Pergamon

Neural Networks, Vol. 10, No. 4, pp. 649–669, 1997

Printed in Great Britain
0893–6080/97 $17.00+.00

PII: S0893-6080(96)00111-6

CONTRIBUTED ARTICLE

Exact ART: A Complete Implementation of an ART Network

MAARTJE E.J. RAIJMAKERS AND PETER C.M. MOLENAAR

University of Amsterdam

(*Received* 13 *January* 1995; *accepted* 22 *August* 1996)

Abstract—*In this article we introduce a continuous time implementation of adaptive resonance theory (ART). ART designed by Grossberg concerns neural networks that self-organize stable pattern recognition categories of arbitrary sequences of input patterns. In contrast to the current implementations of ART we introduce a complete implementation of an ART network, including all regulatory and logical functions, as a system of ordinary differential equations capable of stand-alone running in real time. This means that transient behavior is kept in tact. This implementation of ART is based on ART 2 and is called Exact ART. Exact ART includes an implementation of a gated dipole field and an implementation of the orienting sub-system. The most important features of Exact ART, which are the design principles of ART 2, are proven mathematically. Also simulation studies show that Exact ART self-organizes stable recognition codes that agree with the classification behavior of ART 2. © 1997 Elsevier Science Ltd.*

Keywords—Adaptive resonance theory (ART), Mathematical and computational analysis, Gated dipole field.

1. INTRODUCTION

In the present article we introduce a continuous time implementation of adaptive resonance theory (ART), which is based on Grossberg's and Carpenter's ideas that are elaborated in several articles (Grossberg, 1976a,b, 1980; Carpenter & Grossberg, 1987b). In 1976 Grossberg introduced ART. ART is a mathematical model for the self-organization of stable recognition codes in real time in response to arbitrary sequences of input patterns. The aim of Grossberg (1980) was to develop a mechanism of cognitive coding that is stand-alone. In that article Grossberg does not introduce an implementation of ART, but he describes the basic principles that underlie ART. The development of the model was guided by biological knowledge of the nervous system, (evolutionary) environmental demands, and mathematical transparency. Different implementations of ART in several degrees of complexity have been introduced: ART 1 (Carpenter & Grossberg, 1987a), ART 2 (Carpenter & Grossberg, 1987b), ART 2-A (Carpenter et al., 1991b), ART 3 (Carpenter & Grossberg, 1990), Fuzzy ART (Carpenter et al., 1991a), and variations on adaptive resonance (Ryan & Winter, 1987; Ryan et al., 1987). ART 1 can only handle binary patterns. ART 2, ART 2-A, and ART 3 are more complex and can process analog patterns. Fuzzy ART can also handle analog patterns, but resembles ART 1. All these implementations of ART, however, take also computational complexity into account. As a consequence, some very interesting dynamical aspects of the model that are described by Grossberg (1980), for example those of the gated dipole field (GDF), are approximated by a substitution of the equilibrium behavior.

In this article we will show that the original ideas about ART can be completely implemented as a stand-alone system of ordinary differential equations (ode's). In contrast to the current implementations, the model we will introduce is a complete implementation of an ART network, including all regulatory and logical functions, as a system of ordinary differential equations capable of stand-alone running in real time. This means that transient behavior is kept intact. Since ART 2 agrees most with the original description, we will take this model as a starting point of an implementation of ART as is pointed out by Grossberg (1980).

ART is developed according to so-called design principles. These features include biological plausibility, cognitive psychological plausibility, and environmental constraints. Obviously, the agreement with reality is limited. Models that are completely realistic with regard to biological characteristics of the nervous system up to the smallest level are not only impossible but will also complicate the study of the brain needlessly. However,

Requests for reprints should be sent to Maartje Raijmakers, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands; E-mail: opraijmakers@macmail.psy.uva.nl; Tel: +31-205256826; Fax: +31-206390279.

features which contradict biological or psychological principles have to be discarded (cf., Kentridge, 1994; Sepulchre & Babloyantz, 1991).

The biological plausibility of ART is due to the following three features of the model. First, all dynamic equations of the system are ode's. This means that the model is defined in continuous time. Second, all short term memory (STM) ode's, i.e., the equations that define neuron activity, are derived from membrane equations according to Hodgkin and Huxley (1952). Third, both the STM ode's and long term memory (LTM) ode's use only information available locally in place and time.

An important environmental constraint is expressed by the noise-saturation dilemma (Grossberg, 1980). All systems that receive noisy input have the dilemma between suppression of noise, which may also suppress low level inputs, and amplification of low level inputs, which also amplifies noise by which saturation occurs. Grossberg gives a possible solution of this dilemma by the introduction of on-centre off-surround structured networks. In these competitive networks uniform activity is suppressed, input differences are enhanced, and activities are normalized.

The cognitive psychological reality finds expression in the following features. First, although the network is implemented on a neurological level, learning recognition codes can be interpreted as hypotheses formation and confirmation. Moreover, this search process of learned recognition codes results in direct access if the input is sufficiently familiar. Also some other (cognitive) psychological concepts get an interpretation, e.g., LTM, STM, priming, and attention (Grossberg, 1980; Grossberg & Stone, 1986). Second, classification of input patterns takes place without feedback, but the model can be extended such that supervised and unsupervised learning can be combined (ARTMAP, Carpenter et al., 1991). Most supervised networks, in contrast, can only learn when feedback is available. Third, ART has important characteristics allowing for stand-alone running. It solves the stability–plasticity dilemma. That is, new input does not remove learned information, but will still result in a stable recognition code. Moreover, changes in the input can be detected automatically (Carpenter & Grossberg, 1990). This means that there is no need for careful regulation of input presentation by a teacher. For us the cognitive psychological plausibility of ART is of special interest, since we are working towards a developmental psychological application of ART (Raijmakers et al., 1996; Van der Maas & Molenaar, 1992). A stand-alone and continuous implementation of ART makes it possible to subject the model to bifurcation analysis in order to detect qualitative changes in the dynamics of the system under variation of external and internal parameters.

The outline of the rest of the paper is as follows. In Section 2 we briefly describe ART 2 and the adaptive resonance circuit (ARC, Ryan et al., 1987), and we discuss which adaptations of ART 2 had to be made to define the model completely as a system of ordinary differential equations. One of the major features of our implementation is a GDF, as described in Grossberg (1980). Section 3 discusses the architecture and the features of the resulting neural network, which we call Exact ART. This model reproduces the classification behavior of ART 2. Section 4 presents simulation studies which are compared with ART 2 and which affirm and supplement the analytical elaborations presented in Section 3. In Section 5 we will briefly summarize the features of Exact ART and we will evaluate Exact ART in comparison with ART 2.

2. ART 2 AND ARC

Carpenter and Grossberg (1987b) introduced ART 2 as an implementation of ART that processes analog input patterns. They describe several versions of the model. We base our model, like ART 3, on their figure 10 (page 4930). We will first describe this model briefly, and then we will discuss which features of the model

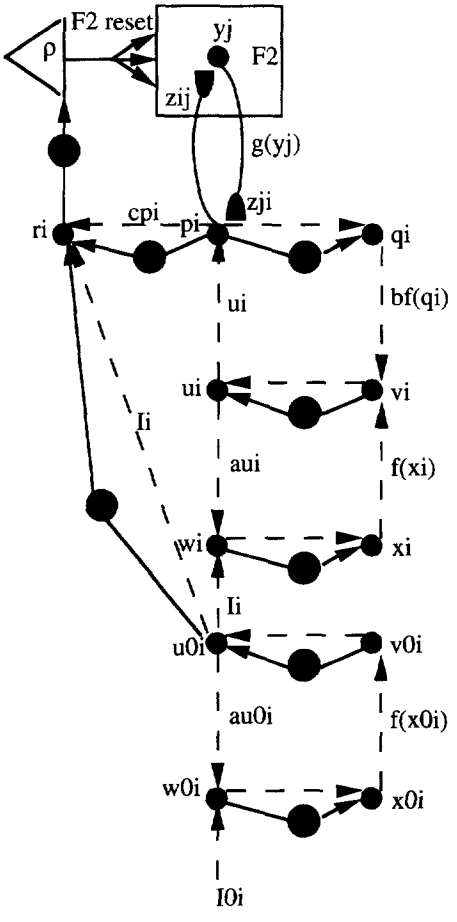

FIGURE 1. A schematic representation of ART 2 after Carpenter & Grossberg (1987b), figure 10. The filled circles denote the off-surround connection structure, that is inhibitory connections between layers. Precise definition of the network is given in the text.

need to be adjusted to keep all transient behavior intact. Ryan et al. (1987) introduced a variation on ART, the property inheritance network which contains the adaptive resonance circuit (ARC). In contrast to ART 2, ARC has a dynamic implementation of the reset procedure, which does not contain a gated dipole field. We will discuss ARC briefly at the end of Section 2.

2.1. ART 2: Architecture Definition

ART 2 consists of four main parts: F0, F1, F2, and the orienting subsystem. Fig. 1 shows a schematic view of ART 2. F0 transforms all inputs, which transformations include normalization, noise suppression, and contrast enhancement by means of competitive nonlinear interaction between layers. Competition is due to an on-centre (drawn as dashed arrows) off-surround (drawn as filled circles) connection structure between layers. F2 selects a learned recognition pattern, which is stored in the LTM connections between F1 and F2 (drawn as half ellipses), by means of a winner-takes-all competition. F1 combines input from these two sources: bottom-up (BU) input from F0 and top-down (TD) expectancies from F2. The F1 activity and the F0 activity are compared by the orienting subsystem. If they match then the LTM connections adapt to the F1 activity pattern. If they do not match then the current active F2-node is quenched, F1 and F2 are reset, and a different F2 node is selected. The search process repeats itself until BU input and TD expectancy match and resonance appears between F1 activity and F2 activity.

All units or neurons of ART 2 obey membrane equations of the following form:

$$\varepsilon \frac{\mathrm{d}}{\mathrm{d}t} V_i = -AV_i + (1 - BV_i)J_i^+ - (C + V_i)J_i^- \qquad (1)$$

J_i^+ is the total excitatory input, J_i^- the total inhibitory input. A is the decay of activity, C is the lower-bound of activity (i.e., inhibitory saturation point). Without any external input the activity becomes zero (i.e., the passive saturation point is zero). The parameter ϵ is the ratio between STM-ode's (that define F0, F1, and F2 activity) and LTM-ode's (that define the change of weights of the connections between F1 and F2): $0 < \varepsilon \ll 1$. If $\mathrm{d}V_i/\mathrm{d}t = 0$, which implies that V_i is in equilibrium, and additionally $B \equiv 0$, $C \equiv 0$, eqn (1) reduces to eqn (2):

$$V_i = \frac{J^+}{A + J^-} \qquad (2)$$

The following equations define the activity of the system in equilibrium.

$$w0_i = I_i + au0_i$$

$$x0_i = \frac{w0_i}{A + \|\mathbf{w0}\|}$$

$$v0_i = f(x0_i)$$

$$u0_i = \frac{v0_i}{A + \|\mathbf{v0}\|}$$

$$w_i = u0_i + au_i$$

$$x_i = \frac{w_i}{A + \|\mathbf{w}\|}$$

$$v_i = f(x_i) + bf(q_i)$$

$$u_i = \frac{v_i}{A + \|\mathbf{v}\|}$$

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

$$q_i = \frac{p_i}{A + \|\mathbf{p}\|}$$

$$f(x) = \left\{ \begin{array}{l} 0 \text{ if } x < \theta \\ x \text{ if } x \geq \theta \end{array} \right\}$$

$\|\mathbf{p}\|$, $\|\mathbf{v}\|$, $\|\mathbf{w}\|$, $\|\mathbf{v0}\|$, and $\|\mathbf{w0}\|$ correspond with J^- in eqn (2). These inhibiting connections between layers normalize the activity within a layer with regard to its Euclidean norm. The activity in F2 is denoted by y_j. The input of a node y_j in F2 is T_j, which is defined by:

$$T_j = \sum_{i=1}^{N} p_i z_{ij}$$

F2 is said to make a choice between the nodes y_j, which implies that only the node with the maximal input that did not cause a mismatch becomes active. This is expressed as follows:

$$g(y_j) = \left\{ \begin{array}{l} d \text{ if } T_j = \max \left\{ T_j \mid \text{the jth node has not been} \right. \\ \left. \text{reset on the current trial} \right\} \\ 0 \text{ otherwise} \end{array} \right\}$$

The LTM equations, the ode's that define the change of z_{ji} and z_{ij}, are given by:

$$(F2 \rightarrow F1): \frac{\mathrm{d}}{\mathrm{d}t} z_{ji} = g(y_j)[p_i - z_{ji}]$$

$$(F1 \rightarrow F2): \frac{\mathrm{d}}{\mathrm{d}t} z_{ij} = g(y_j)[p_i - z_{ij}]$$

The calculation of a match between a TD expectancy ($\mathbf{z}_J$ which is reflected in $\mathbf{q}$) and a BU input ($\mathbf{I}$ which is reflected in $\mathbf{u0}$) takes place by means of units r_i. The equation for r_i in equilibrium is as follows:

$$r_i = \frac{u_i + cp_i}{A + \|\mathbf{u}\| + \|c\mathbf{p}\|}$$

The Euclidean norm of $\mathbf{r}$, $\|\mathbf{r}\|$, is compared to the vigilance parameter r in order to decide whether BU and TD

expectation patterns match. A mismatch takes place if in equilibrium $\|\mathbf{r}\| < (\rho - A)$. If a mismatch is established F1 and F2 activity is reset and recalculated. If a TD expectation pattern matches the BU input pattern, the LTM trace connected to the active F2-node is updated. This sequential procedure for calculating STM and LTM traces is thought to be a proper approximation of simultaneous update for two reasons. First, the LTM-ode's are much slower than the STM-ode's (expressed by a small value of ε in eqn (1)). Second, the period of transient behavior of activities is much shorter than the equilibrium period during one input presentation. A more extended elucidation of ART 2 is given by Carpenter and Grossberg (1987b).

