Neural Networks, Vol. 4, pp. 493–504, 1991
Printed in the USA.

0893-0080/91 $3.00 + .00

ORIGINAL CONTRIBUTION

ART 2-A: An Adaptive Resonance Algorithm for Rapid Category Learning and Recognition

GAIL A. CARPENTER*
STEPHEN GROSSBERG†
DAVID B. ROSEN‡

Boston University

(*Received* 23 *October* 1990; *revised and accepted* 17 *January* 1991)

Abstract—*This article introduces Adaptive Resonance Theory 2-A (ART 2-A), an efficient algorithm that emulates the self-organizing pattern recognition and hypothesis testing properties of the ART 2 neural network architecture, but at a speed two to three orders of magnitude faster. Analysis and simulations show how the ART 2-A systems correspond to ART 2 dynamics at both the fast-learn limit and at intermediate learning rates. Intermediate learning rates permit fast commitment of category nodes but slow recoding, analogous to properties of word frequency effects, encoding specificity effects, and episodic memory. Better noise tolerance is hereby achieved without a loss of learning stability. The ART 2 and ART 2-A systems are contrasted with the leader algorithm. The speed of ART 2-A makes practical the use of ART 2 modules in large scale neural computation.*

Keywords—Neural networks, Pattern recognition, Category formation, Fast learning, Adaptive resonance, ART.

1. INTRODUCTION

Adaptive Resonance Theory (ART) architectures are neural networks that carry out stable self-organization of recognition codes for arbitrary sequences of input patterns. ART first emerged from an analysis of the instabilities inherent in feedforward adaptive coding structures (Grossberg, 1976a, 1976b). More recent work has led to the development of three classes of ART neural network architectures, specified as systems of differential equations. The first class, ART 1, self-organizes recognition categories for arbitrary sequences of binary input patterns (Carpenter & Grossberg, 1987a). A second class, ART 2, does the same for either binary or analog inputs (Carpenter & Grossberg, 1987b). A third class, ART 3, is based on ART 2 but includes a model of the chemical synapse that solves the memory search problem of ART systems embedded in network hierarchies, where there can, in general, be either fast or slow learning and distributed or compressed code representations (Carpenter & Grossberg, 1990).

This article introduces ART 2-A, a simple computational system that models the essential dynamics of the ART 2 analog pattern recognition neural network. The ART 2-A system accurately reproduces the behavior of ART 2 in the fast-learn limit, suggests an efficient method for simulating slow learning, and sharply delineates the essential computations performed by ART 2. ART 2-A runs approximately two to three orders of magnitude faster than ART 2 in simulations on conventional computers, thereby making it easier to use in solving large problems. The ART 2-A algorithm also suggests efficient parallel implementations.

The improved speed of the ART 2-A algorithm is due, in part, to the explicit specification of steady-state variables as a composition of a small number of nonlinear operations. The steady-state equations replace a time-consuming multilayer iterative component of ART 2.

* Supported in part by British Petroleum (89-A-1204), DARPA (AFOSR 90-0083), and the National Science Foundation (NSF IRI-90-00530).

† Supported in part by the Air Force Office of Scientific Research (AFOSR 90-0175 and AFOSR 90-0128), the Army Research Office (ARO DAAL-03-88-K0088), and DARPA (AFOSR 90-0083).

‡ Supported in part by DARPA (AFOSR 90-0083).

Acknowledgements: The authors wish to thank Carol Yanakakis Jefferson for her valuable assistance in the preparation of this manuscript.

Requests for reprints should be sent to Prof. Gail A. Carpenter, Center for Adaptive Systems, 111 Cummington Street, Boston University, Boston, MA 02215, USA.

A second feature of the ART 2-A system is its speed at intermediate learning rates. Intermediate learning rates capture many of the desirable properties of slow learning, including noise tolerance. However, the property of *fast commitment*, or asymptotic learning when a category first becomes active, allows the ART 2-A algorithm to be used as efficiently in this case as in the fast-learn limit. Thus, ART 2 may be needed in some cases not covered by ART 2-A; but ART 2-A can be efficiently substituted for ART 2 in most applications.

Section 2 characterizes ART 2; Section 3 motivates and describes the ART 2-A algorithm; and Section 4 presents the results of simulations comparing ART 2 and ART 2-A with fast learning, and comparing fast and intermediate learning rates in ART 2-A.

2. ANALYSIS OF ART 2 SYSTEM DYNAMICS

Carpenter and Grossberg (1987b) described several ART 2 systems, all having approximately equivalent dynamics. For definiteness, we consider one such system, shown in Figure 1. This ART 2 module includes the principal components of all ART modules, namely an *attentional subsystem*, which contains an input representation field F_1 and a category representation field F_2, and an *orienting subsytem*, which interacts with the attentional subsystem to carry out an internally controlled search process. The two fields are linked by both a bottom-up $F_1 \rightarrow F_2$ adaptive filter and a top-down $F_2 \rightarrow F_1$ adaptive filter. A path from the ith F_1 node to the jth F_2 node contains a long term memory (LTM) trace, or adaptive weight, z_{ij}, a path from the jth F_2 node to the ith F_1 node contains a weight z_{ji}. These weights *gate*, or multiply, path signals between fields.

Figure 1 also illustrates some ART 2 features that are not shared by all ART modules. One such feature is the three layer F_1 field. Both F_1 and F_2, as well as the preprocessing field F_0, are shunting competitive networks that contrast-enhance and normalize their activation patterns.

2.1. The Preprocessing Field F_0

We will now outline how an M-dimensional input vector $\mathbf{I}^0$ is transformed at F_0 and F_1. All equations describe the steady-state values of a corresponding system of differential equations (Carpenter & Grossberg, 1987b). Each layer of the F_0 and F_1 short-term memory (STM) fields carries out two computations: intrafield and interfield inputs to that layer are summed; and the resulting activity vector is then normalized. At the lower layer of F_0, vector $\mathbf{w}^0$ is the sum of an input vector $\mathbf{I}^0$ and the internal feedback

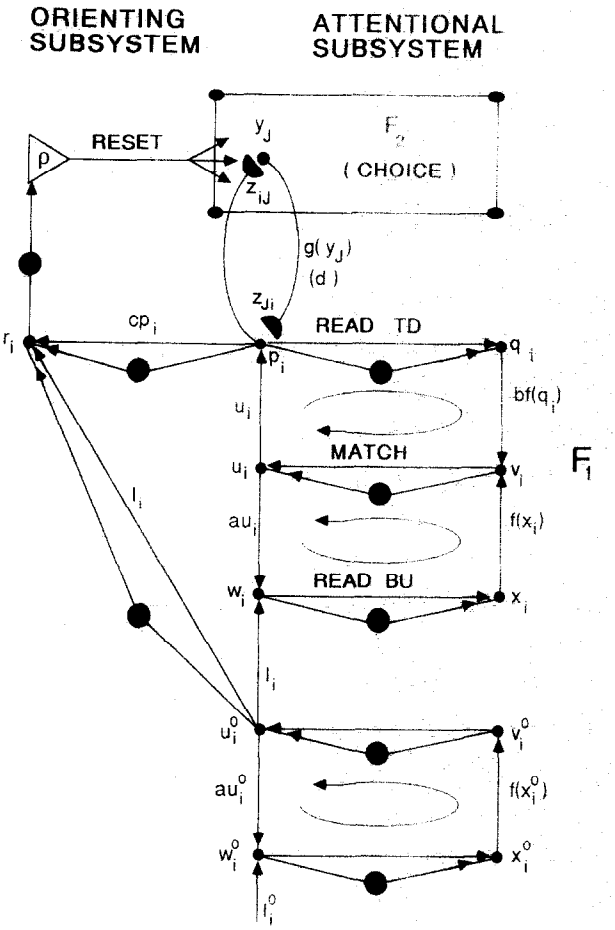

FIGURE 1. ART 2 architecture. Large filled circles represent normalization operations carried out by the network. Adapted from Carpenter and Grossberg (1987b, Figure 10).

