

---

## Neural Architecture and Biophysics for Sequence Recognition

---

J. J. Hopfield and D. W. Tank

---

### I. Introduction

---

There are two conceptually separable ideas having to do with time sequences of events, positions, or actions. One concerns *generating* such sequences, as in the singing of bird song, or the walking of a millipede, or playing the piano rapidly. The other has to do with *recognizing* sequences generated in the external world or by another organism. Such recognition problems occur in vision, where for example one can recognize a friend at a great distance by her/his walk; in hearing, where a sequence of complex sounds is recognized as a spoken word; and in other sensory domains.

Fixed sequence generation is by far the easier problem. A given sequence can be described by an ordered set of states, A followed by B followed by C. . . . If one understands how to generate a doublet sequence, A followed by B, a long and complex sequence can be generated by iterating the doublet mechanism. Sequence recognition is much harder, since the idea is to be able to recognize sequences of great variability as being "the same" sequence. It is not then adequate to represent a sequence as slavishly requiring first one precise event and then another, when no individual element or step can be viewed as reliable. Instead, the sequence must be evaluated overall for its similarity to previously experienced sequences. The difference in difficulty can be illustrated in speech, where it has been possible to generate artificial continuous speech since the invention of the Edison phonograph, while recognizing spoken words in continuous speech is at present still a difficult technological problem.

To some extent, the ability to generate sequences can be used as a means of recognizing sequences (Kleinfeld, 1986), qualitatively like the way that an oscillator circuit can be synchronized by an external signal that is close to the natural frequency of the oscillator. The motor theory of speech perception (Lieberman *et al.*, 1967) explicitly invokes this class

of recognition mechanism. It is theoretically difficult, however, to generate holistic recognition in this manner. Furthermore, it is clear that neurobiology manages to recognize sequences that it cannot generate in a motor sense, as when a dog has a vocabulary of 50 English words. The mechanism we describe does not require any ability to generate the pattern being recognized, either as a motor activity or as an equivalent internal stimulation pattern.

---

## II. The Recognition Problem

---

The problem of syllable or simple word recognition is a general problem of audition, and does not intrinsically involve linguistic ability. The most familiar words can be picked out without preattention in background sounds, as when you suddenly recognize your name spoken in an adjacent conversation in a crowded party. Because this task is preattentive for many words in parallel, it is presumably a low-level task (in analogy to low-level or "early" vision) done by a network of cooperating neurons as a complex form of feature detection. This feature detection necessarily involves information and comparisons over an interval of time. When a very familiar simple sound or word is recognized in this fashion, many neurons are involved in processing the incoming information. The electrophysiological correlate of *recognition* is presumably a strong activity of a few neurons, which (separately as "grandmother cells" or together as an ensemble) represents that recognition. That activity should take place for a short time when adequate information to make a reliable identification has been received, generally at or near the completion of the acoustic stimulus from that word. Our problem is to understand the twofold focusing of the stimulus signal, in time (signals due to different parts of the stimulus arriving at different times must all arrive together at the appropriate recognition instant) and in space (only appropriate cells must be stimulated).

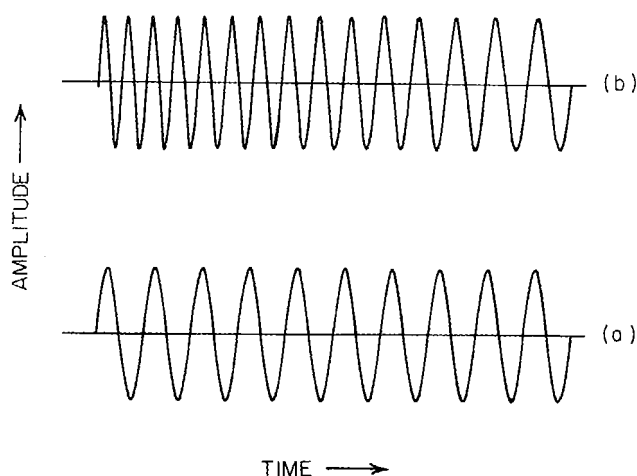
An example of how a simple time-sequence can be decoded into a meaningful signal can be seen in FM (frequency modulated) sonar in some varieties of bats (Suga, 1984). The ideal sonar signal for range determination is a single strong, very short pulse. A small object will generate a weak echo, but the determination of a precise delay time (accurate sonar range) can be done effectively since the return pulse has a high peak power and a time duration short compared to the delay time. Unfortunately, generating short pulses of very high peak power is a difficult requirement on the "transmitter." In bats (and in human-made sonar and radar) a finite duration burst of a carrier frequency is generated. Although a simple pulse of this form (Fig. 1a) solves the transmitter power problems, considerable signal processing is necessary to use the entire signal energy and duration to determine the time of arrival of the front of the weak echo frequency burst.

ally difficult, however, to gener-  
er. Furthermore, it is clear that  
ences that it cannot generate in  
ocabulary of 50 English words.  
quire any ability to generate the  
ctor activity or as an equivalent

## on Problem

recognition is a general problem  
olve linguistic ability. The most  
out preattention in background  
ze your name spoken in an ad-  
Because this task is preattentive  
ably a low-level task (in analogy  
network of cooperating neurons  
his feature detection necessarily  
over an interval of time. When  
recognized in this fashion, many  
incoming information. The elec-  
is presumably a strong activity  
"grandmother cells" or together  
ition. That activity should take  
information to make a reliable  
lly at or near the completion of  
ur problem is to understand the  
in time (signals due to different  
nt times must all arrive together  
and in space (only appropriate