2.2. Lack of Transient Behavior in the Implementation of ART 2

Carpenter and Grossberg (1987b) show that ART 2 possesses the main features of an ART model. However, in the simulations only an approximation of the dynamics, that means equilibrium values of neuron activities, are used. Although the implementation is simplified, the article about ART 2 also presents an outline for a more complete implementation. That means, for the modules of which only equilibria values are implemented, a complete dynamic description is given. Nevertheless, making a completely continuous implementation of an ART model is not straightforward. One reason for this is that the interaction of transient behavior of modules can affect the equilibrium behavior of the model and thus the performed classification unexpectedly. It even appeared that the simultaneous update, instead of a sequential update, of STM and LTM ode's influences the outcome of the classification. Furthermore, some aspects of the behavior, like initializing activities before presentation of an input pattern and generation of an arousal signal, are not defined as dynamic processes.

More specifically, the adjustments that should be made are the following: First, F2 needs an implementation in which both the winner-takes-all dynamics and the suppression of nodes due to reset are defined by ode's. Second, the orienting subsystem should be changed in that both the onset of the inhibition after detection of a mismatch and the generation of inhibition signals are implemented by a system of ode's. Third, the detection of changes in the input should be performed by the orienting subsystem (as in ART 3). Fourth, initialization of activity before the presentation of a new input pattern should not be necessary. Fifth, transients should not destabilize network dynamics. This implies, for example, that transient activity does not initiate a reset if the equilibrium situation will not lead to that. Sixth, all ode's are numerically integrated simultaneously instead of calculating the LTM ode's after STM ode's reached equilibrium.

In the next section, we will introduce Exact ART. Apart from the adjustments of ART 2 that we proposed above, we will introduce some additional ones. As a consequence of these adjustments, we could mathematically derive the important characteristics of the match procedure (Raijmakers & Molenaar, 1994).

2.3. The Dynamic Implementation of the Reset Process in ARC

As we mentioned above, one of the aspects of ART 2 that are not implemented dynamically is the reset process. In their ARC, Ryan et al. (1987) made an elegant dynamic implementation of the reset process without using a GDF. ARC handles binary input patterns without normalization of activity of the input layer and comparison layer. The latter is crucial for the similarity measure, which is used in the reset procedure. The similarity measure of ARC is as follows:

$$s(D, T_J) = \frac{\sum_{i=1}^{N} f(x_i)}{\sum_{i=1}^{N} d_i}$$

where $\{x_1, x_2, ..., x_N\}$ are the nodes of the comparison layer (F1 in ART 2), function f is a threshold function, and $\{d_1, d_2, ..., d_N\}$ is a binary input pattern.

The suppression of mismatching F2-nodes is now implemented by nodes r_j which are governed by:

$$\frac{dr_j}{dt} = Rf(y_j) - r_j$$

$$a_1 = \left\{ \begin{array}{l} 1 \text{ if } s(D, T_J) < \rho \\ 0 \text{ otherwise} \end{array} \right\}, a_2 = \left\{ \begin{array}{l} 1 \text{ if } 1/s(D, T_J) < \rho \\ 0 \text{ otherwise} \end{array} \right\},$$

$$R = R_0 f(a_1 + a_2)$$

where y_j is a node of the F2 layer, ρ is the vigilance parameter, R_0 is a constant, and r_j decays at the LTM rate. Each node y_j of F2 is inhibited by the corresponding node r_j, so that r_j acts as directed arousal. Since we intend to implement an ART network that corresponds with the original description in Grossberg (1980) as much as possible, we implement a GDF with undirected arousal to reset mismatching F2-nodes. Moreover, Exact ART, as ART 2, handles analog input patterns and has normalized activity of the input layer (F0 in ART 2 and Exact ART) and the comparison layer (F1 in ART 2 and Exact ART). Therefore, the similarity measure of ARC in its present form cannot be applied to Exact ART.

3. EXACT ART: NETWORK DEFINITION

Exact ART is a neural network that is completely defined by a system of ode's that is capable of stand-alone running in real time. For the development of Exact ART we followed the design principles of ART 2 and searched for ode implementations of the discrete features, which were

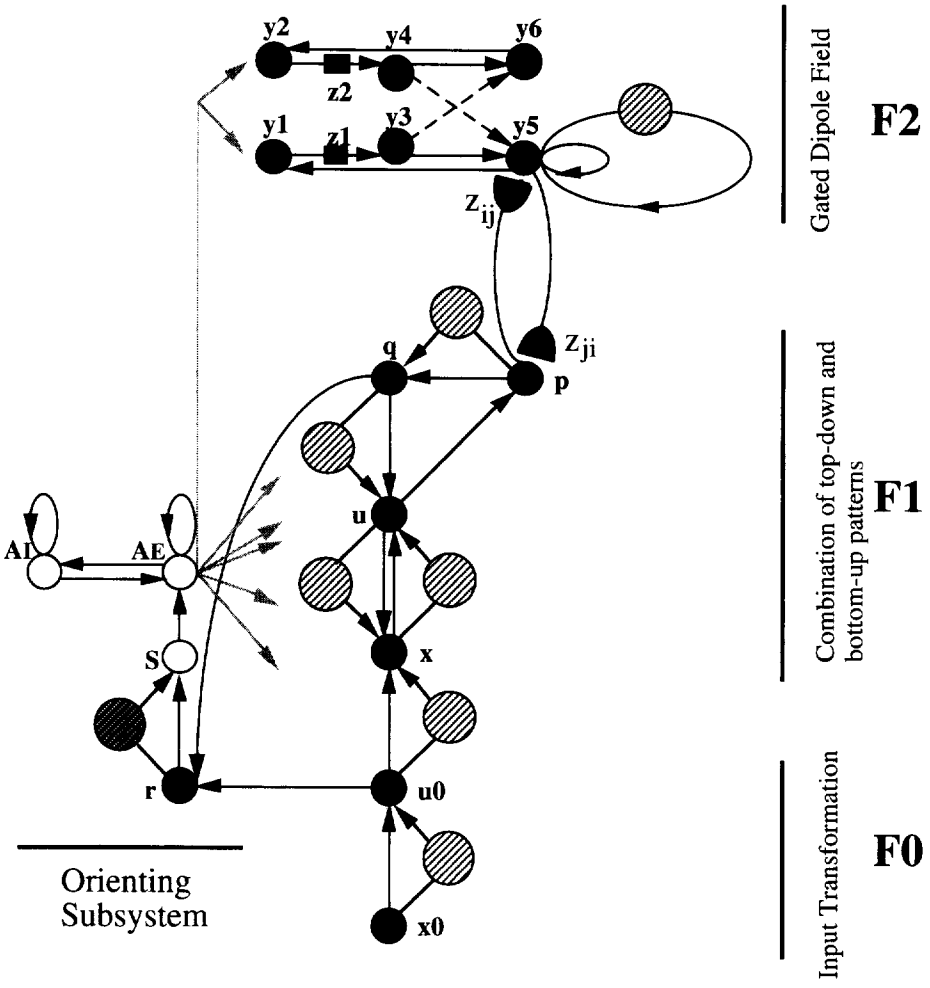

FIGURE 2. This figure shows an outline of Exact ART. The model consists of four major parts: F0, F1, F2, and the orienting subsystem. A full description of the model is given in Section 3. Each filled black circle denotes an element of a layer of equivalent nodes. Each white circle denotes a single unique node. Striped circles denote nonspecific interactions: excitory for the dark circles and inhibitory for the light circles. Black arrows denote specific interactions: excitatory interactions for plain lines and inhibitory interactions for dashed lines. Grey arrows denote nonspecific inhibitory arousal. Filled rectangles and half ellipses denote adaptive weights. Note that the model is designed after the second version of ART 2 (Carpenter & Grossberg, 1987b), and bears a large resemblance to it. The mathematical definition of the model is given in Section 3.1.

mentioned in Section 2. Exact ART resembles the ART 2 architecture in many respects and shows approximately the same classification behavior. We will first present the network by a schematic view (Fig. 2) and by the system of ode's. This will be followed by a discussion of general features of the model and the main differences with ART 2. In Sections 3.2–3.7 the features of individual network modules are described. Mathematical proofs will be given, as far as possible, in the Appendices.

3.1. System of Ordinary Differential Equations

F0 – input transformation:

$$\frac{1}{\phi}\frac{dx0_i}{dt} = -Ax0_i + (B - x0_i)I_i - x0_i\sum_{k\neq i}^{N} I_k \qquad (3)$$

$$\frac{1}{\phi}\frac{du0_i}{dt} = -Au0_i + (B - u0_i)f(x0_i,\theta) - u0_i\left(\sum_{k\neq i}^{N} f(x0_k,\theta) + cf(A_E,\theta)\right) \qquad (4)$$

F1 – combination of bottom-up input and top-down expectation:

$$\frac{dx_i}{dt} = -Ax_i + (B - x_i)(u0_i + au_i) - x_i\left(\sum_{k\neq i}^{N} u0_k + a\sum_{k\neq i}^{N} u_k + cf(A_E,\theta)\right) \qquad (5)$$

$$\frac{du_i}{dt} = -Au_i + (B - u_i)(x_i + bq_i) - u_i\left(\sum_{k\neq i}^{N} x_k + b\sum_{k\neq i}^{N} q_k + cf(A_E,\theta)\right) \qquad (6)$$

$$\frac{dq_i}{dt} = -Aq_i + (B - q_i)p_i - q_i\left(\sum_{k\neq i}^{N} p_k + cf(A_E, \theta)\right) \quad (7)$$

$$\frac{dp_i}{dt} = -p_i + d\sum_{j=1}^{M} y_j z_{ji} + u_i - p_i cf(A_E, \theta) \quad (8)$$

F2 – gated dipole field:

$$\frac{dy1_j}{dt} = -y1_j + y5_j + A_E \quad (9)$$

$$\frac{dy2_j}{dt} = -y2_j + [y6_j]^+ + A_E \quad (10)$$

$$\frac{dy3_j}{dt} = -y3_j + z1_j y1_j \quad (11)$$

$$\frac{dy4_j}{dt} = -y4_j + z2_j y2_j \quad (12)$$

$$\frac{dy5_j}{dt} = -Ay5_j + (B - y5_j) \times \left(h(y5_j) + y3_j + e\sum_{k=1}^{N}\sqrt{z_{kj}p_k}\right) - y5_j\left(\sum_{l\neq j}^{M} h(y5_l) + y4_j\right) \quad (13)$$

$$\frac{dy6_j}{dt} = -y6_j + (y4_j - y3_j) \quad (14)$$

$$\frac{1}{\varepsilon}\frac{dz1_j}{dt} = \beta(\gamma - z1_j) - \delta[y1_j - \Gamma]^+ z1_j \quad (15)$$

$$\frac{1}{\varepsilon}\frac{dz2_j}{dt} = \beta(\gamma - z2_j) - \delta[y2_j - \Gamma]^+ z2_j \quad (16)$$

$$h(x) = x^2 \quad (17)$$

Orienting sub-system – matching and nonspecific arousal:

$$\frac{dr_i}{dt} = -r_i + \frac{1}{2}|u0_i - q_i| \quad (18)$$

$$\frac{1}{\eta}\frac{dA_E}{dt} = -A_E + g\big(a_e(c_1A_E - C_2A_I - \theta_e + S)\big) \quad (19)$$

$$\frac{1}{\eta}\frac{dA_I}{dt} = -A_I + g\big(a_i(c_3A_E - c_4A_I - \theta_i)\big) \quad (20)$$

$$\frac{1}{\iota}\frac{dS}{dt} = -AS + (B - S)\left[\sum_{i=1}^{N} r_i - (1-\rho)\right]^+ - \omega S\left[(1-\rho) - \sum_{i=1}^{N} r_i\right]^+ \quad (21)$$

LTM – connections between F1 and F2:

$$\frac{1}{\alpha}\frac{dz_{ji}}{dt} = df(y_j, \theta_y)(f(p_i, \theta) - z_{ji}) \quad (22)$$

$$\frac{1}{\alpha}\frac{dz_{ij}}{dt} = df(y_j, \theta_y)(f(p_i, \theta) - z_{ij}) \quad (23)$$

Non-linear signal functions:

$$f(x, \theta) = \left\{\begin{array}{l} x \text{ if } x > \theta \\ 0 \text{ if } x \leq \theta \end{array}\right\} \quad (24)$$

$$g(x) = \frac{1}{1 + \exp(-x)} \quad (25)$$

The notation used in this section and the following sections is as follows: Vectors are written in bold characters (**x0**, **u0**, **x**, **u**, **q**, **p**, **y1** – **y6**, **z1**, **z2**, **r**), the weight matrix is also written as a bold character (**z**). Vector elements are denoted by italic letters with an index; i is the index of F1 and F0 elements, j is the index of F2 elements ($x0_i, ..., p_i$, $y1_j$–$y6_j$, $z1_j$, $z2_j$, r_i, z_{ij}, z_{ji}). N is the number of elements within F1 and F0 layers. M is the number of F2 elements, i.e., gated dipoles. The index of the winning node in F2 is J. A_E, A_I, and S denote activities of the orienting subsystem. The sum of activities within a layer is denoted by an italic letter without an index ($x0 = \sum_{i=1}^{N} x0_i$). The external input is denoted by $\mathbf{I} = (I_1, ..., I_i, ..., I_N)$. Further notations used in this chapter are explained in the Appendix. Constants of above equations and their default values are summarized in Appendix C.

