Neural computation by concentrating information in time

(neural network/connectionist/speech recognition/parallel processing)

D. W. TANK[†] AND J. J. HOPFIELD[†‡]

[†]Molecular Biophysics Research Department, AT&T Bell Laboratories, Murray Hill, NJ 07974; and [‡]Divisions of Chemistry and Biology, California Institute of Technology, Pasadena, CA 91125

ABSTRACT An analog model neural network that can solve a general problem of recognizing patterns in a time-dependent signal is presented. The networks use a patterned set of delays to collectively focus stimulus sequence information to a neural state at a future time. The computational capabilities of the circuit are demonstrated on tasks somewhat similar to those necessary for the recognition of words in a continuous stream of speech. The network architecture can be understood from consideration of an energy function that is being minimized as the circuit computes. Neurobiological mechanisms are known for the generation of appropriate delays.

Recognizing the pattern in a time-dependent signal is important in hearing and vision. The tasks of recognizing individual spoken phonemes or words in a continuous stream of speech can be effortlessly and rapidly done by any speaker of a language. If the motion of a walking human is replaced by the pattern of motions of a few dots representing parts of the body we immediately identify the dynamic pattern as "person walking" in spite of our complete inability to recognize "person" from any single frame of such a movie. Although such tasks are simple for us, they are very difficult for present digital computers. Appropriate modeling of neural networks may suggest how nervous systems solve such problems as well as provide a description of artificial neural networks for the task.

Using networks to recognize regularity in a temporal sequence of input stimuli has been previously studied. Reis (1) used time delays to make bandpass filters from simple networks of three or four formal neurons. Kohonen (2) studied neural networks that operate as finite-state machines generating a sequence of states. An input pattern produces an output pattern that is used as a partial input pattern at the next system iteration. A temporal associative network has been studied by Fukushima (3). Related networks have been recently studied by Kleinfeld (4) and Sompolinsky and Kanter (5) as sequence generators, which might be used to recognize simple sequences by supplying inputs to help drive the network from one state to another. Sequence recognition by completion has also been studied in a neural network model by Rumelhart *et al.* (6).

Despite this work, neural network architectures that employ dynamics to continuously recognize sequences distorted in time and form have remained elusive. This problem is addressed in the present work. It is illustrated by two sequences, in which each letter *A*, *B*, *C* represents a particular momentary stimulus state, with the time duration of the stimulus states represented by letter repetitions. Let the first row represent a model known stimulus sequence.

$$S\ S\ E\ E\ E\ E\ R\ I\ I\ E\ Z\ Z\ Z$$

$$T\ H\ H\ S\ S\ \underline{E\ A\ R\ R\ E\ E\ Z}\ S\ S\ O\ O\ U\ N\ N\ D\ Z$$

The second row embeds a distortion of this sequence (underlined) in a continuous data stream. The task is to recognize

the presence of the model known sequence in this stream. (Real problems will generally have continuous time and continuous patterns to deal with, but the general idea is more easily seen in discrete examples.) Difficulties include (*i*) where to start the comparison (word-break problem) and (*ii*) given a starting point, different examples of the sequence may be distorted in time (time-warp problem) and in form (erroneous symbols). Here and in the following, our illustrations of the generic problem use time-dependent symbolic stimulus sequences suggestive of speech, but highly abstracted from real speech.

We would like to recognize sequences in such a data stream *as a whole*. Such recognition implicitly involves comparing known exemplar sequences, distorted in time and form, with the data stream and finding a "best fit." This can be done with serial algorithms (see, for example, refs. 7 and 8) but is computationally intensive. The isolated sequence problem could be rapidly solved if the only problem were time warp, for then the order of stimuli alone is a reliable description of the sequence. But when the symbols are not accurate, elementary sequential search is not useful. Series-to-parallel conversion followed by the use of an associative memory is not a solution since it merely converts the time-warp problem into an equivalent space-warp problem. Using dynamic circuits that sequentially generate the exemplar sequences leads to an inadequate sequential analysis rather than a recognition of the whole.