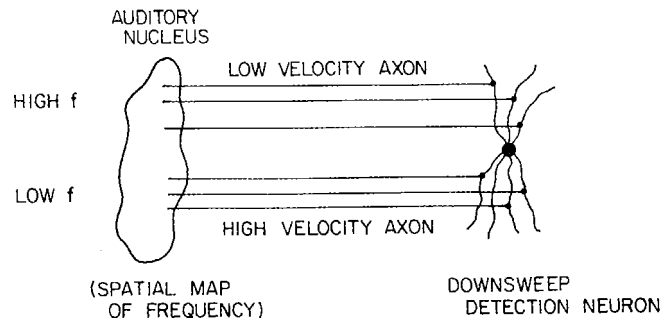
sequence can be decoded into a  
(frequency modulated) sonar in  
the ideal sonar signal for range de-  
pulse. A small object will gener-  
n of a precise delay time (accurate  
nce the return pulse has a high  
compared to the delay time. Un-  
ery high peak power is a difficult  
oats (and in human-made sonar  
carrier frequency is generated.  
(Fig. 1a) solves the transmitter  
processing is necessary to use the  
termine the time of arrival of the  
t.



**Figure 1.**  
(a) The waveform of a sonar pulse with a fixed carrier frequency. (b) The waveform of a down-chirp sonar pulse.

Chirping the pulse leads to a simple method of using the entire echo signal energy to obtain the delay time. The signal burst is sent with an instantaneous frequency which changes in time, as shown in Fig 1b. In this example, the frequency is swept from high to low, a down-chirp. The chirp pulse of an FM bat lasts a few milliseconds, and represents several decimeters in range.

The neural apparatus necessary to concentrate the appropriate information in time could have a simple structure. In early auditory nuclei, there is generally a spatial map of frequency (tonotopic map) available. Suppose that there is a down-sweep recognition cell in a second nucleus connected to the first one by axons with a graded set of propagation velocities, as sketched in Fig. 2. If the axons connecting the high frequencies to a cell in the second nucleus have small diameters and those connecting the low frequencies of larger diameter, then the high-frequency pathways are delayed with respect to the low-frequency pathways. When a chirped signal is received, the high-frequency cells in the first processing nucleus will be activated before the low frequency cells, since the first-arriving part of the reflected signal will be the high-frequency part. The difference of axon propagation time for different frequencies results in the arrival of the high-frequency part and the low-frequency part of the signal at the target cell at the same time. This neuron is strongly driven by a down-chirp, since the delays have been organized to "focus" a down-chirp signal. This same neuron would be insensitive to an up-chirp, for when the low frequencies arrive first the propagation delays would spread out the times of arrival of different parts of the up-chirp. The organization of appropriate time delays is essential to chirp

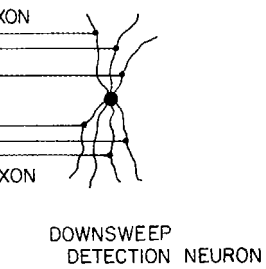


**Figure 2.**

The connections between neurons in a tonotopic map and a pulse recognition neuron. When a down-chirp arrives, the high-frequency neurons are activated earliest, and the activation of lower-frequency neurons is successively later. The delayed signals arriving at the pulse recognition neuron from the different frequency channels all arrive at the same time.

detection. The mechanism of the time delay does not matter. In sound location in the barn owl, a variation of axon length seems to be used for a similar function (Sullivan and Konishi, 1986). The organization of the time delays compresses all the signals of a particular form, essentially a sequence of frequencies occurring over an appreciable time duration, into a short recognition impulse at the end. We use organized time delays, on a rather longer time scale, to produce holistic recognition of speech-like auditory sequences (Tank and Hopfield, 1987a), and illustrate with applications to spoken syllables or simple words (Tank and Hopfield, 1987b).

Figure 3a shows the form of the actual acoustic signal for the spoken phrase "six seven five." The dominant structure visible is due to voicing, a characteristic low-frequency modulation present in sounds like "a" and "en" and lacking in the consonants "v" and "s" (the consonant sounds, not the letter names "vee" and "ess"). The frequency of the fundamental voicing is speaker-specific. Voicing carries some information about the words, but the fact that we also understand whispered words, which lack voicing, make it an inappropriate focus for word recognition. A short-time Fourier transform power spectrum of this speech signal as a function of time is shown in Fig. 3b. This representation makes more of the characteristic patterns of speech sounds visible. Speech is not a random signal, with an arbitrary spectrum. The anatomy and physics of its generation result in spectra of quite restricted forms. One useful way to think about the nature of the speech signal is to consider its spectra in "time bins" of ~10 msec duration. During such a time interval, the spectrum tends to have a stereotype form, and the number of such



otopic map and a pulse  
 arrives, the high-frequency  
 ization of lower-frequency  
 signals arriving at the pulse  
 enacy channels all arrive at the

elay does not matter. In sound  
 on length seems to be used for  
 1986). The organization of the  
 a particular form, essentially a  
 an appreciable time duration,  
 nd. We use organized time de-  
 produce holistic recognition of  
 Hopfield, 1987a), and illustrate  
 simple words (Tank and Hop-

al acoustic signal for the spoken  
 ructure visible is due to voicing,  
 on present in sounds like "a"  
 s "v" and "s" (the consonant  
 l "ess"). The frequency of the  
 oicing carries some information  
 o understand whispered words,  
 iate focus for word recognition.  
 pectrum of this speech signal as  
 his representation makes more  
 ounds visible. Speech is not a  
 um. The anatomy and physics  
 te restricted forms. One useful  
 eech signal is to consider its  
 on. During such a time interval,  
 form, and the number of such

