Proceedings of the 3rd Joint
Sympostuim on Neural
Computation, La Jolla, CA
Vol. 6, 202-221 (1996)

Learning to make predictions in the cerebellum may explain the anticipatory modulation of the vestibulo-ocular reflex (VOR) gain with vergence

Olivier J.M.D. Coenen* Terrence J. Sejnowski†

Computational Neurobiology Laboratory
Howard Hughes Medical Institute
The Salk Institute for Biological Studies
10010 North Torrey Pines Road
La Jolla, CA 92037, U.S.A.

Departments of †Biology and * † Physics
University of California, San Diego
La Jolla, CA 92093, U.S.A
{olivier,terry}@salk.edu

Abstract

Changes in the eye vergence modify the gain of the vestibulo-ocular reflex (VOR). In a previous dynamical model, this modulation was controlled by the cerebellum using vergence angle information (Coenen & Sejnowski; NIPS 96). However, during a vergence eye movement, the change in the VOR gain anticipates the vergence change (Snyder & King; Vision Res. 32:3, 92). We demonstrate here how our previous dynamical model and a predictive cerebellar model may be combined to explain these findings. In the predictive model, the cerebellum receives inputs from vergence-disparity cells in the cortex to construct a prediction of vergence angle. By replacing the regular vergence input in our previous dynamical model by the vergence prediction, results similar to the experimental anticipatory gain changes are observed. The inferior olive in the predictive cerebellar model is responsible for computing a prediction error and for selecting the Purkinje cells to be recruited for learning a particular prediction. The learning model is based on the least-mean square (LMS) algorithm to construct predictions from previous context information. This model is a special case of a more general predictive function for the cerebellum that may provide a consistent explanation for the apparently ubiquitous task participation of the cerebellum. We discuss briefly how this more general model may explain some experimental results observed in the cerebellum with positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). In conclusion, a predictive cerebellar model receiving vergence-disparity cell inputs learns to predict the time course of vergence eye movement and successfully changes the VOR gain in anticipation of vergence changes.

1 Introduction

The cerebellum is involved in motor timing (Ivry and Keele, 1989), motor coordination (Goodkin et al., 1993), motor learning (Molchan et al., 1994; du Lac et al., 1995) and sensorimotor integration (Stone and Lisberger, 1990). For example, cerebellar contributions have been inferred in situations as diverse as timing of the conditioned eyelid response (Perrett and Ruiz, 1993), shifting of attention (Akshoomoff and Courchesne, 1992; Corbetta et al., 1993), adaptation of the vestibulo-ocular reflex (Lisberger et al., 1994) and coordination of eye and hand motor systems (van Donkelaar and Lee, 1994). Some of these studies also suggest that the cerebellum may be involved in some cognitive aspects of information processing (Andreasen et al., 1995; Kim, Uğurbil and Strick, 1994; Leiner, Leiner and Dow, 1993; Middleton and Strick, 1994; Tachibana, 1995; Raichle et al., 1994). Several theories of cerebellar function have been proposed, including the original motor learning theories of Marr (1969), Albus (1971) and others (Bloedel, 1992; Chapeau-Blondeau and Chauvet, 1991; Darlot, 1993; Fujita, 1982; Ito, 1984; Keeler, 1990; Kawato and Gomi, 1992; Leiner, Leiner and Dow, 1989; Llinas and Welsh, 1993; Miall et al., 1993; Paulin, 1989; Thach, Goodkin and Keating, 1992). Few of these theories have attempted to give a consistent view of the role of the cerebellum in the diverse tasks in which the cerebellum participates.

We present a general predictive function for the cerebellum that we argue provides a consistent explanation for the observed phenomena. We propose that the cerebellum is specialized to predict neural signals, and therefore is particularly successful in anticipating the temporal sequences of events experienced repeatedly. Predictions of neural activity can contribute to many aspects of behavior from motor control to cognitive strategies. This could elucidate the apparently ubiquitous participation of the cerebellum and more specifically provide an explanation for the observed decrease of regional cerebral blood flow in the cerebellum with learning through repeated experience, as well as its increase with task complexity and with unexpected events (Flament et al., 1995; Flament et al., 1996; Friston et al., 1992; Molchan et al., 1994; Raichle et al., 1994; Sadato et al., ; Seitz et al., 1994). We first present our general predictive framework, and then demonstrate how it may be applied to model experimental results on the anticipatory modulation of the gain of the vestibulo-ocular reflex (VOR) with vergence eye movements.

2 Cerebellar prediction: Experimental evidence and theory

Evidence that a predictive representation is constructed in the cerebellum is provided by the experimental results of Dugas and Smith (1992). They trained monkeys to grasp, lift and hold an object between thumb and index finger for one second. On selected blocks of trials, a perturbation was applied to the held object to simulate object slip. They clearly showed that Purkinje cells located in the hand representation area lateral to the vermis in lobules IV-VI acquire an anticipatory response to the perturbation in order to stabilize the position of the hand. They also demonstrated a gradual extinction of the anticipatory response after the perturbation was withdrawn.

The contribution of the cerebellum in this task is similar to its role in eyelid conditioning. In eyelid conditioning, the conditioned response timing is apparently controlled by the anterior lobe of the cerebellum and occurs before the unconditioned stimulus (Perrett and Ruiz, 1993; Perrett and Mauk, 1995). Thus, we may infer that the cerebellum participates in constructing a predictive response from the conditioned stimuli.

Purkinje cell responses in the flocculus and ventral paraflocculus are modulated during smooth pursuit eye movements (Stone and Lisberger, 1990). We also know that predictive control of smooth pursuit eye movements occurs for complex two-dimensional tracking trajectories in monkey (Leung and Kettner, 1995) and human (Barnes and Asselman, 1991). The cerebellum is therefore a possible site for the predictive learning of smooth eye movements, and it has been modeled with this perspective (Kettner et al., 1995).

In this section we present the mathematical theory on which the predictive cerebellar model is based.

We clarify the correspondence between the element of this theory and the cerebellum in the next section.

Linear predictor Given a signal $y(t)$, we define its prediction using all data available until time $t-\tau$ as $\hat{y}(t|t-\tau)$. This prediction can be written more generally as the prediction of $y(t)$ available at time $t-\tau_1$ using all data available until time $t-\tau_2$ as $\hat{y}(t-\tau_1;t|t-\tau_2)$. In a system with synaptic and conduction delays, t, $t-\tau_1$, and $t-\tau_2$ are different; we assume $\tau_2 > \tau_1$.

In the linear case, the prediction is constructed from a linear combination of input signals $\mathbf{x}(t) = \{x_1(t), x_2(t), ..., x_n(t)\}$:

$$\hat{y}(t-\tau_1;t|t-\tau_2) = \text{Delay}(\tau_s) \sum_i w_i(t-\tau_2)\, x_i(t-\tau_2) \tag{1}$$

where w_i are adaptive coefficients or weights, and $\text{Delay}(\tau_s)$ is a transport delay operator delaying the terms to the right by τ_s, for example τ_s here may represent synaptic and conduction delays between the location where the x_i originate and $\hat{y}$ is calculated, and by definition $\tau_2 \equiv \tau_1 + \tau_s$.

The weights are adapted using *least-mean-square* (LMS) *algorithm* for the Wiener filter (Eq. 1). Defining the prediction error,

$$\varepsilon(t) = y(t) - \hat{y}(t;t|t-\tau_2) \tag{2}$$

the LMS algorithm is a stochastic gradient algorithm that minimizes the cost function

$$J(t) = \int_0^t \varepsilon^2(\nu)\, d\nu \tag{3}$$

using the learning rule

$$w_i(t) = w_i(t) + \mu_i\, \delta(t-\mathbf{t}_o)\, \varepsilon(t)\, x_i(t-\tau_3) \tag{4}$$

where the learning rate μ_i may be different for each input x_i. Here, weight updates may be performed at times given by $\mathbf{t}_o = \{t_{o_1}, t_{o_2}, ..., t_{o_\infty}\}$ which need not be at equal intervals. Best convergence results will be obtained in general when $\tau_3 = \tau_2$, since for this value the prediction error is correlated with the inputs x_i that directly caused the error (see also Kabal, 1990; Long, Ling and Proakis, 1989; Long, Ling and Proakis, 1992)

In order to compute the prediction error, $\hat{y}(t-\tau_1;t|t-\tau_2)$ must be delayed by τ_1 before a comparison is made since

$$\hat{y}(t;t|t-\tau_2) = \text{Delay}(\tau_1)\, \hat{y}(t-\tau_1;t|t-\tau_2) \tag{5}$$

Note that the value of the inputs at a time τ_3 earlier, $x_i(t-\tau_3)$, must be stored in order to calculate the weight updates.

