Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 2727–2731, May 1987
Biophysics

Neural networks that learn temporal sequences by selection

(neural Darwinism/allosteric receptors/heterosynaptic modulation/Hebb synapse)

STANISLAS DEHAENE*, JEAN-PIERRE CHANGEUX†‡, AND JEAN-PIERRE NADAL§

*Laboratoire de Sciences Cognitives et Psycholinguistique, 54 Boulevard Raspail, F 75270 Paris Cédex 06, France; †Unité de Neurobiologie Moléculaire and Laboratoire Associé au Centre National de la Recherche Scientifique, no. 270, Interactions Moléculaires et Cellulaires, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cédex 15, France; and §Groupe de Physique des Solides, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris Cédex 05, France

Contributed by Jean-Pierre Changeux, December 15, 1986

ABSTRACT **A model for formal neural networks that learn temporal sequences by selection is proposed on the basis of observations on the acquisition of song by birds, on sequence-detecting neurons, and on allosteric receptors. The model relies on hypothetical elementary devices made up of three neurons, the synaptic triads, which yield short-term modification of synaptic efficacy through heterosynaptic interactions, and on a local Hebbian learning rule. The functional units postulated are mutually inhibiting clusters of synergic neurons and bundles of synapses. Networks formalized on this basis display capacities for passive recognition and for production of temporal sequences that may include repetitions. Introduction of the learning rule leads to the differentiation of sequence-detecting neurons and to the stabilization of ongoing temporal sequences. A network architecture composed of three layers of neuronal clusters is shown to exhibit active recognition and learning of time sequences by selection: the network spontaneously produces prerepresentations that are selected according to their resonance with the input percepts. Predictions of the model are discussed.**

The central nervous system does not process information under conditions of static equilibrium, but is in constant dynamic interaction with the outside world and possesses the striking faculties to recognize, store, and produce temporal patterns (1). Yet the many attempts to model neural networks such as the Boltzmann machine (2) or the Hopfield model (3) dealt with systems under static conditions. Models of dynamic behavior have been suggested (4–9), but most are grounded on rather ad hoc assumptions on the architectures of the networks or on the learning rules.

In this communication, we propose a model of formal neural networks that learn temporal patterns on the basis of a set of biologically plausible assumptions. Following our current hypothesis (10, 11), storage of sequences in such networks does not take place as a passive instructive print but is rather the result of active selection (10, 12) among spontaneously generated prerepresentations.

Biological Premises

The model rests upon the following biological premises.

Premise 1. Acquisition of Song Behavior in Birds. Singing behavior in birds like *Melospiza georgiana* or *Melospiza melodia* is acquired in two distinct phases (13, 14): a sensory phase, during which the bird hears and memorizes adult songs, and a sensory–motor (or imitation) phase, during which the bird attempts to imitate the memorized songs. The imitation phase starts with the production of almost unstructured utterances named subsong. The next step is plastic song, where syllables can be identified, but in a number four to five times larger than required to cover the adult species repertoire (overproduction) (13). Finally, the "crystallization" of the adult song is accompanied by the elimination of the superfluous syllables (selective attrition) (13). Although the picture is obviously more complex in the case of human language acquisition, a parallel between baby babbling and bird subsong has been drawn (13).

Premise 2. Sequence-Specific Neurons. In one of the nuclei involved in bird-song production, referred to as HVc, single-cell recordings show discharge patterns temporally linked to syllable production, some of them with a high degree of selectivity for a given syllable. Many neurons of HVc also respond to auditory features. A number of them, called song-specific neurons (13, 15), detect sequences of syllables. Sequence-detecting neurons have also been identified in other systems, such as visual systems (16) or the auditory cortex of the bat (17).

Premise 3. Allosteric Receptors as Regulators of Synaptic Efficacy. The known postsynaptic receptors for neurotransmitters are allosteric proteins (18, 19) that carry several categories of distinct binding sites and may exist under several interconvertible conformations. For instance, the channel-linked nicotinic acetylcholine receptor undergoes conformational transitions between at least four states: a resting state (R), an open, active state (A) (activation reaction), a rapidly desensitized state (I), and a slowly desensitized state (D) (desensitization reactions) where the ion channel is closed. The transitions between the last two states (respectively, potentiation and depression), are slower than the activation reaction and can be modulated by chemical effectors or by the electrical potential. It thus offers a simple mechanism for the regulation of the efficacy of a given synapse by the activity of another synapse (heterosynaptic regulation) (20), a feature that has been experimentally observed in several systems (21, 22).