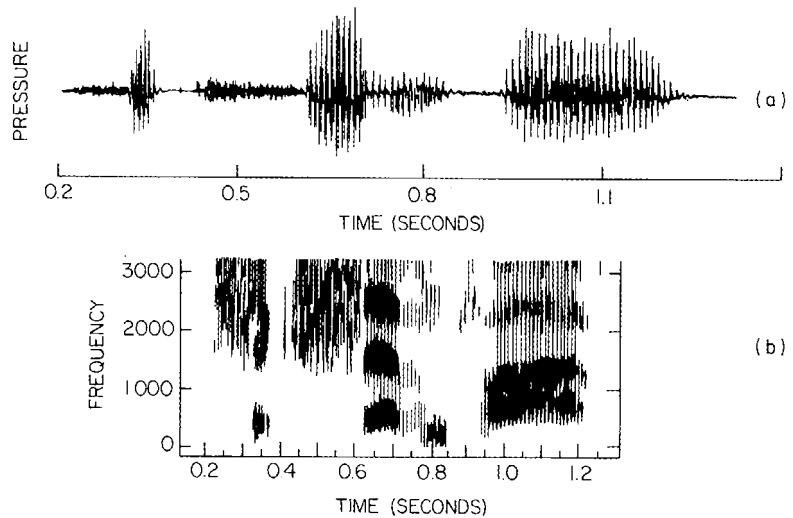


Figure 3.

(a) The electrical signal as a function of time recorded from a microphone responding to the spoken phrase "six seven five." (b) The power spectrum as a function of time for the acoustic signal above. The blackness of the recording indicates the level of acoustic power at the corresponding frequency and time. This calculation corresponds to measuring the power received by a bank of broadband (fast response) filters with varied center frequencies. The time interval between 0.50 and 0.55 sec is an overlap region belonging both to "six" and to "seven." There is no indication that "six" and "seven" are separate words.

forms is limited. For this discussion, ~100 such characteristic forms is adequate, and the speech in a particular time bin will be assigned to the characteristic form (symbol) that it most closely resembles. A short word, lasting 0.5 sec, can then be thought of as a sequence of 50 characteristic forms. In symbolic representation, a particular short vocal utterance can be described as a ~50-letter word, where each letter is taken from an alphabet of ~100 symbols.

The problem of word recognition can be precisely described in this representation. The model utterance for a given word is some particular list of symbols such as ACCHEUUUTUPELVVVVVKHGGGJQWWPPP. When this word is uttered by a speaker, however, this model sequence is not accurately reproduced. Different speakers, or a single speaker under varied circumstances, do not produce invariant sounds, so in this representation there will be symbol substitution errors in the data stream. A second problem is that speech can be produced at different rates, and idiosyncratic variations may even take place within a word. The differences that this "time warp" produces between the model utterance and an actual utterance involve symbol insertions and deletions. Time warp

is a major problem, for it prevents comparing between an actual utterance and model words by a rigid template superposition. Finally, in continuous speech the location of the ends of words are often not indicated in the acoustic signal. For example, the phrase "six seven" is generally spoken together as "sixseven", and there is no indication in the signal or its spectrum of the end of "six" and the beginning of "seven". This word-break problem greatly increases the difficulty of understanding words in continuous speech (see Fig. 3). The neural network approach to word identification (Tank and Hopfield, 1987a) that we will describe deals with all three of these fundamental difficulties.

### III. Model Circuitry Styled on Neurobiology

A simple model anatomy for doing phoneme, syllable, or word recognition tasks is shown in Fig. 4. The model is chosen in an attempt to meet two goals. First, the network must be a recognizable simplification of the kinds of electrophysiology and anatomy seen in mammalian brains. Second, a quantitative analysis of the electrophysiological response of the model must be possible to demonstrate that the network can solve the time series recognition problem. The model should also clearly indicate the kinds of signals and biophysics to be expected in a biological system that recognizes time sequences in the same generic fashion that the model network does. A review of modeling in this style and its relation to neurobiology has been recently published (Hopfield and Tank, 1986).

Area A in Fig. 4 begins the processing in the model network, and might correspond to a brainstem area of the auditory pathway in mammals, having a tonotopic map of the incident sound signal. The activity of a neuron reflects the intensity of the sound in the narrow frequency band to which the cell is primarily responsive. We will not describe the pathways by which this tonotopic representation is produced, but only use the output of such a known area as input to the sequence recognition system.

Area B contains two types of neurons. Its principal neurons are excitatory, and receive excitatory inputs in a tonotopic or direct map fashion from area A. These neurons also have axons projecting to area C. The other neurons in area B are inhibitory interneurons, which receive inputs from the excitatory cells in B and in turn inhibit these excitatory cells (local inhibitory feedback).

Area C is slightly more complex. It receives diffuse, nontotopic inputs from area B. Area C has both excitatory principal neurons and inhibitory interneurons. Some of the interneurons receive direct inputs from area B, and provide prompt feedforward inhibition to the principal

ing between an actual utterance  
perposition. Finally, in contin-  
words are often not indicated  
phrase "six seven" is generally  
e is no indication in the signal  
he beginning of "seven". This  
he difficulty of understanding  
The neural network approach  
(d, 1987a) that we will describe  
l difficulties.

