Dynamics of Encoding in a Population of Neurons

BRUCE W. KNIGHT

From The Rockefeller University, New York 10021

ABSTRACT A simple encoder model, which is a reasonable idealization from known electrophysiological properties, yields a population in which the variation of the firing rate with time is a perfect replica of the shape of the input stimulus. A population of noise-free encoders which depart even slightly from the simple model yield a very much degraded copy of the input stimulus. The presence of noise improves the performance of such a population. The firing rate of a population of neurons is related to the firing rate of a single member in a subtle way.

1. INTRODUCTION

In a nervous system it is usual for extremely precise over-all results to arise from the functioning of a collection of components which have very modest precision in their individual construction and behavior. In the human ear, for example, such prodigies as "perfect pitch" are accomplished by a population of neurons which are somewhat haphazard in morphology, and which individually show ragged firing patterns. Apparently it is the collaboration of a large number of units which is responsible for the precision of the overall result.

In the discussion below, we will examine several models of the process by which a stimulus is encoded to evoke a train of impulses in a single neuron. The behavior of a large population of such neurons will then be explored. The effects that result from variations among members of the population and from irregular behavior of individuals also will be investigated. The most important results will be deduced in section 2, almost without recourse to formal mathematics; the mathematically most difficult results will be presented last. A following paper will compare theoretical results developed here with experiment (Knight, 1972).

This investigation of encoding was undertaken in order to predict quantitatively the inhibitory postsynaptic potential in the visual cell (eccentric cell) of *Limulus*. Here the postsynaptic potential level arises from the pooled effect of nerve impulses arriving from numerous presynaptic neurons. In this well-studied bit of nervous system the dynamics of the various neurological components are known (Knight et al., 1970) well enough to enable us, in principle, to predict the dynamics of the entire eye from stimulus to response. Each small illuminated region of the eye may be conceived as a subpopulation of identical neurons experiencing identical input. In order to determine the effect of this subpopulation upon a particular postsynaptic potential, we must have theoretical tools which enable us to obtain the population response from dynamical laws given initially for individual neurons. The results of this present study yield such tools, and predictions of considerable precision can in fact be made.

The same general problem arises frequently in the consideration of other neural systems. In the visual system of the primate, for example, continuous sensory input apparently is first coded into trains of discrete impulses at the level of the retinal ganglion cells. The next synapse along the major visual pathway, at the lateral geniculate nucleus, apparently is not of the highly convergent type but serves more nearly as a relay station. However, when the geniculate neurons arrive at the visual cortex they give rise to electrophysiological phenomena (Hubel and Wiesel, 1968) which suggest a convergence scheme that bears some close analogies to that of the *Limulus* eccentric cell. Several further layers of population convergence follow, giving rise to neural responses at successive levels of abstraction. Similar statements can be made concerning the secondary visual pathway which conducts impulse trains from the retinal ganglion cells to the superior colliculus.

A similar situation arises in the auditory system. If we conceptually divide the cochlear canal into short sections, we find over the lower half of the frequency range that the mechanical motion of a given section is transcribed into the level of impulse activity in the subpopulation of neurons which arise within it (Brugge et al., 1969), although any given neuron in that subpopulation contributes only a slight fraction of the total activity. The frequency bandwidth for the entire subpopulation greatly exceeds the repetition rate of a single component neuron.

Presumably within the vertebrate central nervous system the remote transmission of information typically is not entrusted to a single neuron, and the multiple channel considerations explored here again will be relevant.

At the motor end of the vertebrate nervous system such considerations again arise. For example, in the spinal stretch reflex circuits of the cat each stretch receptor appears to terminate on all motor neurons of a pool (Mendell and Henneman, 1968).

Three conclusions form the main theme of this paper. The first conclusion is that a particularly simple model for the encoding of a stimulus into nerve impulses yields the result that the variations with time of the firing rate of an entire population can be a perfect time replica of the stimulus. The population firing rate thus has the remarkable property that it may duplicate the stimulus with an indefinitely high degree of fidelity. The second conclusion is that the simple model is essentially unique in this respect, and that more realistic models of encoders are susceptible to spontaneous synchronization, a pathology which makes the temporal variation of the population firing rate a far less useful indicator of the shape of the stimulus. The third conclusion is that this pathology may be thwarted by a population of encoders whose impulse encoding is subject to chance fluctuations. Thus the fact that the encoders are heterogeneous and noisy becomes positively a virtue, which allows the temporal variation of the population firing rate to approach the ideal: a perfect replica of the shape of the stimulus.

In section 2 we will present the "simple integrate-and-fire" model of a neuron. There it will be observed that, in a large population of such neurons, the temporal variation of firing rate of the entire population is a perfect copy of the input stimulus. We will see that this result is still maintained when we introduce individual variations among the members of the population. Finally, we will generalize the model in a way which introduces random fluctuations into the spike train of each individual neuron, and show that the firing rate of the population will still remain a perfect copy of the input stimulus.

In section 3 we investigate the momentary firing rate of a single neuron in the population. If the stimulus is constant, the individual firing rate is proportional to population firing rate. However, if the stimulus is time varying, the single-neuron rate is *not* in fixed proportion to the population rate, nor is it a faithful replica of the input stimulus. It shows two distinct sorts of distortion. The first is nonlinearity in response to large stimulus fluctuations. The second distortion is a phase shift and amplitude attenuation in response to stimulus fluctuations at high frequency. These distortions can be very important when one gathers impulse data from a single nerve fiber, and thereafter tries to deduce the level of impulse activity in an entire population.

In section 4 a determining relationship is discovered between the individual neuron impulse rate and that of a whole population of identical neurons. The result is independent of the impulse encoding model, so long as that model is deterministic (not probabilistic).

Section 5 introduces the "forgetful integrate-and-fire" model for neuron firing. One model feature—infinitely long memory—of the simple integrate-and-fire model, is removed. The introduction of slow decay in memory has slight effect on the single neuron firing rate, but an important new feature appears in the population firing rate: at certain stimulus frequencies, the response of the population is disproportionately large so that the population response is no longer a perfect copy of the stimulus time-course.

Section 6 investigates the response of the forgetful integrate-and-fire model neuron to a periodic stimulus of finite amplitude; emphasis is placed on the phenomenon of "phase locking." Unlike the simple integrate-and-fire case, a population of forgetful neurons will tend to "fall into step," and all fire synchronously at a particular point in the stimulus cycle. The result of section 5, that certain infinitesimal periodic variations in the stimulus lead to disproportionately large responses, was the first hint of this phenomenon.

Section 7 discusses a general theory of the behavior of deterministic encoders in response to periodic stimuli. A ready-made mathematical machinery (created for a different reason) already exists for this problem. A general conclusion emerges: there are two distinct classes of impulse encoders, those which show the tendency to phase lock and those which do not. The class which do not are a slight generalization of the simple integrate-and-fire model, and share the feature of indefinitely extended memory. A population of encoders which do phase lock give the worst possible departure from a perfect copy of the stimulus: their response is in the form of synchronized bursts of impulses, which neither delimit the form of the input stimulus, nor yield more information than does the response of a single encoder. Such encoders also may be brought to the synchronized condition by stimuli which are not periodic.

Section 8 investigates a population of probabilistic (or "stochastic") encoders. A population of such encoders overcomes the phase-locking problem, and in spite of the limitation of finite memory duration, the condition may be approached of a population response rate that gives a perfect representation of the stimulus.

While the following discussion deals with two specific models, these two models do follow from reasonable idealization of the Hodgkin-Huxley equations. In particular, the small-signal frequency response of the impulse encoder in the *Limulus* eccentric cell may be described accurately in terms of a simple integrate-and-fire model (Knight et al., 1970). A slight tendency of this encoder to phase lock to very large signals (see Fig. 5 *b*) suggests that a slightly forgetful integrate-and-fire encoder model would furnish an even more accurate description.

The simple and the forgetful integrate-and-fire models differ in the degree to which a population of such encoders will "fill in" the detailed time profile of a periodic stimulus. (Phase locking is a complete breakdown in this "filling in.") The whole gamut of possibilities in fact arises for different neurons. While at present it is not feasible to record individually from a uniform population of neurons, it is possible to do something equivalent: to record from a single member over repeated stimuli. This has been done, for example, for the retinal ganglion cell (Hughes and Maffei, 1966), in the auditory system (Kiang et al., 1965; Brugge et al., 1969, 1970; Goldberg and Brown, 1969; Rose et al., 1967, 1969; Aitkin et al., 1970), for the innervation of fingertip skin (Mountcastle et al., 1968; Talbot et al., 1967), and for the mammalian muscle spindle (Brown et al., 1967; Matthews and Stein, 1969; Poppele and Bowman, 1970.) (The preceding references are not exhaustive.) The sections that follow should be helpful in the interpretation of these and similar investigations.[1]

2. INTEGRATE-AND-FIRE MODEL

One of the simplest possible impulse encoder models is the following: we imagine a noise-free neuron which contains an internal variable (which we call u) whose value increases at a rate given by the present value of the stimulus [called $s(t)$] which is being encoded. Thus,

$$\frac{du}{dt} = s(t). \qquad (s(t) \geqq 0). \tag{2.1}$$

When u achieves a criterion level (C), a nerve impulse is fired, u is reset to zero, and the process starts again. In general, larger stimuli will encourage higher firing rates. The Hodgkin-Huxley equations may be made to yield this model in a limiting case.[2]

It is easy to see how a large population of such identical and noiseless encoders will respond to a stimulus. Let us define the "density" $\rho(u)$ by the property that in the population there are a number $\rho(u_1)du$ of encoders for which the value of u falls between u_1 and $u_1 + du$. In Fig. 1 the solid curve indicates how $\rho(u)$ might look at a particular moment. A short time later, those encoders which have not fired will have advanced to larger values of u, and the dashed curve will be obtained. The whole curve marches rigidly to the right. According to equation 2.1, its speed of advance is given by $s(t)$. The rate at which firings occur in the entire population will be the rate at which encoders reach the firing point $u = C$. This will depend jointly on the height of the curve $\rho(u)$ at the point $u = C$, and on the rate of the curve's advance. Thus, the population firing rate r is given by

$$r = s(t)\rho(C). \tag{2.2}$$

It is an evident property of this model that, if initially the population are not uniformly distributed over u, this condition will persist forever, and even a constant stimulus s will lead to a periodic fluctuation in the population firing

---

[1] Among neurophysiologists the term "phase locking" is frequently used in an unfortunate colloquial way that blurs the distinction between neurons that phase lock and those that do not. Thus, for example, Rose et al. (1967) state very explicitly (footnote 3 of that paper) that they are observing "phase-preference," although the colloquial usage of "locking" appears in their title. Here throughout we use "phase locking" in the strict sense of seeking a fixed phase with respect to the stimulus. In this sense the simple integrate-and-fire model shows phase preference but does not phase lock.

