Communicated by Stephen Lisberger

Neural Network Model of the Cerebellum: Temporal Discrimination and the Timing of Motor Responses

Dean V. Buonomano*
Department of Neurobiology and Anatomy, University of Texas Medical School, Houston, TX 77225 USA and Departamento de Matemática Aplicada, Instituto de Matemática, Universidade Estadual de Campinas, Campinas, Brasil, and Laboratório de Psicobiologia, Universidade de São Paulo, Ribeirão Preto, Brasil

Michael D. Mauk
Department of Neurobiology and Anatomy, University of Texas Medical School, Houston, TX 77225 USA

Substantial evidence has established that the cerebellum plays an important role in the generation of movements. An important aspect of motor output is its timing in relation to external stimuli or to other components of a movement. Previous studies suggest that the cerebellum plays a role in the timing of movements. Here we describe a neural network model based on the synaptic organization of the cerebellum that can generate timed responses in the range of tens of milliseconds to seconds. In contrast to previous models, temporal coding emerges from the dynamics of the cerebellar circuitry and depends neither on conduction delays, arrays of elements with different time constants, nor populations of elements oscillating at different frequencies. Instead, time is extracted from the instantaneous granule cell population vector. The subset of active granule cells is time-varying due to the granule—Golgi—granule cell negative feedback. We demonstrate that the population vector of simulated granule cell activity exhibits dynamic, nonperiodic trajectories in response to a periodic input. With time encoded in this manner, the output of the network at a particular interval following the onset of a stimulus can be altered selectively by changing the strength of granule → Purkinje cell connections for those granule cells that are active during the target time window. The memory of the reinforcement at that interval is subsequently expressed as a change in Purkinje cell activity that is appropriately timed with respect to stimulus onset. Thus, the present model demonstrates that

*Present address: Keck Center for Integrated Neuroscience, University of California, San Francisco, CA 94143.

a network based on cerebellar circuitry can learn appropriately timed responses by encoding time as the population vector of granule cell activity.

1 Introduction

Generating movements inherently involves producing appropriately timed contractions in the relevant muscle groups. Accordingly, an important aspect of motor control is the timing of motor commands in relation to internal (i.e., proprioceptive cues) and external stimuli. One clear and experimentally tractable example of the timing of responses with respect to an external stimulus is Pavlovian conditioning of eyelid responses. Learned responses are promoted in this paradigm by paired presentation of a cue or conditioned stimulus (CS) and a reinforcing unconditioned stimulus (US). For example, the presentation of a tone (the CS) is reinforced by the copresentation of a puff of air directed at the eye (the US). This promotes the acquisition of learned eyelid responses that are elicited by the CS. These learned eyelid movements are delayed so that they peak near the onset of the potentially harmful US (Schneiderman and Gormezano 1964; Gormezano *et al.* 1983). Previous studies have shown that the timing of the eyelid response is learned and that the underlying neural mechanism involves temporal discrimination during the CS (Mauk and Ruiz 1992). Since appropriately timed responses can be generated for CS–US intervals between 80 and 2000 msec, the neural mechanism appears to be capable of temporal discrimination in this range.

Very little is known about how and where the nervous system encodes time. There is evidence that neurons can use axonal delays to detect temporal intervals. For example, axonal conduction delays appear to contribute to the detection of interaural time delays (Carr and Konishi 1988; Overholt *et al.* 1992). Furthermore, theoretical work (Koch *et al.* 1983) has suggested that dendritic conduction delays could contribute to the detection of temporal delays important for direction selectivity in the visual system. In both these instances, however, the relevant time intervals are below tens of milliseconds. Braitenberg (1967) suggested that the parallel fibers in the cerebellum could function as delay lines that would underlie the timing of movements. Given the conduction velocity of parallel fibers of approximately 0.2 mm/msec it is unlikely that such a mechanism could underlie timing in the tens of milliseconds to second range (Freeman 1969). Other models have been presented in which intervals above tenths of seconds could be stored by a group of oscillating neurons with different frequencies (Miall 1989; Church and Broadbent 1991; Fujita 1982; Gluck *et al.* 1990). As of yet, however, no such population of neurons has been described.

It has been suggested that the cerebellum may play an important role in the timing of movements (Braitenberg 1967; Freeman 1969; Eccles 1973), in addition to being an important component for the learning of movements (Marr 1969; Albus 1971; Fujita 1982; Lisberger 1988). Indeed it has been shown that while cerebellar cortex lesions do not abolish the conditioned response, the timing of the responses is disrupted (McCormick and Thompson 1984; Perrett *et al.* 1993). Whereas numerous studies indicate that output from the cerebellum via cerebellar nuclei is required for Pavlovian eyelid conditioning (Thompson 1986; Yeo 1991; however, see Welsh and Harvey 1989). One class of cerebellar afferents (mossy fibers) appears to convey to the cerebellum the presentation of the CS and a second class of cerebellar afferents (climbing fibers) appears to convey the presentation of the US (e.g., see Thompson 1986). These data suggest that (1) Pavlovian conditioning is a relatively simple, experimentally tractable paradigm for the study of the input/output properties of the cerebellum, (2) the cerebellum is necessary for the appropriate timing of conditioned movements, and thus, (3) Pavlovian conditioning is particularly well suited for the study of cerebellar mechanisms that mediate the ability to discriminate time with respect to the onset of an internal or external stimulus.

The purpose of this paper is to use a neural network model to test a specific hypothesis suggesting how the known circuitry of the cerebellum could make temporal discriminations and mediate the timing of the conditioned response. The model tested here is based on the known circuitry of the cerebellum and does not utilize delay lines, assume arrays of elements with different time constants, or assume arrays of elements that oscillate at different frequencies, but instead shows how the dynamics of cerebellar circuitry could encode time with the population vector of granule cell activity.