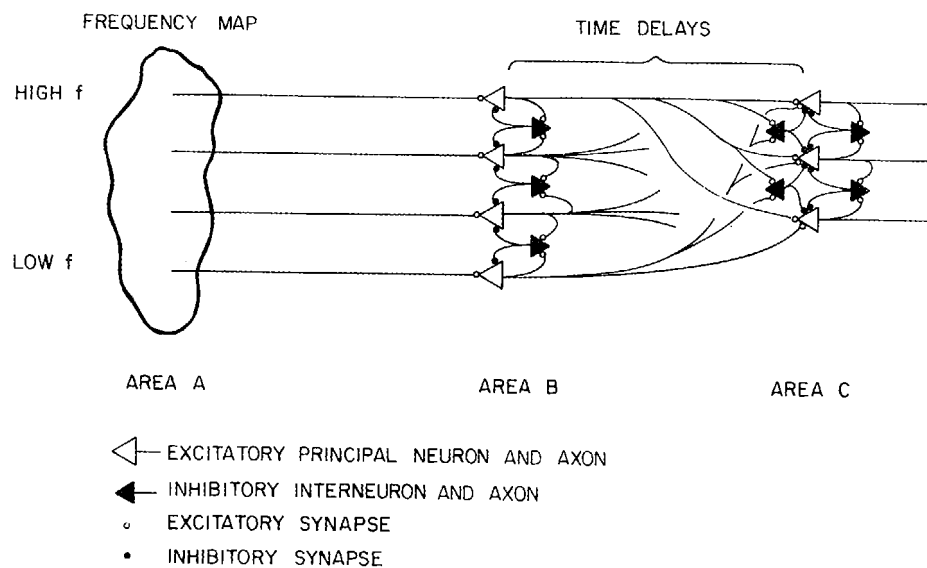
## try Styled \_\_\_\_\_ logy

name, syllable, or word recog-  
del is chosen in an attempt to  
be a recognizable simplification  
omy seen in mammalian brains.  
electrophysiological response of  
rate that the network can solve  
e model should also clearly in-  
cs to be expected in a biological  
n the same generic fashion that  
odeling in this style and its re-  
published (Hopfield and Tank,

ing in the model network, and  
the auditory pathway in mam-  
dent sound signal. The activity  
sound in the narrow frequency  
nsive. We will not describe the  
sentation is produced, but only  
input to the sequence recognition

rons. Its principal neurons are  
s in a tonotopic or direct map  
o have axons projecting to area  
ory interneurons, which receive  
in turn inhibit these excitatory

r receives diffuse, nontonotopic  
xcitatory principal neurons and  
er neurons receive direct inputs  
ward inhibition to the principal



**Figure 4.**

The anatomy of a model neural network for recognizing simple time sequences. Area A contains a tonotopic map, and is the source of signals for the processing areas B and C.

neurons. Other inhibitory cells receive their inputs from the principal neurons in C and produce feedback inhibition. There is also some particular biophysics that must be present in the B-C pathway to generate time delays and that will be described later.

While these two areas are much simpler in architecture and connections than is real cortical anatomy, their structure has generic similarity to the kinds of structures that have been observed in neurobiology. Stability (the absence of oscillations or large-amplitude spontaneous behaviors) is always a problem for neural circuits, whether model or real. There are two simple systems that have no stability problem, feedforward systems (in which there is no indirect return path from the axon of a neuron back to its dendrites) and symmetrically connected networks (Hopfield, 1982, 1984). Areas A and B are not literally symmetric, but if the inhibition pathways are fast compared to the time response of the principal neurons, they can be made equivalent to symmetric systems (Tank and Hopfield, 1986). This architectural design with feedforward connections between equivalently symmetrical areas has stability because it is an elementary hierarchical composition of the two kinds of stable structures.

We next describe the processing done in this system in terms of hypothetical electrophysiological experiments performed on such a system. The most elementary experiment is to record from the principal

neurons in area A while the preparation is treated with a pharmacological agent that suppresses inhibition, using a pure tone as an input signal. Any particular principal neuron would be found to have a tuning curve centered on an optimal sound frequency, and the optimum frequency would change smoothly as a function of physical location. The response of the same cells to the auditory stimulus of the spoken word "one" could also be measured, and typical such peristimulus time histograms (PSTH) for neurons of several different center frequencies (in the absence of inhibition) are shown in Fig. 5a. Most monosyllables will have qualitatively similar PSTH, for all spoken words have very broad power spectra, and will drive neurons having a wide range of center frequencies. There is nothing in such patterns that easily distinguishes one spoken word from another.

When the inhibitory system is also functioning, the single-tone experiments would look qualitatively similar, but with quantitative differences. First, the observed tuning curves would be sharper, due to the (indirect) inhibitory effect of one principal neuron on another. Second, the maximum response would be generally less, due to the operation of the inhibitory system. More elaborate experiments would display qualitatively different effects. Two tone experiments at frequencies  $f_1$ ,  $f_2$  would demonstrate two-tone suppression, in which the response of a principal neuron of center frequency  $f_1$  is suppressed by increasing the auditory signal at frequency  $f_2$  (for this effect,  $f_1$  and  $f_2$  must not be too close together). The inhibitory pathway can produce two-tone suppression. (Two-tone suppression also occurs at the level of the hair cell from mechanical inhibition.) Such suppression is a well-known part of auditory psychophysics and electrophysiology, and is the auditory equivalent of the center-surround receptive fields of early visual processing.

With inhibition functioning, the response of this system to the spoken word "one" is much more dramatic. The PSTH in the presence of inhibition (Fig. 5b) is qualitatively different from that without inhibition. Instead of all neurons being generally active, at any particular time only one or two now tend to be active. Furthermore, the pattern of times in which a neuron is active during a word stimulus is rather characteristic of that word, differing markedly from word to word. The inhibition has a major effect on the information processing, and is responsible for the feature enhancement that will eventually allow the identification of individual words in area C. The fact that we can now see patterns that are identifiable in the PSTH indicates that area C has some reasonable information to work with. While the inhibition reduces the amount of signal that is transmitted from A to the output of B, it reduces the noise and useless information much more than the true signal, and overall makes the pattern easier to identify.