[2] This model was advanced by Partridge (1966). Partridge's comment that "even in a multi-channelled system, considerable distortion could result from the process of pulse rate translation of a dynamic signal" specifically does *not* apply to this model.

rate. The passage of time will not smooth a firing rate which was not smooth initially.

The converse is also true, that if initially the population are uniformly distributed (straight line at ρ_0 in Fig. 1), then they will always remain so. In this case equation 2.2 becomes

$$r = \rho_0 s(t). \tag{2.3}$$

Thus, if the population are uniformly distributed in u, then the population firing rate will be a perfect copy of the stimulus. A more formal derivation of this result will be given near the end of section 4.

The simple preceding discussion deals only with what may be *expected* of an extremely large population of encoders. By choosing a large enough population we may make the fluctuations away from expected behavior arbitrarily small. For a finite population questions concerning fluctuations about expected values demand a far more elaborate methodology than we will develop here. Therefore such questions will not be pursued.

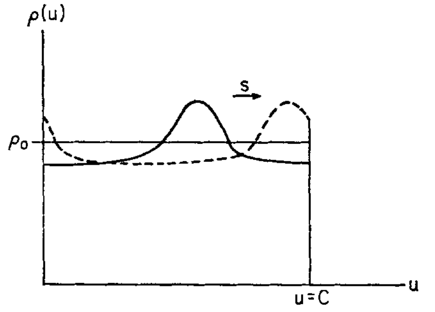

FIGURE 1

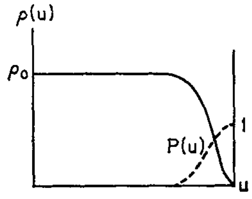

FIGURE 2

FIGURE 1. Time-course of the population density function $\rho(u)$.

FIGURE 2. Steady-state population density function, for a stochastic threshold distributed according to the probability function $P(u)$.

So far the discussion has involved a population of neurons which are identical, in the sense that they all have the same firing threshold level C. The generalization to a heterogeneous population, with a distribution of threshold levels, is immediate. Simply divide the population into subpopulations, according to their values of C. The perfect replica argument holds for each subpopulation, and hence for the total.

A word of caution should be added concerning this heterogeneous population model: to achieve the perfect-replica result, each individual subpopulation must be started with a uniform distribution over the internal variable u. This model does not smooth itself.

The simple integrate-and-fire model may be slightly generalized, to include the feature that neurons fire somewhat unpredictably. The generalization might be described as "the integrate-and-fire model with stochastic threshold." Together with equation 2.1, we assume that there is a probability $P(u_1)$ that a particular neuron will fire before its internal variable u reaches the value u_1. The probability $P(u)$ eventually rises to unity, for large enough u. The population density $\rho(u)$ is shown in Fig. 2. The stochastic threshold condition we have postulated may be expressed as

$$\rho(u) = \rho_0(1 - P(u)), \tag{2.4}$$

which is the content of Fig. 2. According to equation 2.1, individual encoders still advance in u at a speed $s(t)$. In particular, that is their speed of advance at $u = 0$, whence their total rate of appearance at $u = 0$ must be $\rho_0 s(t)$. This must be equal to the population firing rate. Thus equation 2.3 is *still* satisfied, and the population firing rate is again a perfect time replica of the stimulus.

This stochastic model carries one new feature: it tends to randomize the firing times of individual encoders, with respect to one another. Thus, even though the whole population were started synchronously, they would still tend to the time-independent distribution of Fig. 2, and toward the perfect-stimulus-replication behavior.[3]

In sections 5, 6, and 7 we will see that modification of the "simple integrate" law of equation 2.1—even slight modification—will lead to a population of encoders which tend to synchronize among themselves. It is reasonable to suspect that the stochastic feature might offset this tendency. This suggestion will be explored in section 8.

3. INSTANTANEOUS RATE OF A SINGLE UNIT

For a large and homogeneous population of neurons, the "instantaneous rate of a typical single unit" is a well-defined variable at all times, determined by the present state and past history of the entire population. We simply inspect the population for a neuron which currently is firing. The time since its last firing is its instantaneous period, the reciprocal of which is its instantaneous rate.

The single unit rate and the population rate are related in a subtle way. Because it is the single unit rate which usually is observed in the laboratory, and because the single unit rate often is more easily deduced from a theoretical

---

[3] The time of the last firing of a long enough sequence of firings of a given encoder will become uncorrelated with the first firing time. Thus the distribution ρ (u) must become time independent in the limit of long times. It is unreasonable that there should be a second time-independent distribution besides that of equation 2.4. The evolution of the distribution may be reduced to a well-known problem by observing that we are dealing with a so-called "renewal process" in the variable u.

model, we will explore this relationship in this section and the next. The integrate-and-fire model furnishes a start.

Equation 2.1 may be integrated at once, and with the threshold condition, leads immediately to

$$C = \int_{t_n}^{t_{n+1}} dt\, s(t) \tag{3.1}$$

where t_n and t_{n+1} are the times of the nth and $(n + 1)$th impulses, respectively. If the stimulus is constant ($s = s_0$) then

$$C = (t_{n+1} - t_n)s_0 \tag{3.2}$$

or

$$f_0 = s_0/C \tag{3.3}$$

where f_0 is the instantaneous rate of the single unit. Thus the single unit rate is in fixed linear proportion to the stimulus. This also must be true approximately if s changes by only a very small fraction of its value between two impulses. To find the degree of error we express $s(t)$ as a Taylor series

$$s(t) = s(t_n) + \dot{s}(t_n)(t - t_n) + \cdots \tag{3.4}$$

and equation 3.1 becomes

$$C = \frac{1}{f}s + \frac{1}{2}\frac{1}{f^2}\dot{s} + \cdots \tag{3.5}$$

where f is the single unit rate and the time t_n is implied. Multiplying equation 3.5 by f/C gives

$$\begin{aligned} f &= \frac{1}{C}s + \frac{1}{2}\,\frac{1}{C\left(\frac{1}{C}s + \frac{1}{2}\frac{1}{Cf}\dot{s} + \cdots\right)}\,\dot{s} + \cdots \\ &\approx \frac{1}{C}s + \frac{1}{2}\frac{\dot{s}}{s} \end{aligned} \tag{3.6}$$

where the last line assumes $s \gg |\dot{s}/f|$. Now according to equation 3.6, f is no longer a perfect copy of s. It is not even a *linear* copy, in the sense that, for example, doubling the stimulus does not double the rate f. Unlike the population rate, the single unit instantaneous rate is *not* a perfect copy of the stimulus.

Next we investigate the frequency response of the single integrate-and-fire unit's instantaneous rate. Qualitatively, our question is: If the stimulus $s(t)$

fluctuates at a given frequency (the driving frequency), how well does the instantaneous rate $f(t)$ follow? According to equation 3.6 we may at once respond: almost perfectly at very low frequencies; but for driving frequencies which are *not* very low compared to the instantaneous rate, no general answer is known, nor is there known any practical general method for seeking the answer. However, if we confine ourselves to a periodic $s(t)$ which consists of a small fluctuation about a steady mean level then there *is* a general method and an answer in simple terms.

The general method—"linear perturbation theory"—comes in two parts. First, express both the input and output variables as a constant plus a small departure. When these variables are substituted into the mathematical relations which connect them, the strategy will be to ignore all expressions which are small compared to these small departures. An easy example (useful below) will illustrate: instantaneous frequency and instantaneous period [called $T(t)$] are connected by the relation

$$f = 1/T. \tag{3.7}$$

Now let

$$T(t) = T_0 + T_1(t), \quad f(t) = f_0 + f_1(t). \tag{3.8}$$

Note that

$$1/(T_0 + T_1) = 1/T_0 - T_1/T_0^2 + T_1^2/T_0^3 - \cdots. \tag{3.9}$$

If we substitute equation 3.8 into equation 3.7, all that survives is

$$f_1 = -T_1/T_0^2, \tag{3.10}$$

since we knew that $f_0 = 1/T_0$ already, and since T_1^2/T_0^3 and all higher terms are small compared to those in equation 3.10. Note that in equation 3.10 f_1 and T_1 are linearly related. This is a general and important result of the linear perturbation method.

The relation which connects stimulus to period is equation 3.1. In that equation t_{n+1} is the time of the present spike discharge, t_n is the time of the last, and $T = t_{n+1} - t_n$ is the period.

As in equation 3.8, let $s = s_0 + s_1$ and $T = T_0 + T_1$. Now an integral may be interpreted as an area. The relation imposed by equation 3.1 is that the stimulus perturbation s_1 causes a change T_1 in the period just such that the area remains at the unchanged value C. Fig. 3 illustrates this. Equation 3.1 demands ***exactly*** that the two shaded areas must be equal. This is *almost* properly expressed by the relation

$$\int_{t_{n+1}-T_0}^{t_{n+1}} dt'\, s_1(t') = -s_0 T_1. \tag{3.11}$$

The error is that the integral in equation 3.11 includes the little rectangle with edges $s_1 T_1$ in Fig. 3. But this area is small compared to those in equation 3.11, and its neglect amounts to the linear perturbation approximation.