signal vector $a\mathbf{u}^0$, so that

$$\mathbf{w}^0 = \mathbf{I}^0 + a\mathbf{u}^0. \tag{1}$$

Next this vector is normalized to yield

$$\mathbf{x}^0 = \mathfrak{N}\mathbf{w}^0 \tag{2}$$

where the operator

$$\mathfrak{N}\mathbf{w}^0 \equiv \frac{\mathbf{w}^0}{\|\mathbf{w}^0\|} \tag{3}$$

carries out Euclidean normalization. This normalization step, denoted by large filled circles in Figure 1, corresponds to the effects of shunting inhibition in the competitive system of differential equations that describe the full F_0 dynamics. Next, $\mathbf{x}^0$ is transformed to $\mathbf{v}^0$ via a nonlinear signal function defined by

$$\mathbf{v}^0 = \mathfrak{S}_0\mathbf{x}^0, \tag{4}$$

where

$$(\mathfrak{S}_0\mathbf{x}^0)_i \equiv f(x_i^0) = \begin{cases} x_i^0 & \text{if } x_i^0 > \theta \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The threshold θ is assumed to satisfy the constraints

$$0 < \theta \leq \frac{1}{\sqrt{M}}, \tag{6}$$

so that the M-dimensional vector $\mathbf{v}^0$ is always nonzero if $\mathbf{I}^0$ is nonuniform. If θ is made somewhat larger than $1/\sqrt{M}$, input patterns that are nearly uniform will not be stored in STM.

The nonlinearity of the function f, embodied in the positive threshold θ, is critical to the contrast enhancement and noise suppression functions of the STM field. Subthreshold signals are set to zero, while suprathreshold signals are amplified by the subsequent normalization step at the top F_0 layer, which sets

$$\mathbf{u}^0 = \mathfrak{N}\mathbf{v}^0. \tag{7}$$

As shown in Figure 1, vector $\mathbf{u}^0$ equals that the output vector from field F_0 to the orienting subsystem, the internal F_0 feedback signal in (1), and the input vector $\mathbf{I}$ to field F_1:

$$\mathbf{I} = \mathbf{u}^0. \tag{8}$$

2.2. The Input Representation Field F_1

The $F_0 \rightarrow F_1$ input vector $\mathbf{I}$ reaches asymptote after a single F_0 iteration, as follows. Initially all STM variables are zero, so $\mathbf{w}^0 = \mathbf{I}^0$ when $\mathbf{I}^0$ is first presented, by (1). Eqns (3)–(5) next imply that

$$v_i^0 = \begin{cases} I_i^0/\|\mathbf{I}^0\| & \text{if } I_i^0 > \theta\|\mathbf{I}^0\| \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Let Ω denote the *suprathreshold index set*, defined by

$$\Omega = \{i: I_i^0 > \theta\|\mathbf{I}^0\|\}. \tag{10}$$

By (7) and (9), there is a constant $K > 1/\|\mathbf{I}^0\|$ such that

$$u_i^0 = \begin{cases} KI_i^0 & \text{if } i \in \Omega \\ 0 & \text{if } i \notin \Omega \end{cases} \tag{11}$$

on the first F_0 iteration. Next, by (1),

$$w_i^0 = \begin{cases} I_i^0(1 + aK) & \text{if } i \in \Omega \\ I_i^0 & \text{if } i \notin \Omega. \end{cases} \tag{12}$$

Thus, at the second iteration, the suprathreshold portion of $\mathbf{w}^0$ (where $i \in \Omega$) is amplified. The subsequent normalization (2) therefore attenuates the subthreshold portion of the pattern. Hence, the suprathreshold index set remains equal to Ω on the second iteration, and the normalized vector $\mathbf{u}^0$ is unchanged so long as $\mathbf{I}^0$ remains constant. In summary, the $F_0 \rightarrow F_1$ input $\mathbf{I}$ is given by

$$\mathbf{I} \equiv \mathfrak{N}\mathfrak{F}_0\mathfrak{N}\mathbf{I}^0 \tag{13}$$

after a single F_0 iteration. Note that

$$I_i > 0 \quad \text{iff } i \in \Omega, \tag{14}$$

and

$$I_i = 0 \quad \text{iff } i \notin \Omega, \tag{15}$$

where Ω is defined by (10).

The F_0 preprocessing stage is designed to allow ART 2 to satisfy a fundamental ART design constraint; namely, an input pattern must be able to instate itself in F_1 STM, *without triggering reset*, at least until an F_2 category representation becomes active and sends top-down signals to F_1 (Carpenter & Grossberg, 1987a). As described in Section 2.8, the orienting subsystem has the property that no reset occurs if vectors $\mathbf{I}$ and $\mathbf{p}$ are parallel (Figure 1). We will now see that, in fact, $\mathbf{p}$ equals $\mathbf{I}$ so long as F_2 is inactive.

As in F_0, each F_1 layer sums inputs and normalizes the resulting vector. The operations at the two lowest F_1 layers are the same as those of the two F_0 layers. At the top F_1 layer $\mathbf{p}$ sums both the internal F_1 signal $\mathbf{u}$ and all the $F_2 \rightarrow F_1$ filtered signals. That is,

$$p_i = u_i + \sum_j g(y_j) z_{ji}, \tag{16}$$

where $g(y_j)$ is the output signal from the jth F_2 node and z_{ji} is the LTM trace in the path from the jth F_2 node to the ith F_1 node.

2.3. The Category Representation Field F_2

If F_2 is inactive, all $g(y_j) = 0$, so (16) implies

$$\mathbf{p} = \mathbf{u}. \tag{17}$$

An active F_2 competitive field is said to be designed to make a *choice* if only one node ($j = J$) has suprathreshold STM. This is the node that receives the largest total input from F_1. In this case $g(y_J)$ equals a constant d, and the sum in eqn (16) reduces to a single term:

$$p_i = u_i + dz_{Ji}. \tag{18}$$

2.4. F_1 Invariance When F_2 is Inactive

Whether or not F_2 is active, the F_1 vector $\mathbf{p}$ is normalized to $\mathbf{q}$ at the top F_1 layer. At the middle layer, vector $\mathbf{v}$ sums intrafield inputs from the bottom layer, where the $F_0 \rightarrow F_1$ bottom-up input $\mathbf{I}$ is read in, and from the top layer, where the $F_2 \rightarrow F_1$ top-down input is read in. Thus

$$v_i = f(x_i) + bf(q_i), \tag{19}$$

where f is defined as in eqn (5).

Let us now compute the F_1 STM values that evolve when $\mathbf{I}$ is first presented, with F_2 inactive. First, $\mathbf{w}$

(Figure 1) equals **I**. By (13), **x** also equals **I**, since **I** is already normalized. Next, (5), (14), (15), and (19) imply that **v**, too, equals **I**, on the first iteration, when **q** still equals **0**. Similarly, $\mathbf{u} = \mathbf{p} = \mathbf{q} = \mathbf{I}$. On subsequent iterations **w** and **v** are amplified by intrafield feedback, but all F_1 STM nodes remain proportional to **I** so long as F_2 remains inactive.

