J. theor. Biol. (1970) 27, 291–337

Neural Pattern Discrimination

STEPHEN GROSSBERG

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Mass. 02139, *U.S.A.*

(*Received* 8 *July* 1969, *and in revised form* 10 *November* 1969)

Some possible neural mechanisms of pattern discrimination are discussed, leading to neural networks which can discriminate any number of essentially arbitrarily complicated space–time patterns and activate cells which can then learn and perform any number of essentially arbitrarily complicated space–time patterns in response to the proper input pattern. Among the topics that arise in this discussion are: use of non-recurrent inhibitory interneurons for temporal or spatial discrimination tasks which recurrent inhibitory interneurons cannot carry out; mechanisms of temporal generalization whereby the same cells control performance of a given act at variable speeds; a tendency for cells furthest from the sensory periphery to have the most specific response modes and the least ability to follow sensory intensities (e.g. on–off and bimodal responses are common); uses of non-recurrent on–off fields whose signals arrive in waves forming "interference patterns", with the net effect of rapidly choosing at most one behavioral mode from any number of competitive modes, or of non-specifically arousing or suppressing cells which can sample and learn ongoing internal patterns; uses of specific vs. non-specific inhibitory interneurons, axon hillock inhibition, presynaptic inhibition, equal smoothing of excitatory and inhibitory signals, possible production of both excitatory and inhibitory transmitter in a single synaptic knob, blockade of postsynaptic potential response, logarithmic transduction of inputs to spiking frequencies, or saturation of cell body response in non-recurrent on–off fields for purposes of pattern discrimination.

1. Introduction

Recent electrophysiological experiments, such as those of Hubel & Wiesel (1965), Lettvin, Maturana, McCulloch & Pitts (1960) and Sterling & Wickelgren (1969) have demonstrated the existence of nerve cells whose maximal output is triggered by complex input patterns; for example, by a line of light of fixed length moving across a cat's retina with a fixed orientation and a fixed velocity. Many different cellular preferences have already been described, and more are bound to be discovered in the future. The

problem of classifying possible common mechanisms underlying these diverse preferences is therefore an urgent one. Still more important, perhaps, is the problem of discussing how complex cellular pattern discriminations are integrated into the ongoing learning behavior of the animal being discussed.

These problems cannot be separated from the following question, which is suggested by electrophysiological experiments on many sensory modalities. How do cells near the sensory periphery, which individually respond to many different input patterns, nonetheless act together to discriminate one pattern from another with high precision? That is, how does "local non-specificity" co-exist with "global specificity" of cellular response?

This paper discusses some mathematical mechanisms of pattern discrimination that are built up from such familiar neurophysiological ideas as: existence of an axonal spiking threshold, dependence of spiking frequency on cell body membrane potential, additive combination of excitatory and inhibitory inputs at the cell body, exponential averaging of inputs through time by the cell body membrane potential, positive time lags for flow of spikes in axons, and—in some cases—saturation of cell response at finite values. We will find that these simple ideas can be used to create a substantial variety of pattern discriminators if they are arranged in suitable networks, or "anatomies".

It is, of course, very difficult, if not impossible, for a physiological experimentalist to simultaneously record from the millions of nerves that might be used to perform a given pattern discrimination task of even routine behavioral complexity. This limitation will not hamper our mathematical analysis, which is in principle as easily carried out for two as for 10^{10} cells. The independence of our results from considerations of cell population size has the empirical interpretation that *any* pattern resolution can be achieved by repeating the same mechanisms in more cells. It will also appear that knowing the anatomy of a given collection of cells does not characterize the capabilities of this collection as an input filter. One must also know several physiological parameters of the network, such as the relative strengths and onset times of excitatory and inhibitory signals at a given cell, the relative speeds of exponential averaging at different cells, the spatial distribution of spiking threshold values at all cells, the relative specificity of excitatory and inhibitory synaptic fields, etc. These parameters are also difficult, if not impossible, to measure at millions of cells, but again a mathematical analysis will show that certain combinations of these parameters at individual cells are compatible with prescribed tasks carried out by millions of cells of the same type, whereas other parameter combinations are not. Our results therefore begin a catalog of the mathematical possibilities that can achieve

prescribed behavioral discrimination tasks. Armed with such a catalog, the experimentalist can presumably better interpret the behavioral implications of data collected separately from small numbers of cells.

2. Connection with Learning Theory

The results to be described are part of a rigorous theory of learning that has a suggestive psychological, neurophysiological, anatomical, and in some cases biochemical interpretation. The networks of the theory are called *embedding fields* (Grossberg, 1969*a* to *d*). All results herein were motivated and derived to satisfy formal requirements that make efficient learning of complicated skills possible in embedding fields. In this sense, the learning mechanism provides a unifying teleology for constructing pattern discriminators. Some previous papers (Grossberg, 1969*e,f*, 1970) consider this mechanism in detail. Our first results on pattern discrimination will discuss networks in which no learning occurs. If such a network can discriminate a given pattern at one time, it can discriminate the same pattern at future times. These networks will then be connected to networks which can learn. Together the entire network can learn to perform any finite number of output patterns of essentially arbitrary complexity selectively in response to any finite number of input patterns of essentially arbitrary complexity. Once these results are before us, various advantages and disadvantages of including learning, e.g. transmitter potentiation, in the filtering cells themselves can readily be noted.

We will work primarily with networks of the form

$$\dot{x}_i(t) = -\alpha_i x_i(t) + \sum_{k=1}^{n} [x_k(t-\tau_{ki})-\Gamma_{ki}]^+ \beta_{ki} - \sum_{k=1}^{n} [x_k(t-\sigma_{ki})-\Omega_{ki}]^+ \gamma_{ki} + I_i(t), \tag{1}$$

where $[w]^+ = \max(w, 0)$ for any real number w, all constant parameters are non-negative, all initial data and inputs are continuous, and $i = 1, 2, \ldots, n$ with n any fixed positive integer. (1) has the following interpretation.

Let n cell bodies v_i be given with average potential $x_i(t)$, $i = 1, 2, \ldots, n$. If $\beta_{ki} > 0$ ($\gamma_{ki} > 0$), then an excitatory (inhibitory) axon $e_{ki}^+(e_{ki}^-)$ leads from v_k to v_i. Denote the synaptic knob of $e_{ki}^+(e_{ki}^-)$ by $N_{ki}^+(N_{ki}^-)$. Let the spiking frequency which is released from v_k into $e_{ki}^+(e_{ki}^-)$ in the time interval $[t, t+dt]$ be proportional to $[x_k(t)-\Gamma_{ki}]^+\beta_{ki}([x_k(t)-\Omega_{ki}]^+\gamma_{ki})$. Let the time lag for the signal to flow from v_k to $N_{ki}^+(N_{ki}^-)$ be $\tau_{ki}(\sigma_{ki})$, and let the spiking (or signal) threshold of $e_{ki}^+(e_{ki}^-)$ be $\Gamma_{ki}(\Omega_{ki})$. Then by (1), in every time interval $[t, t+dt]$, a signal with size proportional to $[x_k(t)-\Gamma_{ki}]^+\beta_{ki}$ enters e_{ki}^+ from v_k, travels to N_{ki}^+ at finite velocity, and creates a proportional signal at N_{ki}^+ that crosses

the synapse to v_i at time $t+\tau_{ki}$, whence $\dot{x}_i(=dx_i/dt)$ increases proportionately. All excitatory signals from some v_k that reach v_i at time t have an additive effect on $\dot{x}_i$, as the term

$$\sum_{k=1}^{n}[x_k(t-\tau_{ki})-\Gamma_{ki}]^+\beta_{ki}$$

in (1) shows. A similar description holds for inhibitory signals in the axons e_{ki}^-. x_i also decays at the exponential rate α_i, and is perturbed by known inputs $I_i(t)$ that are under control of an experimentalist or independent cells. For my purposes, any synaptic mechanism—whether chemical or electrical—that obeys the above equations will suffice. Equations (1) are supported by substantial experimental evidence (Grossberg, 1969*b*, section 12). For example, they reduce in a special case to the Hartline–Ratliff equation and yield theoretical formulas for the empirical coefficients of that equation (Grossberg, 1969*b*, section 13).

3. Local Temporal Discrimination

It is often necessary for the output of a given cell to have short duration even though its input has long duration. Consider Fig. 1, for example. Figure 1 describes a respondent conditioning paradigm in a simple network.

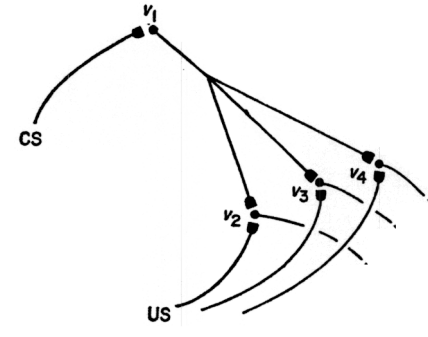


FIG. 1.

A conditioned stimulus (CS) activates the cell v_1, and an unconditioned stimulus (US) activates the cells v_i, $i \geq 2$. It has been shown (Grossberg, 1969*e*, 1970) that by pairing the CS and the US sufficiently often, then a future presentation of the CS alone can reproduce in the outputs of the cells v_i, $i \geq 2$, patterns previously elicited by the US. Grossberg (1969*e*) also shows that the duration of the signal from v_1 must be brief if the US fluctuates rapidly in time, or else the synaptic knobs of v_1 will learn only a

coarse weighted average in time of all the patterns playing on the cells v_i, $i \geq 2$. On the other hand, the CS can have a long duration. For example, in Pavlov's famous experiments on dogs, the duration of the bell (CS) can in principle be very long. Thus we seek a mechanism that shuts off the output from v_1 shortly after it is created by the CS, no matter how prolonged the CS is. Shutting off the v_1 signal while the CS input is still large clearly requires an inhibitory input.

Two main ways exist whereby this inhibitory input might occur, and obvious modifications of them can be readily imagined (Fig. 2).

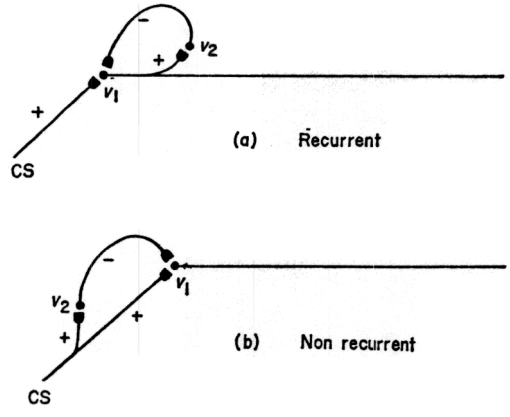


FIG. 2.

In Fig. 2(a), inhibition is *recurrent:* v_1 gives rise to an excitatory axon collateral which perturbs an inhibitory interneuron v_2 that thereupon inhibits v_1. In Fig. 2(b), inhibition is *non-recurrent:* the CS gives rise to an axon collateral which perturbs an inhibitory interneuron v_2 that thereupon inhibits v_1. Only the non-recurrent case has the desired effect, as we now prove.

The systems of Fig. 2 satisfy the following equations.

Recurrent

$$\dot{x}_1(t) = -\alpha x_1(t) - \beta[x_2(t-\sigma) - \Omega]^+ + I_{CS}(t) \tag{2}$$

and

$$\dot{x}_2(t) = -\gamma x_2(t) + \delta[x_1(t-\tau) - \Gamma]^+, \tag{3}$$

for suitable positive parameters and CS input $I_{CS}(t)$. Γ is the spiking threshold of the v_1 axon and Ω is the spiking threshold of the v_2 axon. Γ is also chosen as the spiking threshold of the axon collateral leading from the v_1 axon to

v_2. A larger spiking threshold would only weaken the inhibition exerted by v_2 on v_1, and thereby make our ensuing conclusions easier to prove. A smaller spiking threshold is impossible unless v_1 sends out two independent axon collaterals, instead of one axon with an axon collateral.

A similar interpretation of Γ and Ω holds for the following system.

Non-recurrent

$$\dot{x}_1(t) = -\alpha x_1(t) - \beta[x_2(t-\sigma) - \Omega]^+ + I_{CS}(t) \tag{4}$$

and

$$\dot{x}_2(t) = -\gamma x_2(t) + \delta I_{CS}(t-\tau). \tag{5}$$

To illustrate the essential differences between these networks when the CS is steadily applied for a long time, we consider the case in which the limit $I = \lim_{t\to\infty} I_{CS}(t)$ exists. We also suppose that the CS intensity is sufficiently great to create a signal from v_1 at large times; i.e. $I > \alpha\Gamma$. To avoid an unilluminating discussion of tedious cases, we start the system in equilibrium, i.e. $x_1(t) = 0$, $-\tau \leq t \leq 0$, and $x_2(t) = 0$, $-\sigma \leq t \leq 0$. (Any equilibrium levels P_i for x_i can be discussed in (1) by replacing $-\alpha_i x_i(t)$ by $-\alpha_i(x_i(t) - P_i)$.) [See Grossberg (1969*b*).]

Proposition 1 (*recurrent*). Under the above hypotheses, either the limits $x_i(\infty) = \lim_{t=\infty} x_i(t)$ exist, $i = 1, 2$, with $x_1(\infty) > \Gamma$, or $x_1(t)$ oscillates above Γ infinitely often and at arbitrarily large times.

In short, $x_1(t)$ cannot permanently be driven below threshold if the input is prolonged and sufficiently intense ever to drive $x_1(t)$ above threshold. The proof is given in Appendix A.

The non-recurrent case presents none of these difficulties. For definiteness, three classes of increasingly general inputs will be considered:

(I) *Steady state:* $I_{CS}(t) = I$, $t \geq 0$.

(II) *Monotonely concave:* $\dot{I}_{CS}(t) \geq 0$ and $\ddot{I}_{CS}(t) \leq 0$, $t \geq 0$ (one-sided derivatives are intended where two-sided derivatives do not exist).

(III) *Asymptotically steady state:* $I = \lim_{t\to\infty} I_{CS}(t)$ exists. The heuristic importance of (II) is the following. Let monotonely concave inputs perturb a finite number of cells $\mathscr{V}_1$ in equilibrium, and let these cells send excitatory signals to other cells $\mathscr{V}_2$ along axons having arbitrary time lags and thresholds. Let the cells $\mathscr{V}_2$ in turn send excitatory signals to cells $\mathscr{V}_3$, and so on. Then the potential of a cell body in some cell collection $\mathscr{V}_n$ will, except for usually brief sigmoidal portions of its growth curve, be monotonely concave. Thus an essentially monotonely concave input to v_1 can represent the net effect of rather general excitatory preprocessing. This fact is summarized in the following simple lemmas.

Lemma 1. Let the non-negative functions $x_i(t)$ be monotonely concave, $i = 1, 2, \ldots, n$. Then the function

$$Y(t) = \sum_{i=1}^{n} [x_i(t-\tau_i) - \Gamma_i]^+ \beta_i$$

is also monotonely concave if all parameters are non-negative.

Lemma 2. Define the sequence of functions $x_i(t)$ by

$$x_0(t) = \begin{cases} 0, & t < 0 \\ I, & t \geq 0 \end{cases}$$

$$\dot{x}_i(t) = -\alpha_i x_i(t) + \beta_i [x_{i-1}(t-\tau_i) - \Gamma_i]^+, \tag{6}$$

with $x_i(0) = 0$ and $i = 1, 2, \ldots, n$. Then $x_1(t)$ is monotonely concave, whereas each $x_i(t)$, $i = 2, 3, \ldots, n$ is sigmoidal, i.e. $\dot{x}_i(t) \geq 0$ and $\ddot{x}_i(t)$ changes sign at most once from non-negative to non-positive.