2 Structure of the Model

2.1 Cerebellar Circuitry. In comparison to most brain regions the neural circuitry, cell ratios, and synaptic convergence/divergence ratios of the cerebellum are fairly well established (Eccles *et al.* 1967; Palkovits *et al.* 1971; Ito 1984). The principal six cell types are the granule (Gr), Golgi (Go), Purkinje (PC), basket, stellate, and cells of the cerebellar nuclei. The two primary inputs to the cerebellum are conveyed by mossy fibers (MF) and climbing fibers (CF). Figure 1A illustrates the known connectivity of these cells.

2.2 Classic Cerebellar Theories. Based on the known characteristics of cerebellar organization, Marr (1969) and later Albus (1971) proposed models of the cerebellum suggesting a mechanism to mediate motor learning. In these theories (1) the contexts in which movements take

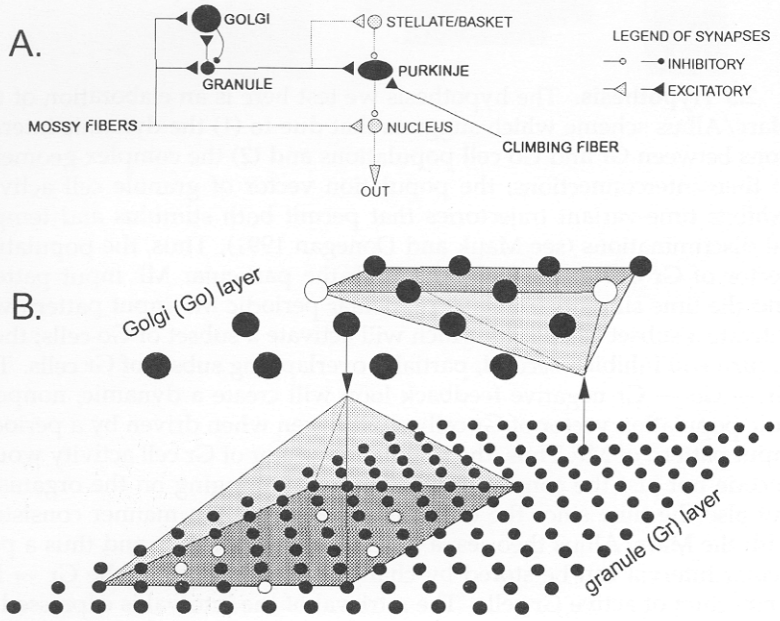

Figure 1: Synaptic organization of the cerebellum. (A) A schematic diagram of the known synaptic connections in the cerebellum. Shown in bold and with solid lines are the components incorporated into the present computer model. (B) A schematic representation of the connectivity in the present model. Small representations of the granule (Gr) and Golgi (Go) layers are shown; there were 10^4 and 900 Gr and Go cells, respectively, in the simulations. The 500 MFs and single PC are omitted for clarity. The arrows and shaded pyramids illustrate the spans or regions to which the cells are eligible to make a synaptic contact. Within the spans connections are made with a uniform distribution. The white cells in each span exemplify cells that receive a synaptic input from the presynaptic cell. Thus, the shape of the spans was designed to reflect the geometry of the projections of each cell type, whereas the number of cells within the span that actually received a connection was a reflection of the convergence and divergence ratios of the synaptic connections.

place are encoded by the population vector of Gr cell activity, (2) the CFs convey error signals indicating that the response requires modification, and (3) this CF error signal modifies active Gr $\rightarrow$ PC synapses such that subsequent motor performance in that context is improved. Marr also

suggested that the Go cell negative feedback stabilizes the amount of Gr cell activity which could maximize the discriminability of the contexts.

2.3 Hypothesis. The hypothesis we test here is an elaboration of the Marr/Albus scheme which suggests that due to (1) the dynamic interactions between Gr and Go cell populations and (2) the complex geometry of their interconnections, the population vector of granule cell activity exhibits time-variant trajectories that permit both stimulus and temporal discriminations (see Mauk and Donegan 1991). Thus, the population vector of Gr cell activity encodes both the particular MF input pattern and the time since its onset. A particular periodic MF input pattern will activate a subset of Gr cells which will activate a subset of Go cells; these in turn will inhibit a second, partially overlapping subset of Gr cells. The Gr $\rightarrow$ Go $\rightarrow$ Gr negative feedback loop will create a dynamic, nonperiodic population vector of Gr cell activity even when driven by a periodic input pattern of MF. Thus, the population vector of Gr cell activity would encode not just the constellation of stimuli impinging on the organism, but also the time since the onset of the stimuli. In a manner consistent with the Marr/Albus theories, a given subset of Gr cells and thus a particular interval can be stored by changing the strength of the Gr $\rightarrow$ PC connection of active Gr cells. The retrieval of the interval is expressed as a change in the activity level of the PC.

