction can be made in the coloring process.
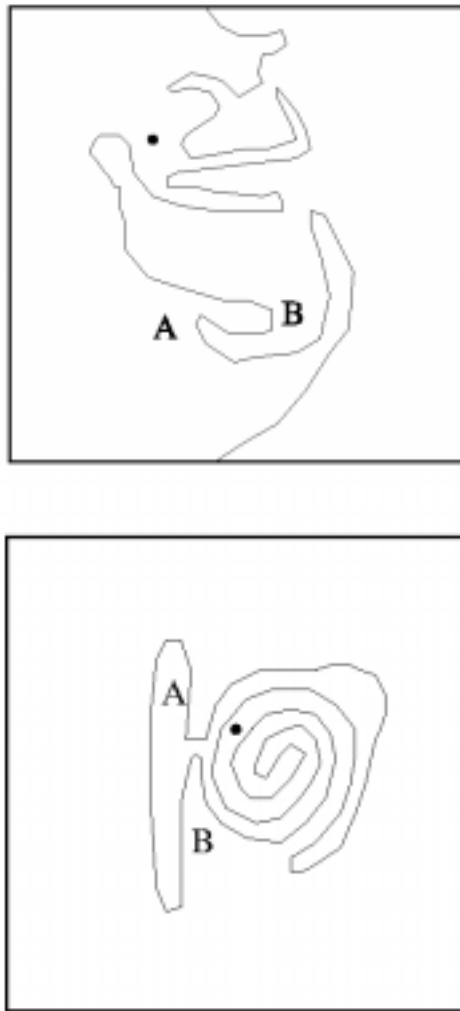
Figure 16. Inside/outside relation (From Chen & Wang 2001). The top frame shows an example where the boundary between area *A* and area *B* is not very convoluted, whereas the bottom frame shows another example with a very convoluted boundary.

We believe, for the following reasons, that the oscillatory correlation theory lays a general foundation to compute a spatial relation between multiple objects. Arbitrary objects can be segmented by a LEGION network, and relevant ones for computing a specific relation can be further selected (see Sect. 4.2). The relevant objects are all activated and yet separated in phase;

this provides a workspace (Baars 1988; Cowan 2001) to calculate attributes from each object and compare them across different ones. Representing each object as an oscillator assembly gives a broad base to derive object-level properties, such as its center, size, spatial extent, etc., which are important for computing geometrical relations (e.g. *left-of*). How geometrical relations can be systematically computed is an important topic for future research.

6.3 Compositionality

The issue of compositionality has received much attention in the debate between connectionism and symbolism (e.g. Fodor & Pylyshyn 1988; Smolensky 1988). The question is whether neural networks possess the representational power to deal with combinatorial structure that is manifested in our ability to process relational and syntactical information. A trivial yes answer can be derived from the fact that neural networks are general-purpose computing devices that can simulate universal Turing machines (Arbib 1995), but such a recourse does not address the critical assessment that neural networks are only an implementation theory and cannot account for cognitive functions. The key issue is whether a fixed network can encode and process hierarchical relations in a flexible way.

We think that the introduction of the time dimension opens an entirely new avenue to address the issue of compositionality. With network architecture fixed, the time dimension provides the critical flexibility: hierarchy could be encoded in time. In the context of range image segmentation, Liu and Wang (1999) showed how a range-defined object can be hierarchically decomposed into its parts (or surfaces) by gradually decreasing the level of global inhibition in a fixed LEGION network; related parts may synchronize at one level of global inhibition and become desynchronized at an increased level. With the ability to select a segment for further analysis as explained in Sect. 4.2, arbitrary hierarchies could be embodied. For example, an embedded tree structure ((*A B*) (*C D*)) could be coded by first forming two assemblies corresponding to (*A B*) and (*C D*), respectively, and then each assembly is selected and further decomposed into two assemblies corresponding to two terminal symbols. This way of representing syntactical structure converts structural complexity into temporal complexity, and time being an infinitely extensible dimension allows the system to have *in principle* unbounded capacity to deal with the compositionality of data structures. This latter property has been argued to be a defining property of symbolic architecture, not shared by connectionist architecture (Fodor & Pylyshyn 1988). Furthermore, embedding combinatorial structure in time makes *processing time* a relevant quantity - problem solving is viewed as a temporal process and one naturally takes more or less time to solve a particular problem, depending on the difficulty of the problem.