Apart from the architecture specified above, additional assumptions are made. The match is denoted by R:

$$R = \sum_{i=1}^{N} r_i$$

$R = 0$ corresponds to a perfect match; $R = 1$ corresponds to a maximum mismatch. At time $t = 0$ TD connections, z_{ji}, are equal to 0 for all $i \leq M$ and $j \leq N$; BU connections, z_{ij} for all $i \leq M$ and $j \leq N$, are initialized randomly from a uniform distribution between 0 and a constant value (i.e., $z_{ij}(0)$). Constant B is set to 1. In most mathematical analyses, A will be taken to be 0 (conforming with Carpenter & Grossberg, 1987b), since its actual value will be very low (0.001). The constant e is very small for reasons given below, (i.e., $e \leq 0.01$).

The procedure to simulate the network on a computer consists of simultaneous calculation of a solution of the whole system of ode's by means of a numerical integration procedure. We use Divpag from the IMSL-library (IMSL, 1991). Section 4 shows some simulations we performed with Exact ART. We will discuss the simulation procedure more extensively in that section.

3.2. Differences Between ART 2 and Exact ART

The structure of the presented system is globally

equivalent to the second version of ART 2 (Carpenter & Grossberg, 1987b, figure 10). This means that the BU input pattern is pre-processed in F0 before it is combined with a TD expectation pattern in F1. A match between a BU input pattern and TD expectation pattern is calculated by the orienting subsystem.

The first striking difference between Exact ART and ART 2 concerns the normalization of activity within layers. In contrast to ART 2, normalization takes place with regard to the summation of activities, instead of the Euclidean norm of the activity vector (Carpenter & Grossberg, 1990, mention this possibility). We use (non-shunting) networks described by Grossberg (1980), which have a constant sum of activity due to on-centre off-surround input. In Fig. 2 the on-centre input, i.e., specific input, is denoted by plain arrows. The off-surround input, i.e., non-specific input, is denoted by light-colored striped circles. The general form of the ode's, which are instances of Hodgkin–Huxley equations, defining these networks is:

$$\frac{\mathrm{d}}{\mathrm{d}t}x_i = -Ax_i + (B - x_i)I_i - (x_i + C)\sum_{k \neq i}^{N} I_k \qquad (26)$$

The F0 and F1 layers, except for **p**, have the form of eqn (26), with $C \equiv 0$. Grossberg (1980) proves that eqn (26) solves the noise-saturation dilemma. In equilibrium (i.e., $\mathrm{d}x_i/\mathrm{d}t = 0$ for all $i \leq N$) the following equation is valid:

$$x_i = \frac{(B + C)I}{A + I}\left(J_i - \frac{C}{B + C}\right) \qquad (27)$$

According to eqn (27), x_i is proportional to J_i ($J_i \equiv I_i/I$)minus a constant (i.e., $C(B + C)^{-1}$), which implies that it is independent of the sum input. That is, also in the presence of noise, the network is not saturated but represents differences between input elements I_i. The sum of activity, denoted by x, equals $[B - (N - 1)C](A + 1)^{-1}$. If $C \equiv 0$, which is the case for the F0 and F1 equations, x does not depend on N and is equal to B (provided that A is taken to be 0). Under various conditions, eqn (26) suppresses noise, which means that uniform input patterns are quenched. As will be explained in Section 3.3, we use another solution for quenching uniform patterns in F0, for reasons given there.

The change of normalization properties in the present implementation has two major consequences. The match, R, heavily depends on the normalization of activities within **q** and **u0**. Hence, the match procedure had to be adjusted (see Section 3.7). Also the input function of F2 needed to change. In order to prevent different **z**-traces with equal norm to cause ties, the input of a F2-dipole is made a nonlinear function of the inner product of **p** and the BU **z**-trace (Section 3.6).

A second difference between ART 2 and Exact ART concerns the nonlinear function f. We moved f from **u** to the LTM-ode's (eqns (22) and (23)) and **u0** (eqn (4)). The reason with regard to the LTM-ode's is to preserve the 2/3-rule (see Section 3.4). The introduction of f in **u0** is necessary for contrast enhancement and for preservation of the perfect match in case $0 \leq J_i \leq q$ for some $i \leq N$ (see Section 3.3).

A third difference concerns the introduction of ode's that implement the orienting subsystem. The dynamics of the orienting subsystem can be described as follows. S increases from 0 to 1 if R exceeds $1 - \rho$ (the vigilance parameter). ρ constitutes the match criterion of a BU input pattern and TD expectation pattern. If S is increased to almost 1, A_E increases suddenly from below zero to 1 and then inhibits several layers (see Section 3.8).

The fourth and major difference concerns the definition of F2. Its global structure is a lumped shunting network with local on-centre and global off-surround connections, which is expressed by eqn (13). As Grossberg (1973, 1980) and Ellias and Grossberg (1975) have shown, this type of networks obeys a winner-takes-all dynamics if the feedback function h is faster than linear. However, their proofs concern shunting networks without enduring external input. In their proofs patterns are presented by initialization of the activity of the nodes at time $t = 0$ instead of an enduring external input. We showed that lowering the effect of the input on units $y5_j$, by means of parameter e ($0 < e \ll 1$), is an alternative way of presenting input without changing the global properties of these networks (Raijmakers & Molenaar, 1994). The elements of the shunting network are gated dipoles. These gated dipoles are based on Grossberg (1980).

In addition to the system of ode's, we changed the solution procedure of Exact ART as well. Carpenter and Grossberg (1987b) use an approximation of a simultaneous solution of STM and LTM ode's. They first compute the equilibrium values of the activity in the network and then, in the fast-learning case, they compute equilibrium values of all LTM ode's based on the former equilibrium. In contrast, we compute (equilibrium values of) all ode's simultaneously by means of a numerical integration procedure. Although the first procedure is thought to approximate the latter, in some cases there appear to be significant differences (Section 3.4).

In the following sections, we will discuss the different modules of Exact ART one-by-one. In this article, not all proven features of the model are elaborated completely. For a more extended description we refer to a technical report (Raijmakers & Molenaar, 1994). Since mathematical analysis of the transient behavior of the network is only discussed in a few cases, simulation studies are important to be performed. In Section 4 we present simulation studies that compare the classification behavior of Exact ART with the classification behavior of ART 2.

3.3. F0: Input Transformation

Preprocessing of the input consists of three transformations: contrast enhancement, noise suppression, and normalization. These transformations are performed by **x0**

and **u0**. First, the input, $\mathbf{I} = (I_1, ..., I_i, ..., I_N)$, is transformed into the normalized pattern **x0**. Then, through a partially linear function f, **x0** is gated in that $u0_i = 0$ if $x0_i \leq q$. The latter appears to be necessary to maintain the 2/3-rule (see Section 3.4). In addition, **u0** activity is normalized with respect to sum activity. As a consequence, the contrast of the pattern is enhanced and uniform input patterns are suppressed. According to Grossberg (1980) uniform input patterns can be interpreted as noise. Hence, noise suppression is performed by F0.

3.4. F1: Combination of Bottom-up Input and Top-down Expectation

F1 receives input from two sources: F0 and F2. The main function of F1 is to combine these inputs. In the match procedure the resulting pattern is compared with the transformed input pattern **u0**. Carpenter and Grossberg (1987b) discuss four main features of F1: Normalization of STM, a TD expectation pattern is normalized before it interacts with a BU input pattern and vice versa, STM is invariant under read-out of matched LTM, and the 2/3-rule.

In Exact ART normalization of STM takes place in **x**, **u**, and **q**, since these layers are structured according to eqn (26).

A BU input pattern is transformed by F0. The result is a normalized activity pattern **u0**, which serves as input for F1. A TD expectation pattern is represented by LTM traces $\mathbf{z}_J = (z_{J1}, ..., z_{Ji}, ..., z_{JN})$. In the fast-learning condition, which means that LTM traces reach equilibrium each time a new input pattern is presented, $\mathbf{z}_J$ is equivalent to **p** (cf., Section 3.6). For each i, $i \leq N$, holds that at equilibrium

$$p_i = \frac{u_i}{1-d} \tag{28}$$

Since **u** is normalized, **p** is normalized (sum of p_i elements $\equiv p = 1/(1-d)$). Hence, $\mathbf{z}_J$ is normalized as well. In equilibrium the sum of LTM elements connected to one F2-node equals $1/(1-d)$. Hence normalization of both TD expectation patterns and BU input patterns takes place before they interact.

In order to check whether STM is invariant under read-out of matched LTM we examine **q** in two cases: First if the TD expectation pattern is zero (i.e., $\Sigma z_{Ji} = z_J = 0$), and second, if a TD expectation pattern matches a BU input pattern perfectly. A perfect match means that $\mathbf{z}_J$ is proportional to **u0**. In both cases, **q** should be equal to **u0**. In Appendix A we show that the following holds if the activities reach equilibrium:

$$q_i = \frac{u0_i + \chi(1+b+ab)Z_{Ji}}{(1+\chi(1+b+ab))}$$

$$Z_{Ji} = \frac{z_{Ji}}{\sum_{k=1}^{N} z_{Jk}} \text{ if } \sum_{k=1}^{N} z_{Jk} > 0, \tag{29}$$

$$\text{otherwise } Z_{Ji} = 0;\ \chi = d \sum_{k=i}^{N} z_{Jk}$$

If no learning took place, which agrees with the first case, $\chi = 0$. Hence eqn (29) reduces to: $q_i = u0_i$ for all i, $i \leq N$. If $\mathbf{z}_J$ is proportional to **u0**, which agrees with the second situation, then $Z_{Ji} = u0_i$ for all $i \leq N$. This implies:

$$q_i = \frac{u0_i(1+\chi(1+b+ab))}{1+\chi(1+b+ab))} \tag{30}$$

from which follows that $q_i = u0_i$ for all i, $i \leq N$, QED.

The 2/3-rule posits that elements in F1 only become active if they receive input from both F0 and F2. This implies two conditions. First, z_{Ji} does not increase if $z_{Ji} = 0$ and $\chi > 0$, even not if $u0_i > 0$. Second, z_{Ji} becomes 0 if $u0_i = 0$, even if initially $z_{Ji} > 0$. Analogous to ART 2, we implement the 2/3-rule implicitly by means of the introduction of parameters a, b, and θ. In contrast to ART 2 we introduced the threshold θ, in the LTM ode's (eqns (22) and (23)).

To prove that the 2/3-rule holds for Exact ART, we start with the first situation. This situation is characterized by the existence of nodes i such that $Z_{Ji} = 0$, $u0_i > 0$, and $\chi > 0$. According to the 2/3-rule, for these i, z_{Ji} should remain zero. On above conditions, for these i eqn (29) becomes:

$$q_i = \frac{u0_i}{1+\chi(1+b+ab)}$$

$$\Leftrightarrow \quad p_i = \frac{u0_i}{1+\chi(1+b+ab)}\left(1+\frac{1}{1-d}\right) \tag{31}$$

Since z_{Ji} is increased if $z_{Ji} < p_i$, $u0_i$ increases both p_i and z_{Ji}. If $a,b \gg 0$, the increase is small, but accumulative because p_i increases if z_{Ji} increases, and vice versa. Now, the difference between a simultaneous and a sequential integration of STM and LTM ode's becomes apparent. By simultaneous integration, the increase accumulates during the presentation of one input pattern. In contrast, by sequential integration accumulations only take place due to several input presentations (cf., simulation of ART 2 with $\theta = 0$ in Carpenter and Grossberg, 1987b). Fig. 3 shows the relation between p_i and χ for $a = b = 20$ and $u0_i = 1$ (its maximum value). It appears that p_i decreases very fast with increasing χ. Hence, a threshold in the LTM ode's prevents z_{Ji} from increasing in the case that some learning took place, $u0_i > 1$, and $z_{Ji} = 0$. In order to preserve a perfect match for each input pattern that selects an uncommitted F2 node, only a simple demand on $u0_i$ is necessary: $u0_i > \theta$, or $u0_i = 0$ for all $i \leq N$. This requirement is satisfied by F0.