2.4 Neural Network. The neural network consisted of 10^4 Gr cells, 900 Go cells, 500 MF inputs, and one PC. The architecture is illustrated schematically in Figure 1B; the 500 MFs and the single PC are omitted for clarity. Each Gr cell received excitatory synaptic inputs from three MFs and inhibitory inputs from three Go cells. Each Go cell received excitatory inputs from 100 Gr cells and 20 MFs. The PC received inputs from all 10^4 Gr cells. In scaling the network to computationally feasible dimensions, it is not possible to maintain both the empirically observed Gr/Go cell ratios and convergence/divergence ratios. The biological Gr/Go cell ratio is approximately 5000/1, and the Gr convergence/divergence ratio is approximately 0.4 (4/10, i.e., each Gr cell receives input from 4 Go cells and sends contacts to 5–10 Go cells; Eccles 1973; Palkovits *et al.* 1971). Thus, a network with 10^4 Gr cells should have only 2 Go cells, yet each of the Gr cells should contact approximately 5–10 Go cells. The underlying assumption we employed in making the compromise between cell ratios and convergence/divergence ratios was that the latter is more important. Thus, the convergence/divergence ratios for the Gr and Go cells were maintained within an order of magnitude of observed experimental values. In the model the Gr/Go ratio was 11.1 (10^4/900) and the Gr convergence/divergence ratio was 0.33 (3/9).

The Gr and Go cells were simulated as modified integrate and fire elements. For example, the voltage of each Go cell (V_i^{Go} was determined by

$$\frac{dV_i^{Go}}{dt} = -g_i^{Go:leak}(V_i^{Go} - E_{leak}) - g_i^{Go:MF}(V_i^{Go} - E_{ex}) - g_i^{Go:Gr}(V_i^{Go} - E_{ex}) \quad (2.1)$$

A spike (S_i^{Go}) was generated if threshold was reached.

$$S_i^{Go} = \begin{cases} 1, & V_i^{Go} \geq \mathrm{Thr}_i^{Go} \\ 0, & V_i^{Go} < \mathrm{Thr}_i^{Go} \end{cases} \quad (2.2)$$

where Thr_i^{Go} = spike threshold for Go cell i. Synaptic currents were simulated with an instantaneous rise and an exponential decay. All inputs of a particular type are summed into a single current that saturates at 1.0 and decays at the rate of τ. Thus the MF → Go cell synaptic current ($g_i^{Go:MF}$) is determined by

$$\frac{dg_i^{Go:MF}}{dt} = \sum_n^{MF} S_n^{MF} \times W^{Go:MF} \times (1 - g_i^{Go:MF}) - g_i^{Go:MF}\tau^{Go:MF} \quad (2.3)$$

where S_n^{MF} represents a spike in a MF, $W^{Go:MF}$ is the synaptic weight of the MF synapses, and $\tau^{Go:MF}$ is the decay time constant.

A relative refractory period and spike accommodation was simulated by increasing threshold to MaxThr^{Go} after each spike with a subsequent exponential decay to the initial value with a time constant of $\tau^{Go:Thr}$. The dynamics of each Gr cell was controlled by similar equations except each Gr cell received an excitatory and an inhibitory synaptic current, $g^{Gr:MF}$ and $g^{Gr:Go}$, respectively. The values and definitions of the constants are given in Table 1.

The single PC received input from all Gr cells in the network. Initially all Gr cells were connected to the PC with the same weight. During the first presentation of the stimulus the weights of the Gr → PC connections for all Gr cells active within a target window were decreased. This decrease in synaptic weight during a particular time window simulates the long-term depression (LTD) of the parallel fiber to PC connection produced by coactivation of parallel and climbing fibers (Ito *et al.* 1982; Linden *et al.* 1991). The ability of the network to generate an appropriately timed response was then tested by monitoring the PC activity during a second presentation of the stimulus. Thus the voltage of the PC represented the weighted summed activity of all the Gr cells in the network with a time constant of 2.5 msec.

In certain versions of the network a conduction delay was incorporated in the Gr → Go connection. Since the delay did not affect the results in a significant manner and to stress that timing in the network emerges from the dynamics of the Gr → Go → Gr feedback loop the simulations presented here did not include conduction delays.

Table 1: Definitions and Values of Constants.[a]

Decay time constants		
$\tau^{Gr:MF}$	MF → Gr synaptic conductance	2.86
$\tau^{Gr:Go}$	Go → Gr synaptic conductance	5.0
$\tau^{Go:MF}$	MF → Go synaptic conductance	2.87
$\tau^{Go:Gr}$	Gr → Go synaptic conductance	2.86
$\tau^{Gr:Thr}$	Gr threshold decay	1.7
$\tau^{Go:Thr}$	Go threshold decay	2.0
Cellular parameters		
E_{leak}	Leak equilibrium potential	−60
E_{ex}	EPSP equilibrium potential	0
E_{inh}	IPSP equilibrium potential	−80
$g^{Gr:leak}$	Leak conductance for Gr	.07
$g^{Go:leak}$	Leak conductance for Go	.07
Synaptic parameters		
$W^{Gr:MF}$	Synaptic weight of MF → Gr	0.15
$W^{Gr:Go}$	Synaptic weight of Go → Gr	0.15
$W^{Go:MF}$	Synaptic weight of MF → Go	0.007
$W^{Go:Gr}$	Synaptic weight of Gr → Go	0.008
Threshold parameters		
Thr^{Gr}	Gr cell minimum threshold	−40
Thr^{Go}	Go cell minimum threshold	−35
$MaxThr^{Gr}$	Maximum Gr threshold	−35
$MaxThr^{Go}$	Maximum Go threshold	−25

[a] In the table and equations the first superscript refers to the postsynaptic cell that the constant or variable applies to, whereas the second superscript refers to the presynaptic cell or to a particular variable.

Input to the network was conveyed by 500 MF. In the simulations described here the presentation of a CS was represented by activation of 20% of the MFs at a frequency of 100 Hz. Driving the network with a constant frequency input insured that any timing exhibited by the network was a result of the intrinsic dynamics of the circuit.

