Temporal Ordering in Neural Networks

J. F. Walsh

November 19, 1988

Abstract

Current theories of neural networks do not provide a mechanism to incorporate the time-ordering of stimuli into the structure of a neural network as a natural part of the network learning and recall process. This is a serious limitation of these models, both from the point of view of their application to problems which deal with temporally ordered stimuli, such as speech and motion recognition, and with respect to their ability to model biological brain behavior. This paper describes a biologically motivated three-dimensional neural network which associates and stores temporal order information with the applied stimuli in the network as an integral part of the learning process. Temporal ordering information is stored primarily in the propagation delay of neuronal signals. In the mammalian brain, such delays may result from interneurons or from the geometry of the axonic and dendritic processes and their synapses. The model can be used to develop either a self-organizing network, or an associative network through the incorporation of pre-synaptic excitation and inhibition.

1 Introduction

This paper describes a three-dimensional neural network, modeled on the one found in the human brain, which has the ability to store the temporal ordering information associated with learned stimuli in a natural fashion. The network is self-organizing, but may be used as an associative memory by incorporating a "reticular activating system" which operates through the mechanism of pre-synaptic inhibition and excitation.

This work is based on the following assumptions:

1. Neurons are distributed over n layers, numbered from $k = 1, \ldots, n$, each layer consisting of N_k neurons.
2. The time it takes the signal that a neuron has fired to transit from the cell body of the initiating neuron into the cell body of the receiving neuron where it may be summed with other inputs, is highly variable and of the

same order as the neuronal firing time. In the mammalian brain, this "propagation time" is affected by the synaptic delay time, the presence of interneurons, the propagation time of a signal through axon and dendrites, membrane permeability, and the geometry of the synapses on the cell body which lead to temporal affects associated with spatial summation.

(Discuss each point in detail and the evidence to support it w/i refs.)

2 Analysis

We start by introducing and generalizing some notation introduced by Kohonen to describe a self-organizing two-dimensional neural network. Kohonen's model differs from ours in that it does not encode temporal ordering information, and deals with a two-dimensional array of neurons, in contrast to our three-dimensional array. Let

$$\eta_i{}^k(t)$$

represent the frequency of firing of the ith neuron on level k at time t. We will also use the notation

$$(i, k)$$

to denote the ith neuron on level k. In other words, (i, k) refers to the neuron itself, and $\eta_i{}^k(t)$ to the rate of firing of that neuron at time t.

We make the convention that the levels k are numbered $1, \ldots, n$, each level consisting of N_k neurons. The neurons themselves on each level k will be numbered $i = 1, \ldots, N_k$. We make the further convention here that "level" $k = 0$ corresponds to the input stimuli. In other words,

$$\eta_i{}^0(t)$$

gives the time course of the ith stimulus to the net. Conceptually, the stimulus vectors are viewed as originating from the 'zero-th' level of the net. This convention allows a convenient and uniform notation, and provides a natural means for describing feedback from the network to the source of the network stimulus, as well as for describing associations developed between "abstractions" formed in different areas of the network, as we will discuss in connection with the "reticular activating system" below.

We assume that each "target neuron" (i, k) is connected to a set of "source neurons" $\{(j, l)\}$. Since the set of neurons is richly interconnected, which neuron is regarded as the "target" for the purpose of this discussion is arbitrary. Neurons whose outputs are connected to the chosen "target" are then classified as source neurons with respect to that target.

Let us suppose that a target neuron (i, k) receives a certain set of inputs at a given observation time t. Because of inter-neuron propagation delays, these signals $\{\eta_j{}^l\}$ would have to originate in the cell bodies of their respective source neurons $\{(j,l)\}$ at some earlier time, say at time $t-\alpha$ for a particular source neuron. In fact, in the mammalian brain, there are many connections between a given source and target neuron pair; in some cases, thousands of such connections. There is no reason to assume *a priori* that signals are propagated through each of these connections with the same time delay since, in general, the paths followed by the respective signals will be completely different even though they share a common origin and destination neuron. We must therefore make allowance for the possibility that a given target neuron (i,k) will be affected by the output which a source neuron had at various times in the past; perhaps even a (semi-)continuous range of such times.