The second situation of the 2/3-rule is characterized by the existence of nodes i such that $Z_{Ji} > 0$, $u0_i = 0$. Provided that $a,b \gg 1$, eqn (29) implies that for these i

$$q_i = \frac{\chi(1+b+ab)Z_{Ji}}{1+\chi(1+b+ab)} \approx Z_{Ji}$$

This means that q_i decreases only very little. As a result of simultaneous updating of LTM and STM ode's the decrease of q_i becomes an accumulative decrease. That is, q_i decreases with regard to z_{Ji}, which makes z_{Ji}

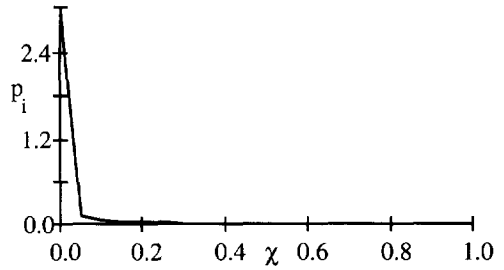

FIGURE 3. This figure shows the dependency of p_i on χ in the case that $Z_{Ji} = 0$, $u0_i = 1$, $a = b = 20$, $d = 0.5$. Because p_i decreases dramatically with increasing c, a threshold in the learn equations can prevent a cumulative growth of Z_{Ji}.

decrease, and then q_i, etc. As a result, those nodes that are not activated by both a TD and a BU pattern are completely suppressed. The latter implies a strong 2/3 rule. Consequently, since the dynamics of **q** and **u** are competitive, the activities of the other nodes are enhanced. If the nodes that are not activated by both the BU and the TD pattern, are not completely suppressed, the 2/3-rule is called weak. It should be noted that the 2/3 rule is only strong if a pattern is learned until equilibrium is reached. This may take a very long time, since the learning rate, α, is relatively low.

3.5. F2: a Gated Dipole Field

As the implementation of the F2 field we chose a GDF as Grossberg (1980) proposed. F2 should fulfil several functions: First, it needs to perform a winner-takes-all dynamics, which means that the activity of node (or dipole) j with the largest input (input of dipole j is named H_j) becomes 1 and that the other activities become 0.

$$H_j = \sum_{i=1}^{N} \sqrt{p_i z_{ji}}$$

Second, the F2-field should depress all nodes j that were active before a reset, by means of undirected arousal. Third, after the undirected arousal, the node (or dipole) with the largest input that was not active before the reset

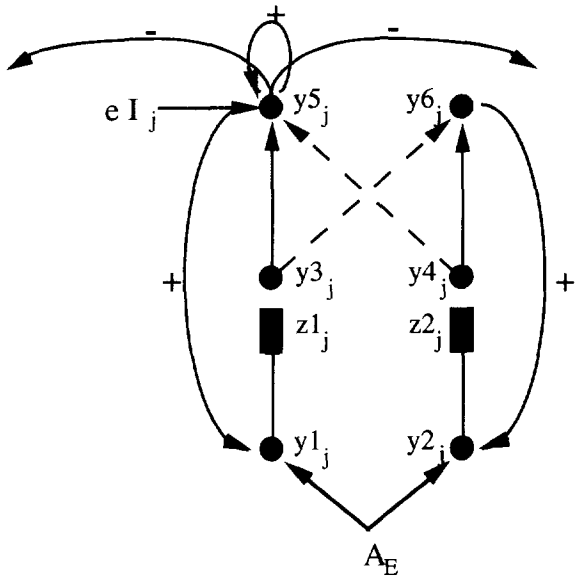

FIGURE 4. A gated dipole that constitutes an element of the F2-layer of Exact-ART.

should become the next winner. Grossberg (1980) describes a GDF as a possible implementation of F2. We adjusted the structure of the proposed dipoles so that the decrease of the amount of transmitter is proportional to the activity of the dipole instead of being proportional to the input (Raijmakers & Molenaar, 1994). The resulting recurrent dipole is shown in Fig. 4. Recurrent dipole connections are also present in the circuit of READ I (Grossberg & Schmajuk, 1987) although the structure of READ is more complex.

To analyse the behavior of the dipole we examine two situations: the GDF without a reset and the GDF during and after a reset. In these analyses we suppose the following: The input of the GDF is such that $1 \geq H_M > H_{M-1} > \ldots > H_1$. $z1_j(0) = \gamma$, $z2_j(0) = \gamma$ for all dipoles j (γ is the maximum and at the same time saturation value of $z1_j$ and $z2_j$). The initial activities of dipole nodes are 0, i.e., $yk_j(0) = 0$, $k \leq 6$, for all dipoles j.

3.5.1. *Situation 1.* Input **H** is on, arousal A_E is off, no arousal event occurred since $t = 0$. Table 1 shows the

TABLE 1
Equilibrium Values of the Elements of a GDF Described by Situation 1

Variable	Dipole 1	Dipole 2...	Dipole *k*...	Dipole *M*
$y1_j$	≈ 0	≈ 0	≈ 0	≈ 1
$y2_j$	0	0	0	0
$y3_j$	≈ 0	≈ 0	≈ 0	$\approx z1_M$
$y4_j$	0	0	0	0
$y5_j$	≈ 0	≈ 0	≈ 0	≈ 1
$y6_j$	≈ 0	≈ 0	≈ 0	$-z1_M < 0$
z_1	$\approx \gamma$	$\approx \gamma$	$\approx \gamma$	$z1_M \rightarrow \beta\gamma/(\beta + \delta(1 - \Gamma)) < \gamma$
z_2	γ	γ	γ	γ

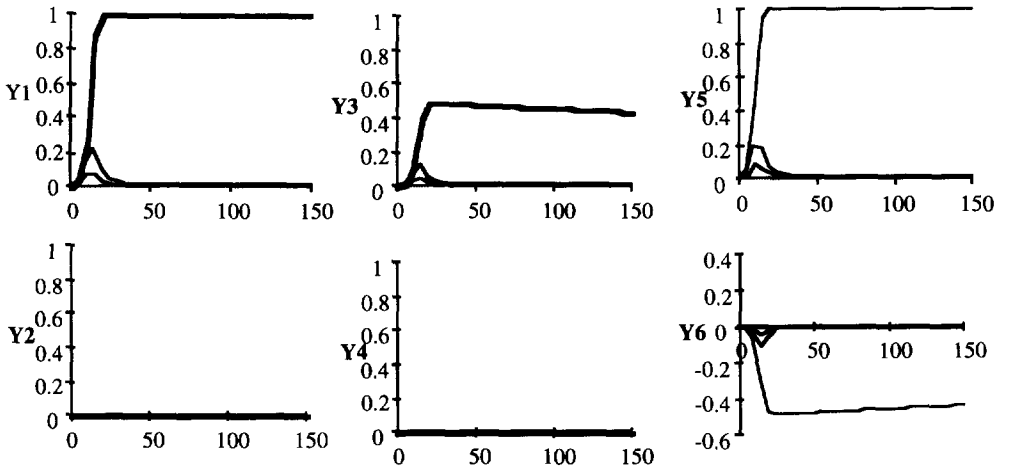

FIGURE 5. Simulation results of a small GDF, $M = 4$, in situation 1. Each figure shows the time course of individual dipole nodes. yk_1 (grey line); yk_2 (thin solid line); yk_3 (medium solid line); yk_4 (thick solid line).

equilibrium values of the elements of the dipoles in the GDF. Equations for the GDF in equilibrium are given in Appendix B, including some additional analyses. Due to recurrent connections node $y5_j$ with the highest input H_j, which is $y5_M$, becomes 1. The other nodes $y5_k$, $k \neq M$ quench. Transmitters $z1_k$ for $k < M$ are decreased only a little due to transient activity. In contrast, $z1_M$ decreases significantly to a lower bound of

$$z1_M = \frac{\beta\gamma}{\beta + \delta[y1_j - \Gamma]^+} = \frac{\beta\gamma}{\beta + \delta(1 - \Gamma)}$$

Transmitters $z2_j$ ($j \leq M$) do not change, since $y2_j = 0$ for all $j \leq M$. This analysis resembles the proofs in Grossberg (1980).

The analysis of Table 1 can be compared with the simulations shown in Fig. 5. In the simulations, $M = 4$, $\mathbf{H} = (0.0, 0.2, 0.4, 0.6)$, $A = 0.001$, $B = 1$, $e = 0.01$, $\varepsilon = 0.001$, $\gamma = 0.5$, $\beta = 0.5$, $\delta = 1$, and $\Gamma = 0.1$. The main thing that we expected, $y5_4 \approx 1$ and $y5_k \approx 0$, for nodes $k < 4$, is clearly shown.

3.5.2. *Situation 2.* Situation 1 just occurred, $\mathbf{H}$ has not changed, arousal $A_E > 0$ for a short period. The main consequence of the onset of arousal $A_E = 1$ is the occurrence of an activity difference between on the one hand $y3_k$ for $k < M$ and on the other hand $y3_M$. This activity difference is due to the difference between $z1_k$ for $k < M$ and $z1_M$. As a result $y5_M$ is depressed more than the other nodes $y5_k$, $k < M$ (Table 2).

Appendix C shows that if e is much smaller than the decrease of $z1_M$ the order of the growth rates of $y5_j$, $j \leq M$ becomes:

$$y5_M < y5_1 < y5_2 ... < y5_{M-1}$$

Moreover, if $y5_M$ decreases sufficiently so that $y5_M \approx y5_j$ for some $j < M$ then $y5_{M-1}$ wins the next competition. This analysis can be repeated for the situation that also $y5_{M-1}$ is reset. Provided that $z1_M$ does not recover too fast compared to its earlier decrease, the ordering becomes:

$$y5_{M-1} < y5_M < y5_1 < y5_2 < ... < y5_{M-2}$$

The changes of $z1_j$ in both directions can be adjusted by means of parameters β and δ in eqn (15).

Fig. 6 shows simulations of a GDF in situation 2. Actually, this simulation is a continuation of the simulation shown in Fig. 5. The most important thing that happens is that after the arousal $y5_3 \approx 1$ and $y5_k \approx 0$, $k \neq 3$.

TABLE 2
Equilibrium Values of the Elements of a GDF Described by Situation 2

Variable	Dipole 1	Dipole 2...	Dipole k...	Dipole M
$y1_j$	$1 < y1_1 < y1_M$	$1 < y1_2 < y1_M$	$1 < y1_k < y1_M$	$y1_M > y1_k$, $k \neq M$
$y2_j$	1	1	1	1
$y3_j$	$\gamma y1_1$	$\gamma y1_2$	$\gamma y1_k$	$< y4_M$
$y4_j$	γ	γ	γ	γ
$dy5/dt$	≥ 0	$\geq dy5_1/dt$	$\geq dy5_{k-1}/dt$	< 0
$y5_j$	≥ 0	$> y5_1$	$> y5_{k-1}$	$\rightarrow < y5_1$
$y6_j$	< 0	< 0	< 0	> 0
$z1_j$	$\approx \gamma$	$\approx \gamma$	$\approx \gamma$	$< \gamma$
$z2_j$	γ	γ	γ	γ

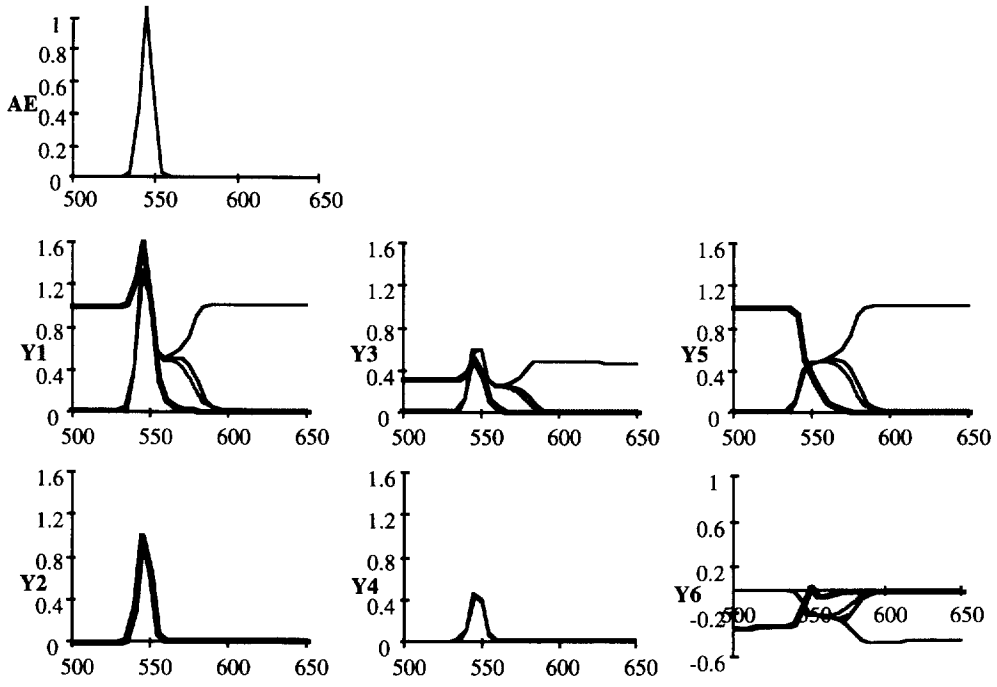

FIGURE 6. Simulation results of a small GDF, $M = 4$, in situation 2. The first figure shows the time course of the arousal. The other figures show the time courses of individual dipole nodes. yk_1 (grey line); yk_2 (thin solid line); yk_3 (medium solid line); yk_4 (thick solid line).

The same process is repeated if also $y5_3$ is quenched by arousal. Fig. 7 shows the course of **y5** including two arousal events at $t = 500$ (i.e., situation 2) and $t = 1000$. Fig. 8 shows the corresponding time course of transmitters $z1_j, j \leq 4$. The difference between transmitters $z1_j$ of active and reset dipoles on the one hand and inactive dipoles on the other hand is shown clearly.

From the analytical studies and simulation studies of an isolated GDF, we can conclude that the GDF as depicted in Fig. 4 behaves according to the three rules we listed at the begin of this section. In Section 4 we will show how the GDF behaves as a part of Exact ART.

