Boston University

OpenBU **http://open.bu.edu**

Cognitive & Neural Systems CAS/CNS Technical Reports

1993-01

Neural Dynamics of Motion Perception, Recognition Learning, and Spatial Attention

https://hdl.handle.net/2144/1978

Downloaded from OpenBU. Boston University's institutional repository.

NEURAL DYNAMICS OF MOTION PERCEPTION, RECOGNITION LEARNING, AND SPATIAL ATTENTION

Stephen Grossberg

January 1993

Revised: October 1993

Revised: January 1994

Technical Report CAS/CNS–93–001

To appear in

Mind as Motion: Dynamics, Behavior, and Cognition

R.F. Port and T. van Gelder (Eds.)

Cambridge, MA: MIT Press, 1995

Permission to copy without fee all or part of this material is granted provided that: 1. the copies are not made or distributed for direct commercial advantage, 2. the report title, author, document number, and release date appear, and notice is given that copying is by permission of the BOSTON UNIVERSITY CENTER FOR ADAPTIVE SYSTEMS AND DEPARTMENT OF COGNITIVE AND NEURAL SYSTEMS. To copy otherwise, or to republish, requires a fee and/or special permission.

Copyright © 1994

Boston University Center for Adaptive Systems and
Department of Cognitive and Neural Systems
111 Cummington Street
Boston, MA 02215

NEURAL DYNAMICS OF MOTION PERCEPTION, RECOGNITION LEARNING, AND SPATIAL ATTENTION

Stephen Grossberg†
Center for Adaptive Systems
and
Department of Cognitive and Neural Systems
Boston University
111 Cummington Street
Boston, MA 02215

To appear in
Mind as Motion: Dynamics, Behavior, and Cognition
R.F. Port and T. van Gelder (Eds.)
Cambridge, MA: MIT Press, 1993

January, 1993

Revised: October, 1993

Revised: January, 1994

Technical Report CAS/CNS-TR-93-01
Boston, MA: Boston University

---

† Supported in part by ARPA (ONR N00014-92-J-4015), the National Science Foundation (NSF IRI-90-24877), and the Office of Naval Research (ONR N00014-91-J-4100).

Acknowledgements: The author wishes to thank Cynthia E. Bradford and Diana J. Meyers for their valuable assistance in the preparation of the manuscript.

1. Introduction

1.1 Motion Perception, Recognition Learning, and What-and-Where Attention

Our brains are designed to control behaviors that are capable of interacting successfully with fluctuating environments whose rules may change unexpectedly through time. They are *self-organizing* systems whereby behaviors may be performed autonomously and adaptively to environmental changes during which no teacher other than the environmental events themselves may be present with correct new answers. The present chapter describes two examples of how the brain may achieve autonomous control in a rapidly changing environment. One example concerns motion perception and object tracking. The other concerns recognition learning, categorization, memory search, and recall. Both examples include dynamical processes which may control attention during cognitive information processing. One process suggests how attention may be used to track *where* objects are moving in space. The other process suggests how attention may delimit *what* the defining features of an object may be. These results thus contribute to an analysis of the What cortical stream, which includes area V4 of visual cortex and temporal cortex, and the Where processing stream, which includes area MT of visual cortex and parietal cortex, that have been the subject of much recent investigation (Desimone and Ungerleider, 1989; Goodale and Milner, 1992; Ungerleider and Mishkin, 1982; Wise and Desimone, 1988).

1.2 The Whole is Greater than the Sum of its Parts

How can effective models of such complex self-organizing brain processes be derived, given that no one type of behavioral or brain data can typically characterize its generative neural mechanisms? The several answers to this question each imply that "the whole is greater than the sum of its parts" when interdisciplinary data and modelling constraints are consistently joined. Even the constraint that the model be self-organizing—namely, that it can autonomously and adaptively respond in real time to its intended range of environmental

challenges—imposes many constraints on system design that are not obvious from a bottom-up analysis of brain data. Modelling self-organizing perception and recognition learning systems requires that several levels of processing, from the behavioral level through the neural system, circuit, cell, and channel levels, be computationally integrated. This is true because such a system uses internal representations that need to achieve behavioral success despite the inability of individual neurons to discern the behavioral meaning of these representations. How are coding errors corrected, or appropriate adaptations to a changing environment effected, if individual neurons do not know that these errors or changes have even occurred? It is often the case that behavioral success can be computed on the level of networks of neurons. That is why neural network models can clarify how properly designed neurons, when embedded in properly designed neural circuits and systems, can autonomously control behavior in a manner that leads to behavioral success.

For example, it is suggested below how properties of variable-speed object tracking and memory consolidation of recognition categories may emerge from system-wide interactions. The computational linkage of multiple organizational levels also leads to new predictions. In particular, properties of preattentive apparent motion processing are linked below to properties of attentive object tracking. It is also suggested how novelty-sensitive processes within the hippocampal formation may modulate the size, shape, and number of recognition categories that are learned by the inferotemporal cortex.

Granted that the emergent properties that have behavioral meaning are typically not properties of single neurons or other individual neuronal components, we can better understand why behavioral and brain processes are so hard to understand. Whereas correctly designed individual neurons are necessary in such a model, they are not sufficient. A multilevel modelling synthesis is needed in which individual components, their intercellular interactions, and their behaviorally significant emergent properties are all crafted together.

Such a multilevel analysis achieves much of its power by focusing on a natural subset of interdisciplinary data and issues—on a "vertical slice" through the space of phenomena. One

never tries to solve "all" the problems at once. In the present instance, these data and issues concern preattentive motion processing, attentive recognition learning, and attentive object tracking. We do not analyse such equally important processes as form and color perception, reinforcement learning, cognitive-emotional interactions, working memory, temporal planning, and adaptive sensory-motor control. On the other hand, larger model systems that integrate aspects of all these processes have been proposed as part of a continuing modelling cycle (Carpenter and Grossberg, 1991; Grossberg, 1982, 1987a, 1987b, 1988, 1993; Grossberg and Kuperstein, 1986, 1989). This cycle has progressively characterized individual modules, and fit them together into larger systems. System constraints that are discovered during this fitting process are used, in turn, to further shape the design of individual modules. The puzzle cannot be finished unless each piece is designed to fit.

These modules are designed to be the minimal models that can explain a targeted data base. They are lumped representations of neural processes in which no process is included unless its functional role is required and clearly understood. The insistence upon functional clarity highlights those data that the model should and should not be able to explain, facilitates the discovery of additional neural processes to explain additional data, and clarifies which species-specific variations of the minimal models are workable and which are not. These discoveries have, in the past, led to the progressive unlumping of the models as they embody ever-more-powerful functional competences for explaining ever-more-encompassing data bases.

2. Modelling Apparent Motion

2.1 The Ecological Significance of Apparent Motion

The first model provides a particularly simple example of emergent properties that are due to dynamically interacting network cells. The example seems simple after you see it, but the data properties that led to its discovery are highly paradoxical and have been known and puzzled about for many years.

These data concern phenomena about apparent motion. One might at once complain that apparent motion phenomena are of no ecological interest. To challenge this impression, consider the task of rapidly detecting a leopard leaping from a jungle branch under a sun-dappled forest canopy. Consider how spots on the leopard's coat move as its limbs and muscles surge. Imagine how patterns of light and shade play upon the leopard's coat as it leaps through the air. These luminance and color contours move across the leopard's body in a variety of directions that do not necessarily point in the direction of the leopard's leap. Indeed, the leopard's body generates a scintillating mosaic of moving contours that could easily prevent its detection. Our perceptual processes can actively reorganize such a scintillating mosaic into a coherent object percept with a unitary direction-of-motion. The leopard as a whole then seems to quickly "pop out" from the jungle background and to draw our attention. Such a perceptual process clearly has a high survival value for animals who possess it.

This description of the leaping leopard emphasizes that the process of motion perception is an active one. It is capable of transforming a motion signal that is generated by a luminance contour into a different motion percept. In this sense, our percepts of moving objects are often percepts of apparent motion, albeit an adaptive and useful form of apparent motion. The task of understanding how we see "real" motion thus requires that we also understand "apparent" motion.

The simplest examples of apparent motion were documented in the 1870's, when Exner provided the first empirical evidence that the visual perception of motion was a distinct perceptual quality, rather than being merely a series of spatially displaced static percepts over time. He did this by placing two sources of electrical sparks close together in space. When the sparks were flashed with an appropriate temporal interval between them, observers reported a compelling percept of continuous motion of a single flash from one location to another, even though neither flash actually moved. At shorter temporal intervals, flashes look simultaneous and stationary. At longer intervals, they look like successive stationary flashes,

with no intervening motion percept. When the spatiotemporal parameters of the display are suboptimal, a "figureless" or "objectless" motion called *phi motion* is perceived, wherein a sense of motion without a clearly defined form is perceived. A smooth and continuous motion of a perceptually well-defined form is called *beta motion*, and typically occurs at a larger interstimulus interval, or ISI, between the offset of one flash and the onset of the next flash.

This classical demonstration of apparent motion was followed by a series of remarkable discoveries, particularly by gestalt psychologists, concerning the properties of motion perception. It was noticed that a decrease in ISI causes the speed of the interpolating motion to increase (Kolers, 1972). A motion percept can also smoothly interpolate flashes separated by different distances, speeding up if necessary to cross a longer distance at a fixed ISI. If a more intense flash follows a less intense flash, the perceived motion can travel backwards from the second flash to the first flash. This percept is called *delta motion* (Kolers, 1972; Korte, 1915). *Gamma motion* is the apparent expansion at the onset of a single flash, or its contraction at its offset (Bartley, 1941; Kolers, 1972). A similar expansion-then-contraction may be perceived when a region is suddenly darkened relative to its background, and then restored to the background luminance.

If a white spot on a gray background is followed by a nearby black spot on a gray background, then motion between the spots can occur while the percept changes from white to black at an intermediate position. Likewise, a red spot followed by a green spot on a white background leads to a continuous motion percept combined with a binary switch from red to green along the motion pathway (Kolers and von Grünau, 1975; Squires, 1931; van der Waals and Roelofs, 1930, 1931; Wertheimer, 1912/1961). These results show that the motion mechanism can combine visual stimuli corresponding to different colors, or even opposite directions-of-contrast. Complex tradeoffs between flash luminance, duration, distance, and ISI in the generation of motion percepts were also discovered. For example, the minimum ISI for perceiving motion increases with increasing spatial separation of the inducing flashes.

This property is sometimes called Korte's Third Law (Boring, 1950; Kolers, 1972; Korte, 1915). A similar threshold decrease with distance occurs in the minimum stimulus onset asynchrony, or SOA, which is the difference between the flash onset times. Interestingly, whereas the minimum ISI decreases with flash duration, the minimum SOA increases with flash duration.

These discoveries raised perplexing issues concerning the nature of the long-range brain interaction that generates a continuous motion percept between two stationary flashes. Why is this long-range interaction not perceived when only a single light is flashed? In particular, why are not outward waves of motion signals induced by a single flash? How does a motion signal get generated from the location of the first flash after the first flash terminates, and only after the second flash turns on? How does the motion signal adapt itself to the variable distances and ISIs of the second flash, by speeding up or slowing down accordingly? In particular, how can the motion signal adapt to the ISI between two flashes even though such adaptation can only begin after the first flash is over? I like to call this the ESP Problem. Moreover, what ecologically useful function do these curious properties realize under more normal perceptual conditions?

The figural organization of motion stimuli can also influence motion percepts. The Ternus displays provide a classical example (Ternus, 1926/1950). In Frame 1 of a Ternus display, three white elements are placed in a horizontal row on a black background (or conversely). After an ISI, in Frame 2 all three elements are shifted to the right so that the two rightward elements in Frame 1 are in the same locations as the two leftward elements in Frame 2. Depending on the ISI, the observer perceives either of four percepts. At very short ISIs, all four elements appear simultaneous. At long ISIs, observers do not perceive motion at all. At ISIs slightly longer than those yielding simultaneity, the leftmost element in Frame 1 appears to jump to the rightmost element in Frame 2. This percept is called *element motion*. At somewhat longer ISIs, all three flashes seem to move together between Frame 1 and Frame 2. This is called *group motion*.

The percept of group motion might suggest that Ternus percepts are due to a cognitive process that groups the flashes into attended objects, and that motion perception occurs only after object perception. Such an explanation is not, however, made easily consistent with the percept of element motion. It has been argued, for example, that at short ISIs, the visual persistence of the brain's response to the two rightmost flashes of Frame 1 continues until the two leftmost flashes of Frame 2 occur (Braddick, 1980; Braddick and Adlard, 1978; Breitmeyer and Ritter, 1986; Pantle and Petersik, 1980). As a result, nothing changes at these two flash locations when Frame 2 occurs, so they do not seem to move. This type of explanation suggests that at least part of the apparent motion percept is determined at early processing stages. It does not, however, explain how we see element motion. In particular, why does not the element motion percept collide with the two stationary flash percepts? What kind of perceptual space can carry element motion across, or over, the stationary flashes?

Reverse-contrast Ternus motion also suggests that motion properties may be determined at early processing stages. In this paradigm, three white spots on a gray background in Frame 1 are followed by three black spots on a gray background in Frame 2 (see Figure 1). At the ISIs where element motion previously occurred, group motion now occurs (Pantle and Picciano, 1976). How does a change of contrast between Frame 1 and Frame 2 obliterate element motion? Does it do so by altering the effects of visual persistence on Frame 2?

Figure 1

A unified answer to all of these questions has recently been developed in a neural model of motion segmentation that clarifies the functional significance of many apparent motion percepts (Grossberg, 1991; Grossberg and Mingolla, 1993; Grossberg and Rudd, 1989, 1992). Perhaps the simplest such model is schematized in Figure 2. It is called a Motion Oriented Contrast-Sensitive Filter, or MOC Filter. The entire model of motion segmentation consists of multiple copies of the MOC Filter, each corresponding to a different range of receptive

field sizes, and each of which inputs to a grouping, or binding, network that is called the Motion Oriented Cooperative Competitive Loop, or MOCC Loop. Taken together, these MOC Filters and MOCC Loops are called the Motion Boundary Contour System, or Motion BCS.

The Motion BCS is designed to possess the minimal number of processing stages that are capable of tracking an object's direction-of-motion independent of whether the object's several parts are darker than or lighter than the background upon which they are moving. Grossberg and Rudd (1992) showed that each of the MOC Filter's processing stages is needed to explain the full corpus of data about beta, gamma, delta, Ternus, and related types of motion. The model's dynamics thereby illustrate how seemingly paradoxical apparent motion data may be explained as emergent properties of ecologically simple design constraints on the tracking of real moving objects.

Figure 2

2.2 Variable Speed Apparent Motion

In this chapter, I will focus on one key process of the MOC Filter; namely, how "large variations in distance are accommodated within a near-constant amount of time" (Kolers, 1972, p. 25). The mechanism that achieves this is posited to exist between Levels 4 and 5 in Figure 2. It is a surprisingly simple mechanism and utilizes components that are generally familiar to psychologists: a Gaussian filter followed by contrast enhancement due to lateral inhibition. Remarkably, in response to temporally successive inputs to the Gaussian filter, a travelling wave can be generated from the first input location to the second input location, and the peak of this wave can be contrast-enhanced by lateral inhibition to generate a focal activation that speeds up or slows down with increases or decreases of distance or ISI just as in the data.

2.3 G-Waves for Long-Range Apparent Motion

How are long-range apparent motion signals generated in such a model? Figure 3 schematizes how a flash at Level 1 (Figure 3a) leads to a focal activation at Level 5 (Figure 3c) after it activates the long-range Gaussian filter that joins Level 4 to Level 5 (Figure 3b). The broad Gaussian activation of Level 5 is sharpened into a focal activation by lateral inhibition, or competition, among the Level 5 cells.

Figure 3

Figure 4 shows how this input activation looks in time. The input to Level 1 (Figure 4a) generates a slowly decaying temporal trace (Figure 4b) that has been called "visual inertia" by Anstis and Ramachandran (1987). When this trace is fed through the Gaussian filter, it generates a spatially distributed input to Level 5 that waxes and wanes through time, without spreading across space (Figure 4c). The maximum value of this input does not move. Hence a single flash does not cause a movement across space.