Concentrating Information in Time

A network that solves sequence recognition problems must temporarily store information about the data stream to make a recognition of the whole. Suppose a holistic decision is made at time *t* that sequence *S* has just occurred, and the data stream of *S* lasts a time τ. Information in the data stream between time $t - \tau$ and time *t* must have been implicitly stored for availability at *t*. If a subset of the units in a network is to be strongly driven, and to turn "on" at time *t* to indicate that *S* occurred, then this information must be concentrated on those units in space and in time. (A similar concentration of information in time is used in chirp-frequency radar ranging.)

The neural network sequence detector we describe here has a set of "delay" filters, one for each known sequence. Each filter has a strong output during a relatively short time at the end of its sequence. These outputs drive a nonlinear neural decision network, whose outputs, changing in time, describe the position of the various sequences in the data stream. The following section of the paper develops the concept of an energy function as a design tool for such a network; the section after presents the network design and illustrates its use.

Time-Dependent Energy Surfaces

An energy function (*E*) is useful in understanding how circuits of neuron-like elements compute solutions to optimization problems (9). The circuits follow trajectories in their state space that minimize *E*. In many cases, minimization of *E* can be related to the minimization of a cost function for an

optimization problem (10, 11). This provides a procedure for designing and understanding a circuit that solves the problem.

The original E function formulation is not applicable for time-dependent inputs. However, the usefulness of an E can be extended to the time-dependent domain. In the more general case in which E has an explicit time dependence

$$E = E(\vec{V}, t) \text{ and } \frac{dE}{dt} = \frac{\partial E}{\partial t} + \nabla_V E \cdot \frac{d\vec{V}}{dt}. \quad [1]$$

The output V_i of neuron i is related to its input u_i by the nonlinear input–output relationship $V_i = g(u_i)$. Choosing a circuit dynamics described by:

$$\frac{d\vec{u}}{dt} \propto - \nabla_V E \quad [2]$$

implies

$$\frac{dE}{dt} = \frac{\partial E}{\partial t} - (g^{-1})' \left[\frac{d\vec{V}}{dt} \right]^2, \quad [3]$$

so that the contribution to dE/dt produced by the *motion* of the circuit (the second term) is still always negative, as in the earlier described case (9, 10), if the inverse input–output function ($u_i = g^{-1}(V_i)$) is monotone increasing. Eq. 2 implies that the "force" on the circuit is always in the generally "downhill" direction. (Eq. 2 is formally equivalent to mechanics with viscous drag, where force is proportional to velocity.) A geometrical picture of computation with time-dependent inputs can be obtained using a "space-time" E surface like the one illustrated in Fig. 1. E is plotted for all points in a combined space of the circuit's state variables and time. In Fig. 1 only one state space dimension is drawn. There are valleys along the state space axis at each time point. The spatial positions of these valleys are determined by the history of data in the problem and change with time in a continuous fashion. Due to the relationship between dynamics and $\nabla_V E$ given in Eq. 2, the circuit will be pushed at each time point toward a valley in its state space. A computation can be performed by a time-dependent E function if the data of the problem produce a channel on the space-time surface that guides the circuit trajectory to a position appropriate to the correct solution.