Model

The proposed model is based on the following assumptions.

Assumption 1. The Synaptic Triad. The efficacy of a synapse of neuron A on neuron B can be influenced by the activity of a third neuron C, called a modulator (20). The ordered triplet of neurons A–B–C will be called a "synaptic triad" (Fig. 1*a*) if the following conditions are met: (*i*) the synapse of neuron A on neuron B is excitatory, (*ii*) prolonged activity of neuron C (on the order of 0.1 sec) causes the A–B synaptic efficacy to increase toward a maximum value, and (*iii*) long-lasting rest of neuron C causes this efficacy to drop toward a minimal value.

Postsynaptic neuron B of a synaptic triad A–B–C behaves as a sequence detector on neurons C and A: neuron A has no influence on the activity of B, unless synapse A–B has been potentiated—that is, unless neuron C has just been active.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡To whom reprint requests should be addressed.

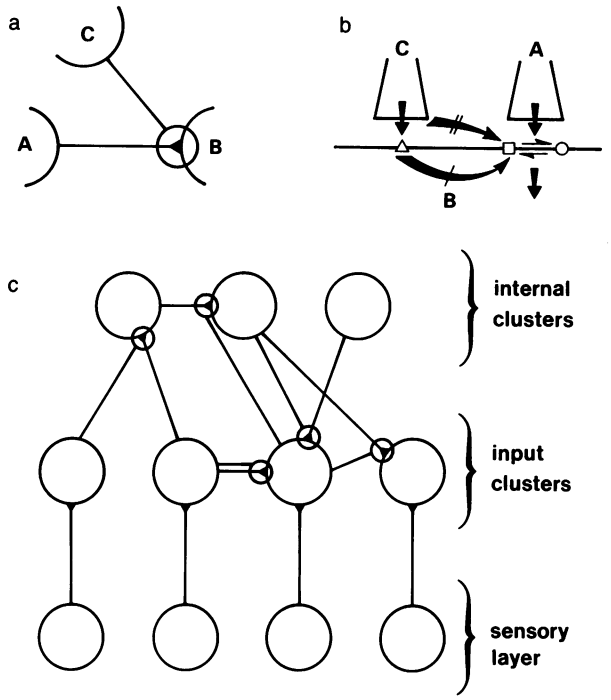

FIG. 1. (*a* and *b*) Formal representation and molecular implementation of a synaptic triad. A, presynaptic neuron; B, postsynaptic neuron; C, modulator. Signals from synapse C–B modulate the efficacy of synapse A–B via allosteric transitions of the postsynaptic receptor. Neuron B detects a transition of activity from neuron C to neuron A. (*c*) Architecture of an active recognition network. Each circle represents a neuronal cluster. Percepts are imposed on the perceptual layer. Internal clusters produce hypotheses or prerepresentations on the incoming signal. Resonance between internal and external signals takes place in the input clusters.

Temporal constraints are (*i*) that activity in neuron C immediately precedes activity in neuron A and (*ii*) that activity in neuron C lasts a critical duration ≈0.1 sec.

A simple molecular illustration of the synaptic triad (for other possibilities, see ref. 22) is the regulation of the conformational states of an allosteric receptor in the postsynaptic membrane of synapse A–B by a signal produced by a second synapse, synapse C–B (Fig. 1*b*). Such a signal may modulate the efficacy of the synapse A–B via diffusible chemical messengers, which may or may not include covalent modification, that would favor the rapidly desensitized (I) state of the receptor. For simplicity, we adopt the idealized situation where synapse C–B has a negligible efficacy on the electrical potential of neuron B.

Again to simplify, we define by activity the binary parameter describing the presence or absence of the high-frequency firing of a given neuron. In the present state of modeling, single action potentials are not considered to be representative parameters.