We use the symbol

$$\mu_{ij}{}^{kl}(\alpha)$$

to denote the strength of the coupling of the output of the jth neuron on level l at time $t-\alpha$ to the input of the ith neuron on level k at time t. In other words, if a signal leaves the cell body of source neuron (j,l) at a time $t-\alpha$, its influence on the target neuron (i,k) at time t will be weighted by a factor of $\mu_{ij}{}^{kl}(\alpha)$. The coupling $\mu_{ij}{}^{kl}(\alpha)$ may be positive or negative depending on whether the effect of the source neuron (j,l) on the on the target neuron (i,k) is excitatory or inhibitory, respectively.

As a final bit of notation, we use the symbol

$$\epsilon_i{}^k$$

to denote the amount of time it takes a set of impulses summed in the cell body of the ith neuron on level k to initiate firing of the neuron. In other words, a signal present on the inputs of a neuron at a time t is assumed to affect the output of the neuron not instantaneously, but at a later time $t+\epsilon_i{}^k$.

Using this notation and, for the time being, neglecting the effect of any non-linearity between target neuron input and output, as well as several other details in the interest of clarity, we have

$$\eta_i{}^k(t+\epsilon_i{}^k)=\sum_{l=0}^{n}\sum_{j=1}^{N_l}\int_{\alpha_{min}}^{\alpha_{max}}\mu_{ij}{}^{kl}(\alpha)\eta_j{}^l(t-\alpha)\,d\alpha \tag{1}$$

In other words, the current output of a given neuron is determined by the outputs of all the neurons in the network at various times in the past. These times are indexed by the time delay factor α, which must be non-negative since current output depends on the past history of the net, not its future behavior. The parameter $\mu_{ij}{}^{kl}(\alpha)$ is determined as part of the network's self-organizing or associative learning process as the time-dependent stimuli to be learned are applied.

Equation 1 is a modification of the widely-used "linear adder" model of the neuron, where the output of the neuron is a function of the weighted sum of the inputs to the neuron. We have included an integration over the time-delay parameter α in order to allow the present behavior of the target neuron to be influenced by the past behavior of the source neurons to which it is connected. Such an integration would appear to be a prerequisite to any system capable of learning time sequences of events. As discussed above, a time delay could be implemented biologically by incorporating a propagation delay between the source and target neurons. In practice, we expect that biological behavior may be more closely modeled by a discrete sum over α rather than by the integral expression given here. We will find it convenient and suggestive, however, to retain this integral formulation in our subsequent analysis.

Note that equation 1, except for the limits of integration, is a convolution. If we make the convention that

$$\mu_{ij}^{kl}(\alpha) = 0, \qquad \alpha < \alpha_{min}, \qquad \alpha > \alpha_{max},$$

where α_{min} and α_{max} are the minimum and maximum time delays, respectively, to which a signal may be subjected in the system under discussion, then we can write

$$\begin{aligned} \eta_i^k(t+\epsilon_i^k) &= \sum_{l=0}^{n}\sum_{j=1}^{N_l}\int_{-\infty}^{+\infty}\mu_{ij}^{kl}(\alpha)\eta_j^l(t-\alpha)\,d\alpha \qquad (2) \\ &= \sum_{l=0}^{n}\sum_{j=1}^{N_l}\mu_{ij}^{kl}(t) * \eta_j^l(t) \qquad . \end{aligned}$$

where we have used the symbol "$*$" to denote the convolution operation.

We will find it convenient in our subsequent discussion to assume that, instead of the neurons' rate of firing $\eta_j^l(t)$ itself, it is the departure of the neuron from its basal rate of firing which is influenced by other neurons and, in turn, influences them. In other words, we will find it expedient to speak in terms of $(\eta_j^l(t) - \bar{\eta}_j^l)$ rather than $\eta_j^l(t)$. This is equivalent to letting the $\eta_j^l(t)$ in the above equations assume positive and negative values. Since negative firing frequencies are unphysical (phase being irrelevant to frequency coding), we choose to speak in terms of $(\eta_j^l(t) - \bar{\eta}_j^l)$ and allow $\eta_j^l(t)$ itself to remain positive definite. Biologically, we find that neurons do, in fact, maintain a non-zero, basal rate of firing, so this assumption is biologically plausible.

