Biol. Cybernetics 22, 203—211 (1976)

On the Stochastic Dynamics of Neuronal Interaction

T. J. Sejnowski

Department of Physics, University of California at Santa Barbara, USA

Received: January 4, 1976

Abstract

The response of many neurons in the nervous system is a nonlinear function of membrane potential. Nevertheless, if the membrane potentials are normally distributed then their covariance satisfies a linear equation. This suggests that information in the nervous system may be processed by correlations between membrane potentials, a hypothesis which is subject to direct experimental test. The acquisition, storage, and retrieval of information in the form of correlations is consistent with present knowledge of human information processing.

Introduction

Remarkably little is known about the neuronal basis of memory. It is generally believed that short-term memory depends upon transient electrical activity, and that long-term memory is the result of relatively permanent structural changes. Before memory can be investigated at the neuronal level, the representation of information in the nervous system must be better understood.

A single neuron has only limited means for processing information, but a collection of neurons, through mutual interaction, can process many channels in parallel and can integrate information over a long time. Although there has been progress in understanding the response of single neurons, the investigation of information coding in parallel channels and information processing by the cooperative interaction between neurons has been hampered by both conceptual and experimental difficulties. Perkel and Bullock (1968, p. 266) have commented that "conceptually, the underlying theory of representation of information by impulses in a plurality of channels, not necessarily independent, is nearly nonexistent."

Correlation has been proposed as a basis for auditory processing (Licklider, 1959) and visual processing (Reichardt, 1962). More recently, similarities between correlation and long-term memory have been pointed out (Longuet-Higgens, 1968; Gabor, 1968). Correlation is, however, difficult to justify in nonlinear nervous systems. An even more fundamental question is how to define correlation at the neuronal level.

Action potentials are one possible carrier of information and certainly all sensory information must initially be coded into spike trains. Although action potentials serve the purpose of long-distance communication, it should not be forgotten that the spatial summation and temporal integration of electrical activity is accomplished by graded membrane potentials. Many neurons, in fact, do not even produce an action potential, and consequently have spatially restricted influence on other neurons.

This article considers the hypothesis that information is coded as correlations between spike trains and processed in a collection of neurons by correlations between membrane potentials. It will be shown that rate coding, which is of primary concern in many experiments, has an important role in controlling how correlation information is processed.

The analysis of neuronal interaction in the next two parts forms the basis for a theory of neuronal information processing. The experimental evidence for the theory is then reviewed and a direct though difficult experimental test is proposed. The last two parts of the article are concerned with a possible neuronal basis for sensory processing and memory. Several experimental concepts, such as spontaneous activity and reverberation, are briefly examined from the present theoretical perspective in the closing discussion.

I. Stochastic Dynamics

The integration of extracellular influence on a neuron takes place throughout its spatial extent and over time. Of all the electrochemical phenomena in a neuron, what should be saved in an analysis of neuronal interaction? Three aspects of functioning neurons are taken into account in the present treatment: first, temporal integration of input by the membrane potential; second, the response of a neuron as a function of membrane potential; and third, the communication of a neuron's response to other neurons.

The membrane potential of a neuron is approximately constant throughout its soma. Because of membrane capacitance and resistance, the membrane potential satisfies

$$\tau \frac{d}{dt} \phi(t) + \phi(t) = 0 ,$$

where τ is the membrane time constant and $\phi = 0$ is the resting potential. The solution is exponential decay

$$\phi(t) = \phi(0) e^{-t/\tau}$$

from an arbitrary initial potential $\phi(0)$.

Allow a collection of neurons to interact according to

$$\tau \frac{d}{dt} \phi_a + \phi_a = \eta_a + \sum_b K_{ab} L_b(\phi_b) , \tag{1}$$

where η_a are the external inputs and K_{ab} are the strengths of connection between neurons. The response of a neuron $L_b(\phi_b)$ represents what one neuron can communicate to other neurons about its internal state. The *interaction equation* (1) reflects the three aspects mentioned earlier: The left-hand side provides temporal integration, while the right-hand side takes into account the response of neurons and their interaction.

For spike producing neurons, let ϕ_a represent the membrane potential induced by external inputs in the absence of the action potential, as illustrated in Fig. 1. Define the *response function*

$$L_a(\phi_a) = r_a(\phi_a)/r_a^*$$

where r_a^* is the maximum rate of firing, and $r_a(\phi_a)$ is the rate of firing when the *induced membrane potential*, produced by a constant input current, is ϕ_a. A typical response function for a spike producing neuron is shown in Fig. 2. The response is zero below a threshold potential, and is bounded for large ϕ_a since the refractory period restricts the maximum rate of firing. The connection matrix K_{ab} gives the average influence of the b-th neuron, firing at maximum rate, on the a-th neuron.