The proof of Lemma 1 is obvious. The simple proof of Lemma 2 is given in Appendix B.

The following theorem shows that the non-recurrent network (4) to (5) can cut off the output produced by a prolonged input. This theorem also studies the number of oscillations in the output and the time needed to shut off the output given sufficiently simple test inputs. To simplify the equations, let $\tau = 0$, which synchronized the CS onset at v_1 and v_2, and thereby causes the inhibitory signal from v_2 to v_1 to lag behind CS onset. In the steady-state case, for each I, $x_1(t)$ and $x_2(t)$ will henceforth be denoted by $x_1(t, I)$ and $x_2(t, I)$, respectively. The functions

$$S(I) = \min \{t: x_2(t-\sigma, I) = \Omega\}$$

$$T(I) = \max \{t: x_1(t, I) = \Gamma\}$$

will be used to denote the onset time of non-recurrent inhibition at v_1, and the time at which $x_1(t, I)$ is finally driven to subthreshold values by inhibition, whenever these times exist.

Theorem 1 (*non-recurrent*).† Let $\beta\delta > \gamma$, $\alpha\delta\Gamma > \gamma\Omega$ and $I > \alpha\Gamma$, and suppose $x_1(t)$ and $x_2(t)$ start out in equilibrium. If $I_{CS}(t)$ is asymptotically steady state, then $x_1(\infty) < \Gamma$. If $I_{CS}(t)$ is monotonely concave, then $\dot{x}_1(t)$ changes sign at most once from non-negative to non-positive. If $I_{CS}(t)$ is steady state, then $dS/dI < 0$ and the limit $T(\infty) = \lim_{I \to \infty} T(I)$ exists, is finite, and if $\alpha \neq \gamma$ satisfies the equation

$$\mu e^{-\gamma T} + \nu e^{-\alpha T} = \omega,$$

† This theorem is proven in Appendix C.

where

$$\mu = \gamma^{-1}(\alpha-\gamma)^{-1}\alpha\beta\delta\, e^{\gamma\sigma}, \qquad \nu = (\gamma-\alpha)^{-1}\, e^{\alpha\sigma} - 1,$$

$$\omega = \gamma^{-1}\beta\delta - 1.$$

Also in the steady-state case, if $\alpha\Gamma \geq \beta\Omega$, then $dT/dI \geq 0$, whereas if $\alpha\Gamma < \beta\Omega$, then $(dT/dI)(I) \leq 0$ for $I \geq I_0$ if $(dT/dI)(I_0) \leq 0$. Moreover

$$(\alpha-\gamma)\frac{d}{dI}(T-S)(I) \leq 0 \quad \text{for } I \geq I_0$$

if

$$(\alpha-\gamma)\frac{d}{dI}(T-S)(I_0) \leq 0.$$

Remark. The intuitive meaning of the inequalities in the Theorem is as follows. The inequality $\beta\delta > \gamma$ keeps $x_1(\infty, I)$ bounded from above as $I \to \infty$. The condition $\alpha\delta\Gamma > \gamma\Omega$ guarantees that inhibition sets in whenever v_1 can transmit a prolonged excitatory signal, and that ultimately the excitatory signal is inhibited away. If $I \leq \alpha\Gamma$, then no prolonged excitatory signal can occur, even without inhibition. Any monotonely concave input creates a single rise and fall in v_1's potential. The duration of suprathreshold response is essentially a monotonic function of input intensity: only one sign change in $\dot{T}(I)$ can occur. A similar statement holds for $(T-S)^{\cdot}(I)$ with the following addition: if inhibition grows more rapidly than excitation ($\alpha < \gamma$), then $(T - S)^{\cdot}$ can change sign only towards the negative, i.e. then increasing input intensity tends to "contract" the time scale of suprathreshold response. This theorem guarantees that a non-recurrent inhibitory interneuron can limit the duration of the v_1 output even in response to an indefinitely prolonged CS input. As a result, the synaptic knobs of v_1 can learn the pattern weights at the cells bodies on which they impinge with arbitrarily good temporal discrimination.

The above conclusions are independent of how we interpret the CS. Suppose we could guarantee that only a prescribed space-time pattern at the sensory periphery ever creates a positive input at v_1. Then by Grossberg (1969*e*, 1970), we could also guarantee that any prescribed output pattern can be learned in response to the given one by letting v_1 be the source of an outstar avalanche. Moreover, if we were given n cells v_i, $i = 1, 2, \ldots, n$, rather than just v_1, and if each of these cells could respond only to a prescribed space–time pattern $\mathscr{P}_i$, then we could guarantee that any output pattern can be learned in response to any of the n patterns $\mathscr{P}_i$ by letting each v_i be a source of an outstar avalanche. For example, such a network can in principle learn to play any sonata in response to any moving picture of external events.

The simplest networks of this kind will perform their discrimination and learning tasks in a wholly ritualistic way. Grossberg (1969*e*) indicates, however, that teleological factors such as "goals", "internal drive states", "paying attention", "novelty", and the like can be introduced into the network by suitably modifying its anatomy. A fuller discussion of such teleological factors will presumably be facilitated by a clear-cut description of the minimal mechanisms needed to discriminate and learn complicated tasks in a ritualistic way.

Non-recurrent inhibitory interneurons can be used as temporal discriminators in many anatomical situations. The following two examples illustrate some of the possibilities.

(A) COMPATIBILITY OF DIFFUSE AROUSAL AND TEMPORAL DISCRIMINATION

Figure 1 can be augmented as in Fig. 3, which is discussed in detail in Grossberg (1969*e*, 1970).

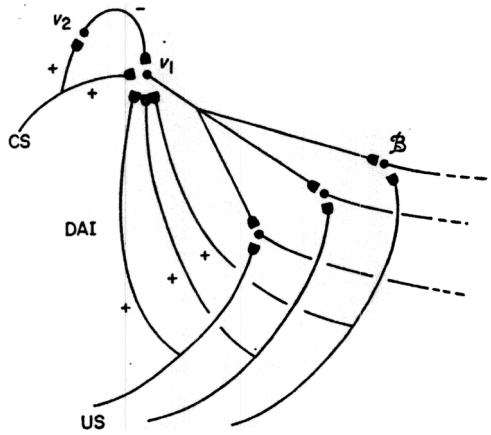


FIG. 3.

In Fig. 3, each US axon send off an excitatory axon collateral to v_1. The inputs delivered to v_1 by these collaterals are called *diffuse arousal inputs* (DAI's), since they arrive at v_1 a fixed time η before the US is received by the cells $\mathscr{B}$, and at all other cells such as v_1 that send axon collaterals to $\mathscr{B}$. The DAI has the following function. Suppose that the US follows the CS by a time lag of T_i on the ith learning trial. If the CS alone can trigger a v_1 output, and if T_i is not independent of i, then a given synaptic knob of v_1 can learn

different parts of the same pattern at $\mathscr{B}$ on different trials, and hence might learn no one pattern well as a result of repeated practice. To avoid this difficulty, require that the CS and the DAI be simultaneously active at v_1 before a signal from v_1 can be created. This is readily done by choosing a sufficiently large v_1 signal threshold once and for all. Then a signal from v_1 will occur approximately η time units before the US onset on all learning trials. Thus a given synaptic knob can practise the same part of the pattern on successive trials, thereby leading to perfect learning.

The inhibitory interneuron v_2 in Fig. 3 furthermore guarantees that the v_1 signal will have a short duration, since by inhibiting the CS, the inhibitory signal also drives v_1 below its signal threshold, given the additivity of all inputs at v_1. Thus synchronizing v_1–US onset and achieving good local temporal discrimination can be simultaneously achieved. This example illustrates an important general theme: spatio-temporally specific *and* diffuse interactions co–existing at the same cells can contribute to the over-all well-being of the organism.

(B) CEREBELLAR PURKINJE CELLS AS NON-RECURRENT TEMPORAL DISCRIMINATORS

In Fig. 4, the non-recurrent inhibitory interneuron is controlled by the US in a gridwork of interactions between perpendicular somatotopic representations of two cell collections having linearly ordered components. See Grossberg (1969*g*) for background details.

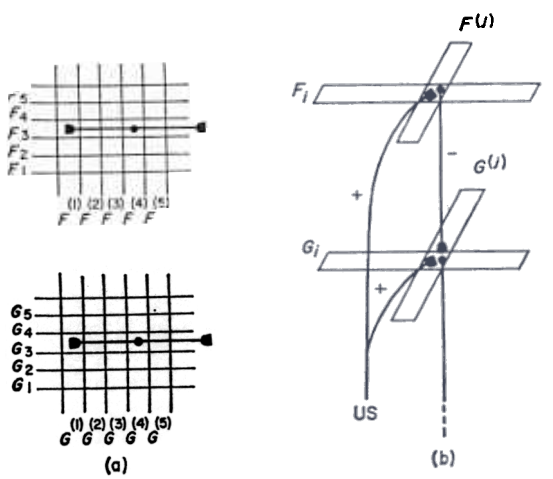


FIG. 4.

Figure 4(a) shows a view from above of two pairs of perpendicular somatotopic representations: $F \equiv \{F_1, F_2, \ldots\}$ and $F^\perp \equiv \{F^{(1)}, F^{(2)}, \ldots\}$, and $G \equiv \{G_1, G_2, \ldots\}$ and $G^\perp \equiv \{G^{(1)}, G^{(2)}, \ldots\}$. The CS is delivered in duplicate to the F and G representations, and the US is delivered to the $F^\perp$ and $G^\perp$ representations. A CS to G_i (and thus to F_i) excites a strip of parallel fibers, as in Fig. 4(b), which carry their signals through the strip and thereupon excite its output cells at different times. The US excites a perpendicular strip $F^{(J)}$. In order to temporally tag the onset of the US, and thereby restrict the class of parallel fibers in G_j which form large cross-correlations with the output cells of $G^{(J)}$, a non-recurrent inhibitory interneuron in $F^{(J)}$ is also activated by the US and rapidly inhibits $G^{(J)}$. After learning occurs, the CS alone should produce an output from $\mathscr{G}$. This output will also be rapidly inhibited due to parallel learning of which inhibitory interneuron in $\mathscr{F}$ should fire to $\mathscr{G}$. Grossberg (1969*g*) points out a possible analogy between the system $\mathscr{F}$ and the cerebellar neocortex. In this analogy, the non-recurrent inhibitory interneurons from $\mathscr{F}$ to $\mathscr{G}$ are cerebellar Purkinje cells, whose role as temporal discriminators would be established were the analogy fully valid.

4. Choices between Incompatible Behaviors: "Majority Rule" in Non-recurrent "Interference Patterns"

In many behavioral situations, a rapid choice between mutually incompatible behavioral modes is called for. Non-recurrent inhibitory interneurons can achieve such a choice between any finite number of behavioral alternatives in one processing step. In Fig. 5(a), the ith input sends an excitatory input to v_i and an equal inhibitory input to v_k, $k \neq i$. In Fig. 5(b), essentially the same process occurs but Dale's Principle is respected, i.e. the synaptic knobs of one cell are either all excitatory or all inhibitory. We will show that the total input of at most one cell v_i can ever be positive at any time, and thus that at most one $x_i(t)$ at a time can be driven up towards suprathreshold values.

The interpretation of the v_i outputs can be very varied. For example, a given v_i can serve as a diffuse arousal source for a large collection of cells $\mathscr{V}_i$ that are used to learn behavior sequences compatible with the ith behavioral mode, e.g. eating or sex. The cells $\mathscr{V}_i$ can be disjoint from all cells $\mathscr{V}_j, j \neq i$, even if the cells $\mathscr{V}_i$ and $\mathscr{V}_j$ lie very close to another. Keeping all cells $\mathscr{V}_i$ and $\mathscr{V}_j$ close might be necessary, for example, to give both behavioral modes equal access to the same motor pathways. On the other hand, cells which can be aroused by several modes can also be readily contemplated, but they would presumably control preparatory precursors of the overt behavioral modes that are mutually compatible.

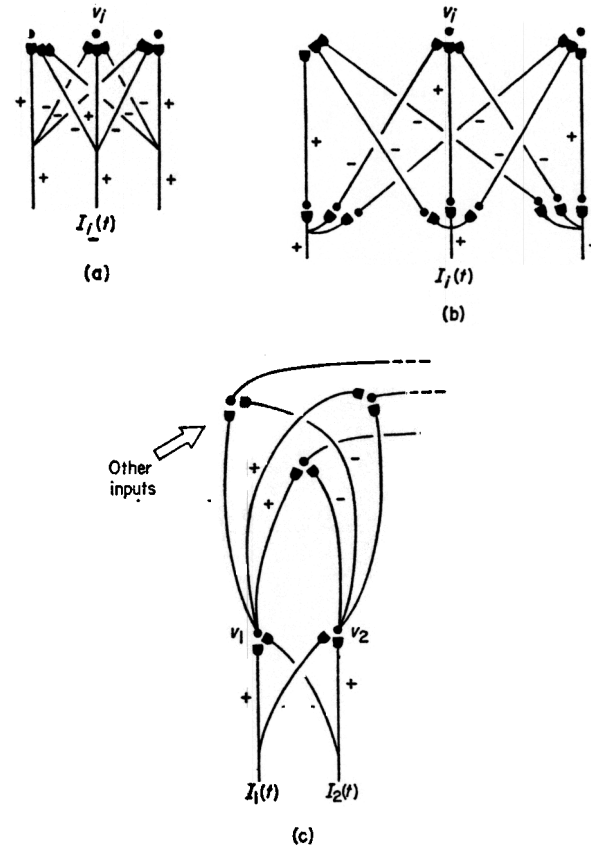


FIG. 5.

Figure 5(c) illustrates a different possibility in the case of two cells v_1 and v_2. Here again only one cell output can be facilitated at any time. Whereas v_1 subliminally arouses a collection of cells, v_2 suppresses these cells. Suppose that exciting v_1 indicates the arrival of inputs $I_1(t)$ which portend good consequences for the organism, whereas exciting v_2 portends bad consequences, e.g. occurrence of food vs. shock. If the instantaneous attraction of food, as manifested by the intensity of $I_1(t)$, is overwhelmed by the instantaneous repulsion of shock, as manifested by the intensity of $I_2(t)$, then the cells which would ordinarily be aroused by, and therefore learn from, the sequence of ongoing events can readily be suppressed by the diffuse v_2 output. The converse is also clearly true. The intensities of $I_1(t)$

and $I_2(t)$ in this case would presumably be partially determined by signals from internal homeostats.

The simplest equations of embedding field type that have these properties are given as follows.

$$\dot{x}_i(t) = -\alpha_i x_i(t) + I_i(t-\tau_i) - \sum_{k \neq i} I_k(t-\tau_k), \tag{7}$$

where the output from v_i has the form

$$O_i(t) = \beta_i[x_i(t) - \Gamma_i]^+, \quad \Gamma_i > 0, \tag{8}$$

$i = 1, 2, \ldots, n$. Note that the arrival time of all excitatory and inhibitory signals with the same source is the same, but that different sources can have different arrival times. In this sense, the individual input sources create incoming "waves", and the excitatory-inhibitory interaction between these waves creates "interference patterns" which give rise to unambiguous "choices". This restriction on arrival times can be guaranteed by choosing the signal velocity in axons proportional to axonal length, i.e. by spatio-temporal self-similarity (Grossberg, 1969*g*).