The time t_{n+1} in equation 3.11 is a perfectly general time, and might as well be called t. When equation 3.11 is solved for T_1 and the result is put into equation 3.10, we obtain

$$f_1(t) = \frac{1}{T_0^2} \int_{t-T_0}^{t} dt' \frac{s_1(t')}{s_0}, \tag{3.12}$$

which says that f_1 depends linearly on the recent past history of s_1, and in fact is proportional to the running average of the stimulus perturbation over the last T_0 time units.

The second part of the general method for finding the frequency response to a small fluctuation is to assume an explicit periodic function for $s_1(t)$. We might assume $s_1(t) = s_1(0) \cos \omega t$, for example, where ω is 2π times the driving frequency. Since equation 3.12 is a linear relationship between s_1 and f_1, we are justified in choosing instead

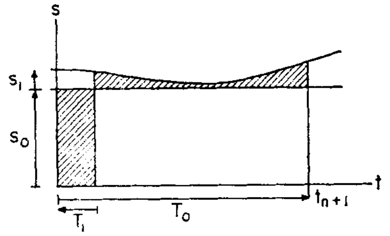

FIGURE 3. Shift T_1 in period due to shift s_1 in stimulus. The two shaded areas must be equal.

$$s_1(t) = s_1(0)e^{i\omega t} \tag{3.13}$$

which simplifies formal manipulations (and, in practiced hands, gives a format closer to one's physical intuition). Substituting equation 3.13 in equation 3.12 leads to a very easy integral, and the result is

$$f_1(t) = \frac{s_1(0)e^{i\omega t}}{T_0^2 s_0} \frac{1 - e^{-i\omega T_0}}{i\omega}. \tag{3.14}$$

Since equation 3.13 reappears in equation 3.14, a bit of rearrangement gives

$$\frac{f_1}{s_1} = \frac{f_0}{s_0} \frac{1 - e^{-i\omega/f_0}}{i\omega/f_0} \equiv \frac{f_0}{s_0} B(\omega/f_0). \tag{3.15}$$

This expression is the frequency response (or so-called "transfer function") for the transduction from s_1 to f_1. We note first that it is a constant, independent of time, and second that it is independent of the amplitude of s_1. Thirdly, it is a complex number, with an amplitude and a phase. Its amplitude is the ratio of the amplitude of the sinusoidal response f_1 to that of the sinusoidal stimulus s_1. Its phase is the phase shift of the crest of the response from the crest of the stimulus. (These facts may be confirmed a bit more laboriously by substituting the cosine form into equation 3.12.) These are all general results of the linear perturbation method. The amplitude and phase of B, over a range of frequencies, are shown in Fig. 4. This sort of frequency response was first recognized in its biological context by Borsellino

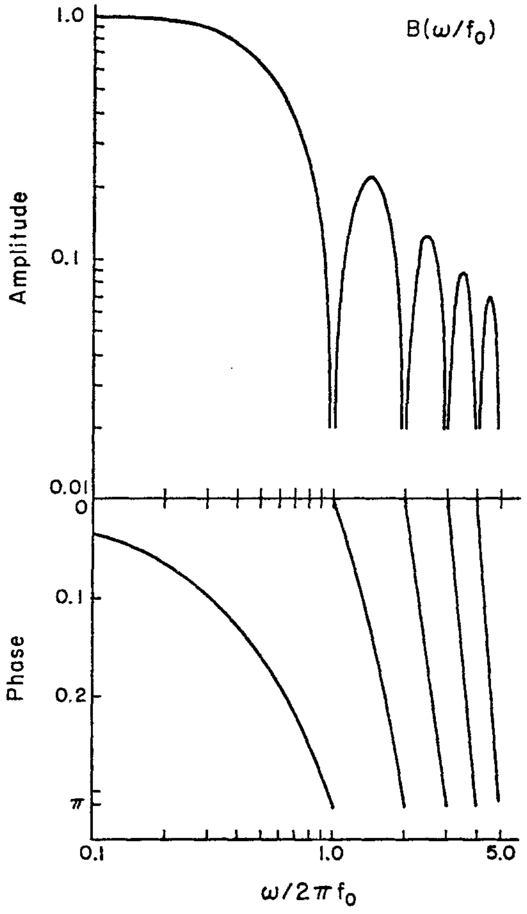

FIGURE 4. The function B, amplitude and phase.

et al. (1965)[4] and by Partridge (1966). It is also the transfer function of a running averager. Three prominent features deserve comment: (*a*) it shows the perfect time-replica feature at very low frequencies, (*b*) a high-frequency cutoff sets in as the driving frequency $\omega/2\pi$ approaches the unperturbed instantaneous rate f_0, and (*c*) it gives a null at any frequency where the sinusoidal input has a running average of zero, as equation 3.12 would predict.

4. THE RELATION BETWEEN UNIT RATE AND POPULATION RATE

Suppose we have a large population of N neurons, all of which are alike, which do not interact, and which encode their impulse trains, from a common stimulus, according to some law that is deterministic rather than probabilistic. We do *not* assume that the deterministic law is the simple integrate-and-fire model. If the encoding law makes any practical sense, then the time at which a neuron fires for the $(n + 1)$th time will be a monotonic (steadily rising) function of the time it fires for the nth time. Since the neurons are identical, this implies that no neuron can straddle two firings of another neuron with two consecutive firings of its own. That is enough to assure that every neuron in the population fires exactly once between the nth and $(n + 1)$th firings of a given neuron. In symbols,

$$N = \int_{t-T(t)}^{t} dt' r(t') \qquad (4.1)$$

where $T(t)$ is the instantaneous period of a single unit, as defined in the first paragraph of section 3, N is the total number of neurons in the population, and $r(t)$ is the population rate as in section 2. Equations 4.1 relates the population rate $r(t)$ to a specified single unit period $T(t)$, without making any assumption (except monotonicity) about the encoding law.

Considering the definition of instantaneous period, it is equally true that equation 3.1 may be written

$$C = \int_{t-T(t)}^{t} dt' s(t'). \qquad (4.2)$$

Now equations 4.1 and 4.2 are identical in form, as N plays the role of C, and r the role of s. The roles of input and output variables have been interchanged, but that is a matter of emphasis, rather than one of mathematics. All the arguments of section 3 apply to equation 4.1. In particular, in the

---

[4] In fact the equipment superimposed such a frequency response upon the biological data. See Poppele and Bowman (1970) for further discussion.

linear perturbation approximation, the frequency response of the population rate to the single unit rate will be

$$\frac{r_1}{f_1} = \frac{r_0}{f_0} \frac{i\omega/f_0}{1 - e^{-i\omega/f_0}}, \tag{4.3}$$

substantially the inverse of equation 3.15 because input and output have interchanged roles.

The most striking feature of the frequency response equation 4.3 is that it becomes infinite at certain frequencies. The denominator vanishes whenever $\omega/f_0 = 2\pi n$ (where n is any nonzero integer). A "resonance" sets in, leading to a huge amplification whenever the driving frequency gets close enough to a multiple of the single-unit unperturbed rate.[5]

For a population of simple integrate-and-fire neurons, what is the frequency response of the population rate to a small periodic fluctuation in the stimulus? We may combine equations 4.3 and 3.15. Thus

$$\frac{r_1}{s_1} = \left\{\frac{r_1}{f_1}\right\}\left\{\frac{f_1}{s_1}\right\} = \frac{r_0}{s_0}. \tag{4.4}$$

This is simply a weakened statement of the perfect time-copy property of a population of simple integrate-and-fire encoders. It could have been derived immediately by applying linear perturbation theory to equation 2.3.

The remarkable feature of equation 4.4 is the perfect cancellation of poles and zeros between the two braced terms. The frequency dependence drops out entirely. This will be in striking contrast to the next section, where a different sort of encoder will be investigated.

If the neurons of equation 4.1 should be simple integrate-and-fire encoders, which satisfy equation 4.2, then a solution of equation 4.1 for $r(t)$ is easily found. Let

$$r(t) = \frac{N}{C} s(t). \tag{4.5}$$

Substitution into equation 4.1 at once yields equation 4.2, which is true by hypothesis. Since $\rho_0 C = N$ (Fig. 1), equation 4.5 is the same as equation 2.3.

Suppose an electronic simple integrate-and-fire circuit is connected to the beam brightener of an oscilloscope. Suppose further that the voltage stimulus,

---

[5] In technical terms, the poles of a frequency response indicate how a system may respond to a vanishingly small stimulus. The poles indicate the "free-running" or undriven behavior of the system. In particular, poles at real ω indicate undamped periodic free responses of the system. In this case the indicated free-running periodic responses are of the sort already mentioned just following equation 2.2. Such periodic free-running responses also may be found in equation 4.1: if T is constant there, r is undetermined to within an additive function which integrates to zero over period T.

a constant voltage plus a sinusoid, also drives the oscilloscope's vertical deflection, and that the horizontal sweep is synchronized to that signal. The experiment should be done at a sweep rate too fast for the eye to follow, and with several cycles of the sinusoid displayed on the scope. If the spike rate is almost identical to the sine frequency, then on the scope face we will see a procession of bright spots voyaging along the sine curve, one to a cycle. The interspot separation will be practically the same when the spots are in the troughs as it is when they are at the crests. But the spots will spend almost all of their time at the crests, and particularly will shun the troughs. The behavior of the single encoder over many cycles shows us what would be the behavior of a population of many encoders over a single cycle.