If the external inputs are stochastic processes, then so are the membrane potentials, as determined by the nonlinear interaction equation. The expectation of (1) is

$$\tau \frac{d}{dt} \hat{\phi}_a + \hat{\phi}_a = \hat{\eta}_a + \sum_b K_{ab} P_b(\hat{\phi}_b) \tag{2}$$

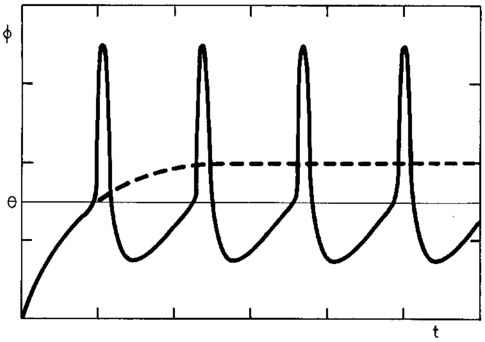

Fig. 1. The response of a neuron to a constant input current as a function of time. The membrane potential ϕ starts at the resting potential; when it exceeds the threshold potential θ, an action potential is released. In the absence of action potentials, the membrane potential follows the dashed line, which represents the induced membrane potential

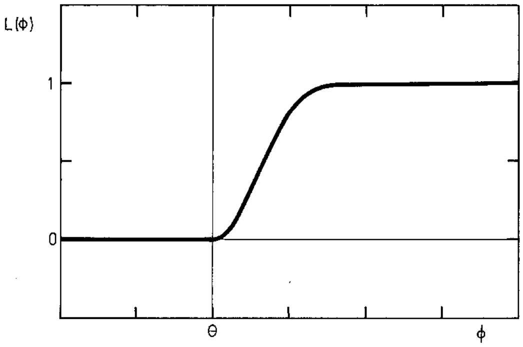

Fig. 2. Typical response of a spike producing neuron as a function of induced membrane potential ϕ. There is no response below a threshold potential θ. The rate of firing approaches a maximum, normalized to $L(\phi) = 1$, as ϕ increases

where

$$\hat{\phi}_a = E(\phi_a)$$
$$\hat{\eta}_a = E(\eta_a)$$
$$P_a = E(L_a(\phi_a)) .$$

Although P_a depends on all the moments of ϕ_a, only the mean will be indicated.

If ϕ_a is weakly stationary, then

$$\frac{d}{dt} \hat{\phi}_a(t) = 0 ,$$

and the mean membrane potentials satisfy

$$\hat{\phi}_a = \hat{\eta}_a + \sum_b K_{ab} P_b(\hat{\phi}_b) . \tag{3}$$

This equation was the basis of a previous study of stationary neuronal interaction (Sejnowski, 1976). If ϕ_a is an ergodic process, then $\hat{\phi}_a$ is equal to the time average for almost all sample functions.

For a neuron capable of producing an action potential, the mean rate of firing is

$$\hat{r}_a(\hat{\phi}_a)=r_a^*P_a(\hat{\phi}_a)\,, \tag{4}$$

where $P_a(\hat{\phi}_a)$ is the probability that the neuron will fire during a time interval $1/r_a^*$, which is approximately the refractory period. The time average rate of firing can only be meaningfully defined over a time interval which is long compared to $1/\hat{r}_a$, the average interspike interval.

The membrane potential of a neuron is the result of a larger number of synaptic events; consequently the membrane potential can be averaged over a far shorter time interval than the time average rate of firing. For this reason, and the fact that not all neurons produce an action potential, the membrane potential will be considered the primary variable.

II. Covariance

If information in the nervous system is processed by coordinated neuronal interaction, then membrane potentials may be correlated. The analysis of covariance in this part is the basis for subsequent study of correlation information coding and processing.