3.6. LTM: F2 and LTM Connections Between F1 and F2

The LTM-traces between F1 and F2 determine the plasticity of the network. In a stand-alone system not only is flexibility important, but also stability is crucial. The stability–plasticity dilemma, which was mentioned in Section 2, is one of the important design principles of ART. In Exact ART, as in ART 2, LTM is only significantly changed if the BU input pattern matches the TD expectation pattern. This means that a learned rare mismatching TD expectancy will not be removed by

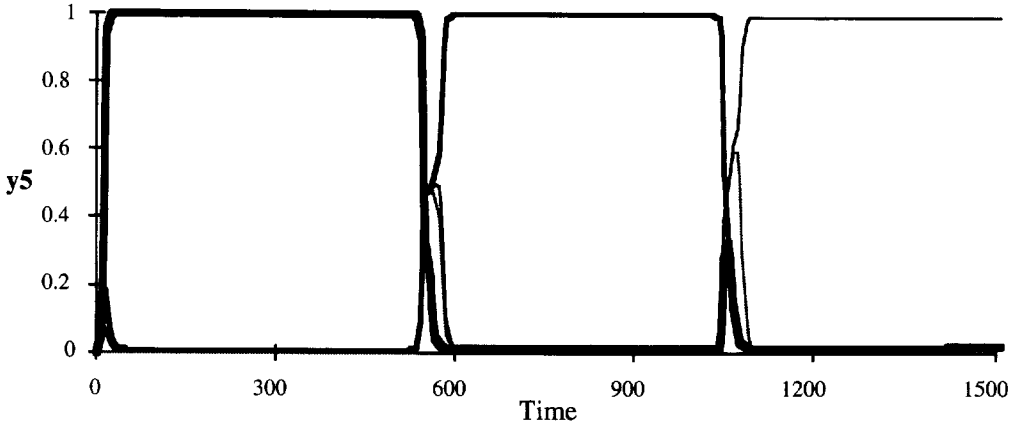

FIGURE 7. The time course of y5 in a gated dipole field with $M = 4$ and input as described in the text. The simulation starts with situation 1 (Time is [0, 500]) which is followed by situation 2 (Time is [551, 1000]). Then, after a stable period, another arousal event occurs. $y5_1$ (grey line); $y5_2$ (thin solid line); $y5_3$ (medium solid line); $y5_4$ (thick solid line).

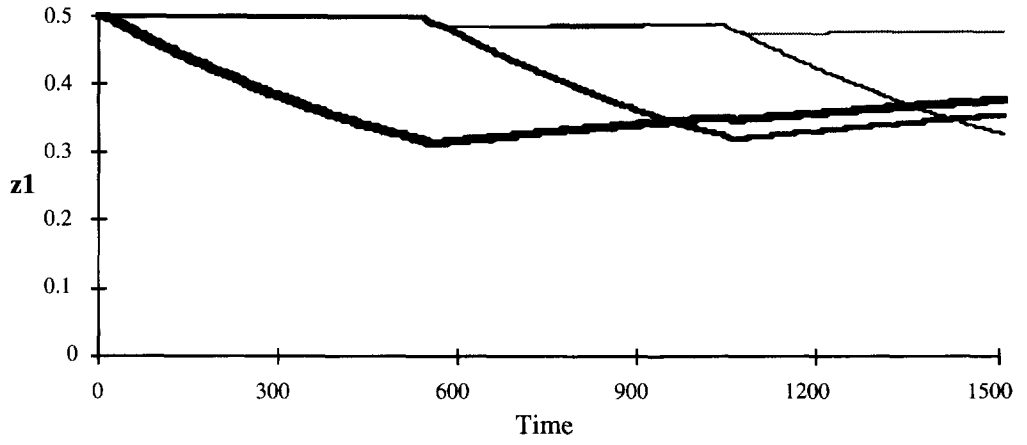

FIGURE 8. The time course of z1 in a gated dipole field with $M = 4$ and input as described previously. The simulation starts with situation 1 which is followed by situation 2. Then, after a stable period, another arousal event occurs. $z1_1$ (grey line); $z1_2$ (thin solid line); $z1_3$ (medium solid line); $z1_4$ (thick solid line).

dissimilar BU inputs. As long as there exist uncommitted nodes j, which are connected to TD $\mathbf{z}_j$-traces that equal to the null-vector, new patterns can be learned.

Another design principle of ART is the search-direct access trade-off. Classification takes place by a parallel search process of learned TD expectation patterns. If an input pattern gets more familiar the best matching TD expectation pattern is accessed directly. This is due to the normalization of LTM vectors and a nonlinear input function for F2-nodes $y5_j$. The maximal sum of weights z_{ij} connected to one F2-node is equal for each F2 node j. If all LTM ode's reach equilibrium, that means $p_i = z_{iJ}$ for all $i \leq N$, the input of $y5_J$ equals:

$$e \sum_{i=1}^{N} \sqrt{z_{iJ} p_i} = e \sum_{i=1}^{N} z_{iJ} = \frac{e}{1-d}$$

A consequence of the non-linear input function for eqn (13) is that different LTM traces never match equally well. Moreover, LTM-traces that match at many sites with low weights match better than LTM traces that match at a few sites with high weights. This means that the most detailed matching TD pattern is chosen first. The same is assured in ART 2 due to the non-linear normalization of LTM-traces.

An important feature with regard to stability of ART is that a F2 choice is stable until reset. Two design principles of Exact ART assure this feature: First, the on-centre off-surround structure of **y5** is very stable. The input of **y5** determines which node $y5_j$ becomes the winner. But the recurrent connections of $y5_j$ nodes determines which node stays the winner, because the activation due to recurrent connections is much larger than the input. The second feature that determines the stability of F2 is that LTM-changes are always in the direction of **p**. Hence, the input of the winning node in F2 can only increase during learning (the latter also holds for ART 2).

3.7. *R*: calculation of a Mismatch

The calculation of a match between a BU input pattern and a TD expectation pattern is implemented by eqn (18). The sum of all elements r_i, named R, is compared to ρ to determine whether the BU and TD patterns match. If in equilibrium $R > 1 - \rho$ the patterns match, otherwise they mismatch. This calculation is performed by S. In order to determine the features of R in equilibrium we derive the following from eqns (29) and (18):

$$R = \frac{1}{2} \sum_{i=1}^{N} |u0_i - q_i| \tag{32}$$

$$\Leftrightarrow R = \frac{1}{2} \frac{\chi(1+b+ab)}{(1+\chi(1+b+ab))} \sum_{i=1}^{N} |u0_i - Z_{Ji}| \tag{33}$$

R should obey four demands: First, without learning the match should be perfect, i.e., $R = 0$. In Section 3.4 we proved that without learning $q_i = u0_i$ for all $i \leq N$. Hence, according to eqn (32), $R = 0$.

Second, in case that $u0_i = Z_{Ji}$ for all i, the match should reach its maximum, that is, $R = 0$. This follows directly from eqn (33).

Third, during learning, the system becomes more sensitive for mismatches. The multiplier of eqn (33) is a monotone increasing function of χ if $a,b > 0$ and $0 \leq \chi$. During learning χ increases, which implies that the sensitivity for differences between $u0_i$ and Z_{Ji} increases.

Fourth, if the following two conditions are true, the mismatch is maximal (i.e., $R \approx 1$): First, $\chi = d/(1-d)$. Second, $u0_i = 0$ if $Z_{Ji} \neq 0$ for all $i \leq N$, which implies that $Z_{Ji} = 0$ if $u0_i \neq 0$ for all $i \leq N$. On the above two

conditions, eqn (33) implies that

$$q_i = \begin{cases} \dfrac{\chi(1+b+ab)}{1+\chi(1+b+ab+} Z_{Ji} & \text{if } u0_i = 0 \\ \dfrac{u0_i}{1+\chi(1+b+ab)} & \text{if } Z_{Ji} = 0 \\ 0 & \text{if } Z_{Ji} = u0_i = 0 \end{cases}$$

$$\Rightarrow R = \frac{1}{2}\sum_{i=1}^{N} \frac{\chi(1+b+ab)}{1+\chi(1+b+ab)} Z_{Ji} + \frac{1}{2}\sum_{i=1}^{N} \frac{\chi(1+b+ab)}{1+\chi(1+b+ab)} u0_i$$

Since $Z_J = u0 = 1$, the above is equivalent to:

$$R = \frac{\chi(1+b+ab)}{1+\chi(1+b+ab)}$$

This implies that $R \approx 1$ (provided that $a,b \gg 1$). In other words: Maximal mismatch between **u0** and **q** occurs if learning is maximal and in addition $(u0_1,..,u0_i,..,u0_N)$ is orthogonal to $(Z_{J1},..,Z_{Ji},..,Z_{JN})$. From the above discussed four issues we can conclude that all requisites of the match procedure are fulfilled.

3.8. Orienting Subsystem: Resetting F1 and F2

The function of the orienting subsystem is to release strong unspecific inhibition if a mismatch is determined. With that, two problems are involved: First, only if a mismatch also occurs in equilibrium, the orienting subsystem should become active. Second, inhibition should remain at least until activity is back to zero. The first problem is solved by means of the introduction of S. S increases to its upper bound 1 (i.e., the default value of B) if $R > (1 - \rho)$. If $R < (1 - \rho)$ S decreases to its lower bound 0. $\frac{1}{\tau}$ determines the rapidity of changes in both directions, ω determines the rapidity of decreases alone. The rapidity of decrease is of importance for the second problem. To solve the second problem we decided to make the inhibition unit, A_E, adaptive to S so that S acts as a control variable of a hysteresis function. This means that A_E is going from low to high activity (i.e., starts inhibiting both F1 and F2) if $S \approx 1$ (e.g., $S = 1 - \varepsilon$, $0 < \varepsilon \ll 1$). Only if S decreases from 1 to approximately 0 (e.g., $S = \varepsilon$) A_E should go back to low activity. Then A_E stops to inhibit F1 and F2. Eqn (19) forms, together with eqn (20), a neural oscillator as defined by Schuster and Wagner (1990). In order to determine the proper parameter values, we performed a bifurcation analysis with two control variables: S (the strength of the external input) and c_3 (the connection strength between A_E and A_I) (Raijmakers & Molenaar, 1994). The resulting parameter values are listed in Appendix D. Fig. 9 shows the behavior of A_E if S is first slowly increased from 0 to 1 (arrow up) and then slowly decreased from 1 back to 0. Simulation studies in Section 4 will show how the orienting subsystem behaves in combination with the other modules of Exact ART.

3.9. Conclusion

Carpenter and Grossberg (1987b) implemented an ART model that classifies analog input patterns: ART 2. We tried to adjust this model in order to implement ART completely by a system of ode's and to keep all transient behavior in tact. We applied the ideas of Grossberg (1980) as much as possible. An alternative system has been described. The presence of all the important features could be proven mathematically: normalization, contrast enhancement, noise suppression, the four demands of the match procedure and the 2/3-rule. In

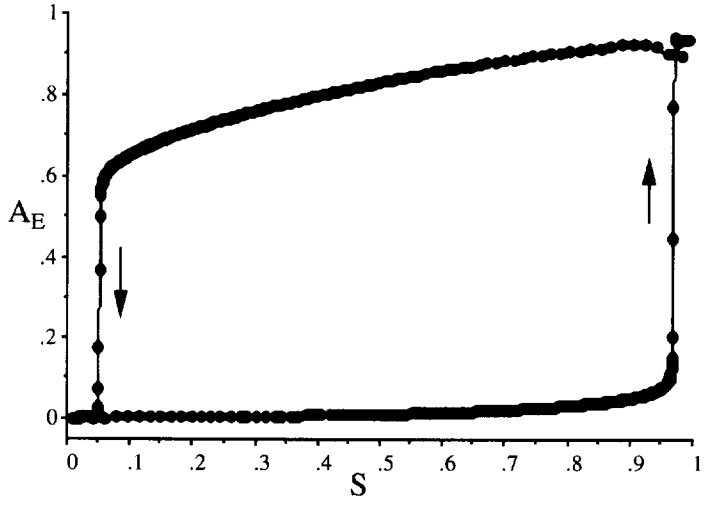

FIGURE 9. Time course of A_E with S varying online. That is, A_E does not reach equilibrium before S is changed again. First S is increased, which causes A_E to increase suddenly if $S \approx 0.95$. Then S is decreased, which causes A_E to decrease suddenly if $S \approx 0.05$. Parameter values are given in Appendix B.

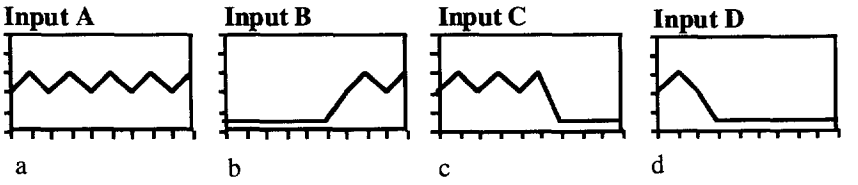

FIGURE 10. The input patterns that are used in the simulation experiments discussed below. The y-axis denotes I_i, the x-axis denotes i. The important features of the input patterns are discussed in the text. These figures are associated with the first column of Table 3.

addition, the activities of both **u0** and **r** are independent of initial values which implies that changes in the input are detected by **r**. The most important addition to ART 2 is a complete implementation of a GDF. The GDF consists of coupled recurrent gated dipoles, which are based on gated dipoles described by Grossberg (1980).