This conceptual framework can be used to design neural circuits based on the intuitive idea described earlier of using a set of delay filters as a sequence detector. We desire an E for which the presentation of a known sequence of stimuli builds a deep pit on the space-time E surface, with a wide valley leading to it. The pit is centered on a time point t^* near the expected completion of the presented sequence. The state space location at the pit is to be that that codes for
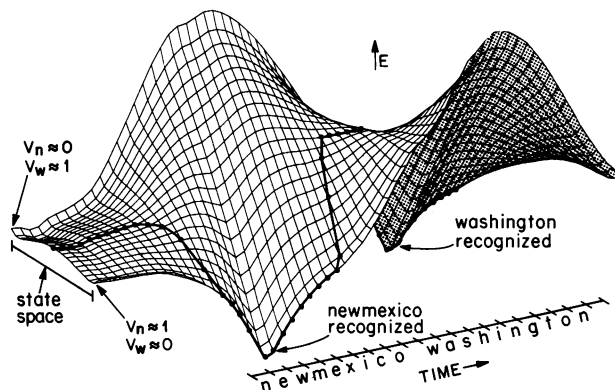
the recognition of the sequence. The appropriate E will provide the prescription for the physical circuit that solves the task.

Let exemplar sequence $S^i(t)$ be described by the stimuli sequence $[S^i(1) = A, S^i(2) = B, \ldots, S^i(n) = E]$, where A, B, $\ldots$, E represent distinct stimulus states in the data stream and where, in general, $S^i(m) = X$ implies that generic stimulus X is present in time bin m in the exemplar. We define the functions $D_X(t)$ as:

$$D_X(t) = 1 \text{ if stimulus } X \text{ is present at } t;$$
$$= 0 \text{ otherwise.} \quad [4]$$

$D_X(t)$ formally represent the outputs of stimulus detectors tuned to different stimuli X from the set $\{A, B, \ldots, E, \ldots\}$. Assume that we want neuron i, with output V_i between 0 and 1, to be activated ($V_i = 1$) when a sequence of stimuli rather like S^i is detected.

When a stimulus X in S^i is detected, an appropriate term in E can contribute to a channel on the space-time E surface leading to the state $V_i = 1$. The necessary term in E, written in a form that combines similar evidence for S^i from other detected stimuli $[D_X(t)]$ is:

$$\delta E = -V_i \sum_X \sum_k \int_0^\infty T_{iX;k} f_k(\tau) D_X(t - \tau) d\tau, \quad [5]$$

with $T_{iX;k} = 1$ if stimulus X occurs in S^i in a time bin k units prior to sequence completion and is 0 otherwise. The $f_k(\tau)$ are *continuous* delay functions sketched in Fig. 2a with the index k proportional to the mean time delay. (See Eq. 11 for a specific example.) The time-independent coefficient $T_{iX;k}$
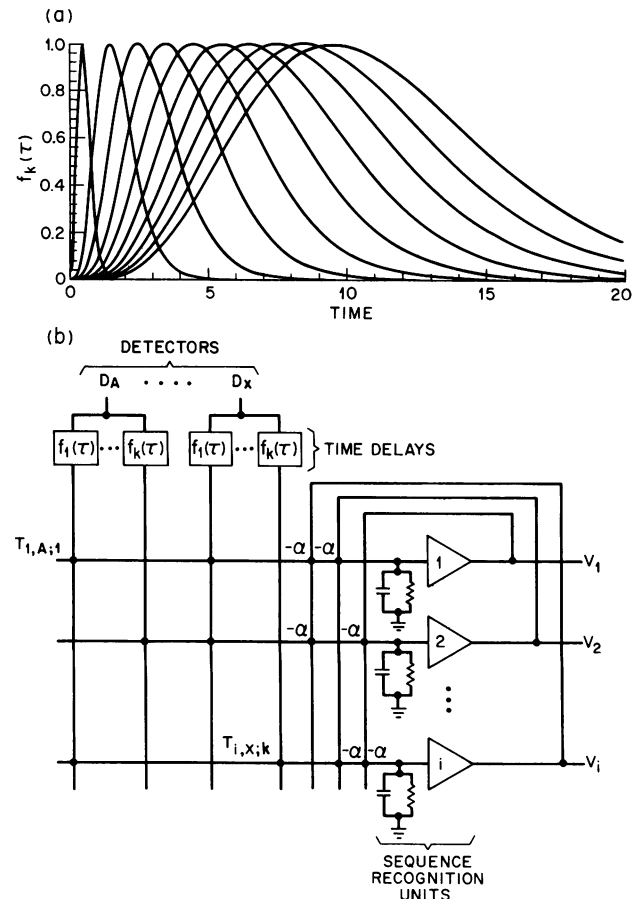

FIG. 2. (a) Set of delay functions used in the simulations described in the text for $k = 1, 2, \ldots, 10$. (b) Sequence recognition circuit described by Eq. 11 in the text.