3 Simulations

3.1 Timing. Figure 2 illustrates the ability of the network to learn a temporal discrimination when driven by the MF input. Following a baseline of low background activity, a subset of mossy fibers was acti-

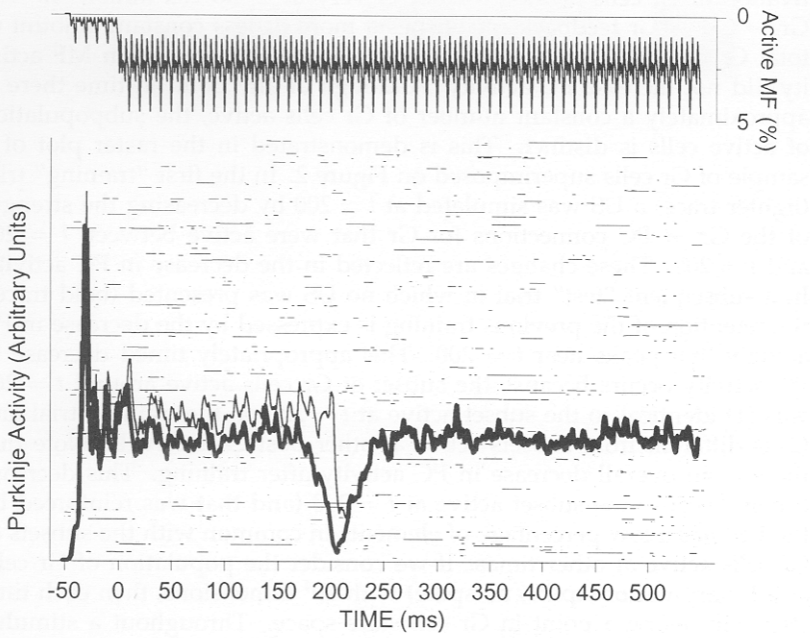

Figure 2: Results from a typical simulation of a Pavlovian conditioning trial in which temporal coding is expressed in the PC activity. Shown along the top is the percent of MFs active each time bin with the onset of the simulated CS marked by the increase in MF activity at time zero. During each trace the network is initially silent due to the absence of MF activity. To provide "background" activity, 5% of the MFs were turned on at $t = -45$, and at CS onset ($t = 0$) an additional 20% of the MFs were activated at 100 Hz (the phase of each MF action potential was random). The lighter continuous trace represents PC activity in the initial training trial in which a climbing fiber input was simulated by decreasing to zero the strength of the Gr $\rightarrow$ PC connections that spiked three or more times between $t = 200$ and 205 msec. The darker continuous trace shows a subsequent test trial in which no US was presented. The retention of the temporal interval between CS and US is expressed as a decrease in PC activity whose onset precedes and whose peak occurs near the time at which the US was previously presented. Finally, a raster plot showing the activity of 600 of the Gr cells as a function of time during the trial is superimposed.

vated as a hypothetical CS presented to the network (total MF activity is shown in the upper panel). The lower panel shows the output of the network as represented by the activity of a PC during two trials. Activity in the PC is proportional to the weighted sum of active Gr cells.

Initially all weights were equal. The initial peak corresponds to the activation of Gr cells in the presence of very little Go cell inhibition. The Gr → Go → Gr feedback establishes a more or less constant amount of total Gr cell activity. Note that at $t = 0$ a 5-fold increase in MF activity did not increase PC activity. Although at any point in time there is approximately a constant number of Gr cells active, the subpopulation of active cells is distinct. This is demonstrated in the raster plot of a sample of Gr cells superimposed on Figure 2. In the first "training" trial (lighter trace) a US was simulated at $t = 200$ by decreasing the strength of the Gr → PC connections for Gr that were active between $t = 200$ and $t = 205$. These changes are reflected in the decrease in PC activity. In a subsequent "test" trial in which no US was presented (bold trace), the retention of the previous training is expressed by the decrease in PC activity that peaks near $t = 200$. This appropriately timed decrease in PC activity occurs because the subset of Gr cells active around $t = 200$ was (1) identical to the subset active at $t = 200$ on the training trial and (2) is different from subsets active at other post-CS intervals. Note that there is an overall decrease in PC activity after training. This decrease occurs because the subset active at $t = 200$ (and that was reinforced by the US) has a low percentage of elements in common with the subsets of Gr cells active at other times. If we consider the population of Gr cells as a binary vector (spike/no spike) with 10^4 dimensions, then each time step will define a point in Gr cell state space. Throughout a stimulus these population vectors will describe a trajectory in this Gr cell state space where different times during the stimulus are encoded by different points along the trajectory. Thus, the degree of overlap (or the average distance in Gr cell state space) determines the signal/noise ratio.

3.2 Ability to Store Multiple Intervals. Simulations similar to those shown in Figure 2 demonstrate that the network is capable of storing multiple intervals in response to the same or different stimuli. Figure 3 illustrates the ability of the network to simulate two different intervals in response to two different stimuli. Each trace represents the response of the network after training with different input stimuli (see figure legend). The first trace (light line) represents the first stimulus that was trained with the US presented 150 msec following the onset of the CS and the second trace (dark line) represents the second stimulus which was trained with a CS–US interval of 350 msec.