Figure 4

Suppose, however, that two locations both input through the same Gaussian receptive field, and that the activation in response to a flash at the first location is decaying while activation is growing in response to a flash at the second location (Figure 5). Under these circumstances, the *total* input to Level 5 from both flashes is the sum of a temporally waning Gaussian plus a temporally waxing Gaussian, as in Figure 6. Under appropriate conditions, this sum represents a wave whose maximum travels continuously in time from the location of the first flash to the location of the second flash.

Figure 5

In summary, the time- and space-averaged responses to individual flashes do not change their positions of maximal activation through time. In this sense, nothing moves. When a series of properly timed and spaced flashes is presented, however, the sum of the temporally

and spatially averaged responses that they generate can produce a continuously moving peak of activity between the positions of the stroboscopic flashes. This is an emergent property of network dynamics, rather than a property of any cell acting alone.

Figure 6

2.4 Motion Speed-Up and Multiscale Coherence

This Gaussian wave, called a G-wave, was discovered and mathematically analysed in Grossberg (1977). These results waited twelve years for publication in Grossberg and Rudd (1989) because it took that long to understand how a long-range Gaussian filter fit into a larger theory of motion perception, such as the Motion BCS, that also includes a role for transient cells and short-range spatial interactions. A G-wave occurs whenever waxing and waning activation traces interact via a spatial Gaussian kernel under appropriate spatiotemporal conditions. The properties of a G-wave correspond closely to properties of long-range apparent motion, including the remarkable properties whereby an apparent motion percept can speed up when the ISI is decreased at a fixed interflash distance, or when the ISI is held constant and the interflash distance is increased.

The basic mathematical framework for proving these properties is very simple. Let flashes occur at positions $i = 0$ and $i = L$. Suppose that

$$\frac{dx_0}{dt} = -Ax_0 + J_0 \tag{1}$$

defines the activity x_0 and input J_0 at position 0, and

$$\frac{dx_L}{dt} = -Ax_L + J_L, \tag{2}$$

does the same at position L, where $x_0(0) = x_L(0) = 0$. Then

$$x_0(t) = \int_0^t e^{-A(t-v)} J_0(v) dv \tag{3}$$

and

$$x_L(t) = \int_0^t e^{-A(t-v)} J_L(v) dv. \tag{4}$$

Let the inputs J_0 and J_L switch on to the constant value J at times 0 and $T+I$ for duration T, as in

$$J_0(t) = \begin{cases} J & \text{if } 0 \le t \le T \\ 0 & \text{if } T < t \end{cases} \tag{5}$$

and

$$J_L(t) = \begin{cases} J & \text{if } T+I \le t \le 2T+I \\ 0 & \text{if } 2T+I < t \end{cases} \tag{6}$$

where I is the ISI between the flashes. Then for $T+I \le t \le 2T+I$,

$$x_0(t) = \frac{J}{A}(1-e^{-AT})e^{-A(t-T)} \tag{7}$$

and

$$x_L(t) = \frac{J}{A}(1-e^{-A(t-T-I)}). \tag{8}$$

Let $x_0(t)$ and $x_L(t)$ interact via a long-range Gaussian filter

$$G_{ji} = \exp[-(j-i)^2/2K^2] \tag{9}$$

as in Figure 2. For simplicity, replace index i by a continuum of cells at positions w in Level 5. Then the total input to position w of Level 5 is

$$T(w,t) = x_0(t)\exp[\frac{-w^2}{2K^2}] + x_L(t)\exp[\frac{-(w-L)^2}{2K^2}]. \tag{10}$$

By (7) and (8),

$$T(w,t) = \frac{J}{A}\Big[(1-e^{-AT})e^{-A(t-T)}\exp[\frac{w^2}{2K^2}] + (1-e^{-A(t-T-I)})\exp[\frac{(w-L)^2}{2K^2}]\Big]. \tag{11}$$

The main result shows under what combinations of parameters the maximum value of $T(w,t)$ moves continuously from position $w=0$ towards position $w=L$ through time. It also characterizes the maximum flash separation that can generate a G-wave in response to a Gaussian with size parameter K in (9).

Theorem 1 (Apparent Motion)

The maximum of $T(w,t)$ moves continuously from position $w = 0$ to position $w = L$ if and only if

$$L < 2K. \tag{12}$$

Proof: The maximum values of $T(w,t)$ occur only at locations $w = w(t)$ such that

$$\frac{\partial T(w,t)}{\partial w} = 0. \tag{13}$$

By (11), such locations obey the equation

$$\frac{e^{A(t-T)} - e^{At}}{1 - e^{-AT}} = \frac{w}{L-w} \exp\Big[\frac{L(L-2w)}{2K^2}\Big]. \tag{14}$$

The function

$$f(t) = \frac{e^{A(t-T)} - e^{At}}{1 - e^{-AT}} \tag{15}$$

is an increasing function of t. We wish to determine when the positions $w = w(t)$ at which $T(w,t)$ is maximal increase as a function of t. In order for this to happen, the right hand side of (14), namely function

$$g(w) = \frac{w}{L-w} \exp\Big[\frac{L(L-2w)}{2K^2}\Big], \tag{16}$$

must also be an increasing function of w, for all $0 \le w \le L$, since then we can solve for

$$w = g^{-1}\Big(f(t)\Big) \tag{17}$$

as an increasing function of w for all $0 \le w \le L$.

Function $g(w)$ is monotone increasing if $g'(w) > 0$, which holds if and only if function

$$h(w) \equiv (L-w)[1 - \frac{Lw}{K^2}] + w \tag{18}$$

satisfies

$$h(w) > 0. \tag{19}$$

In order for (17) to hold for all $0 \leq w \leq L$, the minimum of $h(w)$ for $0 \leq w \leq L$ must be positive. The minimum of $h(w)$ occurs at $w = \frac{L}{2}$, and equals

$$h(\frac{L}{2}) = \frac{L}{2}(2 - \frac{L^2}{2K^2}). \tag{20}$$

The number $h(\frac{L}{2})$ is positive if (12) holds.

The next result proves that the apparent motion signal reaches the position $w = \frac{L}{2}$ midway between positions $w = 0$ and $w = L$ at a time $t_{\frac{1}{2}}$ that is independent of L and K. Independence of L illustrates how the wave speeds up to travel over larger interflash distances.

Theorem 2 (Equal Half-Time Property)

The time at which the motion signal reaches position $w = \frac{L}{2}$ is

$$t_{\frac{1}{2}} = T + \frac{1}{A} ln\Big[e^{AI} + (1 - e^{-AT})\Big]. \tag{21}$$

Proof: By (17), we need to compute $t = f^{-1}(g(w))$ when $w = \frac{L}{2}$, namely

$$t_{\frac{1}{2}} = f^{-1}\Big(g(\frac{L}{2})\Big). \tag{22}$$

By (16),

$$g(\frac{L}{2}) = 1. \tag{23}$$

Equation (21) follows immediately from (23) and (14).

Remarkably, $t_{1/2}$ in (21) also does not depend upon the width K of the Gaussian filter, just so long as the filter is wide enough to support a travelling wave. This means that the speed-up property, which seems so mysterious in itself, also achieves an ecologically useful property; namely, the ability of multiple spatial scales in the motion perception system to generate G-waves that are all spatially coincident (Figure 7). Because of this property, a coherent motion percept may be synthesized from data from all the spatial scales that are activated by the stimulus.

Figure 7

2.5 Spatial Attention Shifts and Target Tracking by the Where Cortical Stream

Another no less useful ecological property of motion speed-up is suggested by the fact that rapidly moving objects may be perceived only intermittently. From this perspective, I suggest that a G-wave may give rise to certain spatial shifts in attention, such as those reported by Ericksen and Murphy (1987), Kwak, Dagenbach, and Egeth (1991), LaBerge and Brown (1989), and Remington and Pierce (1984). For example, if a targeted predator or prey is rapidly moving across a scene, perhaps darting behind protective cover, then an animal may be able to see the target only intermittently. A G-wave can interpolate these temporally discrete views with a continuous motion signal that adapts its speed to the varying speed of the target. Such a continuous motion signal may be used to predict the location and speed of the target, and to command motor responses accordingly. The results of Kwak, Dagenbach, and Egeth (1991) and of Remington and Pierce (1984) are of particular interest, since they report a speed-up of spatial attention to cover variable distances in equal time.

In those cases where motion mechanisms contribute to spatial attention shifts, it needs to be kept in mind that a spatially continuous motion signal is generated only under certain spatiotemporal conditions, the speed of the motion signal is nonuniform in time (see Grossberg and Rudd, 1992), and spatially discrete jumps in activation may occur in cases where continuous motion is not observed; for example, if $L > 2K$ in (12). These properties may help to disentangle some of the apparently conflicting views about how fast attention shifts and whether it does so continuously or discretely.

In thinking about these possibilities, the reader might wonder how a continuous motion signal could be interpolated behind occluding objects in such a way that it is not seen. Two themes need to be developed to understand how this might happen. First, the theory predicts that a boundary segmentation, whether static or moving, is perceptually invisible

within the parvocellular interstripe and magnocellular processing streams of the visual cortex wherein they are predicted to be computed. I like to say that "all boundaries are invisible". Visibility is predicted to be a property of the parallel parvocellular blob cortical stream (Figure 8). Here boundary segmentations define the domains within which visible properties of brightness, color, and depth fill-in surface representations. (See Grossberg and Mingolla (1993) and Grossberg, Mingolla, and Todorović (1989) for a discussion of how this is predicted to happen.) In addition, one needs to analyse how a boundary segmentation, whether static or moving can be completed "behind" an occluding object in such a way that it can influence object recognition without being seen. Examples of such occluded boundary completions are discussed in Grossberg (1994). Bregman (1990, p.23) has also commented upon the possible utility of a motion signal that can interpolate intermittently viewed moving objects. The present theory suggests a dynamical explanation of how this can happen in the brain.

3. Modelling Recognition Learning and Categorization

3.1 Spatial Attention, Featural Attention, and Perceptual Binding in the What and Where Cortical Streams

The hypothesis that a G-wave may give rise to a spatial attention shift is consistent with the fact that the motion perception, or magnocellular, cortical processing stream is part of a larger Where processing stream that includes cortical region MT as well as parietal cortex (Figure 8). The Where processing stream computes the locations of targets with respect to an observer and helps to direct attention and action towards them. In contrast, the form perception, or parvocellular, cortical processing stream is part of a larger What processing stream that includes region V4 as well as inferotemporal cortex (Figure 8). The What processing stream is used to recognize targets based upon prior learning.

Figure 8

The second model of this chapter contributes to the understanding of how humans and

other animals can rapidly learn to recognize and categorize objects in their environments under autonomous learning conditions in real time. Here again, attention plays a key role. It does not predict a spatial location of a target. Rather, it amplifies and binds together feature combinations that are used to categorize environmental events into object representations.

3.2 The Stability-Plasticity Dilemma

An adequate self-organizing recognition system must be capable of *plasticity* in order to rapidly learn about significant new events, yet its memory must also remain *stable* in response to irrelevant or often repeated events. For example, how do we rapidly learn to recognize new faces without risking the unselective forgetting of our parents' faces? In order to prevent the unselective forgetting of its learned codes by the "blooming, buzzing confusion" of irrelevant experience, a self-organizing recognition system must be sensitive to *novelty*. It needs to be capable of distinguishing between familiar and unfamiliar events, as well as between expected and unexpected events.

A class of neural models, called Adaptive Resonance Theory, or ART, models was introduced in 1976 to help understand how this is accomplished (Grossberg, 1976a, 1976b). In ART, dynamical interactions between an attentional subsystem and an orienting subsystem, or novelty detector, self-stabilize learning, without an external teacher, as the network familiarizes itself with an environment by categorizing the information within it in a way that predicts behaviorally successful outcomes (Carpenter and Grossberg, 1991; Grossberg, 1980). ART models combine several types of processes that have been demonstrated in cognitive and neurobiological experiments, but not otherwise synthesized into a model system. Table 1 lists some of the cognitive processes that are joined together in a consistent computational format within ART systems. This synthesis illustrates that learning and information processing mechanisms need to coevolve in order to achieve behaviorally useful properties. It also clarifies how higher-order cognitive processes, such as hypothesis testing of learned top-down expectations, control such apparently lower-order processes as the learning

of bottom-up recognition categories. That is why an analysis of *recognition* needs also to be framed as an analysis of *learning*.

Table 1

3.3 Competitive Learning and Self-Organizing Feature Maps

All the learning goes on in the attentional subsystem. Its processes include activation of short term memory (STM) traces, incorporation through learning of momentary STM information into a longer-lasting long term memory (LTM) traces, and interactions between pathways that carry specific types of information with nonspecific pathways that modulate the specific pathways in several different ways. These interactions between specific STM and LTM processes and nonspecific modulatory processes regulate the stability-plasticity balance during normal learning.

The attentional subsystem undergoes both bottom-up learning and top-down learning between the processing levels denoted by F_1 and F_2 in Figure 9. Level F_1 contains a network of nodes, or cell populations, each of which represents a particular combination of sensory features.

Level F_2 contains a network of nodes that represent recognition codes, or categories, which are selectively activated by the patterns of activation across F_1. Each node in F_1 sends output signals to a subset of nodes in F_2. Each node in F_2 thus receives inputs from many F_1 nodes. The thick arrow from F_1 to F_2 in Figure 9a represents in a concise way the array of diverging and converging pathways shown in Figure 9b. Learning takes place at the synapses denoted by semicircular endings in the $F_1 \rightarrow F_2$ pathways. Pathways that end in arrowheads do not undergo learning. This bottom-up learning enables F_2 nodes to become selectively tuned to particular combinations of activation patterns across F_1 by changing their LTM traces. This basic property of recognition learning is mathematically proved below.

Figure 9

Why does not bottom-up learning suffice? An analysis of this problem was carried out in a type of model – called a self-organizing feature map, competitive learning, or learned vector quantization – that forms part of a larger ART system. Such a bottom-up learning model shows how to combine associative learning and lateral inhibition for purposes of learned categorization. As shown in Figure 10a, an input pattern is normalized and registered as a pattern of activity, or STM, across the feature detectors of level F_1. Each F_1 output signal is multiplied or gated, by the adaptive weight, or LTM trace, in its respective pathway. All these LTM-gated inputs are added up at their target F_2 nodes. Lateral inhibitory, or competitive, interactions across the F_2 nodes contrast-enhance this input pattern. Whereas many F_2 nodes may receive inputs from F_1, lateral inhibition allows a much smaller set of F_2 nodes to store their activation in STM.

Figure 10

Only the F_2 nodes that win the competition and store their activity in STM can influence the learning process. STM activity opens a learning gate at the LTM traces that abut the winning nodes. These LTM traces can then approach, or track, the input signals in their pathways by a process of steepest descent. This learning law is thus often called *gated steepest descent*, or *instar learning*. It was introduced into neural network models in the 1960's (Grossberg, 1969) and is the learning law that was used to introduce ART (Grossberg, 1976a, 1976b). In particular, let x_{1i} denote the STM activity of the ith F_1 node, x_{2j} the STM activity of the jth F_2 node, and z_{ij} the adaptive weight or LTM trace in the bottom-up pathway from node i in F_1 to node j in F_2. Then the rate of change through time of z_{ij}, denoted by $\frac{d}{dt}z_{ij}$, obeys an equation of the form

$$\frac{d}{dt}z_{ij} = f(x_{2j})[-z_{ij} + g(x_{1i})], \tag{24}$$

where f and g are non-negative signal functions. Note that if $f(x_{2j}) = 0$, then $\frac{d}{dt}z_{ij} = 0$. Thus no learning occurs if the gate $f(x_{2j})$ is closed. This can occur either if no inputs perturb F_1 or if node j loses the competition across F_2. If $f(x_{2j}) > 0$, then z_{ij} increases if

$g(x_{1i}) > z_{ij}$ and decreases if $g(x_{1i}) < z_{ij}$.

Such an LTM trace z_{ij} can increase or decrease to track the signal $g(x_{1i})$ in its pathway. It is thus not a Hebbian associative law, which can only increase during a learning episode. Because the adaptive weight z_{ij} can either increase or decrease is size, the same law (24) can control both long term potentiation (LTP) and long term depression (LTD). Equation (24) has been used to model neurophysiological data about hippocampal LTP (Levy, 1985; Levy and Desmond, 1985) and adaptive tuning of cortical feature detectors during the visual critical period (Rauschecker and Singer, 1979; Singer, 1983), lending support to ART predictions that both systems would employ such a learning law (Grossberg, 1976a, 1976b).