To describe spatial relations between objects in a scene, discussed in Sect. 6.2, would require that the system have the capability to deal with syntactical structure. So how to compute spatial relations should have significant bearing on the compositionality issue. In a related study on natural language representation, Shastri and Ajjanagadde (1993) described the use of oscillatory correlation to dynamically bind arguments and constants in order to perform reasoning with predicates and rules. This is a form of instantiation that binds an abstract slot (say "recipient") and a specific filler (say "John"); see von der Malsburg (1999) for a general discussion on instantiation as an application that can benefit from a solution to the binding problem.

6.4 Binding and attention

It has been frequently suggested that selective attention plays the role of binding. In particular, according to the dominant feature integration theory of Treisman and Gelade (1980), the visual system first analyzes a scene in parallel by separate retinotopic feature maps and focal attention integrates the analyses of different feature maps to produce a coherent perceptual object. In other words, attention provides a "spotlight" on the location map to select an object (Treisman 1986). Arguing from the neurobiological perspective, Reynolds and Desimone (1999) also suggested that attention provides a solution to the binding problem. Our theoretical analysis on neural competition

and object selection (Sect. 4.2) suggests instead that selective attention operates on the results of binding. So the key question is whether attention precedes or succeeds binding.

A visual object can have an arbitrary shape and size. This situation creates the following inconsistency in the feature integration theory. On the one hand, it is a location-based theory of attention that binds at the same location individual analyses from different feature maps. On the other hand, to select an object attention spotlight must also have arbitrary shape and size, adapting to a specific object and thus object-based. Without a binding process, what produces such an adaptive spotlight? This is an intrinsic difficulty if focal attention, rather than perceptual organization, is to bind features across different locations. The difficulty is illustrated by the finding of Field et al. (1993) that a group of curvilinear (snake-like) elements stands out from a scene of randomly oriented elements and can be detected by observers, whereas other groups cannot be detected. An analogous effect was found in the monkey cortex (Kapadia et al. 1995). Note that there is virtually an infinite number of "snakes" that can be constructed from orientation elements, and grouping is required to yield a snake pattern to be illuminated by attention spotlight.

The above difficulty does not occur if one adopts the view that focal attention occurs after binding, which provides multiple segments for focal attention to perform sequential analysis. This is fully consistent with the object-based view of visual attention, as mentioned in Sect. 4.2. Though sometimes difficult to tear object-based attention apart from location-based attention, since the former implicitly provides the information for the latter, recent psychophysical and neuropsychological studies support the object-based view (Nakayama et al. 1995; Mattingley et al. 1997; Driver & Baylis 1998). Pertinent to the Field et al. study, the relevant data have been successfully simulated by the oscillation model of Yen and Finkel (1998) discussed earlier.

6.5 Binding and recognition

An issue related to the discussion of Sect. 6.4 is whether binding should be a process separate from recognition or it is simply part of recognition. According to the latter view, binding occurs as a byproduct of recognition, which is typically coupled with some selection mechanism that brings the pattern of interest into focus, and there is really no binding problem so to speak (Riesenhuber & Poggio 1999). For example, Fukushima and Imagawa (1993) proposed a model that performs recognition and segmentation simultaneously by employing a search controller that selects a small area of the input image for processing. Their model is based on Fukushima's neocognitron model for pattern recognition, which is a hierarchical multilayer network, and this model exemplifies the hierarchical coding approach to the binding problem. The model contains a cascade of many layers with both forward and backward connections. The forward path performs pattern recognition that is robust to a range of variations in position and size, and the last layer stores learned patterns. When a scene of multiple patterns is presented, a rough area selection is performed based on feature density of the input, and further competition in the last layer would lead to a winner. The winning unit of the last layer, through backward connections, reinforces the pattern of the input image that is consistent with the stored template. This, in a sense, segments that part of the input image from its background. After a while, the network switches to another area of high feature density and continues the analysis process. Their model has been evaluated on binary images of connected characters. Olshausen et al. (1993) proposed a model that also combines pattern recognition and a model of selective attention. Their attention model is implemented by a shifting circuit that routes information in a hierarchical network while preserving spatial relations between visual features, and recognition is based on a Hopfield model of associative memory. The location and size of an attention blob are determined by competition in a feature saliency map, producing potential regions of interest on an image. This model is viewed by Shadlen and Movshon (1999) as an alternative to the temporal correlation theory. The model is evaluated on binary images with well-separated patterns. A recent model along a similar line was proposed by Riesenhuber and Poggio (1999), and it uses a hierarchical architecture similar to the neocognitron. Their model has been tested on two-object scenes: one is a stored pattern and another is a distractor.