3.3 Sensitivity to Noise. A crucial aspect of any realistic biological model is its ability to perform in the presence of noisy inputs. In the context of the present simulations noise refers to trial-to-trial variation in the initial (pre-CS) state of the model and in the MFs activated by the CS as well as variability in various properties of the individual elements, such as threshold, transmitter release, etc. This noise has the potential deleteri-

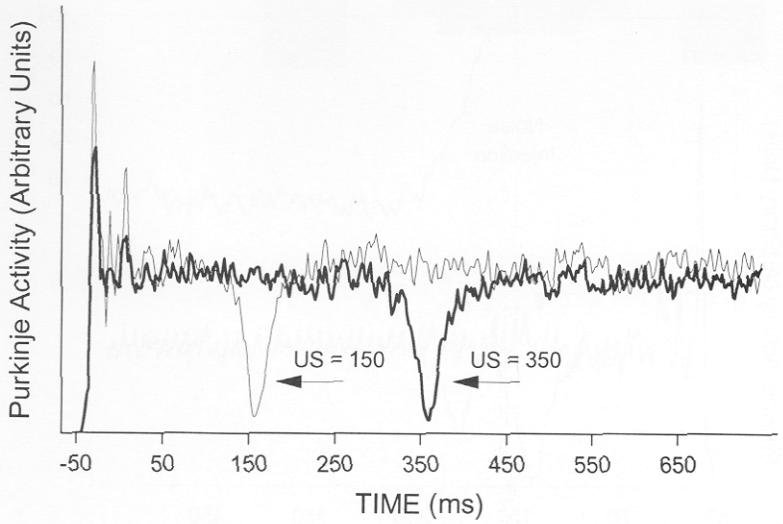

Figure 3: Performing multiple intervals. Simulations similar to those shown in Figure 2 demonstrating that the network is capable of both stimulus and temporal discriminations. The two PC traces represent post-training responses elicited by different CSs (different subsets of MFs were activated for each), each of which was previously reinforced by the US, but at different intervals. The light trace shows the response elicited by the CS previously reinforced at 150 msec and the dark trace was reinforced at 350 msec.

ous effect of reducing the trial-to-trial consistency of the trajectory of the Gr population vector elicited by CS presentations. For example, if each presentation of the same CS promoted a completely different trajectory of the Gr population vector, then the ability to learn an appropriately timed response would be abolished. In the simulations presented thus far no noise was present; the pre-CS state of the network and the CS-evoked MF inputs were identical for each CS presentation. However, since dynamic systems characteristically show sensitivity to noise, the introduction of small errors (noise) leads to increasing divergence of the trajectory.

As a first step toward addressing this issue we have analyzed how the injection of small amounts of noise affects the ability of the network to reproduce a CS-evoked trajectory in the Gr population vector (Fig. 4). Noise was introduced by altering the activity of five Gr cells at $t = 150$ during a CS that had been trained previously. As shown in Figure 4, the introduction of this noise attenuated the PC response and this attenuation

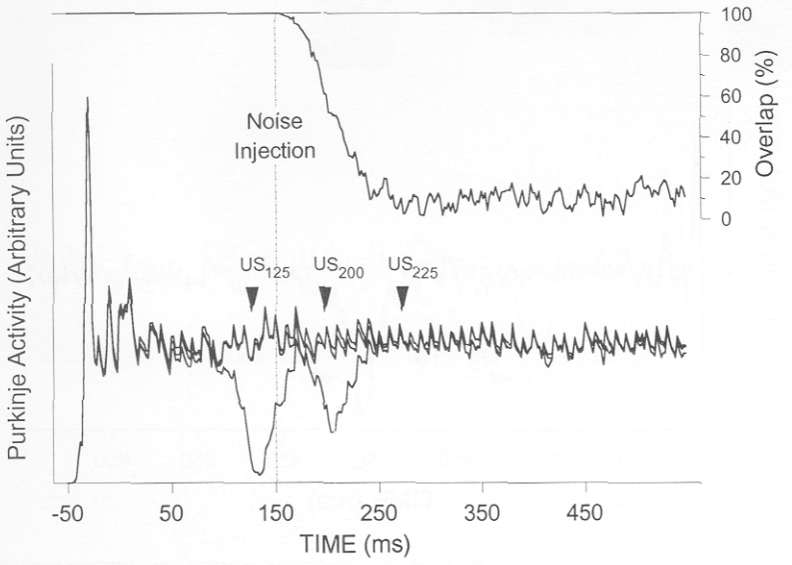

Figure 4: Sensitivity of the network to noise. The lower panel shows PC activity during three separate test trials. Prior to each test trial the network had been trained using a reinforced trial where the US was presented at 125, 175, or 225 msec, respectively, for the three traces. During the test trials, noise was injected at the time shown ($t = 150$) in the form of spurious activity in five randomly chosen Gr cells. The top panel shows the overlap between the Gr cell population vector in the presence and absence of noise. Within approximately 100 msec from the injection of noise ($t = 250$) the overlap decreases to chance. The effect of this decrease on the retention of previous training is illustrated in the PC traces. As the overlap in population vectors of Gr activity between training and test trials decreases, the ability to generate the response deteriorates. Overlap is defined as 1– (Hamming distance ÷ number of active Gr cells), where the Hamming distance is the number of mismatches between the two population vectors.

increased with the time between noise injection and the response. This occurs because the spurious Gr activity changes the trajectory of the network in Gr state space. Since the small initial change leads to further changes in the trajectory, the divergence from the normal trajectory increases with time. With small changes in the trajectory the learned PC response is attenuated, but eventually the divergence is sufficient to eliminate the response completely. The rate at which trajectories are altered

by the introduction of noise can be illustrated by plotting the overlap of Gr cell states during two CS presentations, one with noise and one without (Fig. 4). Before the injection of the noise the overlap between the two Gr cell population vectors is 100%. With the injection of noise the overlap decreases within 100 msec to a baseline level of 10–15%. At this point the ability to generate timed responses is lost. Note that the speed with which the trajectories diverge following noise injection—that is, the slope of the overlap function—can be used as a measure of noise sensitivity (see next section).