Let us assume that the site where $\hat{y}(t)$ originates is different than the site where the prediction error $\varepsilon(t)$ is calculated, that it takes a conduction time τ_1 for $\hat{y}(t)$ to reach the ε site, and that, as above, there is a synaptic delay τ_s between $\mathbf{x}(t)$ and $\hat{y}(t)$. We know that $\tau_2 \equiv \tau_1 + \tau_s$, and that in general for learning to converge we require τ_3 to equal τ_2, i.e. $\tau_3 \equiv \tau_1 + \tau_s$.

The conduction time τ_1 delays the feedback to allow calculation of the prediction error; this determines the prediction time of $\hat{y}$, and also constrains the value of τ_3 in the learning rule. Unlike a Smith predictor model of the cerebellum which learns τ_1 as part of the controller (Miall et al., 1993), in this system τ_1 is not learned but it used by the system to establish predictions that are then available for other purposes to other parts of the whole system. If the system is constructed to have parallel pathways with a range of

different conduction times $\tau_1 = \{\tau_{1_1}, \tau_{1_2}, ..., \tau_{1_m}\}$ between the $\hat{y}$ site and the ϵ site, then a set of predictions $\hat{\mathbf{y}}(t - \tau_1; t|t - \tau_2)$ with different prediction times can be constructed. A selection process could then occur to pick out the most reliable predictions. Given a set τ_1, a similar set τ_3 must also exist such that $\tau_3 = \{\tau_{3_1}, \tau_{3_2}, ..., \tau_{3_m}\}$. Therefore a good correspondence between the units $\hat{\mathbf{y}}$ and the memory mechanism which stores the values $x_i(t - \tau_3)$ must exist, since learning is only guaranteed to converge when $\tau_3 = \tau_1 + \tau_s$ (Haykin, 1994; Long, Ling and Proakis, 1992).

If the learning rates $\mu_i = \mu$ for all i, a condition on the learning rate for convergence of the LMS algorithm can be derived. The LMS algorithm is convergent in the mean square if the learning parameter μ follows the condition (Haykin, 1994):

$$0 < \mu < \frac{2}{\mathrm{tr}[\mathbf{R}_{xx}]} \tag{6}$$

where $\mathrm{tr}[\mathbf{R}_{xx}]$ is the trace of the correlation matrix $\mathbf{R}_{xx} = E[\mathbf{x}(t)\mathbf{x}^T(t)]$, where $\mathbf{x}(t)$ is sampled at regular interval to form a vector. The trace of the correlation matrix $\mathbf{R}_{xx}$ is equal to the total input power measured over all inputs. The input signals used in the predictive filter therefore determine the maximum allowable learning rate to insure convergence. If this type of algorithm applies to the cerebellum then the learning rate must be quite small. (For further convergence results, see Kabal, 1990; Long, Ling and Proakis, 1989; Long, Ling and Proakis, 1992).

The incremental delta-bar-delta (IDBD) algorithm (Sutton, 1992) may be used to adjusts the learning rates μ_i based on previous learning experience. The IDBD algorithm performs better than ordinary LMS and in fact can find the optimal learning rates for nonstationary learning tasks.

There are multiple possibilities within a dynamical system to maintain a short-term memory of $x_i(t-\tau_3)$. One possibility is to use Klopf's eligibility trace (Klopf, 1988; Barto, Sutton and Anderson, 1993). The eligibility trace is a dynamical variable which follows $x_i(t)$ and decreases exponentially when $x_i(t)$ goes to zero. Another possibility is to use a slightly more complicated dynamics such that the variable peaks to a value proportional to its input with a certain time delay. This has been described as a "window of eligibility" (Arbib, Schweighofer and Tach, 1994). (See also the "replacing trace" (Singh and Sutton, 1996)).

3 Cerebellum

3.1 Brief review of cerebellar anatomy

3.1.1 Cerebellum

A brief summary of the principal features of the cerebellum is given in fig. 1. There is a much greater degree of divergence from the mossy fibers to granule cells than shown in the figures. The mossy fibers form 20 to 50 or more rosettes in a folia of the cerebellum. At a glomerulus, Golgi cell axons and one mossy fiber rosette contact the dendrites of about 20 granule cells. Hence one mossy fiber contacts 400-1000 granule cells within one folia. One mossy fiber can also run into another folia. Each granule cells receive inputs from 1 to 7 mossy fibers. There is one Golgi cell per 3-10 Purkinje cells. One Golgi cell contacts all granule cells in its vicinity (4000-6000). From 100 000 to 400 000 parallel fibers pass through the dendritic tree of a Purkinje cell, but only 20% (80 000) have synaptic contacts – this is a calculated estimate based on different assumptions regarding the synaptic contacts. A parallel fiber has a length that ranges from 0.6 mm in the mouse to 2.6 mm in man. It passes through the dendritic tree of about 225 Purkinje cells, and only synapses presumably on about 45. Basket cells are 10-20% more numerous than Purkinje cells. Each basket cell can inhibit up to 50 Purkinje cells over an elliptical area with its main axis along the sagittal plane of the cerebellum. Purkinje cells also send collaterals along that axis. (Nieuwenhuys, Voogd and van Huijzen, 1988 and references therein; Palay and Chan-Palay, 1974; Albus, 1971; Marr, 1969)

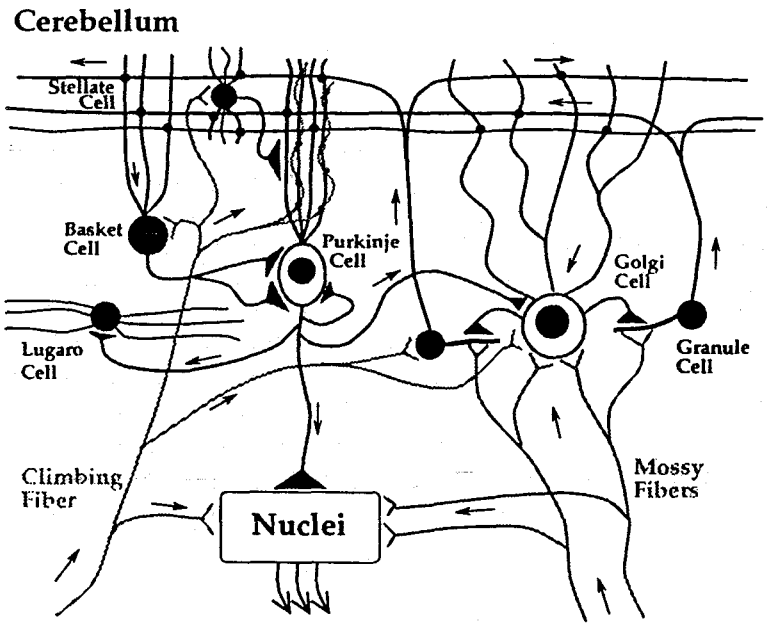

Figure 1: Circuit diagram of the climbing fiber, Purkinje cell, and mossy fiber connections and interneuronal relays. The connections shown in this diagram are based on synaptic contacts that have been confirmed by electron microscopy. All cells in the cerebellum except for granule cells and their parallel fibers are inhibitory cells. Filled triangular synaptic connections are inhibitory. Inputs to the cerebellum enter either as mossy fibers connecting to granule cells or as climbing fibers coming from the inferior olive. The climbing fiber contacts Golgi cells and granule cells in the granular layer. In the molecular layer it has multiple contacts with the Purkinje cell dentritic tree and in addition synapses with basket cells and stellate cells in the surround. The synaptic effect of climbing fibers on interneurons is still undetermined. The mossy fibers contact granule cells, Golgi cells, and the cerebellar nuclei. Granule cell axons synapse with with Golgi cells, Purkinje cells, basket cells, stellate cells and Lugaro cells (connection not shown). Stellate cell and basket cell axons contact the Purkinje cell; Golgi cells synapse with granule cells; Lugaro cell output is yet unknown. The only outflow of the cerebellar cortex is through the Purkinje axons which contact the deep cerebellar nuclei. Recurrent collaterals in the cortex contact Golgi cells, Lugaro cells, and other Purkinje cells mainly along the parasagittal axis (not necessarily itself as shown in this diagram). Each Purkinje cell receive only one climbing fiber; one climbing fiber contact approximately ten Purkinje cells. The parallel fibers drive the Purkinje cells which fire simple spikes at a rate of around 20-100 Hz, generating a modulated inhibition of the deep cerebellar nuclei. The climbing-fiber activity produces complex spikes in Purkinje cells; this short burst (10 ms) of about five spikes strongly inhibits the deep cerebellar neurons. Anatomical connections based on fig. 230 in Palay and Chan-Palay (1974).

3.1.2 Inferior olive

Climbing fibers originate form the contralateral inferior olive and reach the cerebellum where they branch preferentially in the parasagittal plane. Olivocerebellar fibers which terminate as climbing fibers in a particular zone of the cerebellum also send collaterals to the deep cerebellar nucleus which receives its Purkinje cell projections. Collaterals have also been observed to contact inhibitory interneurons and mossy fiber rosettes on granule cells (Nieuwenhuys, Voogd and van Huijzen, 1988 and references therein).