Analysis of (7) is readily accomplished by defining the functions

$$I(t) = \sum_{k=1}^{n} I_k(t-\tau_k)$$

and

$$\theta_i(t) = I_i(t-\tau_i) I^{-1}(t),$$

with the understanding that $\theta_i(t) = 0$ if $I(t) = 0$. Then (7) becomes

$$\dot{x}_i(t) = -\alpha_i x_i(t) + I(t)(\theta_i(t) - \sum_{k \neq 1} \theta_k(t)),$$

and since $\sum_{k=1}^{n} \theta_k(t) = 1$ whenever $I(t) > 0$,

$$\dot{x}_i(t) = -\alpha_i x_i(t) + J_i(t), \tag{9}$$

where

$$J_i(t) = 2I(t)(\theta_i(t) - \tfrac{1}{2}),$$

$i = 1, 2, \ldots, n$. Clearly at most one $J_i(t)$ is positive at any time t, namely the one for which $\theta_i(t) > \frac{1}{2}$, since the $\theta_i(t)$'s form a probability distribution at any time t for which some $J_i(t) \neq 0$. Thus, by (9), at most one $x_i(t)$ at a time can be driven up by a positive input. This accomplishes our goal given *any* collection of non-negative and continuous inputs $I_i(t)$, no matter how oscillatory each input is.

In particular, if the $\theta_i(t)$ are constant in a sufficiently long time interval, then for any positive thresholds Γ_k, all outputs $O_k(t)$ will eventually be zero unless some $\theta_i > \frac{1}{2}$, i.e. no mode is activated. In this latter case, only $O_i(t)$ can eventually be positive. The rate with which this asymptote is approached —indeed whether it ever is reached—depends on the intensity $I(t)$. For

example if $x_i(0) = 0$, then $x_i(t) = \Gamma_i$ at a time $t = T$ for which

$$(2\theta_i - 1) \int_0^T e^{-\alpha_i(T-v)} I(v)\, dv = \Gamma_i.$$

This shows, roughly speaking, that there is an inverse relationship between relative intensity and total intensity of inputs in determining the system's reaction time. The chance that all $\theta_i(t) \leq \frac{1}{2}$ over long time intervals is reduced by assuming that the inputs to different modes build up according to different time scales.

Two interesting papers (Kilmer, Blum & McCulloch, 1969; Kilmer, 1969) came to my attention after these facts were observed. These papers discuss competition of behavioral modes in the reticular formation using computer methods. Systems (7) and (8) represent a highly simplified case of an alternative attack on this phenomenon using embedding fields. Hierarchies of units such as those above can readily be constructed.

The equation (7) can be generalized as follows.

$$\dot{x}_i(t) = -\alpha_i x_i(t) + \beta_{ii} I_i(t-\tau_i) - \sum_{k \neq i} I_k(t-\tau_k)\beta_{ki}.$$

In this general setting, a wide variety of possibilities occurs in which the v_i are not totally incompatible. For example, in all cases, an input of sufficient intensity to a single cell v_i creates an output only from v_i. By choosing the β_{ki} sufficiently large, however, even an input for which $\theta_i(t) \cong 1-\varepsilon$, with ε any prescribed small number, can be so strongly inhibited that v_i emits no signal. Alternatively, two cells v_i and v_j with $\theta_i(t) \cong \frac{1}{2} \cong \theta_j(t)$ can be so weakly coupled by mutual inhibition (i.e. $\beta_{ij} \cong 0 \cong \beta_{ji}$) that they can fire simultaneously.

The system

$$\dot{x}_i(t) = -\alpha_i x_i(t) + I_i(t-\tau_{ii}) - \sum_{k \neq i} I_k(t-\tau_{ki}) \tag{10}$$

can also yield mutually compatible outputs, even though all excitatory and inhibitory signals have equal strength. This is seen by defining

$$I^{(i)}(t) = \sum_{k=1}^{n} I_k(t-\tau_{ki})$$

and

$$\theta_{ki}(t) = I_k(t-\tau_{ki}) I^{(i)}(t),$$

and writing (10) as

$$\dot{x}_i(t) = -\alpha_i x_i(t) + 2I^{(i)}(t)(\theta_{ii}(t) - \tfrac{1}{2}).$$

One then checks that $\theta_{ii}(t) > \frac{1}{2}$ does not imply $\theta_{jj}(t) \leq \frac{1}{2}$ for all $j \neq i$, in general. Hence the restriction that the input sources create "waves" is of some importance, and more fundamentally, spatio-temporal self-similarity is called for.

5. Local Temporal Generalization: Variable Velocities of Motor Performance

Suppose that a given pattern of muscular motion can be reflexively produced at a fixed velocity. Can this pattern also be produced by the same nerve cells at other velocities? Suppose that the pattern is learned at a fixed velocity. Can the pattern be performed at several velocities? The answer to these questions in our networks is "yes". It is also clearly "yes" in many instances chosen from real life: even complicated piano pieces, practised at one speed, can be performed at several speeds.

First consider the simple case of reflexively performing a motor spatial pattern using the following network:

$$\dot{x}_1(t) = -\alpha_1 x_1(t) + I_1(t) \tag{11}$$

and

$$\dot{x}_i(t) = -\alpha x_i(t) + \beta[x_1(t-\tau_1) - \Gamma_1]^+ p_{1i} \tag{12}$$

$i = 2, 3, \ldots, n$, where the cell v_i sends an axon to an idealized muscle group $\mathcal{M}_i$ whose velocity of contraction at time t is $V_i(t) = \gamma x_i(t-\tau)$ (Fig. 6).

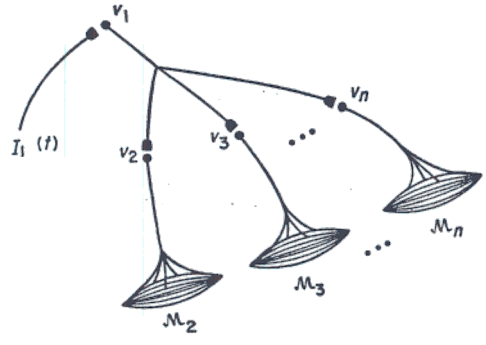


FIG. 6.

Suppose that the system starts out at equilibrium, and perturb the source v_1 with a positive input pulse $I_1(t)$. Then at any time $t \geq 0$, $x_i(t) = p_{1i}J(t)$, where

$$J(t) = \beta \int_0^t e^{-\alpha(t-v)} \left[\int_0^{v-\tau_1} e^{-\alpha_1(v-\tau_1-\xi)} I_1(\xi)\, d\xi - \Gamma_1 \right]^+ dv$$

for $t \geq 0$, and 0 otherwise. Thus $V_i(t) = p_{1i}K(t)$, where $K(t) = \gamma J(t-\tau)$. Either all $V_i(t) = 0$ due to an insufficiently intense input at v_1, or

$$\frac{V_i(t)}{V_j(t)} = \frac{p_{1i}}{p_{1j}},$$

and thus the muscle groups $\mathcal{M}_i$ each contract at fixed relative velocities p_{1i} which characterize the pattern. The absolute velocity of contraction is proportional to $K(t)$, which in turn depends only on the intensity of the input at the source. For example, if in addition to a specific input to v_1, v_1 also receives an arousal input due to some general threat to the organism, then the contraction of the muscles in their prescribed pattern will be speeded up.

This example can be generalized in many ways. Figure 7 provides a simple anatomy for controlling performance of any number of space–time patterns at variable velocities by the same idealized muscle groups.

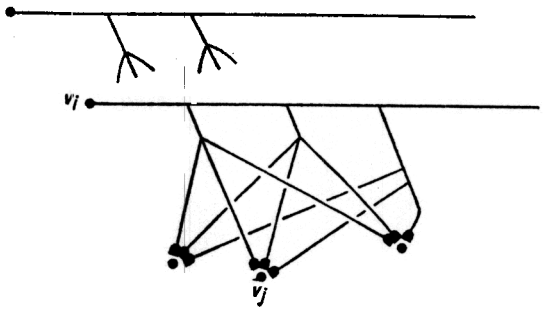


FIG. 7.

Figure 7 describes the system

$$\dot{x}_i(t) = -\alpha_i x_i(t) + I_i(t),$$

and

$$\dot{x}_j(t) = -\alpha x_j(t) + \sum_{i=1}^{n} \sum_{k=1}^{N} [x_i(t-\tau_i-k\xi_i)-\Gamma_i]^+ p_{ijk}, \tag{14}$$

where $i = 1, 2, \ldots, n$ and $j = n+1, n+2, \ldots, n+m$. Let the system start out at equilibrium and suppose that v_i alone is perturbed by a brief but intense input pulse, for some $i \leq n$. Then (14) becomes

$$\dot{x}_j(t) = -\alpha x_j(t) + \sum_{k=1}^{N} [x_i(t-\tau_i-k\xi_i)-\Gamma_i]^+ p_{ijk}.$$

In other words, v_j is perturbed every ξ_i time units by an input with relative weight p_{ijk}. Suppose that the duration of the signal $[x_i(t)-\Gamma_i]^+$ is less than ξ_i and that the decay rate α is large. Then $x_j(t)$ will substantially recover from the kth input burst before the $(k+1)$th burst begins. The response to the kth burst therefore satisfies

$$\frac{x_j(t)}{x_m(t)} \cong \frac{p_{ijk}}{p_{imk}},$$

and thus the muscles can run through the space–time pattern with relative weights p_{ijk}. This is true for any i, so that any number of space–time patterns can be controlled in this way.

Concerning velocity of performance, we can again say that an increase in the absolute size of an axonal signal $[x_i(t) - \Gamma_i]^+$ will speed up performance due to proportionality of muscle contraction rate and $x_j(t-\tau)$ size. Another factor enters in the source signal $[x_i(t)-\Gamma_i]^+$ itself, since increase of $I_i(t)$ decreases the time it takes $x_i(t)$ to exceed Γ_i and to transmit a signal, i.e. decreases the source reaction time. On the other hand, a rate-limiting factor also enters; namely the time interval ξ_i between activation of successive clusters of axon collaterals. This time interval is independent of source energy, and thus pattern performance velocity has a finite upper bound.

All of the above remarks can be carried over to show that a task learned at one speed can be performed at several speeds by varying the source energy. One simply replaces system (13) and (14) by Γ-outstar avalanches, as in Grossberg (1969*e*, 1970). This more complex situation brings with it more interesting possibilities, e.g. a spontaneous speed-up of muscle contraction given a recall input to the source of fixed waveform after a moderate amount of practice (i.e. motor "reminiscence"), and motor manifestations of post-tetanic potentiation, listless response due to disuse, or perfect motor memory until "extinction" occurs on unrewarded trials (Grossberg, 1970).

Clearly Figs 6 and 7 do not describe "muscular" control *per se*, but merely illustrate one way of controlling variable performance rates in general without changing the controlling or learning cells and the pattern they have encoded. One can readily improve the description of the end-organ being controlled in specific cases without necessarily altering these conclusions. For example, in the case of muscular control, at least two improvements in the above discussion are easy to achieve. First, specification of $\mathcal{M}_i$'s contraction in terms of the velocity $V_i(t) = \gamma x_i(t-\tau)$ is insufficient, because when $x_i(t-\tau) = 0$, $V_i(t) = 0$ even if the muscle has not returned to its resting length. Second, some discussion of reciprocity between agonist and antagonist muscles is needed. The following remarks briefly indicate that aspects of these phenomena can be built into the discussion in cases where muscles *per se* are the central concern. These remarks do not, however, pretend to exhaust even some simple macroscopic features of motor control. Our goal is merely to illustrate the flexibility of avalanche and related controllers.

Let $(\mathcal{M}_i^+, \mathcal{M}_i^-)$ be a pair of antagonistic muscles. Let the length of $\mathcal{M}_i^+(\mathcal{M}_i^-)$ at time t be $L_i^+(t)(L_i^-(t))$, and (for simplicity) let the resting length of $\mathcal{M}_i^+$ and $\mathcal{M}_i^-$ be L_{i0}. Introduction of L_{i0} means, in particular, that we will here avoid a discussion of muscle spindles, Golgi tendon organs, and the

structure of limb joints. Let a cell $v_i^+(v_i^-)$ send the excitatory input $\gamma x_i^+(t-\tau)(\gamma x_i^-(t-\tau))$ to $\mathcal{M}_i^+(\mathcal{M}_i^-)$. Then we replace the rule $V_i(t) = \gamma x_i(t-\tau)$ by the pair of equations

$$\dot{L}_i^+(t) = -\delta[L_i^+(t) - L_{i0}] + \gamma[x_i^+(t-\tau) - x_i^-(t-\tau)] \tag{15}$$

and

$$\dot{L}_i^-(t) = -\delta[L_i^-(t) - L_{i0}] + \gamma[x_i^-(t-\tau) - x_i^+(t-\tau)] \tag{16}$$

which describe the non-recurrent inhibitory interaction of Fig. 8(a). Figure 8(b) can also be used.

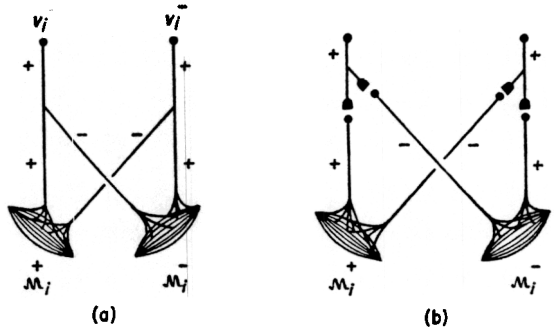


FIG. 8.

These equations clearly generalize the rule for $V_i(t)$ while bringing the muscles back to a resting length in the absence of external forces, and accommodating agonist–antagonist reciprocity. In Fig. 6 the cells v_i can now be replaced by the cell pairs (v_i^+, v_i^-) without changing our conclusions about variable performance velocity. Then we find equations of the form (11), (15), (16),

$$\dot{x}_i^+(t) = -\alpha x_i^+(t) + \beta[x_1(t-\tau_1) - \Gamma_1]^+ p_{1i}^+$$

and

$$\dot{x}_i^-(t) = -\alpha x_i^-(t) + \beta[x_1(t-\tau_1) - \Gamma_1]^+ p_{1i}^-,$$

$i = 2, 3, \ldots, n$. Figure 7 can be similarly generalized.

6. Why are Sensory Pathways in Different Modalities Anatomically Different if Universal Discriminators Exist?

The next section begins a study of the following question. Let n cells v_i, $i = 1, 2, \ldots, n$ be attached to independent peripheral receptors which create the input $I_i(t)$ at v_i. A peripheral environmental event creates a characteristic pattern of inputs $I_i(t)$ on the cells v_i. It is conceivable that

events having very different behavioral significance for the organism all create large inputs to a fixed v_i. In spite of this "local non-specificity" of cellular response, the organism's response to the entire pattern of inputs can be quite specific. Is there a way to construct a cellular network which can discriminate any pattern playing on the cells v_i from any other pattern, to within a prescribed small error, in spite of local non-specificity?

The answer is "yes". Because of this fact, several implications immediately follow. First, the same construction will enable the cells to filter *any* pattern received from the receptors, be its interpretation auditory, tactile, olfactory, visual, etc., i.e. the construction is "universal". It is also simple-minded and uses few cell bodies. Why then do not all the anatomies of sensory pathways in different modalities look alike? One answer seems to be that different modalities preserve different perceptual constancies or invariances, which improve the ability to make "operant" discriminations. For example, in hearing there is a pitch invariance, in vision a size invariance, etc. One can easily in principle imagine networks that include these invariances simply by joining all the input patterns that should create equivalent outputs by some type of "or" switch. Presumably Nature has chosen a more subtle path, if only because connecting up all the requisite "or" switches before they are needed would be very hard to do in a growing brain, and might create undesirable rigidity. Our present construction might therefore exist *in vivo* with the least modification in the most primitive discriminative systems, e.g. smell or taste (Frank & Pfaffmann, 1969). In any case, where anatomically more elaborate discriminators are found, it will henceforth be profitable to ask as a point of departure: why is the universal construction not adequate? Moreover, the same mechanisms are likely to reappear in some form within these more elaborate anatomies.

