Psychol. Belg.
1993, 33-2, 255-269.

NEURAL NETWORKS AND THE REPRESENTATION OF TIME

R.C. MIALL

University of Oxford
Laboratory of Physiology, U.K.

The neurons within artificial neural networks have rather limited temporal behaviour, and the most common mechanisms used to improve their representation of temporal events are simple delay lines or recurrent feedback connections. While there are problems in the match between these properties and the neurophysiological realities, many of these problems are overcome by the benefits of multi-neuron networks. A simple network model is described, more as an example of the network approach than as a serious model of timing. However, it demonstrates that even the most crude network of neurons could be used to count or to accumulate the output of an internal clock.

INTRODUCTION

How is time encoded in the brain? The answer or answers are far from clear, so the aim of this paper is to try to address the question from a different viewpoint. It asks what techniques are used to encode time in artificial neural networks and how plausible are these techniques when mapped onto biological nervous systems. There are many scales at which temporal information affects the brain, and the mechanisms of these clearly differ. Neither very short nor very long durations can easily be encoded by neural systems.

At the finest, sub-millisecond scale, accurate timing is beyond the scope of neurons. This is because the mechanisms for transmission of information along axons and across synapses are limited by membrane and channel dynamics. Hence there is a temporal boundary where signals become lost in the temporal noise of the membrane or the synapse. Signal-to-noise ratio improves if the same signal can be carried by several noisy information channels, but this also introduces additional sources of temporal dispersion. Slight variations in axonal length and diameter will smear the signal in time, as axonal conductance velocities are finite and dependent on axonal diameter. However, while

This work was supported by the Wellcome Trust, by the McDonnell-Pew Centre for Cognitive Neuroscience and by the MRC Centre in Brain and Behaviour. I am grateful to Zoubin Ghahramani and Daniel Wolpert for their insight & suggestions.

this lower limit on timing may affect the resolution and the computing speed of particular brain systems, it may not be critical to questions of gross behavioural responses.

At the other extreme, very long time scales may be encoded by mechanisms that do not depend on the temporal pattern of activity within neurons. Memories of early childhood are unlikely to be recalled by any mechanism based on estimation of the interval between then and now, but retrieved with the help of other partially recalled memories - for example that one lived in a particular house at that time. Gallistel (1990) reviews evidence that many animal behaviours carry a "time-stamp" derived from diurnal or sub-diurnal oscillators. Hence the time-stamp is one of the parameters of the memory, but need not itself be a temporal neural signal.

Somewhere between these two extremes - in the range perhaps of fractions of seconds to minutes - are the sorts of time scales that the nervous system can and does encode. So how do artificial models of the nervous system encode time?

TIME AND ARTIFICIAL NEURAL NETWORKS

In terms of their use as models of timing, artificial neural networks, connectionist or network models may be grouped as:

1. *Iterative update models.* Perhaps the most simple and most common form, in which a computer simulation is iteratively calculated, and at each calculation the activity or synaptic weights of all elements in the model are updated (Rumelhart & McClelland, 1986). At this most basic level the models encode time as a discrete sequence of states and one can talk of time as the iteration rate of the model. Clearly these models are not particularly accurate models of physiology, but they are sufficient for the many tasks in which neural activity can be approximated as the average neural firing rate over a unit of time (Hopfield, 1984).
2. *Continuous models.* Networks can be simulated by sets of continuous differential equations, avoiding the iterative basis of discrete models (Pineda, 1987; Pearlmutter, 1989). The units thus have dynamic, time-dependent properties, giving them some powerful additional behaviour. Essentially these models treat neurons as if they were non-spiking: they have continuously variable activation values, with dynamics analogous to the membrane, synaptic and adaptation conductances (Amari, 1978).

There are also discontinuous versions which add an instantaneous impulse representing an action potential. The action potential is triggered by, and resets the membrane potential, while synapses transmit the action potential (Perkel, 1964; Miall, 1989a; Amit, Evans & Abeles, 1990; Judd & Aihara, 1993).