Olivary neurons have a propensity to oscillate (Llinas and Yarom, 1986; Lampl and Yarom, 1993) and to fire synchronously (Llinas and Yarom, 1981). Their synchronous firing is the result of electrotonic coupling by dendrodentritic gap junctions (Llinas and Yarom, 1981).

The topographical organization of the projections between the cerebellum and the deep nuclei and between the inferior olive and the cerebellum is further maintained by reciprocal connections between the

deep nuclei and the subdivisions of the inferior olive from which they receive their collateral projections. A special population of small GABAergic (inhibitory) neurons of the deep nuclei form the direct and reciprocal projection to the inferior olive. Double labeling studies show that GABAergic fibers from the deep nuclei terminate preferentially to perikarya (cell bodies), and next to gap junctions (De Zeeuw et al., 1989) of olivary neurons. The inferior olive receive excitatory afferent connections from a large number of nuclei. Some parts of the inferior olive also receive a serotonergic projection from the raphe nuclei.

At least 93% of GABAergic neurons in the cerebellar nuclei of rats are retrogradely labeled following injection in the inferior olive (Fredette and Mugnaini, 1991). Purkinje cell terminals have been shown to innervate both GABAergic and non-GABAergic (presumed mostly excitatory from their morphology) neurons in both the cerebellar and vestibular nucleus of the rat (De Zeeuw and Berrebi, 1995). The percentage of Purkinje cell axons observed that contact both GABAergic and non-GABAergic neurons is greater than that of axons which contact only one type. Taking into account this and further experimental evidence, it is likely that most of Purkinje cell axons innervate both GABAergic and non-GABAergic neurons (De Zeeuw and Berrebi, 1995). This evidence further strengthens the existence of olivocerebellar modules formed by three-element loops consisting of specific olivary subnuclei, specific cerebellar sagittal zone of Purkinje cells, and specific cerebellar nuclei (Voogd and Bigaré, 1980).

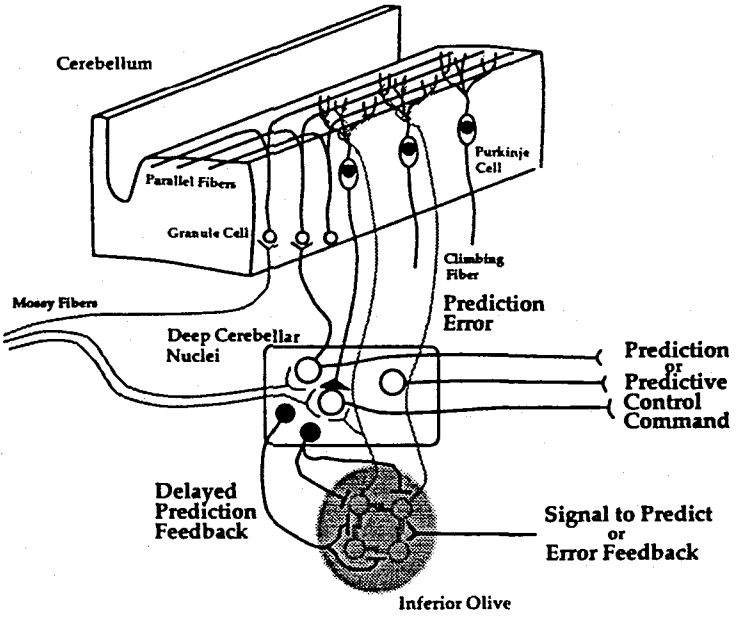

Figure 2: Hypothesis on the functions of the cerebellum and the deep cerebellar nuclei. The deep cerebellar nuclei together with the inhibitory inputs they receive from the cerebellar cortex encode either predictions or predictive motor commands. The actual representation may depend on whether the inferior olive receives direct inhibitory input feedback from the deep cerebellar nuclei or indirect inhibitory inputs from other nuclei. The inferior olivary neurons are electrotonically coupled as indicated by junctional contacts between short dendrites. They received two types of inhibitory GABA-ergic inputs from the deep cerebellar nuclei, one type terminating at the dendro-dendritic gap junctions, and another type terminating at the perikaryon (cell body). The GABA-ergic fibers terminating at the perikaryon are used to calculate the prediction error. The fibers terminating at the gap junctions may be used to modulate the number of olivary neurons firing in synchrony, and therefore the number of climbing fibers reporting a particular prediction error. This consequently may change the number of Purkinje cells involved in a particular prediction.

3.2 Cerebellar Predictions: Implementation

The basic idea of our approach is shown in fig. 2. The deep cerebellar nuclei and the cerebellar cortex form together in our framework a predictive machine that is under the regulatory control of the inferior olive. The predictions being constructed are predictions of neural activities related to the excitatory inputs reaching the inferior olive. Depending on the source of inhibitory inputs to the inferior olive, the output of the cerebellum may be described as predictions or as predictive control commands. We propose that the inhibitory inputs to the inferior olive from the deep cerebellar nuclei carry a delayed feedback of the predictions being established in the deep cerebellar nuclei. The internal circuitry of the deep cerebellar nuclei and feedback connections with other nuclei may provide the time delays required for our hypothesis.

The predictions $\hat{y}(t)$ are built continuously from neural contextual information $x_i(t)$ preceding a "neural event" $y(t)$ corresponding to the activity (firing frequency) of excitatory inputs to the inferior olive. $x_i(t)$ are the activities of both the direct inputs from mossy fibers in the deep nuclei and parallel fibers in the cerebellar cortex. $\hat{y}(t)$ are the activities of deep cerebellar projection neurons.

Considered in isolation from the cerebellar cortex projections, the deep nuclei activity encodes for very immediate predictions (or predictive motor commands – reported as reflexes) with little time between contextual information and signal to be predicted. The cerebellar cortex provides longer time intervals between context and predictions of events than the deep cerebellar nuclei alone would provide. The quality of the predictions is determined by the richness of the contextual information and the repeatability of the neural events observed. The cerebellar cortex by receiving inputs from nearly the entire neuraxis provides this richness of contextual information to provide the best predictions. The short-term predictions given by the deep cerebellar nuclei can be inhibited by the cerebellar cortex and replaced by longer-term predictions made possible by the spectrum of neural activity available in the cerebellar cortex through the large number of granule cells.

3.3 Inferior olive

Role of Synchronicity and Learning The inferior olivary neurons are electrotonically coupled. They receive two types of inhibitory inputs from the deep cerebellar nuclei, one type terminating at the dendrodendritic gap junctions, and another type terminating at the perikaryon (cell body). The GABAergic (inhibitory) fibers terminating at the perikaryon are used in the model to calculate the prediction error $\varepsilon(t)$ (Eq. 2).

The other fibers terminating at the gap junctions may be used to modulate the number of olivary neurons firing in synchrony and therefore change the number of climbing fibers reporting a particular prediction error. This consequently changes the number of Purkinje cells involved in a particular prediction. This adaptive process at the level of the inferior olive may be extremely important in order to recruit the group of parallel fibers ($x_i(t)$) encoding the most relevant signals for constructing the prediction $\hat{y}$.

Consider for example an eyeblink conditioning experiment. During the experiment, a tone or a light (conditioned stimuli, CS) is paired with an air puff (unconditioned stimulus, US). The air puff alone produces the closure of the eyelid (unconditioned response, UR). After learning, the tone or the light alone produces the eyeblink (conditioned response, CR). Information about the air puff reaches the inferior olive, and information about the tone or light is available through mossy fibers at the deep cerebellar nuclei and cerebellar cortex (Gluck and Thompson, 1990). For the purpose of this discussion, we make the simplifying assumption that a prediction of the US by the CS can cause a CR, in the same manner and perhaps with the same circuitry used when a US causes a UR. The shape of the CR in time will be a direct consequence of the shape of the prediction of the US in time.

At the beginning of the experiment, the olivary neurons receiving the air puff sensory signals may not project to Purkinje cells with parallel fiber afferents carrying either tone or light signal. Over a short period

of adaptation by the regulation of their gap junctions, these olivary neurons may fire in synchrony with other olivary neurons that do project to such Purkinje cells. A first stage of "propagation" of the air puff sensory information in the inferior olive may therefore have to occur before learning of the pairing may begin.

The recruitment of olivary neurons may stop as soon as relevant prediction feedback from the deep nuclei inhibits the gap junctions. This presumably can only occur once Purkinje cells receiving tone or light information have started constructing the appropriate prediction of the air puff. The regulatory feedback of gap junctions in the inferior olive may therefore control a search over a group of Purkinje cells to find those receiving appropriate parallel fiber information to construct the prediction.