With this convention, we can write our state equation as

$$(\eta_i^k(t+\epsilon_i^k) - \bar{\eta}_i^k) = \sum_{l=0}^{n}\sum_{j=1}^{N_l}\int_{\alpha_{min}}^{\alpha_{max}}\mu_{ij}^{kl}(\alpha)(\eta_j^l(t-\alpha) - \bar{\eta}_j^l)\,d\alpha \qquad (3)$$

Note that the value of $(\eta_i^k(t+\epsilon_i^k) - \bar{\eta}_i^k)$ on the left-hand-side of this equation can be made arbitrarily large by increasing the "norm" of $(\eta_j^l(t-\alpha)-\bar{\eta}_j^l)$ in the integral. By "norm" here, I mean the length of the "vector" $\{\eta_j^l(t-\alpha)-\bar{\eta}_j^l\}_{j,l,\alpha}$ where the subscript "j, l, α" is used to indicate that the "components" of this "vector" are indexed by these quantities. That is, a "component" of the "vector" $\{\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l\}_{j,l,\alpha}$ is indexed by specifying definite values of j, l, and α. In fact, the "j, l, α"th "component" of this "vector" is just $(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)$.

We have placed the words "norm", "vector" and "component" in quotes because α is a continuous quantity and so these "vectors" actually have uncountably infinitely many components. Such objects are dealt with routinely in Quantum Mechanics, for example, and present no real analytical difficulties. Henceforth we will omit the quotes and simply refer to such sets as vectors.

A variety of norms can of course be defined on our vector space, and the choice of norm in no way affects the validity of our analysis. One norm which is convenient analytically and which we shall show to be required if $\mu_{ij}{}^{kl}(\alpha)$ is to become a matched filter is

$$norm(\{\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l\}_{j,l,\alpha}) = \sqrt{\sum_{l=0}^{n}\sum_{j=1}^{N_l}\int_{\alpha_{min}}^{\alpha_{max}} (\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha} \tag{4}$$

To see that this norm is required for $\mu_{ij}{}^{kl}(\alpha)$ to become a matched filter, we note that, if $\mu_{ij}{}^{kl}(\alpha)$ is a matched filter for the detection of a signal $\{\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l\}_{j,l,\alpha}$ at time t, then $\mu_{ij}{}^{kl}(\alpha) \propto \eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l$. Using A to denote the constant of proportionality, we have by substituting for $\mu_{ij}{}^{kl}(\alpha)$ in equation 3

$$\eta_i{}^k(t+\epsilon_i{}^k)-\bar{\eta}_i{}^k = A\sum_{l=0}^{n}\sum_{j=1}^{N_l}\int_{\alpha_{min}}^{\alpha_{max}} (\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha$$

which implies

$$\mu_{ij}{}^{kl}(\alpha) = \frac{(\eta_i{}^k(t+\epsilon_i{}^k)-\bar{\eta}_i{}^k)(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)}{\sum_{l=0}^{n}\sum_{j=1}^{N_l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}$$

where here $\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l{}_{j,l,\alpha}$ is the signal to be recognized. The square root sign is introduced when we divide up the normalising factor between equation 3 and our definition for $\mu_{ij}{}^{kl}(\alpha)$.

In the following discussion, we will use the norm defined in equation 4 for the sake of concreteness, but as pointed out above nothing in our subsequent analysis relies on or limits us to this choice of norm.