Second, the number n of receptive cells can be chosen arbitrarily large. Surely if n equals a small number, such as two, there might exist many different environmental events whose effect on two receptors will be the same. As n is taken to large values, however, even if individual receptors differ only slightly in their specificity of response, the chance that the inputs $I_i(t)$ really characterize the environmental event steadily improves. For any fixed choice of n, the network to be constructed can discriminate the pattern—as the organism perceives it—with arbitrarily good accuracy.

7. Unselective Filtering of Spatial Patterns by Excitatory Networks

Suppose that a given collection of n receptor cells v_i, $i = 1, 2, \ldots, n$ is presented with an arbitrary sequence of spatial patterns at widely separated onset times. What is the simplest embedding field network having m output

cells that each respond to at most one spatial pattern, within some prescribed margin of error? Henceforth, we will set $m = 1$, since once the problem of filtering one pattern is overcome, any finite number of patterns can readily be filtered. Suppose for example that the given pattern is $I_i(t) = \theta_i I(t)$ with weights $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$. We will construct a network such that the output cell fires only if the pattern weights $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_n)$ playing on the receptors satisfy the inequalities

$$\theta_i - \varepsilon < \tilde{\theta}_i < \theta_i + \varepsilon \tag{17}$$

$i = 1, 2, \ldots, n$ for some arbitrarily small $\varepsilon > 0$, and if, moreover, the input pattern is presented with sufficient intensity over a sufficiently long time interval, where of course the minimum effective input intensity and duration will tend to vary inversely with respect to one another.

The simplest networks, conceptually speaking, contain only excitatory interactions. A routine example will show, however, that such a network cannot conveniently achieve the selectivity of response that we seek. A mixture of excitatory and inhibitory interactions will hereby be called for, and the deficiencies of the purely excitatory network will readily suggest procedures for connecting the excitatory and inhibitory components.

Let the cells v_i receive the spatial pattern $I_i(t) = \tilde{\theta}_i I(t)$, $i = 1, 2, \ldots, n$. We suppose that all $\tilde{\theta}_i$ are positive without loss of generality, since otherwise we could simply delete the receptors receiving no input from our consideration. Denote the output cell which will be responsive to this pattern by v_{n+1}. v_{n+1} must be responsive to all weights $\tilde{\theta}_i$ of the pattern if ever it could succeed in discriminating this pattern from others. Signals from each v_i must therefore ultimately reach v_{n+1}. If we restrict ourselves to excitatory transmissions, then either an excitatory signal traverses an edge $e_{i,n+1}$ directly from v_i to v_{n+1}, or else several intermediate stages of excitatory processing will be juxtaposed between v_i and v_{n+1}. These intermediate stages can smooth the input, or sum it up, or truncate it using thresholds. We will suppose for simplicity that only direct interactions exist, since the deficiencies which arise in this case can only be made worse by intermediate processing. In the direct case, we find the equations

$$\dot{x}_i(t) = -\alpha_i x_i(t) + \tilde{\theta}_i I(t), \tag{18}$$

$$\dot{x}_{n+1}(t) = -\alpha_{n+1} x_{n+1}(t) + \sum_{k=1}^{n} [x_k(t-\tau_k) - \Gamma_k]^+ \beta_k, \tag{19}$$

where the output from v_{n+1} has the form

$$O_{n+1}(t) = [x_{n+1}(t) - \Gamma_{n+1}]^+ \beta_{n+1}. \tag{20}$$

Let the pattern play upon the receptors with a total intensity that eventually

becomes steady state. Then $I = \lim_{t \to \infty} I(t) > 0$ exists, and by (18) and (19),

$$\lim_{t \to \infty} x_{n+1}(t) = \frac{1}{\alpha_{n+1}} \sum_{k=1}^{n} [\alpha_k^{-1} \tilde{\theta}_k I - \Gamma_k]^+ \beta_k. \tag{21}$$

By (20) and (21), an output will eventually arise from v_{n+1} if I is so large that

$$\sum_{k=1}^{n} [\alpha_k^{-1} \tilde{\theta}_k I - \Gamma_k]^+ \beta_k > \alpha_{n+1} \Gamma_{n+1}.$$

This shows, however, that a sufficiently large input intensity I can create an output from v_{n+1} for *any* choice of pattern weights $\tilde{\theta}_i$. Excitatory networks are therefore unselective if a wide range of total input intensities exists, a perhaps obvious conclusion, but one which when taken seriously has non-trivial consequences.

8. Two Stages of Non-recurrent Inhibition for Pattern Discrimination

The difficulties encountered in the excitatory network are twofold, and can be overcome by two stages of non-recurrent inhibitory interaction.

(A) PATTERN NORMALIZATION AND LOW-BAND FILTERS

In the direct excitatory network, as the input intensity I increases, the asymptotic membrane potentials $x_i(\infty)$ increase linearly, and thus an output from v_{n+1} can be created by an ever less restrictive class of patterns. To avoid this, the potentials $x_i(t)$ must eventually approach a finite asymptote, even if $I \to \infty$. Moreover, the mechanism that creates the asymptote must not distort the recording of pattern weights $\tilde{\theta}_i$ in each $x_i(t)$. We will therefore seek a mechanism such that $x_i(t) \cong \tilde{\theta}_i \Omega$, for some finite constant Ω, and large times t, as $I \to \infty$. This phenomenon will be called *pattern normalization.*

Suppose that we have somehow achieved pattern normalization. This fact along with the existence of positive spiking thresholds now permits us to forbid an output from v_i unless the pattern weight $\tilde{\theta}_i$ satisfies $\tilde{\theta}_i > \theta_i - \varepsilon$, for some $\varepsilon > 0$. Simply choose the spiking threshold Γ_i of the axon emitted by v_i to equal

$$\Gamma_i = \Omega(\theta_i - \varepsilon). \tag{22}$$

Since $x_i(t) \cong \theta_i \Omega(t)$ for some function $\Omega(t)$ which satisfies $\Omega(t) \leq \Omega$ for $t \geq 0$, clearly $x_i(t) \leq \Gamma_i$ unless $\tilde{\theta}_i > \theta_i - \varepsilon$. One-half of the inequalities (17) can hereby be achieved.

(B) HIGH-BAND FILTERS

Now the temptation is great to try completing our construction of a selective spatial pattern by the following simple procedure. Let the outputs

from each v_i converge on a cell v_{n+1}, and choose the spiking threshold Γ_{n+1} so high that v_{n+1} cannot emit a signal unless it receives a positive signal simultaneously from *all* cells v_i. Unfortunately, this device does not suffice. v_{n+1} will also be able to emit signals in response to patterns with weights very different from θ.

To see this, note the following facts. v_{n+1} must surely produce an output if the pattern weights satisfy $\tilde{\theta}_i = \theta_i$, $i = 1, 2, \ldots, n$. In this case, the total input to v_{n+1} is asymptotically approximately

$$\sum_{i=1}^{n} [\theta_i \Omega - \Gamma_i]^+ = n\varepsilon\Omega, \tag{23}$$

by (19), where we have also let all $\beta_i = 1$ for simplicity. Suppose that $\theta_k = \min \{\theta_i : i = 1, 2, \ldots, n\}$ and present the pattern with weights

$$\tilde{\theta}_k = 1, \quad \tilde{\theta}_i = 0, \quad i \neq k. \tag{24}$$

The asymptotic input to v_{n+1} produced by this pattern is approximately

$$\begin{aligned} \sum_{i=1}^{n} [\tilde{\theta}_i \Omega - \Gamma_i]^+ &= \tilde{\theta}_k \Omega - \Gamma_k \\ &= \Omega(1 - \theta_k + \varepsilon). \end{aligned} \tag{25}$$

If the weights θ_i are not close to the weights in (24), than v_{n+1} should not respond. By (23) and (25), therefore, whenever the pattern θ is not concentrated at the one cell v_k,

$$n\varepsilon\Omega \geq \Omega(1 - \theta_k + \varepsilon),$$

or

$$\min \{\theta_i : 1 \leq i \leq n\} \geq 1 - \varepsilon(n-1). \tag{26}$$

Fix n at any finite value. To achieve arbitrarily good pattern discrimination, an arbitrarily small choice of ε should be possible. But then (26) implies

$$\min \{\theta_i : 1 \leq i \leq n\} \cong 1,$$

which is clearly impossible unless $n = 1$, and if $n = 1$ no pattern discrimination whatsoever occurs.

A related deficiency of this approach is seen as follows. To achieve selective filtering of *any* pattern, the right-hand side of (26) must be nonpositive. Then $\varepsilon \geq 1/(n-1)$. But the total input (23) always satisfies

$$\begin{aligned} n\varepsilon\Omega &= \sum_{i=1}^{n} [\theta_i \Omega - \Gamma_i]^+ \\ &\leq \sum_{i=1}^{n} \theta_i \Omega = \Omega, \end{aligned}$$

or $\varepsilon \leq 1/n$. These two inequalities imply the contradiction $n - 1 \geq n$. Thus selective pattern filtering of all patterns is impossible for any n.

These facts show that an additional mechanism is needed to shut off outputs from v_{n+1} in response to patterns for which some $\tilde{\theta}_i > \theta_i + \varepsilon$. In other words, if the signal from any v_i to v_{n+1} becomes too large, it must be shut off or competitively inhibited before $x_{n+1}(t)$ can exceed threshold. If this is achieved, then each v_i can transmit to v_{n+1} either no signal at all or a signal with a finite upper bound. Consequently, we can now choose the threshold Γ_{n+1} so that v_{n+1} transmits a signal only if it receives signals almost simultaneously from all v_i. Since $\Gamma_i = \Omega(\theta_i - \varepsilon)$, the patterns transmitting these signals satisfy $\tilde{\theta}_i > \theta_i - \varepsilon$. By shutting off the signal from v_i to v_{n+1} if it exceeds 2ε in size, these patterns also satisfy $\tilde{\theta}_i < \theta_i + \varepsilon$.

The two stages of input processing: pattern completion, which leads to low-band filtering, and high-hand filtering can both be accomplished by non-recurrent inhibition. An important heuristic lesson of this construction will be that the very same *local* inhibitory mechanisms acting at two different stages of input processing can have profoundly different effects on the *global* transformation of the input at each stage. Of course, only one stage of inhibition is needed to discriminate a pattern of *absolute* input intensities.

9. Specific vs. Non-specific Inhibitory Interneurons, Inhibition at the Axon Hillock, Presynaptic Inhibition, Equal Smoothing and Dale's Principle

The title of this section lists some of the more detailed considerations that will arise while constructing our filter. They are listed here to avoid losing them later in technical details.

Pattern normalization can be accomplished by "non-specific", or "diffusely projecting", non-recurrent inhibitory interneurons, whereas high-band filtering can be done by "specific" interneurons that are excited by one cell and inhibit one cell. These specific inhibitory interneurons can transform an input that varies over a large intensity range into an essentially "on-off" or bimodal output response, the alternative depending on the relative strengths of excitatory and inhibitory inputs. Analogous input–output transformations have been found experimentally in the ventral cochlear nucleus, for example (Whitfield, 1967, p. 80).

It is often important that excitatory and inhibitory signals interact only after they have been smoothed an equal number of times by prior stages of cellular processing. Otherwise, it is hard to achieve the proper relative onset times of excitatory and inhibitory signals, or the proper relative strengths of these signals, for purposes of pattern discrimination. Analogously, equal smoothing is also useful in processing the inputs to pairs of antagonistic "muscles", as in Fig. 8, so that these muscles act in synchrony.

An alternative to equal smoothing exists. The inhibitory interneuron of Fig. 2(b) can be used if it exponentially smooths its input with a decay rate

which is large relative to the fluctuation rate of the input, *and* if the input is magnified before it is smoothed. Then the inhibitory output will be approximately as smooth as the excitatory input. Hence almost equal smoothing is possible by two pathways with different numbers of intermediate cells.

Small, rapidly responding, non-recurrent inhibitory interneurons can accomplish high-band filtering. Rapid response is needed to forbid build-up of the postsynaptic potential to large values. Small interneuronal cell bodies can achieve rapid growth of interneuronal membrane potential by avoiding the dilution of interneuronal input in a large cellular volume. In principle, high-band filtering can be accomplished without an inhibitory interneuron, as the next paragraph notes.

High-band filtering can also be achieved if local postsynaptic response gets blocked as presynaptic spiking frequency increases. Also a switch-over as spiking frequency increases from net release of excitatory transmitter to net release of inhibitory transmitter would be a very effective mechanism. This last mechanism violates Dale's principle, but its efficiency could be so great that it should be kept in mind.

Non-specific inhibitory interneurons can produce pattern normalization if they terminate either at suitable cell bodies, or at the axon hillocks of prior cells, or even at the synaptic knobs of prior cells. The latter two locales for inhibitory interaction are, at least formally, better than the cell body termination for two reasons. First, a layer of cell bodies can then be eliminated. Second, the axonal response rate to inputs is presumably at least as rapid as the response at the axon's cell body, because the axons in our network faithfully replicate in their spiking frequencies the "slow potentials" fluctuating in the cell bodies. The inhibitory input is consequently less smoothed by axon hillock and synaptic knob potentials than by cell body potentials, and thus the cell's net output is more faithfully tuned to input events. In a similar fashion, axon hillock and synaptic knob inhibition is advantageously located to block totally the signal of a large cell body; the same inhibitory signal acting directly at the cell body could be lost in an ocean of excitatory influences. Experimental reports of axon hillock and synaptic knob inhibition have appeared (Eccles, 1964; Eccles, Ito & Szentagothai, 1967).

The mathematical results below on pattern filtering are true under weak constraints on cell body parameters. The main constraints are: the build-up of inhibition is at least as rapid as the build-up of excitation; the inhibitory time lag—discounting threshold effects—is no longer than the excitatory time lag; the inhibitory threshold is higher than the excitatory threshold; and the axonal time scales are no slower than the time scales for slow potential fluctuation in cell bodies. Given these constraints, one can almost

say that simply by "throwing together" excitatory and inhibitory components, some patterns will be selectively filtered. To guarantee that many patterns will be filtered, one can, for example, organize the cells into successive layers whose profusely branching axons flow mainly away from the periphery with a wide distribution of spiking thresholds. Some network examples which illustrate the above remarks are listed below, where $i = 1, 2, \ldots, n$.

Type I

$$\dot{x}_i(t) = -\alpha x_i(t) + \beta I_i(t),\ I_i(t) = \tilde{\theta}_i I(t), \tag{27}$$

$$\dot{x}_{n+1}(t) = -\gamma x_{n+1}(t) + \delta I(t),\ I(t) = \sum_{k=1}^{n} I_k(t), \tag{28}$$

$$\dot{x}_{n+1+i}(t) = -\zeta x_{n+1+i}(t) + \eta[x_i(t-\tau_1) - \Gamma_i]^+ - \kappa[x_{n+1}(t-\tau_2) - \Gamma_{n+1}]^+ \tag{29}$$

$$\dot{x}_{2n+1+i}(t) = -\lambda x_{2n+1+i}(t) + \mu[x_{n+1+i}(t-\tau_3) - \Gamma_{n+1+i}]^+, \tag{30}$$

$$\dot{x}_{3n+1+i}(t) = -\nu x_{3n+1+i}(t) + \xi[x_{n+1+i}(t-\tau_4) - \Gamma_{n+1+i}]^+, \tag{31}$$

$$\dot{x}_{4n+1}(t) = -\rho x_{4n+1}(t) + \sigma \sum_{k=1}^{n} [x_{2n+1+k}(t-\tau_5) - \Gamma_{2n+1+k}]^+ - \chi \sum_{k=1}^{n} [x_{3n+1+k}(t-\tau_6) - \Gamma_{3n+1+k}]^+, \tag{32}$$

and

$$O_{4n+1}(t) = \omega[x_{4n+1}(t) - \Gamma_{4n+1}]^+$$

This network is pictured in Fig. 9, where $n = 2$.