If the signals are appropriately delayed prior to arriving at area C so that all the information relevant to a particular word arrives at once,

treated with a pharmacological pure tone as an input signal. found to have a tuning curve and the optimum frequency physical location. The response of the spoken word "one" peristimulus time histograms center frequencies (in the absence of monosyllables will have equal-ordered have very broad power wide range of center frequencies. easily distinguishes one spoken

functioning, the single-tone experiment, but with quantitative differences would be sharper, due to the effect of one neuron on another. Second, the response would be less, due to the operation of the system. The experiments would display a systematic organization of activity in which the response of the neurons is suppressed by increasing the number of active neurons. The effect,  $f_1$  and  $f_2$  must not be too close together. The system can produce two-tone suppression at the level of the hair cell from the stimulus. This is a well-known part of auditory processing and is the auditory equivalent of the sequential processing of early visual processing.

The response of this system to the spoken word "one" is shown in Fig. 5. The PSTH in the presence of inhibition is quite different from that without inhibition. In the presence of inhibition, at any particular time only one frequency band is active. Furthermore, the pattern of times in which the neurons are active is rather characteristic of the spoken word. The inhibition has a systematic organization, and is responsible for the sequential processing. It allows the identification of individual syllables. As we can now see patterns that emerge from the data, that area C has some reasonable amount of inhibition. This inhibition reduces the amount of information in the output of B, it reduces the noise in the signal, and overall the signal is more clearly displayed prior to arriving at area C. At any particular word arrives at once,

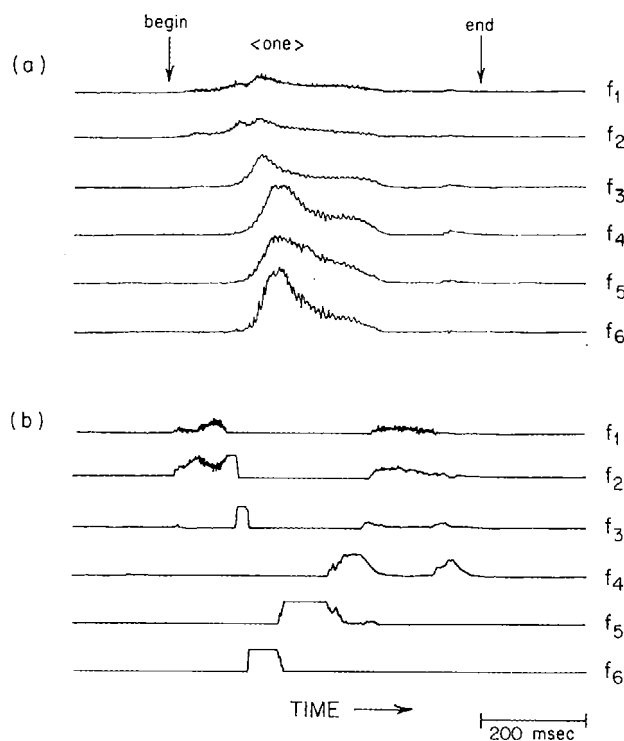


Figure 5.

(a) The peristimulus time histogram (PSTH) of the principal neurons of area B during the presentation of the spoken word "one" as a stimulus. Inhibition in area B has been suppressed. (b) Stimulus and recording as in (a), but in the presence of normal inhibition. An easily visible systematic organization of activity has been created.

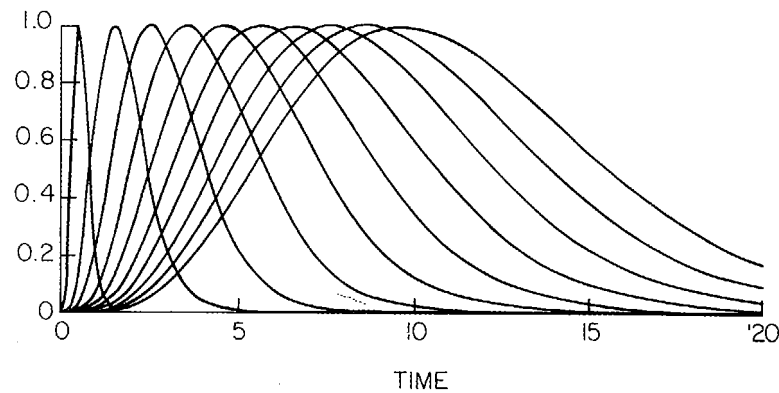
word recognition can be carried out in that processing area on the basis of the total information relevant to the word. We have already seen how to do this for a frequency sweep, which is one of the readily visible features of Fig. 5b, indicated by the sequential activation of  $f_6$ ,  $f_5$ ,  $f_4$ , and  $f_3$ .

The architecture of the model network performs the recognition of words in three stages. Area A provides the basic signal decomposition into frequency bands. Area B does feature extraction. The connections between area B and area C use time delays in a systematic fashion to organize the signals arriving at C. For simple word or syllable recognition, the requisite time delays lie in the interval 0.05–0.5 sec. To recognize a word as a whole, it is necessary to store in one way or another the information about the earlier parts of the utterance until the latter parts of the utterance have arrived. A delay mechanism is an elementary physical method for information storage. (In earlier times, electronic

computers used mercury delay lines—sound waves propagating in liquid mercury—for fast memory storage.)