FIG. 1. Representation of an E surface for space time, illustrating the computation performed in recognizing two sequences.

represents the "synaptic" strength of the coupling, with mean delay k, between the output of D_X and the input of neuron i. The "force" on neuron variable V_i due to the term in Eq. 5 is

$$-\frac{\partial E}{\partial V_i} = \sum_X \sum_k \int_0^\infty T_{iX;k} f_k(\tau) D_X(t - \tau) d\tau, \qquad [6]$$

so that a presentation of stimulus X results in a postsynaptic effect on neuron i given by the convolution of the X detector output and the delay function $f_k(\tau)$, scaled by $T_{iX;k}$. When $T_{iX;k}$ is positive, the corresponding term in δE in Eq. 5 contributes to a pit on the space-time E surface spatially centered in the subspace $V_i = 1$ and elongated in time due to the temporal spread in $f_k(\tau)$. Since the breadth of the $f_k(\tau)$ is longer for larger k, corresponding to longer mean delay times, the early stimuli in the sequence produce longer contributions on the E surface. The scaling of latency with position in the exemplar sequence insures that each stimulus can contribute to the same pit. Because the delays have a continuous "probability" amplitude, a valley of considerable depth occurs at sequence completion even if the presented sequence is a time distortion of the exemplar. (The form of the delay functions can be related to a Bayesian analysis of expected exemplar distortions—see below.) The pit depth is a measure of the evidence that sequence S^i was presented to the network. Evidence from detected stimuli against a given exemplar can be similarly treated by inhibitory connections $(T_{iX;k} < 0)$ that build mounds on the E landscape. In this way, erroneous symbols in the data stream of a distorted exemplar only add noise to the energy landscape, gracefully degrading the network performance. The contributions to E in Eq. 5 above describe the synaptic connections with a delayed postsynaptic effect from a set of stimulus detectors $[D_X(t)]$ to a set of "grandmother" neurons whose state reflects the evidence that the stimulus train to which it is tuned has driven the system of detectors. A mutual inhibition contribution to $E(V, t)$ of the form

$$\delta E = +\frac{\alpha}{2} \sum_i \sum_{j \neq i} V_i V_j \qquad [7]$$

insures that only one sequence at a time is detected with certainty $(V_i = 1)$. To prevent nonexemplar sequences from activating the network, appropriate global inhibition (Eq. 8) will set a threshold for activation:

$$\delta E = +\gamma \sum_i V_i. \qquad [8]$$

The total E function for the sequence detector network is

$$E = -\sum_i \sum_X \sum_k V_i \int_0^\infty T_{iX;k} f_k(\tau) D_X(t - \tau) d\tau$$

$$+ \frac{\alpha}{2} \sum_i \sum_{j \neq i} V_i V_j + \sum_i \gamma V_i + \sum_i \frac{1}{R} \int_0^{V_i} g_i^{-1}(V) dV. \qquad [9]$$

The significance of the gain term (last term in Eq. 9) has been discussed (10). E is the cost function for a time-dependent optimization problem of choosing the best match of the known exemplar sequence S^i to the presented signal, which is embodied in the pattern in time provided by the detector outputs $D_X(t)$. The network will report out a decision only when the evidence is adequately large. At that time, E is minimized by the state $V_i \approx 1$ for the best-match sequence S^i, and all other $V_j \approx 0$. At other times, all V_i will be small.