Self-organizing feature map models were introduced and characterized computationally in Grossberg (1972, 1976a, 1976b, 1978), Malsburg (1973), and Willshaw and Malsburg (1976). These models were subsequently applied and further developed by many authors (Amari and Takeuchi, 1978; Bienenstock, Cooper, and Munro, 1982; Cohen and Grossberg, 1987; Grossberg, 1982, 1987a, 1987b; Grossberg and Kuperstein, 1986; Kohonen, 1984; Linsker, 1986; Rumelhart and Zipser, 1985). They exhibit many useful properties, especially if not too many input patterns, or clusters of input patterns, perturb level F_1 relative to the number of categorizing nodes in level F_2. It was shown that under these sparse environmental conditions, category learning is stable; the LTM traces track the statistics of the environment, are self-normalizing, and oscillate a minimum number of times (Grossberg, 1976a, 1976b, 1978). In addition, it was observed that the category selection rule tends to minimize error, as in a Bayesian classifier. These are the basic properties that have been used in all subsequent applications.

It was also proved, however, that under arbitrary environmental conditions, learning becomes unstable. If our own learned categorizations exhibited this property, we could forget our parents' faces. Although a gradual switching off of plasticity can partially overcome this problem, such a mechanism cannot work in a recognition learning system whose plasticity needs to be maintained throughout adulthood. This memory instability is due to basic

properties of associative learning and lateral inhibition. ART models were introduced to incorporate self-organizing feature maps in such a way as to stabilize their learning in response to an arbitrary stream of input patterns.

3.4 Feature Binding and Attentional Focusing

In an ART model (Carpenter and Grossberg, 1987, 1991), learning does not occur as soon as some winning F_2 activities are stored in STM. Instead activation of F_2 nodes may be interpreted as "making a hypothesis" about an input at F_1. When F_2 is activated, it quickly generates an output pattern that is transmitted along the top-down adaptive pathways from F_2 to F_1 . These top-down signals are multiplied in their respective pathways by LTM traces at the semicircular synaptic knobs of Figure 10b. The LTM-gated signals from all the active F_2 nodes are added to generate the total top-down feedback pattern from F_2 to F_1. This pattern plays the role of a learned expectation. Activation of this expectation "tests the hypothesis," or "reads out the prototype," of the active F_2 category. As shown in Figure 10b, ART networks are designed to match the "expected prototype" of the category against the bottom-up input pattern, or exemplar, to F_1. Nodes that are activated by this exemplar are suppressed if they do not correspond to large LTM traces in the top-down prototype pattern. The matched F_1 pattern encodes the cluster of input features that are relevant to the hypothesis based upon the network's past experience. This resultant activity pattern, called $\mathbf{X}^*$ in Figure 10b, encodes the pattern of features to which the network starts to "pay attention."

If the expectation is close enough to the input exemplar, then a state of *resonance* develops as the attentional focus takes hold. The pattern $\mathbf{X}^*$ of attended features reactivates the F_2 category $\mathbf{Y}$ which, in turn, reactivates $\mathbf{X}^*$. The network locks into a resonant state through a positive feedback loop that dynamically links, or binds, $\mathbf{X}^*$ with $\mathbf{Y}$. Damasio (1989) has used the term "convergence zones" to describe such a resonant process. The resonance binds spatially distributed features into either a stable equilibrium or a synchronous

oscillation (Eckhorn and Schanze, 1991; Grossberg and Somers, 1991, 1992) with properties much like synchronous feature binding in visual cortex (Eckhorn *et al.*, 1988; Gray and Singer, 1989; Gray *et al.*, 1989).

In ART, the resonant state, rather than bottom-up activation, drives the learning process. The resonant state persists long enough, at a high enough activity level, to activate the slower learning process; hence the term *adaptive resonance* theory. The resonance process shows how dynamical properties, such as differences in the faster STM rates and slower LTM rates, are exploited by the system as a whole. Fast information processing in STM is altered by previously learned LTM traces, even if the LTM traces do not undergo new learning due to the STM patterns that they help to create. When an STM resonance is maintained through a feedback exchange of bottom-up and top-down signals, it lasts long enough for the slower LTM traces to respond to the resonating STM activities and to undergo new learning. In effect, the resonance embodies a global system-wide consensus that the resonating STM patterns are worth learning about.

ART systems learn prototypes, rather than exemplars, because the attended feature vector $\mathbf{X}^*$, rather than the input exemplar itself, is learned. These prototypes may, however, also be used to encode individual exemplars. How the matching process achieves this is described below. If the mismatch between bottom-up and top-down information is too great, then resonance cannot develop. Instead the F_2 category is quickly reset before erroneous learning can occur, and a bout of hypothesis testing, or memory search, is initiated to discover a better category. This combination of top-down matching, attention focusing, and memory search is what stabilizes ART learning and memory in an arbitrary input environment. The top-down matching process suppresses those features within an exemplar that are not expected and starts to focus attention on the features $\mathbf{X}^*$ that are shared by the exemplar and the active prototype. The memory search chooses a new category on a fast time scale, before an exemplar that is too different from the prototype can destabilize its previous learning. How these matching and search operations work will now be summarized.

3.5 Phonemic Restoration, Priming, and Consciousness

The ART attentive matching process is realized by combining bottom-up inputs and top-down expectations with a nonspecific arousal process that is called attentional gain control (Carpenter and Grossberg, 1987, 1991). An F_1 node can be fully activated only if two of the three input sources that converge on the node send positive signals to the node at a given time. This constraint is called the 2/3 Rule. A bottom-up input pattern turns on the attentional gain control channel in order to instate itself in STM at F_1 (Figure 10a). A top-down expectation turns off the attentional gain control channel (Figure 10b). As a result, only those input features that are confirmed by the top-down prototype can be attended at F_1 after an F_2 category is selected.

The 2/3 Rule enables an ART network to solve the stability-plasticity dilemma. Carpenter and Grossberg (1987) proved that ART learning and memory are stable in arbitrary environments, but become unstable when 2/3 Rule matching is eliminated. Thus the matching law that guarantees stable learning also enables the network to pay attention. This type of insight could never be derived without an analysis of the dynamics of autonomous learning in real time.

Matching by the 2/3 Rule in the brain is illustrated by experiments on "phonemic restoration" (Repp, 1991; Samuel, 1981a, 1981b; Warren, 1984; Warren and Sherman, 1974). Suppose that a noise spectrum replaces a letter sound, or phonetic segment, in a word heard in an otherwise unambiguous context. Then subjects hear the correct phonetic segment, not the noise, to the extent that the noise spectrum includes the acoustic signal of the phones. If silence replaces the noise, then only silence is heard. Top-down expectations thus amplify expected input features while suppressing unexpected features, but do not create activations not already in the input, just as in the 2/3 Rule.

The 2/3 Rule for matching also explains paradoxical reaction time and error data from priming experiments during lexical decision and letter gap detection tasks (Grossberg and

Stone, 1986; Schvaneveldt and MacDonald, 1981). Although priming is often thought of as a residual effect of previous bottom-up activation, a combination of bottom-up activation and top-down 2/3 Rule matching was needed to explain the complete data pattern. This analysis combined bottom-up priming with a type of top-down priming; namely, the top-down activation that prepares a network for an expected event that may or may not occur. The 2/3 Rule hereby clarifies why priming, by itself, is subliminal and unconscious, even though it can facilitate supraliminal processing of a subsequent expected event. Only the resonant state can support a conscious event in the model.

These examples illustrate how data from a variety of experimental paradigms can emerge from computational properties that are designed to accomplish quite different functions than the paradigm itself might disclose; in this case, fast and stable recognition learning in response to a rapidly changing environment.

3.6 Memory Search, Vigilance, and Category Generalization

The criterion of an acceptable 2/3 Rule match is defined by the model parameter ρ that is called *vigilance* (Carpenter and Grossberg, 1987, 1991). The vigilance parameter is computed in the orienting subsystem $\mathcal{A}$. Vigilance weighs how similar an input exemplar $\mathbf{I}$ must be to a top-down prototype $\mathbf{V}$ in order for resonance to occur. It does so by comparing the total amount of inhibition from the attentional focus at F_1 with the total amount of excitation from the input pattern $\mathbf{I}$ (Figure 10b). In cases where binary features are processed, the 2/3 Rule implies that the attentional focus $\mathbf{X}^*$ equals the intersection $\mathbf{I} \cap \mathbf{V}$ of the bottom-up exemplar $\mathbf{I}$ and the top-down prototype $\mathbf{V}$. Resonance occurs if $\rho|\mathbf{I}| - |\mathbf{X}^*| \leq 0$. This inequality says that the F_1 attentional focus $\mathbf{X}^*$ inhibits $\mathcal{A}$ more than the input $\mathbf{I}$ excites it. If $\mathcal{A}$ is inhibited, then a resonance has time to develop between F_1 and F_2.

Vigilance calibrates how much novelty the system can tolerate before activating $\mathcal{A}$ and searching for a different category. If the top-down expectation and the bottom-up input are too different to satisfy the resonance criterion, then hypothesis testing, or memory search, is

triggered, because the inhibition from F_1 to $\mathcal{A}$ is no longer sufficient to prevent the excitation due to $\mathbf{I}$ from activating $\mathcal{A}$. Nonspecific arousal from $\mathcal{A}$ to F_2 resets the active category at F_2 and initiates the memory search. Memory search leads to selection of a better category at level F_2 with which to represent the input features at level F_1. During search, the orienting subsystem interacts with the attentional subsystem, as in Figures 10c and 10d, to rapidly reset mismatched categories and to select other F_2 representations with which to learn about novel events, without risking unselective forgetting of previous knowledge. Search may select a familiar category if its prototype is similar enough to the input to satisfy the vigilance criterion. The prototype may then be refined by 2/3 Rule attentional focussing. If the input is too different from any previously learned prototype, then an uncommitted population of F_2 cells is rapidly selected and learning of a new category is initiated.

3.7 Supervised Learning of Many–to–One Maps from Categories to Names

Because vigilance can vary across learning trials, recognition categories capable of encoding widely differing degrees of generalization or abstraction can be learned by a single ART system. Low vigilance ρ leads to broad generalization and abstract prototypes because exemplars $\mathbf{I}$ that differ greatly from an active prototype $\mathbf{V}$ can satisfy $\rho|\mathbf{I}| - |\mathbf{X}^*| \leq 0$. High vigilance leads to narrow generalization and to prototypes that represent fewer input exemplars, even a single exemplar. Thus a single ART system may be used, say, to recognize abstract categories that encode higher-order invariants of faces and dogs, as well as individual faces and dogs. ART systems hereby provide a new answer to whether the brain learns prototypes or exemplars. Various authors have realized that neither one nor the other alternative is satisfactory, and that a hybrid system is needed (Smith, 1990).

Supervised ART, or ARTMAP systems can perform this hybrid function in a manner that is sensitive to environmental demands (Figure 11). In an ARTMAP system, predictive errors can be used to trigger searches for new categories. As a result, many categories in one modality (e.g., visual recognition categories) may be learned and associated with each

category in another modality (e.g., auditory naming categories), just as there may be many different visual fonts that all have the same name "A". A predictive error in naming increases the vigilance ρ in the visual categorization network just enough to satisfy $\rho|\mathbf{I}| - |\mathbf{X}^*| > 0$ and thereby to activate $\mathcal{A}$ and initiate a memory search for a better visual category with which to predict the desired name (Carpenter and Grossberg, 1992; Carpenter, Grossberg, and Reynolds, 1991; Carpenter *et al.*, 1992). Since low vigilance leads to learning of the most general categories, this operation, which is called *match tracking*, sacrifices the minimum amount of visual generalization on each learning trial in order to correct a naming error.

Table 2 summarizes how such a supervised ART system performs relative to other machine learning, genetic algorithm, and back propagation networks in benchmark simulations. These benchmarks indicate that models of biological learning enjoy computational advantages over more traditional approaches. Such benchmarks are described more fully in Carpenter, Grossberg, and Reynolds (1991) and Carpenter *et al.* (1992).

Table 2

3.8 Memory Consolidation as an Emergent Property of Network Dynamics

As inputs are practiced over learning trials, the search process eventually converges upon stable categories that access the corresponding category directly, without the need for search. The category that is selected is the one whose prototype provides the globally best match to the input pattern at the system's present state of knowledge. In this way, "familiar" patterns can resonate with their category without the need for search, much as Gibson (1979, p. 249) may have intended when he hypothesized that the perceptual system "resonates to the invariant structure or is attuned to it". If both familiar and unfamiliar events are experienced through time, familiar inputs can directly activate their learned categories, even while unfamiliar inputs continue to trigger adaptive memory searches for better categories, until the network's memory capacity is fully utilized (Carpenter and Grossberg, 1987, 1991).

This process whereby search is gradually and automatically disengaged may be inter-

preted as a form of memory consolidation. This type of memory consolidation is an emergent property of network interactions. It is, once again, a property that can only be understood by studying the network's dynamics. Emergent consolidation does not, however, preclude structural forms of consolidation, since persistent resonance may also be a trigger for other learning-dependent processes, such as transmitter production and protein synthesis, at individual cells.

3.9 Face Recognition and Inferotemporal Cortex

How do components of the ART model map onto brain mechanisms? To begin with, level F_2 properties may be compared with properties of cell activations in inferotemporal cortex (IT) during recognition learning in monkeys. The ability of F_2 nodes to learn categories with different levels of generalization clarifies how some IT cells can exhibit high specificity, such as selectivity to views of particular faces, while other cells respond to broader features of the animal's environment (Desimone, 1991; Desimone and Ungerleider, 1989; Gochin *et al.*, 1991; Harries and Perrett, 1991; Mishkin, 1982; Mishkin and Appenzeller, 1987; Perrett, Mistlin, and Chitty, 1987; Schwartz *et al.*, 1983; Seibert and Waxman, 1991). In addition, when monkeys are exposed to easy and difficult discriminations (Spitzer, Desimone, and Moran, 1988), "in the difficult condition the animals adopted a stricter internal criterion for discriminating matching from nonmatching stimuli ... the animals' internal representations of the stimuli were better separated, independent of the criterion used to discriminate them ... increased effort appears to cause enhancement of the responses and sharpened selectivity for attended stimuli" (pp. 339–340). These are also properties of model cells in F_2. Prototypes represent a smaller set of exemplars at higher vigilance levels, so a stricter matching criterion is learned. These exemplars match their finer prototypes better than do exemplars which match a coarser prototype. This better match more strongly activates the corresponding F_2 nodes.

Data from IT support the hypothesis that unfamiliar or unexpected stimuli nonspecif-

ically activate level F_2 via the orienting subsystem. According to Desimone (1992), "the fact that IT cortex has a reduced level of activation for familiar or expected stimuli suggests that a high level of cortical activation may itself serve as a trigger for attentional and orienting systems, causing the subject to orient to the stimulus causing the activation. This link between the mnemonic and attentional systems would 'close the loop' between the two systems, resulting in orienting behavior that is influenced by both current stimuli and prior memories. Such a mechanism has a number of similarities to the adaptive resonance theory" (p. 359). Properties of IT cells during working memory tasks suggest that active reset occurs at the end of each trial (Miller, Li, and Desimone, 1991; Riches, Wilson, and Brown, 1991). Reset of F_2 is also a key ART operation.

These recent neurophysiological data about IT cells during recognition tasks are thus reflected in level F_2 properties. Additional data suggest that the pulvinar may mediate aspects of attentional gain control (Desimone, 1992; Robinson and Peterson, 1992). Data that correlate IT and pulvinar recordings are needed to critically test this hypothesis. Carpenter and Grossberg (1993) have suggested that the orienting operations whereby vigilance is controlled may take place in the hippocampal formation. They support this hypothesis by showing how a formal lesion of the orienting system in ART creates a set of properties strikingly like symptoms of medial temporal amnesia in human patients with hippocampal lesions. This linkage suggests the prediction that operations which make the novelty-related potentials of the hippocampal formation more sensitive to input changes may trigger the formation of more selective inferotemporal recognition categories. Such a correlation may be sought, for example, when monkeys learn easy and difficult discriminations. The hypothesis also suggests that operations which block hippocampal novelty potentials may lead to the learning of coarser recognition categories, with amnesic symptoms as the limiting case when the hippocampal formation is completely inoperative.