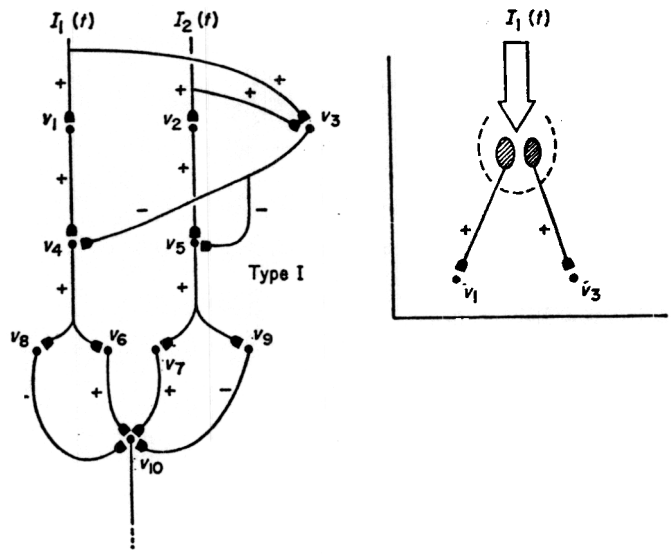
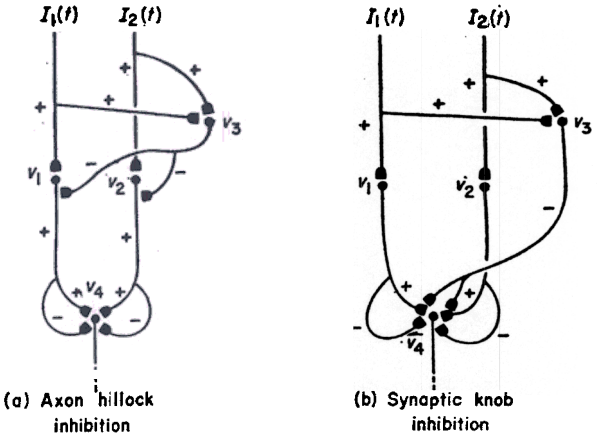


FIG. 9.



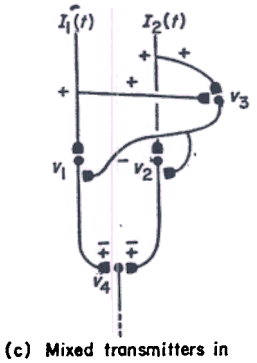


FIG. 10.

Type II illustrates what happens if some of the exponential averaging steps occur so quickly that they can be approximated by algebraic transformations.

Type II

Use (27), (28)

$$O_i(t) = \eta[x_i(t-\tau_1)-\Gamma_i]^+ - \kappa[x_{n+1}(t-\tau_2)-\Gamma_{n+1}]^+, \tag{34}$$

$$\dot{x}_{n+2}(t) = -\lambda x_{n+2}(t)+\mu\sum_{k=1}^{n}[O_k(t-\tau_3)-\Gamma_{n+1+k}]^+ - \nu\sum_{k=1}^{n}[O_k(t-\tau_4)-\Gamma_{2n+1+k}]^+, \tag{35}$$

and

$$O_{n+2}(t) = \xi[x_{n+2}(t) - \Gamma_{n+2}]^+. \tag{36}$$

Some Type II networks are illustrated in Fig. 10, where $n = 2$. All axonal or synaptic knob inhibitions are presumed in Fig. 10 to occur very rapidly compared to exponential averaging rates at the cell bodies. Double synaptic knob inhibition is also possible, if (35) is changed to

$$\dot{x}_{n+2}(t) = -\lambda x_{n+2}(t) + \mu \sum_{k=1}^{n} [O_k(t-\tau_3) - \nu[O_k(t-\tau_4) - \Gamma_{2n+1+k}]^+ - \Gamma_{n+1+k}]^+.$$

The resultant network is shown in Fig. 11.

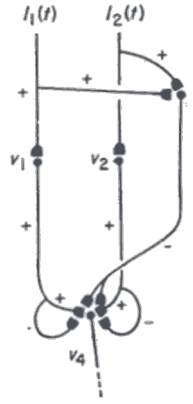


FIG. 11.

Type III and IV networks illustrate some possibilities if *all* exponential averaging steps occur so quickly relative to the input fluctuation rate that they can be approximated by algebraic transformations.

Type III

Let

$$P_i(t) = \alpha[I_i(t-\tau_1) - \Gamma_i]^+ - \beta\left[\sum_{k=1}^{n} I_k(t-\tau_2) - \Gamma_{n+1}\right]^+$$

$$Q_i(t) = \gamma[P_i(t-\tau_3) - \Gamma_{n+1+i}]^+ - \delta[P_i(t-\tau_4) - \Gamma_{2n+1+i}]^+,$$

and let the output be

$$O(t) = \zeta\left[\sum_{k=1}^{n} Q_k(t) - \Gamma_{3n+1}\right]^+$$

Type IV

Use (37) and (39) along with

$$Q_i(t) = \gamma[P_i(t-\tau_3) - \delta[P_i(t-\tau_4) - \Gamma_{2n+1+i}]^+ - \Gamma_{n+1+i}]^+$$

A study of Type I networks will readily suggest how the other, simpler types behave. In a Type I network, the pattern $I_i(t) = \tilde{\theta}_i I(t)$, already neurally encoded, is transferred in a "specific" point-to-point representation along the excitatory pathways

$$I_i(t) \overset{+}{\rightarrow} v_i \overset{+}{\rightarrow} v_{n+1+i}.$$

Each $I_i(t)$ also reaches the "non-specific" inhibitory interneuron v_{n+1} via axon collaterals. Hence in (27), the input to v_i is $\theta_i I(t)$, whereas in (28), the input to v_{n+1} is $\sum_{k=1}^{n} \theta_k I(t) = I(t)$. The non-specific cell v_{n+1} thereupon inhibits each v_i, and the total input to v_{n+1+i} in (29) is given by

$$J_i(t) = \eta[x_i(t-\tau_1)-\Gamma_i]^+ - \kappa[x_{n+1}(t-\tau_2)-\Gamma_{n+1}]^+, \tag{41}$$

which is also the net output $O_i(t)$ from v_i in (34). Each v_{n+1+i} sends excitatory signals to v_{2n+1+i} and v_{3n+1+i}. Each v_{2n+1+i}, in turn, excites the output cell v_{4n+1}, whereas each v_{3n+1+i} inhibits v_{4n+1}, as in (32). The output from v_{4n+1} is given in (33). The inset in Fig. 9 points out that the input receptor itself can be a compound organ with excitatory and inhibitory components, cf. Ratliff (1965).

The general remarks of section 10 are realized in Type I networks as follows:

(A) NON-SPECIFIC VS. SPECIFIC CELLS IN PATTERN NORMALIZATION

The cells v_i and v_{n+1+i} are "specific" cells, since they preserve the spatial separation and ordering of the individual inputs $I_i(t)$. The cell v_{n+1} is a "non-specific" cell, since it averages all inputs $I_i(t)$ and then diffusely inhibits all cells v_{n+1+i}. This non-specific cell will produce the pattern normalization.

(B) OUTPUT THRESHOLDS AS LOW-BAND FILTERS

The cut-off of outputs for which $\tilde{\theta}_i < \theta_i - \varepsilon$ will be accomplished by the thresholds $\Gamma_i = \Gamma_{n+1}(\theta_i - \varepsilon)$ in (29).

(C) SPECIFIC INHIBITORY INTERNEURONS ACTING ON A COMMON OUTPUT CELL AS HIGH-BAND FILTERS

Equations (29) to (32) will forbid any $I_i(t)$ from creating a positive input to v_{4n+1} if $\tilde{\theta}_i > \theta_i + \varepsilon$. The cut-off is accomplished by the inhibitory interneurons v_{3n+1+i}. This could also be accomplished by (34) and (35), or (37) and (38), etc.

Thus a substantial conceptual difference exists between the non-specific inhibitory interneuron v_{n+1} and the specific ones v_{3n+1+i}: v_{n+1} normalizes patterns, whereas v_{3n+1+i} creates a high-band filter.

(D) THE CONTINUUM BETWEEN "AND" AND "OR" IN THE TERMINAL THRESHOLD

The threshold Γ_{4n+1} in (33) can be chosen so large that the output from v_{4n+1} is positive only if all $I_i(t)$ create positive signals to v_{4n+1}. Then Γ_{4n+1} acts like an "and" switch. Were we to let Γ_{4n+1} approach zero, the "and" switch would be smoothly deformed into an "or" switch when $\Gamma_{4n+1} = 0$, and $O_{4n+1}(t)$ would be positive if any signal to v_{4n+1} were positive. This is one reason why the anatomy of the network by itself does not suffice to tell us what tasks the network is performing. In a similar way, all our results would change were the inhibitory signals too weak, or too slow, etc.

(E) EQUAL SMOOTHING

The excitatory cells v_i and the inhibitory cell v_{n+1} each smooth the signals reaching v_{n+1+i} once. The excitatory cells v_{2n+1+i} and the inhibitory cells v_{4n+1+i} each smooth the signals reaching v_{4n+1} once. Type II, III and IV networks replace some of these smoothing steps by input amplification and rapid decay.

(F) MULTIPLE-THRESHOLD VS. SINGLE-THRESHOLD CELLS

A paper by Wickelgren (1969) appeared as this paper was being written, and describes some possible uses of neurons with multiple thresholds. From the vantage point of the present formalism, anything that a multiple-threshold neuron can discriminate or learn can also be discriminated or learned by a suitable juxtaposition of single-threshold units. The main formal advantage of a multiple-threshold unit is that, once its design as an input filter is perfected, it can readily be replicated wherever it is needed, e.g. the pyramidal cells of the cerebral cortex. The question therefore seems to be one of evolutionary efficiency and miniaturization of control rather than of absolute formal superiority of one or the other type of system. Our network operations can, in fact, be interpreted as interactions between small membrane patches and associated cell volume segments in multiple-threshold neurons if one so desires. If this is done, then several filtering steps could take place in the dendrites of the multiple-threshold cell, or in any region whose time scale is faster than the scale of slow potential fluctuations.

10. Pattern Normalization in Type I Networks

To prove that pattern normalization is accomplished by equations (27) to (28), we study the input $J_i(t)$ to v_{n+1+i}, as defined in (41), when the spatial pattern $\tilde{\theta}$ is presented. Supposing that x_i and x_{n+1} have zero initial data

(i.e. start out in equilibrium), $J_i(t)$ can be written in the form

$$J_i(t) = \eta[J(\alpha, \tilde{\theta}_i I, t-\tau_1)-\Gamma_i]^+ - \kappa[J(\gamma, I, t-\tau_2)-\Gamma_{n+1}]^+$$

by redefining several parameters and using the notation

$$J(\omega, K, t) = \begin{cases} 0, & t < 0 \\ \int_0^t e^{-\omega(t-v)} K(v)\, dv, & t \geq 0. \end{cases}$$

The following Lemma illustrates pattern normalization in $J_i(t)$ for a convenient choice of parameters.

Lemma 3. Let $\alpha = \gamma$, $\tau_1 = \tau_2$, $\kappa \geq \eta$, and $\Gamma_i = \Gamma_{n+1}(\theta_i - \varepsilon)$. Let $\tau_1 = 0$ for convenience. Let $I(t)$ be any bounded, monotone non-decreasing, and continuous function with $I(0) = 0$.

If $\tilde{\theta}_i \leq \theta_i - \varepsilon$, then $J_i(t) \leq 0$ for all $t \geq 0$.

Suppose $\tilde{\theta}_i > \theta_i - \varepsilon$. Then $J_i(t) = 0$ until the first time $t = T_1$ at which

$$J(\alpha, I, t) = \tilde{\theta}_i^{-1}\Gamma_{n+1}(\theta_i - \varepsilon) \tag{42}$$

and

$$\frac{d}{dt} J(\alpha, I, t) > 0. \tag{43}$$

Thereafter $J_i(t)$ is monotone non-decreasing and satisfies the equation

$$J_i(t) = \eta[\tilde{\theta}_i J(\alpha, I, t) - \Gamma_{n+1}(\theta_i - \varepsilon)]^+ \tag{44}$$

until the first time $t = T_2$ at which (43) and

$$J(\alpha, I, t) = \Gamma_{n+1} \tag{45}$$

hold. For $t > T_2$, $J_i(t)$ is monotone non-increasing. Thus

$$J_i(t) \leq \eta\Gamma_{n+1}[\tilde{\theta}_i - (\theta_i - \varepsilon)]^+ \tag{46}$$

for $t \geq 0$, and

$$\lim_{t\to\infty} J_i(t) = \eta[\alpha^{-1}\tilde{\theta}_i I - \Gamma_{n+1}(\theta_i - \varepsilon)]^+ - \kappa[\alpha^{-1} I - \Gamma_{n+1}]^+, \tag{47}$$

where $I = \lim_{t\to\infty} I(t)$.

Lemma 3 can easily be modified to include inputs which rise so rapidly that both excitation and inhibition set in before the input begins to decay, as proposition 2 will show. The parameter choices $\alpha = \gamma$ and $\tau_1 = \tau_2$ will ultimately be generalized to $\alpha \geq \gamma$ and $\tau_1 \geq \tau_2$, since the excitatory and inhibitory cells cannot *in vivo* be certain to have equal parameters and times lags. The present case illustrates some basic phenomena with a minimum of technical detail, and is studied in Appendix D.

By Lemma 3, the input to v_{n+1+i} is positive only if $\tilde{\theta}_i > \theta_i - \varepsilon$ and if $I(t)$ is sufficiently intense that $\sup_t J(\alpha, I, t) > \Gamma_{n+1}(\theta_i - \varepsilon)$. Even for arbitrarily large I, the input is bounded above by $\Gamma_{n+1}[\tilde{\theta}_i - (\theta_i - \varepsilon)]^+$ and

oscillates at most once. Thus pattern normalization and low-band filtering have occurred.

For a general total input $I(t)$, the following proposition holds.

Proposition 2. Let $\alpha = \gamma$, $\tau_1 = \tau_2 = 0$, $\kappa \geq \eta$, and $\Gamma_i = \Gamma_{n+1}(\theta_i - \varepsilon)$. Let $I(t)$ be any bounded, non-negative, and continuous function with $I(0) = 0$.

If $\tilde{\theta}_i \leq \theta_i - \varepsilon$, then $J_i(t) \leq 0$ for all $t \geq 0$.

Suppose $\tilde{\theta}_i > \theta_i - \varepsilon$. Then corresponding to every rise and fall in $I(t)$, $J(\alpha, I, t)$ can rise and fall at most once, and thus the following oscillations in $J_i(t)$ can occur.

(a) (Unimodal). Suppose $J(\alpha, I, t)$ continues to rise until (42) and (43) hold, but falls before (45) holds. Then $J_i(t)$ rises and falls with

$$\operatorname{sign} \frac{\mathrm{d}}{\mathrm{d}t} J_i(t) = \operatorname{sign} \frac{\mathrm{d}}{\mathrm{d}t} J(\alpha, I, t).$$

(b) (Bimodal). Suppose $\eta\tilde{\theta}_i < \kappa$ and that $J(\alpha, I, t)$ continues to rise until (43) and (45) hold. Thereafter $J_i(t)$ decreases as $J(\alpha, I, t)$ increases, and $J_i(t)$ increases as $J(\alpha, I, t)$ decreases until (45) holds again, after which $J_i(t)$ decreases towards $J_i(t) = 0$.