2.5. F_1 Invariance During New Code Learning

With **p** equal to **I**, ART 2 satisfies the design constraint that no reset occur when F_2 is inactive. Another ART design constraint specifies that there be no reset when a new F_2 category representation becomes active. That is, no reset should occur when the LTM traces in paths between F_1 and an active F_2 node have not been changed by pattern learning on any prior input presentation. When F_2 is designed to make a choice and when the active F_2 node with index $j = J$ has never been active previously, we say that the active node is *uncommitted*. After learning occurs, this node is said to be *committed*.

Suppose that the active F_2 node is uncommitted. One ART 2 system hypothesis specifies that the top-down LTM traces are initially equal to zero. Recall that $\mathbf{p} = \mathbf{I}$ when F_2 is inactive. By (18), **p** remains equal to **I** immediately after F_2 becomes active as well. The no-reset constraint will continue to be satisfied if the ART 2 learning laws are chosen so that **p** remains proportional to **I** during learning by an uncommitted node. We will now see that this is the case.

The ART 2 top-down adaptive filter is composed of a set of *outstars* (Grossberg, 1967). That is, when the Jth F_2 node is active, top-down weights in paths fanning out from node J learn the activity pattern at the border of this star-like formation. In ART 2, an active $F_2 \rightarrow F_1$ outstar learns the F_1 activity pattern. That is, while the Jth F_2 node is active

$$\frac{dz_{Ji}}{dt} = p_i - z_{Ji}. \tag{20}$$

By (18), therefore,

$$\frac{dz_{Ji}}{dt} = (1 - d)\left[\frac{u_i}{1 - d} - z_{Ji}\right], \tag{21}$$

where $0 < d < 1$. At the start of learning, **u** equals **I**. Since p_i is a linear combination of u_i and z_{Ji}, p_i will remain proportional to I_i during learning by an uncommitted node if z_{Ji} remains proportional to u_i. By (21), this will be true since the $F_2 \rightarrow F_1$ LTM traces from an uncommitted node are initially zero.

In summary, during learning by an uncommitted node J, the normalized F_1 STM vectors **q**, **u**, and **x** remain identically equal to **I**, while the remaining STM vectors **p**, **v**, and **w** remain proportional to **I**. During ART 2 learning, moreover, the top-down LTM weight vector approaches **p**. By (18), when J is an uncommitted node, the norm of **p** rises from 1 toward $1/(1 - d)$. By (20), the norm of the top-down LTM weight vector rises from zero toward $1/(1 - d)$ while

$$z_{Ji} \rightarrow \frac{I_i}{1 - d}. \tag{22}$$

2.6. F_2 Activation: Code Selection

The $F_2 \rightarrow F_1$ input is a sum of weighted path signals, as in (16). The $F_1 \rightarrow F_2$ input is also a sum of weighted path signals, the input to the jth F_2 node being proportional to the sum

$$\sum_i p_i z_{ij}. \tag{23}$$

When F_2 is inactive, the $F_1 \rightarrow F_2$ input is proportional to

$$\sum_i I_i z_{ij}. \tag{24}$$

When F_2 is designed to make a choice, the Jth node becomes active if

$$\sum_i I_i z_{iJ} = \max_j \left\{\sum_i I_i z_{ij}\right\}. \tag{25}$$

In ART 2, all $F_1 \rightarrow F_2$ LTM traces to an uncommitted node are initially chosen randomly around a constant value. This constant needs to be small enough so that, after learning, an input will subsequently select its own category node over an uncommitted node. Larger values of this constant bias the system toward selection of an uncommitted node over another node whose LTM vector only partially matches the input. The initial choice of LTM values includes small random noise so that not all terms (24) to uncommitted nodes are exactly equal.

2.7. $F_1 \rightarrow F_2$ Learning

If an uncommitted node does become active, **p** remains proportional to **I** throughout learning (Section 2.5). The top-down filter performs outstar learning (20). The bottom-up filter performs *instar* learning (Grossberg, 1976a), which is dual to outstar learning in the sense that, when the Jth F_2 node is active, bottom-up weights in paths fanning in to node J learn the activity pattern from the border into the center of this star-like formation. In ART 2, an active $F_1 \rightarrow F_2$ instar learns the F_1 activity pattern. That is, while the Jth F_2 node is active

$$\frac{dz_{iJ}}{dt} = p_i - z_{iJ}. \tag{26}$$

Thus if J is an uncommitted node,

$$z_{iJ} \rightarrow \frac{I_i}{1 - d} \tag{27}$$

during learning, as in (22) for the top-down LTM traces.

2.8. Match and Reset

While the initial F_2 node selection is determined by (25), the LTM trace pattern of the chosen category may or may not be considered a good enough pattern match to the input $\mathbf{I}$. If not, the orienting subsystem resets the active category, thus protecting that category from adventitious recoding. The match and reset process proceeds as follows.

Let $\mathbf{z}_J$ denote the vector of top-down LTM traces. The vector $\mathbf{r}$ (Figure 1) monitors the degree of match between the F_1 bottom-up input $\mathbf{I}$ and the top-down input $d\mathbf{z}_J$. System reset occurs iff

$$\|\mathbf{r}\| < \rho, \tag{28}$$

where ρ is a dimensionless *vigilance parameter* between 0 and 1. Vector $\mathbf{r}$ obeys the equation

$$\mathbf{r} = \frac{\mathbf{I} + c\mathbf{p}}{\|\mathbf{I}\| + \|c\mathbf{p}\|}, \tag{29}$$

where $c > 0$. Thus

$$\|\mathbf{r}\| = \frac{[\|\mathbf{I}\|^2 + 2c\|\mathbf{I}\|\,\|\mathbf{p}\|\cos(\mathbf{I}, \mathbf{p}) + c^2\|\mathbf{p}\|^2]^{1/2}}{\|\mathbf{I}\| + c\|\mathbf{p}\|}. \tag{30}$$

If $\mathbf{p}$ is proportional to $\mathbf{I}$, $\|\mathbf{r}\| = 1$, so reset does not occur. This is always the case when J is an uncommitted node (Section 2.5).

Suppose, on the other hand, that J is a committed node. By (21), $\mathbf{z}_J$ has previously converged toward the vector $\mathbf{p} = \mathbf{u}/(1 - d)$ which was active at F_1 when node J was active at F_2. We will illustrate how $\|\mathbf{r}\|$ reflects the degree of match between $\mathbf{I}$ and $\mathbf{z}_J$ by analyzing a special case of ART 2 dynamics. Consider the fast-learn limit, in which LTM convergence is complete on each input presentation, and assume that parameter d is close to 1. Then, in the sum

$$\mathbf{p} = \mathbf{u} + d\mathbf{z}_J, \tag{31}$$

the norm of the first term on the right is 1 while the norm of the second term is $d/(1 - d)$, which is much greater than 1. In this case,

$$\mathbf{p} \approx d\mathbf{z}_J. \tag{32}$$

Then, since $\|\mathbf{I}\| = 1$ and $\|\mathbf{p}\| \approx d/(1 - d)$, (30) and (31) imply that

$$\|\mathbf{r}\| \approx \frac{[1 + 2\sigma\cos(\mathbf{I}, \mathbf{z}_J) + \sigma^2]^{1/2}}{1 + \sigma}, \tag{33}$$