In summary, the resulting system of ode's is mathematically simple compared to ART 2 and obeys all our present requirements: a fully stand alone network of which all the transient network behavior is kept intact. Since mathematical analysis of the transient behavior of the network is only discussed in a few cases, simulation studies are important to perform.

4. EXACT ART: SIMULATION STUDIES

4.1. A Small Classification Problem

Carpenter and Grossberg (1987b) performed a sequence of simulation experiments (figure 8, p. 4928), which they discuss extensively, since it examines important characteristics of the model with regard to stability and plasticity. Most of these characteristics, like the 2/3-rule, were proven to hold for Exact ART. However, some aspects of the model, mainly with regard to transient behavior, are not fully understood yet. Simulation experiments may give further insight and are a crucial test for the performance of the model. The classification problem comprises four specially designed input patterns (Fig. 10), which are presented in a particular sequence.

The following features of the input patterns are essential to test the behavior of Exact ART with regard to stability. First, the sum of activity (i.e., I) differs between input patterns, which requires normalization of patterns. Second, patterns B, C, and D are subset patterns of A in that all highly activated units of B, C, and D have values equivalent to Pattern A and, in addition, some elements of A have activities that are much higher than the corresponding elements of B, C, and D. Third, pattern D is a subset pattern of C, and patterns B and C are exclusive.

The sequence of presentation is A, B, C, A, D. This sequence is presented twice on trials 1–5 and 6–10. A simulation experiment consists of the numerical integration of the entire system of ode's. In case of fast learning, the presentation of each input pattern lasts 500,000 time-units (i.e. the system of ode's is integrated numerically from $t = 0$ to 500,000), which implies that all LTM-ode's reach equilibrium. Some experiments do not include an initialization of activity before the presentation of a following input pattern, apart from the start. Otherwise, initialization consists of setting all activity vectors to zero and setting all transmitter values $z1_j$ and $z2_j$ to γ. Parameter values are listed in Appendix C. Simulations are performed with several values of the vigilance parameter ρ: 0.0, 0.1, 0.5, and 0.7. The numerical integration method we used, Divpag, comes from IMSL (1991, p. 755) and is appropriate for both stiff and unstiff initial value problems.

The input patterns are transformed by layers **x0** and **u0**. Fig. 11a–d shows **u0** after presentation of each of the four input patterns. This figure shows the effect of normalization and contrast enhancement performed by F0. Since **u0** does not depend on learned LTM vectors $\mathbf{z}_j$, these vectors are invariant across the experiments described below.

Table 3 lists the sequence of F2-nodes that become active during each trial and the resulting degree of mismatch. The last number in each cell outside the brackets denotes the F2-node connected to the LTM-trace that represents the input pattern. The preceding numbers are the sequence of F2-nodes that became active but caused a reset. The value inside the brackets, is the mismatch, R, at the end of the trial. In contrast to the last two columns, columns 2–5 show results of Exact ART obtained with

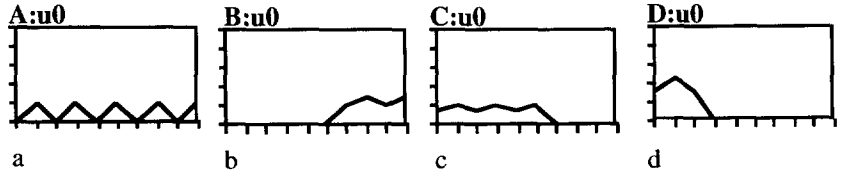

FIGURE 11. Activity vectors u0 after presentation of input patterns *A*, *B*, *C* and *D*, respectively.

TABLE 3
Classification of Input Patterns with Different Values of ρ

Input patterns	With initialization				Without initialization	
	$\rho = 0.7$	$\rho = 0.5$	$\rho = 0.1$	$\rho = 0.0$	$\rho = 0.1$	$\rho = 0.0$
A	1 (0)	1 (0)	1 (0)	2 (0)	3 (0)	3 (0)
B	1, 3 (0)	1, 2 (0)	1 (0.4)	2 (0.4)	3 (0.4)	3 (0.4)
C	1, 4 (0)	1 (0.4)	3 (0)	3 (0)	3, 2 (0)	3 (1)
A	1 (0)	1 (0.4)	1 (0.6)	2 (0.6)	2 (0.4)	3 (0.6)
D	4, 1, 2 (0)	1, 3 (0)	3 (0)	3 (0)	2 (0.6)	3 (1)
A	1 (0)	1 (0.4)	1 (0.6)	2 (0.6)	2 (0.8)	3 (0.6)
B	3 (0)	2 (0)	1 (0.4)	2 (0.4)	2, 3 (0.4)	3 (0.4)
C	4 (0)	1 (0.4)	3 (0.5)	3 (0.5)	3, 2 (0.8)	3 (1)
A	1 (0)	1 (0.4)	1 (0.6)	2 (0.6)	2 (0.8)	3 (0.6)
D	2 (0)	3 (0)	3 (0)	3 (0)	2 (0.6)	3 (1)
Fig. 10	Fig. 12	Fig. 13	Fig. 15	Fig. 15	Fig. 16	Fig. 17

The numbers outside the brackets denote the sequence of active F2-nodes. The numbers inside the brackets are the values of the mismatch R after learning. Figs 12–17 show the associated learned $\mathbf{z}_J$ patterns.

initialization at the beginning of each trial. Without initialization the system detects changes in the input automatically. We will first present some general results and then discuss the individual simulations.

Different values of ρ lead to a different number of formed categories. It turned out that if $\rho = 0.1$ two categories are formed, three categories are formed with $\rho = 0.5$, and each pattern is classified in a distinct category if $\rho = 0.7$. Initialization of the network may also influence the number of categories. In case no initialization takes place and $\rho = 0$ the patterns are classified in only one category. Whereas initialization of activity causes the network to form two categories if $\rho = 0$. It appears that under all performed circumstances the choice of F2-node on the second five trials is equal to the final choice on trials 2–5. This implies that the dynamics of the system leads to a stable recognition code that will be reproduced if the same input sequence is repeated. In addition, the access of the matching F2-node is direct during trials 6–10 if initialization takes place.

In Figs 12–17 learned $\mathbf{z}_J$ traces that are associated with the simulation results in Table 3 are presented. The horizontal axis denotes unit index i. In all figures the range of values on the vertical axis is $[0,1/(1 - d) = 2]$. z_{1i} in Fig. 12a, for example, denotes values of both the BU and the TD $\mathbf{z}_J$-traces, which are equal, after representing input pattern A in z_1 with $\rho = 0.7$. Fig. 12a agrees with the first entry of the second column of Table 3.

With $\rho = 0.7$, each input pattern is represented in a different F2-node. The first time an F2-node is active $\mathbf{u0}$ is copied to $\mathbf{z}_J$. Hence, the corresponding mismatches are zero. This simulation shows also that the search process, which takes place during the first five trials, becomes a direct access during the second five trials. Moreover, the search process does not necessarily include all learned $\mathbf{z}_j$ traces before a new node becomes active (cf., pattern D with $\rho = 0.7$; the fifth entry of the second column of Table 3). The latter depends on the initial values of BU $\mathbf{z}_j$ traces (i.e., $z_{ij}(0)$).

If $\rho = 0.5$ the classification process gives a clear illustration of the 2/3-rule. During the presentation of input pattern A $\mathbf{u0}$ is copied to $\mathbf{z}_1$. If input pattern C is presented the TD expectancy pattern (i.e., $\mathbf{z}_1$) is a superset pattern of the BU input pattern (i.e., $\mathbf{u0}$ in Fig. 11c). In agreement with the 2/3-rule the resulting $\mathbf{z}_1$ pattern at the end of the trial is a section of both patterns. The same holds for $\mathbf{z}_1$ and $\mathbf{u0}$ during the second presentation of input pattern A. Now the TD pattern is a subset pattern and the BU pattern is a superset pattern.

In contrast to fast learning, if LTM-traces do not reach equilibrium the 2/3-rule is weak. This is illustrated by Fig. 14, which shows the results of a simulation equivalent to Fig. 13 except for the presentation time of input

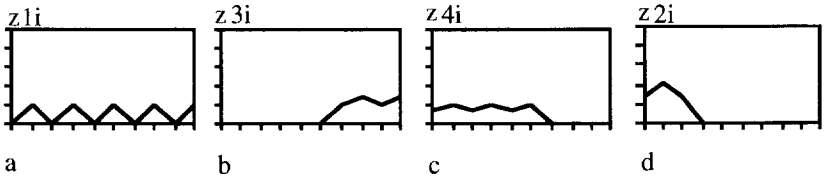

FIGURE 12. (a)–(d) show z_J-traces after classifying patterns *A*, *B*, *C*, and *D* respectively, with $\rho = 0.7$ and with initialization of activity before presentation of each input pattern. The z_J-traces do not change after the first presentation of the input sequence. This figure is associated with the second column of Table 3.

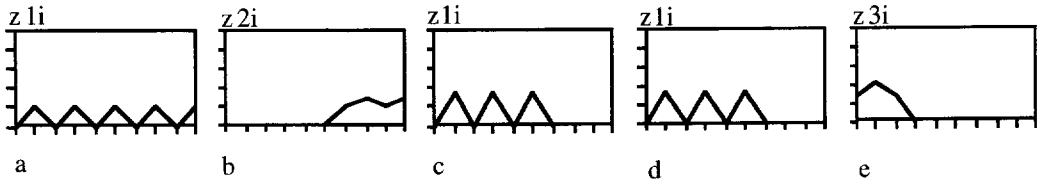

FIGURE 13. (a)–(e) show z_J-traces after classifying patterns *A*, *B*, *C*, *A*, *D* respectively, with $\rho = 0.5$ and with initialization of activity before presentation of each input pattern. The traces z_1, z_2, and z_3 do not change after the first presentation of the input sequence. This figure is associated with the third column of Table 3.

patterns (50,000 instead of 500,000). Fig. 14a shows $\mathbf{z}_1$ after the first trial; $\mathbf{u0}$ is copied in $\mathbf{z}_1$. Fig. 14b shows the same LTM-trace after presentation of input pattern C. It appears that the nodes (and consequently also the corresponding $\mathbf{z}_1$ traces) that did not receive TD input become zero. In contrast, the nodes that only received BU input are quenched only a little. Fig. 14c shows the $\mathbf{z}_1$-trace after the second presentation of input pattern A. Now the nodes that received less TD input than BU input are depressed with regard to $\mathbf{u0}$ and enhanced with regard to $\mathbf{z}_1$.

In all cases learning should stabilize rapidly. Fig. 15 shows the results of simulations with $\rho = 0$. In that case no reset takes place. Nevertheless, a stable recognition code, consisting of two classes is formed. If $\rho = 0.1$ the same recognition code is formed.

Figs 16 and 17 show results of simulations without initialization of activity at the beginning of each trial. As expected, changes in the input are detected automatically, but only if these changes are large enough to cause a mismatch. Hence, the recognition code shown in Fig. 16 ($\rho = 0.1$ without initialization) differs from the recognition code shown in Fig. 17 ($\rho = 0.1$ with initialization). If $\rho = 0$ no resets can take place. Hence, all input patterns are classified in one category even if the BU input pattern and the learned $\mathbf{z}_J$-trace cause a maximum mismatch (i.e., $R = 1$). Fig. 17 shows that even in the latter case learned recognition codes are not removed.

From the above analyses we can conclude that the classification behavior of Exact ART agrees with that of ART 2 with regard to stability and plasticity properties. Although we mainly discussed the equilibrium behavior the most important conclusion can be drawn about transient behavior: transient behavior does not seriously affect the stability of the system. One exception is, however, that the F1-activity is sensitive to either the rapidity of the LTM-ode's or the rapidity of F0-ode's. If the LTM ode's are calculated too fast or F0-ode's are calculated too slow then F1 performs an additional contrast enhancement. Nevertheless, classifications will still be performed, but the requirements of the match procedure are not assured. To solve this problem, we decided to introduce a parameter ϕ which regulates the rapidity of F0 ode's.

After examination of the transient behavior of individual modules in Section 3 we will now take a closer look at these modules working together. Fig. 18 shows the values of R, S, $_J$ (the index of the winning node $y5_J$ in F2), and A_E during the first classification of input pattern D with initialization and $\rho = 0.7$. Two resets are involved. Fig. 18 illustrates some properties of transient behavior. First, a reset only occurs if R is stabilized. At the beginning of the trial and just after each reset (between 0–10, 160–170, and 375–385) R is temporarily increased above $1 - \rho$. Due to the buffer in S this does not induce a reset. Second, a reset is faster induced if R is larger. This follows directly from eqn (21).

4.2. A Larger Classification Problem

In the former section we studied a very small classification problem with which we could examine important aspects of the stability and plasticity of Exact ART during fast learning. Only for illustration, also the outcomes of a larger classification problem are shown. The input set contains 50 patterns. All patterns are presented several times in a fixed sequence. Each pattern is presented during 1500 time units. This means that fast learning is not assured for each presentation of an input pattern. With respect to Appendix C, the following parameter values are changed: $N = 25$, $M = 26$, $\theta = 0.04$, $\alpha = 0.01$, $z_{ij}(0) = 0.04$. Two classifications are made: one

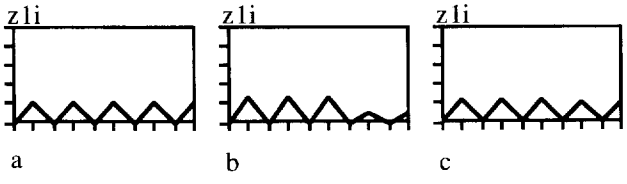

FIGURE 14. (a)–(c) show z_J-traces after classifying patterns *A*, *C*, *A* respectively and with $\rho = 0.5$, with initialization of activity before presentation of each input pattern. In contrast to Fig. 13 learning is not fast.