Applying the Circuit to Noisy and Time-Warped Sequences

By Eq. 2, the E function describes the physical circuit in Fig. 2b. The input potentials u_i of the recognition neurons are determined by the continuous dynamics:

$$C \frac{du_i}{dt} = -\frac{u_i}{R} - \sum_{j \neq i} \alpha V_j - \gamma$$

$$+ \sum_X \sum_k \int_0^\infty T_{iX;k} f_k(\tau) D_X(t - \tau) d\tau. \qquad [10]$$

C and R are the input capacitance and resistance of the recognition neurons. The delay functions $f_k(\tau)$ and the $T_{iX;k}$ are chosen to detect particular exemplar sequences S^i. The first term in E (Eq. 9) results merely in a time-dependent input $[I_i(t)]$ in the circuit equation [10].

We studied a specific circuit designed to recognize the sequences of letters in the names of the 50 states of the United States. For simplicity, the functional form of the delay functions is chosen as:

$$f_k(t) = \beta \left(\frac{t}{k}\right)^n e^{-n\left(\frac{t}{k}\right)}. \qquad [11]$$

$\beta = e^n$ normalizes the amplitude of the peak response. (In these experiments, $n = 5-10$.) As before, we let the letters A, B, C represent momentary stimulus states X, with one time unit per letter. The exemplar sequences determine the connection strengths $(T_{iX;k})$ by Eq. 5. For example, consider the exemplar sequence $(ARIZONA)$ and assign unit i with output V_i as the word recognizer unit for this signal. Since A is the first entry in the sequence, $T_{iA;6} = 1$ provides a delayed input to i centered on 6 time units. Since A is also in position 7, $T_{iA;0} = 1$ provides additional input to unit i with no $(k = 0)$ delay. This procedure also generates the set of connections $T_{iA;6} = T_{iA;0} = T_{iR;5} = T_{iI;4} = T_{iZ;3} = T_{iO;2} = T_{iN;1} = 1$. The construction is repeated for each "state" sequence and its corresponding recognition unit. All other connections were set to a small negative value providing weak inhibition for exemplar detection by the presence of erroneous symbols, not part of the exemplar, in the data stream.

Computer simulations of the network were performed by numerically integrating Eq. 10 for different input sequences. (See Fig. 3 legend for specific parameter values.) In Fig. 3a, the outputs V_i are plotted versus time for presentation of the stimulus pattern $(NEWMEXICO \ WASHINGTON)$. The outputs of all 50 word units V_i are superimposed, with those of the most active units individually labeled. All units remain
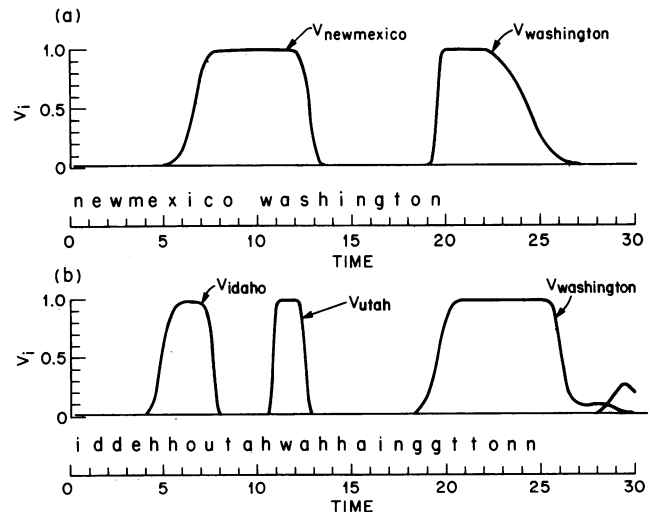

FIG. 3   (*a* and *b*) Outputs of the sequence recognition neurons for indicated input sequences. The outputs of all 50 neurons are superimposed. For units that are activated by the sequence, the exemplar sequences to which they are tuned are indicated. The circuit parameters were: $C = 1.0$; $\tau = RC = 0.5$; $\alpha = 3.0$; $\gamma = 2.5$; $V_i = g(u_i) = 1/2[1 + \tanh(u_i/0.5)]$. For each neuron i, the excitatory inputs $T_{i,X;k}$ described in the text were scaled by $1/l_i$, where l_i = length of sequence S^i. Inhibitory connections were scaled by $-0.5/l_i$.