where

$$\sigma \equiv \frac{cd}{1 - d}. \tag{34}$$

Thus $\|\mathbf{r}\|$ is an increasing function of $\cos(\mathbf{I}, \mathbf{z}_J)$ such that

$$\frac{(1 + \sigma^2)^{1/2}}{1 + \sigma} \leq \|\mathbf{r}\| \leq 1, \tag{35}$$

and $\|\mathbf{r}\| = 1$ iff $\cos(\mathbf{I}, \mathbf{z}_J) = 1$. In fact, by (28) and (33), reset occurs iff

$$\cos(\mathbf{I}, \mathbf{z}_J) < \rho^*, \tag{36}$$

where

$$\rho^* = \frac{\rho^2(1 + \sigma)^2 - (1 + \sigma^2)}{2\sigma}. \tag{37}$$

Note that $\rho^* = 1$ iff $\rho = 1$ and that $\rho^* < 0$ if $\rho = 0$. Since all components of $\mathbf{I}$ and $\mathbf{z}_J$ are non-negative, reset never occurs if $\rho^* \leq 0$, thereby eliminating the search/reset process altogether. On the other hand, reset would always occur if ρ^* were greater than 1. Thus, by hypothesis, $0 \leq \rho^* \leq 1$.

Remark. ART 2 includes the additional constraint

$$\sigma \leq 1. \tag{38}$$

This implies that $\|\mathbf{r}\|$ in (33) is a decreasing function of σ for each fixed value of $\cos(\mathbf{I}, \mathbf{z}_J)$ (Carpenter & Grossberg, 1987b, Figure 7). In ART 2, (38) implies that, during fast or slow learning, $\|\mathbf{r}\|$ in (30) decreases as $\|\mathbf{z}_J\|$ increases, all other things being equal. This corresponds to the idea that $\|\mathbf{z}_J\|$ reflects the *degree of commitment* of category J. For a given pattern match, i.e., for a fixed value of $\cos(\mathbf{I}, \mathbf{p})$, the matching criterion defined jointly by (28) and (30) becomes stricter as $\|\mathbf{z}_J\|$ grows toward its asymptotic limit of $d/(1 - d)$. In fast learning, this limit is reached on a single input presentation. With slow learning, constraint (38) implies that more learning by a committed node carries a greater tendency for mismatched bottom-up and top-down vectors to trigger reset and hence greater permanence of that node's category LTM representation. For both the fast learning and the intermediate learning cases considered below, $\|\mathbf{z}_J\| \approx d/(1 - d)$ once J becomes a committed node. This is why constraint (38) does not appear in the ART 2-A algorithm.

2.9. Search and Resonance

Once one F_2 node is reset, ART 2 activates the F_2 node J with the next highest input (24). As above, the search process will cease if J is uncommitted. Among committed nodes, the order of search is determined by the product of the norm of the bottom-up LTM vector times the cosine of the angle between

I and that vector. With slow learning, bottom-up weights may be small if little coding has already occurred at that node. In this case an extended search may ensue. However, in the special case where weights are normalized by the end of each input presentation, the search process may be replaced by an abbreviated algorithm, as follows. Note first that the bottom-up weight vector of each committed node j equals the corresponding top-down weight vector $\mathbf{z}_j$, by (20) and (26). By (24) the order of search among committed nodes is determined by the size of terms

$$\|\mathbf{I}\| \, \|\mathbf{z}_j\| \cos(\mathbf{I}, \mathbf{z}_j). \tag{39}$$

The order of search therefore depends on $\cos(\mathbf{I}, \mathbf{z}_j)$ alone, since $\|\mathbf{I}\| = 1$ and $\|\mathbf{z}_j\| = 1/(1 - d)$. By (36), if the first chosen node resets then all other committed nodes will also reset if chosen. Eventually, either an uncommitted node will be chosen and coded, or, if no uncommitted nodes remain, the system has exceeded its capacity and the input $\mathbf{I}^0$ is not coded. Thus if one reset occurs, algorithmic search immediately selects an uncommitted node at random.

In all cases, *resonance* is the state in which the system retains a constant code representation over a time interval that is long relative to the transient time scale of F_2 activation and search.

2.10. ART 2 Fast Computation

The abbreviated ART 2 search process described in Section 2.9 is insufficient in general. Search of committed nodes may be necessary with slow learning, in order to allow a given input access to a given node, until weights grow toward their asymptotic size. In addition, the ART reset process is used for other functions besides search: It can signal the presence of a new input for classification, or it can be modulated by reinforcing or other evaluative inputs. These various cases, as well as a neural implementation of the search process, are the primary focus of ART 3 (Carpenter & Grossberg, 1990).

The purpose of the present article, in contrast, is to consider cases in which ART 2 dynamics can be approximated by efficient algorithms, such as the fast-search algorithm of Section 2.9. One of these special cases is the fast-learn limit. However, fast learning may be too drastic for certain applications, as when the input-set is degraded by high noise levels. ART 2 slow learning is better able to cope with noise, but has not previously been amenable to rapid computation. In the present article, we develop an efficient algorithm that approximates ART 2 dynamics not only for fast learning but also for a much larger set of cases that we here call *intermediate learning*. Intermediate learning permits partial recoding of the LTM vectors on each input presentation, thus retaining the increased noise-tolerance of slow learning. In addition, however, an ART 2 intermediate learning system operates in a range where algorithmic approximations enable rapid computation. Dynamics of ART 2 with both fast learning and intermediate learning are approximated by the algorithmic system ART 2-A described in Section 3.

3. ART 2-A

3.1. Fast Learning With Linear STM Feedback

ART 2-A approximates the STM and LTM dynamics of an ART 2 system with choice at F_2. The ART 2-A equations are partially motivated by the following theorem about fast-learn ART 2 with the signal function threshold θ set equal to 0 in F_0 and F_1. Note that the key ART 2 hypothesis (6) is violated here, and the F_1 signal function therefore is linear.

Theorem 1 states that when the F_1 feedback function has zero threshold, the LTM vectors of the active category approach a vector proportional to **I**. In fast learning, the system retains no trace of previous inputs coded in this category.

THEOREM 1. *Consider fast-learn ART 2 with the F_1 signal threshold θ set equal to 0. Then, after an F_2 node J has coded an input **I**, both bottom-up and top-down LTM vectors are proportional to **I**. In fact*

$$\mathbf{z}_J = \frac{1}{1 - d}\mathbf{I}. \tag{40}$$

Theorem 1 is proved in the Appendix.

Remark. Figure 8(e) of Carpenter and Grossberg (1987b) shows an ART 2 simulation with $\theta = 0$, in which nonzero components of LTM vectors after learning retain traces of previous inputs rather than fully tracking the relative values of the current input, in contradiction to Theorem 1. That simulation illustrates an intermediate learning situation in which LTM traces are approaching, but have not yet reached, equilibrium when a committed node is chosen. Some of these traces approach zero when the current input component is zero. With $\theta = 0$, the ART 2 system allows traces that are approaching zero, but have not reached it, to grow again during subsequent input presentations.