Besides the conceptual difficulties with the hierarchical coding discussed in Section 2, it is unclear how these models can be extended to computationally analyze scenes where complex

objects are arranged in arbitrary ways. Again, snake-like patterns studied by Field et al. (1993) illustrate computational problems of binding as recognition. The prohibitively large number of snake shapes makes it infeasible to search for all possible snake patterns. Even with a prespecified pattern, Field et al. (1993) demonstrate that observers cannot identify the pattern when its elements are not arranged in a curvilinear fashion.

6.6 Read out

An issue often cited as a problem to the oscillatory correlation representation concerns how a synchronized code is decoded by a later processing stage (see Ghose & Maunsell 1999). The readout problem is not really unique to temporal coding and, as stated by Roskies (1999), it is "one of the most puzzling and fundamental problems for systems neuroscience in general". If later processing, say recognition, requires information across a large part of the visual field, whether that information is encoded via temporal correlation or any other means, it must somehow be decoded in a corresponding way.

While the readout issue in temporal coding is an open question in neuroscience, some computational considerations may be helpful. As illustrated in Fig. 8, the basic claim of oscillatory correlation is that each segment pops out at a distinct time from the network and different segments alternate in time. In the case of LEGION dynamics, a segment is in the active phase when it pops out. As a result, all of the features of the segment, but none of the features from competing segments, are simultaneously available for later processing tasks such as selective attention and recognition. This way of encoding a pattern, i.e. activating all of its features, is most commonly used in neural models for pattern recognition, e.g. perceptrons. The model of Wang and Liu (2002), discussed in Sect. 4.4, shows a concrete way in which LEGION-based segmentation is coupled with an associative memory model for recognition. Their model performs scene analysis in a closed loop.

7. Conclusion: Versatile computing requires the time dimension

The substrate for diverse mental functions in perception, reasoning, and action is a gigantic network of neurons whose common language is a neuronal signal. The fundamental claim of the temporal (and oscillatory) correlation theory is that binding is manifested in the time structure of such a signal.

If there is one difference that stands out between natural intelligence and artificial intelligence, it is the versatility of the former. Furthermore, natural intelligence emerges from a concrete neural network - an individual brain - whose architecture is more or less fixed after development. As pointed out by von der Malsburg (1999), a typical practice in neural computation, and artificial intelligence in general, is that "give me a concrete problem and I will devise a network that solves it." This is the principle of universality, in the sense of universal Turing machines or multilayer perceptrons. The problem that faces the brain is a rather different one: "given the concrete network learn to cope with situations and problems as they arise." Let me call this the principle of versatility (see also Singer 1999). In other words, the difference between universality and versatility comes down to "first the problem then the network" versus "first the network then the problem".

How can such a network give rise to versatility as wide-ranging as from sensory response, perceptual organization, to language processing and long-term planning? I believe that time provides a necessary dimension for the network to fulfill its various functional requirements. The time dimension is flexible and infinitely extensible - a characteristic not shared by spatial organization of the network, no matter how complex it is.

Appendix: On the number of connected patterns

For a 1-D image R with n binary pixels, it is easy to see that the number of connected patterns is $n(n+1)/2$. For a 2-D R, the number of connected patterns has been stated to be an exponential relation with respect to $|R|$ (p. xiv, Hinton & Sejnowski 1999; p. 132, Wang 2000). However, I cannot find a proof to this conclusion in the literature, and it is not straightforward to give a precise number of connected patterns on an image of the size mxn, where $m > 1$. Thus, I furnish in this appendix a proof on a 2xn figure, and from it to an nxn image.

Theorem 1. The number of connected patterns on a 2xn binary image, R, increases exponentially with $|R|$.