3. *Neural mimics.* The most physiologically accurate models are simulations of individual neurons, modelling membrane properties with cable equations or compartmental models, and often also accurately modelling the dynamics of the action potential as well (review: MacGregor, 1988). These are so computationally exhaustive that very few have been put to use for the sorts of tasks for which discrete artificial neural networks are so popular (Buhmann & Schulten, 1986; Pearson, Finkel & Edelman, 1987; see also Koch & Segev, 1989).

Timing schemes in artificial neural networks. The vast majority of published examples of artificial networks are simulated as iterative processes, and time is represented by the iteration rate. This fact only becomes meaningful however, if the temporal pattern of inputs and outputs is important - for many static applications the activity level of individual units is simply taken to represent the mean firing rate of a neuron during the presentation of the stimulus. The most simple way to extend a network from a static to a temporal representation is to effectively map sequential samples from the inputs into a "spatial pattern" and thus train the network to recognise certain regular features in this sequence of changing spatial patterns. This was the basis of the early NET-talk simulator (Sejnowski & Rosenberg, 1987), that could read aloud by learning the mapping between common letter groups and phonemes. Even in applications like speech recognition, these networks can be valuable tools, given some form of pre-processor that effectively converts the incoming temporal signal into a discrete set of inputs which can then be treated as a spatial pattern.

A popular form of pre-processor, which may also be used on connections within a network, is the delay line. Imagine a set of 5 samples from a time series, perhaps speech, sampled at 1 millisecond intervals, each passed into a delay line which delays the signal by between 0 and 5 ms. If the early samples are delayed the most, and the later samples delayed the least, they will exit the delay line simultaneously, and provide a spatially distributed snapshot of 5 ms of the time series (Lang, Waibel & Hinton, 1990). Typically a tapped delay line, from

which the delayed signal can be read off at many different points, or multiple delays lines of various lengths are employed. Thus the optimum delay for solving a particular signal processing problem need not be known in advance but is reached by synaptic selection between the available delay line options (Unnikrishnan, Hopfield & Tank, 1992). More complex varieties of delay line are worth mentioning - dispersive, Gaussian and adaptive delays. In dispersive and Gaussian delay lines, the output is a time-weighted version of the input (Tank & Hopfield, 1987; Bodenhausen & Waibel, 1991). In other words, if the input is a discrete pulse, the output may be a temporally blurred version whose output rises and falls around some median time delay (review: de Vries & Principe, 1992). Adaptive delays have also been used, in which the delay time or the parameters of the dispersion are modified, again avoiding prior knowledge of the optimal form (review Mozer, in press). And lastly, nets have been designed with "slow" and "fast" synapses, the slow one again delaying signals between neurons (Sompolinksy & Kantor, 1986; Kleinfeld & Sompolinksy, 1988).

In the previous paragraph, I have shifted the emphasis from "feed-forward" to "recurrent" networks in which connections within the network feed activity back from outputs to inputs. This again gives the networks dynamic scope, as activity within the net is no longer governed simply by the inputs (whether or not delayed) but can be self-sustaining. Imagine a unit with a self-excitatory connection - if the strength of this connection is high enough then activity in the unit in response to a pulse input will feedback and re-excite itself, and this activity will be maintained long after the input pulse has ended. Thus a recurrent network can have behaviour similar to that of a dispersive delay line, but can also have more complex dynamic behaviours. The timing of the network is again often implicit in the iteration rate of the model (Jordan, 1986). These recurrent networks have been used widely for signal processing, motor control, sequence production and the like; very many applications are described in Touretzky (1989-1992).

A few final properties of artificial neural networks should be mentioned. There are nets that incorporate dynamic thresholds (Horn & Usher, 1989; Heskes & Gielen, 1992), so that their responses vary with constant inputs. The parameters controlling the threshold can endow the network with very rich behaviour. A trivial example might be an oscillatory threshold, so that the unit acts much like a pace-maker; another could be a bistable membrane, switching from one

state to another. Alternatively, the synapses may adapt over time (Dong & Hopfield, 1992) or with activity from "modulators" (Dehaene, Changeux & Nadal, 1987). The level of activity within one set of synapses thus sets the moment-to-moment weight of another set of connections (Dehaene, Changeux & Nadal, 1987; Schmidhuber, 1992).