3.2. Fast Learning With Nonlinear STM Feedback

Consider now a fast-learn ART 2 system with $\theta > 0$, and hence the nonlinear signal function f (5) at F_0 and F_1. As in Section 2.8, assume that parameter d is close to 1, so that $\mathbf{p} \approx d\mathbf{z}_J$ when a committed node J is active, as in (32). In this case, to a first approximation,

$$q_i \le \theta \quad \text{iff} \quad z_{Ji} \le \theta/(1 - d). \tag{41}$$

where $\mathbf{q}$ is the normalized STM vector in the top F_1 layer (Figure 1). When $q_i \leq 0$, $f(q_i) = 0$ in (19). The ART 2 internal F_1 feedback parameters a and b are assumed to be large enough so that, if the ith F_1 node receives no top-down amplification via $f(q_i)$, then STM at that node is quenched, even if I_i is relatively large. As in (41), this property allows the system to satisfy the ART design constraint that, once a trace z_{Ji} falls below a certain positive value, it will decay permanently to zero.

In (10), we defined an index set Ω which has the property that $i \in \Omega$ iff $I_i > 0$. The preceding discussion leads us now to define analogous index sets Ω_J. During resonance on a given input presentation in which the *committed* node J is active, let

$$i \in \Omega_J \quad \text{iff} \quad z_{Ji}^{(\text{old})} > \frac{\theta}{1 - d}, \tag{42}$$

where $\mathbf{z}_J^{\text{old}}$ denotes the top-down LTM vector at the *start* of the input presentation. Intuitively, Ω_J is the index set of "critical features" that define category J. Set Ω_J corresponds approximately to the ART 1 template index set $\mathbf{V}^{(J)}$ (Carpenter & Grossberg, 1987a).

Since all features can *a priori* be coded by an *uncommitted* node, each set

$$\Omega_J = \{i : i = 1, 2, \ldots, M\} \tag{43}$$

on the first input presentation in which node J is active.

In fast-learn ART 2, the set Ω_J can shrink when J is active, but Ω_J can never grow. This monotonicity property is necessary for overall code stability. On the other hand, z_{Ji} learning is still possible for $i \in \Omega_J$ when J is active. This observation leads to the following conjecture.

CONJECTURE 1. *Consider fast-learn ART 2, with $\theta > 0$, when an F_2 node J is coding a fixed F_1 input $\mathbf{I}$. Let Ω denote the $F_0 \rightarrow F_1$ input index set*

$$\Omega = \{i : I_i > 0\}, \tag{44}$$

which in ART 2 is equivalent to

$$\Omega = \{i : I_i > \theta\}. \tag{45}$$

Let Ω_J denote the category index set, as follows. If J is an uncommitted node, let

$$\Omega_J = \{i : i = 1, 2, \ldots, M\}. \tag{46}$$

If J is a committed node, let

$$\Omega_J = \{i : z_{Ji}^{(\text{old})} > 0\} \tag{47}$$

where $z_{Ji}^{(\text{old})}$ denotes the $F_2 \rightarrow F_1$ LTM vector at the start of the input presentation. In ART 2, (47) is equivalent to

$$\Omega_J = \left\{i : z_{Ji}^{(\text{old})} > \frac{\theta}{1 - d}\right\}. \tag{48}$$

Define the vector $\boldsymbol{\Psi}$ by

$$\Psi_i = \begin{cases} I_i & \text{if } i \in \Omega_J \\ 0 & \text{if } i \notin \Omega_J. \end{cases} \tag{49}$$

Then, during learning, both the bottom-up and the top-down LTM vectors approach a limit vector proportional to $\boldsymbol{\Psi}$. At the end of the input presentation,

$$\mathbf{z}_J = \mathbf{z}_J^{(\text{new})} = \frac{\mathfrak{N}\boldsymbol{\Psi}}{1 - d}. \tag{50}$$

Moreover

$$\Omega_J^{(\text{new})} = \Omega_J^{(\text{old})} \cap \Omega. \tag{51}$$

By characterizing fast-learn ART 2 system dynamics, Conjecture 1 directly motivates the fast-learn limit of the ART 2-A algorithm. On a given input presentation, the algorithm partitions the F_1 index set into two classes, and defines different dynamic properties for each class. If $i \notin \Omega_J$, z_{Ji} remains equal to 0 during learning; that is, it retains its memory of the past, independent of the present F_1 input I_i. In contrast, if $i \in \Omega_J$, z_{Ji} nearly forgets the past by becoming proportional to I_i. The only reflection of past learning for $i \in \Omega_J$ is in the proportionality constant.

3.3. Intermediate Learning: Fast Commitment With Slow Recoding

The fast-learn limit is important for system analysis and is useful in many applications. However, a finite learning rate is often desirable in ART 2 to increase stability and noise tolerance, and to make the category structure less dependent on input presentation order. Here, we consider intermediate learning rates, which provide these advantages, and show how they can be approximated by an ART 2-A algorithm that includes fast learning as a limiting case.

The ART 2-A intermediate learning algorithm embodies the properties of fast commitment and slow recoding. These properties are based on an analysis of ART 2 dynamics. In particular, the ART 2 LTM vectors tend to approach asymptote much more quickly when the active node J is uncommitted than when J is committed; and once J is committed, $\|\mathbf{z}_J\|$ stays close to $1/(1 - d)$. For convenience let $\mathbf{z}_J^*$ denote the scaled LTM vector

$$\mathbf{z}_J^* \equiv (1 - d)\mathbf{z}_J. \tag{52}$$

The approximations (i)–(iii) below characterize the value of $\mathbf{z}_J^*$ at the end of an input presentation during which the F_2 node J is in resonance:

(i) If J is an uncommitted node, $\mathbf{z}_J^*$ is set equal to $\mathbf{I}$.

(ii) If J is a committed node, $\mathbf{z}_J^*$ is set equal to a convex combination of its previous value and the vector $\mathfrak{N}\boldsymbol{\Psi}$ defined by (3) and (49).

(iii) $\mathbf{z}_J^*$ is renormalized so that its magnitude always equals 1.

The fast-learn limit corresponds to setting $\mathbf{z}_J^*$ equal to $\mathfrak{N}\mathbf{\Psi}$ in (ii). Slower ART 2 learning corresponds to keeping $\mathbf{z}_J^*$ closer to its previous value in (ii). Previous simplified versions of ART 2, such as that of Ryan (1988), have included computations similar to setting $\mathbf{z}_J^*$ equal to a convex combination of $\mathbf{I}$ and the previous $\mathbf{z}_J^*$ vector. ART 2-A uses $\mathfrak{N}\mathbf{\Psi}$ in (ii), rather than $\mathbf{I}$. The vector $\mathbf{\Psi}$, defined by equation (49), endows ART 2-A with the critical stability properties of ART 2.