The covariance between membrane potentials is defined as

$$R(\phi_a(t),\,\phi_b(s))=E(\phi_a'(t)\phi_b'(s))\,,$$

where

$$\phi_a'=\phi_a-\hat{\phi}_a\,.$$

By virtue of the interaction equation, the covariance satisfies

$$\tau\frac{d}{dt}R(\phi_a(t),\,\phi_b(s))+R(\phi_a(t),\,\phi_b(s))$$
$$=R(\eta_a(t),\,\phi_b(s))+\sum_c K_{ac}R(L_c(\phi_c(t)),\,\phi_b(s))\,.$$

The analysis of these equations is made difficult by their nonlinearity and their large number. In primate cerebral cortex, for example, there are approximately 10^{10} neurons, and a typical neuron may receive afferent connections from 10^4 others. In order to gain further insight, some simplification must be sought which makes the problem tractable without, at the same time, doing serious violence.

According to the central limit theorem, the sum of a large number of independent random variables is approximately Gaussian. This suggests that the membrane potential of a neuron, which is the result of an extremely large number of synaptic events along many afferents, might be approximately Gaussian. However, the synaptic events are unlikely to be strictly independent.

In a neuron which produces an action potential, the timing of spike initiation depends on local conditions, but the response function $L_a(\phi_a)$ does not include these details. Therefore, other sources of random influence, not properly reflected in the equations, may tend to produce Gaussian processes. For the remainder of this article, let us assume that the membrane potentials have a joint normal distribution. The nonlinear terms in the covariance equation are then significantly simplified by the following:

Theorem. (Bussgang, 1952) Let X and Y be jointly Gaussian random variables, and L(Y) any bounded function. Then

$$R(X,\,L(Y))=R(X,\,Y)P'(\hat{Y})\,,$$

where

$$P'(\hat{Y})=\frac{\partial}{\partial\hat{Y}}E(L(Y))\,.$$

With the help of the above theorem, the covariance equation becomes, in matrix notation,

$$(\tau\frac{d}{dt}+A(t))R(\phi(t),\,\phi(s))=R(\eta(t),\,\phi(s))\,, \tag{5}$$

where

$$A(t)=I-K'(t)\,,$$

and

$$K_{ab}'(t)=K_{ab}P_b'(\hat{\phi}_b(t))\,. \tag{6}$$

The covariance is coupled with the mean membrane potentials (2), and the two equations must be solved simultaneously.

The *interaction matrix* $K_{ab}'(t)$ depends on the membrane potentials. Consequently, not all connections between neurons enter equally in determining the covariance. In a neuron which produces an action potential, the factor $P_b'(\hat{\phi}_b)$ is only significant when the mean membrane potential is near threshold; the interaction matrix is then a skeleton network connecting only neurons whose rates of firing are neither very low nor very high. These *critical neurons* are maximally sensitive to correlated inputs and make the largest contribution to the covariance equation.

The output from one collection of neurons often serves as the input to another. If C_{ab} is the connection matrix between two areas, and ζ_a are the inputs, then

$$\zeta_a=\sum_b C_{ab}L_b(\phi_b(t))\,,$$

where time delays have not been taken into account. Applying the above theorem, the input covariance is

$R_\zeta(t, s) = C'(t) R_\phi(t, s) C'^*(s)$,

where

$R_\phi(t, s) = R(\phi(t), \phi(s))$,

$C'_{ab}(t) = C_{ab} P'_b(\hat{\phi}_b(t))$,

and C'^* is the transpose of C'. Thus, even though neurons may have a nonlinear response, the internal transfer of covariance from one area to another is linear.

In the above analysis, the membrane potentials only entered indirectly, through their means and covariance. The form of the covariance equation suggests that the membrane potentials might themselves satisfy a similar equation, although this is not apparent from Eqs. (1) and (2).

Define the vector stochastic process ϕ'_a by the stochastic differential equation

$$(\tau \frac{d}{dt} + A(t))\phi'(t) = \eta'(t) , \tag{7}$$

with

$\eta'(t) = \eta(t) - \hat{\eta}(t)$.

It can be verified that ϕ'_a and $\phi_a - \hat{\phi}_a$ have identical first and second order moments; if, as assumed, they are both Gaussian, then they are in fact identical processes. Similarly, the internal transfer of information between two areas is

$$\zeta'(t) = C'(t)\phi'(t) . \tag{8}$$

Under stationary conditions the covariance depends only on time differences

$R_\phi(s - t) = R(\phi(t), \phi(s))$.

Also, since A and $R_\phi(0)$ are constant,

$$A R_\phi(0) + R_\phi(0) A^* = R_{\eta\phi}(0) + R_{\phi\eta}(0) , \tag{9}$$

where A^* is the transpose of A.
The variances are implicitly coupled with the mean membrane potentials, and Eqs. (3) and (9) must be solved simultaneously.