We have then that

$$\begin{aligned}
\eta_i{}^k(t+\epsilon_i{}^k)-\bar{\eta}_i{}^k &= \sum_{l=0}^{n}\sum_{j=1}^{N_l}\int_{\alpha_{min}}^{\alpha_{max}} \mu_{ij}{}^{kl}(\alpha)\frac{(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)}{norm(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)}\,d\alpha \qquad (5)\\
&= \sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}} \mu_{ij}{}^{kl}(\alpha)\frac{(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}\,d\alpha
\end{aligned}$$

This is the equation we will use in our subsequent analysis. Note that we are still neglecting the effect of any sigmoidal or other non-linearity between neuron input and output response.

The presence of a convolution operation in the fundamental equation describing time-dependent neuronal activity in our network calls to mind an analogy with Optical Image Processing. If we were to encounter a similar formula in optical image processing—which is done over the space domain x rather than the time domain t—the variable $(\eta_j^l(x) - \bar{\eta}_j^l)$ would play the role of the "object", $\mu_{ij}^{kl}(x)$ would specify the transfer function of the lens or other optical processing system illuminated by the "object", and $(\eta_i^k(x + \epsilon_i^k) - \bar{\eta}_i^k)$ would be the "image" formed of this object by the system, displaced by the amount ϵ_i^k. The summation over j and l in the right hand side of equation 3 means the image formed is a composite of several objects each passing through their own "optical processing system". In fact, the weighting coefficients μ_{ij}^{kl} have been referred to in the literature for some time as a "transfer function", but to my knowledge the precise sense in which $\mu_{ij}^{kl}(\alpha)$ serves as a transfer function has not been stated because the treatment of the interneuronal propagation delays α discussed here is novel.

Since we are interested in employing our neural network to 'recognize' when a particular sequence of temporally-ordered stimuli is applied to it, and because of the formal similarity between this network and optical image processing noted above, it is natural to examine optical image processing schemes for image recognition. One of the most elegant and successfully employed optical image recognition schemes is called "Complex Matched Spatial Filtering", and employs Fourier Transform Holography. In this technique, a filter is formed which records the optical interference pattern between an illuminated transparency of the object to be recognized and a second transparency which contains an image of the desired recognition signal. Generally, but not essentially, the desired recognition signal is chosen to be a point source—a spatial "delta" function. When the filter so formed is illuminated by the object to be recognized, an image is formed of the recognition signal. Illumination of the filter by objects different from the one used to create the filter yields, in general, a diffuse, unfocused image. This technique works because when an interference pattern is illuminated by one of the interfering objects, the phase of the light transmitted by the pattern is identical to that which would issue from the missing object.

The model of optical complex matched spatial filtering suggests that, in order to devise a recognition scheme, we should search for an analog to "interference" in the context of our neural network. Once the notion of interference in a neural network is defined, a "complex matched *temporal* filter" can be created by forming the interference pattern between the signals to be recognized—the set of time-ordered stimuli $(\eta_j^l(t - \alpha) - \bar{\eta}_j^l)$—and the desired recognition signal, which we choose to be the triggering of the target neuron (i, k) at time $t + \epsilon_i^k$. This complex interference pattern can then be realized in the transfer function $\mu_{ij}^{kl}(\alpha)$. The set of axons, dendrites and synapses which connect the two neurons, whose connection strength are specified by $\mu_{ij}^{kl}(\alpha)$, will serve as a complex matched temporal filter between the source neuron (j,l) and the target neuron (i,k). We will see that the interference pattern so generated is real-valued, with positive and negative coefficients.

I propose that we use the concept of the "correlation" of neuronal firings to define interference phenomena. The correlation idea—but not its analogy to optical interference—was pioneered by Hebb and is frequently cited. We propose adopting Hebbian learning and using it as our model for interference in the following sense:

1. If the firing of a source and target neuron are correlated, accounting for propagation delay, then their "interference" is said to be constructive. In other words, if $(\eta_j{}^l(t-\alpha) - \bar{\eta}_j{}^l)$ and $(\eta_i{}^k(t+\epsilon_i{}^k) - \bar{\eta}_i{}^k)$ have a positive correlation in time, then they are said to "interfere constructively" and, in our recognition scheme where μ_{ij}^{kl} is used to record the interference pattern between the source and target neurons, $\mu_{ij}^{kl}(\alpha)$ is positive. This is the case where, over time, both neurons either tend to fire at an above-average rate simultaneously, or at a below-average rate simultaneously. This is the analog, optically, of the crests or the troughs of two waves coinciding at the plane where the interference pattern is measured.