Assumption 2. Clusters of Synergic Neurons and Intercluster Bundles. The neural network considered here is formed of juxtaposed clusters of synergic neurons [see also Edelman's "groups" (12)]. For instance, each cluster is composed of ≈100 neurons, which are densely interconnected by excitatory synapses. Activity of a cluster is defined as the fraction of active neurons in the cluster and is, therefore, a continuous, not a binary parameter. Intracluster excitatory synapses generally provide two stable states of activity: most neurons are active, or most neurons are resting. These states are weakly sensitive to the presence of noise or the loss of single neurons or synapses.

Extracluster synapses link clusters into a neuronal assembly. The level of description chosen is that of ensembles of synapses. The ensemble of all nonmodulated synapses A–B with all presynaptic neurons A in cluster **A** and all postsynaptic neurons B in cluster **B** will be called bundle A–B. Clusters **A** and **B** are, respectively, the anterior and the posterior clusters of the bundle. Similarly, the ensemble of all synaptic triads A–B–C with neurons A, B, and C belonging to the same clusters **A**, **B**, and **C** will be called a modulated bundle **A–B–C**, where cluster **C** becomes the modulator cluster. A bundle is characterized by its efficacy, the average of the efficacies of its synapses.

Within an assembly, it will be useful to separate input clusters, which receive direct bundles from clusters outside the assembly, from output clusters, which send bundles outside, and from internal clusters [hidden units (2)], which are not directly connected to the exterior. We will consider only the following type of connectivity between clusters: within an assembly, (*i*) inhibitory, nonmodulated bundles that provide lateral inhibition (23) and (*ii*) excitatory modulated bundles that are temporarily desensitized or potentiated as a function of internal activity and, between assemblies, (*iii*) excitatory nonmodulated bundles that propagate the outputs of intraassembly computations.

Assumption 3. Learning Rule for Synaptic Triads. We consider modifications of synaptic efficacy that apply to adult and developmental learning and do not require any modification of the graph of connections [yet if one assumes that low-efficacy synapses degenerate, our rule also accounts for epigenetic disconnection (10)]. We thus take as initial state a network with already segregated neuronal clusters and deal with subsequent stages of learning, where only modulated bundles can be modified.

The local Hebbian (24) learning rule adopted for a synaptic triad A–B–C is the following: (*i*) If synapse A–B has recently contributed to the postsynaptic potential of neuron B, that is if neurons C and A have been activated in that order, then (*ii*) the maximum efficacy of synapse A–B is modified. It increases toward an absolute maximum when, after integration of the postsynaptic potentials generated by synapse A–B, postsynaptic neuron B was activated, and its firing rate reached a threshold value. If on the contrary neuron B remained silent, then the maximum efficacy of synapse A–B drops toward zero.

This rule causes the selection of synaptic triads that stabilize the ongoing temporal activity of the network, and the elimination of those that perturb it. The rule applies only to excitatory modulated synapses constituting synaptic triads. The expression of the rule is the same for a modulated bundle as for a single triad.

The rule involves only modifications of maximum synaptic efficacies, not of time constants. A simple implementation in terms of allosteric receptors (18, 19) is that the postsynaptic receptor of synapse A–B is able to exist under a slowly desensitized state D in addition to the fast desensitized state I. The absolute maximum of synaptic efficacy is determined by the total number of receptor molecules in the synapse. At time t, the synaptic efficacy is bounded by the number of molecules in the states R, A, and I, the transitions to and from I state taking place in ≈0.1 sec. According to our definition, learning would take place on the slowly desensitized state D and consist of the regulation, by chemical messengers (including covalent modifications) or by electrical activity, of the fraction of receptor molecules present in that state for seconds, minutes, or even hours.

Formalization

The model is formalized at the level of clusters of neurons and synaptic bundles. The activity of a cluster i is represented by a continuous variable $S_i(t)$ between 0 and 1. Nonmodulated bundles between clusters are represented by the connection

matrix ($\mathbf{V_{ij}}$), where V_{ij} is the efficacy of the bundle with anterior cluster j and posterior cluster i. This static connectivity is such that each cluster excites itself with efficacy V_α and inhibits others with efficacy V_β. Thus for a totally connected assembly, matrix ($\mathbf{V_{ij}}$) has values V_α on the diagonal and values V_β elsewhere. Modulated bundles are triplets of clusters (anterior, posterior, and modulator). Each modulated bundle b has an instantaneous efficacy $W_b(t)$, which varies between 0 and a maximum value $W_b^m(t)$, which is also variable between bundles.