Proof. Let us arrange R in a 2-row n-column layout, as shown in Figure 17A. Denote the number of connected patterns on such R as $N(n)$. Thus, we have $N(1) = 3$, $N(2) = 13$. For $n > 2$, consider R as formed by appending a 2x(n–1) image with an additional column at the right of the image (see Figure 17A). One can divide all connected figures into two sets, those containing no black pixel in column n–1 and the remainder. The size of the first set is simply $N(n-2) + N(1)$, counting those to the left of column n–1 and those to the right. The second set must have at least one black pixel in column n–1, and is determined by possible ways of appending column n. There are four different ways of expanding to column n, as illustrated in Figure 17B. First, it involves no black pixel in column n and the number of such patterns is just $N(n-1) - N(n-2)$. Second, it involves expanding when the upper pixel in column n–1 is black and the lower one is white, and this leads to two distinct subsets of connected patterns depending on whether column n has one or two black pixels; see Figure 17B. The third way is a symmetrical case when the upper pixel in column n is white but the lower one is black. The fourth way involves expanding when both pixels in column n–1 are black, and this leads to three distinct sets of connected patterns, as shown in Figure 17B. Even counting only two such subsets, the total number of connected patterns that involve black pixels in both column n–1 and column n is $2[N(n-1) - N(n-2)]$. Thus we have the inequality

$$N(n) > [N(n-2) + N(1)] + [N(n-1) - N(n-2)] + 2[N(n-1) - N(n-2)]$$
$$= 2N(n-1) + [N(n-1) - 2N(n-2)] + 3$$
$$> 2N(n-1) + [N(n-1) - 2N(n-2)]$$