Although the stationary covariance equation is linear, it should be borne in mind that the constant matrix A nonetheless depends on external inputs. If the homogeneous solutions are

$(\tau \frac{d}{dt} + A)T(t) = 0$,

with

$T(0) = I$,

then the solution of the covariance equation is

$$R_\phi(s - t) = \int_{-\infty}^{t} dt' \int_{-\infty}^{s} ds' T(t - t') R_\eta(s' - t') T^*(s - s') , \tag{10}$$

where T^* is the transpose of T.

The stationary transition matrix $T(t)$ is well-known and can be constructed from the generalized eigenvectors of the interaction matrix:

$(\lambda_n I - K') m \psi^{nl}_m = 0$

$\psi^{*nl}_m (\lambda_n I - K')^{m_{nl} - m + 1} = 0$,

where λ_n are the eigenvalues of K', and the l-th eigenvector of λ_n has algebraic multiplicity m_{nl}. Within the dual null spaces, the two sets of generalized eigenvectors can be chosen biorthonormally, so that

$(\psi^{*nl}_m, \psi^{n'l'}_{m'}) = \delta_{nn'} \delta_{ll'} \delta_{mm'}$.

Each set of generalized eigenvectors forms a basis for the space, though not necessarily an orthonormal one.

The transition matrix is

$$T(t) = \sum_n e^{(\alpha_n - 1)t/\tau} \sum_k (t/\tau)^{k-1}/(k-1)! \cdot \{\cos(\beta_n t/\tau) \mathrm{Re}\, E^n_k - \sin(\beta_n t/\tau)\, \mathrm{Im}\, E^n_k\} , \tag{11}$$

with

$\alpha_n = \mathrm{Re}\, \lambda_n , \quad \beta_n = \mathrm{Im}\, \lambda_n$,

and

$E^n_k = \sum_l \sum_{m=k}^{m_{nl}} \psi^{nl}_{m-k+1} \otimes \psi^{*nl}_m$.

The state equation (7) together with the output (8) defines a linear dynamical system. Under stationary conditions, the solution is

$$\phi'(t) = \int_{-\infty}^{t} dt' T(t - t') \eta'(t') , \tag{12}$$

which has the form of a multidimensional linear filter (Wiener, 1949). Although the covariance satisfies a linear equation, linear superposition does not generally hold.

III. Neuronal Information Processing

Correlated input spike trains produce correlations between membrane potentials in a collection of neurons. This form of information coding and information processing will be called *ensemble correlation.* When all channels are independent and the membrane potentials are uncorrelated, ensemble correlation reduces to a single channel code, such as rate coding.

Ensemble correlation has several advantages: first, a more efficient use of the information capacity of available channels; second, means whereby exquisitely detailed sensory information can be coded by sensory afferents in apparently noisy activity; and third, the versatility of linear processing, as discussed in the next two parts.

Whether or not the nervous system uses ensemble correlation is an experimental question. Indirect evidence in favor of the possibility will be reviewed and a direct experimental test proposed.

Evoked potentials and EEG recordings reflect large-scale regularities in the central nervous system. If, as generally believed, these extracellular potentials are the result of the average postsynaptic potentials from many neurons, then any signal which survives the gross average may indicate large-scale correlations between membrane potentials. The regularities in these slow potentials are consistent with ensemble correlation.

There is evidence that spike trains are correlated in the peripheral nervous system. Representative data from different sensory systems include phase-locked discharge in the cochlear nucleus (Kiang *et al.*, 1965) and the somato-sensory system (Mountcastle, 1967), and preferred intervals between spikes in retinal ganglion cells (Ogawa *et al.*, 1966). However suggestive, this is weak evidence for ensemble correlation; there is no experimental indication of whether such information is decoded or used centrally.

Julesz (1971) has shown that stereopsis can be induced by random spatial patterns which are binocularly correlated. Although it is not yet clear how this information is represented or decoded, the fact of perception proves that it can be decoded. Some aspects of auditory perception, in particular "periodicity pitch", may indicate correlation information processing in the auditory system (Roederer, 1974).

The hypothesis that information is processed by correlations between membrane potentials can be tested by correlating intracellular recordings. The action potential in spike producing neurons can be eliminated by biasing the membrane potential with a hyperpolarizing current. Although the membrane potential measured under this condition reflects the passive integration of input currents, it may not accurately represent the induced membrane potential under normal conditions. Fortunately, the problem can be avoided by recording from neurons which do not produce an action potential.