The V_{ij} are never modified, even during learning. All the variables $S_i(t)$, $W_b(t)$, and $W_b^m(t)$ obey a discrete time-step dynamics: $S_i(t)$ is updated according to

$$S_i(t+1) = F\left[\sum_j V_{ij}S_j(t) + \sum_{b=(a,i,m)} W_b(t)S_a(t) + N\right], \quad [1]$$

where F is the sigmoid function

$$F(x) = \frac{1}{1+e^{-x}}. \quad [2]$$

The first summation takes static connectivity into account. The second summation is made over all modulated bundles b with posterior cluster i. $S_a(t)$ is then the activity of the anterior cluster of bundle b. N is a noise term with the uniform distribution $[-n,n]$. Thresholds are omitted for simplicity.

The instantaneous efficacy $W_b(t)$ of each bundle b is modified according to the level of activity of the modulator cluster m. If this activity exceeds a threshold of 0.5, $W_b(t)$ increases toward the maximum $W_b^m(t)$, otherwise it drops toward zero:

$$W_b(t+1) = \begin{cases} \alpha_p W_b(t) + (1-\alpha_p)W_b^m(t), & \text{if } S_m(t) > 0.5; \\ \alpha_d W_b(t), & \text{if } S_m(t) < 0.5. \end{cases} \quad [3]$$

α_p and α_d are constants related to the time constants T_p and T_d for potentiation and desensitization:

$$\alpha_p = e^{-1/T_p};\ \alpha_d = e^{-1/T_d}. \quad [4]$$

During learning, the maximum efficacies $W_b^m(t)$ of modulated bundles $b = (a, i, m)$ are themselves updated according to the learning rule. Modifications of $W_b^m(t)$ occur only if the recent contribution of anterior cluster a to posterior activity is large enough, namely if

$$W_b(t-2)S_a(t-2) > 0.5\ W_b^m(t). \quad [5]$$

$W_b^m(t)$ is then modified according to the level of activity of posterior cluster i in the very same way as $W_b(t)$ is modified according to modulator activity:

$$W_b^m(t+1) = \begin{cases} \beta_1 W_b^m(t) + (1-\beta_1)W', & \text{if } S_i(t) > 0.5; \\ \beta_2 W_b^m(t), & \text{if } S_i(t) < 0.5. \end{cases} \quad [6]$$

W' is the maximum possible value of modulated efficacies, independent of time. β_1 and β_2 are constants of the learning process, in the interval $]0,1[$. β_2 is chosen to be large in comparison to β_1. This results in a conservative selective rule that causes the inactivation of connections only if they systematically perturb ongoing activity.

Various network architectures were simulated on a PDP 11/73 computer with a maximum of 50 clusters, which corresponds to several thousand neurons. The scaling of the properties with the size of the network was not investigated. Typical values for the parameters are as follows: $V_\alpha = 13$, $V_\beta = -8$, $T_p = 20$, $T_d = 15$, $W' = 13$, $\beta_1 = 0.75$, and $\beta_2 = 0.998$. The behavior of the model was shown to be independent of variations in these values within a reasonable range. It was, however, noticed that for some input conditions most modulated bundles become potentiated and block the network into a hyperactive state.

Properties of the Model

Property 1. Passive Recognition and Production of Temporal Patterns: Role of Synaptic Triads. This section analyzes the consequences of the presence of synaptic triads in a network. The postsynaptic neuron of an individual synaptic triad detects a transition of activity. Similarly, posterior clusters of modulated bundles function as passive transition detectors. However, at the cluster level, a property appears: as a consequence of excitatory self-connections, clusters can maintain a self-sustained stable activity. Thus, a trace of previous transitions will remain in the assembly until it becomes progressively erased by lateral inhibition. Transition detectors may then work on the output of other transition detectors, in a linear as well as in a hierarchical fashion. The end clusters become complex sequence detectors that respond to arbitrary long sequences which may include repetitions.