Once a coarse prediction has been establish, presumably a smaller number of olivary neurons reporting the prediction error will be activated than during the search. If we assume that the firing rate of the olivary neurons increases with prediction error, the activity of this smaller set of olivary neurons will decrease with learning. Therefore we expect the activation of the inferior olive to go through two distinct phases with a smooth and continuous transition between the two. The first phase is a search to select the appropriate deep cerebellar neurons and Purkinje cells which received appropriate afferent information to construct the prediction, and the second phase consists of fine tuning the responses of these neurons. The inferior olive activity is therefore expected to be large at the onset of an experiment, in the naive state, and to decrease gradually to baseline with learning.

Since climbing fiber firing produces complex spikes in Purkinje cells, the frequency of complex spikes in a Purkinje cell increases with the firing frequency of the corresponding olivary neuron. Furthermore since one climbing fiber contacts approximately ten Purkinje cells, the number of Purkinje cells firing complex spikes synchronously is approximately ten times larger than the number of olivary neurons firing in synchrony. In the light of the previous paragraph, we therefore expect complex spike activity in an area of cerebellar cortex to be vastly larger in a naive state than after learning.

The precise relationship between regional cerebral blood flow (rCBF) in fMRI and PET and neural activity has not so far been clearly determined. Nevertheless there may be some direct relationship between the two. It may be that in the cerebellar cortex the greatest change of rCBF during an experiment is correlated with complex spikes activity since the simple spike activity is as likely to go up or down during the course of learning, and therefore are more likely to average out than the complex spike rCBF related activity. This could explain the observed decreases of rCBF in the cerebellum with learning, as well as the observed increases with task complexity and unexpected events. The presumption is that in more complicated tasks, good predictions are harder to establish than in simpler task; prediction errors would persist for a longer period, and the search in the inferior olive would recruit a larger area of cerebellar cortex. For unexpected events, since no previous predictions exist the prediction errors would also be large. To test this hypothesis, experiments should be designed in which the predictability of the events are varied but where the amount of movement in motor responses stays constant.

4 Application: Anticipatory VOR modulation

In this section we apply the general framework above to the oculomotor system, specifically to the modulation of the VOR.

4.1 The vestibulo-ocular reflex is modulated with vergence.

A head motion induces a compensatory eye movement to keep the eyes fixed on an object to stabilize its image on the retina and therefore reduce blur. The ratio of eye velocity over head velocity is called the gain of the vestibulo-ocular reflex (VOR). The gain of the VOR in monkeys may be modulated by gaze direction (eccentricity), linear head motion, and viewing distance (fig. 3) (Snyder & King, 1992). Therefore

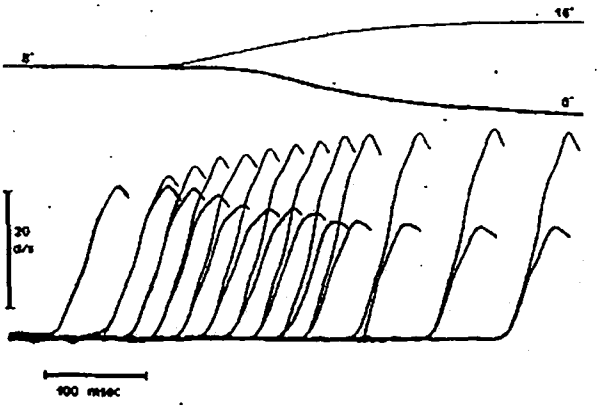

Figure 3: VOR responses elicited during 8 deg vergence eye movements. Upper traces show vergence eye movements from a target located at 8 deg vergence angle to a near target (upper trajectory, 16 deg vergence) or to a far target (lower trajectory, 0 deg vergence). Lower traces show eye velocity responses evoked by brief head rotation at different times during vergence eye movement. The upper superimposed trajectories correspond to convergence (from 8 deg to 16 deg) and the lower trajectories to divergence (from 8 deg to 0 deg). The modulation of the VOR gain with vergence is clearly apparent. The gain increases during convergence and conversely decreases during divergence. Figure from Snyder & King (1992).

the VOR gain may be modulated by many neural inputs including eye position signals, otolith signals and vergence signals. We have developed a dynamical model of the VOR modulation in which these neural inputs are combined nonlinearly to reproduce the data of Snyder & King (1992) (Coenen and Sejnowski, 1996). The dynamical model is consistent with cerebellar anatomy and physiology and suggests a plausible cerebellar contribution to the modulation of the VOR. Although the gain of the VOR depends on viewing distance, the neural correlate of viewing distance that modulates the VOR is not known. We assumed in our dynamical model that a vergence signal was available and was combined with other sensory informations in the cerebellum to modulate the VOR.

Snyder, Lawrence and King (1992) have shown that changes in VOR actually anticipate changes in vergence angle in monkey on average by 50 ms and up to 200 ms in some instances (fig. 4). They suggested that a central command signal rather than a copy of vergence eye position is used to modulate the VOR. Vergence eye movements can be initiated and directed by visual disparity and blur information (Cumming and Judge, 1986). Disparity selective cells in monkey primary visual cortex form the first stage of visual disparity information (Poggio et al., 1985). More than 80% of disparity selective cells are also modulated by viewing distance (Trotter, 1995; Trotter et al., 1992) . We suggest that signals derived from the dynamical response of disparity selective cells (DeAngelis, Ohzawa and Freeman, 1995) may be used by the cerebellum to provide an anticipatory signal for vergence eye position. This anticipatory vergence signal may then reenter the cerebellum to modulate the VOR. The cerebellum may therefore anticipate the outcome of the slower vergence plant to predictively modulate the VOR.

4.2 VOR learning: Indirect Method

The problem of learning an anticipatory modulation of the VOR gain may be solved with a direct or an indirect approach. In the direct method, the VOR error or retinal slip drives learning directly, and the cerebellum does not build any intermediate representation of the system (Coenen, Sejnowski and Lisberger, 1993). The vestibular nucleus and cerebellum select and modify sensorimotor inputs and feedback to reduce the retinal slip directly. In the indirect method, an internal and predictive representation of individual neural signals

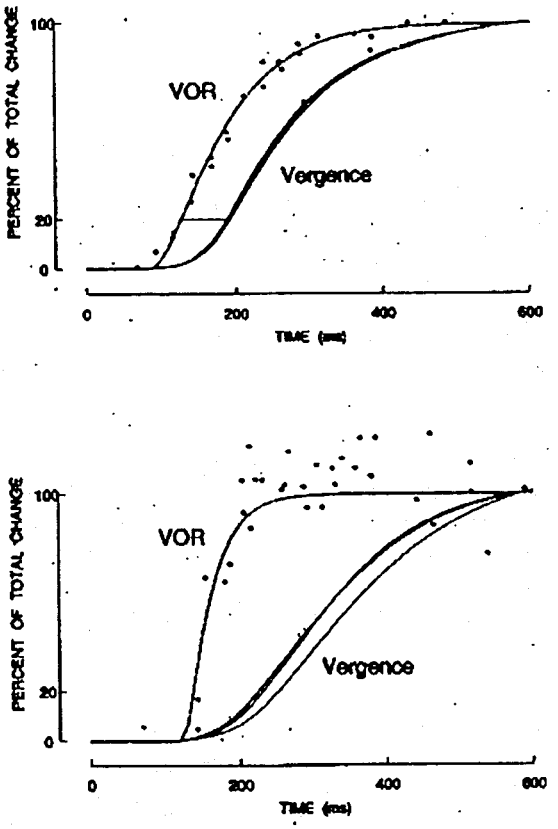

Figure 4: Percentage of total change of the VOR gain and vergence angle occurring during 8 deg vergence eye movements. Visual stimuli for the vergence eye movements occurs at time zero. The solid lines plots the normalized vergence angle as a function of time obtained on three test days. The dashed line shows a fit to the normalized eye velocity peaks obtained from brief head rotation at different times during vergence eye movements (see fig. 3). Upper graph uses the data from the experiment shown in fig. 3. Lower graph shows data from a second monkey. Taken from Snyder & King (1992).

may be built to solve a larger goal. In this case, the overall problem of reducing retinal slip may be decomposed into smaller subproblems which may be easier to build and update. One possibility is that a predictive representation of vergence angle is first constructed by the cerebellum, and then used through feedback to the cerebellum to modulate the VOR. It is the feasibility of this approach that we study here.