Although intracellular recording is a fairly common technique in many preparations, there do not appear to be any correlation data in the literature. Published intracellular records show wavelike activity from some areas, such as the hippocampus (Fujita and Sato, 1964) and cerebral cortex (Elul, 1968). When correlating intracellular records, care is required in choosing a sufficiently long stationary interval.

In sensory areas, the strongest correlations should occur between neurons whose receptive fields overlap. Generally, the farther removed a neuron from sensory receptors, the greater the size of its receptive field; therefore, the coherence distance and coherence time for correlations should increase with each layer of processing. Those areas of the brain which are responsible for memory should exhibit correlations over a time scale comparable to that for short-term storage.

How does ensemble correlation affect average rates of neuron firing? Correlation information can be decoded, that is, converted into rate information for the purpose of motor commands, by introducing time delays in afferent. Parallel fibers in the cerebellum may serve this function.

IV. Sensory Processing

Without a specific model for a particular area of the brain, comparison between structure and function is not possible. Nevertheless, it would be worthwhile to show that general brain functions, such as sensory processing and memory, can be accounted for within the theory.

In humans, sensory information is preserved for only a few tenths of a second, during which particular features are extracted for short-term storage. Attention and expectation affect which features are extracted and depend on an associative long-term memory. The remainder of this article examines whether these concepts, which derive from psychological experiments (Norman, 1969) and have no known physiological basis, are consistent with the present theory of neuronal information processing.

The external inputs to a collection of neurons in an early stage of sensory processing may arise from both sensory afferents and the central nervous system. Consider an impulse input along sensory afferents

$$\eta'(t)=\eta'(t_0)\delta(t-t_0)\,.$$

Although this form of input serves mainly as a clear example, visual input, which is driven by microsaccades (Ditchburn, 1973), may be impulse modulated for preliminary processing.

An impulse input is not stationary. However, the means and variances of membrane potentials may remain approximately stationary over a short time interval. If so, then the interaction matrix remains constant and the stationary approximation is valid. Under these conditions, the response (12) of the membrane potentials to the impulse is

$$\phi'(t) = \begin{cases} T(t-t_0)\eta'(t_0) & t \geqq t_0 \\ 0 & t < t_0 \end{cases} .$$

The impulse response is a linear superposition of terms, each of which is exponentially damped and sinusoidally modulated. The envelope of the k-th term for the eigenvalue λ_n, owing to the factor t^{k-1}, has a peak of amplitude $1/(1-\alpha_n)^{k-1}$ occurring at $(k-1)\tau/(1-\alpha_n)$ after t_0. Thus, there is a succession of peaks at equally spaced intervals. The peaks decrease in amplitude if $\alpha_n < 0$; otherwise the successive peaks increase. Stability requires that $\alpha_n < 1$.

The subspaces defined by

$$\mathscr{F}^n_k = (\psi^{nl}_m)^{l=1,2,\dots}_{m=1,2,\dots,m_{nl}-k+1}$$

are nested

$$\mathscr{F}^n_{k+1} \subset \mathscr{F}^n_k$$

and are invariant subspaces of the linear operators

$$E^n_k : \mathscr{F}^n_k \rightarrow \mathscr{F}^n_k$$

which compose the transition matrix. The projection operators

$$E^n_1 E^m_1 = \delta_{nm} E^n_1$$

form a resolution of the identity.

$$\sum_n E^n_1 = I .$$

Let us call $\mathscr{F}^n_1$ the *feature subspace* of λ_n. Then the above description of the impulse response can be restated as follows: The impulse input η' is projected into the feature subspaces, each with its own characteristic time constant and modulation frequency. A *feature* is a spiral (in the degenerate case a straight line), first outgoing, then ingoing, in the plane spanned by the real and imaginary parts of ψ^{nl}_m. Different combinations of features in each feature subspace emerge sequentially, first $E^n_1\eta'$, then $E^n_2\eta'$, and so forth.

The information in the impulse is initially preserved in the covariance between membrane potentials. As features decay, information is lost. Feature subspaces with long time constants and high algebraic multiplicities remain longest.