The existence of distinct ART 2 operating modes, fast commitment and slow recoding, can be explained as follows. By (21) and (52),

$$\frac{d\mathbf{z}_J^*}{dt} = (1 - d)(\mathbf{u} - \mathbf{z}_J^*). \tag{53}$$

By (53), $\mathbf{z}_J^*$ approaches $\mathbf{u}$ at a fixed rate. As described in Section 2.5, when J is an uncommitted node, $\mathbf{u}$ remains identically equal to $\mathbf{I}$ throughout the input presentation. Thus vector $\mathbf{z}_J^*$ approaches $\mathbf{I}$ exponentially, and $\mathbf{z}_J^* \approx \mathbf{I}$ at the end of the input presentation if the presentation interval is long relative to $1/(1 - d)$. On the other hand, if J is a committed node, as in Section 2.8, $\mathbf{u}$ is close to $\mathbf{z}_J^*$. In other words,

$$\mathbf{u} = \mathfrak{N}(\varepsilon\mathfrak{N}\mathbf{\Psi} + (1 - \varepsilon)\mathbf{z}_J^*), \tag{54}$$

where $\mathbf{\Psi}$ is defined by (49) and $0 < \varepsilon \ll 1$. Since ε is small,

$$\mathbf{u} \approx \varepsilon\mathfrak{N}\mathbf{\Psi} + (1 - \varepsilon)\mathbf{z}_J^*. \tag{55}$$

Thus, (53) and (55) imply

$$\frac{d\mathbf{z}_J^*}{dt} \approx \varepsilon(1 - d)(\mathfrak{N}\mathbf{\Psi} - \mathbf{z}_J^*). \tag{56}$$

Hence, $\mathbf{z}_J^*$ begins to approach $\mathfrak{N}\mathbf{\Psi}$ at a rate that is slower, by a factor ε, than the rate of convergence of an uncommitted node. In ART 2, the size of ε is determined by the parameters a and b (Figure 1). The normal ART 2 parameter constraints that a and b be large conspire to make ε small.

In summary, if the ART 2 input presentation time is large relative to $1/(1 - d)$, the LTM vectors of an uncommitted node J converge to $\mathbf{I}$ on the first activation of that node. Subsequently, the LTM vectors remain approximately equal to a vector $\mathbf{z}_J$, where

$$(1 - d)\,\|\mathbf{z}_J\| \equiv \|\mathbf{z}_J^*\| \approx 1. \tag{57}$$

Because $\mathbf{z}_J^*$ is normalized when J first becomes committed, and, by (53), it approaches $\mathbf{u}$, which is both normalized and approximately equal to $\mathbf{z}_J^*$, $\mathbf{z}_J^*$ remains approximately normalized during learning. Thus, the rapid-search algorithm (Section 2.9) remains valid for intermediate learning as well as for fast learning. Finally, (53) and (54) suggest that a (normalized) convex combination of the $\mathfrak{N}\mathbf{\Psi}$ and $\mathbf{z}_J^*$ vector values at the start of an input presentation gives a reasonable first approximation to $\mathbf{z}_J^*$ at the end of the presentation. The ART 2-A algorithm summarized in the next section includes both the fast and the intermediate learning cases.

3.4. Summary of the ART 2-A Algorithm

Eqs (58)–(70) summarize the ART 2-A system for both intermediate and fast learning rates. The heart of the ART 2-A algorithm is an update rule that adjusts LTM weights in a single step for each presentation interval during which the input vector is held constant.

Input

Given a nonuniform M-dimensional input vector $\mathbf{I}^0$ to F_0, the input $\mathbf{I}$ to F_1 satisfies

$$\mathbf{I} = \mathfrak{N}\mathfrak{T}_0\mathfrak{N}\mathbf{I}^0 \tag{58}$$

where

$$\mathfrak{N}\mathbf{x} \equiv \frac{\mathbf{x}}{\|\mathbf{x}\|}, \tag{59}$$

and

$$(\mathfrak{T}_0\mathbf{x})_i \equiv \begin{cases} x_i & \text{if } x_i > \theta \\ 0 & \text{otherwise.} \end{cases} \tag{60}$$

Threshold θ in (60) satisfies the inequalities

$$0 < \theta \leq 1/\sqrt{M}. \tag{61}$$

Eqns (58)–(61) imply that $\mathbf{I}$ is nonzero.

F_2 activation

The input to the jth F_2 node is given by

$$T_j = \begin{cases} \alpha \Sigma_i I_i & \text{if } j \text{ is an uncommitted node} \\ \mathbf{I} \cdot \mathbf{z}_j^* & \text{if } j \text{ is a committed node.} \end{cases} \tag{62}$$

The constant α in (62) satisfies

$$\alpha \leq \frac{1}{\sqrt{M}}. \tag{63}$$

Intially, all F_2 nodes are *uncommitted*. The set of committed F_2 nodes and the scaled LTM vectors $\mathbf{z}_j^*$ are defined iteratively below.

Choice function

The initial choice at F_2 is one node with index J satisfying

$$T_J = \max_j(T_j). \tag{64}$$

If more than one node is maximal, choose one at random. After an input presentation on which node J is chosen, J becomes *committed*.

Resonance or reset

The node J initially chosen by (64) remains constant if J is uncommitted or if J is committed and

$$T_J \geq \rho^*, \tag{65}$$

where ρ^* is constrained so that

$$0 \leq \rho^* \leq 1. \tag{66}$$

If J is committed and

$$T_J < \rho^*, \tag{67}$$

then J is reset to the index of an arbitrary uncommitted node. Because the Euclidean norms of $\mathbf{I}$ and $\mathbf{z}_j^*$ are all equal to 1 for committed nodes, T_j in (62) equals the cosine of the angle between $\mathbf{I}$ and $\mathbf{z}_i^*$.

Learning

At the end of an input presentation, $\mathbf{z}_J^*$ is set equal to $\mathbf{z}_J^{*(\text{new})}$ defined by

$$\mathbf{z}_J^{*(\text{new})} = \begin{cases} \mathbf{I} & \text{if } J \text{ is an uncommitted node} \\ \mathfrak{N}(\beta \mathfrak{N} \boldsymbol{\Psi} + (1 - \beta)\mathbf{z}_J^{*(\text{old})}) & \text{if } J \text{ is a committed node} \end{cases} \tag{68}$$

where, if J is a committed node, $\mathbf{z}_J^{*(\text{old})}$ denotes the value of $\mathbf{z}_J^*$ at the start of the input presentation,

$$\boldsymbol{\Psi}_i \equiv \begin{cases} I_i & \text{if } z_{Ji}^{*(\text{old})} > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{69}$$

and

$$0 \leq \beta \leq 1. \tag{70}$$

3.5. Contrast With the Leader Algorithm

The ART 2-A weight update rule (68) for a committed node is similar in form to eqn (54). However, (54) describes the STM vector $\mathbf{u}$ immediately after a node J has become active, before any significant learning has taken place, and parameter ε in (54) is small. ART 2-A approximates a process that integrates the form factor (54) over the entire input presentation interval. Hence, β ranges from 0 to 1 in (70). Setting β equal to 1 gives ART 2-A in the fast-learn limit. Setting β equal to 0 turns ART 2-A into a type of *leader algorithm* (Hartigan, 1975, Ch. 3), with the weight vector $\mathbf{z}_J^*$ remaining constant once J is committed. Small positive values of β yield system properties similar to those of an ART 2 slow learning system. Fast commitment obtains, however, for all values of β. Note that β could vary from one input presentation to the next, with smaller values of β corresponding to shorter presentation intervals and larger values of β corresponding to longer presentation intervals.

Parameter α in (62) corresponds to the initial values of LTM components in an ART 2 $F_1 \rightarrow F_2$ weight vector. As described in Section 2.6, α needs to be small enough, as in (63), so that if $\mathbf{z}_J^* = \mathbf{I}$ for some J, then J will be chosen when $\mathbf{I}$ is presented. Setting α close to $1/\sqrt{M}$ biases the network toward selection of an uncommitted node over category nodes that only partially match $\mathbf{I}$. In the simulations described below, α is set equal to $1/\sqrt{M}$. Thus even when $\rho^* = 0$ and reset never occurs, ART 2-A can establish several categories. Instead of randomly selecting any uncommitted node after reset, the value α for all T_j in (62) could be replaced by any function of j, such as a ramp or random function, that achieves the desired balance between selection of committed and uncommitted nodes and a determinate selection of a definite uncommitted node after a reset event.