4.3 Vergence and accommodation

Vergence eye movements can be initiated and directed by visual disparity and blur information (Cumming and Judge, 1986). When a visual target is moved abruptly from a mid-distance to a near or far distance, the disparity of a visual target changes suddenly. This change in disparity can initiate a vergence eye movement and accommodation to focus the target. We model this behavior using a dynamical model of accommodation and convergence (Schor, 1992). The model uses cross-coupling between accommodation and vergence to accurately tune these two motor systems. Fig. 5 shows the vergence response of the model to an initial disparity step of 8 deg; the following time course of disparity is also plotted. The model is based on human data, and the parameters described in Schor (1992) were used; this explains the longer time course compared

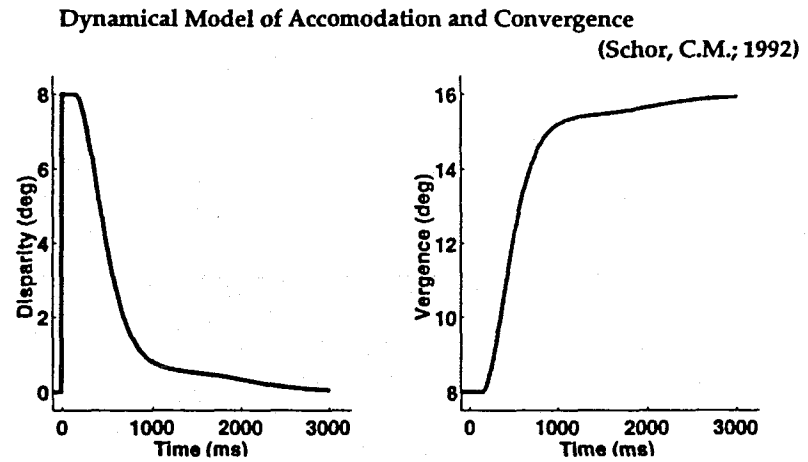

Figure 5: Response generated by the dynamical model used to simulate vergence eye movement (Schor, 1992). At time zero, a simulated visual target located 18 cm in front of an hypothetical subject is moved to 9 cm. This produces a disparity step of 8 deg (left) which initiates a vergence eye movement about 200 ms later (right). The time course of disparity (left) is shown during the following vergence eye movement (right) to foveate the target.

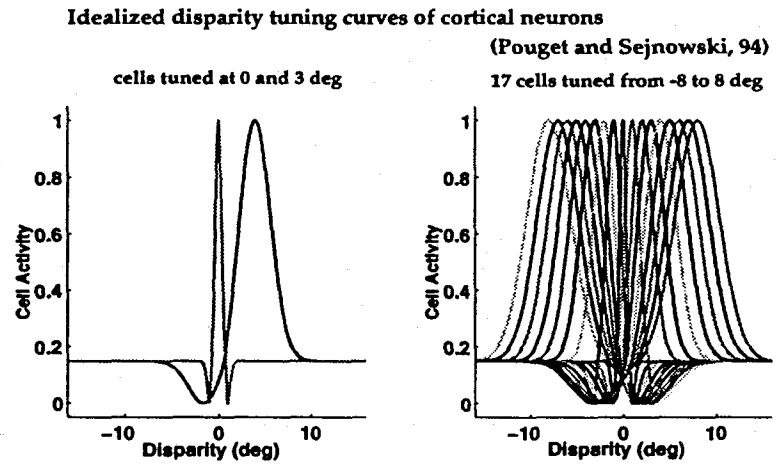

Figure 6: Idealized disparity tuning curve of cortical neurons used in the model (Pouget and Sejnowski, 1994). On the left the tuning curves of two cells tuned at 0 and 3 deg are plotted individually for clarity. On the right, the tuning curves of the 17 cells tuned between -8 and 8 deg used in the model are shown.

to rhesus monkeys (fig. 3), otherwise binocular vision in the two species is very similar (Harwerth, Smith 3rd and Siderov, 1995).

4.4 Visual inputs: disparity-vergence cells

In this section we introduce the tuning curves and the dynamics of the visual input cells that we used in the model. We first describe their tuning curves to disparity and vergence and then specify their dynamics.

Many cells in the primary visual cortex and extrastriate areas are selective to horizontal disparity (Poggio et al., 1985). Originally these cells were classified into three groups: near, tuned and far cells. Recent psychophysical studies have provided evidence that in fact there exists disparity selective cells responding over

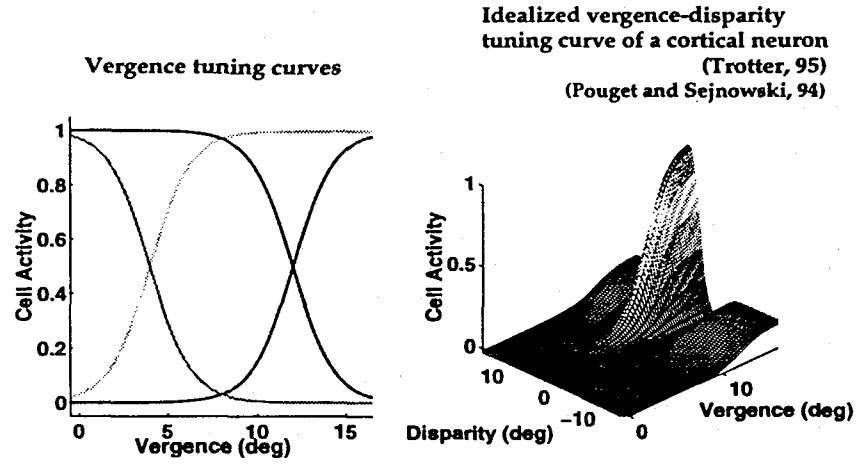

Figure 7: Idealized vergence-disparity tuning curve of a cortical neuron. The left diagram shows the four vergence tuning curves used for the cells in the model. On the right, the idealized vergence-disparity tuning curve is plotted by combining the disparity tuning curve of a cell tuned at 3 deg (fig. 6) with the tuning curve of a cell tuned at 16 deg vergence (left diagram). There are $17 \times 4 = 68$ cells with different vergence-disparity tuning curves in the model.

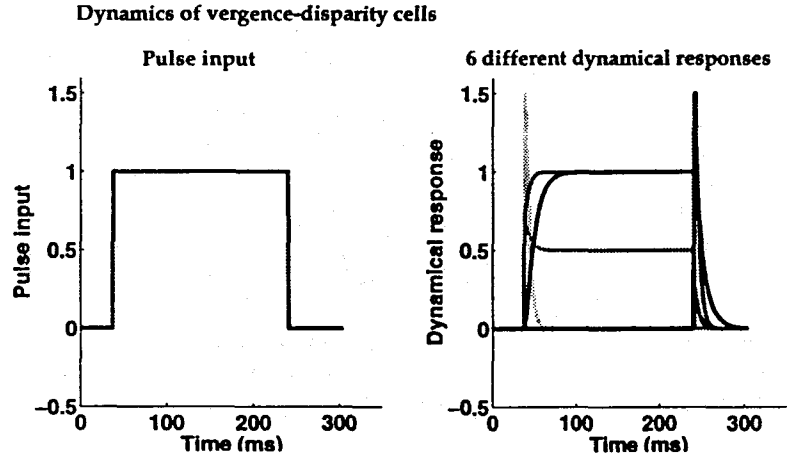

Figure 8: Dynamics used for the vergence-disparity cells. A pulse input is shown on the left. On the right the responses to the pulse are shown. There is one ON-phasic response, two tonic responses with different time constants, two OFF-phasic responses of different amplitudes, and one phasic-tonic response.

a continuous range of disparity (Stevenson et al., 1992). Pouget and Sejnowski (1994) used a continuum of idealized disparity selective cells in their model with disparity ranging from -4 to 4 deg. For theoretical reasons, and noting that vergence eye movement can be performed at least over a range from 0 to 16 deg vergence angle, we decided to extend the range of selective disparity cells in the model to -8 to 8 deg (fig. 6).

Trotter (1992) showed that most cells in V1 tuned for horizontal disparity are also tuned broadly for vergence. Fig. 7 shows the vergence and vergence-disparity tuning curve of a typical cell used in the model (Pouget and Sejnowski, 1994).

Since cortical cells have time-varying responses (DeAngelis, Ohzawa and Freeman, 1995), we consid-

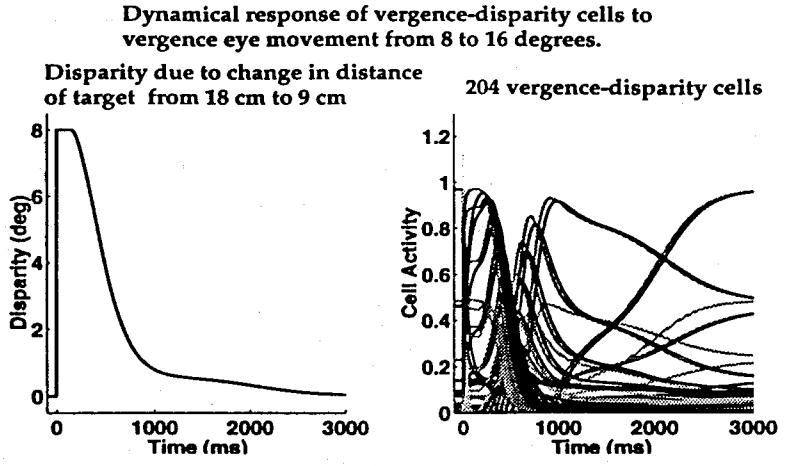

Figure 9: Dynamical response of vergence-disparity cells to vergence eye movement used as input to the cerebellum. The disparity change due to the movement of a target from 18 cm to 9 cm followed by a convergence eye movement is shown on the left. The vergence angle varies from 8 to 16 deg during the movement. On the right the responses to the disparity change and vergence eye movement of 204 vergence-disparity cells are shown.

ered six different dynamical responses (fig. 8) whose Laplace transforms were as follow:

$$\frac{1}{50s+1} \qquad \frac{15s+1}{20s+1} \qquad \Theta(\pm\frac{20s+0.5}{30s+1}) \qquad \Theta(\pm\frac{30s}{20s+1}) \tag{7}$$

where $\Theta(x)$ is a zero threshold function ($\Theta(x) = x$ for $x > 0$, and 0 otherwise).