TIME DEPENDENT BEHAVIOUR IN BIOLOGICAL NEURAL NETWORKS

Can we specify which of these many mechanisms are more or less likely biologically? The simple answer is no, as it is increasingly clear that every trick that the computer modellers dream up and probably many more are used in biological nervous systems (Harris-Warwick et al., 1992). We can however make some statements about neurons that weaken some mechanisms.

Neurons in general operate fast. A typical estimate for the membrane time constant is about 5 to 20 ms (at least within the vertebrate brain); synapses delay signals by about 1-2 ms; and axonal delays may add another 10-20 ms. Thus single neurons cannot delay signals for any length of time, whereas we are interested in a range of seconds to minutes. Nor do long chains of neurons acting as delay lines (Moore, Desmond & Berthier, 1989) look promising, as few neurons in the vertebrate brain have synapses powerful enough to ensure signal transmission down a single neural chain. Larger chains comprising many linked groups of neurons could be used instead, although temporal smearing of the signals would be expected because of individual differences in synaptic delays and conduction velocities. There is of course evidence for a wide variety of dynamic responses (with slow exponential decay, oscillatory or rebound behaviour, bistable membrane potentials switching etc), which could be used instead. However, as a general rule, such schemes need at least some neurons operating with time scales as long as the time interval to be represented (e.g., Church & Broadbent, 1990 a,b). Re-entrant pathways have also be proposed to form "delay lines" (e.g., Longuet-Higgins, 1968), but the limits on neural accuracy to be discussed next are difficult to reconcile with these schemes.

Neurons are inaccurate. They operate stochastically, both at the membrane and the synapse (Holden, 1976). Even from the great deal

of work on long-term potentiation and depression of synapses, it is not clear over what range synaptic weights normally vary, the precision with which they are set, nor the precision with which they transmit single impulses. However, the resolution of synaptic weights is probably quite low, perhaps 4 - 5 bits, or 16 - 32 separable levels. There is also an upper bound on how accurately neurons can encode quantities with a change in membrane potential prior to rate-coding as action potentials, because of the signal-to-noise ratio of the potential; estimates of 4 - 7 bits (16 - 128 separable levels) are sometimes given (Laughlin, 1989; Attwell & Tessier-Lavigne, 1989). In the steady state, a neuron's firing rate represents its activation level, but the time over which the rate is assessed may need to be quite long for any degree of accuracy. First, the rate can only be measured as an inter-spike interval, so by default one must at least wait for the second spike. Bailek and colleagues (1991, 1992) estimate about 3 bits of information per spike; using about a 30 ms window allows the signal to be recovered but with a signal-noise ratio of perhaps only 5:1 (i.e. 20% errors). Second, the firing rate measured over tens or even hundreds of milliseconds may fluctuate widely about the average (look at the trial-trial variability in any neurophysiological experiment). This also limits the precision with which slow membrane potential changes can signal events over long time intervals. For example, if a very slow oscillation is used as the basis of some timing function, then the duration of any phase of the oscillation cycle cannot be given precisely: neither the moment at which the rising potential reaches threshold and triggers a spike nor the length of the burst of spikes can be known precisely. Finally, many neurons adapt to a steady input, and the firing rate falls with time. Thus quantity cannot be represented with infinite accuracy, either as a synaptic weight nor as an activation level. For related reasons, one would not expect that recurrent networks could act as "re-entrant" delay line modules in which they gradually cycle through a number of steps before emitting an accurate but delayed copy of their input. The demands on maintaining an accurate representation of the signal as it repeatedly passes through many synapses seem too great.

In contrast, many artificial neural networks require high precision (more than 12 bits or 4096 levels). During the gradual training which is typical of the backpropagation algorithm, synaptic weights are changed by minute amounts. The units' activation levels are also typically of high precision. There is an active debate about the extent to which this is necessary (Durbin, 1987), as at least in some cases

the networks can learn successfully with low resolution synapses and useful networks have been implemented with binary (on-off) neurons and synapses.