3.4 Effects of the MF → Go Connection on Timing and Sensitivity to Noise. The ability of the network to make temporal discriminations was fairly insensitive to most parameters. The parameters best analyzed were the connection strengths from the MF to Gr cells ($W^{Gr:MF}$) and the reciprocal connection strengths between the Gr and Go cells ($W^{Gr:Go}$, $W^{Go:Gr}$). Within reasonable boundaries timing was not dramatically affected by these parameters, although changes in the signal/noise ratio and the absolute number of cells active in the network were observed. We were particularly interested in the MF to Go cell connection ($W^{Go:MF}$), since neither in our model nor in previous cerebellar models is the functional consequence of this connection entirely clear. To obtain insights as to the possible function of this connection we performed a parametric analysis of the effect of the strength of the MF → Go connection on temporal discrimination and on sensitivity to noise.

Figure 5 illustrates the effect of different values of the MF → Go connection on temporal discrimination and sensitivity to noise. We define temporal discrimination as the ability to generate a temporally specific response to the CS. We observed that temporal discrimination is best with low $W^{Go:MF}$ or when the MF → Go connection is omitted ($W^{Go:MF} = 0.0$). With higher $W^{Go:MF}$ values temporal discrimination deteriorates. Conversely the network is more resistant to noise injected into the Gr cells with higher $W^{Go:MF}$ values. At low $W^{Go:MF}$ the dynamics of the network are determined mostly by the Gr → Go → Gr loop, resulting in an unstable but highly variable Gr cell population trajectory (robust temporal discrimination but high sensitivity to noise). At high $W^{Go:MF}$ values the periodic MF input can entrain the network by dominating the input to the Go cells, resulting in little or no temporal discrimination but better resistance (low sensitivity) to Gr cell noise. Thus the MF → Go cell connection may play a role in making the dynamics of the network more resistant to Gr cell noise, and the effectiveness of the MF → Go cell connection in the cerebellum may reflect a compromise between temporal discrimination and resistance to noise. It should be stressed that this conclusion pertains to instances in which the noise is present in the Gr cell activity. It remains to be determined whether the same holds true in the presence of noisy MF inputs.

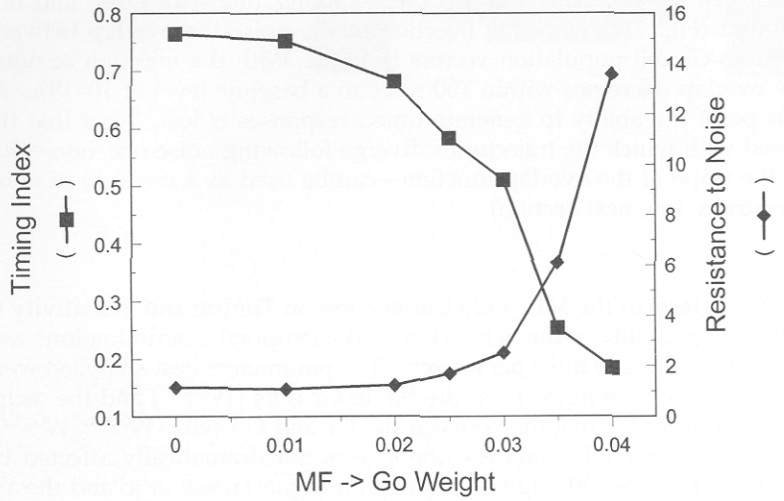

Figure 5: The influence of the MF → Go synaptic strength on the sensitivity of the network to noise and on a measure of the ability of the network to perform temporal discriminations. The injection of noise was identical to the simulations shown in Figure 4. With low MF → Go synaptic strength (or no connection, $W^{Go:MF} = 0.0$) the quality of temporal discrimination or timing was relatively high but resistance to noise was relatively low. As the MF → Go connection was increased, noise resistance increased, but temporal discrimination decreased. Resistance to noise was defined as the reciprocal of the slope (absolute value) of the overlap function shown in Figure 4. Thus, small values reflect a sharp slope of the overlap function, indicating sensitivity to noise. The timing index was defined as the amplitude of the learned decrease in PC activity minus the decrease in the baseline activity following learning. This difference was then normalized to the prelearning baseline activity. Thus the timing index ranges from 0 (no temporal discrimination) to 1 (maximal amplitude of the timed learned response with no change in baseline).

4 Discussion

We have demonstrated that an artificial neural network whose circuitry is based on the synaptic organization of the cerebellar cortex is capable of temporal discriminations and the generation of appropriately timed responses over intervals of tens of milliseconds to seconds. In this network time is encoded in the population vector of Gr cell activity—during

a stimulus the subset of granule cells that is active varies in a characteristic manner. To use Gr cell dynamics to encode particular time windows, there has to be an efficient way to select or to tag the subpopulation of Gr cells that encode a particular time window. It turns out that LTD gated by climbing fiber inputs to PCs provides an effective way to accomplish this. However, it is important to note that the timing mechanism we propose does not necessarily require LTD, any plasticity rule based on temporally correlated climbing fibers and Gr cell inputs would be effective. With LTD, storage of a particular interval is induced by decreasing the Gr → PC connection strength during the desired time window. The retrieval of the interval is expressed as a decrease in the activity level of the PC. Any particular time window can be encoded in this manner because the population of active Gr cells is dynamic. Chapeau-Blondeau and Chauvet (1991) also proposed a model in which time could be encoded by the population of active Gr cells, although their model stresses the importance of the parallel fibers functioning as delay lines.