In fig. 9 we show the time courses of the responses during a vergence eye movement for all the input V1 vergence-disparity cells used in the model. The distance of a visual target was changed from 18 cm to 9 cm at time zero. The sudden disparity change causes a vergence eye movement to begin after some time delay. The responses of the cells varied according to their dynamics, their tuning curve to disparity and their tuning curve to vergence. The disparity and vergence inputs to the cells are as shown in fig. 5. The value of vergence angle during the eye movement is used to modulate the visual cell responses. This requires either an efference copy or extraocular muscle proprioception source of vergence to modulate cell responses in V1. Evidence shows that the latter source is implicated in normal development of binocular vision and therefore is a likely candidate (Trotter, 1995).

In total there were 408 vergence-disparity cells (17 disparity tuning curves $\times$ 4 vergence tuning curves $\times$ 6 dynamical responses). To increase simulation speed we only included cells that changed their responses during the vergence eye movement. Since in this example the eye movement was from 8 to 16 deg vergence, only two vergence tuning curves (the two curves having 0.5 activity at 12.5 deg vergence in fig. 7) were used for a total of 204 input cells.

4.5 Vergence prediction

In this section we present how the responses of the vergence-distance cells that we constructed in the previous section may be used by a cerebellar model to learn the prediction of vergence angle during a vergence eye movement.

To study the feasibility of constructing the prediction of vergence angle, we chose the simplest network possible. The network had only one cell in the deep cerebellar nuclei, and only one Purkinje cell with dynamics as shown in fig. 10.

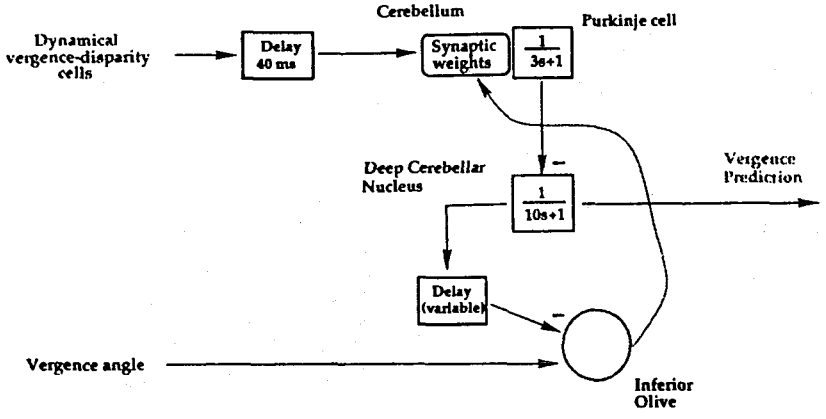

Figure 10: Cerebellar neural network model used in the prediction of vergence eye movements. The adaptation of the synaptic weights follows the LMS algorithm. The delay of 40 ms represents the time taken for the visual system to process the information and to relay the information to the brainstem. The delay in the projection from the deep nuclei was set to either 60 ms, 100 ms or 225 ms to obtain predictions with the different time advance. The activity of vergence disparity cells $x_i(t)$ was also delayed by 60 ms, 100 ms or 225 ms to calculate the correlation with prediction error for the synaptic weight updates (Eq. 4).

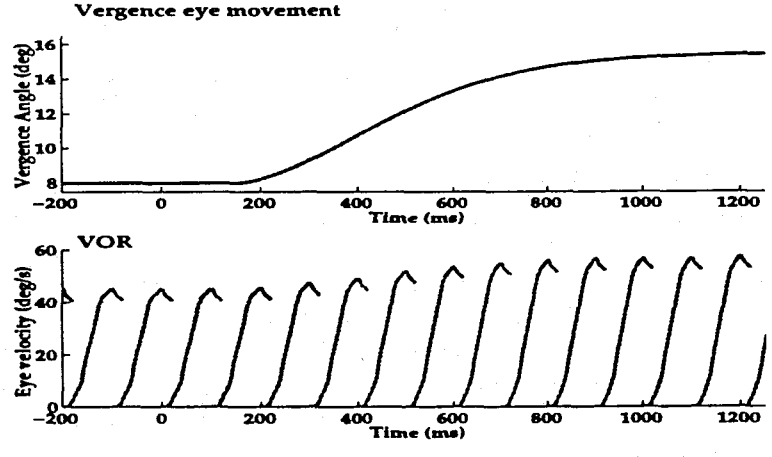

Figure 11: Simulated eye velocity responses of the VOR during a vergence eye movement from 8 degrees to 16 degrees vergence. Upper trace shows the vergence eye movement from a target located at 8 deg vergence angle to a near target at 16 deg vergence. The dynamical model of accommodation and convergence was used to produce the vergence eye movement. The lower traces show eye velocity responses evoked by simulated brief head rotation at different times during the vergence eye movement. A dynamical model of VOR modulation was used to simulate the eye velocities (Coenen and Sejnowski, 1996). The dynamical VOR model was modified to use the vergence angle prediction of the present model (fig. 10) instead of the regular (non-predictive) vergence input normally used. For this figure, the lead time of vergence prediction was 60 ms. The increase of the VOR gain with increase in vergence angle is clearly apparent. (Compare with fig. 3).

The dynamical vergence-disparity cell responses (fig. 9) were delayed by 40 ms and projected to the Purkinje cell through a set of adaptive synaptic weights. This delay represents the time taken for the visual system to process the visual information and to relay it to the brainstem. The deep nuclei neuron only received Purkinje cell input. The synaptic weights were modified according to the LMS rule in Eq. 4, and

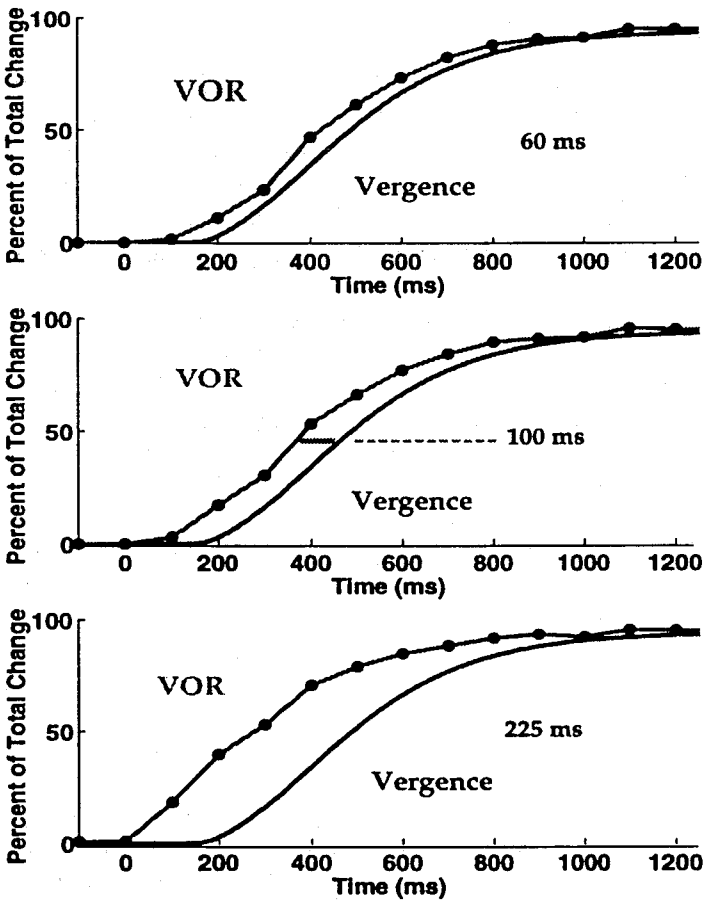

Figure 12: Anticipatory VOR gain modulation with vergence for three different time delays (60, 100, 225 ms) used in the model during learning. Percentage of total change of the VOR gain and vergence angle occurring during an 8 deg (8 to 16 deg) vergence eye movement. At time zero, a simulated visual target moved abruptly from 18 cm to 9 cm (see fig.9). The solid lines plot the normalized vergence angle as a function of time. The linked points are the normalized peak eye velocities obtained from brief head rotation at different times during vergence eye movement. The values of the peak velocities from VOR eye responses as shown in fig. 11 were used in this plot (the eye responses are different for the three delays). Note that the VOR modulation in the bottom plot does not start at zero, in reality it starts with a time delay of about 50 ms. This is a consequence of the sampling interval used in fig. 11. Compare these results with the experimental observations in fig. 4.

updated every 5 ms, at $t_o = \{5, 10, 15, ...\}$ms. In order to compare the model with experimental data from different monkeys which showed different lead times in the anticipation of the VOR with vergence (fig. 4), we used three different delays for the nucleoolivary feedback pathway: 60, 100 and 225 ms. The inputs x_i used in Eq. 4 were also delayed by either 60, 100 and 225 ms before being correlated with the prediction error. The learning rate μ used in the simulations was 0.00006, which was small enough to insure convergence (Eq. 6). Learning was stopped when changes of the vergence prediction during a vergence eye movement were less than 1%. It took fewer than 800 vergence eye movements to learn the prediction for each time delay.