5. FORGETFUL INTEGRATE-AND-FIRE MODEL

Equation 2.1 is the simplest example of the more general relationship

$$\frac{du}{dt} = F(u, s(t)) \tag{5.1}$$

which might describe the internal dynamics of an encoding neuron. If we interpret u in equation 5.1 as a set of internal variables, and F as a set of functional relationships, the Hodgkin-Huxley equations take this form, with $s(t)$ the input current; the four components of u are the voltage and the three conductance-determining parameters. In designing a neuron encoding model, we should, according to common sense, pick $F(u, s(t))$ in equation 5.1 in such a way that the present value of u depends more strongly on the immediate past history of $s(t)$ than on its more distant past. The Hodgkin-Huxley equations have this property. The simplest example in the form of equation 5.1 with this property is the one-component equation

$$\frac{du}{dt} = -\gamma u + s(t) \tag{5.2}$$

which may be got from the Hodgkin-Huxley equations in a limiting case less drastic than that which yielded equation 2.1. Equation 5.2 carries the feature that the effect of s at time t' upon u at time t will have decremented by a factor of $\exp[-\gamma(t - t')]$.

We complete our encoder model by imposing a firing threshold at the criterion level $u = C$, as before. Equation 5.2 is easily integrated for u, whereupon the threshold condition yields

$$C = \int_{t_n}^{t_{n+1}} dt' e^{-\gamma(t_{n+1}-t')} s(t') \tag{5.3}$$

which should be compared to equation 3.1. Following section 3, if the stimulus is constant ($s = s_0$) integration gives

$$C = +\gamma^{-1}\{1 - \exp[-\gamma(t_{n+1} - t_n)]\}s_0 \tag{5.4}$$

which may be solved for f_0, giving

$$f_0 = -\gamma/\lg[1 - (\gamma C/s_0)]. \tag{5.5}$$

We notice that, as we decrease the stimulus s_0, the firing rate f_0 falls to zero at the finite stimulus level $s_0 = \gamma C$. This could be found at once by putting $u = \text{const}$ into equation 5.2, to find the asymptotic value to which u will rise if no threshold is crossed. We see that u rises asymptotically toward the value s_0/γ, which will be below firing threshold if $s_0 < \gamma C$. To compare equation 5.5 to equation 3.3, we expand equation 5.5 about large s_0 :

$$f_0 = s_0/C - \gamma/2 + \gamma \cdot 0(\gamma C/s_0). \tag{5.6}$$

If $s_0 = 2\gamma C$, the first two terms in equation 5.6 differ from the exact result in equation 5.5 by only about 14%. Thus, except very near threshold, the only effect of forgetfulness on the response to a steady stimulus is that the single unit firing rate is offset by a constant amount $-\gamma/2$.

In order to find the single-unit frequency response, it is convenient to express equation 5.3 as

$$C = \int_{t-T(t)}^{t} dt' e^{-\gamma(t-t')} s(t'). \tag{5.7}$$

The graphical argument of section 3 corresponded to assuming linear perturbations $s = s_0 + s_1(t)$, $T = T_0 + T_1(t)$ in this expression. Substitution into equation 5.7 yields (with $\gamma T_1 \ll 1$)

$$0 = T_1(t)e^{-\gamma T_0}s_0 + \int_{t-T_0}^{t} dt' e^{-\gamma(t-t')} s_1(t') \tag{5.8}$$

in analogy to equation 3.11; the equation analogous to equation 3.12, which follows, is

$$f_1(t) = \frac{e^{\gamma T_0}}{T_0^2} \int_{t-T_0}^{t} dt' e^{-\gamma(t-t')} \frac{s_1(t')}{s_0}. \tag{5.9}$$

Notice that f_1 is proportional to a weighted average over past values of s_1, with weights biased in favor of the most recent past.

The second part of the frequency response calculation proceeds much as in section 3, and yields the transfer function

$$\frac{f_1}{s_1} = \frac{f_0}{s_0} e^{\gamma/f_0} \frac{1 - e^{-(i\omega+\gamma)/f_0}}{(i\omega + \gamma)/f_0}, \tag{5.10}$$

which corresponds to equation 3.15. Equation 5.10 shows two new features. First, if γ is comparable in size to f_0, then even when ω is near zero the perfect copy feature has been lost: f_1/s_1 departs from f_0/s_0 by a factor $(\gamma/f_0)^{-1}(\exp(\gamma/f_0) - 1)$. This reflects the fact that near firing threshold the f_0 vs. s_0 relation in equation 5.5 is nonlinear. Well away from threshold ($\gamma/f_0 \ll 1$) the frequency response in equation 5.10 looks extremely similar to equation 3.15. However, a second slight discrepancy exists, which will prove important: the frequency response does not quite null at the resonance points $\omega = 2\pi n f_0$, and there equation 5.10 becomes

$$\frac{f_1}{s_1} = \frac{f_0}{s_0}\frac{-i}{2\pi n}\frac{\gamma}{f_0}. \tag{5.11}$$

By hypothesis this is small, but it is not zero.

The frequency response of the population rate to a small periodic fluctuation in stimulus may be found as in equation 4.5:

$$\frac{r_1}{s_1} = \left\{\frac{r_1}{f_1}\right\}\left\{\frac{f_1}{s_1}\right\} = \frac{r_0}{s_0}\left(\frac{i\omega}{i\omega + \gamma}\right)\frac{e^{\gamma/f_0} - e^{-i\omega/f_0}}{1 - e^{-i\omega/f_0}}. \tag{5.12}$$

For small γ this simplifies to

$$\frac{r_1}{s_1} = \frac{r_0}{s_0}\left\{1 + \frac{\gamma/f_0}{1 - e^{-i\omega/f_0}}\right\}. \tag{5.13}$$

The additional term, which is not in equation 4.5, is small under most circumstances, but goes to infinity whenever $\omega/2\pi$ approaches a resonant frequency. If ω is near $2\pi n f_0$ we find

$$\frac{r_1}{s_1} = \frac{r_0}{s_0}\left\{1 - \frac{i\gamma}{\omega - 2\pi n f_0}\right\}. \tag{5.14}$$

Thus the population frequency response of the forgetful integrate-and-fire encoder model, well above threshold, is of the flat perfect time-copy type *except* near the resonant frequencies where it is enormously amplified. The approach of equation 5.14 to equation 4.5, as $\gamma \to 0$, is nonuniform: the peak gets narrower but no less tall. The feature of response climbing to infinity, as the frequency approaches resonance, survives no matter how small a finite value of γ we choose.

We close this section with a comment about the general equation 5.1 with which we started. The degree of forgetfulness it exhibits may be built in by rewriting it as

$$\frac{du}{dt} = F(\gamma u, s(t)) \tag{5.15}$$

where γ is the "forgetfulness parameter" in the sense that if $\gamma = 0$ then equation 5.15 reduces to the simple integrate-and-fire model with input stimulus $F(0, s(t))$. If we assume that γ is small and also that $s(t) = s_0 + s_1(t)$ where s_1 is small, then we may expand F in both small quantities and obtain

$$\frac{du}{dt} = a + bs_1(t) + c\gamma u + \cdots \tag{5.16}$$

where all the further terms are second or higher order in smallness. This approximate equation closely resembles equation 5.2, and may be put through the same logical procedures to yield essentially the population frequency response equation 5.13. Thus we see that the feature of resonant amplification is not peculiar to the forgetful integrate-and-fire model of equation 5.2, but rather is a common feature intimately associated with the general property of forgetfulness in a deterministic encoder.

6. PHASE LOCKING IN THE FORGETFUL MODEL

For a simple integrate-and-fire encoder, there is no fixed relationship between the phase of a periodic input and the moment at which the encoder fires an impulse. This is so even if the frequency of the periodic stimulus is identical to the firing rate of the encoder.

In equation 3.1 we may add to the stimulus $s(t)$ any other stimulus $s'(t)$ which integrates to zero between the firing times, and the equation will still be satisfied with the firing times unchanged. If all the t_m of equation 3.1 are evenly spaced, and $s(t)$ is periodic over that spacing, then we may let

$$s'(t) = s(t + \tau) - s(t) \tag{6.1}$$

where τ is arbitrary. The effect of adding equation 6.1 to $s(t)$ is to shift the stimulus pattern by an arbitrary time τ, without shifting the firing times.

For the forgetful integrate-and-fire encoder the situation is altogether different. We may ask under what conditions spike firings will keep in step with a periodic stimulus. Suppose we apply a stimulus which is a constant s_0 plus a sinusoid of fractional amplitude m:

$$s(t) = s_0\{1 + m \operatorname{Re} e^{i(\omega t+\phi)}\}. \tag{6.2}$$

Since an undetermined phase ϕ has been included in the stimulus, we may start the integral in equation 5.3 at $t_n = 0$ without loosing generality.

$$C = \int_0^T dt' e^{-\gamma(T-t')} s_0\{1 + m \operatorname{Re} e^{i(\omega t'+\phi)}\}. \tag{6.3}$$

In this equation we intend to see if we can pick the phase ϕ of the driving signal in such a way that the firing period T will be the reciprocal of the driving frequency $\omega/2\pi$. The reason for writing $\cos(\omega t + \phi)$ in that peculiar form in equation 6.2 is that later operations are facilitated and the integral becomes easy:

$$C = e^{-\gamma T} s_0 \left\{ \frac{e^{\gamma T} - 1}{\gamma} + m \operatorname{Re} \left(e^{i\phi} \frac{e^{(\gamma + i\omega)T} - 1}{\gamma + i\omega} \right) \right\}. \tag{6.4}$$

It will be convenient also to write

$$\frac{1}{\gamma + i\omega} = \frac{1}{\sqrt{\gamma^2 + \omega^2}} e^{i\beta} \tag{6.5}$$

where β is the known angle

$$\beta = -\arctan(\omega/\gamma). \tag{6.6}$$

Now we will impose the condition that $s(t)$ is periodic over the firing period T:

$$e^{i\omega T} = 1. \tag{6.7}$$

Under this assumption, can a fixed firing phase ϕ be found such that equation 6.3 or 6.4 is satisfied? Equation 6.4 reduces to

$$C = s_0 \left\{ \frac{1 - e^{-\gamma T}}{\gamma} + \frac{m}{\sqrt{\gamma^2 + \omega^2}} (1 - e^{-\gamma T}) \operatorname{Re} e^{i(\phi + \beta)} \right\}. \tag{6.8}$$

The only unknown in this expression is the phase ϕ, and that appears only in one place. We rearrange equation 6.8 to isolate the unknown, and find

$$\left\{ \frac{C/s_0}{1 - e^{-\gamma T}} - \frac{1}{\gamma} \right\} \frac{\sqrt{\gamma^2 + \omega^2}}{m} = \cos(\phi + \beta). \tag{6.9}$$

Since the cosine is an even function, the extremes of which lie at ± 1, there will be no solution to equation 6.9 for the phase ϕ if the left-hand side is greater than unity in absolute value, but if it is less there will be two solutions.