We have developed and tested (Tank and Hopfield, 1987a) a model of how these time delays must be organized in order to recognize patterns as a whole in the presence of distortions. It was based on the idea that if some particular recognition neuron or neurons are to indicate recognition of a word by firing strongly for a short period of time immediately after the completion of the word, then the diverse signals that make up the word must all arrive coherently at that time. Looking at Fig. 5b, we see that if the neuron with optimal frequency  $f_2$  is connected to a recognition cell by a pathway that delays the signal for 0.6 seconds,  $f_3$  for 0.5 sec,  $f_4$  for 0.4 sec, 0.3 sec, and 0.1 sec, and  $f_5$  with 0.4 sec, then all these signals will arrive at the recognition neuron at the same time and drive that neuron strongly at that time, the termination of the utterance "one." These delays should not be precise. Because the word "one" might last anywhere from 0.6 to 0.75 sec, in order for the signals to add up coherently even in the presence of such time warp, a signal denoting the recognition of the feature (in  $f_2$ ) that might indicate the beginning of the word "one" should be sent to the recognition neurons with a variety of delays spread over this range. The appropriate response of the delay pathways to signals of short duration is shown in Fig. 6. The increasing width of these responses as a function of the mean delay is what makes the system able to cope with the time distortions typical of spoken sounds or other sequential recognition problems.

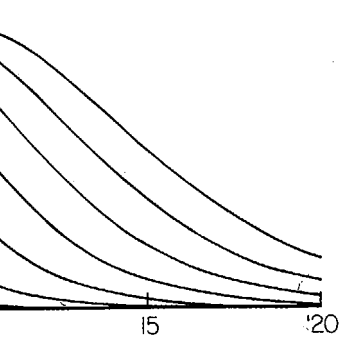
The feedforward inhibition in area C serves the function of de-



**Figure 6.**

The impulse response of the delayed signal propagation from the output of the principal neurons in B to the effect which these outputs have on the neurons in area C. Response curves for five different mean time delays are shown. The peak location indicates the mean delay. For a particular mean delay, a distribution of actual delays is needed, representing the fact that the time duration of a segment of a speech utterance has a probability distribution.

and waves propagating in liquid  
 (Tank and Hopfield, 1987a) a model  
 designed in order to recognize patterns  
 of signals. It was based on the idea that  
 neurons are to indicate recog-  
 nition over a short period of time immediately  
 after the diverse signals that make up  
 the utterance at time. Looking at Fig. 5b, we  
 see that frequency  $f_2$  is connected to a rec-  
 ognizer signal for 0.6 seconds,  $f_3$  for  
 0.4 sec, and  $f_5$  with 0.4 sec, then all  
 neurons are active at the same time and  
 the termination of the utterance  
 occurs. Because the word "one"  
 is recognized, in order for the signals to add  
 up to each time warp, a signal denoting  
 the beginning of the utterance might indicate the beginning  
 of the recognition neurons with a  
 delay. The appropriate response of  
 the recognition neurons is shown in Fig. 6. The  
 duration is shown in Fig. 6. The  
 response as a function of the mean delay is  
 shown in Fig. 6. The time distortions typical of  
 natural speech are shown in Fig. 6. The  
 response of area C serves the function of de-



...al propagation from the output  
 which these outputs have on the  
 ...ve different mean time delays are  
 ...ean delay. For a particular mean  
 ...eded, representing the fact that  
 ...h utterance has a probability

scribing negative evidence. For example, since the neuron  $f_2$  is not active  
 0.4 sec before the end of the word "one," an inhibitory pathway with  
 a delay of 0.4 sec from  $f_2$  to a neuron that is to recognize "one" can be  
 made. In computational terms, activation of  $f_2$  0.4 sec earlier is evidence  
 that a "one" recognition should not occur at the present time. The feed-  
 back inhibition in area C indirectly generates inhibitory pathways be-  
 tween neurons that are strongly activated by different words. Thus the  
 activity of principal neurons in C will be dominated at any time by those  
 neurons that are associated with a single word. The feedback inhibition  
 is the physical representation of the logical idea that at any particular  
 time, two different words cannot have been simultaneously completed.

The model network has been described in a neurobiological met-  
 aphor, and its processing specified. All the elements present in Fig. 4  
 can be implemented in simple analog electronic hardware, using only  
 resistors, capacitors, and operational amplifiers. We have built such a  
 network to recognize 10 short words (Tank and Hopfield, 1987b). With  
 connections chosen on the basis of the previous discussion it is capable  
 of recognizing its vocabulary of a few words in continuous speech. Figure  
 7 shows the acoustic signal  $S(t)$  and output of several different recognition  
 units (electronic model "neurons") in area C when the phrase "six net-  
 work repeat six" was spoken. Each recognition neuron becomes strongly  
 driven for a short time near the completion of its corresponding word,  
 and is inactive at other times. The PSTH shown in Fig. 5 for the utterance  
 "one" was actually obtained from the voltages in the corresponding

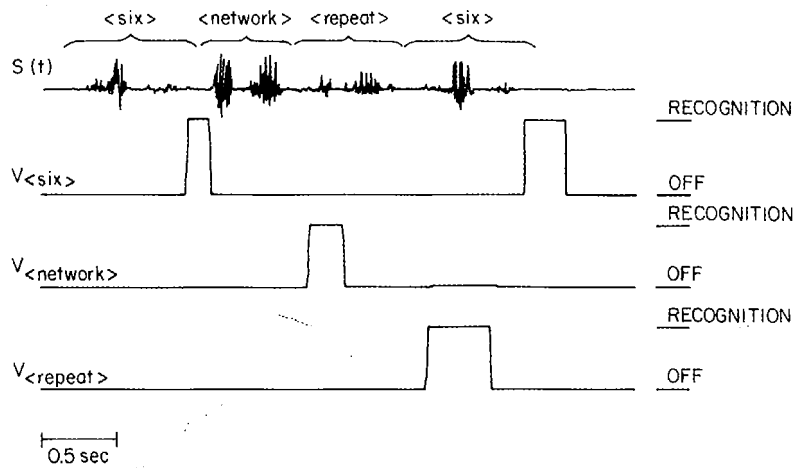


Figure 7.