The vergence angle input to the inferior olive could either be a projection of an efferent copy or extraocular muscle proprioception of vergence angle. Neurons firing in correlation with vergence angle and vergence velocity have been observed in the mesencephalic reticular formation (Mays, 1984; Mays et al., 1986). The maximum lead time observed for convergence burst cells (vergence velocity) was 42 ms. They are therefore not likely candidates for causing predictive VOR changes which occur with 50 and up to 200 ms lead time. For this reason, they have not been included in the inputs used to construct the vergence prediction, but they may provide the vergence efferent copy to the inferior olive required in the model. Cells firing linearly with vergence angle have also been observed in the nucleus reticularis tegmenti pontis (NRTP), but unfortunately the lead times were not determined precisely (Gamlin and Clarke, 1995). The NRTP has reciprocal connections with the cerebellum and receives input from a number of cortical and subcortical areas. In the cerebellum, the NRTP projects to the flocculus, oculomotor vermis, paravermis, and to the fastigial and interpositus nuclei. The model may therefore represent one of these areas.

Our previous dynamical model of VOR modulation was used to simulate the eye velocities (Coenen and Sejnowski, 1996). The dynamical VOR model was modified to use the vergence angle prediction of the present model (fig. 10) instead of the regular (non-predictive) vergence input normally used. The result is shown in fig. 11 (compare to fig. 3). The values of the peak velocities from VOR eye responses as shown in fig. 11 were used in the plots of fig. 12. The eye responses were different for the three delays; the delay used in fig. 11 was 60 ms. Compare these results with the experimental observations in fig. 4.

5 Conclusion

We have shown that a dynamical cerebellar model that includes realistic visual disparity-vergence cell inputs can learn to anticipate the vergence eye position up to 225 ms in advance. In the model, this predictive signal modulates the VOR through the cerebellum in anticipation of final vergence eye position. The cerebellum in the model is also responsible for combining nonlinearly sensory inputs with eye position signals to modulate the VOR for different target positions and rotation axes. We have proposed that this nonlinear interaction is learned in the cerebellum.

The model is incomplete in many respects and an expanded model that learns the VOR modulation directly from feedback error will be the focus of future research. In particular, the important role of the coupling between cells in the inferior olive, and the effects of synchronous firing of these cells on the cerebellum, will be integrated into the model.

6 Acknowledgments

The first author was supported by a McDonnell-Pew Graduate Fellowship during this research. This research was supported by the Howard Hughes Medical Institute.

References

Akshoomoff, N. A. and Courchesne, E. (1992). A new role for the cerebellum in cognitive operations. *Behavioral Neuroscience*, 106(5):731–738.

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

Andreasen, N. C., O'Leary, D. S., Arndt, S., Cizadlo, T., Hurtig, R., Rezai, K., Watkins, G. L., Ponto, L. L. B., and Hichwa, R. D. (1995). Short-term and long-term verbal memory: A positron emission tomography study. *Proc. Natl. Acad. Sci.*, 92:5111–5115.

Arbib, M., Schweighofer, N., and Tach, W. (1994). Modeling the role of cerebellum in prism adaptation. In Cliff, D., Husbands, P., Meyer, J.-A., and Wilson, S., editors, *From Animals to Animats 3. Proceedings of the Third International Conference on Simulation of Adaptive Behavior*, pages 36–44, Cambridge, MA, USA. MIT Press.

Barnes, G. and Asselman, P. (1991). The mechanism of prediction in human smooth pursuit eye movements. *Journal of Physiology*, 439:439–461.

Barto, A. G., Sutton, R. S., and Anderson, C. W. (1993). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13:834–846.

Bloedel, J. (1992). Functional heterogeneity with structural homogeneity: How does the cerebellum operate? *Behavioural and Brain Science*, 15:666–78.

Chapeau-Blondeau, F. and Chauvet, G. (1991). A neural network model of the cerebellar cortex performing dynamic associations. *Biological Cybernetics*, 65:267–279.

Coenen, O. J.M. D., Sejnowski, T. J., and Lisberger, S. (1993). Biologically plausible local learning rule for the adaptation of the vestibulo-ocular reflex. In Giles, C., Hanson, S., and Cowan, J., editors, *Advances in Neural Information Processing Systems 5*, San Mateo, CA. Morgan Kaufmann Publishers.

Coenen, O. J.M. D. and Sejnowski, T. J. (1996). A dynamical model of context dependencies for the vestibulo-ocular reflex. In Touretzky, D., Mozer, M., and Hasselmo, M., editors, *Advances in Neural Information Processing Systems 8*, Cambridge, MA. MIT Press.

Corbetta, M., Miezin, F. M., Shulman, G. L., and Petersen, S. E. (1993). A PET study of visuospatial attention. *Journal of Neuroscience*, 13(3):1202–1226.

Cumming, B. and Judge, S. (1986). Disparity-induced and blur-induced convergence eye movement and accommodation in the monkey. *Journal of Neurophysiology*, 55:896–914.

Darlot, C. (1993). The cerebellum as a predictor of neural messages -I. The stable estimator hypothesis. *Neuroscience*, 56:617–646.

De Zeeuw, C. and Berrebi, A. (1995). Postsynaptic targets of purkinje cell terminals in the cerebellar and vestibular nuclei of the rat. *European Journal of Neuroscience*, 7(11):2322–33.

De Zeeuw, C. I., Hostege, J. C., Ruigrok, T. J., and Voogd, J. (1989). Ultrastructural study of the GABAergic, cerebellar, and mesodiencephalic innervation of the cat medial accessory olive: Anterograde tracing combined with immunocytochemistry. *The Journal of Comparative Neurology*, 284:12–35.

DeAngelis, G., Ohzawa, I., and Freeman, R. (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neurosciences*, 18:451–8.

du Lac, S., Raymond, J. L., Sejnowski, T. J., and Lisberger, S. G. (1995). Learning and memory in the vestibulo-ocular reflex. *Annual Review of Neuroscience*, 18:409–41.

Dugas, C. and Smith, A. M. (1992). Responses of cerebellar Purkinje cells to slip of a hand-held object. *Journal of Neuroscience*, 67:483–495.

Flament, D., Ellermann, J., Kim, S.-G., Ugurbil, K., and Ebner, T. (1995). Functional magnetic resonance imaging (fmri) of cerebellar activation while learning to correct for visuomotor errors. In *25th Annual Meeting Society for Neuroscience Abstracts*. Society for Neuroscience.

Flament, D., Ellermann, J., Kim, S.-G., Ugurbil, K., and Ebner, T. (1996). Temporal changes in cerebellar activation during motor errors and the learning of a visuomotor dissociation task, revealed using high-field functional magnetic resonance imaging. To appear.

Fredette, B. J. and Mugnaini, E. (1991). The GABAergic cerebello-olivary projection in the rat. *Anat. Embryol.*, 184:225–243.

Friston, K., Frith, C., Passingham, R., Liddle, P., and Frackowiak, R. (1992). Motor practice and neurophysiological adaptation in the cerebellum: a positron tomography study. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 248(1323):223–8.

Fujita, M. (1982). Adaptive filter model of the cerebellum. *Biological Cybernetics*, 45:195–206.

Gamlin, P. D. R. and Clarke, R. J. (1995). Single-unit activity in the primate nucleus reticularis tegmenti pontis related to vergence and ocular accomodation. *Journal of Neurophysiology*, 73(5):2115–2119.

Gluck, M. A. and Thompson, R. F. (1990). Adaptive signal processing and the cerebellum: Models of classical conditioning and VOR adaptation. In Gluck, M. A. and Rumelhart, D. E., editors, *Neuroscience and Connectionist Theory*, chapter 4, pages 131–185. Lawrence Erlbaum Associates.

Goodkin, H., Keating, J., Martin, T., and Thach, W. (1993). Preserved simple and impaired compound movement after infraction in the territory of the superior cerebellar artery. *The Canadian Journal of Neurological Sciences*, 20(Suppl. 3):S93–S104.