3.10 Feature Discovery by Competitive Learning

The above properties of ART systems have been computationally demonstrated and mathematically proved in a series of articles by Gail Carpenter and myself in collaboration with several students. The core articles are brought together in Carpenter and Grossberg (1991). In the present chapter, some of the most important mathematical properties of competitive learning and self-organizing feature maps are reviewed. These properties are important both in themselves and as a stepping stone to a mathematical study of ART systems.

Perhaps the simplest competitive learning system is defined by the following equations. Let I_i be the input to the i node in F_1. Let fast competitive interactions within F_1 normalize this input. There are several possible types of normalization. In the simplest type, the normalized activity x_{1i} of the ith node, or cell population, in F_1 satisfies

$$x_{1i} = \theta_i \equiv \frac{I_i}{\Sigma_k I_k}, \tag{25}$$

so that

$$\Sigma_i x_{1i} = 1. \tag{26}$$

Property (1) is called L^1 normalization. In L^p normalization, $\Sigma_i x_{1i}^p = 1$. The effects of choosing different values of p, notably $p = 2$, were first described in Grossberg (1978). Here we analyse the case $p = 1$, as in the Grossberg (1976a) model that was also used by Rumelhart and Zipser (1985).

The normalized signals θ_i are multiplied by adaptive weights z_{ij} and added to generate the total input

$$S_j = \Sigma_i \theta_i z_{ij} \tag{27}$$

to each node j in F_2. A competition between nodes in F_2 rapidly chooses the activity x_{2j} whose total input S_j is maximal for storage in STM, while normalizing the total activity of F_2. Such a network is often said to carry out a winner-take-all (WTA) operation. How a WTA competitive network may be designed was first described in Grossberg (1973). This

fast competitive dynamical process may be approximated by the algebraic equation

$$x_{2j} = \begin{cases} 1 & \text{if } S_j > \max(\epsilon, S_k : k \neq j) \\ 0 & \text{if } S_j \leq \max(\epsilon, S_k : k \neq j) \end{cases} \tag{28}$$

where ϵ is a threshold that all inputs must exceed in order to trigger the STM choice and storage process.

Learning in the LTM traces takes place more slowly than the STM processes (25) and (28). Hence learning cannot be approximated by an algebraic equation. Rather, it obeys the instar, or gated steepest descent, differential equation (24). In the present case, this reduces to

$$\frac{d}{dt} z_{ij} = x_{2j}(-z_{ij} + \theta_i). \tag{29}$$

In order to get a sense of how competitive learning and self-organizing feature maps work, suppose that a single input pattern θ perturbs F_1 through time and activates node $j = J$ in F_2. As a result, the total inputs S_j to F_2 nodes in (27) obey the inequalities $S_J > S_k$, $k \neq J$, so $x_{2J} = 1$ and $x_{2k} = 0$, $k \neq J$, by (28). By (29), only the vector $z_J = (z_{1J}, z_{2J}, \ldots, z_{nJ})$ of LTM traces that abut node J undergo learning, because only $x_{2J} > 0$ in (29). During learning, the LTM vector z_J is attracted towards the normalized input vector $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ as each z_{iJ} is "tuned" by the ith feature activity θ_i. As learning proceeds, the Euclidean length $||z_J|| = \sqrt{\Sigma_i z_{iJ}^2}$ of the LTM vector is normalized as it approaches the length $||\theta|| = \sqrt{\Sigma_i \theta_i^2}$ of the input vector, and the total input S_J in (27) increases to its maximal possible value.

This result was proved in Grossberg (1976a) and is reviewed below. Its significance is clarified by noting that the total input S_J to node J in (27) can be rewritten as

$$S_J = \Sigma_i \theta_i z_{iJ} = ||\theta||\ ||z_J||\ cos(\theta, z_J), \tag{30}$$

because S_J is the dot product, or inner product, of the vectors θ and z_J. Given that $||z_J||$ approaches $||\theta||$ during learning, (30) shows that the maximal S_J, among all S_j, is the one whose $cos(\theta, z_J)$ is maximal. This quantity corresponds to that vector z_J which is most

parallel to θ. In other words, the STM competition (28) at F_2 chooses the node J whose vector z_J is most parallel to θ, and learning tunes z_J to become even more parallel to θ. In addition, the fact that $||z_J||$ approaches $||\theta||$ shows that choosing all input vectors so that their length $||\theta|| = 1$ eliminates the effects of variable input length on the category choice process. Then L^p normalization with $p = 2$ replaces the L^1 normalization in (26). This implication of the learning theorem was noted in Grossberg (1978) and used by Kohonen (1984) in his subsequent applications of self-organizing feature maps. Grossberg (1976a) also proved that, when multiple inputs activate F_1, then the LTM weight vectors tend to track the statistics of the input environment. This property was also exploited by Kohonen (1984), among others.

Theorem 3 (Adaptive Tuning of Feature Maps)

Given a pattern θ, suppose that there exists a unique $j = J$ such that

$$S_J(0) > \max\{\epsilon, S_k(0) : k \neq J\}. \tag{31}$$

Let θ be practiced during a sequence of nonoverlapping intervals $[U_m, V_m], m = 1, 2, \ldots$. Then the angle between $z^{(J)}(t)$ and θ monotonically decreases, the signal $S_J(t)$ is monotonically attracted towards $||\theta||^2$, and $||z^{(J)}||^2$ oscillates at most once as it pursues $S_J(t)$. In particular, if $||z^{(J)}(0)|| \leq ||\theta||$, then $S_J(t)$ is monotone increasing. Except in the trivial case that $S_J(0) = ||\theta||^2$, the limiting relations

$$\lim_{t\to\infty} ||z^{(J)}(t)||^2 = \lim_{t\to\infty} S_J(t) = ||\theta||^2 \tag{32}$$

hold if and only if

$$\sum_{m=1}^{\infty} (V_m - U_m) = \infty. \tag{33}$$

Proof of Theorem 1. Consider the case in which

$$||\theta||^2 > S_J(0) > \max\{\epsilon, S_k(0) : k \neq J\}. \tag{34}$$

The case in which $S_J(0) \geq \|\theta\|^2$ can be treated similarly. First it will be shown that if the inequalities

$$\|\theta\|^2 > S_J(t) > \max\{\epsilon, S_k(t) : k \neq J\} \tag{35}$$

hold at any time $t = T$, then they hold at all future times $t \geq T$. By (34), $x_{2J}(T) = 1$ and $x_{2k}(T) = 0$, $k \neq J$. Consequently, by (29), at any time $t = T$ during a learning interval $[U_m, V_m]$,

$$\frac{d}{dt} z_{iJ}(T) = -z_{iJ}(T) + \theta_i \tag{36}$$

and

$$\frac{d}{dt} z_{ik}(T) = 0 \tag{37}$$

for $k \neq J$ and $i = 1, 2, \ldots, n$. By (27) and (35)–(37),

$$\begin{aligned} \frac{d}{dt} S_J(T) &= \sum_i \theta_i \frac{d}{dt} z_{iJ} \\ &= \sum_i \theta_i(-z_{iJ} + \theta_i) \\ &= -S_J(T) + \|\theta\|^2. \end{aligned} \tag{38}$$

Thus by (34), from the first moments of learning onwards,

$$\frac{d}{dt} S_J(T) > 0 = \frac{d}{dt} S_k(T), \; k \neq J. \tag{39}$$

Since $S_J(t)$ continues to grow while all $S_k(t)$ remain constant, $k \neq J$, inequality (34) continues to hold for all $t \geq T$ throughout the learning process. Since all $x_{2j} = 0$ whenever no learning occurs, there is no change in any z_{iJ} or S_J during these times. Thus (34) holds for all $t \geq T$. Moreover, by (38), $S_J(t)$ converges monotonically towards $\|\theta\|^2$.

To show that $S_J(t)$ converges to $\|\theta\|^2$ only if (33) holds, integrate (38) throughout the mth learning interval $[U_m, V_m]$. Then

$$S_J(V_m) = S_J(U_m)e^{-(V_m - U_m)} + \|\theta_J\|^2(1 - e^{-(V_m - U_m)}). \tag{40}$$

Since no learning occurs between time $t = V_m$ and time $t = U_{m+1}$, $S_J(U_{m+1}) = S_J(V_m)$. Using notation $S_{Jm} = S_J(U_m)$ and $W_m = V_m - U_m$ for simplicity, (40) may thus be written as

$$S_{J,m+1} = S_{Jm}e^{-W_m} + ||\theta_J||^2(1 - e^{-W_m}). \tag{41}$$

Equation (41) can be solved recursively by setting $m = 1, 2, \ldots$ in (41) to find that

$$S_{J,m+1} = S_{Jm}e^{-\sum_{k=1}^{m} W_k} + ||\theta_J||^2(1 - e^{-\sum_{k=1}^{m} W_k}), \tag{42}$$

from which it follows that $S_J(t)$ converges to $||\theta_J||^2$ only if $\sum_{k=1}^{\infty} W_m = \infty$.

To show that S_J *increases* towards $||\theta||^2$ if

$$||z^{(J)}(0)|| \leq ||\theta||, \tag{43}$$

we need only to show that $S_J(0) \leq ||\theta||^2$, since then $\frac{d}{dt}S_J(0) \geq 0$ by (38). Were

$$S_J(0) > ||\theta||^2, \tag{44}$$

then by (30), (43), and (44),

$$||\theta||\ ||z^{(J)}(0)|| \geq S_J(0) > ||\theta||^2 \geq ||\theta||\ ||z^{(J)}(0)|| \tag{45}$$

and thus

$$||z^{(J)}(0) > ||\theta|| \geq ||z^{(J)}(0)||, \tag{46}$$

which is a contradiction.

To show that the learning process normalizes the vector of LTM traces, let us use the notation $N_J = ||z^{(J)}||^2 = \sum_i z_{iJ}^2$. By (36),

$$\begin{aligned} \frac{d}{dt}N_J &= 2\sum_i z_{iJ}\frac{d}{dt}z_{iJ} \\ &= 2\sum_i z_{iJ}(-z_{iJ} + \theta_i) \\ &= 2(-N_J + S_J). \end{aligned} \tag{47}$$

Equations (38) and (47) show that N_J tracks S_J as S_J tracks $||\theta||^2$. Consequently the norm $||z^{(J)}|| = \sqrt{N_J}$ approaches $||\theta||$ as learning proceeds. In addition, since S_J monotonically approaches $||\theta||^2$, N_J oscillates at most once.

Finally, let us verify that the angle between $Z^{(J)}$ and θ closes monotonically during learning even while $||z^{(J)}||$ is converging to $||\theta||$. To do this, the notation $C_J = cos(z^{(J)}, \theta)$ is convenient. It is sufficient to prove that $C_J(t)$ increases towards 1 as learning proceeds. By (30), C_J can be rewritten as

$$C_J = \frac{S_J}{||\theta||\sqrt{N_J}}. \tag{48}$$

Differentiating (48), we find that

$$\frac{d}{dt}C_J = \frac{N_J^{\frac{1}{2}}\frac{d}{dt}S_J - \frac{1}{2}S_J N_J^{-\frac{1}{2}}\frac{d}{dt}N_J}{||\theta||N_J}. \tag{49}$$

Substituting (38) and (47) into (49) and cancelling the term $N_J S_J$ in two places leads to the equation

$$\begin{aligned}\frac{d}{dt}C_J &= \frac{N_J||\theta||^2 - S_J^2}{||\theta||N_J^{\frac{3}{2}}} \\ &= \frac{||\theta||}{\sqrt{N_J}}(1 - \frac{S_J^2}{||\theta||^2 N_J}),\end{aligned} \tag{50}$$

which by (48) is the same as

$$\frac{d}{dt}C_J = \frac{||\theta||}{\sqrt{N_J}}(1 - C_J^2) \geq 0. \qquad 51)$$

Equation (51) shows that C_J increases monotonically towards 1.

4. Concluding Remarks

This chapter has described two models whose explanations of complex data can be understood by analysing their interactive dynamics in real time. The first model illustrates how classical data about apparent motion can be rationalized in terms of the spatiotemporal dynamics of a long-range Gaussian filter followed by a contrast-enhancing stage of lateral

inhibition. The model suggests how paradoxical properties of motion speed-up can be linked to the functionally useful property of synthesizing motion data from multiple spatial scales, and to the property of predictively interpolating intermittent motion signals in a way that is capable of continuously tracking a target moving at variable speeds with a focus of spatial attention.

The second model illustrates how a combination of bottom-up adaptive filtering, top-down learned expectations, attentive resonance, and novelty-sensitive memory search can control rapid learning of recognition categories whose shape and coarseness can be matched to complex environmental demands, including culturally imposed demands. The model clarifies how top-down expectations can stabilize the learning process in an arbitrary environment and, in so doing, focus attention upon and coherently bind those prototypical feature clusters that are used in object categorization. Properties such as memory consolidation arise as dynamical properties of network interactions, and data about such varied phenomena as phonemic restoration, priming, and the dynamics of inferotemporal cortex were linked to emergent properties of the network model.

These results suggest that basic neural mechanisms, such as contrast-enhancing lateral inhibition, play a role in multiple neural systems, whether to select the peak of a motion wave, as in Figure 6, or to choose a recognition category, as in equation (28). On the level of the system design itself, the results support the view that two distinct types of attention may modulate visual information processing, one a form of spatial attention that arises in the Where processing stream through MT and parietal cortex, and the other a form of featural attention that arises within the What processing stream through V4 and temporal cortex (Figure 8). How these two types of attention interact during our daily experiences with rapidly changing mixtures of familiar and unfamiliar events remains an important subject for future research.

REFERENCES

Amari, S.-I. and Takeuchi, A. 1978. Mathematical theory on formation of category detecting nerve cells. *Biological Cybernetics*, 29:127–136.

Anstis, S. and Ramachandran, V.S. 1987. Visual inertia in apparent motion. *Vision Research*, 27:755–764.

Bartley, S.H. 1941. *Vision, a study of its basis.* New York: D. Van Nostrand.

Bienenstock, E.L., Cooper, L.N., and Munro, P.W. 1982. Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2:32–48.

Boring, E.G. 1950. *A history of experimental psychology.* Englewood Cliffs, NJ: Prentice-Hall.

Braddick, O. 1980. Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society (London)*, 290B:137–151.

Braddick, O. and Adlard, A. 1978. Apparent motion and the motion detector. In *Visual psychophysics and psychology*, edited by J.C. Armington, J. Krauskopf, and B.R. Wooten. New York: Academic Press.

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

Breitmeyer, B.G. and Ritter, A. 1986. Visual persistence and the effect of eccentric viewing, element size, and frame duration on bistable stroboscopic motion percepts. *Perception and Psychophysics*, 39(4):275–280.

Carpenter, G.A. and Grossberg, S. 1987. A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37:54–115.

Carpenter, G.A. and Grossberg, S. 1990. ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, 3:129–152.

Carpenter, G.A. and Grossberg, S., eds. 1991. *Pattern recognition by self-organizing neural networks.* Cambridge, MA: MIT Press.

Carpenter, G.A. and Grossberg, S. 1992. Fuzzy ARTMAP: Supervised learning, recognition, and prediction by a self-organizing neural network. *IEEE Communications Magazine*, 30:38–49.

Carpenter, G.A. and Grossberg, S. 1993. Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences*, 16:131–137.

Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., and Rosen, D.B. 1992. Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, 3:698–713.

Carpenter, G.A., Grossberg, S., and Reynolds, J.H. 1991. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, 4:565–588.

Cohen, M.A. and Grossberg, S. (1987). Masking fields: A massively parallel architecture for learning, recognizing, and predicting multiple groupings of patterned data. *Applied Optics*, **26**, 1866-1891.

Damasio, A.R. 1989. The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, 1:123–132.

Desimone, R. 1991. Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience*, 3:1–8.

Desimone, R. 1992. Neural circuits for visual attention in the primate brain. In *Neural networks for vision and image processing*, edited by G.A. Carpenter and S. Grossberg. Cambridge, MA: MIT Press, 343–364.

Desimone, R. and Ungerleider, L.G. 1989. Neural mechanisms of visual processing in monkeys. In *Handbook of neuropsychology, Volume 2*, edited by F. Boller and J. Grafman. Amsterdam:

Elsevier Science Publishers, 267–299.

DeYoe, E.A. and van Essen, D.C. 1988. Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences*, 11:219–226.