An easy illustrative example is the case in which the term in braces in equation 6.9 vanishes. According to equation 5.4, that is the case where T is the free-running firing period in the absence of modulation. In this case equation 6.9 is evidently solved by

$$\phi = -\beta \pm \pi/2. \tag{6.10}$$

Since the left-hand side of equation 6.9 need only lie between ± 1, a period somewhat off from the free-running period also will permit a solution of equation 6.9, but a period that is off badly will only solve if the modulation m is made large enough.

If there is only slight memory loss over the firing period ($\gamma T \ll 1$), equation 6.9 becomes

$$\left\{\frac{Cs_0}{T} - 1\right\}\frac{\omega}{\gamma m} = \cos(\phi + \beta). \tag{6.11}$$

The braced term now vanishes if T is the free-running period of the simple integrate-and-fire encoder. We must stick close to that period if equation 6.11 is to have solutions. And, so to speak, twice as close if the encoder becomes only half as forgetful—or we must double the modulation.

In deriving equation 6.11, we have used the implication of equation 6.7 that

$$\omega = 2\pi n/T \tag{6.12}$$

for some integer n. Since ω is a property of the cause, and T a property of the effect, it is of course ω which determines T through equation 6.12, provided equation 6.9 has a solution. The fact that equation 6.9 can be solved for a range of T about the free-running period shows that the firing rate can be "pulled" away from its free-running value by a driving frequency that lies close to a multiple of the free-running firing rate.

What happens if equation 6.9 has no solution? If the modulated part of the stimulus undergoes, for example, two periods of oscillation in a time closely similar to three periods of the free-running encoder, then "frequency pulling" may still occur if the modulation depth is sufficient. More generally, if the sum of k free-running periods falls close to n periods of the driving frequency, a repeating time pattern of k impulses over every n driving cycles may be established. This behavior was noticed and has been treated in detail by Rescigno et al. (1970). The general "n/k" case is very much more difficult in detail[6] than is the "$n/1$" case which led to equation 6.9 above, although very similar conclusions are reached.

If equation 6.9 has any solutions at all, typically it will have two solutions. Both give points in the stimulus cycle at which firings of the encoder will continue in step with the stimulus. These are called "fixed points" because firings continue to occur at them, cycle after cycle. It is not difficult to follow what happens if the encoder is initially fired at a point in the cycle slightly off a solution of equation 6.9 (see Rescigno et al. 1970). The conclusion is that one solution of equation 6.9 yields a stable fixed point, and the other an unstable one. The fixed point on the rising part of the stimulus is stable and

---

[6] Rescigno et al. (1970) is the definitive study of the forgetful model stimulated by a finite sinusoid. The final bit of section 5 in that work must be approached with caution, however. It is unclear whether the final inequality implies necessity or sufficiency, and a result seems to emerge which is in discord with Denjoy's general result (which we meet in section 7 of this paper) concerning "continuity of the turning angle" which claims that "n/k" may take on irrational values.

the one on the falling part is not. An encoder which initially is fired at an arbitrary point in the stimulus cycle will, over a sequence of subsequent firings, choose a sequence of points in the cycle which converge to the stable phase solution of equation 6.9.[7] Thus the stimulus tends to "lock" the firing of the encoder to a fixed phase of the stimulus's own rhythm.

The consequence for a population of forgetful integrate-and-fire encoders is evident and dramatic. The entire population will fall into step with the stimulus, at the stable phase-lock point given by equation 6.9. We saw in the last section that an infinitesimal periodic modulation of the stimulus leads, at the resonant frequencies, to an indefinitely large population response. For a *finite* stimulus modulation the same sort of thing happens not only at the resonant frequencies, but in the whole neighborhoods around those frequencies—the frequency-pulling range—over which equation 6.9 has a solution.

If we view the population of encoders as a transducer the output of which is the population firing rate, this is a very serious matter. The output gives information only on the frequency of the input, plus the fact that the modulation was strong enough for phase locking to occur at that frequency. One or a few encoders could deliver as much information.[8]

7. GENERAL THEORY OF ENCODER PHASE LOCKING[1]

The forgetful integrate-and-fire and the simple integrate-and-fire models show a striking contrast in one feature: in the one model the population of encoders tend to fall into step and fire simultaneously; in the other model they do not. These contrasting behaviors are not specific to the two models we have chosen to analyze in detail. Indeed these two models may be regarded as prototypes of two distinct classes of encoders. This section will show that in fact no further classes exist, so long as we confine ourselves to encoders which are deterministic and depend only on input since the last impulse. Thus the results of the previous sections should be applicable, except for quantitative details, to a wide variety of neural encoders.

Topological methods of a very general nature, discussed in the present section, lead to two conclusions: (a') the most general deterministic impulse encoder which does *not* phase lock to any periodic signal is equivalent to an

---

[7] In the near neighborhood of either the stable fixed point or the adjacent unstable fixed point, the successor to a given impulse steps toward the phase of the stable point and away from the phase of the unstable point. The same must be true over the entire span of phase in between these neighborhoods: the phase of the successor is a continuous function of the phase of the given impulse; the signature of the phase difference can only reverse by passing through zero, which would define another fixed point between the adjacent fixed points. This topological argument illustrates the power of the methods which will be discussed in the next section.

[8] In particular applications phase locking should be advantageous; for example, in the sound direction sense at low frequencies, which depends on the accurate measurement of phase differences between the two ears.

arbitrary continuous transducer followed by a simple integrate-and-fire encoder; and (*b′*) an encoder which *does* phase lock to some specific periodic stimulus will also phase lock to distinct periodic stimuli which are sufficiently similar, and in particular it will phase lock over a finite span of frequencies.

At present the general theory is not in a definitive final form. Two more conclusions strongly suggested but as yet unproven in general are: (*c′*) no forgetful encoder (in the sense of section 5, paragraph 1) can mimic an encoding scheme the last stage of which is an infinite-memory encoder, hence, by conclusion (*a′*) any forgetful encoder must show phase-locking behavior; and (*d′*) any pulse-encoding scheme which shows phase locking in response to some periodic stimulus will also show population synchronization in response to a wide class of reasonable aperiodic stimuli. The tentative conclusions (*c′*) and (*d′*) stand up in explicit cases examined to date. If they hold universally, then they imply that a homogeneous population of deterministic encoders designed on *any* forgetful encoding scheme must eventually fall into a synchronized condition. The remainder of this section outlines the path of reasoning that leads to conclusions (*a′*) and (*b′*).

The general theory of deterministic encoder response to periodic stimuli corresponds to the topological theory of the continuous one-to-one mappings of the circumference of the circle onto itself. In particular the classification of encoders is closely related to the classification of such mappings. Suppose a periodic stimulus $s(t)$ (not necessarily sinusoidal) has a period $T = 2\pi/\omega$. The variable

$$x = t/T \tag{7.1}$$

ranges from zero to unity over one cycle of the stimulus. We can imagine the ascending values of x as points arrayed around the circumference of a circle. Equation 3.1 or 5.3 is an implicit relation which determines x_{n+1} once x_n has been specified. Both are examples of the general form

$$x_{n+1} = \phi(x_n) \tag{7.2}$$

where x_{n+1} is some new value which we may place between zero and unity by adopting the obvious cyclic convention. Equation 7.2 expresses a "mapping" in that any point x_n on the circle is mapped uniquely onto a new point x_{n+1}. The position on the circle of the kth successor to x_n will be given by

$$x_{n+k} = \phi(\phi(\cdots\phi(x_n)\cdots)) \equiv \phi^k(x_n). \tag{7.3}$$

An evident extension of this notation is, for example,

$$x_n = \phi^{-1}(x_{n+1}), \tag{7.4}$$

and if two mappings are ϕ, ψ, we will sometimes write

$$x' = \chi(x) = \psi(\phi(x)) \quad \text{as} \quad \chi = \psi * \phi. \tag{7.5}$$

The successive application of two one-to-one mappings is a one-to-one mapping; each mapping has an inverse, and there is an identity mapping (do nothing). Technically the mappings form a "group" and their natural classification is in terms of "equivalence classes" in the group theoretic sense.

There is a second way (besides that of equation 7.2) of looking upon the functional relation

$$x' = \psi(x) \tag{7.6}$$

on the circle. It may be looked upon as a reexpression of the same point x in terms of a new coordinate. For example, x might be distance as measured around a circle of unit circumference by an accurate tape measure, and x' that distance as measured by an inaccurate tape measure which is stretched and shrunk over different parts of its range between zero and unity. Then the transformation ψ^{-1} is the correction table to be used with the inaccurate measure.

If x gets changed, as in equation 7.2, by a mapping $\phi(x)$, how does x' get changed? How does the point-to-point mapping, ϕ, look in the primed coordinate system? What is, say, the corresponding $\phi'(x')$? Answer: first change x' to x with the coordinate change ψ^{-1}, then move x to $\phi(x)$ with ϕ, then change the new x back to the new x' with the coordinate change ψ, whence

$$\phi' = \psi * \phi * \psi^{-1}. \tag{7.7}$$

Any two mappings related as ϕ and ϕ' above are said to belong to the same equivalence class. They are, so to speak, the same mapping expressed in terms of alternative coordinates.