Harwerth, R., Smith 3rd, E., and Siderov, J. (1995). Behavioral studies of local stereopsis and disparity vergence in monkeys. *Vision Research*, 35(12):1755–70.

Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. Macmillan.

Ito, M. (1984). *The Cerebellum and Neural Control*. Raven Press, New York.

Ivry, R. B. and Keele, S. W. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience*, 1(2):136–152.

Kabal, P. (1990). The stability of adaptive minimum mean square error equalizers using delayed adjustments. *In IEEE Trans. Comm.*, Com-31(3):430–32.

Kawato, M. and Gomi, H. (1992). The cerebellum and VOR/OKR learning models. *Trends in Neuroscience*, 15:445–453.

Keeler, J. D. (1990). A dynamical system view of cerebellar function. *Physica D*, 42:396–410.

Kettner, R., Mahamud, S., Leung, H.-C., Barto, A., Houk, J., and Peterson, B. (1995). Predictive smooth pursuit eye movements are generated by a cerebellar model. In *25th Annual Meeting Society for Neuroscience Abstracts*. Society for Neuroscience.

Kim, S.-G., Uğurbil, and Strick, P. (1994). Activation of a cerebellar output nucleus during cognitive processing. *Science*, 265:949–951.

Klopf, A. H. (1988). A neuronal model of classical conditioning. *Psychobiology*, 16(2):85–125.

Lampl, I. and Yarom, Y. (1993). Subthreshold oscillations of the membrane potential: A functional synchronizing and timing device. *Journal of Neurophysiology*, 70:2181–2186.

Leiner, H., Leiner, A., and Dow, R. (1989). Reappraising the cerebellum: what does the hindbrain contribute to the forebrain? *Behavioral Neuroscience*, 103(5):998–1008.

Leiner, H., Leiner, A., and Dow, R. (1993). Cognitive and language functions of the human cerebellum. *Trends in Neurosciences*, 16(11):444–447. Debate over the evidence for non-motor function of the cerebellum.

Leung, H.-C. and Kettner, R. (1995). Predictive control of smooth pursuit eye movements along complex two-dimensional trajectories in monkey. In *25th Annual Meeting Society for Neuroscience Abstracts*. Society for Neuroscience.

Lisberger, S. G., Pavelko, T. A., Bronte-Stewart, H. M., and Stone, L. S. (1994). Neural basis for motor learning in the vestibuloocular reflex of primates. II. Changes in the responses of horinzontal gaze velocity Purkinje cells in the cerebellar flocculus and ventral paraflocculus. *Journal of Neurophysiology*, 72(2):954–973.

Llinas, R. and Welsh, J. P. (1993). On the cerebellum and motor learning. *Current Opinion in Neurobiology*, 3:958–965.

Llinas, R. and Yarom, Y. (1981). Electrophysiology of mammalian inferior olivary neurons in vitro. *Journal of Physiology*, 315:549–567.

Llinas, R. and Yarom, Y. (1986). Oscillatory properties of guinea-pig inferior olivary neurones and their pharmacological modulation: an *in vitro* study. *Journal of Physiology*, 376:163–182.

Long, G., Ling, F., and Proakis, J. (1989). The LMS algorithm with delayed coefficient adaptation. *IEEE Trans. Acoust. Speech & Signal Proc.*, 27(9):1397–1459.

Long, G., Ling, F., and Proakis, J. (1992). Corrections to 'the LMS algorithm with delayed coefficient adaptation'. *IEEE Transactions on Signal Processing*, 40(1):230–2.

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

Mays, L. E. (1984). Neural control of vergence eye movements: Convergence and divergence neurons in midbrain. *Journal of Neurophysiology*, 51:1091–1108.

Mays, L. E., Porter, J. D., Gamlin, P., and Tello, C. A. (1986). Neural control of vergence eye movements: Neurons encoding vergence velocity. *Journal of Neurophysiology*, 56:1007–1021.

Miall, R. C., Weir, D. J., Wolpert, D. M., and Stein, J. F. (1993). Is the cerebellum a smith predictor ? *Journal of Motor Behavior*, 25(3):203–216.

Middleton, F. A. and Strick, P. L. (1994). Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science*, 266:458–461.

Molchan, S. E., Sunderland, T., McIntosh, A. R., Herscovitch, P., and Schreurs, B. G. (1994). A functional anatomical study of associative learning in humans. *Proc. Natl. Acad. Sci.*, 91:8122–8126.

Nieuwenhuys, R., Voogd, J., and van Huijzen, C. (1988). *The human central nervous system*. Springer-Verlag.

Palay, S. L. and Chan-Palay, V. (1974). *Cerebellar Cortex, Cytology and Organization*. Springer-Verlag.

Paulin, M. G. (1989). A Kalman filter theory of the cerebellum. In Arbib, M. A. and Amari, S.-i., editors, *Dynamic Interactions in Neural Networks: Models and Data*, chapter III, pages 239–259. Springer-Verlag, New York. Research Notes in Neural Computing, Vol. 1.

Perrett, S. P. and Mauk, M. D. (1995). Extinction of conditioned eyelid responses requires the anterior lobe of cerebellar cortex. *Journal of Neuroscience*, 15:2074–2080.

Perrett, S. P. and Ruiz, B. P. (1993). Cerebellar cortex lesions disrupt learning-dependent timing of conditioned eyelid responses. *Journal of Neuroscience*, 13:1708–1718.

Poggio, G., Motter, B., Squatrito, S., and Trotter, Y. (1985). Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Research*, 25:397–406.

Pouget, A. and Sejnowski, T. J. (1994). A neural model of the cortical representation of egocentric distance. *Cerebral Cortex*, 4:314–329.

Raichle, M. E., Fiez, J. A., Videen, T. O., Macleod, A.-M. K., Pardo, J. V., Fox, P. T., and Petersen, S. E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex*, 4:8–26.

Sadato, N., Campbell, G., Ibáñez, V., Deiber, M.-P., and Hallett, M. Complexity affects regional cerebral blood flow change during sequential finger movements. Journal of Neuroscience (To appear).

Schor, C. M. (1992). A dynamical model of cross-coupling between accomodation and convergence: Simulations of step and frequency responses. *Optometry and Vision Science*, 69(4):258–269.

Seitz, R., Canavan, A., Yaguez, L., Herzog, H., Tellmann, L., Knorr, U., Huang, Y., and Homberg, V. (1994). Successive roles of the cerebellum and premotor cortices in trajectorial learning. *Neuroreport*, 5(18):2541–4.

Singh, S. and Sutton, R. (1996). Reinforcement learning with replacing eligibility traces. *Machine Learning*, 22:123–58.

Snyder, L. H. and King, W. M. (1992). Effect of viewing distance and location of the axis of head rotation on the monkey's vestibuloocular reflex I. eye movement response. *Journal of Neurophysiology*, 67(4):861–874.

Snyder, L. H., Lawrence, D. M., and King, W. M. (1992). Changes in the vestibulo-ocular reflex (VOR) anticipate changes in vergence angle in monkey. *Vision Research*, 32(3):569–575.

Stevenson, S., Cormack, L., Schor, C., and Tyler, C. (1992). Disparity tuning in mechanisms of human stereopsis. *Vision Research*, 32(9):1685–94.

Stone, L. S. and Lisberger, S. (1990). Visual response of Purkinje cells in the cerebellar flocculus during smooth-pursuit eye movements in monkeys. I. simple spikes. *Journal of Neurophysiology*, 63:1241–1261.

Sutton, R. (1992). Adapting bias by gradient descent: an incremental version of delta-bar-delta. In *AAAI-92. Proceedings Tenth National Conference on Artificial Intelligence*, pages 171–6, San Jose, CA, USA. AAAI Press.

Tachibana, H. (1995). Event-related potentials in patients with cerebellar degeneration: electrophysiological evidence for cognitive impairment. *Cognitive Brain Research*, 2:173–180.

Thach, W. T., Goodkin, H. P., and Keating, J. G. (1992). The cerebellum and the adaptive coordination of movement. *Annual Review of Neuroscience*, 15:403–442.

Trotter, Y. (1995). Cortical representation of visual three-dimensional space. *Perception*, 24:287–298.

Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S., and Imbert, M. (1992). Modulation of neural stereoscopic processing in primate area V1 by the viewing distance. *Science*, 257:1279–81.

van Donkelaar, P. and Lee, R. G. (1994). Interactions between the eye and hand motor systems: Disruptions due to cerebellar dysfunction. *Journal of Neurophysiology*, 72:1674–1685.

Voogd, J. and Bigaré, F. (1980). Topographical distribution of olivary and corticonuclear fibers in the cerebellum. a review. In Courville, J., de Montigny, C., and Lamarre, Y., editors, *The Inferior Olivary Nucleus*, pages 207–235. Raven Press, New York.