Neural networks for timing and counting

If we take these limitations on the speed and accuracy of single neurons seriously, then we should ask what schemes can be used to accurately time events over the time scale of seconds or minutes. An early scheme was proposed by Longuet-Higgins (1968) using banks of frequency filters (actually visualised as oscillatory neurons) to store a brief temporal signal. A somewhat different scheme for timing intervals used networks of oscillating neurons, which relied on the phenomenon of "beating" (Miall, 1989b, 1992). This was robust, and made no unreasonable physiological demands, but in fact did not match the psychophysics of timing well: it showed no increase in variance for long intervals and it's recall was either perfect or disastrous, never merely inaccurate. Church and Broadbent (1990 a,b) proposed a related scheme based on summing the activity of a set of phase-locked oscillators, which could match the psychophysical data. However, like Longuet-Higgins' scheme, their scheme required oscillators with a very broad range of oscillation periods, up to and including the time period to be stored. The beating scheme (Miall, 1989b, 1992) avoided this problem, and could store arbitrarily long intervals.

There is, however, considerable psychophysical and behavioural evidence for some form of counter, so that intervals can be estimated by accumulating or integrating the ticks of a internal clock for comparison with a reference value (Gibbon & Allan, 1984). Individual neurons from many frontal areas of the cortex show a gradual increase or decrease in their averaged activity during timing tasks, as expected of cells functioning as an integrator or counter (Niki & Watenabe, 1979). However, for reasons mentioned above, one would not expect a single cell to be able to integrate accurately over long periods, and in general the activity of these neurons is not smooth from moment-to-moment within any one trial. Furthermore, neurons are leaky integrators, and cannot maintain a constant activation level indefinitely. To counteract the leak requires a positive recurrent connection that precisely balances the decay, but again accuracy is a problem: too little feedback merely slows the decay down, while too

much leads to saturation. Robinson (1989) has also argued this point for the neural integrator involved in oculomotor control. His solution is to propose that networks of neurons with lateral inhibitory connections can overcome their individual variations, and achieve positive feedback by self-disinhibition through their neighbours (Arnold & Robinson, 1991). However, the oculomotor system may be a rather special case, in that the neurons involved in the integrator show remarkably consistent activity profiles, quite unlike those recorded in cortical areas during delayed response tasks (see e.g., Mauritz & Wise, 1986; Niki & Watenabe, 1979). Furthermore, Zipser (1991) has shown that a deterministic network able to produce activity profiles like the averaged responses of neurons in a delayed task may not be stable over sufficiently long periods if neuronal noise is added.

An alternative, based on a suggestion by Zoubin Ghahramani at M.I.T. (personal communication), is to simply consider the count to be the total activity within a population of noisy neurons, none of which can individually accumulate or integrate. I will briefly describe this idea and consider the accuracy of the counter in relation to biological timing.

Networks to time. Consider a group of N neurons, all receiving the same input from an internal clock periodically emitting a pulse of activity. Let each "neuron" or unit have a low probability of being switched on by any one clock tick ($p_{on} < 1$). Once on, it remains on, but because neurons may not maintain prolonged activity faithfully, we can also add a low probability of each active unit switching itself off at any moment ($p_{off} < p_{on}$). Now, let each of these N units excite some other system (perhaps another neuron or network of neurons), such that the total excitation represents the accumulated measure of time. Obviously we need this measure of total activity to rise monotonically with the number of clock ticks. On the first tick, a small number of units will be activated, and some of these inactivated. The total activity will be:

$$A_1 = N \cdot p_{on} \cdot (1 - p_{off}).$$

At the second tick, more of the inactive units will be activated:

$$A_2 = A_1 \cdot (1 - p_{off}) + (N - A_1) \cdot p_{on} \cdot (1 - p_{off}), \text{ and so on.}$$

If we consider the number of active units as a Markov process governed by the two probabilities, then the number of active and inactive units at time $t+1$ is given by:

$$\begin{vmatrix} N_{on}(t+1) \\ N_{off}(t+1) \end{vmatrix} = \begin{vmatrix} (1-p_{off}) & p_{on}\cdot(1-p_{off}) \\ p_{off}\cdot(1-p_{on}) & (1-p_{on}) \end{vmatrix} \cdot \begin{vmatrix} N_{on}(t) \\ N_{off}(t). \end{vmatrix}$$