Top trace: the waveform of the utterance "network six repeat network."  
 Second, third, and fourth traces: The activity of amplifiers playing the role  
 of principal neurons in area C. The first of these amplifiers is a recognizer  
 for the word "network," the second is for the word "repeat," and the  
 third for "six."

electronic circuit when the word "one" was presented. [In the analog electronic system, the output voltages of amplifiers correspond to the instantaneous firing rates of the biological neurons, and the output voltage as a function of time is equivalent to a PSTH (Hopfield and Tank, 1986)]

#### IV. What the Model Implies for Neurobiology

The computational problem of word recognition might be stated as follows. The recognizing system has information about a few words. For any short sequence of sound, and given an understanding of how words are likely to be distorted in ordinary speech, there exists a probability  $P_{\text{six}}$  that, if "six" were spoken, it would result in the observed sound segment. The recognition system should compute  $P_{\text{six}}$ ,  $P_{\text{network}}$ , etc. for all the words it knows (based on the input and the distortion model), find the largest such probability, and if that probability is not too small, then produce an output signal signifying that a particular word has just been spoken (Levinson *et al.*, 1983; Bahl *et al.*, 1983). The feature extraction, time delay, and competitive recognition neurons are a dynamic system that carries out the essence of this computational procedure for continuous speech in the presence of time warp and symbol substitution errors.

Since we understand how this model network responds and functions, we can ask what might be learned about its processing from some of the classic ways of studying a neural system. Consider for example an "electric shock to a tract" applied to the entire pathway between the tonotopic nucleus A and area B. Area B will give a brief response, which need not resemble any response elicited by a natural stimulus. Responses will be recorded in area C for times from 0 to 0.8 sec after the shock stimulus. Since these time-delayed responses are the essential mechanism by which the information in the time-dependent signal is organized for recognition, the discovery of these time delays would be of considerable significance. But the computational meaning of these delays could not be deduced from shock stimuli, nor would it even be evident whether these delays had *any* meaning in terms of information processing.

In this network, the feature extraction done in area B requires the inhibitory interneurons. Eliminating this inhibition results in unfamiliar and unrecognizable information being sent to area C. The information processing aspects of inhibition are equally important within area C. In our model circuit, the inhibition does not simply "keep the total activity down." The specific form of the inhibition, the synaptic interconnection topology and time course, are essential physical components of the computational structure. While decreasing the level of inhibition—for ex-

was presented. [In the analog of amplifiers correspond to the al neurons, and the output volt- to a PSTH (Hopfield and Tank,

## l Implies for \_\_\_\_\_ ogy

ognition might be stated as fol- mation about a few words. For an understanding of how words oeech, there exists a probability d result in the observed sound d compute  $P_{\text{six}}$ ,  $P_{\text{network}}$ , etc. for nput and the distortion model), that probability is not too small, g that a particular word has just hl *et al.*, 1983). The feature ex- cognition neurons are a dynamic his computational procedure for ne warp and symbol substitution

del network responds and func- l about its processing from some l system. Consider for example the entire pathway between the will give a brief response, which by a natural stimulus. Responses rom 0 to 0.8 sec after the shock nses are the essential mechanism ependent signal is organized for delays would be of considerable eaning of these delays could not ould it even be evident whether s of information processing.

ction done in area B requires the is inhibition results in unfamiliar sent to area C. The information ally important within area C. In ot simply "keep the total activity ion, the synaptic interconnection physical components of the com- g the level of inhibition—for ex-

ample, by pharmacological agents—can be a useful way to study elec- troanatomy, it has disastrous consequences for information processing in such a circuit.

We turn finally to the relation between these ideas and synapses, synapse dynamics, and synapse plasticity. The time scale of delays re- quired are so long that mechanisms other than a soma RC (resistance- capacitance) time constant or propagation delays would seem necessary. A variety of known mechanisms could produce such delays. Llinas and Yarom (1981) have studied posthyperpolarization rebound  $\text{Ca}^{2+}$  spikes, delayed by up to 0.3 sec after the release of hyperpolarization. Byrne (1980, 1982) has noted that an ordinary neuron with a fast potassium channel activation mechanism (A current) will have a delayed response to excitation, since depolarizing conductances will be masked by the  $\text{K}^+$  conductance until the potassium channel inactivates. Indirect pathways (Kehoe and Marty, 1980) by which the binding of a transmitter molecule activates a chain of biochemistry that terminates in the phosphorylation of a channel protein and increasing the open time of the channel could have delays of seconds. Thus there are many known mechanisms by which an appropriate length of delay could be generated between the activation of a presynaptic neuron and its electrical effect on a postsyn- aptic neuron. Of course, short delays could also be concatenated into longer delays by multicellular pathways.

Delay mechanisms like axon propagation delays (due to different propagation velocities or different lengths) and slow synaptic potentials are directly observable with straightforward electrophysiological tech- niques. Their action results in a direct change in a current or membrane potential in a neuron in some section of the delay pathway and can be observed with a microelectrode.