To illustrate we look at a particular important class: the "equivalence class of rigid rotations." In equation 7.7 above let ϕ be, in particular,

$$x_X \equiv \phi(x) = x + X \tag{7.8}$$

so that each point x is advanced around the circle rigidly by a constant increment X. Equation 7.7 becomes

$$x'_X = \phi'(x') = \psi(\phi(\psi^{-1}(x'))) = \psi(\psi^{-1}(x') + X) = \psi(x + X) \tag{7.9}$$

which is the generic form of the equivalence class of rigid rotations through the angle X.

There is a remarkable implicit restatement of equation 7.9. Equation 7.8 is trivially rearranged to

$$X = \int_x^{x_X} dx. \tag{7.10}$$

In equation 7.6 let $\psi = \sigma^{-1}$. Then we have $dx/dx' = \dot{\sigma}(x')$ where the dot on $\dot{\sigma}$ stands for differentiation. In terms of the x' coordinate equation 7.10 becomes

$$X = \int_{x'}^{x'_X} dx' \dot{\sigma}(x'), \tag{7.11}$$

the implicit restatement of equation 7.9. But if $\dot{\sigma}$ is regarded as a stimulus this is exactly the equation that determines the phase of the successor impulse x'_X of a simple integrate-and-fire encoder which last fired at x'. The mappings obtained by simple integrate-and-fire encoders, with all possible periodic inputs ($\dot{\sigma}$) and threshold levels (X), are the same as the equivalence classes of the rigid rotations through the various rotation angles X.

There are also equivalence classes distinct from those of the rigid rotations. Equation 5.3, for example, led to equation 6.9, which was solved for the "fixed points" of the mapping; that is, those phase points in the stimulus cycle such that successive impulses appeared at unchanged phase, or in the present language, those points left unmoved by the mapping of equation 7.2. The existence of a fixed point is a co-called "topological property"—it is independent of changes in coordinate system such as equation 7.7. Hence the whole equivalence class will have two fixed points, as surely as any equivalence class of rigid rotations has no fixed points at all (except if X is an integer).

In passing we note that topological considerations demand that if there is a stable fixed point then there must also be an unstable fixed point. If there is a fixed point which the transformation makes other points step *toward* from both sides, somewhere on the circle there must be a fixed point with the opposite property.

Classes also exist which are distinct from rotations and have no fixed points. Consider for example the situation in which equation 5.3 permits phase locking but with two impulses per stimulus cycle. The transformation ϕ has no fixed points, but ϕ^2 has *four* isolated fixed points, as either stable firing position in the cycle will repeat after two firings (see Fig. 5). More generally, an encoding situation which allows a stable time pattern of k spikes to first repeat after n stimulus cycles will yield a mapping whose kth iterate has $2k$ isolated fixed points.

These simple facts will be used in conjunction with a set of deeper results, mostly due to Denjoy (1932; see also Coddington and Levinson, 1955, chapter 17; and Moser, 1968,[9] pp. 41–77), which we cite without proof.

(*a*) Every mapping $\phi(x)$ has associated with it a number (called the "turning angle") defined by

$$\alpha(\phi) = \lim_{n \to \infty} \frac{1}{n} \phi^n(x) \tag{7.12}$$

---

[9] Moser, J. 1968. Notes On Dynamical Systems. Courant Institute, New York University (unpublished).

which is finite and independent of x. (For purposes of equation 7.12 we do not impose the cyclic convention mentioned at equation 7.2, but alternatively "unroll" the circle along the infinite line.) Given that the limit of equation 7.12 exists, it is easy to see that all members of an equivalence class have the same value of α. It is also instructive to substitute equation 7.8 into equation 7.12 and to calculate directly that the turning angle of a rigid rotation is indeed X.

(*b*) If the mapping ϕ depends continuously on a parameter (in the encoder case the circular driving frequency ω of the stimulus will do) then the turning angle $\alpha(\phi)$ also depends continuously on that parameter.

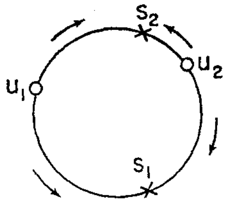

FIGURE 5 *a*

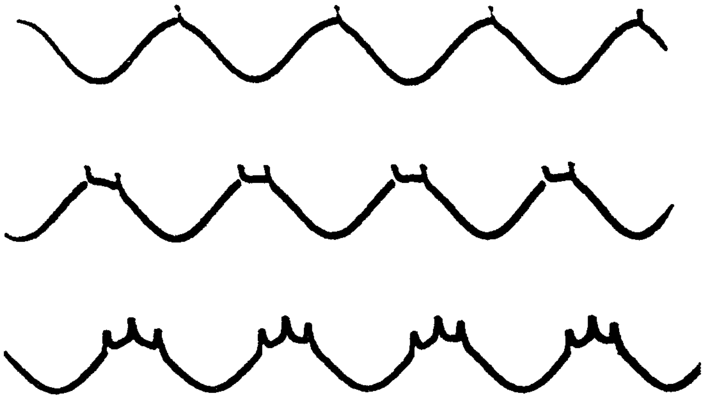

FIGURE 5 *b*

FIGURE 5 *a*. Circle showing two stable fixed points (s_1, s_2) and two unstable fixed points (u_1, u_2) of the mapping ϕ^2.

FIGURE 5 *b*. Phase locking in a sensory neuron (*Limulus* eccentric cell) in response to intracellularly injected current. Top frame: ϕ has one stable fixed point. Middle frame: ϕ^2 has two stable fixed points (as in Fig. 5 *a*). Bottom frame: ϕ^3 has three stable fixed points.

(*c*) The turning angle $\alpha(\phi)$ is either an irrational or a rational number. If α is irrational then ϕ belongs to the equivalence class of rigid rotations with turning angle α.

(*d*) If $\alpha(\phi)$ is rational, say n/k, there are two subcases: either every point x of ϕ^k (x) is a fixed point, in which case $\phi(x)$ again belongs to an equivalence class of rigid rotations, or

(*e*) $\phi^k(x)$ has a discrete set of fixed points. In this case a small finite change of a parameter (see [*b*] above) in at least one direction will leave $\alpha(\phi)$ unchanged.[10]

The deep and difficult statement of the lot is (*c*), which asserts that we did not overlook any additional kinds of equivalence classes in our earlier enumeration.

---

[10] We bar one exceptional case which is unimportant in the application to encoders.

An encoder which never exhibits contingency (*e*) always corresponds to a mapping in the equivalence class of some rotation, and hence to a simple integrate-and-fire encoder receiving some stimulus with the proper period. This substantiates conclusion (a') at the beginning of the section.

The contingency (*e*) is the phase-locking situation, as in Fig. 5, for example. A small enough parameter change simply shifts the positions of the fixed points of ϕ^n, and may be regarded as a coordinate transformation. Hence the equivalence class remains unchanged, and so does the turning angle. The phase-lock condition will likewise persist.[11] Thus any encoder which will phase lock at all will do so over a finite range of frequency (or amplitude, etc.—this is conclusion [b'] at the beginning of this section). If we choose a frequency (and amplitude, etc.) at random, we stand a finite chance to draw a phase-locked condition.

We close with a remark concerning tentative conclusions (c') and (d') made at the beginning of this section: suppose two identical encoders, which share the same input signal, fire initially at times that are only slightly separated. Let us follow these encoders through a large number of firings. If the encoders are forgetful in the sense we have used above, then their latest firing times will be more strongly influenced by their very similar recent past histories than by their different initial conditions. Hence we anticipate that their firing times will draw together as their total number of firings increases. We expect them to "fall into step." This heuristic argument explicitly demands the property of forgetfulness, and nowhere asserts that the input signal is periodic.

8. STOCHASTIC ENCODERS

We have seen that one indication of the synchronization phenomenon in a population of deterministic encoders is the resonant amplification of an infinitesimal periodic fluctuation in the stimulus, as shown in equation 5.14. We suggested in sections 1 and 2 that the inclusion of a random process in each encoder should tend to break up this synchronization.[12] In the present section we will verify our suggestion to the extent of showing quantitatively how fluctuations in the firing rate which are nondeterministic, or stochastic, suppress the infinite resonant peaks in the population frequency response.

The frequency response of the population firing rate we will determine in two steps. Following the development for the deterministic case in section 4, first we will derive a relation between the population rate and a complete specification of the firing periods of the individual encoders. The second step

---

[11] We note that $\psi * (\phi^n) * \psi^{-1} = (\psi * \phi * \psi^{-1})^n$ where ψ corresponds to the parameter change. The transformation $\psi * \phi * \psi^{-1}$ has the same turning angle as ϕ does (of course ϕ^n has fixed points and hence turning angle zero). The fact that a ψ may be found to represent the parameter change is called Pliss's theorem and is discussed by Moser.[9]

[12] This idea has been advanced by Stein (1970), and Stein and French (1969).

will be to derive the firing period information from a stochastic model of an individual encoder.

Let us generalize the monotonicity postulate, with which we started section 4, in the following way: we postulate a homogeneous population of stochastic encoders which are such that we *expect* each to fire once between any two consecutive firings of any specified member of the population. The expected total number of firings between two firings of one member is N, the number of encoders in the population. Let $n(T, t)$ be the number of firings between the times $t - T$ and t, and let $Q(T, t)$ be the probability density that an encoder which fires at t also had its last firing at $t - T$. Then our postulate states that

$$N = \int_0^\infty dTQ(T, t)n(T, t). \qquad (8.1)$$

Since the number of firings in the span T is related to the population firing rate $r(t)$ by

$$n(T, t) = \int_{t-T}^{t} dt'r(t'), \qquad (8.2)$$

equation 8.1 becomes

$$N = \int_0^\infty dTQ(T, t) \int_{t-T}^{t} dt'r(t') \qquad (8.3)$$

which is the relationship that determines the population firing rate $r(t)$ from the specified encoder period distribution $Q(T, t)$. It is the stochastic analogue of equation 4.1. In the deterministic limit

$$Q(T, t) = \delta(T - T_s(t)), \qquad T_s(t) \text{ a specified function,} \qquad (8.4)$$

equation 4.1 is recovered. Or if we assume $r = r_0$ is constant, and $Q = Q_0(T)$, equation 8.3 gives

$$N = r_0\bar{T}_0 \qquad (8.5)$$

where

$$\bar{T}_0 = \int_0^\infty dTQ_0(T)T \qquad (8.6)$$

is the mean firing period. Equation 8.5 gives the steady rate r_0 in terms of only the first moment $\bar{T}_0$ of $Q_0(T)$.