near $V_i = 0$ early in the sequence presentation. As the presentation of the sequence *NEWMEXICO* proceeds, the *NEWMEXICO* unit that correctly corresponds to the detection of this sequence is activated. Shortly after termination of *NEWMEXICO*, the activity in the *NEWMEXICO* unit drops to baseline. Near the completion of the *WASHINGTON* sequence, the *WASHINGTON* recognition unit similarly becomes briefly activated. This perfect recognition in the absence of distortions is observed for all 50 state exemplar sequences.

Fig. 3b illustrates that the circuit is insensitive to time warp, erroneous symbols, and lack of word breaks. The actual sequence presentation is based upon the string of exemplar sequences [*(IDAHO)(UTAH)(WASHINGTON)*]. However, the expected sequence of letters has been distorted in time and form by repetition, deletion, and insertion (e.g., the exemplar *IDAHO* is distorted to *IDDEHO*) and there are no spaces or breaks between sequences. Despite time warp, erroneous symbols, and the absence of spaces between the presentation of the individual sequences, the correct word recognition units are activated.

Why Does the Circuit Work?

Fig. 1 shows a projection of the time-dependent E surface explicitly calculated for the network already described when the sequence *NEWMEXICO WASHINGTON* is presented. There should be 51 dimensions to this surface (50 space dimensions, one for each recognition variable, and 1 time dimension), but we can plot only a two-dimensional surface. During the stimuli presentation, V_j other than $V_{NEWMEXICO}$ (V_N) or $V_{WASHINGTON}$ (V_W) have negligible values. The three axes drawn in Fig. 1 are E, time, and $(V_N - V_W)$ for [(V_N + V_W) = 1]. The latter axis corresponds to a projection along the line between $V_N = 1$, $V_W = 0$ and $V_N = 0$, $V_W = 1$, the two points representing the detection of *NEWMEXICO* and *WASHINGTON*, respectfully. At early times, there are weak valleys at the two ends of this coordinate due to the lateral inhibition, but there is a true deep minimum in E in the four-dimensional space time (V_N, V_W, t, E) at $V_N = V_W$ = 0 (not shown on this projection). As time goes on, a deep minimum begins to form near $V_N - V_W = 1$ due to the fact that the initial incoming data look like *NEWMEXICO*, and a maximum forms at the other end of this coordinate because the early data do not correlate with *WASHINGTON*. These trends continue until a coherent minimum is obtained at the completion of the first known string. After that, the signal from the early data decays, and the additional data begin to correlate with *WASHINGTON*. This lowers the energy at the *WASHINGTON* end of the spatial coordinate, near the time for the expected completion of *WASHINGTON*. The trajectory drawn on the energy surface is the projection of the motion of the state of the system [$V_N(t) - V_W(t)$] on the energy surface and shows sequential recognition of the two sequences. Considerable reductions in depth due to time distortions can be tolerated before the energy landscape is so modified that the wrong decision is made or a word is not recognized.