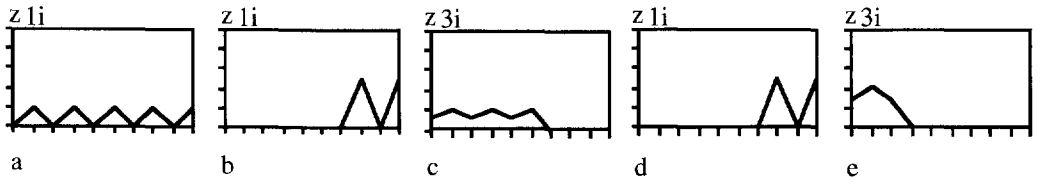

FIGURE 15. (a)–(e) show z_J-traces after classifying patterns *A*, *B*, *C*, *A*, *D* respectively, with $\rho = 0.1$ and with initialization of activity before presentation of each input pattern. The traces z_1, z_2, and z_3 do not change after the first presentation of the input sequence. This figure is associated with the fourth and fifth column of Table 3.

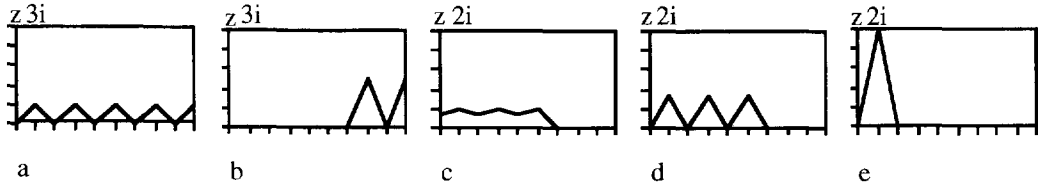

FIGURE 16. (a)–(e) show z_J-traces after classifying patterns *A*, *B*, *C*, *A*, *D* respectively, with $\rho = 0.1$ and without initialization of activity before presentation of each input pattern. The traces z_2, and z_3 do not change after the first presentation of the input sequence. This figure is associated with the sixth column of Table 3.

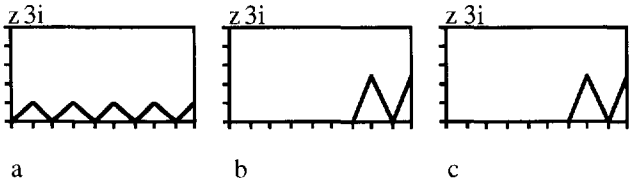

FIGURE 17. z_3-trace after presenting *A*, *B*, *C*, *A* and *D* in sequence with $\rho = 0$, without initialization of activity when a following input pattern is presented. This figure is associated with the last column of Table 3. (a) z_3 after the first presentation of *A*. (b) z_3 after learning Pattern *B*. (c) All further patterns are classified by node 3; the z_3-trace changes no more. Note that u0 during presentation of input pattern *D* and z_3 are orthogonal. Nevertheless, the learned z_3-trace is not changed. That is, learned information is not removed.

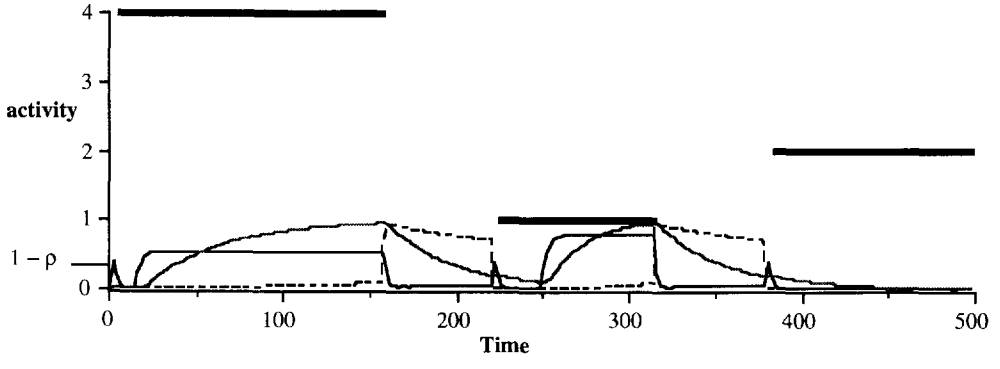

FIGURE 18. Time plot of *R* (plain black line), *S* (plain grey line), the index of the winning y5-node (thick lines), and A_E (dashed line) as a function of time. The figure shows the trial of the first classification of pattern *D*. Time = 0 at the beginning of the trial. $\rho = 0.7$ and initialization takes place at the beginning of the trial.

classification with $\rho = 0.8$, and one classification with $\rho = 0.6$. Fig. 19 shows the input patterns grouped by the category to which they belong during the third presentation with $\rho = 0.6$. After the second presentation only two patterns were classified in a different category. This means that a stable recognition code is formed rapidly. Fig. 20 shows the input patterns grouped by the category to which they belong during the fourth presentation with $\rho = 0.8$. After the third presentation only one pattern was classified in a different category.

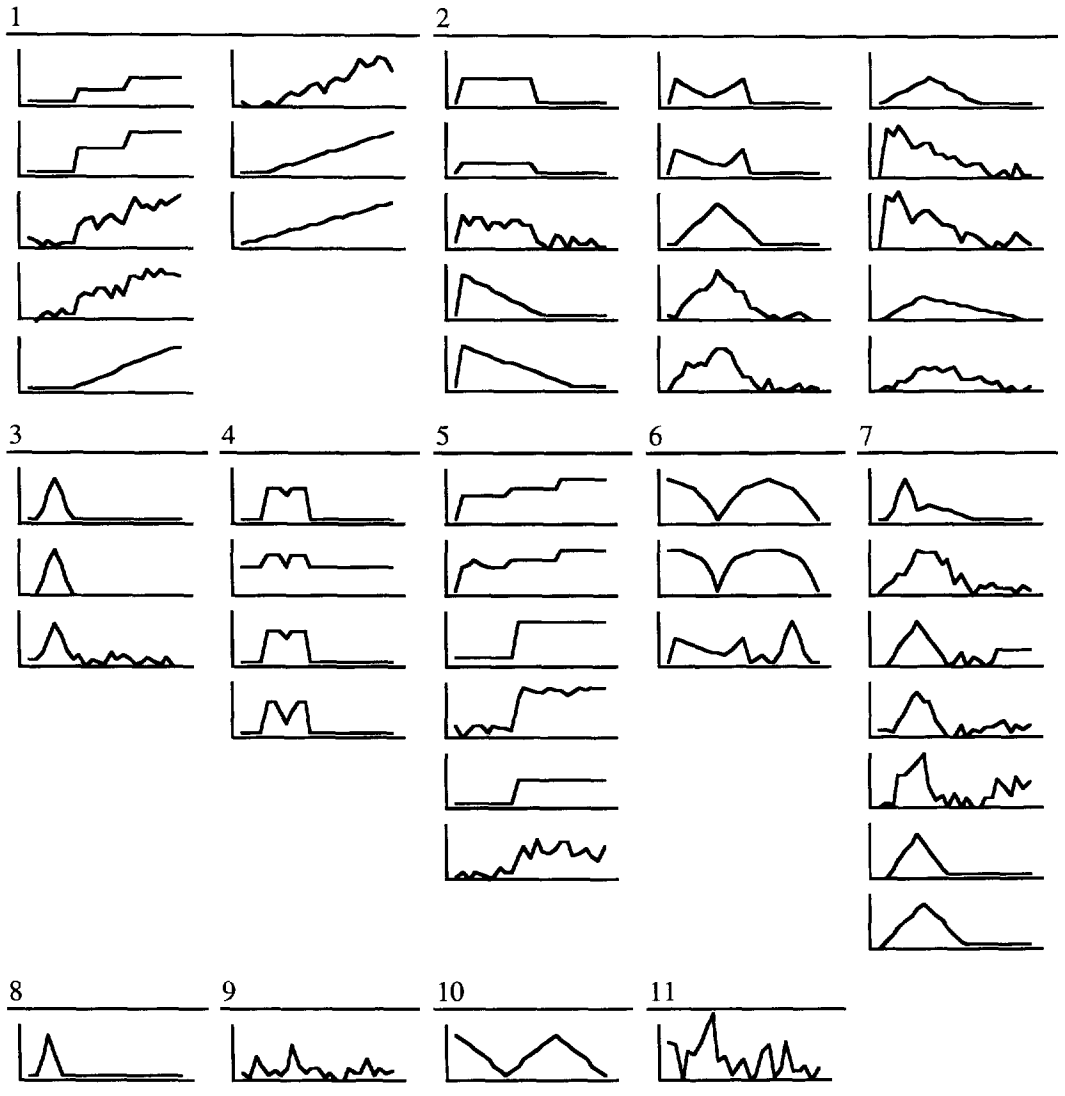

FIGURE 19. The resulting classification of 50 input patterns with $\rho = 0.6$. All patterns are presented in a specified sequence. This sequence is presented three times. Learning is not fast and the network is initialized at the beginning of each trial.

5. CONCLUSION

In this article we introduced Exact ART, an implementation of ART. Exact ART is completely defined by a system of ode's, which does not hold for the current implementations of ART. Exact ART is primarily based on ART 2 and constitutes an attempt to implement the theoretical framework in Grossberg (1980). This results in a completely stand-alone network that self-organizes stable recognition codes for analog input patterns. In Sections 3 and 4 we showed by means of analytical methods and simulation studies that Exact ART obeys all demands with regard to stability and plasticity that counted as design principles for ART 2 (Carpenter & Grossberg, 1987b). Moreover, the classification of analog input patterns is equivalent to the classification made by ART 2. Below, the main differences between ART 2 and Exact ART are summarized.

First of all, the system of ode's, which defines Exact ART completely, is stand-alone. In contrast to ART 2, the orienting subsystem and F2 are implemented by ode's. The structure of the orienting subsystem is a neural oscillator. F2 is a GDF. The introduction of a GDF was the major change we made. Furthermore, being stand-alone means that initialization of activity before the presentation of each input pattern is not necessary (the latter also holds for ART 3). As a consequence of the complete implementation, however, the computing of larger classification problems requires a lot of CPU-seconds and RAM. This is mainly due to the application of a numerical-integration algorithm that is appropriate for stiff problems. Since it was not our aim to develop an efficient model, but to study the dynamics of a realistic model, we do not consider the expensive computing effort as a problem. Moreover, since the classification algorithm is relatively simple in the fast learning case, a classification of a set of input patterns can be calculated quickly.

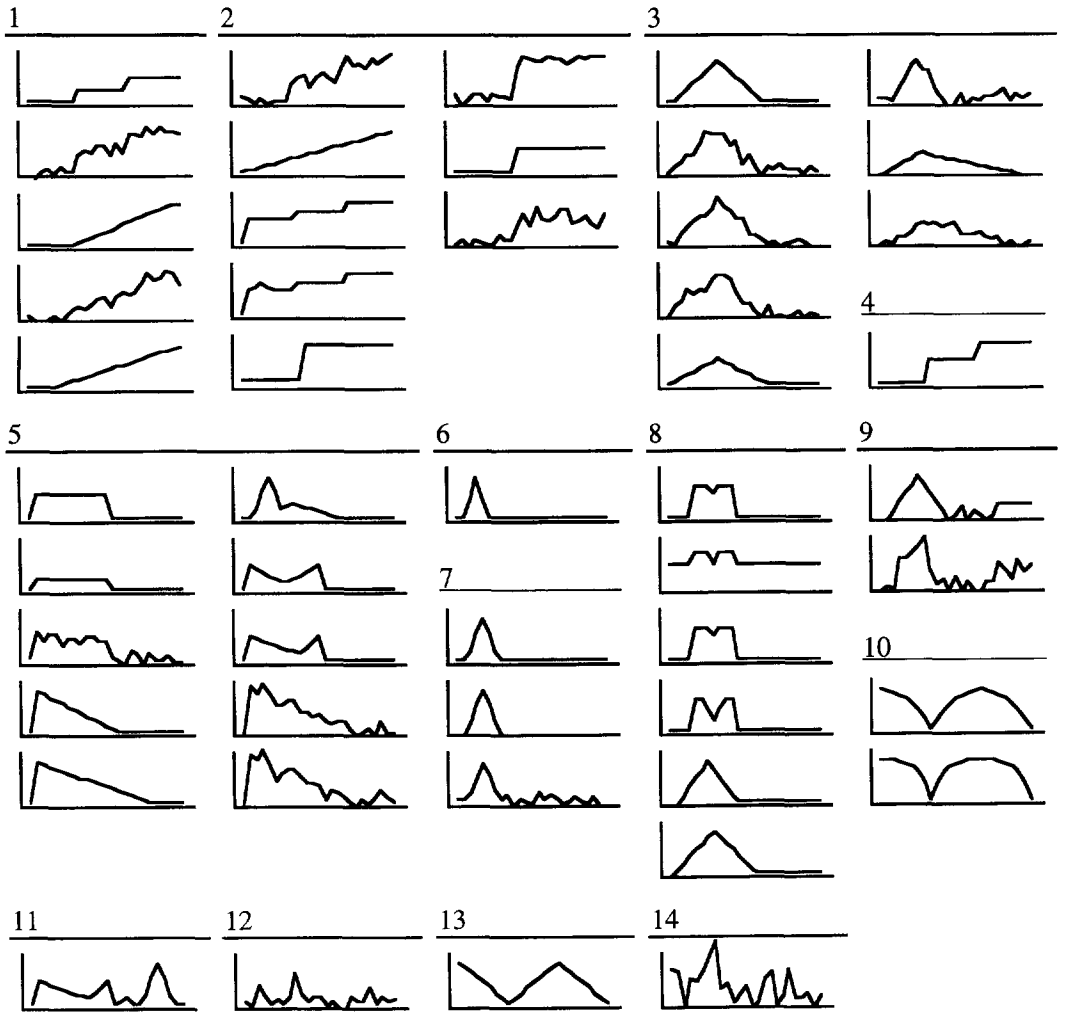

FIGURE 20. The resulting classification of 50 input patterns with $\rho = 0.8$. All patterns are presented in a specified sequence. This sequence is presented three times. Learning is not fast and the network is initialized at the beginning of each trial.