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., and Reitbock, H.J. 1988. Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics*, 60:121–130.

Eckhorn, R. and Schanze, T. 1991. Possible neural mechanisms of feature linking in the visual system: Stimulus-locked and stimulus-induced synchronizations. In *Self-organization, emerging properties, and learning*, edited by A. Babloyantz. New York: Plenum Press.

Eriksen, C.W. and Murphy, T.D. 1987. Movement of attentional focus across the visual field: A critical look at the evidence. *Perception and Psychophysics*, 42:29–305.

Exner, S. 1875. Ueber das sehen von bewegungen und die theorie des zusammengesetzen Auges. *Sitzungsberichte Akademie Wissenschaft Wien*, 72:156–190.

Gibson, J.J. (1979). The ecological approach to visual perception. Boston, MA: Houghton Mifflin.

Gochin, P.M., Miller, E.K., Gross, C.G., and Gerstein, G.L. 1991. Functional interactions among neurons in inferior temporal cortex of the awake macaque. *Experimental Brain Research*, 84:505–516.

Goodale, M.A. and Milner, A.D. 1992. Separate visual pathways for perception and action. *Trends in Neurosciences*, 15:20–24.

Gray, C.M., Konig, P., Engel, A.K., and 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. and Singer, W. 1989. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences*, 86:1698–1702.

Grossberg, S. 1969. On learning and energy-entropy dependence in recurrent and nonrecurrent

signed networks. *Journal of Statistical Physics*, 1:319–350.

Grossberg, S. 1972. Neural expectation: Cerebellar and retinal analogs of cells fired by learnable or unlearned pattern classes. *Kybernetik*, 10:49–57.

Grossberg, S. 1973. Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52:217-257. Reprinted in: S. Grossberg. 1972.*Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control.* Dordrecht: Kluwer Academic Publishers.

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

Grossberg, S. 1976b. Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, 23:187–202.

Grossberg, S. 1977. Apparent motion. Unpublished manuscript.

Grossberg, S. 1978. A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plan. *Progress in theoretical biology, Volume 5*, edited by R. Rosen and F. Snell. New York: Academic Press. Reprinted in: S. Grossberg. 1982. *Studies of mind and brain.* Boston, MA: Reidel Press.

Grossberg, S. 1980. How does a brain build a cognitive code? *Psychological Review*, 1:1–51.

Grossberg, S. 1982. *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control.* Boston: Reidel Press.

Grossberg, S., ed. 1987a. *The adaptive brain, I: Cognition, learning, reinforcement, and rhythm.* Amsterdam: Elsevier/North-Holland.

Grossberg, S. ed. 1987b. *The adaptive brain, II: Vision, speech, language, and motor control.* Amsterdam: Elsevier/North-Holland.

Grossberg, S. ed. 1988. *Neural networks and natural intelligence.* Cambridge, MA: MIT Press.

Grossberg, S. 1991. Why do parallel cortical systems exist for the perception of static form

and moving form? *Perception and Psychophysics*, 49:117–141.

Grossberg, S. 1994. 3-D vision and figure-ground separation by visual cortex.*Perception and Psychophysics*, January, in press.

Grossberg, S. and Kuperstein, M. 1986. *Neural dynamics of adaptive sensory-motor control: Ballistic eye movements.* Amsterdam: Elsevier/North-Holland (1989, expanded edition, Elmsford, NY: Pergamon Press).

Grossberg, S. and Mingolla, E. 1993. Neural dynamics of motion perception: Direction fields, apertures, and resonant grouping. *Perception and Pyschophysics*, 53:243–278.

Grossberg, S., Mingolla, E., and Todorović, D. 1989. A neural network architecture for preattentive vision. *IEEE Transactions on Biomedical Engineering*, 36:65-84.

Grossberg, S. and Rudd, M. 1989. A neural architecture for visual motion perception: Group and element apparent motion. *Neural Networks*, 2:421–450.

Grossberg, S. and Rudd, M.E. 1992. Cortical dynamics of visual motion perception: Short-range and long-range apparent motion. *Psychological Review*, 99:78–121.

Grossberg, S. and Somers, D. 1991. Synchronized oscillations during cooperative feature linking in a cortical model of visual perception. *Neural Networks*, 4:453–466.

Grossberg, S. and Somers, D. 1992. Synchronized oscillations for binding spatially distributed feature codes into coherent spatial patterns. In *Neural networks for vision and image processing*, edited by G.A. Carpenter and S. Grossberg. Cambridge, MA: MIT Press, 385–405.

Grossberg, S. and Stone, G.O. 1986. Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review*, 93:46–74.

Harries, M.H. and Perrett, D.I. 1991. Visual processing of faces in temporal cortex: Physiological evidence for a modular organization and possible anatomical correlates. *Journal of Cognitive Neuroscience*, 3:9–24.

Kohonen, T. 1984. *Self-organization and associative memory.* New York: Springer-Verlag. Second edition, 1987.

Kolers, P.A. 1972. *Aspects of motion perception.* Elmsford, NY: Pergamon Press.

Kolers, P.A. and von Grünau, M. 1975. Visual construction of color is digital. *Science*, 187:757–759.

Korte, A. 1915. Kinematoskopische Untersuchungen. *Zeitschrift für Psychologie*, 72:194–296.

Kwak, H.-W., Dagenbach, D., and Egeth, H. 1991. Further evidence for a time-independent shift of the focus of attention. *Perception and Psychophysics*, 49:473–480.

LaBerge, D. and Brown, V. 1989. Theory of attentional operations in shape identification. *Psychological Review*, 96:101–124.

Levy, W.B. 1985. Associative changes at the synapse: LTP in the hippocampus. In *Synaptic modification, neuron selectivity, and nervous system organization*, edited by W.B. Levy, J. Anderson and S. Lehmkuhle. Hillsdale, NJ: Erlbaum Associates, 5–33.

Levy, W.B. and Desmond, N.L. 1985. The rules of elemental synaptic plasticity. In *Synaptic modification, neuron selectivity, and nervous system organization*, edited by W.B. Levy, J. Anderson, and S. Lehmkuhle. Hillsdale, NJ: Erlbaum Associates, 105–121.

Linsker, R. 1986. From basic network principles to neural architecture: Emergence of spatial-opponent cells. *Proceedings of the National Academy of Sciences*, 83:8779–8783.

Malsburg, C. von der. 1973. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14:85–100.

Miller, E.K., Li, L., and Desimone, R. 1991. A neural mechanism for working and recognition memory in inferior temporal cortex. *Science*, 254:1377–1379.

Mishkin, M. 1982. A memory system in the monkey. *Philosophical Transactions of the Royal Society of London*, 298:85–95.

Mishkin, M. and Appenzeller, T. 1987. The anatomy of memory. *Scientific American*, 256:80–89.

Pantle, A.J. and Petersik, J.T. 1980. Effects of spatial parameters on the perceptual organi-

FIGURE CAPTIONS

Figure 1. The Ternus display. (a) Three spots are presented in each frame in such a way that the two leftwardmost spots in Frame 2 occupy the same positions as the two rightwardmost spots in Frame 1. The two frames are repeatedly cycled with ISIs inserted between them. At very short ISIs, all dots appear to flicker in place. At longer ISIs the dots at shared positions appear to remain stationary, while apparent motion occurs between the leftwardmost spot in Frame 1 and the rightwardmost spot in Frame 2 ("element motion"). At still longer ISIs, the three dots appear to move from Frame 1 to Frame 2 and back as a group ("group motion"). (b) When the dots in successive frames have opposite contrast with respect to the frame, only group motion occurs at the ISIs where element motion occurred in (a). [Reprinted with permission from Grossberg and Rudd (1992).]

Figure 2. The simplest one-dimensional MOC Filter. The input pattern at Level 1 is spatially and temporally filtered by sustained response cells at Level 2. The sustained cells have oriented receptive fields that are sensitive to the direction-of-contrast in the image, either dark-to-light or light-to-dark. Level 2 cells play the role of a short-range spatial filter. Spatial and temporal averaging are also carried out by transient response cells at Level 3. The transient cells have unoriented receptive fields that are sensitive to the direction of contrast change in the cell input. The upward arrow denotes transient on-cells that are activated by a transition from dark to light. The downward arrow denotes transient off-cells that are activated by a transition from light to dark. Level 4 cells combine sustained and transient cell signals multiplicatively and are thus rendered sensitive to both direction-of-motion and direction-of-contrast. Level 5 cells sum across space via a long-range Gaussian spatial filter, and across the two types of Level 4 cells. Level 5 cells are thus sensitive to direction-of-motion but insensitive to direction-of-contrast. [Reprinted with permission from Grossberg and Rudd (1992).]

Figure 3. Spatial responses at various levels of the MOC Filter to a point input. (a) Sustained activity of a Level 2 cell. (b) Total input pattern to Level 5 after convolution with a Gaussian kernel. (c) Contrast-enhanced output of Level 5 centered at the location of the input maximum. [Reprinted with permission from Grossberg and Rudd (1992).]

Figure 4. Temporal response of the MOC Filter to a point input. (a) The input is presented at a brief duration at location 1. (b) Sustained cell activity at 1 gradually builds after the input onset, then decays after offset. (c) Growth of the input pattern to Level 5 with transient cell activity held constant. The activity pattern retains a Gaussian shape centered at the location of the input, that waxes and wanes through time without spreading across space. [Reprinted with permission from Grossberg and Rudd (1992).]

Figure 5. Temporal response of the sustained cells $x_{ik}(t)$ at Level 2 to two brief successive point inputs $I_{ik}(t)$ at locations $i = 0$ and $i = W$. For an appropriately timed display, the decaying response at position 0 overlaps in time the rising response at position W. Parameter k is defined in the full model. [Reprinted with permission from Grossberg and Rudd (1992).]

Figure 6. Simulated MOC Filter response to a two-flash display. Successive rows correspond to increasing times following the Frame 1 offset. (a) The two lower curves in each row depict the total input R_i at position i of Level 5 due to each of the two flashes. The input due to the left flash decreases while the input due to the right flash increases. The summed input due to both flashes is a traveling wave whose maximum value across space moves continuously between the two flash locations. (b) Position over time of the contrast-enhanced Level 5 response. [Reprinted with permission from Grossberg and Rudd (1992).]

Figure 7. Motion paths generated by MOC Filters with different Gaussian filter kernel widths K in (9). The motion paths are plotted in a space-time diagram wherein each rectangle indicates the spatiotemporal boundaries of one flash in a two flash display. All the

motion paths intersect at a point halfway between the two flash locations. [Reprinted with permission from Grossberg and Rudd (1992)].

Figure 8. Schematic diagram of anatomical connections and neuronal selectivities of early visual areas in the macaque monkey. LGN = lateral geniculate nucleus (parvocellular and magnocellular divisions). Divisions of V1 and V2: blob = cytochrome oxidase blob regions; interblob = cytochrome oxidase-poor regions surrounding the blobs; 4B = lamina 4B; thin = thin (narrow) cytochrome oxidase strips; interstripe = cytochrome oxidase-poor regions between the thin and thick stripes; thick = thick (wide) cytochrome oxidase strips; V3 = visual area 3; V4 = visual area(s) 4; MT = middle temporal area. Areas V2, V3, V4, MT have connections to other areas not explicitly represented here. Area V3 may also receive projections from V2 interstripes or thin stripes. Heavy lines indicate robust primary connections, and thin lines indicate weaker, more variable connections. Dotted lines represent observed connections that require additional verification. Icons: rainbow = tuned and/or opponent wavelength selectivity (incidence at least 40%); angle symbol = orientation selectivity (incidence at least 20%); spectacles = binocular disparity selectivity and/or strong binocular interactions (V2) (incidence at least 20%); pointing hand = direction of motion selectivity (incidence at least 20%). [Adapted with permission from DeYoe and van Essen (1988).]

Figure 9. Interactions between the attentional and orienting subsystems of an adaptive resonance theory (ART) circuit: Level F_1 encodes a distributed representation of an event to be recognized via a short term memory (STM) activation pattern across a network of feature detectors. Level F_2 encodes the event to be recognized using a more compressed STM representation of the F_1 pattern. Learning of these recognition codes takes place at the long term memory (LTM) traces within the bottom-up and top-down pathways between levels F_1 and F_2. The top-down pathways can read-out learned expectations whose prototypes are matched against bottom-up input patterns at F_1. Mismatches in response to novel events

activate the orienting subsystem $\mathcal{A}$ on the right side of the figure, depending on the value of the vigilance parameter ρ. Arousal emitted from $\mathcal{A}$ resets the recognition codes that are active in STM at F_2 and initiates a memory search for a more appropriate recognition code. Output from subsystem $\mathcal{A}$ can also trigger an orienting response. (a) Block diagram of circuit. [Reprinted with permission from Carpenter and Grossberg (1990).] (b) Individual pathways of circuit, including the input level F_0 that generates inputs to level F_1. The gain control input to level F_1 helps to instantiate the 2/3 Rule (see text). Gain control to level F_2 is needed to instate a category in STM. [Reprinted with permission from Carpenter, Grossberg, and Reynolds (1991).]

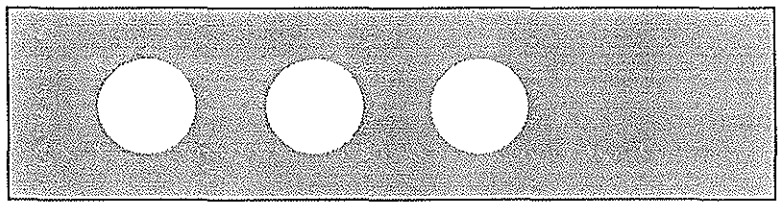
Figure 10. ART search for an F_2 recognition code: (a) The input pattern $\mathbf{I}$ generates the specific STM activity pattern $\mathbf{X}$ at F_1 as it nonspecifically activates the orienting subsystem $\mathcal{A}$. $\mathbf{X}$ is represented by the hatched pattern across F_1. Pattern $\mathbf{X}$ both inhibits $\mathcal{A}$ and generates the output pattern $\mathbf{S}$. Pattern $\mathbf{S}$ is transformed by the LTM traces into the input pattern $\mathbf{T}$, which activates the STM pattern $\mathbf{Y}$ across F_2. (b) Pattern $\mathbf{Y}$ generates the top-down output pattern $\mathbf{U}$ which is transformed into the prototype pattern $\mathbf{V}$. If $\mathbf{V}$ mismatches $\mathbf{I}$ at F_1, then a new STM activity pattern $\mathbf{X}^*$ is generated at F_1. $\mathbf{X}^*$ is represented by the hatched pattern. Inactive nodes corresponding to $\mathbf{X}$ are unhatched. The reduction in total STM activity which occurs when $\mathbf{X}$ is transformed into $\mathbf{X}^*$ causes a decrease in the total inhibition from F_1 to $\mathcal{A}$. (c) If the vigilance criterion fails to be met, $\mathcal{A}$ releases a nonspecific arousal wave to F_2, which resets the STM pattern $\mathbf{Y}$ at F_2. (d) After $\mathbf{Y}$ is inhibited, its top-down prototype signal is eliminated, and $\mathbf{X}$ can be reinstated at F_1. Enduring traces of the prior reset lead $\mathbf{X}$ to activate a different STM pattern $\mathbf{Y}^*$ at F_2. If the top-down prototype due to $\mathbf{Y}^*$ also mismatches $\mathbf{I}$ at F_1, then the search for an appropriate F_2 code continues until a more appropriate F_2 representation is selected. Then an attentive resonance develops and learning of the attended data is initiated. [Reprinted with permission from Carpenter and Grossberg (1990).]

Figure 11.(a) Many-to-one learning combines categorization of many exemplars into one category, and labeling of many categories with the same name. (b) In an ARTMAP architecture, the ART_a and ART_b networks form recognition categories of the separate streams of input vectors labelled **a** and **b**, as in the case of visual categories and their auditory naming categories. The Map Field learns an associative map from categories in ART_a to categories in ART_b. When a predicted output in ART_b is mismatched by an input vector **b**, the Match Tracking process increases the ART_a vigilance value ρ_A until $\rho_A|\mathbf{a}| - |\mathbf{x}^a| > 0$, thereby triggering memory search for a better set of features in **a** with which to build a category that can correctly predict **b**. [Reprinted with permission from Carpenter and Grossberg (1992)].