An interesting feature that emerged from the model is the anticipatory response, that is, after training the response begins before the US onset and peaks during the US. This type of anticipatory response is observed in eyelid conditioning of the rabbit (e.g., Mauk and Ruiz 1992). In the model the anticipatory response occurs because the Gr cells that code for the US interval tend to have a higher probability of being active shortly before and after the US interval. In other words as the Gr cell population trajectory approaches the time of US onset the anticipatory response is generated.

The major weakness of the model is its sensitivity to noise. Spurious activity in a few Gr cells can degrade the trajectory of the network such that within 100–200 msec the population dynamics is altered to a degree that the timing signal is lost. Since timing is generated by the dynamics of the Gr → Go → Gr loop, and encoded in the population of active Gr cells, activity in spurious Gr cells is amplified throughout the network. We did not perform a formal mathematical analysis to determine if the network was chaotic, although the sensitivity of the network to noise suggests that this may be the case.

It has been proposed (Mauk and Donegan 1991) that sensitivity to noise could be responsible for the dependence of Pavlovian conditioning on the interstimulus interval (ISI); conditioning requires that CS onset precede the US by at least 80–100 msec, but not by more than 2–3 s (the so called ISI function). Thus acquisition could require trial-to-trial consistency in the subset of Gr → PC synapses modified by the US, and the ISI function could reflect a within-trial variation of the across-trials CS-evoked trajectory. The present simulations demonstrate that this notion might be feasible. Noise injected early in a CS leads to a time-dependent increase in the variation of the CS-evoked Gr cell trajectories. This increase is consistent with the decrease in conditioning as the ISI increases.

It is clear that the network presented here is too sensitive to noise since changes in a small percentage of Gr cells lead to rapid divergence of the Gr cell trajectory and thus loss of timing. There are several factors that may contribute to the extreme sensitivity of this network to noise. (1) We simulated timing under the worse-case condition in which the network is driven by MF inputs that convey no temporal information. It seems likely that the MFs may convey a small degree of temporal coding—perhaps some MFs fire tonically to a stimulus while others fire phasically at stimulus onset and offset. (2) The parameters used in this model favored robust temporal discrimination at the possible expense of resistance to noise. As shown in Figure 5, with increases in the MF → Go cell connection strength it is possible to improve the resistance to noise and maintain some degree of temporal discrimination. (3) We simulated "learning" with a single trial, it may be that with multiple trials the trade-off between resistance to noise and temporal discrimination may be more forgiving. (4) Aspects of cerebellar physiology not incorporated into the network, such as the correct Gr/Go cell ratios or cellular compartmentalization, may be important. Indeed, preliminary simulations in which the excitatory/inhibitory interactions of the Gr cells were compartmentalized in each dendrite improved the resistance of the network to noise.

Several neural-like models have been proposed to account for the ISI function and for the learned timing of conditioned responses. For example, Moore and colleagues (Moore *et al.* 1986, 1989) presented a neural model that used an eligibility period to obtain the ISI function and tapped delay lines to obtain response timing. Grossberg and Schmajuk (1989) proposed a model using an array of CS-activated elements with different time constants to obtain response timing. Also, Gluck *et al.* (1990) obtained response timing with an array of CS-activated elements that oscillate at different frequencies and phases. Clearly, from a computational point of view it is not difficult to develop hypothetical systems that generate time-varying CS representations. The important questions concerning each candidate mechanism are its overall biological plausibility in general and its ability to map specifically onto the anatomy and physiology of particular brain regions. We suggest that the present approach differs from previous models in that there are no parameters or hypothetical mechanisms included that specifically code for timing or the ISI function. Instead, we have attempted to capture—qualitatively if not quantitatively—the basic properties of the organization of certain parts of cerebellar cortex. Our simulations show that behavioral properties such as response timing and the ISI function can emerge from this organization.

The present model gives rise to several expectations regarding the activity of granule cells that would be elicited by the presentation of a particular stimulus: (1) the presentation of a stimulus should elicit activity in a subset of granule cells, (2) the activity in each cell should follow the onset of the stimulus by a fairly consistent interval, and (3) assum-

ing recordings from a sufficient number of cells, all conditionable intervals should be represented by activity in at least some cells. Assuming again that recordings from a large number of granule cells could be obtained, the possibility that Pavlovian eyelid conditioning occurs only for a limited range of interstimulus interval can be explained by within-trials variation in the across-trials consistency of the population vector of active granule cells Mauk and Donegan (1991) suggests an additional prediction: (4) the trial-to-trial consistency in the stimulus-evoked activity in the granule cells should vary throughout the stimulus in a manner that parallels the ability of different interstimulus intervals to support conditioning. In particular, as the duration of the stimulus (e.g., the ISI) increases the trial-to-trial consistency in the stimulus-evoked granule cell activity should decrease. Thus, given recordings from a sufficiently large subset of granule cells during the presentation of a stimulus, the population vector of the activity of those cells should provide for temporal coding throughout the stimulus and the consistency of this temporal code should parallel the effectiveness of each interval to support Pavlovian eyelid conditioning. While these predictions of the present model are quite concrete, we acknowledge the difficult and time-consuming nature of the experiments. This further highlights the value of combining biologically inspired computer simulations with empirical approaches to investigate the information-processing mechanisms of the nervous system.

Acknowledgments

This research was supported by NIMH Fellowship F31 MH-09895 (DVB), FAPESP Fellowship 91/5111-4 (DVB), NIMH Grant MH46904-02, and Scholars Awards from the National Down Syndrome Society (MDM) and the McKnight Foundation (MDM). We would like to thank John Byrne, Len Cleary, Carlos Tomáz, and Vincent Buonomano for providing the necessary conditions for this research, and Garrett T. Kenyon for helpful comments.