This can be resolved to give the activity at time t as:

$$A_t = N\cdot p_{on}\cdot(1-p_{off})\cdot[1-(1-p_{on})^t\cdot(1-p_{off})^t]\,/\,[\,p_{on}\cdot(1-p_{off})+p_{off}].$$

As t goes towards infinity, this asymptotes towards $N\cdot p_{on}\cdot(1-p_{off})\,/\,[p_{on}\cdot(1-p_{off})+p_{off}]$, so the network has an upper limit to the number of time steps it can encode. Figure 1A shows the averaged results

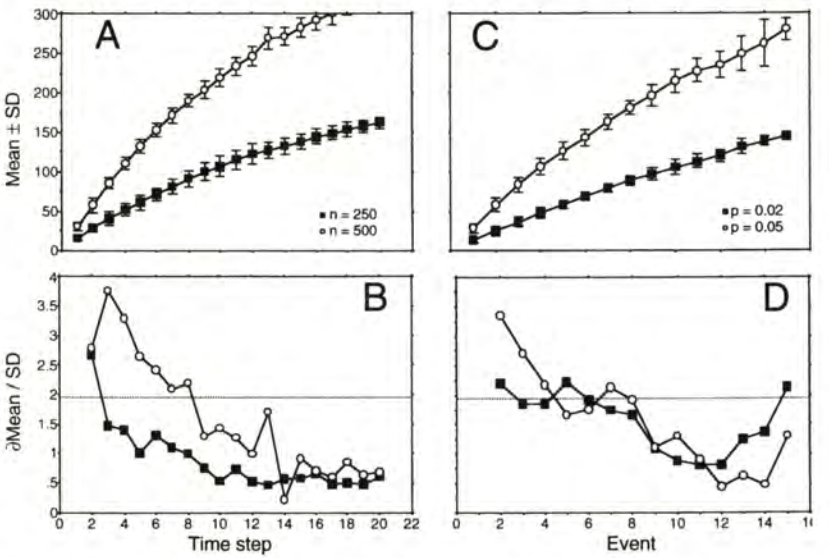

Figure 1. Timing and counting. A: *Timing*. Mean number of active units (± SD) against time step. Hollow symbols: a net of 500 units; solid symbols: a net of 250 units. **B:** The increment in mean activity over standard deviation for data from Fig. 1A. The horizontal line is at 1.96 (95% confidence limit). **C:** *Counting*. Mean activity against event number. Hollow symbols: counting probability p=0.05; solid symbols, p=0.02. **D:** Incremental activity over standard deviation for data from Fig. 1C.

from 10 runs of a simple computer simulation. The net had either N = 500 or 250 units; p_{on} was set to 0.05; and $p_{off} = 0.0001$ per iteration (since the clock ticks arrived every 100 iterations, $p_{off} = 0.25p_{on}$). The total activity does indeed increase with time, but the variance also rises. This is particularly obvious for the smaller net (solid symbols, N = 250). Clearly the fidelity of the system depends on the incremental increase in activity with each clock tick in respect to the variance. Figure 1B shows this ratio, and a cut-off threshold of 1.96, which

would represent the 95% confidence limit for accurate detection of the increase in activity per time step. For the larger net, this limit is reached after about 9 clock ticks, whereas the smaller net fails earlier.

Networks to count. Figure 1 C shows the same network put to the task of counting. In this case, the timing of the events (or clock ticks) was random, arriving with a probability of 0.005 per iteration. The net had $N = 500$ units, and was tested with $p_{on} = 0.05$ or 0.02; p_{off} remained at 0.0001 per iteration. Again the total activity increased with the number of events, approaching the asymptote more gradually the lower the value of p_{on} but in each case with increasing variance. Figure 1D shows the ratio of incremental increase in activity to standard deviation; the curves for both nets hover around or above the 95% confidence limit until a count of 8, before falling quite sharply.