2. If the firing of a source and target neuron are anticorrelated, accounting for propagation delay, then their "interference" is said to be destructive. In other words, if $(\eta_j{}^l(t-\alpha) - \bar{\eta}_j{}^l)$ and $(\eta_i{}^k(t+\epsilon_i{}^k) - \bar{\eta}_i{}^k)$ are negatively correlated in time, $\mu_{ij}^{kl}(\alpha)$ will be negative. This is the case where, over time, when one neuron fires at an above-average rate, the other tends to fire at a below-average rate. This is the analog, optically, of the crest of one wave coinciding with the trough of another wave.

3. If there is no correlation between the firing of a source and target neuron, accounting for propagation delay, then the result of their "interference" will be some intermediate value. In other words, if $(\eta_j{}^l(t-\alpha) - \bar{\eta}_j{}^l)$ and $(\eta_i{}^k(t+\epsilon_i{}^k) - \bar{\eta}_i{}^k)$ have no correlation in time, then $\mu_{ij}^{kl}(\alpha)$ will be near zero. Optically, this is the analog of some part of two waves which are neither crests or troughs coinciding.

Mathematically, we can express the "interference" relationship embodied in $\mu_{ij}^{kl}(\alpha)$ as

$$\mu_{ij}^{kl}(\alpha) = C\int_{-\infty}^{+\infty} (\eta_i{}^k(t+\epsilon_i{}^k) - \bar{\eta}_i{}^k)\frac{(\eta_j{}^l(t-\alpha) - \bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}} (\eta_j{}^l(t-\alpha) - \bar{\eta}_j{}^l)^2\, d\alpha}}\, dt \tag{6}$$

where the constant of proportionality C is to be determined.

Intriguingly, we can actually arrive at equation 6 and compute the value of our constant of proportionality C without reference to the idea of interference or Hebbian learning. We can do this solely by operating mathematically on our state equation, equation 6. If we multiply both sides of equation 6 by

$$\frac{(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}$$

and integrate with respect to t between $-\infty$ and $+\infty$, we obtain

$$\int_{-\infty}^{+\infty}(\eta_i{}^k(t+\epsilon_i{}^k)-\bar{\eta}_i{}^k)\frac{(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}\,dt =$$

$$\sum_{p=0}^{n}\sum_{q=1}^{N_p}\int_{\alpha_{min}}^{\alpha_{max}}\mu_{ip}{}^{kq}(\beta)$$

$$\cdot\int_{-\infty}^{+\infty}\frac{(\eta_p{}^q(t-\beta)-\bar{\eta}_p{}^q)}{\sqrt{\sum_{p,q}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_p{}^q(t-\beta)-\bar{\eta}_p{}^q)^2\,d\beta}}\frac{(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}\,dt\,d\beta$$

Here we've let the dummy variables on the RHS of equation 6 be p, q and β instead of j, l and α as stated in the original equation. We have also assumed that $\mu_{ij}{}^{kl}(\alpha)$ is not a function of time; that is, that the network is mature and no learning is taking place. We will relax this assumption in the section on training below, but for now we will focus on mature network values.