References

Albus, J. S. 1971. A theory of cerebellar function. *Math. Biosci.* **10**, 25–61.

Braitenberg, V. 1967. Is the cerebellar cortex a biological clock in the millisecond range? *Prog. Brain Res.* **25**, 334–336.

Carr, C. E., and Konishi, M. 1988. Axonal delay lines for time measurement in the owl's brainstem. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8311–8316.

Chapeau-Blondeau, F., and Chauvet, G. 1991. A neural network model of the cerebellar cortex performing dynamic associations. *Biol. Cyber.* **65**, 267–279.

Church, R. M., and Broadbent, H. A. 1991. A connectionist model of timing. In *Neural Network Models of Conditioning and Action*, M. L. Commons, S. Grossberg, and J. E. R. Staddon, eds., pp. 225–240. Erlbaum, Hillsdale, NJ.

Eccles, J. C. 1973. The cerebellum as a computer: Patterns in time and space. *J. Physiology* **229**, 1–32.

Eccles, J. C., Ito, M., and Szentágothai, J. 1967. *The Cerebellum as a Neuronal Machine*. Springer-Verlag, New York.

Freeman, J. A. 1969. The cerebellum as a timing device: An experimental study in the frog. In *Neurobiology of Cerebellar Evolution and Development*, R. Llinas, ed., pp. 397–420. American Medical Association, Chicago, IL.

Fujita, M. 1982. An adaptive filter model of the cerebellum. *Biol. Cyber.* **45**, 195–206.

Gluck, M. A., Reifsnider, E. S., and Thompson, R. F. 1990. Adaptive signal processing and the cerebellum: Models of classical conditioning and VOR adaptation. In *Neuroscience and Connectionist Theory*, M. A. Gluck and D. E. Rumelhart, eds., pp. 131–186. Erlbaum, Hillsdale, NJ.

Gormezano, I., Kehoe, E. J., and Marshall, B. S. 1983. Twenty years of classical conditioning with the rabbit. *Prog. Psychobiol. Physiol. Psychol.* **10**, 197–275.

Grossberg, S., and Schmajuk, N. A. 1989. Neural dynamics of adaptive timing and temporal discrimination during associative learning. *Neural Networks* **2**, 79–102.

Ito, M. 1984. *The Cerebellum and Neuronal Control*. Raven Press, New York.

Ito, M., Sakurai, M., and Tonogroach, P. 1982. Climbing fibre-induced depression of both moody fibre responsiveness and glutamate sensitivity of cerebellar Purkinje cell. *J. Physiol.* **324**, 113–134.

Koch, C., Poggio, T., and Torre, V. 1983. Nonlinear interactions in a dendritic tree: Localization, timing, and role in information processing. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2799–2802.

Linden, D. J., Dickenson, M. H., Smeyne, M., and Connor, J. A. 1991. A long-term depression of AMPA currents in cultured cerebellar Purkinje neurons. *Neuron* **7**, 81–89.

Lisberger, S. G. 1988. The neural basis for learning simple motor skills. *Science* **242**, 728–735.

Marr, D. 1969. A theory of cerebellar cortex. *J. Physiol.* **202**, 437–470.

Mauk, M. D., and Donegan, N. H. 1991. A model of eyelid conditioning based on the cerebellum. *Neurosci. Abstr.* **17**, 869.

Mauk, M. D., and Ruiz, B. P. 1992. Learning-dependent timing of Pavlovian eyelid responses: differential conditioning using multiple interstimulus intervals. *Behav. Neurosci.* **106**, 666–681.

McCormick, D. A., and Thompson, R. F. 1984. Cerebellum: Essential involvement in the classically conditioned eyelid response. *Science* **223**, 296–299.

Miall, C. 1989. The storage of time intervals using oscillating neurons. *Neural Comp.* **1**, 359–371.

Moore, J. W., Desmond, J. E., Berthier, N. E., Blazis, D. E., Sutton, R. S., and Barto, A. G. 1986. Simulation of the classically conditioned nictitating membrane response by a neuron-like adaptive element: Response topography, neural firing, and interstimulus intervals. *Behav. Brain Res.* **21**, 143–154.

Moore, J. W., Desmond, J. E., and Berthier, N. E. 1989. Adaptively timed conditioned responses and the cerebellum: A neural network approach. *Biol. Cybern.* **62**, 17–28.

Overholt, E. M., Rubel, E. W., and Hyson, R. L. 1992. A circuit for coding interaural time differences in the chick brainstem. *J. Neurosci.* **12**, 1698–1708.

Palkovits, M., Magyar, P., and Szentágothai, J. 1971. Quantitative histological analysis of cerebellum in cat. III. Structural organization of the molecular layer. *Brain Res.* **34**, 1–18.

Perrett, S. P., Ruiz, B. P., and Mauk, M. D. 1993. Cerebellar cortex lesions disrupt learning-dependent timing of conditioned eyelid responses. *J. Neurosci.* **13**(4), 1708–1718.

Schneiderman, N., and Gormezano, I. 1964. Conditioning of the nictitating membrane of the rabbit as a function of CS-US interval. *J. Comp. Physiol. Psychol.* **57**, 188–195.

Thompson, R. F. 1986. The neurobiology of learning and memory. *Science* **233**, 941–947.

Welsh, J. P., and Harvey, J. A. 1989. Cerebellar lesions and the nictitating membrane reflex: Performance deficits of the conditioned and unconditioned response. *J. Neurosci.* **9**, 299–311.

Yeo, C. 1991. Cerebellum and classical conditioning of motor responses. *Ann. N. Y. Acad. Sci.* **627**, 292–304.

Received October 15, 1992; accepted April 9, 1993.