A second major difference is a simultaneous instead of sequential update of STM and LTM ode's. As a consequence the 2/3-rule is strong instead of weak in the fast-learning case. A strong 2/3-rule implies that TD expectancies stabilize in less presentations of input sequences. Moreover, the classification procedure seems easy to describe mathematically, since learning agrees with transforming the active TD expectancy pattern ($\mathbf{z}_J$) by taking the section with the active BU pattern ($\mathbf{u0}$). It would be interesting to make a mathematical description of the classification process and to compare this description with other unsupervised classification procedures, for example cluster analysis.

A third difference between ART 2 and Exact ART is that the characteristics of the match procedure and the 2/3 rule are derived mathematically. The proofs are relatively simple because we changed the normalization features of activity layers and we moved the partially linear function f from F1 to F0 and the LTM ode's.

In general, we can say that Exact ART is a proper implementation of ART that maintains the features of ART with regard to biological plausibility, psychological plausibility, and environmental constraints. Moreover, it shows very complex sequential behavior that is completely regulated by a system of ode's which does not contain explicit descriptions of the sequence of events that should take place. The sequential behavior seems to agree with (cognitive) psychological concepts as "hypothesis testing", "arousal events", and STM. At the same time, Exact ART obeys very basic physical and biological demands like stand-alone running in real time.

REFERENCES

Carpenter, G.A., Grossberg, S., & Rosen, D.B. (1991a). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, **4 (6)**, 759–771.

Carpenter, G.A., Grossberg, S., & Rosen, D.B. (1991b). ART 2-A: An

adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks*, **4 (4),** 493–504.

Carpenter, G., & 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., & Grossberg, S. (1987). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, **26 (23),** 4919–4930.

Carpenter, G., & Grossberg, S. (1990). ART 3: Hierarchical search using Chemical Transmitters in Self-Organizing Pattern Recognition Architectures. *Neural Networks*, **3,** 129–152.

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

Ellias, S.A., & Grossberg, S. (1975). Pattern formation, contrast control, and oscillations in the short term memory of shunting on-centre off-surround networks. *Biological Cybernetics*, **20,** 69–78.

Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, **52 (3),** 213–257.

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

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

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, **87,** 1–51.

Grossberg, S., & Schmajuk, N.A. (1987). Neural dynamics of attentional modulated pavlovian conditioning: conditioned reinforcement, inhibition, and opponent processing. *Psychobiology*, **15,** 195–240.

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

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

IMSL (1991). *Fortran subroutines for mathematical applications* ((Math/Library)). Houston, TX: IMSL inc.

Kentridge, R.W. (1994). Critical dynamics of neural networks with spatially localized connections. In M. Oaksford & G.D.A. Brown (Eds), *Neurodynamics and psychology* (pp. 181–214). London: Academic Press.

Raijmakers, M.E.J., & Molenaar, P.C.M. (1994). Exact ART: *A complete implementation of an ART network, including all regulatory and logical functions, as a system of differential equations capable of stand-alone running in real time. Internal Report*: Amsterdam: Department of Developmental Psychology, University of Amsterdam.

Raijmakers, M.E.J., Van Koten, S., & Molenaar, P.C.M. (1996). On the validity of simulating stagewise development by means of PDP-networks: Application of catastrophe analysis and an experimental test for rule-like network performance. *Cognitive Science*, **20 (1),** 101–136.

Ryan, T.W., & Winter C.L. (1987). Variations on Adaptive Resonance. In M. Caudill & C. Butler (Eds.), *IEEE First International Conference on Neural Networks*. San Diego, CA: IEEE, (pp. 767–775).

Ryan, T.W., Winter, C.L., & Turner, C.J. (1987). Dynamic control of an artificial neural system: the property inheritance network. *Applied Optics*, **26 (23),** 4961–4971.

Schuster, H.G., & Wagner, P. (1990). A model of neural oscillations in the visual cortex. *Biological Cybernetics*, **64,** 77–82.

Sepulchre, J.A., & Babloyantz, A. (1991). Spatio-temporal patterns and network computation. In A. Babloyantz (Ed.), *Self-organization, emergent properties and learning. Nato ASI Series B: Physics (Vol. 260)*. New York: Plenum Press.

Van der Maas, H., & Molenaar, P.C.M. (1992). Stagewise cognitive development: An application of catastrophe theory. *Psychological Review*, **99 (3),** 395–417

APPENDIX

Notation

The length of F0 and F1 vectors	N
The length of F2 vectors	M
A vector name is written bold, elements italic:	$\mathbf{x} = (x_1,...,x_i,...,x_N)$
Activity of an individual unit i in F1:	x_i
Activity vector at time t:	$\mathbf{x}(t) = (x_1(t),..., x_i(t),..., x_N(t))$
Sum of the elements of vector **x0**, **u0**, **x**, **u**, **q**, **p** is respectively	$x0$, $u0$, x, u, q, p
Relative activity of an unit x_i, i.e., $\frac{x_i}{x}$, is written in capitals	X
The input vector is	$\mathbf{I} = (I_1,...,I_i,...,I_N)$
Sum of the elements of vector **I** is	I
Sum of the elements of vector **r**, the match measure, is	R
Relative value of input unit i, $\frac{I_i}{I}$:	Ji
Activity vectors of F2, $k \leq 6$:	$\mathbf{yk} = (yk_1,...,yk_j,...,yk_M)$
The subscript of the active $y5_j$ node is	J
LTM connection between a unit $y5_j$ of F2 and p_i of F1:	z_{ji}
LTM connection between an unit p_i of F1 and $y5_j$ of F2:	z_{ij}
The LTM vector between F1 and the active F2-node $_J$ can be both top-down (TD):	$\mathbf{z}_J = (z_{J1},...,z_{Ji},...,z_{JN})$
and bottom-up (BU):	$\mathbf{z}_J = (z_{1J},...,z_{iJ},...,z_{NJ})$
The sum of elements of a vector $\mathbf{z}_J$ multiplied by parameter d:	χ
Vector of relative weights:	$\mathbf{Z}_J = (Z_{J1},...,Z_{Ji},...,Z_{JN})$
or:	$\mathbf{Z}_J = (Z_{1J},...,Z_{iJ},...,Z_{NJ})$

A. Proofs Concerning F1 and the Orienting Subsystem

From Section 3.1 follows that in equilibrium:

$$p_i = dz_{Ji} + u_i \tag{A1}$$

$$q_i = \frac{p_i}{\sum_{k=1}^{N} p_k} = \frac{u_i + dz_{ji}}{\sum_{k=1}^{N} u_k + d\sum_{k=1}^{N} z_{Jk}} \tag{A2}$$

eqn (A2) is equivalent to:

$$q_i = \frac{u_i + \chi Z_{Ji}}{\sum_{k=1}^{N} u_k + \chi} \tag{A3}$$

$$Z_{Ji} = \frac{z_{Ji}}{\sum_{k=1}^{N} z_{Jk}} \text{ if } \sum_{k=1}^{N} z_{Jk} > 0,\ Z_{Ji} = 0 \text{ otherwise}$$

$$\chi = d\sum_{k=1}^{N} z_{Jk}$$

Furthermore:

$$u_i = \frac{b^* q_i + x_i}{b\sum_{k=1}^{N} q_k + \sum_{k=1}^{N} x_k} \tag{A4}$$

$$x_i = \frac{a^* u_i + u0_i}{a\sum_{k=1}^{N} u_k + \sum_{k=1}^{N} u0_k} \tag{A5}$$

Since in equilibrium holds that

$$\sum_{k=1}^{N} u_k = \sum_{k=1}^{N} q_k = \sum_{k=1}^{N} u0_k = \sum_{k=1}^{N} x_k = 1$$

eqn (A3) is reduced to:

$$u_i = q_i(1+\chi) - \chi Z_{Ji} \tag{A6}$$

Substitute eqn (A6) in eqn (A4):

$$(b+1)(q_i(1+\chi) - \chi Z_{Ji}) = bq_i + x_i \tag{A7}$$

Substitute eqn (A6) in eqn (A5):

$$X_i = \frac{u0_i + a(q_i(1+\chi) - \chi Z_{Ji})}{1+a} \tag{A8}$$

Substitute eqn (A8) in eqn (A7):

$$(1+a)[(b+1)(q_i(1+\chi) - \chi Z_{Ji}) - bq_i] = u0_i + a(q_i(1+\chi) - \chi Z_{Ji})$$

$$\Leftrightarrow \quad q_i = \frac{u0_i + \chi(1+b+ab)Z_{Ji}}{(1+\chi(1+b+ab))} \tag{A9}$$

B. The Gated Dipole Field

Assuming that no arousal has occurred, in equilibrium the following equalities hold for the GDF (Section 3.1):

$$y1_j = y5_j \tag{A10}$$

$$y2_j = y6_j \tag{A11}$$

$$y3_j = zl_j yl_j \tag{A13}$$

$$y4_j = z2_j y2_j \tag{A14}$$

$$y5_j = \left\{ \begin{array}{l} 1 \text{ if } I_j > I_k \text{ for all } k \neq j \\ 0 \text{ otherwise} \end{array} \right\} \tag{A15}$$

$$y6_j = y4_j - y3_j \tag{A16}$$

$$zl_j = \frac{\beta\gamma}{\beta + \delta[yl_j - \Gamma]^+} \tag{A17}$$

$$z2_j = \frac{\beta\gamma}{\beta + \delta[y2_j - \Gamma]^+} \tag{A18}$$

If an arousal event occurs after stabilization of a winner in **y5**, the following analysis concerning the transient activity can be made. The analysis concerns situation 2 described in Section 3.5.2. Rewriting eqn (13), defined in Section 3.1, gives:

$$\frac{dy5_j}{dt} = B(h(y5_j) + y3_j + eI_j) - y5_j \left(\sum_{k=1}^{M} h(y5_k) + y4_j + y3_j + eI_j + A \right) \tag{A19}$$

If $y5_j \approx 0$, which holds for $j \leq M-1$ then

$$\frac{dy5_j}{dt} \approx B(y3_j + eI_j) \approx B(zl_j A_E + eI_j) \tag{A20}$$

If $y5_j \approx B \equiv 1$, which holds for $j = M$, then

$$\frac{dy5_j}{dt} \approx -B(y4_j + A), \text{ since } h(y5_j) \approx \sum_{k=1}^{M} h(y5_k) \tag{A21}$$

eqn (A21) implies that $dy5_M/dt$ becomes negative, and hence $y5_M$ is depressed. In contrast, $y5_k$, $k < M$, increases. If $y5_k \approx y5_l$ for $k, l \leq M$:

$$\frac{dy5_k}{dt} - \frac{dy5_l}{dt} \approx (B - y5_k)((y5_k + A_E)(zl_k - zl_l) + eI_k - eI_l) \tag{A22}$$

From eqn (A22) we can deduce the resulting order of the grow rates of nodes $y5_j$. The latter is described in Section 3.5.

C. Default Parameter Values

F0 and F1 parameters	
Number of F0-units and F1-units (N)	10
Decay (A)	0.001
Upper bound of activation (B)	1.0
Pos. connection from q to u (b)	10.0
Pos. connection from u to x (a)	10.0
Rapidity of F0 equations (ϕ)	100.0
F2 parameters	
Number of F2-units (M)	4
Strength of input to F2 (e)	0.01
Strength of input from F1 to F2 (d)	0.5
Increase of transmitter (β)	0.5
Decrease of transmitter (δ)	5.0
Threshold of transmitter (Γ)	0.1
Upperbound of transmitter (γ)	0.5
Rapidity of dipole-transmitter equations (ε)	0.001
LTM parameters	
Initial weights ($z_{ij}(0)$)	0.01
Initial weights ($z_{ji}(0)$)	0.0
Threshold in LTM equations (θ_y)	0.9
Rapidity of LTM (F1 < - > F2) (α)	0.005
OSS parameters	
Vigilance (ρ)	0.5
Rapidity of OSS equations (η)	100.0
Rapidity of S equation (ι)	0.1
Decrease of S (ω)	1.0
General parameters	
Threshold in signal function (θ)	0.1
Strength of arousal (c)	10.0
Integration parameters	
Duration of 1 trial	500,000
Tolerance of numerical integration	0.0001