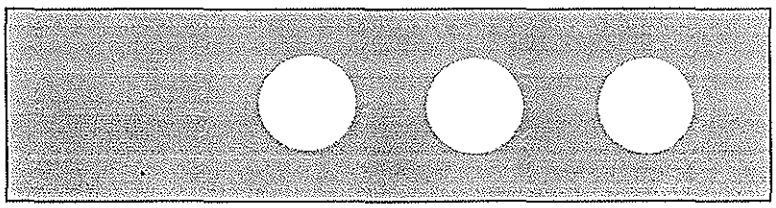
Table 1. Complementary cognitive and neural properties that are synthesized into a self-consistent computational format within individual ART recognition models. The cumulative computational constraints imposed by these properties force design decisions that are not evident in either cognitive or neural data by themselves.

Table 2. Some machine learning benchmark studies which compare the performance of supervised ART, or ARTMAP, models with that of alternative models. These benchmarks describe how well these systems predict test sets when they experience equivalent training sets (as in benchmarks 1–4) and the number of epochs, or repetitions of the training set, that are needed to reach the same level of accuracy (benchmark 5) [Reprinted with permission from Carpenter and Grossberg, 1992].

FRAME 1

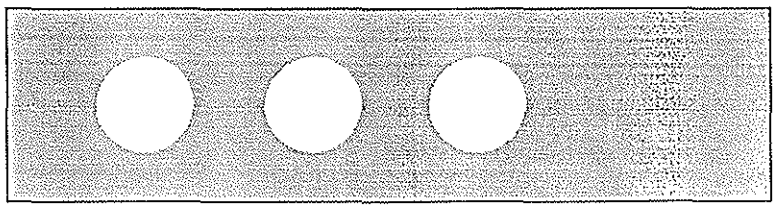


FRAME 2

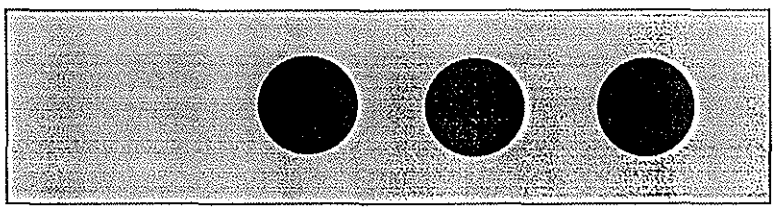


(a)

FRAME 1



FRAME 2



(b)

Figure 1

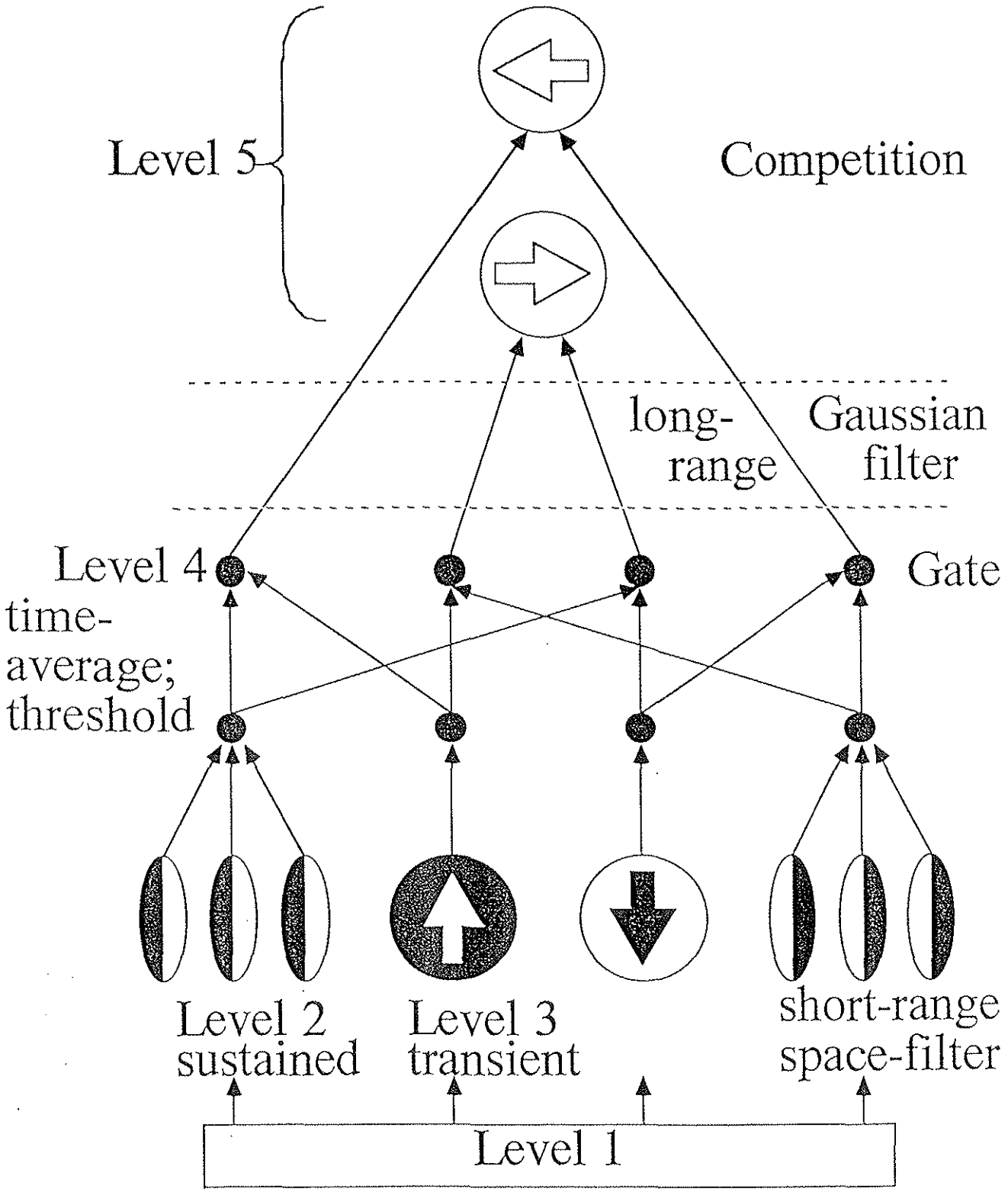

Level 5
Competition
long-
range
Gaussian
filter
Level 4
time-
average;
threshold
Gate
Level 2
sustained
Level 3
transient
short-range
space-filter
Level 1

Figure 2

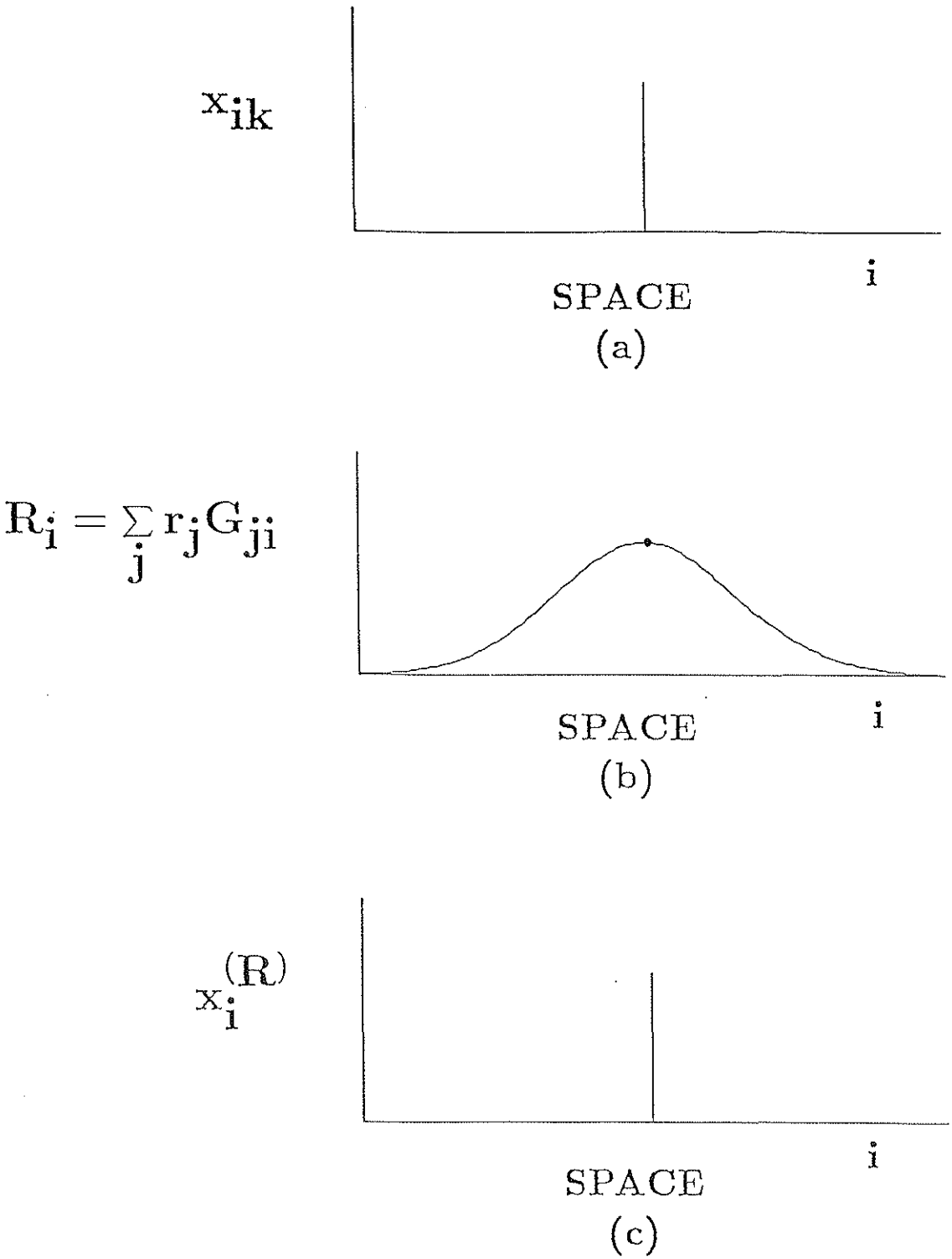


Figure 3

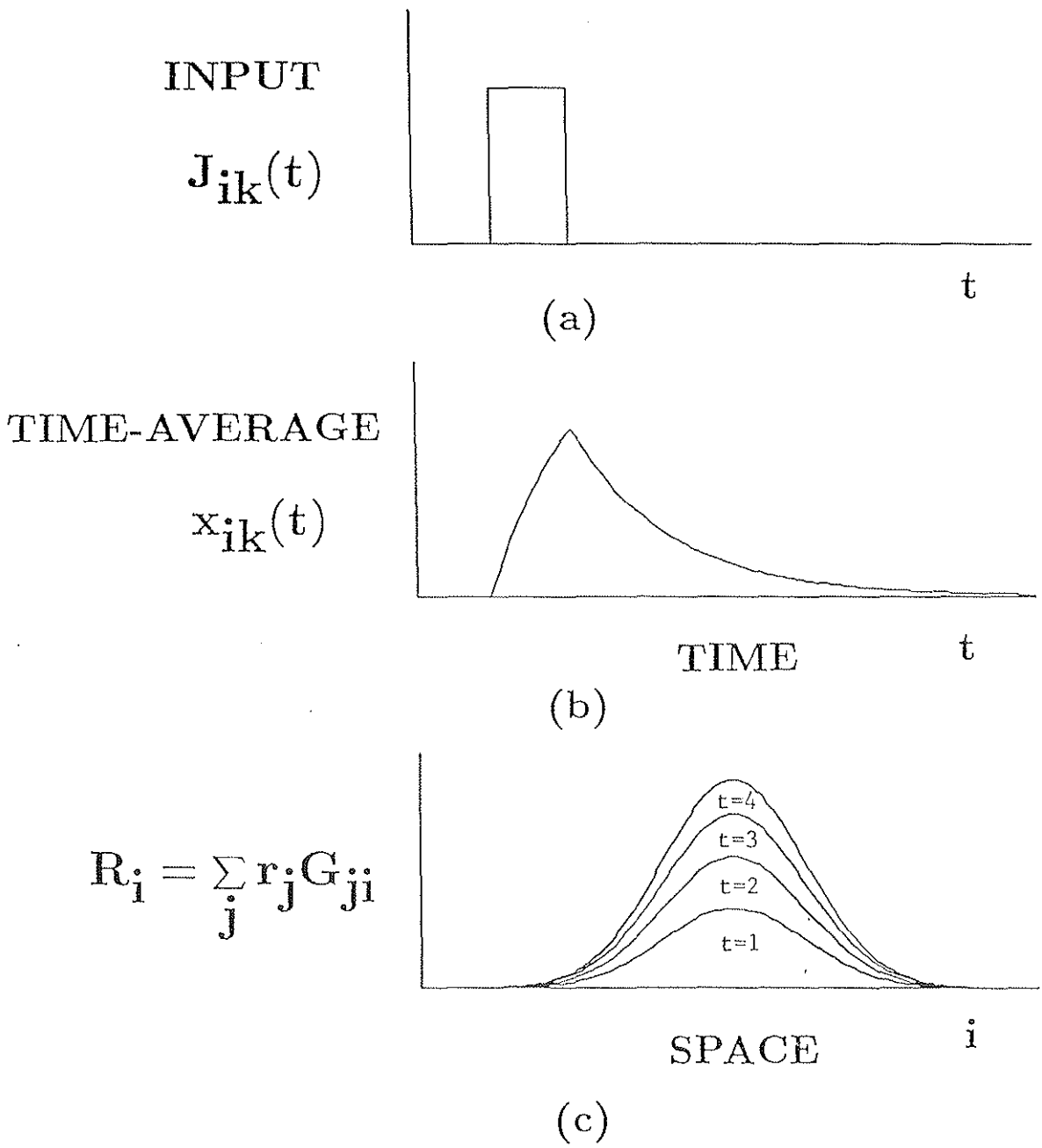

INPUT
$J_{ik}(t)$
t
(a)
TIME-AVERAGE
$x_{ik}(t)$
TIME
t
(b)
$R_i = \sum_j r_j G_{ji}$
t=4
t=3
t=2
t=1
SPACE
i
(c)

Figure 4

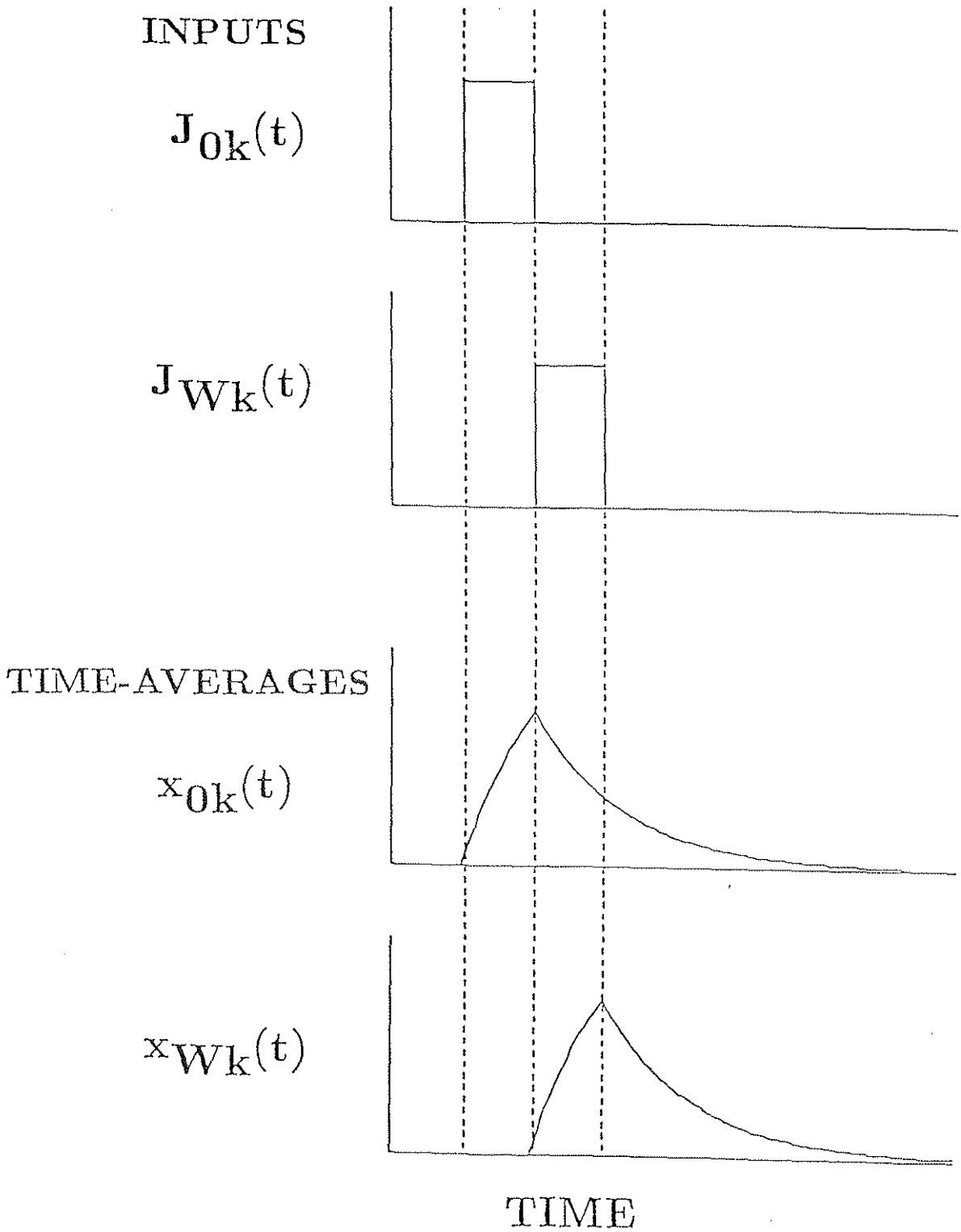


Figure 5

LONG-RANGE INTERACTION

$$R_i = \sum_j r_j G_{ji}$$

SHARP MOTION SIGNAL

$$x_i^{(R)} = \begin{cases} 1 & \text{if } R_i > R_j, j \neq i \\ 0 & \text{otherwise} \end{cases}$$