These are very simple models and the match or mismatch to the psychophysics of timing and counting should not be overstated; by selecting other parameter values the networks can be made much more robust, or much less. However, they do raise several interesting points for discussion.

First, it is clear that a population of very low quality neurons can form the accumulator required for an internal timer or to count randomly timed events. It is only the source of the input that determines whether the network output reflects time or quantity; and thus the same network could be used to time or to count. It also means that timing is feasible with irregular clock pulses, as allowed for by scalar timing (Church & Gibbon, 1982) and behavioural timing theories (Killeen & Fetterman, 1988). In this simulation the individual units were either on or off, all synaptic weights were fixed (and uniform), and all neurons were stochastic both in their activation and inactivation. Any one neuron may not respond on any given trial, and only the total activity in the net correlates with the passing time or number of events. Hence, no obvious clue to the timing ability of the network would be gained by recording from any one neuron on any one trial - the evidence for accumulation can only be seen in the population response, or in the responses averaged over many trials. This is not unreasonable biologically, but one could easy add more power to the network, for example allowing low resolution changes in activation level. Second, the net has increasing variance with increasing events or time, and therefore at first approximation fits the psychophysical data. Killeen (1992) reviews evidence that the counter is indeed inaccurate

in accumulating clock ticks. Furthermore, the asymptotic nature of the curves in Figure 1A and 1C imply that there would be systematic errors of timing and counting - underestimating long intervals or high counts or overestimating short intervals and counts - features which are typical of time psychophysics (see again Killeen, 1992; Church & Gibbon, 1982). More detailed simulations will be needed to test this fit further, and test the other predictions of scalar timing theory. Third, the ratio of differential increase in activity to the variance (Figures 1B and 1D) implies that there is an upper limit to the number of accumulated clock intervals or the number of events that can be distinguished. This also fits the behavioural evidence: Kristofferson (1980) reported that the clock period seems to change with increase in duration, such that a maximum of 8 to 16 clock ticks would be accumulated; and Killeen (1992) suggests that some small number of counts is usually reached when mentally counting off time. Gallistel (1990) reviews evidence that animals may count their own behavioural responses, perhaps up to limits of 40 or 50 events, although with considerable variance. Fourth, the network as currently presented acts only as the accumulator, and for timing must receive inputs from a clock or pacemaker; for counting, it simply receives discrete pulses from a sensory system detecting the events to be counted. Thus the effects of drugs, temperature, information load, etc. on the perception of time could affect the clock and not the pacemaker. One way that this model could be affected, however, is for the probabilities p_{on} and p_{off} to change. An increase in p_{on} would mean that the output of the network would be greater for a given number of inputs, and thus the net would seem to count fast; an increase in p_{off} would tend to have the reverse effect. Interestingly, these changes would also affect the variance of the responses, so one could test the question of whether it was the clock or the accumulator (or both) that was affected.

Lastly, I should briefly address how it could be that only a small random subset of otherwise identical neurons might respond to a common input ($p_{on} < 1.0$). One possibility would be to require coincident input from the timer and from another random input. For example, if each neuron received an independent series of action potentials uncorrelated with the common timer input, then only those neurons with coincident arrival of both inputs would become activated. Alternatively, one might imagine that the whole population of neurons has asynchronous sub-threshold oscillatory behaviour. Then the ad-

ditional excitatory timing input arriving in-phase with this oscillation could switch them to their active state; again, if the population oscillations were uncorrelated, the selected group would be just those few at the peak phase of their oscillation period.

Conclusion

The three main points of this paper are that artificial neural networks have limited temporal behaviour, and there are problems in the match between these properties and the neurophysiological realities. However, many of these problems can be overcome by the benefits of multi-neuron networks. A simple network making no unreasonable demands on biology could be used to count or accumulate the output of an internal timer. Such network solutions are typically hard to interpret, but then so too are the records from electrophysiological experiments.

References

Amari, S. (1978). *Mathematical theory of neural networks*. Tokyo: Sangyo-tosyo.

Amit, D. J., Evans M. R., & Abeles, M. (1990). Attractor neural networks with biological probe records. *Network*, *1*, 381-405.