Synaptic triads clearly have the potentialities to perform production of time sequences. A modulated bundle whose anterior and modulator clusters are identical works as a delay line: it propagates activity with a time delay. A ring of such connections can produce a temporal sequence of activity, associating each state with its successor. In general, however, a sequence will include repetitions of the same items in different contexts. This precludes the direct designation of the successor from the current state (1). In general the correct successor can be determined only by the knowledge of states prior to the current one. Production networks must thus keep traces of the anterior productions. We call this fundamental property the remanence of previous states. It suggests that a production network is able to recognize and memorize its own outputs. According to the span of its memory, the network will be able to produce sequences of various complexities. Producing sequence 1–2–3–1–2–3, . . . does not require any memory; producing sequence 1–2–1–2–1–3, . . . requires a span of three previous states. The degree of a sequence will be defined as the minimal memory span required to produce it.

A ring of "delay-line" synaptic triads produces sequences of degree zero. In that case the modulator, which initiates the transition, is the same as the anterior cluster. This need not be the case because a transition from anterior to posterior can be initiated by higher-order information. If the modulator is a complex sequence-detector cluster, the following state can be chosen according to information on former productions. Remanence can thus be accounted for with synaptic triads.

Property 2. Learning Rule and Genesis of Internal Organization. Synaptic triads organized in hierarchical architectures thus perform passive sequence detection and production. Simulations show that these architectures actually develop with the learning rule out of initial disorder and do not require ad hoc wiring.

We took as initial state an architecture with two layers. The first one was composed of input clusters, the activity of which was imposed to study the consequences of the learning rule. Each input cluster represented one note in the sequence to be recognized, and only one of them could be activated at a time. The second layer consisted of internal clusters, randomly connected to other clusters via many modulated bundles (two upper layers of Fig. 1*c*).

Differentiation of Sequence Detectors. In a first simulation, only internal clusters received modulated bundles with an-

terior and modulator clusters chosen at random among input or internal clusters. This gives them the possibility to detect various sequences in the input signal. Indeed, most clusters initially respond to several sequences of the imposed melody. Conversely, each transition is often encoded by several clusters. With application of the learning rule, some of this redundancy disappears, and the activity patterns become much more reproducible. Each cluster finally codes for fewer but more specific sequences. This process of differentiation is illustrated by the activity of internal clusters in Fig. 2. The learning rule selects among the events that each cluster can detect and chooses those that match the teaching signal. At a high initial diversity, clusters detecting sequences as long as four notes were identified. Hierarchies of detectors may thus develop.

Stabilization of Ongoing Activity. As a second step in the simulation, modulated bundles were added, with the anterior and posterior clusters chosen among input clusters and with the modulator randomly chosen among input or internal clusters. This provided a means to produce sequences of high degree. Consistent with the above analysis, the network spontaneously produced quasirandom sequences. We tested whether the learning rule results in a stabilization of ongoing activity. First, a sequence was imposed while learning occurred. Later on, we checked whether the sequence had been stabilized by allowing reproduction without imposing activity. The following rules were established: "Delay-line" bundles between input clusters are modified first and encode the zero-order regularities of the sequence. Bundles that are modulated by internal clusters are modified later on, after activity of internal clusters has stabilized. These last bundles encode higher-order information such as "if the current state is 1 and preceding state was 2, then the next state should be 3." *A priori*, sequences of arbitrary degree can be stabilized provided enough diversity is initially present. Fig. 3*A* shows the production of the first-degree sequence 1–2–1–3–1–4.