Bayesian Interpretation of the Connections

A systematic choice of time-delay functions is possible if information about the nature of the likely time distortions is available. A typical known exemplar sequence S^i is sketched in Fig. 4. Typical time-warp distortions σ' and σ'' differ from S^i in that individual stimuli X in the sequence differ in duration. Let $P(\sigma^{sig}|S^i)$ be the conditional probability that σ^{sig} is generated when a sequence S^i was intended, and $P(S^i)$ be the probability of occurrence of S^i. When σ^{sig} is received, the maximum likelihood choice of which sequence S^i was
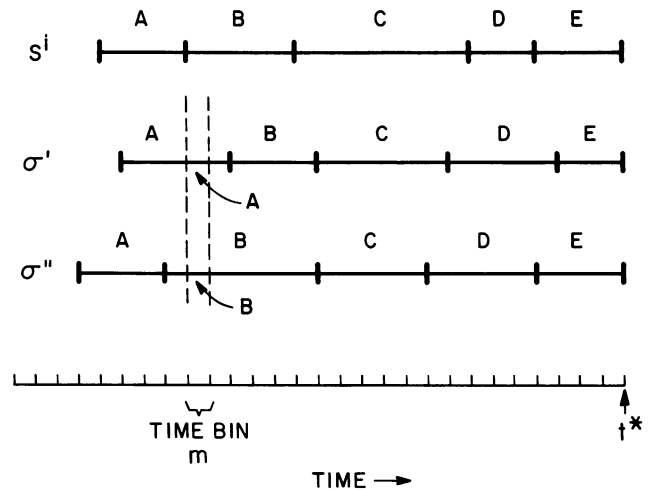

FIG. 4. An exemplar sequence and two distortions that indicate the method of computing appropriate time delay functions.

sent is that for which $P(\sigma^{sig}|S^i)*P(S^i)$ is largest. Equivalently, one can use the maximum of $\ln[P(\sigma^{sig}|S^i)] + \ln[P(S^i)]$. {The $\ln P[(S^i)]$ term is an amplitude factor that does not alter the pattern of connections.} Let a sequence σ^{sig}, consisting of the stimuli sequence (A at bin 1, B at bin 2, . . ., X at bin m) in a set of time bins 1, 2, 3, . . . m, be received in the data stream. We can approximate the conditional probability distribution for distorting S^i to σ^{sig} as the product of m conditional probability distributions of finding the various symbols in the time bins, as

$$P(\sigma^{sig}|S^i) = P(A \text{ at bin 1}, B \text{ at bin 2}, . . ., X \text{ at bin } m|S^i)$$

$$\approx P_1(A|S^i)*P_2(B|S^i)*. . .*P_m(X|S^i), \qquad [12]$$

where $P_m(X|S^i)$ is the probability of finding stimulus X in time bin m for the ensemble of distortions expected for S^i. At time t, let the input to each sequence recognition unit i, due to the detection of stimulus X at an earlier time bin m, be proportional to $\ln[P_m(X|S^i)]$. Then the total inputs to the recognition units will represent the probability that the sequence of stimuli immediately preceding t is a distortion of its exemplar sequence S^i. Unlikely distortions [$P_m(X|S^i) \approx$ 0] contribute strong inhibition, whereas common distortions are much less inhibitory (i.e., relatively excitatory).

The width of the delay functions can be given meaning in this representation. Suppose S^i contained stimulus A in time bin m and different stimuli in nearby bins before and after m. Then $\ln[P_{m-n}(A|S^i)]$, viewed as a function of n, describes the appropriate shape of the delay function for time bin m with respect to the termination of sequence i. For this model of distortion, it is clear the inputs to the recognition units should be summed over the sequence element detectors, time-delayed excitatory and inhibitory connections are both expected in the kind of pattern used, and the width and height of the time delay functions are directly related to statistical models of the distortion functions. The same line of discussion can be extended to include symbol errors and more sophisticated networks derived for more accurate approximations than that given in Eq. 12.

Discussion

The circuits described use an organized pattern of time delays to build an energy surface space–time structure that channels the circuit trajectory to correct solutions at the times when adequate information has been assembled. The concept of the time-dependent energy and its geometrical interpretation should lead to useful ways to program other computations. Although the behavior described is easiest to

understand in an adiabatic view of the motion on the time-dependent energy surface [the input information changes slowly compared with the basic time scale of the network response (see also ref. 5)], the best computational behavior need not be adiabatic. Circuits based on the idea of concentrating information in time but with self-input produced by feedback over several time scales could be used to build robust sequence generators.