4. SIMULATIONS

4.1. Comparative Simulations of ART 2-A and ART 2 Fast-Learn Systems

The simulation summarized in Figure 2 illustrates how ART 2-A groups 50 analog input patterns. The ART 2-A simulation gives a result essentially identical to the simulation result of a fast-learn ART 2 system with comparable parameters. The input set

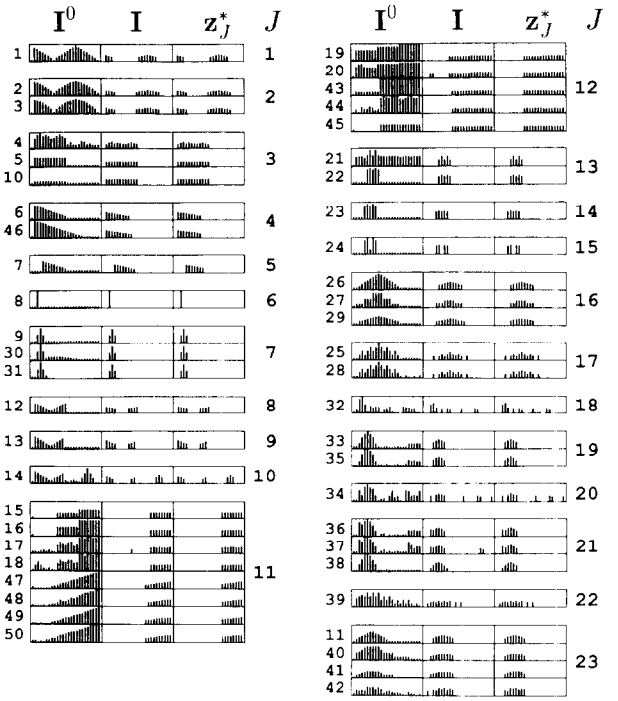

FIGURE 2. ART 2-A fast-learn simulation. I^0 is the input to F_0. I is the input to F_1. z_J^* is the scaled LTM vector of the winning F_2 category node J at the end of each input presentation interval. The numbers in the left column index the input vectors and give their order of presentation. The vertical axes of the inputs I^0 all have the same scale, which is arbitrary due to the initial normalization in F_0. The vertical axes for I and z_J^* run from 0 to 1.

consisted of the 50 patterns used in the original ART 2 simulations (Carpenter & Grossberg, 1987b). The inputs, indexed in the left column of Figure 2, were repeatedly presented in the order $1, 2, \ldots, 50$ until the category structure stabilized.

Table 1 shows the parameters used for one of the fast-learn simulations (Carpenter & Grossberg, 1987b, Figure 11). Since fast-learn LTM components approach but never reach a limit on each input presentation, each ART 2 simulation requires selection of a convergence criterion. As described below, different criteria can produce slight variations in category structure.

The ART 2-A parameters for Figure 2 (see Table 2) correspond to the ART 2 parameters. For example, eqn (37) is used to set $\rho^* = .92058$ when $\rho = .98$ and $\sigma = cd/(1 - d) = .9$. Since ART 2-A gives formula (68) for the LTM limit, no convergence criterion is necessary.

The ART 2 and ART 2-A simulations give identical partitions of the 50 patterns into 23 recognition categories (Figure 2). Each component of the final LTM vectors differs at most by 0.5%. The difference between the two results decreases as the convergence criterion on the ART 2 simulation is tightened.

For both ART 2 and ART 2-A, the category structure stabilizes to its asymptotic state during the second presentation of the entire input set. However, the suprathreshold LTM components continue to track the relative magnitudes of the components in the most recent input. The inputs and final templates of the ART 2-A simulation are shown in Figure 2. Inputs are shown grouped according to the F_2 node category J chosen during the second and subsequent presentations of each input. Category 23 shows how $\mathbf{z}_J^*$ tracks the suprathreshold analog input values in feature set Ω_J while ignoring input values outside that set. The corresponding figure for the ART 2 simulation is indistinguishable from Figure 2.

The earlier ART 2 simulation (Carpenter & Grossberg, 1987b, Figure 11) had one fewer category than Figure 2, even though the model parameters

TABLE 1
ART 2 simulation parameters
(Carpenter & Grossberg, 1987b, Figure 11)

Parameter	Value
M	25
$z_{ij}(0)$	$\frac{1}{(1-d)\sqrt{M}} = 2$
θ	$\frac{1}{\sqrt{M}} = .2$
ρ	.98
a	10
b	10
c	.1
d	.9

TABLE 2
ART 2-A simulation parameters for Figures 2–4

Parameter	Value		
M	25		
α	$\frac{1}{\sqrt{M}} = .2$		
θ	$\frac{1}{\sqrt{M}} = .2$		
	Figure 2	Figure 3	Figure 4
ρ^*	.92058	0	0
β	1	1	.01

were the same as in Table 1. This difference appears to be due to different convergence criteria.

The ART 2-A fast-learn simulation in Figure 2 used only four seconds of Sun 4/110 CPU time to run through the 50 patterns three times. The corresponding ART 2 simulation took 25 to 150 times as long, depending on the fast-learn convergence criterion imposed. This speed-up occurred even using a fast integration method for ART 2, in which LTM values were allowed to relax to equilibrium alternatively with STM variables. Carpenter and Grossberg (1987b) employed a slower integration method, in which LTM values changed only slightly for each STM relaxation. Compared to this latter method, the ART 2-A speed-up is even greater. Finally, integration of the full ART 2 dynamical system would take longer still.

4.2. Comparative Simulations of ART 2-A Fast-Learn and Intermediate-Learn Systems

Simulation results of ART 2-A with fast learning (Figure 3) and intermediate learning (Figure 4) use the same 50 input patterns as in Figure 2, but the inputs are now presented randomly, rather than cyclically. This random presentation regime simulates a statistically stationary environment in which each member of a fixed set of patterns is encountered with equal probability at any given time. In addition, ρ^* was set to zero in these simulations, making the number of categories more dependent on parameter α than when ρ^* is larger. Other parameters are given in Table 2.

Figures 3 and 4 show the asymptotic category structure and scaled LTM weight vectors established after an initial transient phase of 2,000 to 3,000 input presentations. Figure 3 illustrates that category nodes may occasionally be abandoned after a transient encoding phase (see nodes $J = 1$, 6, and 7). Figure 3 also includes a single input pattern (39) that appears in two categories ($J = 12$ and 15). In the simulation, input 39 was usually placed in category 12. However, when the most recent input to category

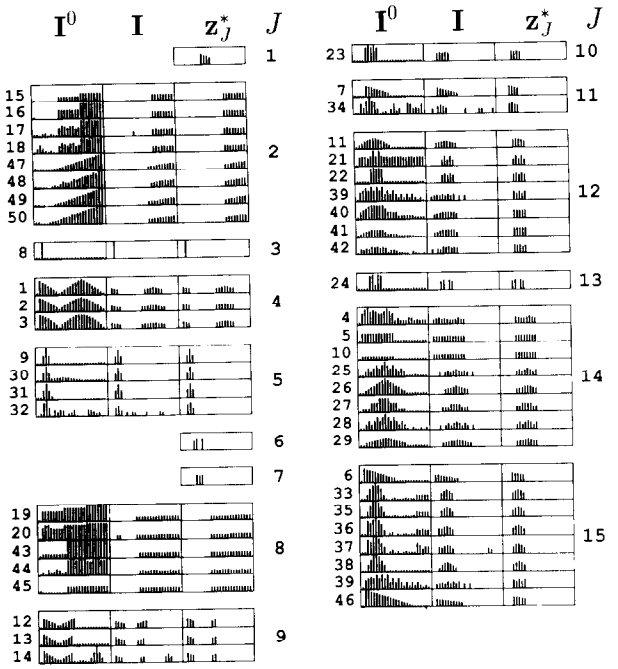

FIGURE 3. ART 2-A fast-learn simulation. Input presentation order is random and $\rho^* = 0$. Otherwise the system is the same as in Figure 2. The three categories ($J = 1$, 6, and 7) showing no inputs were coded only during early presentations. Pattern 39 appears in both categories 12 and 15.