Because $N(2) - 2N(1) > 0$, we have the following recurrence inequality,

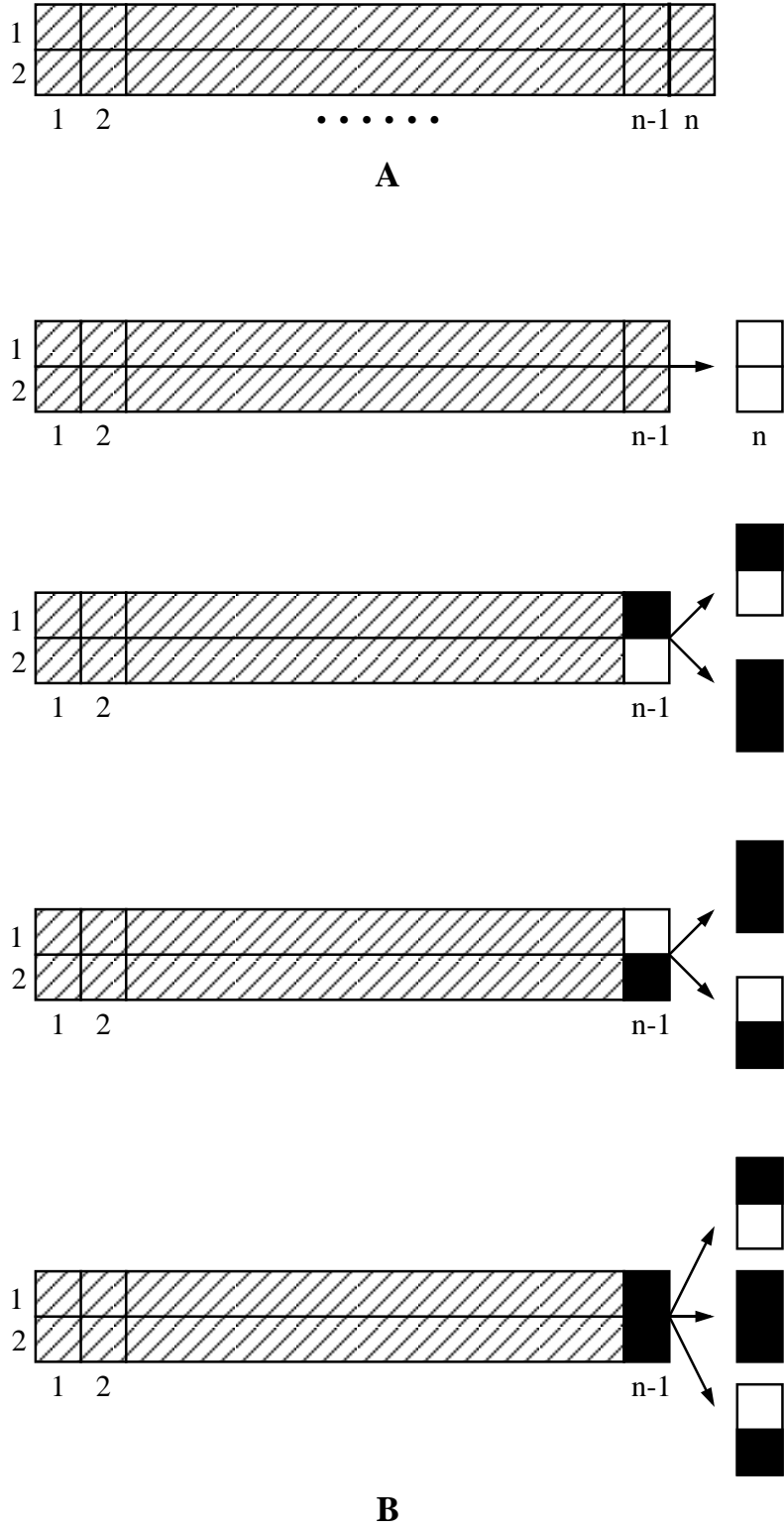$$N(n) > 2N(n-1) \tag{A1}$$

Thus, we have $N(n) > 2^n$. This completes the proof.

Given that all connected patterns on a 2xn image are also connected patterns on an nxn image for $n \geq 2$, we have the following corollary:

Corollary 1. The number of connected patterns on an nxn image, R, increases exponentially with $|R|$.

Figure 17. Counting the number of connected patterns on a 2xn image. **A**. A 2xn grid. **B**. Four different ways of expanding from a 2x(n–1) image to include column n.

References

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

Arbib, M. A. (Ed.). (1995). *Handbook of brain theory and neural networks*. Cambridge, MA: MIT Press.

Arbib, M. A. (Ed.). (2002). *Handbook of brain theory and neural networks* (2nd ed.). Cambridge MA: MIT Press.

Baars, B. J. (1988). *A cognitive theory of consciousness*. London: Cambridge University Press.

Barlow, H. B. (1972). Single units and cognition: A neurone doctrine for perceptual psychology. *Perception*, **1**, 371-394.

Barth, D. S., & MacDonald, K. D. (1996). Thalamic modulation of high-frequency oscillating potentials in auditory cortex. *Nature*, **383**, 78-81.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford U.K.: Oxford University Press.

Black, M. J., & Anandan, P. (1996). The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *Computer Vision, Graphics, and Image Processing: Image Understanding*, **63**, 75-104.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge MA: MIT Press.

Brown, G. J. (2002). Auditory scene analysis. In M. A. Arbib (Ed.), *Handbook of brain theory and neural networks*, 2nd ed. (in press). Cambridge MA: MIT Press.

Brown, G. J., & Wang, D. L. (1997). Modelling the perceptual segregation of double vowels with a network of neural oscillators. *Neural Networks*, **10**, 1547-1558.

Campbell, S. R., Wang, D. L., & Jayaprakash, C. (1999). Synchrony and desynchrony in integrate-and-fire oscillators. *Neural Computation*, **11**, 1595-1619.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, **27**, 115-127.

Castelo-Branco, M., Goebel, R., Neuenschwander, S., & Singer, W. (2000). Neural synchrony correlates with surface segregation rules. *Nature*, **405**, 685-689.

Cesmeli, E., Lindsey, D. T., & Wang, D. L. (in press). An oscillatory correlation model of visual motion analysis. *Perception & Psychophysics*.

Cesmeli, E., & Wang, D. L. (2000). Motion segmentation based on motion/brightness integration and oscillatory correlation. *IEEE Transactions on Neural Networks*, **11**, 935-947.

Cesmeli, E., & Wang, D. L. (2001). Texture segmentation using Gaussian-Markov random fields and neural networks. *IEEE Transactions on Neural Networks*, **12**, 394-404.

Chen, K., & Wang, D. L. (2001). Perceiving geometric patterns: from spirals to inside-outside relations. *IEEE Transactions on Neural Networks*, **12**, 1084-1102.

Chen, K., Wang, D. L., & Liu, X. (2000). Weight adaptation and oscillatory correlation for image segmentation. *IEEE Transactions on Neural Networks*, **11**, 1106-1123.

Chen, L. (1982). Topological structure in visual perception. *Science*, **218**, 699-700.

Chen, L. (1990). Holes and wholes: A reply to Rubin and Kanwisher. *Perception & Psychophysics*, **47**, 47-53.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of Acoustical Society of America*, **25**, 975-979.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

Cowan, N. (2001). The magic number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, **24**, 87-185.

Crick, F. (1984). Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Sciences of USA*, **81**, 4586-4590.

deCharms, R. C. (1998). Information coding in the cortex by independent or coordinated populations. *Proceedings of the National Academy of Sciences of USA*, **95**, 15166-15168.

deCharms, R. C., & Merzenich, M. M. (1996). Primary cortical representation of sounds by the coordination of action-potential timing. *Nature*, **381**, 610-613.