The first part of the impulse response has the character of a sensory information buffer; the latter part resembles short-term memory (Norman, 1969). The successive unfolding of each feature subspace is associative. If a single feature ψ^{nl}_m were present in the external input, the output would follow the sequence: ψ^{nl}_m, ψ^{nl}_{m-1}, $\psi^{nl}_{m-2} \dots \psi^{nl}_1$, and terminate. That is, the original feature would be followed by a sequence of associated features, a process which will be called *feature association.*

Feature association for an impulse input, if averaged over a large population of neurons, resembles the sensory evoked response (John, 1972). In the general case, sensory input arrives continuously; feature association then provides continuous filtering of the information.

The means and variances of membrane potentials will be called the *background.* For spike-producing neurons, the background determines the mean rates of firing (4). The background also affects the interaction matrix (6), thereby selecting features and controlling feature association. By internal adjustment of the background, different features of sensory data can be selectively filtered and attended. Similarly, an adjustable filter can aid in searching for expected features. The background is a physiological equivalent of context.

Non-specific afferents to pyramidal cells in the cerebral cortex, arising from the reticular formation, basal ganglia, and elsewhere, make synaptic connections on distal dendrites. Specific afferents to primary sensory cortex arise from thalamic relay nuclei and synapse near the soma. The former afferents may provide background input; the latter are well-placed for providing correlation information.

The mathematical concept of feature introduced here is similar to previous suggestions that resonant modes of linearly interacting neurons may be identified with "distinctive features" (Greene, 1962; Anderson, 1974). The present treatment demonstrates how and under what conditions linear analysis is justified. The essential nonlinear character of neuronal interaction, however, cannot be neglected; it is because of nonlinearity that many different linear systems can be embedded in the same collection of neurons.

V. Central Processing

Several models of memory (Longuet-Higgens, 1968; Gabor, 1968) are based on correlation and are equivalent to linear filters (Borsellino and Poggio, 1972; Pfaffelhuber, 1975). Although these models were offered as analogies of memory, without any physiological basis, they nonetheless anticipate the present theory. The previous discussion of sensory processing is extended here to central processing.

The impulse input considered in the previous part was nonstationary, perhaps more so than for most sensory input. Nonstationary input produces nonlinear coupling, which may reduce the coherence time for correlations. Information could be processed in a linear fashion if the background were approximately stationary. Central processing, and in particular the retrieval of information from long-term memory, requires such stable conditions.

Large-scale correlations in afferents to an area produce correlations in membrane potentials; incoming information serves as the content address for the information subsequently generated by the interacting neurons. Information is retrieved from long-term memory as ensemble correlation, the same form in which information is retained in short-term storage. Ensemble correlation thus provides a close connection between information acquisition, storage, and retrieval. It is even possible that short-term storage occurs in the same areas of the brain which are responsible for long-term memory.

In addition to information retrieval in the spatial and temporal domains, feature association also allows retrieval in the frequency domain. If the afferents to a collection of neurons are modulated at a specific frequency

$$\eta'(t) = \eta'(0)e^{i\omega t/\tau},$$

then $\phi'(t)$ is given by the Fourier transform of the impulse response matrix

$$\phi'(t) = S(\omega)\eta'(t)$$

$$S(\omega) = \sum_n \sum_{k=1} \frac{E_k^n}{(1-\lambda_n+i\omega)^k},$$

where it is understood that only the real parts of $\eta'(t)$ and $\phi'(t)$ are considered. The main contribution comes from those feature subspaces with $\operatorname{Im}\lambda_n$ near ω, and $\operatorname{Re}\lambda_n$ near 1. Resonances occur as ω is scanned across the spectrum.

Is there any way to estimate the interaction matrix spectrum? The major observed rhythms of the EEG may be within the same frequency bands which are used for information processing and retrieval. A rhythm of frequency v is related to the eigenvalue spectrum by

$$2\pi v = \operatorname{Im}\lambda_n/\tau .$$

For a membrane time constant of $\tau \sim 15$ msec, the alpha rhythm corresponds to

$$\operatorname{Im}\lambda_n \sim 1 .$$

A more direct estimate of the spectrum may be obtained from correlation of intracellular recordings (Elul, 1968).

Information is stored on a long-term basis in the global structure of connections between neurons. Feature association for a given background depends only on those neuronal connections which make a significant contribution to the interaction matrix. New information can therefore be selectively stored by selectively altering connections in the interaction matrix. As more backgrounds are developed, interference between them increases, thereby limiting the storage capacity. How connections could be altered to store correlation information will be examined elsewhere.