We undertake a perturbation analysis of equation 8.3, and assume

$$r(t) = r_0 + r_1(t), \qquad Q(T, t) = Q_0(T) + Q_1(T, t) \qquad (8.7)$$

which gives

$$0 = \int_0^\infty dTQ_0(T) \int_{t-T}^t dt' r_1(t') + \int_0^\infty dTQ_1(T, t) \int_{t-T}^t dt' r_0 \qquad (8.8)$$

whence

$$\int_0^\infty dTQ_0(T) \int_{t-T}^t dt' r_1(t') = -r_0 T_1(t) \qquad (8.9)$$

where the perturbation in mean firing period $T_1(t)$ has a definition analogous to equation 8.6. Again, the perturbation in the population rate depends only on the first moment of the perturbation in the period distribution.

To find the frequency response we assume

$$T_1(t) = T_1(0)e^{i\omega t}, \qquad r_1(t) = r_1(0)e^{i\omega t} \qquad (8.10)$$

and substitution into equation 8.9 gives, with one easy integration,

$$r_1 \int_0^\infty dTQ_0(T) \frac{1 - e^{-i\omega T}}{i\omega} = -r_0 T_1. \qquad (8.11)$$

Since $Q_0(T)$ is a probability density, and integrates to unity, equation 8.11 in turn at once leads to

$$\frac{r_1}{T_1} = -r_0 \frac{i\omega}{1 - \tilde{Q}_0(i\omega)}, \qquad (8.12)$$

where

$$\tilde{Q}_0(i\omega) = \int_0^\infty dTQ_0(T)e^{-i\omega T} \qquad (8.13)$$

is the average value of $\exp(-i\omega T)$ over T, or in probability terminology, the "characteristic function" of Q_0.

Equation 8.12, the frequency response of r_1 to T_1, is the main result of this final theoretical section. As the resonant poles of the deterministic models first arose from the denominator of equation 4.3, we compare the denominator of equation 8.12 to that expression. We notice that a term of the form $\exp(-i\omega T)$ has been replaced by its average over a collection of periods T determined by chance. The difference is very important: although $\exp(-i\omega T)$ has unit length on the complex plane and touches the unit circle, the average of $\exp(-i\omega T)$ over different values of T, its "center of gravity" $\tilde{Q}_0(i\omega)$, must fall *within* the unit circle. Hence the denominator of equation 8.12 cannot vanish. The stochastic feature has taken care of the infinite resonance problem.

Finally, we analyze a specific model which combines the forgetfulness feature of section 5 with the stochastic threshold feature with which we concluded section 2. We assume an internal variable u related to the input stimulus by equation 5.2, and a distribution of firing thresholds C characterized by a probability density $P'(C)$ where P' is the derivative of the probability P that was shown in Fig. 2. As in equation 5.7, the firing threshold C and the period T are related by

$$C = \int_{t-T}^{t} dt'\, e^{-\gamma(t-t')} s(t') \tag{8.14}$$

and the probability density $Q(T)$ of the random variable T may be found from $P'(C)$ and from equation 8.14 by

$$Q(T) = P'(C(T)) \frac{dC(T)}{dT} \tag{8.15}$$

according to ordinary probability theory. In equation 8.15 we regard as fixed parameters the time t and the whole past stimulus history $s(t')$ up to that time. To specify our model fully, we may specify either $P'(C)$ or $Q(T)$, as they are related by equation 8.15. For finding the frequency response it is convenient to let $s = s_0$ in equation 8.14 and specify the unperturbed period distribution $Q_0(T_0)$. The reason is that the random variable T_1 is given most conveniently as a function of the random variable T_0. Using $T_0 = 1/f_0$, $T_1 = -T_0^2 f_1$ (equation 3.10), and equation 5.10, we see that

$$T_1 = -\frac{e^{\gamma T_0} - e^{-i\omega T_0}}{i\omega + \gamma} \frac{s_1}{s_0}. \tag{8.16}$$

The mean value $\bar{T}_1$ is thus

$$\bar{T}_1 = \int_0^\infty dT_0\, Q_0(T_0) T_1(T_0) = -\frac{\tilde{Q}_0(-\gamma) - \tilde{Q}_0(i\omega)}{i\omega + \gamma} \frac{s_1}{s_0} \tag{8.17}$$

where $\tilde{Q}_0(-\gamma)$ follows the definition of equation 8.13. Using both equations 8.12 and 8.17, we find that the frequency response of the population rate to the stimulus is

$$\frac{r_1}{s_1} = \frac{r_0}{s_0} \left(\frac{i\omega}{i\omega + \gamma} \right) \frac{\tilde{Q}_0(-\gamma) - \tilde{Q}_0(i\omega)}{1 - \tilde{Q}_0(i\omega)}, \tag{8.18}$$

which should be compared to the deterministic result of equation 5.12.

A small stochastic effect corresponds to a period distribution $Q_0(T_0)$ which is peaked sharply around the mean period $\bar{T}_0$. We may write

$$\begin{aligned} e^{-i\omega T_0} &= e^{-i\omega \bar{T}_0} e^{-i\omega(T_0 - \bar{T}_0)} \\ &= e^{-i\omega \bar{T}_0} \left\{ 1 - i\omega(T_0 - \bar{T}_0) - \frac{\omega^2}{2!}(T_0 - \bar{T}_0)^2 + \cdots \right\} \end{aligned} \tag{8.19}$$

and substitution into equation 8.13 yields

$$\tilde{Q}_0(i\omega) \approx e^{-i\omega\bar{T}_0}\left(1 - \frac{\omega^2}{2}\tau^2\right) \tag{8.20}$$

where

$$\tau = \sqrt{\int_0^\infty dT_0(T_0 - T_0)^2 Q_0(T_0)} \tag{8.21}$$

measures the distribution's half-width. There is an analogous result for $\tilde{Q}_0(-\gamma)$.

What is the effect of small stochastic fluctuations upon the frequency response near resonance in a population of slightly forgetful encoders? If we assume both γ and τ are small in equation 8.18, we find

$$\frac{r_1}{s_1} = \frac{r_0}{s_0}\left\{1 - \frac{i\gamma}{(\omega - 2\pi n f_0) - i\frac{f_0}{2}(2\pi n f_0 \tau)^2}\right\} \tag{8.22}$$

by the same approximation that led to equation 5.14. (We have set $f_0 = 1/T_0$.) As we anticipated, the response at resonance is finite, and equation 8.22 converges uniformly to the early perfect time-copy result of equation 4.4 as γ approaches zero. This uniformity is in contrast to the nonuniform convergence found at equation 5.14. The size of the response at resonance is

$$\frac{r_1}{s_1} = \frac{r_0}{s_0}\left\{1 + \frac{2}{(2\pi n)^2}\frac{\gamma/f_0}{(\tau f_0)^2}\right\} \tag{8.23}$$

and represents a contest between deterministic forgetfulness and irregular firing. We note that the *square* of the small number τf_0 appears in the denominator of equation 8.25, so that in this limit a relatively substantial stochastic spread in firing periods is necessary to control a relatively much smaller degree of forgetfulness, if the perfect time-copy property is to be approximated at resonance. As a rough example consider the case where the internal memory variable u relaxes 10% between typical spikes ($\gamma/f_0 = 0.1$) and the spike periods have a root mean square scatter of 10% about their mean ($\tau f_0 = 0.1$). At the fundamental resonance ($n = 1$), even though the coefficient $2/(2\pi)^2 \approx 0.05$ is small, the resonant response is about 1.5 times the response well away from resonance.

We remark that even though equation 8.22 was derived from a specific model, that result is model independent to within a multiplicative scale factor on γ; the effect of a forgetfulness parameter should first appear through γ to the first power, and the denominator in equation 8.22 followed from equation 8.12 which was model independent.

The half-width τ (equation 8.21) at most may be made equal to the mean $\bar{T}_0$, and this happens when $Q_0(T_0)$ is the Poisson distribution

$$Q_0(T_0) = f_0 e^{-f_0 T_0} \tag{8.24}$$

which has the property that successive firing times are completely uncorrelated. In this case it is easy to evaluate the frequency response of equation 8.18 exactly, and the result is

$$\frac{r_1}{s_1} = \frac{r_0}{s_0} \frac{1}{1 - \gamma f_0}, \tag{8.25}$$

which is independent of frequency. In this case the perfect time-copy property is actually achieved for the variation in any stimulus which departs only slightly from its mean.

The frequency response of equation 8.18 also may be calculated exactly for the general "gamma" distribution

$$Q_0(T_0) = \frac{((n + 1)f_0)^{n+1}}{n!} T_0^{\,n} \, e^{-((n+1)f_0)T_0} \tag{8.26}$$

for which the Poisson and deterministic cases are opposite limits $n = 0$ and $n \to \infty$. We find $n + 1 = (f_0\tau)^{-2}$, and the characteristic function is

$$\tilde{Q}_0(i\omega) = \left(1 + i(f_0\tau)^2 \frac{\omega}{f_0}\right)^{-1/(f_0\tau)^2} \tag{8.27}$$

Fig. 6 gives examples of equation 8.18 which fall between the limiting cases.[13] We see for a very forgetful encoder that an rms stochastic scatter comparable to the forgetfulness coefficient suppresses the resonances very effectively. In the case of $\gamma/f_0 = \tau f_0 = 0.1$, the exact expression of equation 8.27 yields a response ratio of 1.51 between the first resonant peak and zero frequency, as compared to 1.5 calculated from the approximate equation 8.22.