Notice that if we make the assumption that

$$\int_{-\infty}^{+\infty}\frac{(\eta_p{}^q(t-\beta)-\bar{\eta}_p{}^q)}{\sqrt{\sum_{p,q}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_p{}^q(t-\beta)-\bar{\eta}_p{}^q)^2\,d\beta}}\frac{(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}\,dt \propto \delta_{pj}{}^{ql}\,\delta(\alpha-\beta) \quad (7)$$

that is, if we assume that all of the input "channels" are uncorrelated, then equation 7 becomes

$$\int_{-\infty}^{+\infty}(\eta_i{}^k(t+\epsilon_i{}^k)-\bar{\eta}_i{}^k)\frac{(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}\,dt =$$

$$\mu_{ij}{}^{kl}(\alpha)\int_{-\infty}^{+\infty}\left(\frac{(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}\right)^2 dt. \quad (8)$$

This implies that

$$\mu_{ij}{}^{kl}(\alpha) \;=\; \left[\int_{-\infty}^{+\infty}\left(\frac{(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}\right)^2 dt\right]^{-1} \quad (9)$$

$$\cdot \int_{-\infty}^{+\infty} (\eta_i{}^k(t+\epsilon_i{}^k) - \bar{\eta}_i{}^k) \frac{(\eta_j{}^l(t-\alpha) - \bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l} \int_{\alpha_{min}}^{\alpha_{max}} (\eta_j{}^l(t-\alpha) - \bar{\eta}_j{}^l)^2 \, d\alpha}} \, dt$$

which is exactly our equation 6 with the constant of proportionality defined.

It is impossible for a physical function $(\eta_j{}^l(t) - \bar{\eta}_j{}^l)$ to meet the exact requirements of equation 7 in terms of having an autocorrelation function which is a Dirac delta function; to do so would require infinite energy. In practice, then, we require that $(\eta_j{}^l(t) - \bar{\eta}_j{}^l)$ have a time autocorrelation function which is "sharply peaked". This is a requirement on optical transparencies for successful optical image recognition as well, so it is not surprising that it should appear in the context of our temporally ordering neural network recognition scheme as well. One way this requirement can be met is by requiring that neural activity be "sparse"; that is, than an individual neuron not fire at a rate significantly different from its basal rate very often. This requirement is, interestingly, met by biological neural networks.

Equation 10, together with equation 6, form our final system of equations. These equations are valid provided that the input channels to the network are uncorrelated, in the sense described by equation 7. These equations describe the neural network analog of an optical fourier transform holography complex matched filter, and describe a network which has the property of associating a specific target neuron (i,k) with a given time-ordered set of stimuli $\{\eta_j{}^l(t)\}$. In the remainder of this paper, we shall refer to a network which obeys the above equations as a "timenet". We shall discuss the applications of timenets below.

3 Training the Network

The above equations for a "timenet" (equations 6 and 10) describe a mature, fully developed network. In practical applications—such as, perhaps, a mammalian brain—learning will take place and the values of the weight coefficients $\mu_{ij}{}^{kl}(\alpha)$ will alter in time.

We can imagine the network being trained in two ways, each of which will be discussed in more detail below:

1. A self-organizing approach where pre-existing, randomly assigned connections, implemented as values of $\mu_{ij}{}^{kl}(\alpha)$, are either made more positive or more negative depending on whether a given source neuron tends to excite or inhibit the firing of a target neuron.

2. An associative approach where a selected target neuron is excited by an outside agency at an appropriate time. Connections then "grow" toward the target neuron from source neurons whose firings are correlated or anti-correlated with that of the target neuron in time, allowing for propagation delays.

It is the latter mechanism that I believe is responsible for forming early connections in the brain, with the self-organizing approach, perhaps, taking over at a later stage of development. In the associative approach, the "outside agency" used to excite a target neuron could be from another level of processing, *e.g.*. For example, in a baby learning to speak, the centers of the brain responsible for muscle movements could trigger a "target" neuron in the speech center when a particular muscle group is activated. In this manner, the receptor neurons triggered by the baby listening to his own speech would become associated with the muscle movements needed to produce that speech sound—an essential first step to speech recognition. Self-organizing behavior might then be responsible for correlating speech sounds produced by other persons with the child's own muscle movements.

In order to develop a "law" by which the network can "learn", we are subject to the constraint that to be biologically plausible, any such learning law must be "local" in both space and time. In other words, any law we choose to alter the value of the weight coefficients cannot presuppose anything which would not be "known" by a particular neuron at a particular time. For example, knowledge of the instantaneous state of the network as a whole or of stimuli to be received in the future cannot be components of a local law. Only the state of the target neuron or other information which is available to it at a given moment of time may be incorporated into a local learning law.