(c) (On–Off). Suppose $\eta\tilde{\theta}_i = \kappa$; i.e. $\tilde{\theta}_i = 1$ and $\eta = \kappa$. Let $J(\alpha, I, t)$ rise and fall as in (b). Then $J_i(t)$ rises with $J(\alpha, I, t)$ until

$$J(\alpha, I, t) \geq \Gamma_{n+1}. \tag{49}$$

For all t such that (49) holds,

$$J_i(t) = \eta\Gamma(1 - \theta_i + \varepsilon).$$

Thereafter $J_i(t)$ decreases until zero is reached. See Fig. 12.

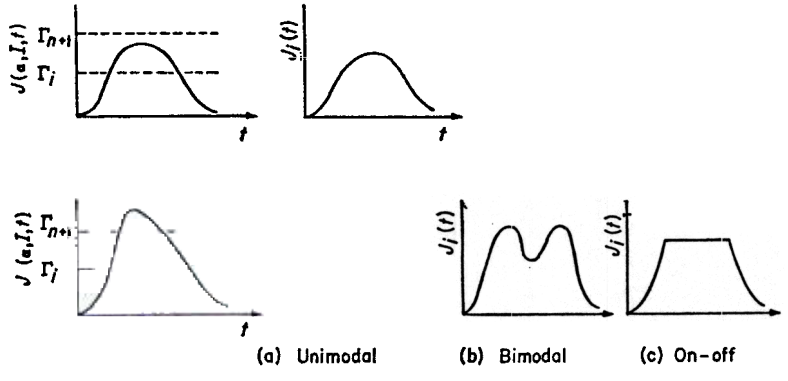


FIG. 12.

Proposition 2 is proved using the method of Lemma 3. The condition $\Gamma_i = \Gamma_{n+1}(\theta_i - \varepsilon)$ in these statements amounts to nothing more than the inequality $\Gamma_i < \Gamma_{n+1}$ written in a convenient form that shows which pattern weights can pass through.

The general case $\alpha \leq \gamma$, $\tau_1 \geq \tau_2$, $\kappa \geq \eta$, and $\Gamma_i < \Gamma_{n+1}$ requires a study of the input

$$J_i(t) = \eta[\tilde{\theta}_i J(\alpha, I, t-\tau_1) - \Gamma_{n+1}(\theta_i - \varepsilon)]^+ - \kappa[J(\gamma, I, t-\tau_2) - \Gamma_{n+1}]^+. \quad (50)$$

The condition $\tau_1 \geq \tau_2$ is needed to avoid a gap of $\tau_2 - \tau_1$ time units during which $J_i(t)$ can approach ∞ as $I \to \infty$ before inhibition sets in. The condition $\alpha \geq \gamma$ guarantees that the inhibitory potential grows no slower than the excitatory potential, and therefore will be sufficiently strong to eventually drive $J_i(t)$ to zero if $\tau_1 \geq \tau_2$. The absolute, as well as the relative, sizes of α, γ, τ_1, and τ_2 and influence the form of $J_i(t)$. If $\tau_1 = \tau_2$ but $\alpha \geq \gamma$, for example, then the inequality

$$J(\alpha, I, t-\tau_1) \leq J(\gamma, I, t-\tau_1),$$

which holds for any non-negative and continuous input I, guarantees that pattern normalization occurs. On the other hand, the difference $[J(\gamma, I, t-\tau_1) - J(\alpha, I, t-\tau_1)]$ at any time t can depend strongly on the past shape of the input, and thus $J_i(t)$ can remain non-positive even if the input is at times intense. $J_i(t)$ will better reflect instantaneous fluctuations of the input if the input is magnified and rapidly smoothed. The next proposition proves that this will happen if the input fluctuations have a certain amount of regularity; for example, if the input is the output of another cell.

Proposition 3. Given any $t \geq 0$ and $\varepsilon > 0$ suppose that there exist functions $R(\varepsilon) > 0$ and $T(t, \varepsilon) \geq 0$ such that

$$t \geq R(\varepsilon) + T(t, \varepsilon) \quad (51)$$

and

$$|I(v) - I(t)| \leq \varepsilon \quad \text{if } v \in [T(t, \varepsilon), t]. \quad (52)$$

Let $B(\alpha)$ be any non-negative and continuous function such that $\omega = \lim_{\alpha \to \infty} \alpha^{-1} B(\alpha)$ exists with $0 < \omega < \infty$. Then

$$|\alpha J(\alpha, I, t) - I(t)| \leq \|I\| e^{-\alpha R(\varepsilon)} + \varepsilon, \quad (53)$$

where $\|I\| = \sup\{|I(t)| : t \geq 0\}$, and obviously

$$\lim_{\alpha \to \infty} J(\alpha, B(\alpha)I, t) = \omega I(t). \quad (54)$$

Remarks: The function $B(\alpha)$ magnifies the input as the decay rate α increases. (53) shows that if $I(t)$ fluctuates ever more slowly, then less magnification and slower decay will suffice to keep $J(\alpha, B(\alpha)I, t)$ close to $\omega I(t)$. The simple proof is given in Appendix E.

Propositions 2 and 3 are applied to $J_i(t)$ below, where we vary $\eta = \eta(\alpha)$ and $\kappa = \kappa(\gamma)$ as functions of α and γ, respectively.

Corollary 1. Let $\alpha \geq \gamma$, $\kappa \geq \eta$, and $\Gamma_i = \Gamma_{n+1}(\theta_i - \varepsilon)$. Suppose that the finite positive limits $\mu = \lim_{\alpha \to \infty} \alpha^{-1}\eta(\alpha)$ and $\nu = \lim_{\gamma \to \infty} \gamma^{-1}\kappa(\gamma)$ exist, and let

$$0 \leq \tau_1 - \tau_2 \leq \min\left\{R\left(\frac{\varepsilon}{3}\right), R\left(\frac{\varepsilon}{3\nu}\right)\right\}. \tag{56}$$

Then for α and γ sufficiently large, $|J_i(t) - M_i(t)| \leq \varepsilon$, where $M_i(t) \leq 0$ for all $t \geq 0$ unless $\tilde{\theta}_i > \alpha\gamma^{-1}(\theta_i - \varepsilon)$. If $\tilde{\theta}_i > \alpha\gamma^{-1}(\theta_i - \varepsilon)$, $M_i(t)$ is either unimodal, biomodal, or on–off on any one oscillation of $I(t - \tau_1)$.

The proof is given in Appendix F. Corollary 1 shows that increasing the relative growth rate of inhibition to excitation increases the minimal pattern weights $\tilde{\theta}_i$ which give rise to an output signal from v_i, unless the relative threshold size $\Gamma_i\Gamma_{n+1}^{-1}$ is changed to compensate. Also decreasing the input fluctuation rate allows an increase in the maximal permissible gap between the inhibitory and excitatory time lags. The decrease in input fluctuation rate can be achieved, other things equal, by letting the input source be an ever larger cell body. In other words, an increase in the spatial scale of the input source can allow an increase in the temporal scale of the relative excitatory and inhibitory signal onset times.

11. High-band Filters

Specific inhibitory interneurons will now be used to cut off signals to the output cell when these signals become too large. It suffices to let $\Gamma_{n+1+i} = \Gamma^{(1)}$, $\Gamma_{2n+1+i} = \Gamma^{(2)}$ and $\Gamma_{3n+1+i} = \Gamma^{(3)}$ for all $i = 1, \ldots, n$. For convenience of exposition, we also introduce the notation $\tau^{(2)} = \tau_3 + \tau_5$ and $\tau^{(3)} = \tau_4 + \tau_6$. Then (30) to (32) can be written in terms of the functions

$$K_i(t) = [x_{n+1+i}(t) - \Gamma^{(1)}]^+ \tag{57}$$

and

$$L_i(t) = \sigma[\mu J(\lambda, K_i, t) - \Gamma^{(2)}]^+ - \chi[\xi J(\nu, K_i, t + \tau^{(2)} - \tau^{(3)}) - \Gamma^{(3)}]^+$$

as

$$x_{2n+1+i}(t) = \mu J(\lambda, K_i, t - \tau_3), \tag{59}$$

$$x_{3n+1+i}(t) = \xi J(\nu, K_i, t - \tau_4), \tag{60}$$

and consequently

$$x_{4n+1}(t) = J(\rho, \sum_i L_i, t - \tau^{(2)}), \tag{61}$$

if x_{2n+1+i}, x_{3n+1+i}, and x_{4n+1} have zero initial data.

Suppose we could show that each summand $L_i(t)$ in (61) will be nonpositive if K_i becomes too large. Then by choosing the threshold Γ_{4n+1} sufficiently large, the output $O_{4n+1}(t)$ in (33) would remain zero if any K_i

is too large. We therefore consider $L_i(t)$ as a functional of K_i. $L_i(t)$ can be rewritten in the form

$$L_i(t) = \eta[J(\lambda, K_i, t) - \Omega]^+ - \kappa[J(\nu, K_i, t+\tau) - \Gamma]^+$$

where $\eta = \sigma\mu$, $\kappa = \chi\xi$, $\Omega = \sigma\Gamma^{(2)}$, $\Gamma = \chi\Gamma^{(3)}$, and $\tau = \tau^{(2)} - \tau^{(3)}$. To accomplish our aim, we approximate $L_i(t)$ by an input of the form

$$N_i(t) = \eta[\lambda^{-1}K_i(t)(1-e^{-\lambda t}) - \Omega]^+ - \kappa[\nu^{-1}K_i(t)(1-e^{-\nu(t+\tau)}) - \Gamma]^+.$$

As in Proposition 3, this can be done if $K_i(t)$ has uniformly small oscillations in very small intervals, by magnifying $K_i(t)$, smoothing it quickly, and choosing τ sufficiently small. In fact, for very fast smoothing rates, $N_i(t)$ can be further approximated in sufficiently small intervals by

$$P_i(t) = \eta[\lambda^{-1}I(1-e^{-\lambda t}) - \Omega]^+ - \kappa[\nu^{-1}I(1-e^{-\nu(t+\tau)}) - \Gamma]^+,$$

i.e. by a constant input I smoothed once by the excitatory and inhibitory cells. The approximation $P_i(t)$ holds even for slow smoothing rates if the input $K_i(t)$ is of "on-off" type, as we can guarantee by Proposition 3 at the pattern normalizing stage of non-recurrent inhibition. Hence we consider the function $P_i(t)$ below.

Proposition 4. Suppose $\lambda \geq \nu$, $\Gamma > \Omega$, $\nu\Gamma > \lambda\Omega$, $\kappa \geq \eta\, e^{\nu\tau}$, and $\tau > 0$. Then $P_i(t) \leq 0$ for all $t \geq 0$ unless $I < I_0$, where

$$\left(1 - \frac{\lambda\Omega}{I_0}\right)^{1/\lambda}\left(1 - \frac{\nu\Gamma}{I_0}\right)^{-1/\nu} = e^{\tau}. \tag{62}$$

The proof is given in Appendix G.

At least partial high-band filtering is also possible if the integrals

$$\int_0^t e^{-\omega(t-v)}\, K_i(v)\, dv$$

with $\omega = \lambda$, ν are not very close to $\omega^{-1}I(1-e^{-\omega t})$, due to the following simple inequalities, which hold whenever $\lambda \geq \nu$ and $K_i(t)$ is monotone non-decreasing:

$$J(\lambda, K_i, t) \leq J(\lambda, K_i, t+\tau) \leq J(\nu, K_i, t+\tau)$$

and

$$\frac{d}{dt}J(\lambda, K_i, t) \leq \frac{d}{dt}J(\nu, K_i, t).$$

Nonetheless, a precise control of maximal output size is not readily available in the general case.

12. Discrimination of Space–Time Patterns

The above construction yields cells that can respond only to a given spatial pattern, or to patterns differing from the given one by a prescribed

error, if these patterns are presented with sufficient intensity and duration. It is now simple in principle to construct cells which respond only to a prescribed space–time pattern. Let a space–time pattern with weights

$$\theta_i(t) = I_i(t)\left[\sum_{k=1}^{n} I_k(t)\right]^{-1}$$

be given. For sufficiently small values of $\xi > 0$, the continuous function $\theta_i(t)$ can be arbitrarily well approximated by a sequence

$$\{\theta_i(k\xi): k = 0, 1, 2, \ldots, N\}$$

of its values. For each k, the weights

$$\theta(k\xi) = \{\theta_i(k\xi);\ i = 1, 2, \ldots, n\}$$

form a spatial pattern. To guarantee a good approximation of $\theta(t)$ by the patterns $\theta(k\xi)$ let ξ be chosen such that each $\theta_i(t)$ changes slowly in time intervals of length ξ. Given such a ξ, let total input intensities be specified—the "suprathreshold" intensities—which can create an output signal from a cell $v^{(k)}$ in response to the pattern $\theta(k\xi)$, and to no distinct pattern. In other words, the cells $v^{(k)}$, $k = 1, 2, \ldots, N$, divide all spatial patterns into $N+1$ classes—the N classes which are close to some pattern $\theta(k\xi)$, and the class of all the other patterns.

It is now readily seen that the output cells $v^{(k)}$, $k = 1, 2, \ldots, N$, for any finite N, can receive their inputs from the same receptive cells (or "retina") v_i, $i = 1, 2, \ldots, n$, as in Fig. 13, where we have chosen $n = N = 2$ for simplicity. In Fig. 13, each receptive cell v_{n+1+i} sends out an axon with two axon collaterals. One collateral from each v_{n+1+i} will lead towards one of the cells $v^{(k)}$. The same normalizing cell v_{n+1} can be used to normalize all the

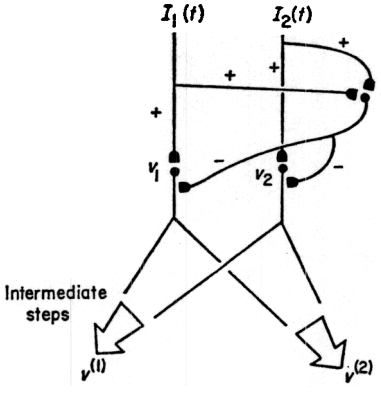


FIG. 13.

spatial patterns. This is not necessary, however; one normalizing cell can normalize several patterns, or several normalizing cells can normalize one pattern. Given a specific spatial distribution of normalizing cells, one can then determine from the thresholds of the v_{n+1+l} axon collaterals which pattern class will excite each $v^{(k)}$.

In a clear sense, this construction uses the fewest possible cells to filter spatial patterns, The n cells at the first layer characterize the level of sensory discrimination that is desired. The N cells at the last layer characterize the number of pattern classes to be discriminated. All intermediate cells can, if desired, be replaced by processes of smoothing, input additions and subtractions, and threshold cut-offs taking place on dendrites, axons, or synaptic knobs of these $n+N$ cells, perhaps at the price of violating Dale's principle.

Now that any finite number of spatial patterns $\theta(k\xi)$ can be discriminated by cells $v^{(k)}$ leading from the same receptors, discrimination of the space–time pattern $\theta(t)$ can be achieved by constructing a cell $\tilde{v}$ which produces an output signal only if all the cells $v^{(k)}$ are stimulated one after the other with a time lag of ξ. This can be done in many ways. The most direct way is to let $v^{(k)}$ send an axon to $\tilde{v}$ whose time lag is $\tau^{(k)}$. Suppose $\tau^{(k)} = \tau^{(k-1)} - \xi \geq 0$, $k = 1, 2, \ldots, N$, and use non-recurrent inhibitory interneurons to guarantee that $\tilde{v}$ fires only if it receives signals (almost) simultaneously from all cells $v^{(k)}$. This completes the construction.