Property 3. Active Recognition and Learning of Time Sequences by Selection. Selective theories of learning and

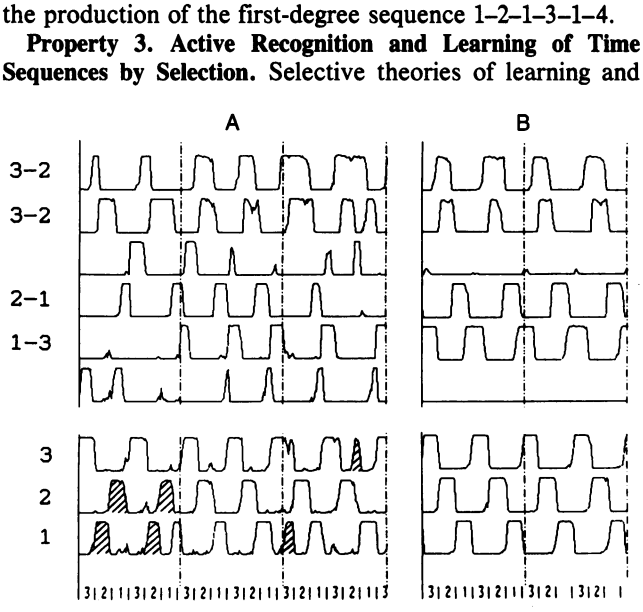

FIG. 2. Simulation of the entire active recognition network. Time flows from left to right (vertical bars every 100 update cycles). Each trace represents the evolution of the activity of one cluster with time. Bottom traces, input clusters. Top traces, internal clusters (sequence detectors). Numbers at bottom represent the sequence imposed on the perceptual layer. This simulation studies the evolution of the internal activity with learning. (*A*) Initial responses to sequence 3–2–1, ... Periods of resonance (plain) and dissonance (striped) are clearly visible in the input units. Internal activity is chaotic and coding is redundant. (*B*) Responses after learning. "Dissonant" prerepresentations have been eliminated. Internal coding is now stable and reproducible. The network completes sequences where items are missing (sequence 3–2–blank at the end).

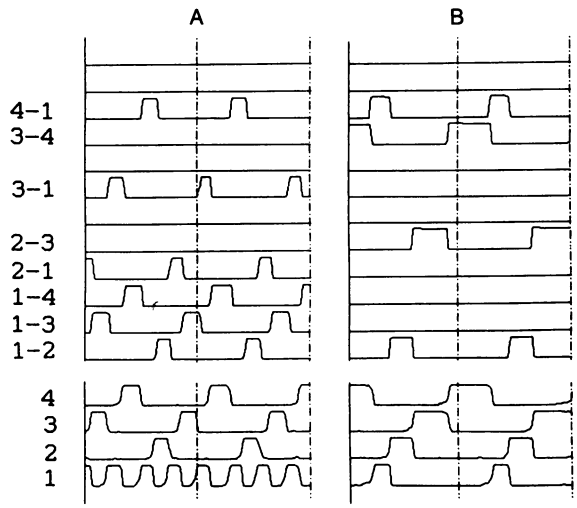

FIG. 3. Simulations of networks that produce sequences. The two upper layers of Fig. 1*c* were simulated. (*A*) Production of sequence 1–2–1–3–1–4 in the input clusters. This complex sequence of degree one was stabilized with the standard learning rule. (*B*) Control of duration. The standard learning rule drives each synaptic efficacy toward its lower or higher bound, leading to notes of constant duration. This sequence with variable durations was stabilized with the following modified learning rule (replacing Eqs. **5** and **6**) that applies if

$$W_b(t-2)S_a(t-2) > \theta; \qquad [7]$$

then the following modification:

$$W_b^m(t+1) = \begin{cases} \Phi[W_b^m(t) + \delta], & \text{if } S_p(t) > 0.5; \\ \Phi[W_b^m(t) - \delta], & \text{if } S_p(t) < 0.5. \end{cases} \qquad [8]$$

Φ is a truncation function that keeps $W_b^m(t)$ in $[0,W']$, θ is a threshold value, and δ is a small increment of the order of 0.1. Under some conditions on θ, this rule can be mathematically shown to drive the efficacy $W_b^m(t)$ to a value that creates the right delay before switching.

recognition are based upon the internal production, out of the initial structure of the network, of prerepresentations that are compared with the incoming signal or percept. This in turn implies that (*i*) an "active recognition" network should internally produce temporal sequences and that (*ii*) percepts alone should not be able to impose the state of the network, which should depend on an interaction between external and internal constraints.