Arnold, D. B., & Robinson, D. A. (1991). A learning network model of the neural integrator of the oculomotor system. *Biological Cybernetics*, *64*, 447-454.

Attwell, D., & Tessier-Lavigne M. (1989) Designing synaptic connections in the retina. In R. M. Durbin, R. C. Miall, & G. J. Mitchison (Eds.), *The computing neuron* (pp. 337-354). Wokingham, UK: Addison-Welsey.

Bailek, W., & Rieke, F. (1992). Reliability and information transmission in spiking neurons. *Trends in Neuroscience*, *15*, 428- 434.

Bailek, W., Rieke, F., de Ruyter van Steveninck, R. R., & Warland, D. (1991). Reading a neural code. *Science*, *252*, 1854-1857.

Bodenhausen, U., & Waibel, A. (1991). The Tempo 2 algorithm: Adjusting time-delays by supervised learning. In R. P. Lippmann, J. E. Moody, & D. S. Touretzky (Eds.), *Advances in neural information processing systems* (Vol. 3, pp. 155-161). San Mateo, CA: Morgan Kaufmann.

Buhmann, J., & Schulten, K. (1986). Associative recognition and storage in a model network of physiological neurons. *Biological Cybernetics*, *54*, 319-335.

Church, R. M., & Broadbent, H. A. (1990a). A connectionist model of timing. In M. L. Commons, S. Grossberg, & J. E. R. Staddon (Eds.), *Quantitative models of behaviour: Neural networks and conditioning* (pp. 225-240). Hillsdale, NJ: Erlbaum.

Church, R. M., & Broadbent, H. A. (1990b). Alternative representations of time, number and rate. *Cognition, 37*, 55-81.

Church, R. M., & Gibbon, J. (1982). Temporal generalization. *Journal of Experimental Psychology: Animal Behavior Processes, 8*, 165-186.

Dehaene, D., Changeux, J.-P., & Nadal, J. P. (1987). Neural networks that learn temporal sequences by selection. *Proceedings of the National Academy of Sciences of the United States of America, 84*, 2727-2731.

de Vries, B., & Principe, J. C. (1992). The gamma model - a new neural model for temporal processing. *Neural Networks, 5*, 565-576.

Dong, D. W., & Hopfield, J. J. (1992). Dynamic properties of neural networks with adapting synapses. *Network, 3*, 267-283.

Durbin, R. M. (1987). Backpropagation with integers. *Abstracts of the meeting, Neural Networks for Computing*, Snowbird, UT.

Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: M.I.T. Press.

Gibbon, J., & Allan, L. G. (Eds.). (1984). Time and time perception. *Annals of the New York Academy of Sciences, vol. 423*.

Harris-Warwick, R. M., Marder, E., Selverston, A. I., & Moulins, M. (Eds.). (1992). *Dynamic biological networks: The stomatogastric nervous system*. Cambridge, MA: M.I.T. Press.

Heskes, T. M., & Gielen, S. (1992). Retrieval of pattern sequences at variable speeds in a neural network with delays. *Neural Networks, 5*, 145-152

Holden, A. V. (1976). *Models of stochastic activity of neurons*. Berlin: Springer.

Hopfield, J. J. (1984). Neurons with graded responses have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences of the United States of America, 81*, 3088-3092.

Horn, D., & Usher, M. (1989). Neural networks with dynamical thresholds. *Physical Review A, 40*, 1036-1044.

Jordan, M. I. (1986). Attractor dynamics and parallelism in a connectionist sequential machine. *Proceedings of the eighth annual conference on Cognitive Science* (pp. 531-546). Hillsdale, NJ: Erlbaum.

Judd, K. T., & Aihara, K. (1993). Pulse propagation networks: A neural network model that used temporal coding by action potentials. *Neural Networks, 6*, 203-215.

Killeen, P. R. (1992). Counting the minutes. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (NATO ASI Series D, Vol. 66, pp. 203-214). Dordrecht: Kluwer.

Killeen, P. R., & Fetterman, J. G. (1988). A behavioural theory of timing. *Psychological Reviews, 95*, 274-295.