Since any combination of events at the sensory periphery is a space–time pattern, any combination of events can be discriminated by an application of the above simple mechanisms. These mechanisms will not be used in precisely the given form in all sensory filters, if only due to differences in perceptual constancies between modalities, and because the above "passive" discriminations must be supplemented by "operant" discriminations in realistic behavioural interactions. Such refinements will be considered in another place, in suitable idealized cases.

13. Velocity and Orientation Detectors

To illustrate the above construction, we sketch a possible anatomy for two hypothetical cells that can respond only to lines of fixed length moving with a prescribed velocity and orientation (Fig. 14). Each dendrite in the dendritic bush D_1 receives inputs from a different "retinal" cell, and the retinal cells fall under a straight line of fixed length and orientation on the retina. Each input $I_i(t)$ sends an excitatory signal to some dendrite and a normalizing non-recurrent inhibitory signal to the region R_1 at which local spike potentials carry excitatory signals from the cell

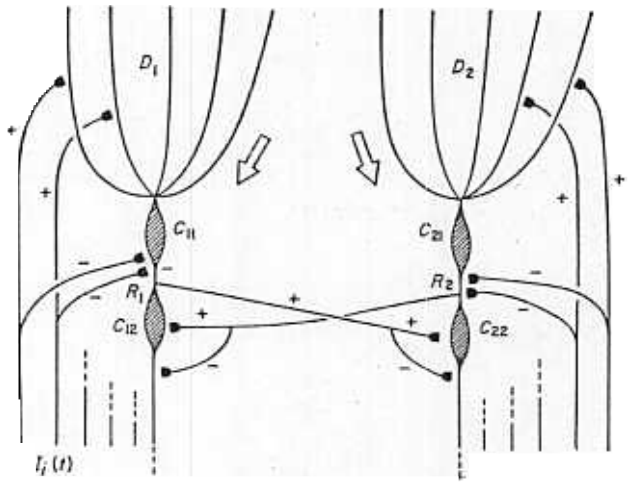


FIG. 14.

body enlargement C_{11} to C_{12}. The threshold of R_1 can thus be chosen so that a signal reaches C_{12} from C_{11} only if all dendrites in D_1 have been almost simultaneously excited, i.e. if a line of prescribed length and orientation has perturbed the retinal cells. R_1 also sends an excitatory signal to C_{22}, which is normalized at the C_{22} axon hillock by a non-recurrent inhibitory interneuron. The time lag of this signal is ξ. The threshold of C_{22} is chosen so that an output occurs only if signals from R_1 and R_2 arrive almost simultaneously at C_{22}. Suppose that the line moves from the retinal cells that perturb D_1 to those that perturb D_2 in ξ time units—say at a velocity $V(\xi)$. Then $v^{(2)}$ will fire. If the motion is from D_2 receptors to D_1 receptors, then $v^{(1)}$ will fire. No other input can fire these cells.

One can in a similar way get $v^{(1)}$ to fire in response to a stationary line, $v^{(2)}$ to fire only in response to a line moving from $D^{(1)}$ to $D^{(2)}$ receptors at velocity $V(\xi)$, and $v^{(1)}$ *not* to fire if a line moves from $D^{(2)}$ to $D^{(1)}$ at velocity $V(\xi)$ but will fire if the motion proceeds at other velocities. Only minor changes in threshold, and the use of non-recurrent inhibition from R_2 to C_{12} are needed. More complicated discriminators can be imagined with equal ease.

14. Alternative Mechanisms of Pattern Normalization: Saturating Potentials in an On–Off Field, or Logarithmic Transducers

Pattern normalization by non-specific inhibitory interneurons can formally be replaced by at least two other mechanisms. In other words, a variety of mechanisms might seek a common functional goal *in vivo*.

(A) SATURATING POTENTIALS IN AN ON–OFF FIELD

Given n cells v_i, $i = 1, 2, \ldots, n$, let the excitatory input $I_i(t)$ to v_i be delivered with equal strength as an inhibitory input to all v_j, $j \neq i$ as in Fig. 5(a). Suppose that $x_i(t)$ responds linearly to the excitatory input $I_i(t)$ when $x_i(t)$ has small values, but approaches a finite constant M if $I_i(t)$ persists with very large values in the absence of competing inhibition. Also let $x_i(t)$ respond linearly to the inhibitory couplings $-x_i(t)I_j(t)$, $j \neq i$, and let $x_i(t)$ decay exponentially to equilibrium ($= 0$) in the absence of inputs. Then the output of each v_i is normalized.

By hypothesis,

$$\dot{x}_i(t) = [M - x_i(t)]I_i(t) - x_i(t)\left[\alpha + \sum_{k \neq i} I_k(t)\right]. \tag{63}$$

Let $I_i(t) = \theta_i I(t)$, and define the variables

$$x(t) = \sum_{k=1}^{n} x_k(t)$$

$$X_i(t) = x_i(t)x^{-1}(t).$$

Then (63) yields the equations

$$\dot{X}_i(t) = A(t)[\theta_i - X_i(t)] \tag{64}$$

and

$$\dot{x}(t) = MI(t) - [\alpha + I(t)]x(t), \tag{65}$$

where $A(t) = MI(t)x^{-1}(t)$. To illustrate our main point let $I(t)$ be monotone non-decreasing with $I = \lim_{t \to \infty} I(t) > 0$, and start the system in equilibrium. Now by (64), $X_i(t)$ monotonically approaches the limit θ_i. In particular, after initial transients decay,

$$x_i(t) \sim \theta_i \frac{MI(t)}{\alpha + I(t)} \leq \frac{\theta_i M I}{\alpha + I} \leq \theta_i M,$$

showing that pattern normalization occurs.

(B) LOGARITHMIC TRANSDUCERS

This alternative is ironic in that it works well formally, is compatible with some data about logarithmic transduction from inputs to frequencies in individual cells (Granit, 1955), but seems hard to build into the interaction between cells unless linear and logarithmic transduction laws are mixed.

Suppose that the peripheral input $I_i(t)$ is logarithmically transduced when it reaches v_i, $i = 1, 2, \ldots, n$. Also let all the peripheral inputs combine linearly at a cell v_{n+1}, whereupon the total input is logarithmically transduced. Let v_{n+1} linearly inhibit each v_i. Then the output from each v_i is normalized.

By hypothesis, the transduced input to v_i is

$$\begin{aligned} \tilde{I}_i(t) &= \log I_i(t) \\ &= \log \theta_i I(t) \\ &= \log \theta_i + \log I(t). \end{aligned}$$

The total input to v_{n+1} is $\sum_{k=1}^{n} I_k(t) = I(t)$, which is logarithmically transduced yielding $\tilde{I}_{n+1}(t) = \log I(t)$. $\tilde{I}_{n+1}(t)$ then linearly inhibits v_i yielding the net potential

$$\tilde{I}_i(t) - \tilde{I}_{n+1}(t) = \log \theta_i,$$

which is normalized. The non-linearity of the transformation $\theta_i \to \log \theta_i$ does not interfere with the selectivity of the filter. It is only necessary that the transformation be one-to-one.

The preparation of this work was supported in part by the National Science Foundation (G.P. 9003), the Office of Naval Research (N00014–67–A–0204–0016) and the A. P. Sloan Foundation.

REFERENCES

Eccles, J. C. (1964). "The Physiology of Synapses." Berlin: Springer Verlag.

Eccles, J. C., Ito, M. & Szentagothai, J. (1967). "The Cerebellum as a Neuronal Machine." New York: Springer Verlag.

Frank, M. & Pfaffmann, C. (1969). *Science, N.Y.* **164**, 1183.

Granit, R. (1955). "Receptors and Sensory Perception." New Haven: Yale University Press.

Grossberg, S. (1969*a*). *J. math. Psych.* 6(2), 209.

Grossberg, S. (1969*b*). *Math. Biosci.* 4, 255.

Grossberg, S. (1969*c*). *J. theor. Biol.* 22, 325.

Grossberg, S. (1969*d*). *Math. Biosci.* 4, 201.

Grossberg, S. (1969*e*). *J. Math. Mech.* 19(1), 53.

Grossberg, S. (1969*f*). *J. stat. Phys.* 1(2), 319.

Grossberg, S. (1969*g*). *Stud. appl. Math.* XLVIII (2), 105.

Grossberg, S. (1970). *Stud. appl. Math.* (in the press).

Hubel, D. H. & Wiesel, T. N. (1965). *J. Neurophys.* **28**, 229.

Kilmer, W. L., Blum, J. & McCulloch, W. S. (1969). The Reticular Formation, Part I: Modeling Studies of the Reticular Formation, Interim Sci. Report No. 3, Division of Engineering Research, Mich. State.

Kilmer, W. L. (1969). The Reticular Formation, Part II: The Biology of the Reticular Formation. Interim Sci. Report No. 3, Division of Engineering Research, Mich. State.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S. & Pitts, W. H. (1960). *J. gen. Physiol.* 43, (2), 129.

Ratliff, F. (1965). "Mach Bands: Quantitative Studies on Neural Networks in the Retina." San Francisco: Holden-Day.

Sterling, P. & Wickelgren, B. G. (1969). *J. Neurophysiol.* **32**, 1.

Whitfield, I. C. (1967). "The Auditory Pathway." London: Edward Arnold Ltd.

Wickelgren, W. A. (1969). *Bull. math. Biophys.* **31**, 123.

APPENDIX A

Proof of Proposition 1

Suppose $x_1(\infty)$ and $x_2(\infty)$ exist. If $x_1(\infty) \leq \Gamma$, then by (3), $x_2(\infty) = 0$. But then by (2), $x_1(\infty) = \alpha^{-1}I$, which is a contradiction since $I > \alpha\Gamma$ by hypothesis. Hence $x_1(\infty) > \Gamma$ if $x_1(\infty)$ and $x_2(\infty)$ exist.

In a similar fashion, $x_1(\infty) > \Gamma$ if either limit $x_1(\infty)$ or $x_2(\infty)$ exists, since the existence of one limit clearly implies the existence of the other by integrating (2) and (3).

Suppose neither limit $x_1(\infty)$ or $x_2(\infty)$ exists. Then by the above remarks, the inequality $x_1(t) > \Gamma$ must hold at arbitrarily large times, or else by integration of (3), the existence of $x_2(\infty)$ will follow. In fact $x_1(t)$ must oscillate above Γ infinitely often: if $\dot{x}_1(t)$ has fixed sign at all large t, then $x_1(\infty)$ exists, since the solution of (2) and (3) is clearly bounded and continuous.

APPENDIX B

Proof of Lemma 2

Clearly $x_0(t)$ is monotonely concave for $t \geq 0$. Integrating (6) yields

$$x_i(t) = \beta_i \int_0^t e^{-\alpha_i(t-v)} [x_{i-1}(v-\tau_i)-\Gamma_i]^+ \, dv. \tag{B1}$$

Given any integral of the form

$$f(t) = \int_0^t e^{-\alpha(t-v)} g(v) \, dv$$

with $g(v)$ piecewise twice-differentiable and $g(0) = 0$, successive integration by parts yields

$$\dot{f}(t) = \int_0^t e^{-\alpha(t-v)} \dot{g}(v) \, dv$$

and

$$f(t) = \dot{f}(0) e^{-\alpha t} + \int_0^t e^{-\alpha(t-v)} g(v) \, dv.$$

By (B2), if $\dot{g}(v) \geq 0$ for $v \geq 0$, then $\dot{f}(0) \geq 0$ for $t \geq 0$. By (B3), $f(0) = f(0) \geq 0$. Thus if $g(v)$ changes sign at most once from non-negative to non-positive the same is true of $f(t)$. Applying (B2) and (B3) to $f(t) = x_i(t)$ and $g(t) = \beta_i[x_{i-1}(v-\tau_i)-\Gamma_i]^+$ iteratively for $i = 1, 2, \ldots, n$ proves Lemma 2.

APPENDIX C

Proof of Theorem 1

Let

$$J(t) = I_{CS}(t) - \beta[K(t) - \Omega]^+,$$

where

$$K(t) = \delta\, e^{-\gamma(t-\sigma)} \int_0^{t-\sigma} e^{\gamma v} I_{CS}(v)\, dv$$

for $t \geq \sigma$, and 0 otherwise. By (4) and (5)

$$x_1(t) = \int_0^t e^{-\alpha(t-v)} J(v)\, dv$$

since $x_1(0) = 0$.

If $I_{CS}(t)$ is asymptotically steady state, then (C3) implies

$$x_1(\infty) = \alpha^{-1} I(1 - \gamma^{-1}\beta\delta) + \alpha^{-1}\beta\Omega.$$

We must check that $x_1(\infty) < \Gamma$. Given $\beta\delta > \gamma$, this is equivalent to

$$I(\gamma^{-1}\beta\delta - 1) > \beta\Omega - \alpha\Gamma,$$

which is true for all $I > \alpha\Gamma$ if

$$\alpha\Gamma(\gamma^{-1}\beta\delta - 1) > \beta\Omega - \alpha\Gamma,$$

or $\alpha\delta\Gamma > \gamma\Omega$, as hypothesized.

Suppose $I_{CS}(t)$ is monotonely concave. To show that $\dot{x}_1(t)$ changes sign at most once from non-negative to non-positive, note by (C3) that

$$\dot{x}_1(t) = \int_0^t e^{-\alpha(t-v)} \dot{J}(v)\, dv,$$

since $x_1(0) = 0$. It thus suffices by Lemma 2 to show that $J(0) \geq 0$ and that $J(t) \leq 0$ for $t \geq t_0$ if $J(t_0) = 0$. Since $J(t) = \dot{I}_{CS}(0) \geq 0$ for $t \in [0, \sigma]$, $J(0) = \dot{I}_{CS}(0) \leq 0$. The inhibitory term is eventually positive since $I > \delta^{-1}\gamma\Omega$. Let $t_1 = \max\{t: K(t) \leq \Omega\}$. Note that for all $t \leq t_1$, $K(t) \leq \Omega$, since by (11),

$$\dot{K}(t) = \delta\, e^{-\gamma(t-\sigma)} \int_0^{t-\sigma} e^{\gamma v} \dot{I}_{CS}(v)\, dv \geq 0.$$

For $t \geq t_1$, (C1) implies

$$\dot{J}(t) = \dot{I}_{CS}(t) - \beta\delta\, e^{-\gamma(t-\sigma)} \int_0^{t-\sigma} e^{\alpha v} \dot{I}_{CS}(v)\, dv.$$

Suppose $\dot{J}(t_0) = 0$ for some t_0. Then $\dot{J}(t) \leq 0$ for $t \geq t_0$ if $\ddot{J}(t_0) \leq 0$. Since

$$\ddot{J}(t) = \ddot{I}_{CS}(t) - \beta\delta\dot{I}_{CS}(t-\sigma) + \gamma\beta\delta\, e^{-\gamma(t-\sigma)} \int_0^{t-\sigma} e^{\gamma v} \dot{I}_{CS}(v)\, dv,$$

(C4) implies

$$\ddot{J}(t_0) = \ddot{I}_{CS}(t_0) + \gamma \dot{I}_{CS}(t_0) - \beta\delta \dot{I}_{CS}(t_0 - \sigma).$$

Since $\ddot{I}_{CS}(t) \leq 0$ for all $t \geq 0$,

$$\ddot{J}(t_0) \leq \gamma(1 - \gamma^{-1}\beta\delta)\dot{I}_{CS}(t_0 - \sigma),$$

and since $\beta\delta > \gamma$ and $\dot{I}_{CS}(t) \geq 0$, $\ddot{J}(t_0) \leq 0$. Thus $\dot{x}_1(t)$ changes sign at most once towards the non-positive.