Let us use the expression $\mu_{ij}{}^{kl}(\alpha; t)$ to denote the value of $\mu_{ij}{}^{kl}(\alpha)$ at time t. Evidently

$$\mu_{ij}{}^{kl}(\alpha;t) = \left[\int_{\alpha_{min}}^{t}\left(\frac{(\eta_j{}^l(t'-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t'-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}\right)^2 dt'\right]^{-1} \cdot \int_{\alpha_{min}}^{t}(\eta_i{}^k(t'+\epsilon_i{}^k)-\bar{\eta}_i{}^k)\frac{(\eta_j{}^l(t'-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t'-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}\,dt' \tag{10}$$

converges to the mature value of $\mu_{ij}{}^{kl}(\alpha)$ as specified in equation 10 in the limit $t \to \infty$, if we make the convention $(\eta_j{}^l(t)-\bar{\eta}_j{}^l) = 0$ for $t < 0$.

Equation 15 is local in the sense discussed above. The only quantities which need be known by the target neuron (i, k) are the values of the inputs to and outputs of the neuron. In other words, each point at which a neuron is connected to a stimulus $(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)$ —we term such points "synapses"—must 'know' simultaneously the value $(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)$ of input to that synapse and the response $(\eta_i{}^k(t+\epsilon_i{}^k)-\bar{\eta}_i{}^k)$ of the neuron. The neuron as a whole must also have the capability of globally normalizing its inputs—as a biological neuron could do by having a fixed number of receptor sites, e.g., and allocating these sites to the inputs in proportion to their relative magnitudes. The susceptibility of a given synapse to further modification decreases in proportion to the variance of the input stimulus $(\eta_j{}^l(t-\alpha)$ about the basal rate of firing $\bar{\eta}_j{}^l$ of the source neuron. In other words, the more a given synapse is stimulated, the less sensitive that synapse becomes to learning. Each synapse must therefore have some way of 'remembering' the total amount of stimulation it has received, and decreasing its 'plasticity' in response.

Notice that our state equation—equation 6—can be interpreted as the "inner product" of the two vectors $\{\mu_{ij}{}^{kl}(\alpha)\}_{j,l,\alpha}$ and $\{(\eta_j{}^l(t-\alpha) - \bar{\eta}_j{}^l)\}_{j,l,\alpha}$ (normalized), provided that we define the inner product as

$$(v_j{}^l(\alpha), w_j{}^l(\alpha)) \equiv \sum_{l=0}^{n} \sum_{j=1}^{N_l} \int_{\alpha_{min}}^{\alpha_{max}} v_j{}^l(\alpha) w_j{}^l(\alpha)\, d\alpha.$$

This interpretation of our state equation suggests the picture that the weight vector $\{\mu_{ij}{}^{kl}(\alpha)\}_{j,l,\alpha}$ defines a fixed direction in "signal space" which completely characterizes a target neuron (i,k). A particular signal being applied to the network is specified by a point $\{(\eta_j{}^l(t-\alpha) - \bar{\eta}_j{}^l)\}_{j,l,\alpha}$ (properly normalized) in signal space. Because such signals are normalised, each signal can be represented by a point on the unit sphere in signal space. The output of a given target neuron (i,k) will then simply be the magnitude of the vector $\{\mu_{ij}{}^{kl}(\alpha)\}_{j,l,\alpha}$ times the cosine of the angle between the weight vector characterising the target neuron and the vector representing the signal being applied.

In other words, the weight vectors characterising the different neurons of the network form the basis for a mapping from "signal space" to the "space" of neurons. A given signal being applied to the network is decomposed by the network into its components with respect to the weight vectors characterizing the neurons. The output of a neuron (i,k) is simply the component of the incoming signal vector which lies in the direction of the weight vector $\{\mu_{ij}{}^{kl}(\alpha)\}_{j,l,\alpha}$. The application of a signal to a "timenet" such as we describe is therefore equivalent to a spectral decomposition, or change-of-basis operation in linear algebra, with the resulting basis having reduced dimensionality. The reduction in dimensionality comes about because each neuron integrates a range of time delays.