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*. Ph.D dissertation, Stanford University.

Driver, J., & Baylis, G. C. (1998). Attention and visual object recognition. In R. Parasuraman (Ed.), *The attentive brain* (pp. 299-326). Cambridge MA: MIT Press.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). New York: Wiley & Sons.

Eckhorn, R., *et al* (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex. *Biological Cybernetics*, **60**, 121-130.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, **2**, 704-716.

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

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Research*, **33**, 173-193.

FitzHugh, R. (1961). Impulses and physiological states in models of nerve membrane. *Biophysical Journal*, **1**, 445-466.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, **28**, 3-71.

Freeman, W. J. (1978). Spatial properties of an EEG event in the olfactory bulb and cortex. *Electroencephalographical and Clinical Neurophysiology*, **44**, 586-605.

Fukushima, K., & Imagawa, T. (1993). Recognition and segmentation of connected characters with selective attention. *Neural Networks*, **6**, 33-41.

Galambos, R., Makeig, S., & Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proceedings of the National Academy of Sciences of USA*, **78**, 2643-2647.

Ghose, G. M., & Maunsell, J. (1999). Specialized representations in visual cortex: a role for binding? *Neuron*, **24**, 79-85.

Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. *Neuron*, **9**, 1-13.

Gilbert, C. D., & Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience*, **9**, 2432-2442.

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

Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, **338**, 334-337.

Gray, C. M., & McCormick, D. A. (1996). Chattering cells: superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex. *Science*, **274**, 109-113.

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

Hinton, G., & Sejnowski, T. J. (Ed.). (1999). *Unsupervised learning: Foundations of neural computation*. Cambridge MA: MIT Press.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology (London)*, **117**, 500-544.

Horn, D., & Opher, I. (1999). Collective excitation phenomena and their applications. In W. Maass & C. M. Bishop (Ed.), *Pulsed neural networks* (pp. 297-320). Cambridge MA: MIT Press.

Horn, D., & Usher, M. (1991). Parallel activation of memories in an oscillatory neural network. *Neural Computation*, **3**, 31-44.

Jevons, W. S. (1871). The power of numerical discrimination. *Nature*, **3**, 281-282.

Joliot, M., Ribary, U., & Llinas, R. (1994). Human oscillatory brain activity near to 40 Hz coexists with cognitive temporal binding. *Proceedings of the National Academy of Sciences of USA*, **91**, 11748-11751.

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a 'preattentive' feature search task. *Nature*, **387**, 805-807.

Kahneman, D., Treisman, A., & Gibbs, B. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology*, **24**, 175-219.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron*, **15**, 843-856.

Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition*, **56**, 263-269.

Kaufman, E., Lord, M., Reese, T., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology*, **62**, 498-525.

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

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt.

Lee, S. H., & Blake, R. (1999). Visual form created solely from temporal structure. *Science*, **284**, 1165-1168.

Leonards, U., Singer, W., & Fahle, M. (1996). The influence of temporal phase differences on texture segmentation. *Vision Research*, **36**, 2689-2697.

Lindsey, D. T., & Todd, J. T. (1996). On the relative contributions of motion energy and transparency to the perception of moving plaids. *Vision Research*, **36**, 207-222.

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.

Lisman, J. E., & Idiart, M. A. P. (1995). Storage of 7±2 short-term memories in oscillatory subcycles. *Science*, **267**, 1512-1515.

Liu, X., Chen, K., & Wang, D. L. (2001). Extraction of hydrographic regions from remote sensing images using an oscillator network with weight adaptation. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 207-211.

Liu, X., & Wang, D. L. (1999). Range image segmentation using a relaxation oscillator network. *IEEE Transactions on Neural Networks*, **10**, 564-573.

Llinás, R., & Ribary, U. (1993). Coherent 40-Hz oscillation characterizes dream state in humans. *Proceedings of the National Academy of Sciences of USA*, **90**, 2078-2082.

Lourenco, C., Babloyantz, A., & Hougardy, M. (2000). Pattern segmentation in a binary/analog world: unsupervised learning versus memory storing. *Neural Networks*, **13**, 71-89.

MacGregor, J. N. (1987). Short-term memory capacity: Limitation or optimization? *Psychological Review*, **94**, 107-108.

Maldonado, P. E., & Gerstein, G. L. (1996). Neuronal assembly dynamics in the rat auditory cortex during reorganization induced by intracortical microstimulation. *Experimental Brain Research*, **112**, 431-441.