Suppose $I_{CS}(t)$ is steady state. Then (C3) becomes

$$x_1(t, I) = \alpha^{-1} I(1 - e^{-\alpha t}) - \beta\, e^{-\alpha t} \int_0^t e^{\alpha v} [x_2(v - \sigma, I) - \Omega]^+ \, dv,$$

where by (5),

$$x_2(t, I) = \gamma^{-1}\delta I(1 - e^{-\gamma t}). \tag{C6}$$

Substitution of (C6) into (C5) and integration of (C5) for $t \geq S \equiv S(I)$ yields

$$x_1(t, I) = \alpha^{-1} I(1 - e^{-\alpha t}) - \alpha^{-1}\beta(\gamma^{-1}\delta I - \Omega)(1 - e^{-\alpha(t-S)}) + \gamma^{-1}(\alpha - \gamma)^{-1}\beta\delta I\, e^{-\gamma(S-\sigma)} [e^{-\gamma(t-S)} - e^{-\alpha(t-S)}], \tag{C7}$$

where we have chosen $\alpha \neq \gamma$ to avoid the simpler case. By (C6) and the definition of S,

$$\gamma^{-1}\delta I[1 - e^{-\gamma(S-\sigma)}] = \Omega,$$

and thus

$$e^{-\gamma(S-\sigma)} = \delta^{-1} I^{-1}(\delta I - \gamma\Omega). \tag{C9}$$

Substituting (C9) in (C7), we find

$$x_1(t, I) = \alpha^{-1} I(1 - e^{-\alpha t}) + \gamma^{-1}\beta(\delta I - \gamma\Omega)[-\alpha^{-1}(1 - e^{-\alpha(t-S)}) + (\alpha - \gamma)^{-1}(e^{-\gamma(t-S)} - e^{-\alpha(t-S)})]. \tag{C10}$$

To compute dS/dI, differentiate (C9) and find

$$\frac{dS}{dI} = -I^{-1}(\delta I - \gamma\Omega)^{-1}\Omega < 0. \tag{C11}$$

To compute dT/dI, set $x_1(t, I) = \Gamma$ in (C10), let $t = T$, and differentiate with respect to I. The result, after some rearrangement of terms, is

$$\frac{dT}{dI} = AB^{-1},$$

where

$$A = -\alpha^{-1}(1 - e^{-\alpha T}) + \gamma^{-1}\beta\delta[\alpha^{-1}(1 - e^{-\alpha(T-S)}) + (\alpha - \gamma)^{-1}(e^{-\alpha(T-S)} - e^{-\gamma(T-S)})] - (\alpha - \gamma)^{-1} I^{-1}\beta\Omega[e^{-\alpha(T-S)} - e^{-\gamma(T-S)}]$$

and

$$B = I\, e^{-\alpha T} + (\alpha - \gamma)^{-1}\beta(\delta I - \gamma\Omega)[e^{-\alpha(T-S)} - e^{-\gamma(T-S)}].$$

A can be further simplified, since

$$-IA = \alpha^{-1}I(1-e^{-\alpha T})+\gamma^{-1}\beta(\delta I-\gamma\Omega)[-\alpha^{-1}(1-e^{-\alpha(T-S)})+ +(\alpha-\gamma)^{-1}(e^{-\gamma(T-S)}-e^{-\alpha(T-S)})]-\alpha^{-1}\beta\Omega(1-e^{-\alpha(T-S)}),$$

and thus by (C10)

$$A = -I^{-1}[\Gamma-\alpha^{-1}\beta\Omega(1-e^{-\alpha(T-S)})].$$

B can be simplified as well, since by (C10),

$$B = \frac{\partial}{\partial t}x_1(T,I) \tag{C15}$$

and thus $B < 0$ whenever T is defined. Letting

$$\text{sign}\,(w) = \begin{cases} 1, & w > 0 \\ 0, & w = 0 \\ -1, & w < 0, \end{cases}$$

(C12), (C14) and (C15) yield

$$\text{sign}\,\frac{dT}{dI} = \text{sign}\,[\Gamma-\alpha^{-1}\beta\Omega(1-e^{-\alpha(T-S)})]. \tag{C16}$$

In particular, $\alpha\Gamma \geq \beta\Omega$ implies $dT/dI \geq 0$. In general $(dT/dI)(I) \leq 0$ for $I \geq I_0$, if $(dT/dI)(I_0) \leq 0$, since by (C16) it suffices to show that

$$\frac{d}{dI}[\Gamma-\alpha^{-1}\beta\Omega(1-e^{-\alpha(T-S)})] \leq 0$$

if $dT/dI = 0$, or that $(d/dI)(T-S) \geq 0$ if $dT/dI = 0$, which follows from (C11). In both cases, $T(I)$ is monotonic for large I. Thus $T(\infty) \equiv \lim_{I\to\infty} T(I) \leq \infty$ exists. To show that $T(\infty) < \infty$, set $t = T$ in (C10) and let $I \to \infty$. If $T(\infty) = \infty$, this yields the contradiction $\Gamma = -\infty$, since $\beta\delta > \gamma$.

To study the dependence of $(d/dI)(T-S)$ on the sign of $\alpha-\gamma$, manipulate (C11) to (C14) to find

$$\frac{d}{dI}(T-S) = -CB^{-1},$$

where

$$C = -(\delta I-\gamma\Omega)^{-1}\Omega\, e^{-\alpha T}+I^{-1}\{\Gamma-\alpha^{-1}\beta\Omega(1-e^{-\alpha(T-S)})+ +(\alpha-\gamma)^{-1}\beta\Omega[e^{-\gamma(T-S)}-e^{-\alpha(T-S)}]\} \tag{C17}$$

and thus

$$\text{sign}\left[\frac{d}{dt}(T-S)\right] = \text{sign}\,C. \tag{C18}$$

To prove that

$$\text{sign}\left[(\alpha-\gamma)\frac{d}{dI}(T-S)(I_0)\right] \leq 0$$

implies

$$\text{sign}\left[(\alpha-\gamma)\frac{\mathrm{d}}{\mathrm{d}I}(T-S)(I)\right] \leq 0$$

for $I \geq I_0$, it suffices to show that

$$\text{sign}\left(\frac{\mathrm{d}C}{\mathrm{d}I}\right) = \text{sign}\,(\gamma-\alpha) \quad \text{if } C = 0, \tag{C19}$$

since before $(\mathrm{d}/\mathrm{d}I)(T-S)$ can change sign from ± 1 to ∓ 1, it must pass through zero, where, by (C19), it will be deflected to the sign of $\gamma-\alpha$.

To prove (C19), write the first term on the right in (C17) as

$$(\delta I-\gamma\Omega)^{-1}\Omega\,\mathrm{e}^{-\alpha T} = (\delta I-\gamma\Omega)^{-1}\Omega\,\mathrm{e}^{-\alpha(T-S)}\,\mathrm{e}^{-\alpha S}$$

and apply (C9). Then

$$(\delta I-\gamma\Omega)^{-1}\Omega\,\mathrm{e}^{-\alpha T} = (\delta I)^{-\alpha/\gamma}\Omega\,\mathrm{e}^{-\alpha\sigma}(\delta I-\gamma\Omega)^{(\alpha/\gamma)-1}\,\mathrm{e}^{-\alpha(T-S)}. \tag{C20}$$

Differentiate (C20) subject to the constraint $(\mathrm{d}/\mathrm{d}I)(T-S) = C = 0$. Then

$$-\frac{\mathrm{d}}{\mathrm{d}I}\left[(\delta I-\gamma\Omega)^{-1}\Omega\,\mathrm{e}^{-\alpha T}\right]$$

$$= \gamma^{-1}(\delta I-\gamma\Omega)^{-1}\Omega\,\mathrm{e}^{-\alpha T}\left[\alpha I^{-1}-(\alpha-\gamma)\delta(\delta I-\gamma\Omega)^{-1}\right]$$

$$= I^{-1}(\delta I-\gamma\Omega)^{-1}\Omega\,\mathrm{e}^{-\alpha T}\left[1+(\gamma-\alpha)\Omega(\delta I-\gamma\Omega)^{-1}\right].$$

Also note by (C17) that $C = 0$ implies

$$\frac{\mathrm{d}}{\mathrm{d}I}\left\{I^{-1}\left[\Gamma-\alpha^{-1}\beta\Omega(1-\mathrm{e}^{-\alpha(T-S)})+(\alpha-\gamma)^{-1}\beta\Omega(\mathrm{e}^{-\gamma(T-S)}-\mathrm{e}^{-\alpha(T-S)})\right]\right\}$$

$$= -I^{-1}(\delta I-\gamma\Omega)^{-1}\Omega\mathrm{e}^{-\alpha T}.$$

Combining these identities shows that

$$\frac{\mathrm{d}C}{\mathrm{d}I} = (\gamma-\alpha)\Omega^2\,\mathrm{e}^{-\alpha T}\,I^{-1}(\delta I-\gamma\Omega)^{-2}$$

if $C = 0$, which yields (C19).

APPENDIX D

Proof of Lemma 3

Since $I(t)$ is monotone non-decreasing,

$$\frac{\mathrm{d}}{\mathrm{d}t}J(\alpha, I, t) \geq 0.$$

Thus once either of the terms

$$R_i(t) = \left[\tilde{\theta}_i J(\alpha, I, t)-\Gamma_{n+1}(\theta_i-\varepsilon)\right]^+$$

or

$$S_i(t) = [J(\alpha, I, t) - \Gamma_{n+1}]^+$$

in $J_i(t)$ becomes positive, it remains positive.

If $\bar{\theta}_i \leq \theta_i - \varepsilon$, then $S_i(t)$ becomes positive no later than $R_i(t)$ does. Until $R_i(t) > 0$, $J_i(t) \leq 0$. Once $R_i(t)$ becomes positive,

$$\frac{d}{dt} J_i(t) = (\eta\bar{\theta}_i - \kappa) \frac{d}{dt} J(\alpha, I, t) \leq 0, \tag{D1}$$

so that $J_i(t) \leq 0$ for all $t \geq 0$ in this case.

If $\bar{\theta}_i > \theta_i - \varepsilon$, then $S_i(t)$ becomes positive before $R_i(t)$ does. In fact, $S_i(t)$ becomes positive after $t = T_1$, and $R_i(t)$ becomes positive after $t = T_2$. The equation (44) thus holds for $t \leq T_2$, whence $J_i(t)$ is non-decreasing during these times, whereas for $t > T_2$, (D1) holds. The remaining assertions of the lemma follow readily from these facts.

APPENDIX E

Proof of Proposition 3

First we prove (53). Clearly

$$\alpha[J(\alpha, I, t) - I(t)J(\alpha, 1, t)] = \alpha J(\alpha, I - I(t), t).$$

Thus

$$|\alpha J(\alpha, I, t) - I(t)| \leq ||I|| \, e^{-\alpha t} + \alpha J(\alpha, I - I(t), t). \tag{E1}$$

Denoting $T(t, \varepsilon)$ by T, note that

$$\alpha J(\alpha, I - I(t), t) < \varepsilon + \|I\| [e^{-\alpha(t-T)} - e^{-\alpha t}],$$

which along with (E1) implies (53).

By (E1),

$$|\alpha J(\alpha, \omega I, t) - \omega I(t)| \leq \omega \|I\| \, e^{-\alpha R(\varepsilon)} + \omega\varepsilon$$

for any $\omega > 0$. Given any $\delta > 0$, choose $\varepsilon \leq \delta/2\omega$ and then

$$\alpha > R^{-1}(\varepsilon) \log (2\delta^{-1}\omega \|I\|)$$

to find that

$$|\alpha J(\alpha, \omega I, t) - \omega I(t)| \leq \delta;$$

i.e.

$$\lim_{\alpha \to \infty} \alpha J(\alpha, \omega I, t) = \omega I(t).$$

Since also

$$J(\alpha, B(\alpha) I, t) = \alpha J(\alpha, \alpha^{-1} B(\alpha) I, t)$$

and clearly

$$\lim_{\alpha \to \infty} \alpha J(\alpha, \alpha^{-1} B(\alpha) I, t) = \lim_{\alpha \to \infty} \alpha J(\alpha, \omega I, t),$$

(54) follows.

APPENDIX F

Proof of Corollary 1

By proposition 3, for sufficiently large α and γ,

$$|J_i(t)-K_i(t)| \leq \frac{\varepsilon}{3},$$

where

$$K_i(t) = \eta[\tilde{\theta}_i\alpha^{-1}I(t-\tau_1)-\Gamma_{n+1}(\theta_i-\varepsilon)]^+ - \kappa[\gamma^{-1}I(t-\tau_2)-\Gamma_{n+1}]^+.$$

Define

$$M_i(t) = \eta[\tilde{\theta}_i\alpha^{-1}I(t-\tau_1)-\Gamma_{n+1}(\theta_i-\varepsilon)]^+ - \kappa[\gamma^{-1}I(t-\tau_1)-\Gamma_{n+1}]^+.$$

Since by (51), (52) and (56),

$$|I(t-\tau_1)-I(t-\tau_2)| \leq \frac{\varepsilon}{3}\min(1, \nu^{-1}),$$

for sufficiently large α and γ,

$$|K_i(t)-M_i(t)| \leq \frac{2\varepsilon}{3}.$$

We have therefore shown that $|J_i(t)-M_i(t)| \leq \varepsilon$ for sufficiently large α and γ. The properties of $M_i(t)$ as $I(t-\tau_1)$ oscillates once are easily proved as in proposition 2.

APPENDIX G

Proof of Proposition 4

First we show that $P_i(t) > 0$ for some $t \geq 0$ only if the excitatory signal arrives before the inhibitory signal does, i.e. only if there exists a time $T \leq t$ such that

$$\lambda^{-1}I(1-e^{-\lambda T}) > \Omega$$

and

$$\nu^{-1}I(1-e^{-\nu(T+\tau)}) \leq \Gamma.$$

These two equations are equivalent to the inequality $f(I) > e^{\tau}$, where

$$f(I) = \left(1-\frac{\lambda\Omega}{I}\right)^{1/\lambda}\left(1-\frac{\nu\Gamma}{I}\right)^{-1/\nu}.$$

The only other way for $N_i(t)$ to become positive is for inhibition to set in before excitation does, but not so strongly as to keep $N_i(t)$ non-positive. If this were to happen, then $N_i(t)$ would remain non-positive until the first time $t = T$ at which

$$\lambda^{-1}I(1-e^{-\lambda T}) = \Omega$$

and

$$\nu^{-1} I(1-e^{-\nu(T+\tau)}) \geq \Gamma.$$

At such a time, however, we readily find that

$$\begin{aligned} \text{sign } \dot{N}_i(t) &= \text{sign } [\eta\, e^{-\lambda T} - \kappa\, e^{-\nu(T+\tau)}] \\ &\leq \text{sign } [\eta - \kappa\, e^{-\nu\tau}] \leq 0, \end{aligned}$$

whence $N_i(t)$ is always non-positive. Thus $N_i(t) > 0$ for some $t \geq 0$ only if $f(I) > e^{\tau}$.

It is readily checked that for $I \geq \lambda\Omega$,

$$\text{sign } \dot{f}(I) = \text{sign } [\lambda\Omega - \nu\Gamma] < 0.$$

Thus if there exists an $I = I_0$ such that $f(I_0) = e^{\tau}$, as in (62), then $f(I) > e^{\tau}$ only if $I < I_0$. Such an I_0 exists, since $f(\nu\Gamma) = \infty$, $f(\infty) = 1$, and $\tau > 0$.