A perhaps richer class of delay mechanisms is beginning to be characterized that can be distinguished from the above by the fact that the mechanisms are in a sense "hidden": a biochemical change can occur in ion channel proteins with a time course similar to one of our spread- out delays (Fig. 6) but that will cause no observable change in current or membrane potential unless "tested." For example, Johnson and Ascher (1987) observed an interaction of glycine with the *N*-methyl-D-aspartate (NMDA) receptor. Application of glycine dramatically potentiates the membrane current produced by NMDA. The potentiation has a delay that lasts several seconds (the unbinding rate appears to be slow; the glycine binding rate has not yet been reported). A mechanism such as this is an elementary form of memory, and any such mechanism in con- junction with other machinery could be used to form the kind of delay and organize sequential information in the general fashion we have dis- cussed. But the temporary change in the nervous system that happens to the NMDA receptor upon glycine application is hidden to the inves- tigator unless probed with an appropriate electrophysiological experi-

ment. The simple circuit that we have described could be generalized to make use of such "hidden" delays in place of the more overt or direct delays.

Appropriate delayed connections would need to be made when an animal learns to recognize a particular sound or word. Such plasticity could be of two forms. In the simplest case, a distribution of intrinsic time delays would always be present and located before the modifiable synapses in the recognition network. To learn new sounds, ordinary synapses could be modified, and the necessary modification algorithm would be qualitatively similar to that described by Hebb (see Chapter 6, this volume). Alternatively, the synapse modification process could change the delay time of already existing connections. In either case, the modification paradigm would necessarily involve signals that arrive at the sensory system at different times, and then converge onto a synapse in such a fashion that together they alter its efficacy or the delay it produces.

The complex anatomy, biophysics, and biochemistry of the central nervous system contain a wealth of details that could be used as the basis for information and signal processing. Every experimenter finds immensely more facts and peculiarities than ever can be published, and selects for publication those that seem to be the most significant or interpretable. One function of this kind of modeling is to indicate how some of the unusual details of neurobiology may be used for processing information in unexpected ways. L. Kitzes (private communication) has observed long delays (among other unpublished complex behaviors) in A<sub>1</sub> cortex of awake monkeys. T. M. McKenna and co-workers (1988) have shown that the majority of neurons in primary auditory cortex of alert cats show tuning effects due to prior tones of different frequency which arrive 0.3–1.6 sec earlier. We have shown in a detailed model stylized on neurobiology that such delays in signalling between cells can be a general and simple biological method to organize sequence information, and can be central to understanding aspects of audition.

## References

- Bahl, L. R., Jelinek, F., and Mercer, R. L. (1983). *IEEE Trans. Pattern. Anal. Mach. Int. PAMT-5*, 179–190.
- Byrne, J. H. (1980). *J. Neurophysiol.* **43**, 630, 651–668.
- Byrne, J. H. (1982). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **41**, 2147–2152.
- Hopfield, J. J. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554–2558.
- Hopfield, J. J. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3088–3092.
- Hopfield, J. J., and Tank, D. W. (1986). *Science* **233**, 626–633.
- Johnson, J. W., and Ascher, P. (1987). *Nature (London)* **325**, 529.
- Kehoe, J. S., and Marty, A. (1980). *Annu. Rev. Biophys. Bioeng.* **9**, 437.
- Kleinfeld, D. (1986). *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9469.
- Levinson, S. E., Rabiner, L. R., and Sondhi, M. M. (1983). *Bell Syst. Tech. J.* **62**, 1035–1074.

e described could be generalized  
a place of the more overt or direct

would need to be made when an  
r sound or word. Such plasticity  
t case, a distribution of intrinsic  
and located before the modifiable  
To learn new sounds, ordinary  
necessary modification algorithm  
described by Hebb (see Chapter  
apse modification process could  
ting connections. In either case,  
ssarily involve signals that arrive  
s, and then converge onto a syn-  
ney alter its efficacy or the delay

s, and biochemistry of the central  
etails that could be used as the  
ssing. Every experimenter finds  
than ever can be published, and  
to be the most significant or in-  
of modeling is to indicate how  
bology may be used for processing  
ützes (private communication) has  
published complex behaviors) in  
McKenna and co-workers (1988)  
ons in primary auditory cortex of  
rior tones of different frequency  
have shown in a detailed model  
ys in signalling between cells can  
thod to organize sequence infor-  
anding aspects of audition.

## ces

- (1983). *IEEE Trans. Pattern. Anal. Mach. Int.*  
-668.  
*J. Biol.* **41**, 2147-2152.  
*J. A.* **79**, 2554-2558.  
*J. A.* **81**, 3088-3092.  
**233**, 626-633.  
*(London)* **325**, 529.  
*Biophys. Bioeng.* **9**, 437.  
**83**, 9469.  
I. M. (1983). *Bell Syst. Tech. J.* **62**, 1035-

- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967).  
*Psychol. Rev.* **74**, 431.  
Llinas, R., and Yarom, Y. (1981). *J. Physiol. (London)* **315**, 569.  
Suga, N. (1984). *Trends NeuroSci.* **7**, 20 (1984).  
Sullivan, W. E., and Konishi, M. (1986). *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8400-8404.  
Tank, D. W., and Hopfield, J. J. (1986). *IEEE Trans. Circuits and Systems* **33**, 533-541.  
Tank, D. W., and Hopfield, J. J. (1987a). *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1896-1900.  
Tank, D. W., and Hopfield, J. J. (1987b). *Proc. Int. Conf. Neurol Networks, 1st IV*, 455-468.