Mandler, G., & Shebo, B. J. (1982). Subitizing: an analysis of its component processes. *Journal of Experimental Psychology: General*, **111**, 1-22.

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

McClelland, J. L., & Rumelhart, D. E. (Ed.). (1986). *Parallel distributed processing, vol. 2: Psychological and biological models*. Cambridge MA: MIT Press.

Mermin, N. D., & Wagner, H. (1966). Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic Heisenberg models. *Physical Review Letters*, **17**, 1133-1136.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, **63**, 81-97.

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

Minsky, M. L., & Papert, S. A. (1969). *Perceptrons*. Cambridge MA: MIT Press.

Minsky, M. L., & Papert, S. A. (1988). *Perceptrons (Expanded edition)*. Cambridge MA: MIT Press.

Moore, B. C. J. (1997). *An introduction to the psychology of hearing* (4th ed.). San Diego, CA: Academic Press.

Morris, C., & Lecar, H. (1981). Voltage oscillations in the barnacle giant muscle fiber. *Biophysical Journal*, **35**, 193-213.

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.

Nagumo, J., Arimoto, S., & Yoshizawa, S. (1962). An active pulse transmission line simulating nerve axon. *Proceedings of the Institute of Radio Engineers*, **50**, 2061-2070.

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 (Ed.), *An invitation to cognitive science* (pp. 1-70). Cambridge MA: MIT Press.

Niebur, E., & Koch, C. (1998). Computational architectures for attention. In R. Parasuraman (Ed.), *The attentive brain* (pp. 163-186). Cambridge, MA: MIT Press.

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, **13**, 4700-4719.

Palmer, S. E. (1999). *Vision science*. Cambridge MA: MIT Press.

Parasuraman, R. (Ed.). (1998). *The attentive brain*. Cambridge MA: MIT Press.

Pashler, H. E. (1998). *The psychology of attention*. Cambridge MA: MIT Press.

Pöppel, E., & Logothetis, N. (1986). Neuronal oscillations in the human brain. *Naturwissenschaften*, **73**, 267-268.

Popper, A. N., & Fay, R. R. (Ed.). (1992). *The mammalian auditory pathway: Neurophysiology*. New York: Springer-Verlag.

Pylyshyn, Z. (1994). Some primitive mechanisms of spatial attention. *Cognition*, **50**, 363-384.

Reynolds, J. H., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron*, **24**, 19-29.

Rhouma, M. B. H., & Frigui, H. (2001). Self-organization of pulse-coupled oscillators with application to clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**, 180-195.

Ribary, U., *et al* (1991). Magnetic field tomography of coherent thalamocortical 40-Hz oscillations in humans. *Proceedings of the National Academy of Sciences of USA*, **88**, 11037-11041.

Riesenhuber, M., & Poggio, T. (1999). Are cortical models really bound by the "binding problem"? *Neuron*, **24**, 87-93.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, **65**, 386-408.

Rosenblatt, F. (1962). *Principles of neural dynamics*. New York: Spartan.

Roskies, A. L. (1999). The binding problem. *Neuron*, **24**, 7-9.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Ed.), *Parallel distributed processing* (pp. 318-362). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (Ed.). (1986). *Parallel distributed processing, vol. 1: Foundations*. Cambridge MA: MIT Press.

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

Sejnowski, T. J., & Hinton, G. E. (1987). Separating figure from ground with a Boltzmann machine. In M. A. Arbib & A. R. Hanson (Ed.), *Vision, brain, and cooperative computation* (pp. 703-724). Cambridge MA: MIT Press.

Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron*, **24**, 67-77.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, **16**, 417-494.

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

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, **11**, 1-74.

Somers, D., & Kopell, N. (1993). Rapid synchrony through fast threshold modulation. *Biological Cybernetics*, **68**, 393-407.

Sompolinsky, H., & Tsodyks, M. (1994). Segmentation by a network of oscillators with stored memories. *Neural Computation*, **6**(4), 642-657.

Sporns, O., Tononi, G., & Edelman, G. M. (1991). Modeling perceptual grouping and figure-ground segregation by means of active re-entrant connections. *Proceedings of the National Academy of Sciences of USA*, **88**, 129-133.

Stoner, G. R., Albright, T. D., & Ramachandra, V. S. (1990). Transparency and coherence in human motion perception. *Nature*, **344**, 153-155.