Kleinfeld, D., & Sompolinsky, H. (1988). Associative neural network model for the generation of temporal patterns. *Biophysical Journal, 54*, 1039-1051.

Koch, C., & Segev, I. (Eds.). (1989). *Methods in neuronal modelling: From synapses to networks*. Cambridge, MA: M.I.T. Press.

Kristofferson, A. B. (1980). A quantal step function in duration discrimination. *Perception & Psychophysics, 27*, 300-306.

Lang, K., Waibel, A. H., & Hinton, G. E. (1990). A time-delay neural network architecture for isolated word recognition. *Neural Networks, 3*, 23-44.

Laughlin, S. (1989). The reliability of single neurons and circuit design: A case study. In R. M. Durbin, R. C. Miall, & G. J. Mitchison (Eds.), *The computing neuron* (pp. 322-336). Wokingham, UK: Addison-Welsey.

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

MacGregor, R. J. (1988). *Neural and brain modeling*. San Diego, CA: Academic Press.

Mauritz, K. H., & Wise, S. P. (1986). Premotor cortex of the rhesus monkey - Neuronal activity in anticipation of predictable environmental events. *Experimental Brain Research, 61*, 229-244.

Miall, R. C. (1989a). The diversity of neuronal properties. In R. M. Durbin, R. C. Miall, & G. J. Mitchison (Eds.), *The computing neuron* (pp. 11-34). Wokingham, UK: Addison-Welsey.

Miall, R. C. (1989b). The storage of time intervals using oscillating neurons. *Neural Computation, 1*, 359-371.

Miall, R. C. (1992). Oscillators, predictions and time. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (NATO ASI Series D, Vol. 66, pp. 215-227). Dordrecht: Kluwer.

Moore, J. W., Desmond, J. E., & Berthier, N. E. (1989). Adaptively timed conditioned responses and the cerebellum: A neural network approach. *Biological Cybernetics*, 62, 17-28.

Mozer, M. (in press). Neural net architectures for temporal sequence processing. In A. Weigend & N. Gershenfeld (Eds.), *Predicting the future and understanding the past* Redwood City, CA: Addison-Welsey.

Niki H., & Watenabe M. (1979). Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Research, 171*, 213-224.

Pearlmutter, B.A. (1989). Learning state space trajectories in recurrent neural networks. *Neural Computation, 1*, 263-269.

Pearson, J. C., Finkel, L. H., & Edelman, G. M. (1987). Plasticity in the organization of adult cerebral cortical maps: A computer simulation based on neuronal group selection. *Journal of Neuroscience, 7*, 4209-4223.

Perkel, D. M. (1964). *A digital computer model of nerve cell functioning*. Technical Report RM-4132-NIH, Rand Corp., CA.

Pineda, F. (1987). Generalization of backpropagation to recurrent neural networks. *Physical Review Letters, 19*, 2229-2232.

Robinson, D. A. (1989). Integrating with neurons. *Annual Review of Neuroscience, 12*, 33-45.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing.* Cambridge, MA: M.I.T. Press.

Schmidhuber, J. (1992). Learning to control fast-weight memories: An alternative to dynamic recurrent networks. *Neural Computation, 4*, 131-139.

Sejnowski, T. J., & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems, 1*, 145-168.

Sompolinsky, H., & Kanter, I. (1986). Temporal association in asymmetric neural networks. *Physical Review Letters, 57*, 2861-2864.

Tank, D. W., & Hopfield, J. J. (1987). Neural computation by concentrating information in time. *Proceedings of the National Academy of Sciences of the United States of America, 84*, 1896-1900.

Touretzky, D. S. (Ed.). (1989-92). *Advances in neural information processing systems* (Vols. 1-4). San Mateo, CA: Morgan-Kaufmann Publishers.

Unnikrishnan, K. P., Hopfield, J. J., & Tank, D. W. (1992). Speaker-independent digit recognition using a neural network with time-delayed connections. *Neural Computation, 4*, 108-119.

Zipser, D. (1991). Recurrent network model of the neural mechanism of short-term active memory. *Neural Computation, 3*, 179-193

Parks Road
Oxford OX1 4PT
United Kingdom