In our model, we meet the second point by decomposing the active recognition network in the following two assemblies (Fig. 1*c*): a sensory layer, on which percepts are merely imposed, and an internal production network, whose structure is similar to the one described in the previous sections and comprises input and internal clusters. Both the sensory layer and the internal clusters innervate the input clusters. Dendritic summation thus combines perceptual as well as internal components. Resonance (10) will be defined as the matching of these internally and externally produced activities. Simulations confirm the potentialities of this architecture for learning by selection. In the absence of sensory inputs, starting from any initial condition, sequences are spontaneously produced. Initially these prerepresentations are quasirandom, although they partially reveal internal connectivity, but very small sensory weights (inferior to noise level) suffice to influence these productions. With a weak perceptual input, periods of resonance and dissonance are clearly visible in the activity of input clusters (Fig. 2*A*).

Learning occurs only in the periods of high and stable activity or resonance. The learnable sequences must thus

belong both to the prerepresentations and to the sensory percepts received. Consistent with the "attrition" phenomena noticed during acquisition of song by birds (13), a reduction of the repertoire of prerepresentations accompanies learning: the periods of dissonance progressively disappear, until activity optimally matches the sensory signal (Fig. 2*B*). Slight learning-induced modifications may occur in the absence of percepts, and sequences sometimes self-stabilize. Yet the initial architecture and connectivity fundamentally restrict what can be perceived and learned: an organism can not learn more than is initially present in its prerepresentations. As an illustration, a network with no delay line between clusters 1 and 2 was simulated. Spontaneous activity shows no trace of transition 1–2. When the network is stimulated with sequence 1–2–3, . . . , it is not equipped to perceive it faithfully. Input clusters indeed play sequence 1–(3)–2–3, which is the only one that can possibly be learned. Such effects might account for part of the selectivity of bird song learning to conspecific material (13).

Sequences of degree up to one were learned by selection in the network. During active recognition, perception of the beginning of a sequence causes the internal production of the remaining part, thus demonstrating capacities for anticipation and signal restoration (Fig. 2*B*). When many choices are possible (for example, when several learned sequences share the same beginning), the maximum efficacies remain high for those transitions that are ambiguous. Differences in the frequencies of presentations of the sequences result in the most frequent one being chosen as the default. Other transitions remain possible with a slightly higher threshold. During recognition, prerepresentations are proposed by the production network and are progressively selected and modified, up to a point where there is only one sequence that is still compatible with the percept. Recognition is then achieved, since the correct sequence can be unambiguously predicted. The model thus incorporates notions of frequency effect (25) and point of identification (26) analogous to those developed for lexical access in psycholinguistics.

Functioning of the network relies on a delicate balance of credit between internal and external constraints. The network can be adapted to diverse tasks by modifying this balance. Letting percepts impose the internal state leads to the neglect of already formed categories and to the processing of "instructions" novel to the network but compatible with its structure; reducing the weight of percepts leads the network to "autism," or simply to false alarms in recognition. Attention may play a role in regulating this balance.

Conclusions

The proposed theroretical network, which recognizes, stores, and produces time sequences, differs from other models in which time is just considered as another spatial dimension (27) or that rely either on an ad hoc architecture (4, 5, 7, 28) or on very finely tuned synaptic efficacies (3). It comes closer to models that rely on specific properties of the neuronal dynamics, for example, adaptation (6, 29), short-term changes of synaptic efficacy (8, 9, 30), interactions between three neurons (31), or other mechanisms for sequence detection (32).

From an experimental point of view, the model points to the crucial role of synaptic triads in sequence detection and production. It predicts the occurrence of hierarchies of sequence detectors, the remanence of previous productions, and the specialization of sequence-coding neurons in the course of learning. It also suggests that allosteric transitions of postsynaptic receptors might be involved in sequence detection and elementary learning. Finally, the notions that storage and recognition of temporal sequences take place by selection among internally produced prerepresentations and, more tentatively, that the degree of a sequence correlates with its perceived complexity, could be tested in birds and humans by behavioral and/or psychophysical methods.

The authors thank Gérard Toulouse for insightful comments at various stages of this research, and Elie Bienenstock, Jacques Mehler, and Pierre Peretto for reviewing the manuscript. This work was partially supported by Grant 84C1390 from the Ministère de la Recherche et des Universitès and ATP 5207 from the Centre National de la Recherche Scientifique to Jacques Mehler, Laboratoire de Sciences Cognitives et Psycholinguistique.