With this picture in mind, we suggest an associative training procedure. In order to associate a particular set of output values $\{(\eta_i{}^k(t+\epsilon_i{}^k) - \bar{\eta}_i{}^k)\}_{i,k}$ with a given normalized signal $\{(\eta_j{}^l(t-\alpha) - \bar{\eta}_j{}^l)\}_{j,l,\alpha}$ we force neuron (i,k) to have the value at time t'

$$\begin{aligned}(\eta_i{}^k(t'+\epsilon_i{}^k) - \bar{\eta}_i{}^k) \quad = \quad & (\eta_i{}^k(t+\epsilon_i{}^k) - \bar{\eta}_i{}^k) \\ & \cdot\left(\frac{(\eta_j{}^l(t'-\alpha) - \bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}} (\eta_j{}^l(t'-\alpha) - \bar{\eta}_j{}^l)^2\, d\alpha}}, \right. \\ & \left. \frac{(\eta_j{}^l(t-\alpha) - \bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}} (\eta_j{}^l(t-\alpha) - \bar{\eta}_j{}^l)^2\, d\alpha}}\right) \qquad (11)\end{aligned}$$

where $\{(\eta_j{}^l(t'-\alpha) - \bar{\eta}_j{}^l)\}_{j,l,\alpha}$ is the signal applied at that time. Notice that if we follow this training procedure, $\mu_{ij}{}^{kl}(\alpha)$ will become a matched filter for the recognition of the normalised signal $\{(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)\}_{j,l,\alpha}$. We can see this by substituting for $(\eta_i{}^k(t'+\epsilon_i{}^k)-\bar{\eta}_i{}^k)$ in equation 15:

$$\mu_{ij}{}^{kl}(\alpha;t) = \left[\int_{\alpha_{min}}^{t}\left(\frac{(\eta_j{}^l(t'-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t'-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}\right)^2 dt'\right]^{-1} \tag{12}$$

$$\cdot\int_{\alpha_{min}}^{t}(\eta_i{}^k(t'+\epsilon_i{}^k)-\bar{\eta}_i{}^k) \tag{13}$$

$$\cdot\frac{(\eta_j{}^l(t'-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t'-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}\,dt'$$

$$= \left[\int_{\alpha_{min}}^{t}\left(\frac{(\eta_j{}^l(t'-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t'-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}\right)^2 dt'\right]^{-1}$$

$$\cdot\sum_{p,q}\int_{\alpha_{min}}^{\alpha_{max}}\frac{(\eta_p{}^q(t-\beta)-\bar{\eta}_p{}^q)}{\sqrt{\sum_{p,q}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_p{}^q(t-\beta)-\bar{\eta}_p{}^q)^2\,d\beta}}$$

$$\cdot\int_{\alpha_{min}}^{t}\frac{(\eta_p{}^q(t'-\beta)-\bar{\eta}_p{}^q)}{\sqrt{\sum_{p,q}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_p{}^q(t'-\beta)-\bar{\eta}_p{}^q)^2\,d\beta}} \tag{14}$$

$$\cdot\frac{(\eta_j{}^l(t'-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t'-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}}\,dt'\,d\beta$$

$$= (\eta_i{}^k(t+\epsilon_i{}^k)-\bar{\eta}_i{}^k)\frac{(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)}{\sqrt{\sum_{j,l}\int_{\alpha_{min}}^{\alpha_{max}}(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)^2\,d\alpha}} \tag{15}$$

where we have used equation 7 above. This, of course, is the definition of a matched filter for the recognition of the normalized signal $\{(\eta_j{}^l(t-\alpha)-\bar{\eta}_j{}^l)\}_{j,l,\alpha}$ weigted by the desired value of $eta_i{}^k$ at time $t+\epsilon_i{}^k$, which confirms the validity of this procedure.

4 Experimental Results