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

Treisman, A. (1999). Solutions to the binding problem: progress through controversy and convergence. *Neuron*, **24**, 105-110.

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

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perceptron of objects. *Cognitive Psychology*, **14**, 107-141.

Ullman, S. (1984). Visual routines. *Cognition*, **18**, 97-159.

Ullman, S. (1996). *High-level vision*. Cambridge MA: MIT Press.

Usrey, W. M., & Reid, R. C. (1999). Synchronous activity in the visual system. *Annual Review of Physiology*, **61**, 435-456.

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

van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, **21**, 1-14.

van Gelder, T., & Port, R. (1995). It's about time: an overview of the dynamical approach to cognition. In R. Port & T. van Gelder (Ed.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 1-43). Cambridge MA: MIT Press.

Varela, F., Lachaux, J. P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, **2**, 229-239.

von der Malsburg, C. (1981). *The correlation theory of brain function*. Internal Report 81-2 (Reprinted in *Models of neural networks* II, E. Domany, J.L. van Hemmen, and K. Schulten, eds., Berlin: Springer, 1994), Max-Planck-Institute for Biophysical Chemistry.

von der Malsburg, C. (1999). The what and why of binding: the modeler's perspective. *Neuron*, **24**, 95-104.

von der Malsburg, C., & Schneider, W. (1986). A neural cocktail-party processor. *Biological Cybernetics*, **54**, 29-40.

Wallach, H. (1935). Uber visuell wahrgenommene bewegungsrichtung. *Psychologische Forschung*, **20**, 325-380.

Wang, D. L. (1993a). Modeling global synchrony in the visual cortex by locally coupled neural oscillators. In *Proc. of 15th Ann. Conf. Cognit. Sci. Soc.* (pp. 1058-1063). Boulder CO:

Wang, D. L. (1993b). Pattern recognition: Neural networks in perspective. *IEEE Expert*, **8**, 52-60, August.

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

Wang, D. L. (1996). Primitive auditory segregation based on oscillatory correlation. *Cognitive Science*, **20**, 409-456.

Wang, D. L. (1999a). Object selection based on oscillatory correlation. *Neural Networks*, **12**, 579-592.

Wang, D. L. (1999b). Relaxation oscillators and networks. In J. Webster (Ed.), *Encyclopedia of electrical and electronic engineers* (pp. 396-405, also available on the web at www.cis.ohio-state.edu/~dwang). New York: Wiley.

Wang, D. L. (2000). On connectedness: a solution based on oscillatory correlation. *Neural Computation*, **12**, 131-139.

Wang, D. L. (2002). Temporal pattern processing. In M. A. Arbib (Ed.), *Handbook of brain theory and neural networks*, 2nd ed. (in press). Cambridge MA: MIT Press.

Wang, D. L., & Brown, G. J. (1999). Separation of speech from interfering sounds based on oscillatory correlation. *IEEE Transactions on Neural Networks*, **10**, 684-697.

Wang, D. L., Buhmann, J., & von der Malsburg, C. (1990). Pattern segmentation in associative memory. *Neural Computation*, **2**, 95-107 (Reprinted in *Olfaction*, J.L. Davis and H. Eichenbaum, ed., Cambridge MA: MIT Press, pp. 213-224, 1991).

Wang, D. L., Kristjansson, A., & Nakayama, K. (2001). *Efficient visual search without top-down or bottom-up guidance: A putative role for perceptual grouping*. Technical Report 26, Ohio State University Center for Cognitive Science.

Wang, D. L., & Liu, X. (2002). Scene analysis by integrating primitive segmentation and associative memory. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, **32**, 254-268.

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 Comp.*, vol. 9, pp. 1623-1626, 1997).

Wersing, H., Steil, J. J., & Ritter, H. (2001). A competitive-layer model for feature binding and sensory segmentation. *Neural Computation*, **13**, 357-388.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II. *Psychologische Forschung*, **4**, 301-350.

Wrigley, S. N., & Brown, G. J. (2001). A neural oscillator model of auditory attention. *Lecture Notes in Computer Science*, **2130**, 1163-1170.

Yen, S.-C., & Finkel, L. H. (1998). Extraction of perceptually salient contours by striate cortical networks. *Vision Research*, **38**, 719-741.

Zhao, L., & Macau, E. E. N. (2001). A network of dynamically coupled chaotic maps for scene segmentation. *IEEE Transactions on Neural Networks*, **12**, 1375-1385.