The circuit form we have described can be directly tested on real speech recognition problems. Refinements will be needed, such as a hierarchical set of networks detecting exemplar sequences on several different time scales (e.g., a subphoneme recognizer whose output would drive a phoneme recognizer, and so on), or additional time-delay connections to represent the fact that the probability distributions are not truly factorable. For some networks of the form described here, simple learning rules establish the correct pattern of connections.

The kind of processing we have described requires a particular form of temporary information storage that makes available now (through a time-delay mechanism) information that was presented at various earlier times. Our ability to understand speech must be based on temporary storage over time scales at least as long as syllables and short words. Similar comparisons may occur on different time scales in other computations and other modalities.

Mechanisms are known within neurobiology for generating time delays of relevant durations. In avian binaural audition, differences in axon path length between contralateral and ipsilateral fibers may generate time delays on the 10- to 100-μsec time scale that are used for sound localization (12). Action potential propagation delays are also likely to be of relevance in cortical areas, such as the prepyriform cortex in the rabbit (13, 14), where afferent and associational fibers terminate in widely separated fields and differential delays can be of order 10 msec. Synaptic delays in functional "chains" of neurons could also contribute to propagation delays. On longer time scales, slow synaptic transmission (15), ion channel desensitization, and other electrochemical and biochemical changes occur in nerve cells that could be effectively used as time delays. Responses delayed up to 0.2 sec have been observed following postinhibitory rebound (16). Latencies as long as 0.5 sec following brief auditory stimulation have been measured for neurons in AI cortex of behaving monkeys (L. Kitzes, unpublished data). Our results suggest an important computational function for these time-de-lay generating systems in processing signals that vary in time and suggest a thorough study of long-delay generators in auditory and early postauditory areas.

The model described was based on "grandmother cell" sequence recognizers to simplify the understanding of concentrating information in time and time-dependent energy surfaces. More distributed representations of the individual sequences could have been used, and such representations may be relevant to biology and circuit design.

1. Reis, R. F. (1964) in *Neural Theory and Modeling*, ed. Reis, R. F. (Stanford Univ. Press, Stanford, CA), pp. 105–137.
2. Kohonen, T. (1984) *Self-Organization and Associative Memory* (Springer, Berlin), pp. 16–20.
3. Fukushima, K. (1973) *Kybernetic* **12**, 58–63.
4. Kleinfeld, D. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9469–9473.
5. Sompolinsky, H. & Kanter, I. (1986) *Phys. Rev. Lett.* **57**, 2861–2864.
6. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986) in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, eds. Rumelhart, D. E. & McClelland, J. L. (Bradford Books/MIT Press, Cambridge, MA), Vol. 1, pp. 358–361.
7. Levinson, S. E., Rabiner, L. R. & Sondhi, M. M. (1983) *Bell Syst. Tech. J.* **62**, 1035–1074.
8. Bahl, L. R., Jelinek, F. & Mercer, R. L. (1983) *IEEE Trans. Patt. Anal. Mach. Int.* **5**, 179–190.
9. Hopfield, J. J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3088–3092.
10. Hopfield, J. J. & Tank, D. W. (1985) *Biol. Cybern.* **52**, 141–152.
11. Tank, D. W. & Hopfield, J. J. (1986) *IEEE Trans. Circuits Syst.* CAS-33, 533–541.
12. Sullivan, W. E. & Konishi, M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8400–8404.
13. Biedenbach, M. A. & Stevens, C. F. (1969) *J. Neurophysiol.* **32**, 204–214.
14. Haberly, L. B. & Shepherd, G. M. (1973) *J. Neurophysiol.* **36**, 789–802.
15. Kehoe, J. S. & Marty, A. (1980) *Annu. Rev. Biophys. Bioeng.* **9**, 437–465.
16. Llinas, R. & Yarom, Y. (1981) *J. Physiol. (London)* **315**, 569–584.