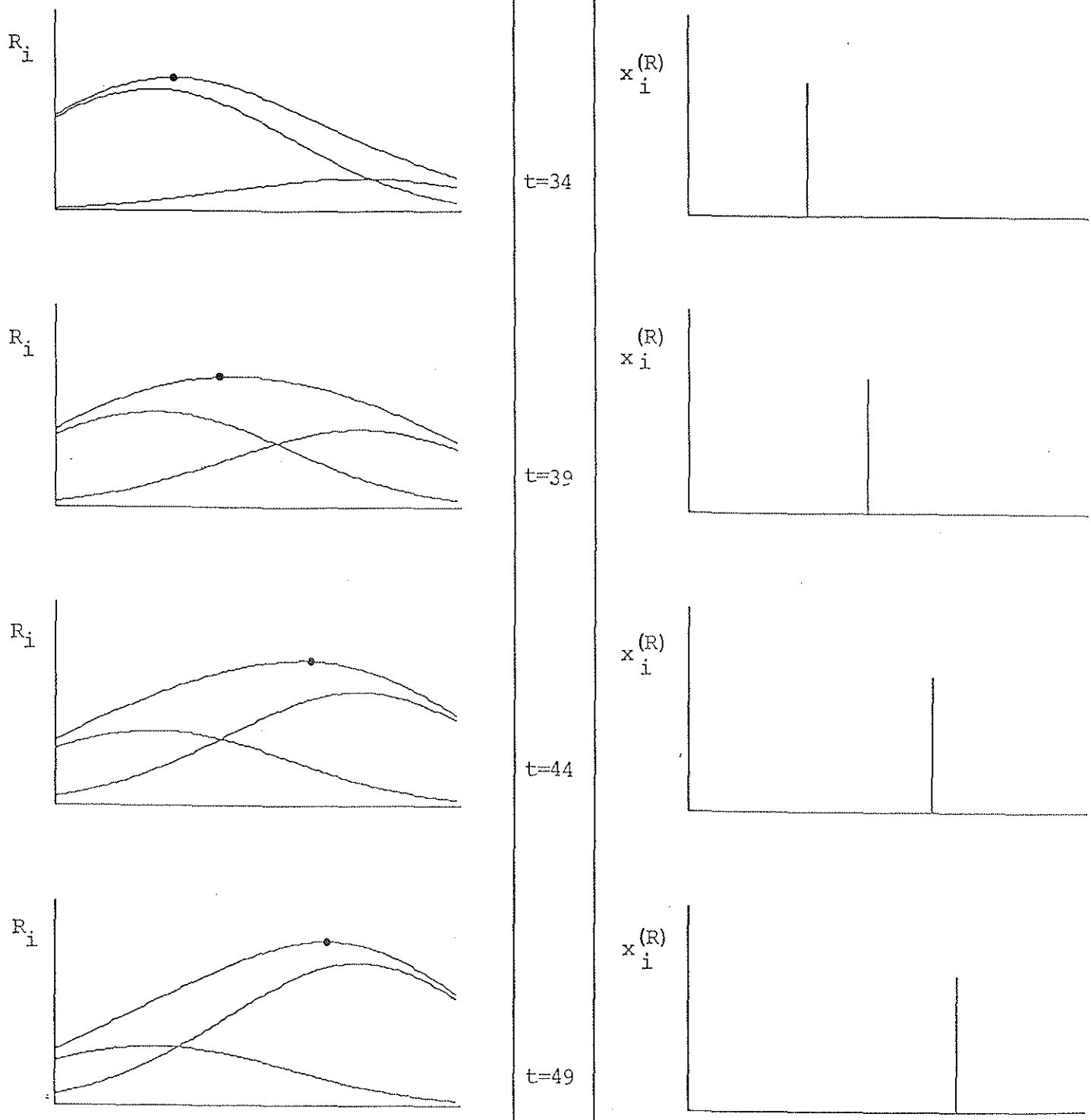


Figure 6

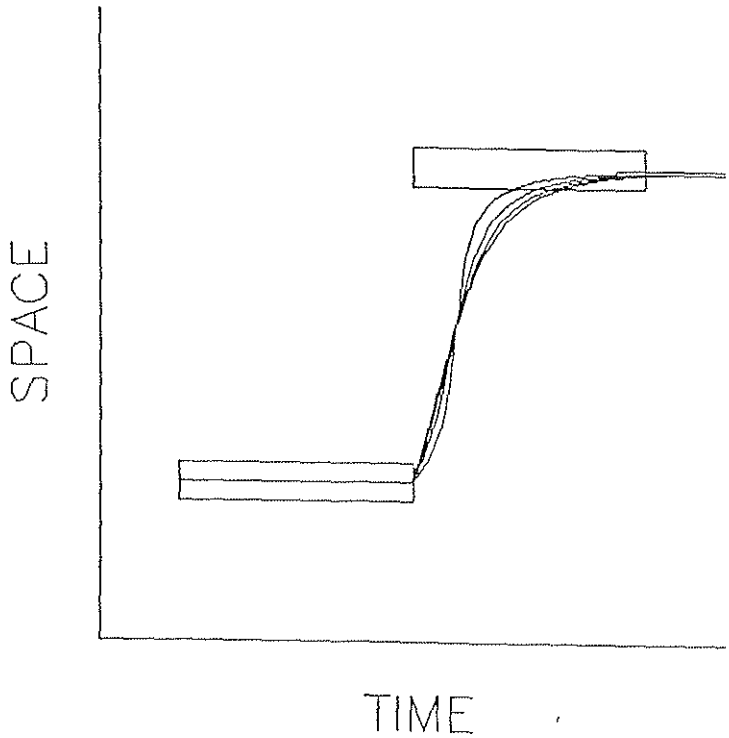


Figure 7

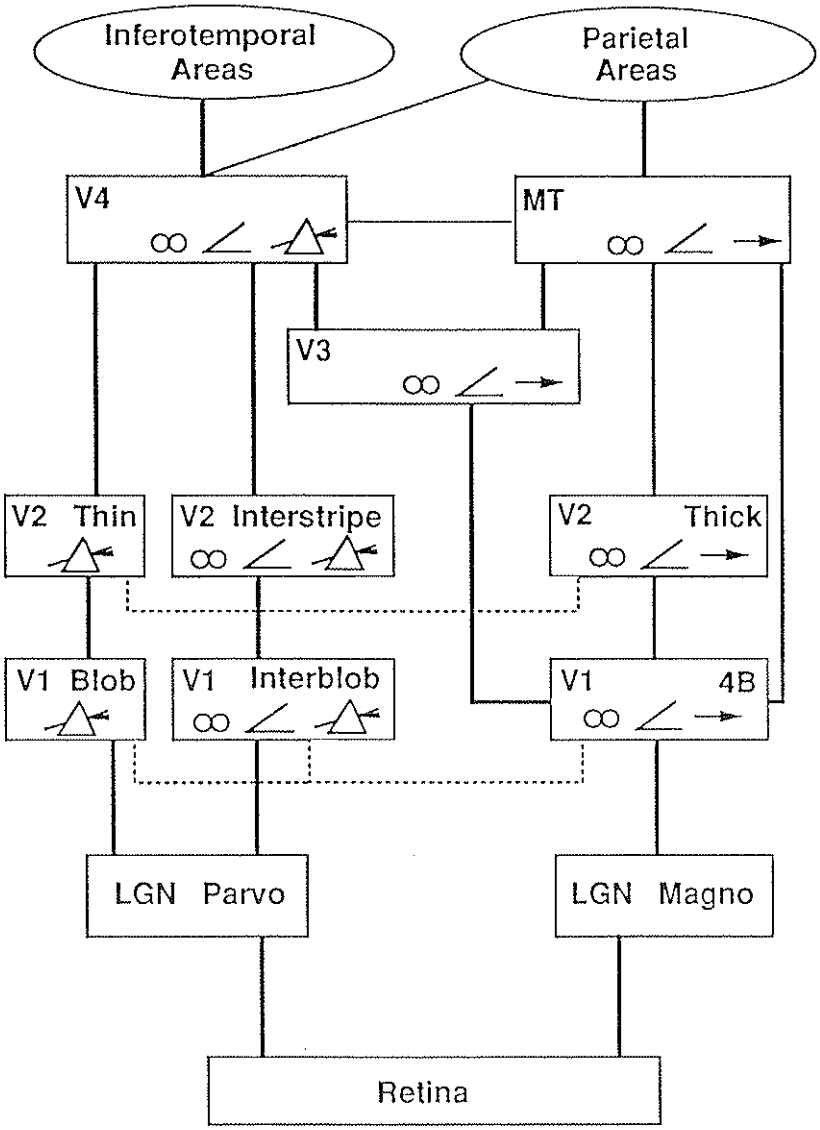

Inferotemporal Areas
Parietal Areas
V4
MT
V3
V2 Thin
V2 Interstripe
V2 Thick
V1 Blob
V1 Interblob
V1 4B
LGN Parvo
LGN Magno
Retina

Figure 8

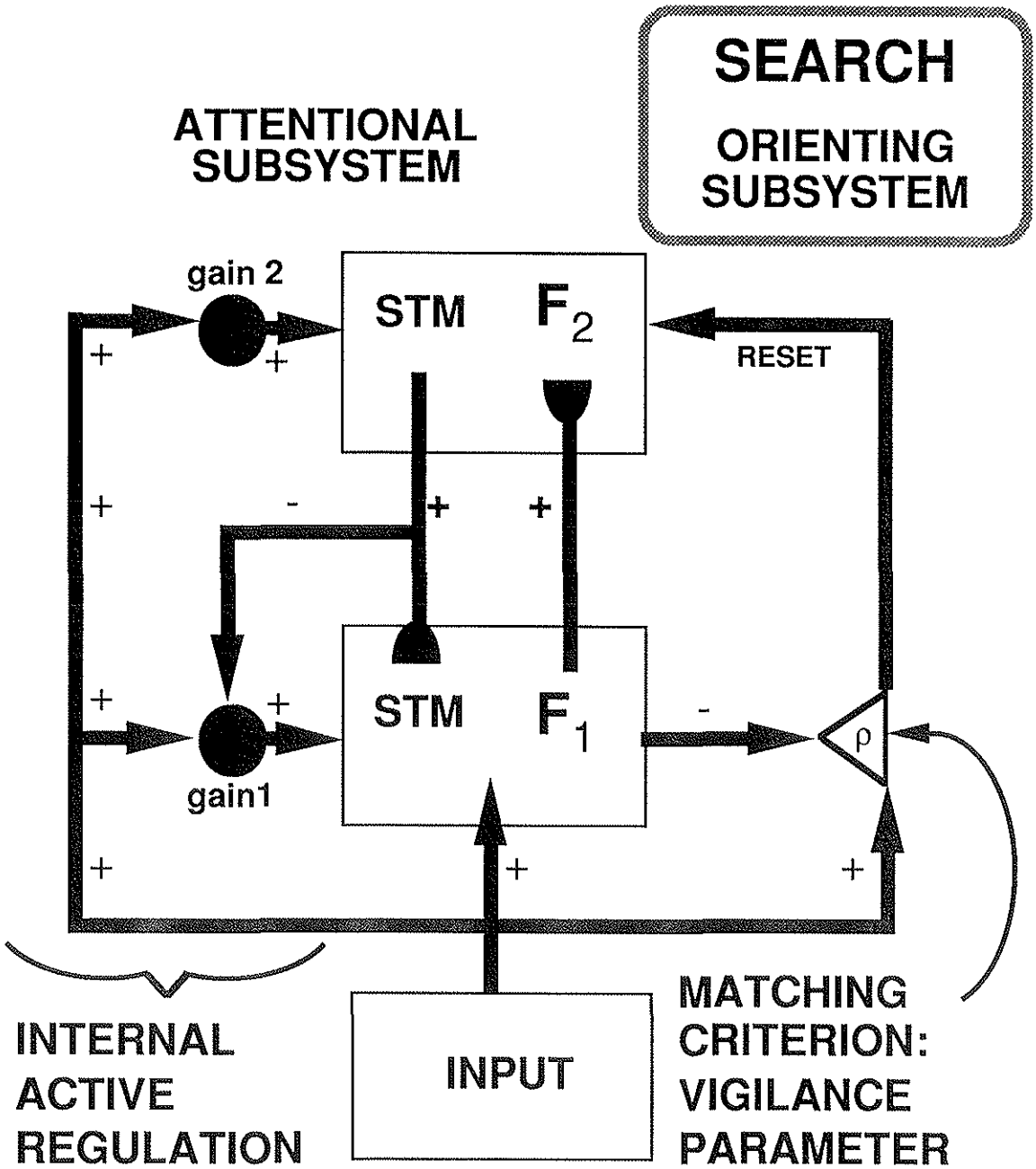

SEARCH
ORIENTING SUBSYSTEM
ATTENTIONAL SUBSYSTEM
gain 2
STM
F2
RESET
STM
F1
gain1
ρ
INTERNAL ACTIVE REGULATION
INPUT
MATCHING CRITERION: VIGILANCE PARAMETER