The time scale for consolidation of long-term memory is longer than that for short-term dynamics; modification of feature association (Sejnowski, 1976) for a given background can be treated adiabatically. Feature association, in the sense shown below, varies smoothly with adiabatic change to the strengths of connection between neurons.

Assume that conditions are stationary and let the interaction matrix depend analytically on ε so that

$$A(\varepsilon) = I - K'(\varepsilon) .$$

In order to determine how the transition matrix depends on ε, take the Laplace transform of (7) and obtain

$$\mathscr{L}[T(\varepsilon)] = (s\tau I + A(\varepsilon))^{-1} .$$

Therefore, by the inverse Laplace transform,

$$T(\varepsilon) = \frac{1}{2\pi i}\int_{\mathscr{C}} ds(s\tau I + A(\varepsilon))^{-1}e^{st},$$

where the contour $\mathscr{C}$ is parallel to the imaginary axis and lies to the right of all singularities. Since the integrand is analytic in ε everywhere on the contour (Kato, 1966), $T(\varepsilon)$ must be analytic as well. Neither the eigenvalues nor the eigenvectors of the interaction matrix, however, are necessarily analytic.

Many questions concerning the neuronal basis of memory remain to be explored. What is the time scale for coherent correlations between membrane potentials in particular areas? How do the correlations vary with the positions of neurons within an area? How do correlations between neurons affect the strengths of connection between them? The experimental answers to these questions may provide a foundation for a neuronal theory of memory.

Discussion

Throughout this article membrane potentials have played the primary role in representing and processing information, but clearly the role of spike trains in transmitting information cannot be neglected. The correlations between membrane potentials in one area are encoded as correlations in parallel spike trains for use in other areas. In a sense, the action potential is a shadow, representing in summary fashion the local cellular integration.

Several other possible neural codes (Perkel and Bullock, 1968) are special cases of a correlation code. For example, special patterns of firing along a single fiber can be described by autocorrelation, which is a diagonal term of the correlation matrix. An interval code along a single fiber, in turn, is a special case of autocorrelation. Impulse coding in a pair of fibers can be accounted for by cross-correlation, an off-diagonal term of the correlation matrix. Single channel codes, such as rate coding, are always present as the background, even when all channels are independent and there are no correlations between membrane potentials.

Although from an experimental point of view all these possibilities must be considered separately, there is conceptual advantage to thinking of correlation as an ensemble property, either as ensemble correlation in spike trains along parallel fibers, or as ensemble correlation between membrane potentials in a cellular assembly.

Apparently random spontaneous activity can be found in most parts of the nervous system. Why should such noisy conditions be maintained? There has been a general feeling that spontaneous activity must have something to do with the basic operation of the brain. For example, Shepherd (1974, p. 205) has written:

"Both regions [retina and cerebellum] process information against a background of steady activity; the transient responses, therefore, are in the form of perturbations of ongoing activity. This apparently is a more precise mode of information transfer than is transmission by excitatory responses against little or no background. Since background activity may be adjusted under different behavioral conditions, information may be transmitted about steady state conditions, and transient inputs interpreted relative to those states."

The equations which describe neuronal interaction in a collection of neurons are nonlinear, owing mainly to the threshold for firing and the refractory period of spike producing neurons. When treated as a stochastic equation, however, and under reasonable assumptions, the equation for covariance between membrane potentials is linear. Randomness plays an essential role in achieving this simplification.

A neuron is maximally sensitive to input correlations when its average membrane potential is near threshold; if the nervous system uses a correlation code then there is advantage to maintaining spontaneous activity in neurons.

Another striking aspect of the nervous system is the extensive interconnection between neurons. How is neuronal interaction organized and coordinated? Early neuronanatomical evidence for feedback loops between neurons led to the idea of reverberation and reverberatory circuits (Lorente de Nó, 1938). Although circulatory pathways are known to exist, it is not clear what, exactly, is circulated. In many areas of the brain, a single action potential is insufficient, by itself, to give rise to an action potential in another neuron. What remains of reverberation is the intuitive feeling that highly interconnected neurons should, in some sense, show repetitive patterns of interaction.

Correlations between membrane potentials are a fine measure of coordination within a collection of neurons. Significantly, the solution of the stationary covariance equation is a linear superposition of damped oscillatory terms. In view of the random nature of electrical activity of neurons, it might be worthwhile to identify reverberation with correlations between membrane potentials; that is, to consider reverberation as a statistical property rather than a fixed repetition of any particular pattern.