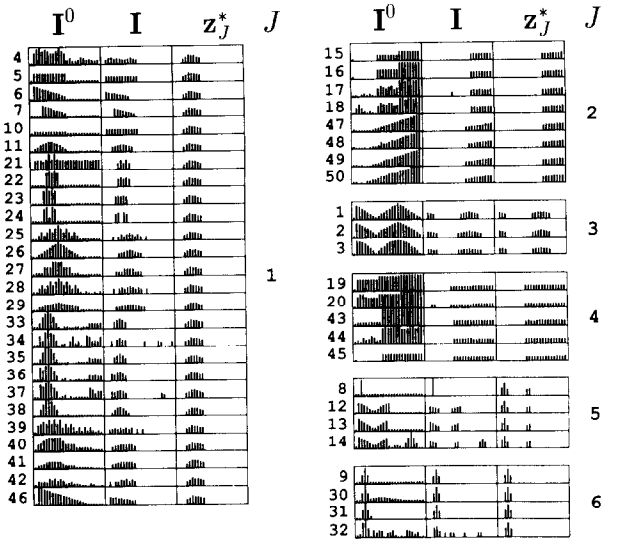

FIGURE 4. ART 2-A intermediate-learn simulation. The learning rate parameter β is set equal to .01. Otherwise the system is the same as in Figure 3, including a zero value of vigilance that leads to coarse, but stable, categories.

12 was pattern 21, category 15 could win in response to input 39, though whether or not it did depended on which pattern category 15 had coded most recently as well. In addition to depending on input presentation order, the instability of pattern 39 is promoted by the system being in the fast-learn limit with a small value of ρ^*, here $\rho^* = 0$. A corresponding ART 2 system gives similar results.

These anomalies did not occur in the intermediate-learn case, in which there is not such drastic recoding on each input presentation. Similarly, intermediate learning copes better with noisy inputs than does fast learning. Figure 4 illustrates an ART 2-A simulation run with the inputs and parameters of Figure 3, except that the learning rate parameter is small ($\beta = .01$). The analog values of the suprathreshold LTM components do not vary with the most recent input nearly as much as the components in Figure 3. A slower learning rate helps ART 2-A to stabilize the category structure by making coding less dependent on order of input presentation.

5. CONCLUSION

ART 2 fast-learn and intermediate-learn systems combine analog and binary coding functions. The analog portion encodes the recent past while the binary portion retains the distant past. On the one hand, LTM traces that fall below threshold remain below threshold at all future times. Thus once a feature is deemed "irrelevant" in a given category, it will remain irrelevant throughout the future learning experiences of that category in that such a feature will never again be encoded into the LTM of that category, even if the feature is present in the input pattern. For example, the color features of a chair may come to be suppressed during learning of the category "chair" if these color features have not been consistently present during learning of this category.

On the other hand, the suprathreshold LTM traces track a time-average of recent input patterns, even while they are being renormalized due to suppression of other components. Intuitively, a feature that is consistently present tracks the most recent amplitudes of that feature, eventually forgetting subtle differences of its past exemplars, much as in word frequency effects, encoding specificity effects, and episodic memory (Mandler, 1980; Underwood & Freund, 1970), which are qualitatively explained in terms of a time-averaged ART learning equation analogous to (68) in Grossberg and Stone (1986).

The ART 2-A algorithm incorporates these coding features while achieving an increase in computational efficiency of two to three orders of magnitude over the full ART 2 system.

REFERENCES

Carpenter, G. A., & Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition

machine. *Computer Vision, Graphics, and Image Processing*, **37**, 54–115.

Carpenter, G. A., & Grossberg, S. (1987b). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, **26**, 4919–4930.

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

Grossberg, S. (1967). Nonlinear difference-differential equations in prediction and learning theory. *Proceedings of the National Academy of Sciences (USA)*, **58**, 1329–1334.

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

Hartigan, J. A. (1975). *Clustering algorithms*. John Wiley & Sons, New York.

Mandler, G. (1980). Recognizing: The judgement of previous occurrence. *Psychological Review*, **87**, 252–271.

Ryan, T. W. (1988). The resonance correlation network. *Proceedings of the IEEE International Conference on Neural Networks*, **I**, 673–680.

Underwood, B. J., & Freund, J. S. (1970). Word frequency and short term recognition memory. *American Journal of Psychology*, **83**, 343–351.

APPENDIX: PROOF OF THEOREM 1

During resonance with the Jth F_2 node active, when the threshold θ equals 0, the ART 2 STM vector $\mathbf{u}$ satisfies the implicit equation

$$\mathbf{u} = \mathfrak{N}\left| \mathfrak{N}\left(\mathbf{I} + a\mathbf{u}\right) + b\mathfrak{N}\left(\mathbf{u} + \frac{d}{1-d}\mathbf{z}\right)\right|. \tag{71}$$

When the scaled LTM vector $\mathbf{z}^* = (1-d)\mathbf{z}$ reaches equilibrium, it equals $\mathbf{u}$. Then, denoting $\mathbf{z} = \mathbf{z}^*$,

$$\begin{aligned} \mathbf{z} &= \mathfrak{N}\left| \mathfrak{N}\left(\mathbf{I} + a\mathbf{z}\right) + b\mathfrak{N}\left(\mathbf{z} + \frac{d}{1-d}\mathbf{z}\right)\right| \\ &= \mathfrak{N}(\mathfrak{N}(\mathbf{I} + a\mathbf{z}) + b\mathbf{z}| \\ &= \left(\frac{\mathbf{I} + a\mathbf{z}}{\|\mathbf{I} + a\mathbf{z}\|} + b\mathbf{z}\right)\left\|\frac{\mathbf{I} + a\mathbf{z}}{\|\mathbf{I} + a\mathbf{z}\|} + b\mathbf{z}\right\|, \end{aligned} \tag{72}$$

from which it follows that

$$(1 - A(Ba + b))\mathbf{z} = AB\mathbf{I} \tag{73}$$

where

$$A = \left\|\frac{\mathbf{I} + a\mathbf{z}}{\|\mathbf{I} + a\mathbf{z}\|} + b\mathbf{z}\right\| \tag{74}$$

and

$$B = \|\mathbf{I} + a\mathbf{z}\|. \tag{75}$$

Since $A \neq 0$ and $B \neq 0$,

$$\mathbf{I} = \frac{(1 - A(Ba + b))}{AB}\mathbf{z}. \tag{76}$$

Since also $\|\mathbf{I}\| = \|\mathbf{z}\| = 1$, it follows from (76) that $\mathbf{I} = \mathbf{z}$, which completes the proof.