Figure 9A

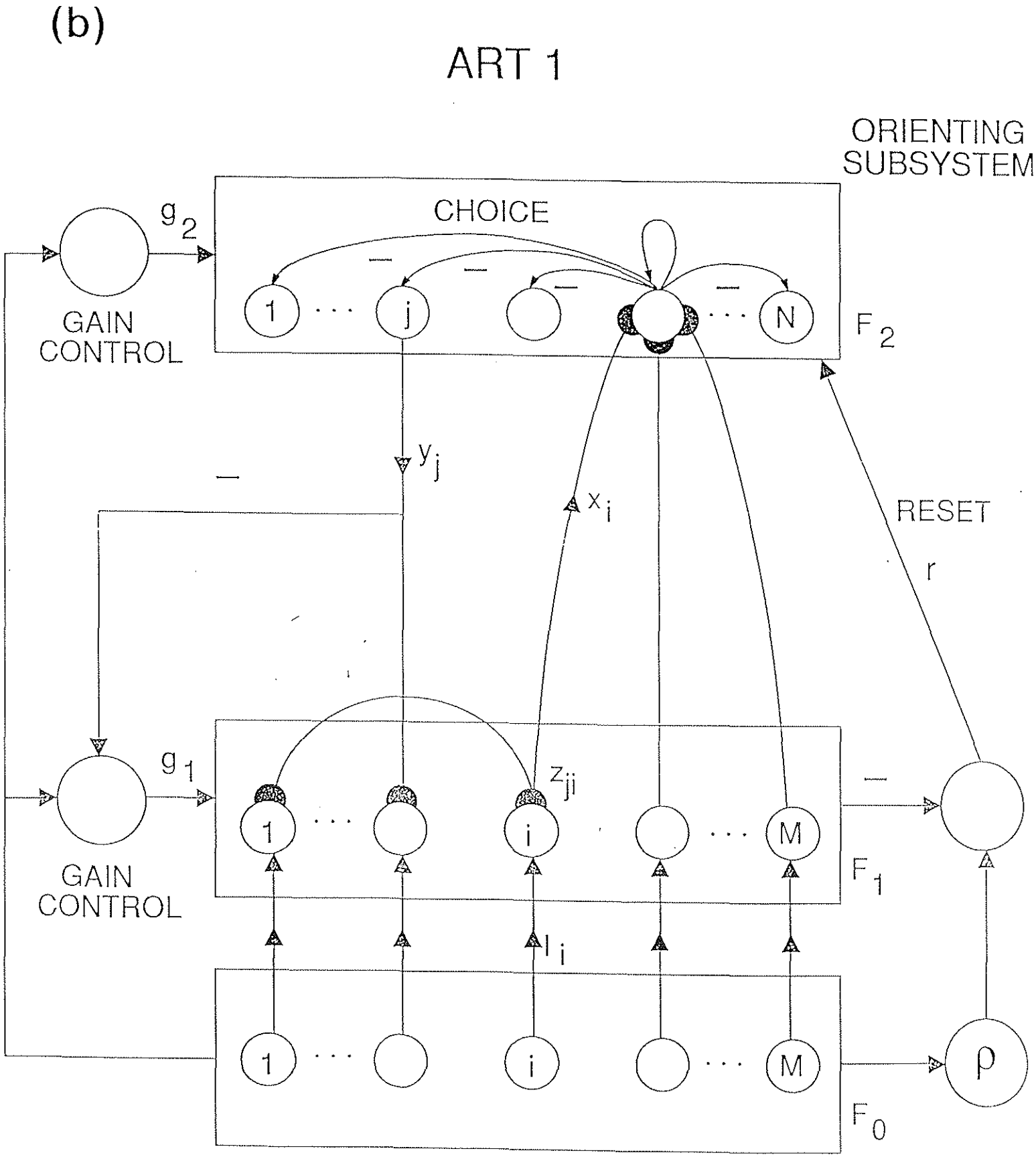

(b)
ART 1
ORIENTING SUBSYSTEM
CHOICE
g_2
GAIN CONTROL
1
j
N
F_2
y_j
x_i
RESET
r
g_1
z_{ji}
1
i
M
F_1
GAIN CONTROL
I_i
1
i
M
F_0
ρ

Figure 9B

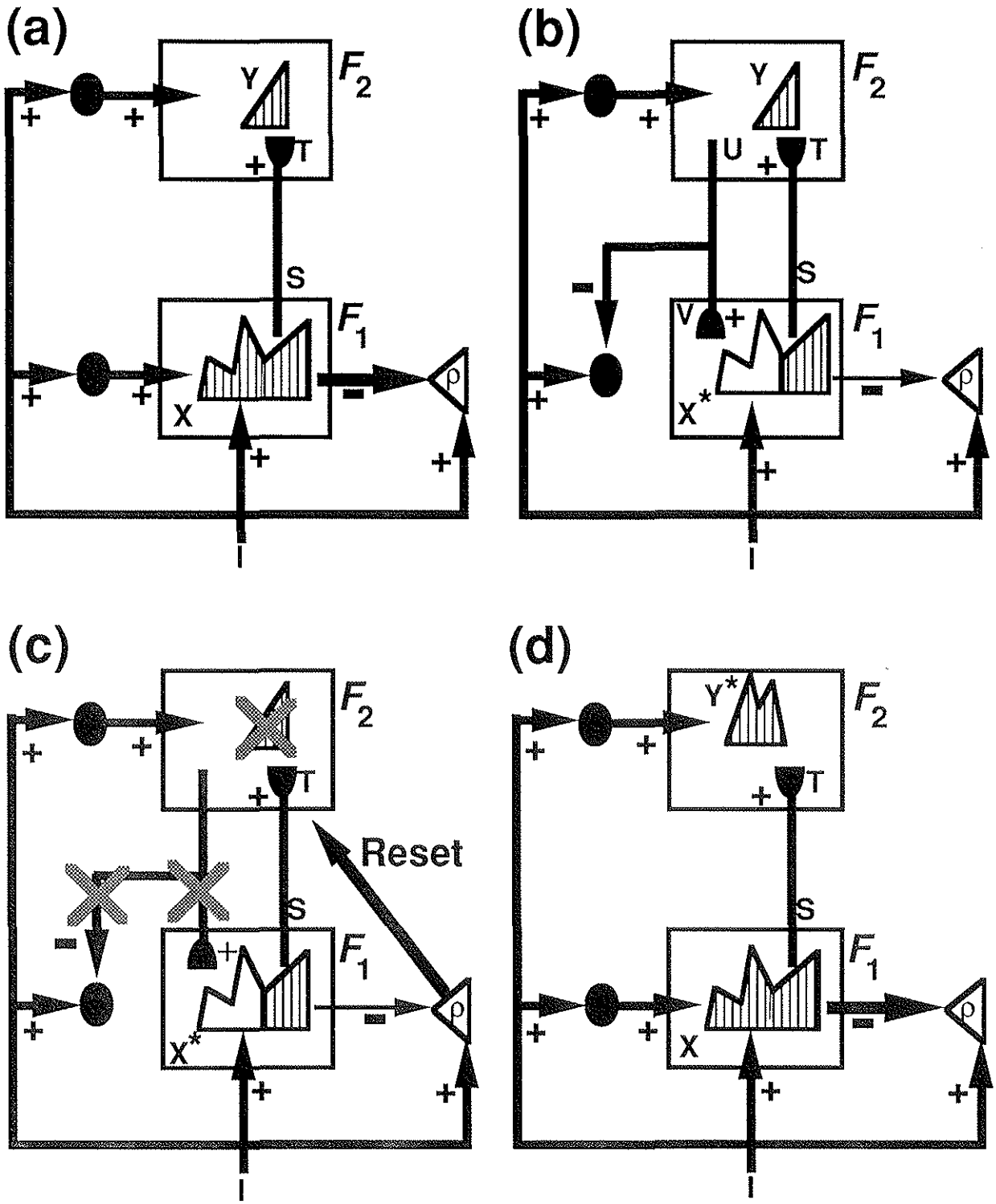

(a)
Y
F2
T
S
F1
X
ρ
I
(b)
Y
F2
U
T
S
V
F1
X*
ρ
I
(c)
F2
T
Reset
S
F1
X*
ρ
I
(d)
Y*
F2
T
S
F1
X
ρ
I

Figure 10

MANY-TO-ONE MAP

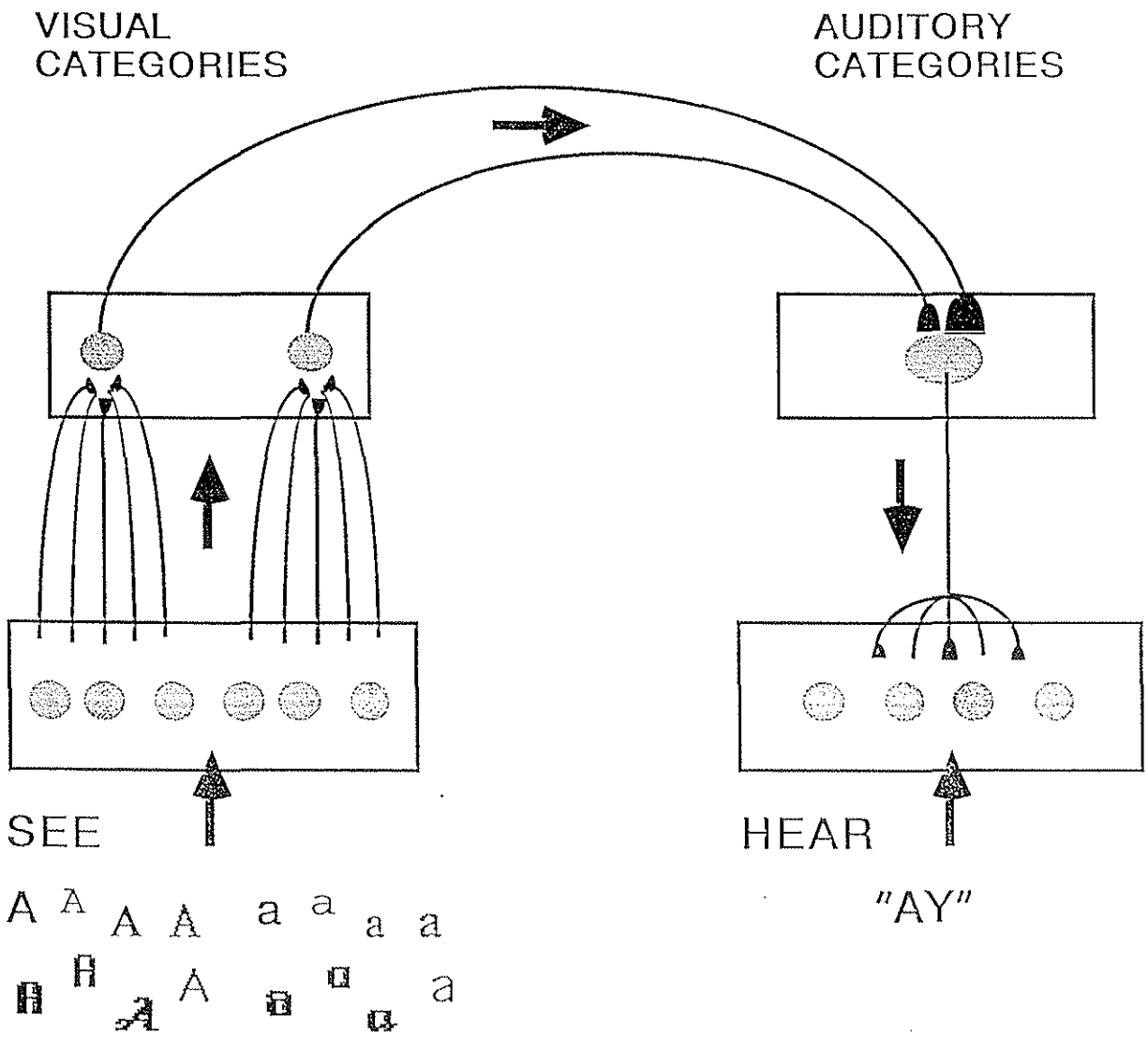


Figure 11A

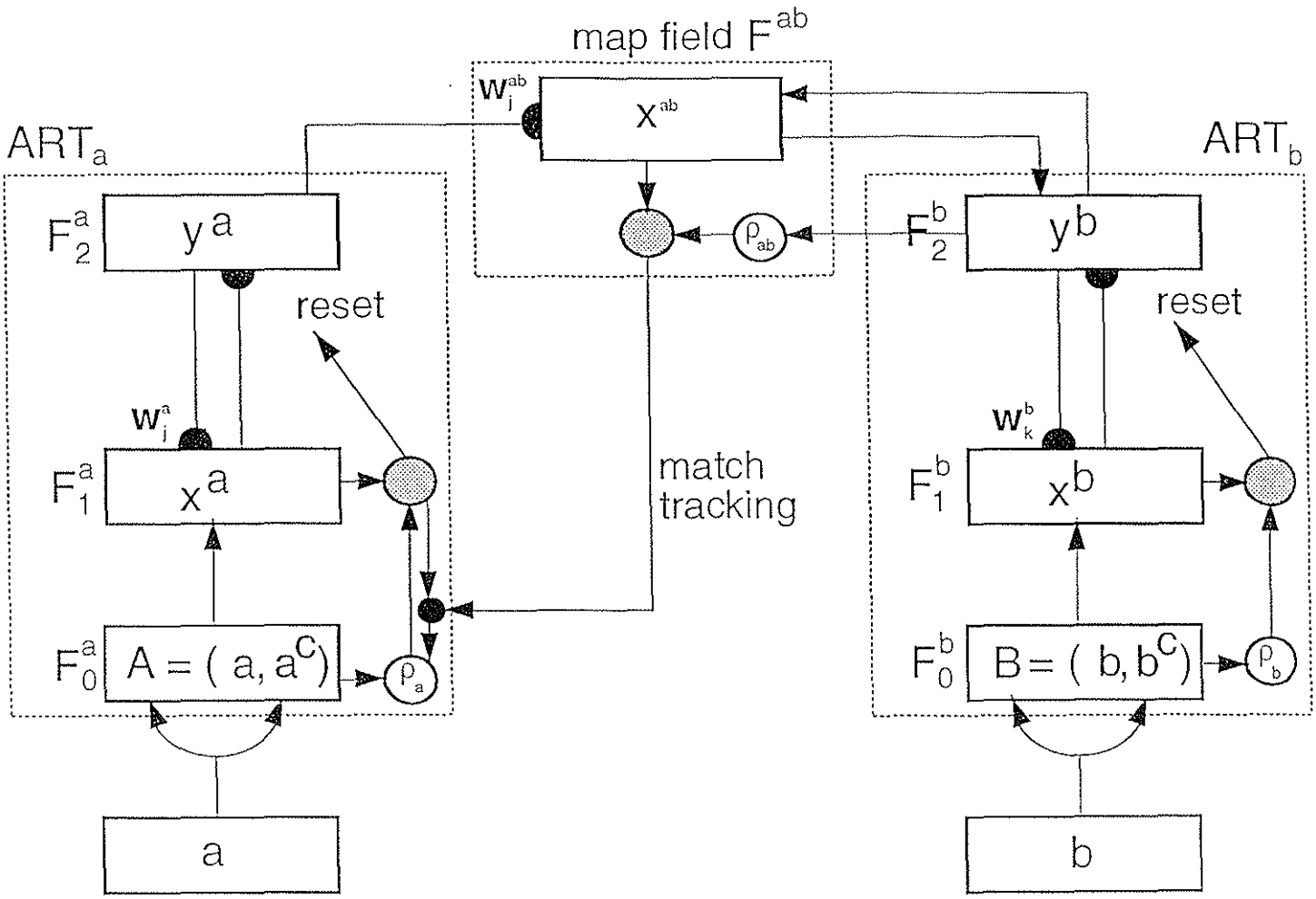

map field F^{ab}
w_j^{ab}
x^{ab}
ART_a
ART_b
F_2^a
y^a
F_2^b
y^b
ρ_{ab}
reset
reset
w_j^a
w_k^b
F_1^a
x^a
F_1^b
x^b
match
tracking
F_0^a
$A = (a, a^c)$
ρ_a
F_0^b
$B = (b, b^c)$
ρ_b
a
b

Figure 11B

COMPLEMENTARY PROPERTIES OF AN ART RECOGNITION ARCHITECTURE

stability	plasticity
attention	orientation
short term memory	long term memory
bottom-up	top-down
match	mismatch
direct access	memory search
resonance	reset
supraliminal	subliminal
specific	nonspecific
local features	global patterns
attentional gain	vigilance
cooperation	competition
distributed	compressed

Table 1

ARTMAP BENCHMARK STUDIES

1. Medical database - mortality following coronary bypass grafting (CABG) surgery

   FUZZY ARTMAP significantly outperforms:

   LOGISTIC REGRESSION

   ADDITIVE MODEL

   BAYESIAN ASSIGNMENT

   CLUSTER ANALYSIS

   CLASSIFICATION AND REGRESSION TREES

   EXPERT PANEL-DERIVED SICKNESS SCORES

   PRINCIPAL COMPONENT ANALYSIS

2. Mushroom database

   DECISION TREES (90-95 % correct)

   ARTMAP (100% correct)

   Training set an order of magnitude smaller

3. Letter recognition database

   GENETIC ALGORITHM (82% correct)

   FUZZY ARTMAP (96% correct)

4. Circle-in-the-Square task

   BACK PROPAGATION (90% correct)

   FUZZY ARTMAP (99.5% correct)

5. Two-Spiral task

   BACK PROPAGATION (10,000 - 20,000 training epochs)

   FUZZY ARTMAP (1-5 training epochs)

Table 2