The question arises to what extent the results of this article depend on simplifying assumptions. The idealized neurons considered here do not include axonal latency, the time development of the postsynaptic potential, or dendritic electrotonus. As will be shown elsewhere, these refinements do not alter the main results. The key assumption that membrane potentials are normally distributed can be tested by experiment.

Although the present theory was inspired by a biological invention, any system which satisfies the equations for neuronal interaction is, of course, subject to the same analysis, regardless of its material realization.

Acknowledgement. Discussions with David Bender and Alan Gelprin have helped clarify the ideas in this article.

References

Anderson, J.A.: What is a distinctive feature? Technical report number 74–1 from the Center for Neural Studies, Brown University (1974)

Borsellino, A., Poggio, T.: Holographic aspects of temporal memory and optomotor response. Kybernetik **10**, 58–60 (1972)

Bussgang, J.J.: Crosscorrelation functions of amplitude distorted Gaussian signals. Res. Lab. Elec. MIT Tech. Rep. **216**, 1–14 (1952)

Ditchburn, R.W.: Eye-movements and visual perception. Oxford: Oxford University Press 1973

Elul, E.: Brain waves: Intracellular recording and statistical analysis help clarify their physiological significance. In: Enslein, K. (Ed.) Data acquisition and processing in biology and medicine. Pp. 93–115. Oxford: Pergamon 1968

Fujita, Y., Sato, T.: Intracellular records from hippocampal pyramidal cells in rabbit during theta rhythm activity. J. Neurophysiol: **27**, 1012–1025 (1964)

Gabor, D.: Holographic model of temporal recall. Nature **217**, 584 (1968)

Greene, P.H.: On looking for neural networks and "cellular assemblies" that underlie behavior: I. Mathematical model. Bull. Math. Biophys. **24**, 247–275 (1962)

John, E.R.: Switchboard versus statistical theories of learning and memory. Science **177**, 850–864 (1972)

Julesz, B.: Foundations of cyclopean vision. Chicago: University of Chicago Press 1971

Kato, T.: Perturbation theory for linear operators. Berlin-Heidelberg-New York: Springer 1966

Kiang, N.Y.-S., Watanabe, T., Thomas, E.C., Clark, L.F.: Discharge patterns of single fibers in the cat's auditory nerve. Cambridge: MIT Press 1965

Licklider, J.C.R.: Three auditory theories. In: Koch. S. (Ed.): Psychology, a study of science, Vol. I. Pp. 41–144. New York: McGraw Hill Book Co. 1959

Longuet-Higgins, H.C.: Holographic model of temporal recall. Nature **217**, 104 (1968)

Lorente de Nó, R.: Analysis of the activity of chains of internuncial neurons. J. Neurophysiol. **1**, 207–244 (1938)

Mountcastle, V.B.: The problem of sensing and neural coding. In: Quarton, G.C., Melnechuk, T., Schmitt, F.O. (Eds.): The neurosciences: A study program. Pp. 393–407. New York: Rockefeller University Press 1967

Norman, D.A.: Memory and attention: An introduction to human information processing. New York: Wiley 1969

Ogawa, T., Bishop, P.O., Levick, W.R.: Temporal characteristics of response to photic stimulation by single ganglion cells in the unopened eye of the cat. J. Neurophysiol. **29**, 1–30 (1966)

Perkel, D.H., Bullock, T.H.: Neural coding. Neurosci. Res. Progr. Bull. **6**, 221–348 (1968)

Pfaffelhuber, E.: Correlation memory models – A first approximation in a general learning scheme. Biol. Cybernetics **18**, 217–223 (1975)

Reichardt, W.: Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In: Rosenblith, W. (Ed.): Sensory communication. Pp. 303–317. Boston: MIT Press 1962

Roederer, J.G.: Auditory processing in the nervous system. In: Conrad, M., Guttinger, W., Dal Cin, M. (Eds.): Physics and mathematics of the nervous system. Lecture notes in biomathematics, Vol. 4. Pp. 211–227. Berlin-Heidelberg-New York: Springer 1974

Sejnowski, T.J.: On global properties of neuronal interaction. Biol. Cybernetics **22**, 85–95 (1976)

Shepherd, G.M.: The synaptic organization of the brain. Oxford: Oxford University Press 1974

Wiener, N.: Extrapolation, interpolation, and smoothing of stationary time series. Cambridge: MIT Press 1949

T. J. Sejnowski
Dept. of Psychology
Princeton University
Princeton, N. J. 08540, USA