1. Lashley, K. S. (1951) in *Cerebral Mechanisms in Behavior*, ed. Jeffress, L. A. (Wiley, New York), pp. 112–136.
2. Ackley, D. H., Hinton, G. E. & Sejnowski, T. J. (1985) *Cognit. Sci.* **9,** 147–169.
3. Hopfield, J. J. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 2554–2558.
4. Amari, S. I. (1972) *IEEE Trans. Comput.* **C-21,** 1197–1206.
5. Fukushima, K. (1973) *Kybernetik* **12,** 58–63.
6. Willwacher, G. (1982) *Biol. Cybern.* **43,** 115–126.
7. Kohonen, T. (1984) *Self-Organization and Associative Memory* (Springer, Berlin).
8. Kawamoto, A. H. & Anderson, J. A. (1985) *Acta Psychol.* **59,** 35–65.
9. Peretto, P. & Niez, J. J. (1986) in *Disordered Systems and Biological Organization*, eds. Bienenstock, E., Fogelman, F. & Weisbuch, G. (Springer, Berlin), pp. 171–185.
10. Changeux, J.-P., Heidmann, T. & Patte, P. (1984) in *The Biology of Learning*, eds. Marler, P. & Terrace, H. (Springer, Berlin), pp. 115–133.
11. Toulouse, G., Dehaene, S. & Changeux, J.-P. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1695–1698.
12. Edelman, G. & Finkel, L. (1984) in *Dynamic Aspects of Neocortical Function*, eds. Edelman, G., Gall, W. & Cowan, W. M. (Wiley, New York), pp. 653–695.
13. Marler, P. & Peters, S. (1982) *Dev. Psychobiol.* **15,** 369–378.
14. Konishi, M. (1985) *Annu. Rev. Neurosci.* **8,** 125–170.
15. Margoliash, D. (1983) *J. Neurosci.* **3,** 1039–1057.
16. Nakayama, K. (1985) *Vision Res.* **25,** 625–660.
17. Suga, N. (1984) in *Dynamic Aspects of Neocortical Function*, eds. Edelman, G., Gall, W. E. & Cowan, W. M. (Wiley, New York), pp. 315–373.
18. Changeux, J.-P. (1981) *Harvey Lect.* **75,** 85–254.
19. Changeux, J.-P. & Heidmann, T. (1987) in *New Insights into Synaptic Function*, eds. Edelman, G., Gall, W. E. & Cowan, W. M. (Wiley, New York), in press.
20. Heidmann, T. & Changeux, J.-P. (1982) *C.R. Acad. Sci. Ser. 2* **295,** 665–670.
21. Ito, M., Sakurai, M. & Tongroach, P. (1982) *J. Physiol. (London)* **324,** 113–134.
22. Hawkins, R. & Kandel, E. (1984) *Psychol. Rev.* **91,** 375–391.
23. Ratliff, F. (1965) *Mach Bands: Quantitative Studies of Neural Networks in the Retina* (Holden-Day, New York).
24. Hebb, D. O. (1949) *The Organization of Behavior* (Wiley, New York).
25. Morton, J. (1969) *Psychol. Rev.* **76,** 165–178.
26. Marslen-Wilson, W. D. & Tyler, L. K. (1980) *Cognition* **8,** 1–86.
27. McClelland, J. L. & Elman, J. L. (1986) *Cognit. Psychol.* **18,** 1–86.
28. Abeles, M. (1982) *Local Cortical Circuits* (Springer, New York).
29. Stent, G. S., Kristan, W. B., Friesen, W. O., Ort, C. O., Poon, M. & Calabrese, R. L. (1978) *Science* **200,** 1348–1357.
30. von der Malsburg, C. & Bienenstock, E. (1986) in *Disordered Systems and Biological Organization*, eds. Bienenstock, E., Fogelman, F. & Weisbuch, G. (Springer, Berlin), pp. 247–272.
31. Ballard, D. H. (1986) *Behav. Brain Sci.* **9,** 67–120.
32. Koch, C., Poggio, T. & Torre, V. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2799–2802.