9. CONCLUDING REMARKS

In this investigation we have paid particular attention to the population firing rate of a collection of neurons. Our motivation has been that this rate is essentially what is seen by a postsynaptic neuron. We have confined our consideration to "very large" neuron populations. In a practical sense "very

---

[13] The theoretical results of Fig. 6 may be compared with the hardware analogue results of Stein (1970), and Stein and French (1969).

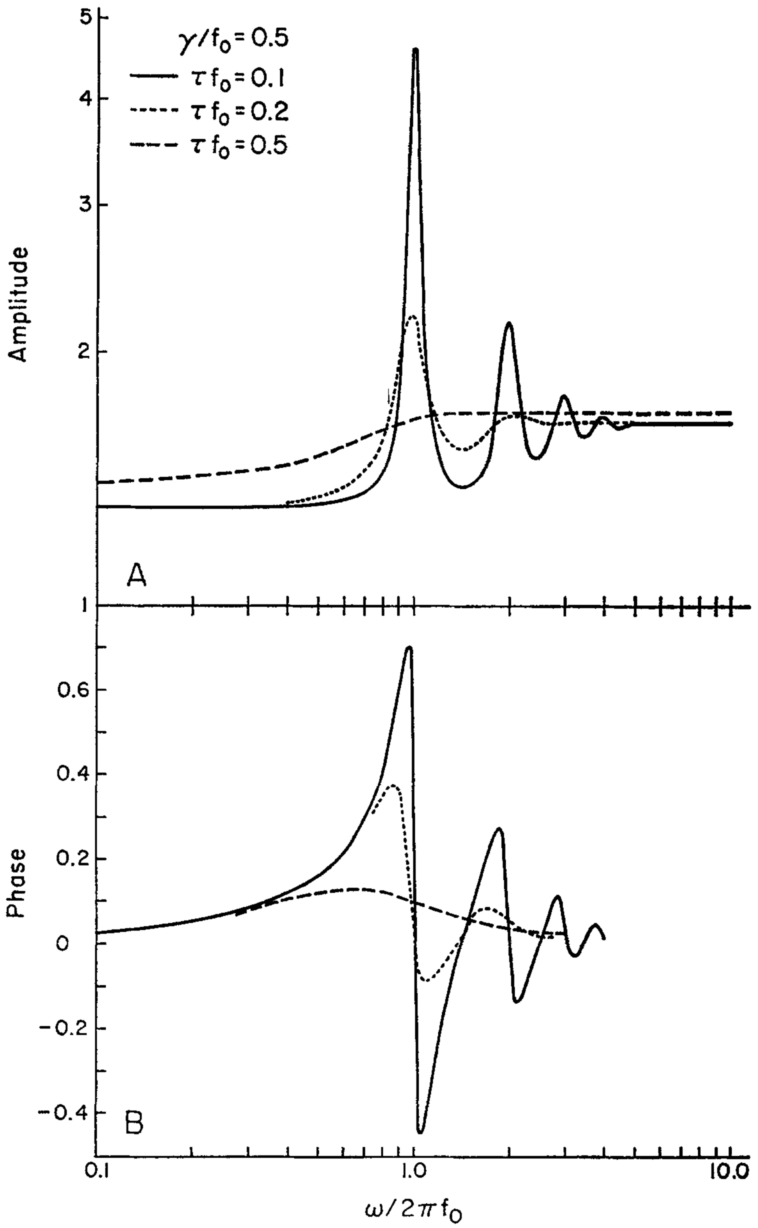

FIGURE 6. Amplitude and phase of population rate frequency response, showing contending effects of forgetfulness and stochastic scatter. Phase is given in radians.

large" means that the population firing rate must exceed the intracellular voltage resolution time of the postsynaptic neuron. The spike encoding schemes we have considered all involve only the stimulus history since the previous spike; this assumption is reasonable for a neuron whose intracellular voltage resolution time is limited by its electrical characteristic time rather than by the decay time of a chemical mediator.

We have reached the following conclusions.

(*a*) A simple encoder model, which is a reasonable idealization of known electrophysiology, yields a population firing rate which is a perfect replica of the input stimulus.

(*b*) A population of noise-free encoders which depart even slightly from the simple encoder model show a tendency to fall into step, and eventually yield a bursting type of population firing which yields a very much degraded copy of the input stimulus.

(*c*) The presence of noise in the encoders counteracts the tendency to synchronize. A slight noise level will retrieve a faithful population response for encoders which depart slightly from the simple model. Large stochastic fluctuations will do the same for a population which departs substantially from the simple model.

In developing these conclusions we have noted that there is a subtle quantitative relation between the firing rate of a single unit and the firing rate of the population from which it is drawn. This relation must be taken into account when the behavior of a population is deduced from observations made on a single cell.

I owe a particular debt to Richard Poppele and Richard Stein, and to Jun-ichi Toyoda who set me on the track of the material in section 2. Others whose thought helped are M. Arbib, E. Bayly, F. A. Dodge, Jr., David Lange, Henry McKean, Jurgen Moser, Richard L. Purple, Morris Schreiber, and Charles F. Stevens.

This research was supported in part by research grants (EY 00188) from the National Eye Institute and (GM1789) from the National Science Foundation.

Received for publication 31 December 1970.

REFERENCES

AITKIN, L. M., D. J. ANDERSON, and J. E. BRUGGE. 1970. Tonotopic organization and discharge characteristics of single neurons in nuclei of the lateral lemniscus of the cat. *J. Neurophysiol.* **32**:421.

BORSELLINO, A., R. E. POPPELE, and C. A. TERZUOLO. 1965. Transfer functions of the slowly adapting stretch receptor organ of Crustacea. *Cold Spring Harbor Symp. Quant. Biol.* **30**:581.

BROWN, M. C., I. ENGBERG, and P. B. C. MATTHEWS. 1967. The relative sensitivity to vibration of muscle receptors of the cat. *J. Physiol.* (*London*). **192**:773.

BRUGGE, J. F., D. J. ANDERSON, and L. M. AITKIN. 1970. Responses of neurons in the dorsal nucleus of the lateral lemniscus of cat to binaural tonal stimulation. *J. Neurophysiol.* **33**:441.

BRUGGE, J. F., D. J. ANDERSON, J. E. HIND, and J. E. ROSE. 1969. Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. *J. Neurophysiol.* **32**:386.

CODDINGTON, E. A., and N. LEVINSON. 1955. Theory of ordinary differential equations. McGraw-Hill Book Company, New York.

DENJOY, A. 1932. Sur les courbes definies par les equations differentielles a la surface du tore. *J. Math. Pure Appl.* **11**:333.

GOLDBERG, J. M., and P. B. BROWN. 1969. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J. Neurophysiol.* **32**:613.

Hubel, D. H., and T. N. Wiesel. 1968. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* (*London*). **195**:215.

Hughes, G. W., and L. Maffei. 1966. Retinal ganglion cell response to sinusoidal light stimulation. *J. Neurophysiol.* **39**:333.

Kiang, N. Y.-S., T. Watanabe, E. C. Thomas, and L. F. Clark. 1965. Discharge Patterns of Single Fibers in the Cat's Auditory Nerve. The M.I.T. Press, Cambridge, Mass.

Knight, B. W. 1972. The relationship between the firing rate of a single neuron and the level of activity in a population of neurons. Experimental evidence for resonant enhancement in the population response. *J. Gen. Physiol.* **59**:767.

Knight, B. W., J.-I. Toyoda, and F. A. Dodge. 1970. A quantitative description of the dynamics of excitation and inhibition in the eye of *Limulus*. *J. Gen. Physiol.* **56**:421.

Matthews, P. B. C., and R. B. Stein. 1969. The sensitivity of muscle spindle afferents to small sinusoidal changes of length. *J. Physiol.* (*London*). **200**:723.

Mendell, L. M., and E. Henneman. 1968. Terminals of single Ia fibers: distribution within a pool of 300 homonymous motor neurons. *Science* (*Washington*). **160**:96.

Mountcastle, V. B., W. H. Talbot, H. Sakata, and J. Hyvarinen. 1968. Cortical neuronal mechanisms in flutter-vibration studied in unanesthetized monkeys. Neuronal periodicity and frequency discrimination. *J. Neurophysiol.* **32**:452.

Partridge, L. 1966. A possible source of nerve signal distortion arising in pulse rate encoding of signals. *J. Theor. Biol.* **11**:257. (See also Erratum. 1968. *J. Theor. Biol.* **21**:292.)

Poppele, R. E., and R. J. Bowman. 1970. Quantitative description of linear behavior of mammalian muscle spindles. *J. Neurophysiol.* **33**:59.

Rescigno, A., R. B. Stein, R. L. Purple, and R. E. Poppele. 1970. A neuronal model for the discharge patterns produced by cyclic inputs. *Bull. Math. Biophys.* **32**:337.

Rose, J. E., J. F. Brugge, D. J. Anderson, and J. E. Hind. 1967. Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *J. Neurophysiol.* **30**:770.

Rose, J. E., J. F. Brugge, D. J. Anderson, and J. E. Hind. 1969. Some possible neural correlates of combination tones. *J. Neurophysiol.* **32**:402.

Stein, R. B. 1970. The Role of Spike Trains in Transmitting and Distorting Sensory Signals. *In* The Neurosciences. The Rockefeller University Press, New York. 597.

Stein, R. B., and A. S. French. 1969. Models for the Transmission of Information by Nerve Cells. *In* Excitatory Synaptic Mechanisms. Oslo University Press, Oslo. 247.

Talbot, W. H., I. Darian-Smith, H. H. Kornhuber, and V. B. Mountcastle. 1967. The sense of flutter-vibration: comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey hand. *J. Neurophysiol.* **31**:301.