Parallel Fiber Coding in the Cerebellum for Life-Long Learning

OLIVIER J.-M.D. COENEN* AND MICHAEL P. ARNOLD
*Computational Neurobiology Laboratory, Howard Hughes Medical Institute, Salk Institute,
La Jolla, CA 92037, USA*
mikea@salk.edu

TERRENCE J. SEJNOWSKI
*Computational Neurobiology Laboratory, Howard Hughes Medical Institute, Salk Institute, La Jolla,
CA 92037, USA and Department of Biology, UCSD, La Jolla, CA 92093, USA*
terry@salk.edu

MARWAN A. JABRI
*Electrical and Computer Engineering, Oregon Graduate Institute, 20000 N.W. Walker Road,
Beaverton, OR, 97006-8921, USA*
marwan@ece.ogi.edu

Abstract. Continuous and real-time learning is a difficult problem in robotics. To learn efficiently, it is important to recognize the current situation and learn appropriately for that context. To be effective, this requires the integration of a large number of sensorimotor and cognitive signals. So far, few principles on how to perform this integration have been proposed. Another limitation is the difficulty to include the complete contextual information to avoid destructive interference while learning different tasks.

We suggest that a vertebrate brain structure important for sensorimotor coordination, the cerebellum, may provide answers to these difficult problems. We investigate how learning in the input layer of the cerebellum may successfully encode contextual knowledge in a representation useful for coordination and life-long learning. We propose that a sparsely-distributed and statistically-independent representation provides a valid criterion for the self-organizing classification and integration of context signals. A biologically motivated unsupervised learning algorithm that approximate such a representation is derived from maximum likelihood. This representation is beneficial for learning in the cerebellum by simplifying the credit assignment problem between what must be learned and the relevant signals in the current context for learning it. Due to its statistical independence, this representation is also beneficial for life-long learning by reducing the destructive interference across tasks, while retaining the ability to generalize. The benefits of the learning algorithm are investigated in a spiking model that learns to generate predictive smooth pursuit eye movements to follow target trajectories.

Keywords: cerebellum, learning, plasticity, Bayesian, maximum likelihood, information theory, sparse, statistically independent, independent component analysis, ICA, coding, code, model, spiking, granule cell, mossy fiber, Golgi cell, glomerulus, parallel fiber, granular layer, robotics, smooth pursuit, saccade, motor control, destructive interference

*Current affiliation: Sony Computer Science Laboratory, 6, rue Amyot, 75005 Paris, France. E-mail: coenen@csl.sony.fr

1. Introduction

Continuous and real-time learning is a difficult problem in robotics. One reason is that general criteria and techniques for the integration and classification of large number of continuously changing sensory, motor and other higher *cognitive* signals have been difficult to produce. Another reason is that destructive interference across tasks occurs when using function approximators and neural networks that lack specialized representations for the context of the situation in which learning takes place. One solution has been to approximate learning of nonlinear functions with spatially localized linear models in order to minimize interference (Atkeson et al., 1997). Interestingly, the cerebellar architecture and its functionality have points in common with such models.

Movements in robotics often follow *desired* stereotyped trajectories whereas motor control in humans and animals is often tailored to the prevailing motor context (Blakemore et al., 1998). The ability to generate appropriate motor behavior remains therefore contingent upon the ability to recognize the current context and to adjust motor commands accordingly. Evidence suggest that the cerebellum is a key participant in this process (Blakemore et al., 1999).

The cerebellum has been known to be important for sensorimotor coordination in vertebrates for quite some time, yet its specific function has remained elusive. The cerebellum has afferent and efferent projections carrying a large variety of signals including sensory, motor and higher cognitive signals from most major areas of the brain.

The cerebellum (Cb) has been suggested to act as a short-term predictive engine in the brain (Miall et al., 1993) whose processing may require three distinct functional stages: 1) the transformation and combination of cerebellar inputs in a pre-processed form appropriate for predictions (granular layer, Fig. 1), 2) the identification and selection of the pre-processed inputs that define the context of the prediction and that can anticipate the neural activity to predict (molecular and Purkinje layer) and 3) the construction of the predictions themselves at the Purkinje and deep cerebellar nuclei cells (Coenen, 1998). The computational strength of the cerebellum may be to produce predictions or predictive neural commands dependent on precise contextual information.

Most of the modeling and theoretical work in the literature on learning and adaptation in the cerebellum has
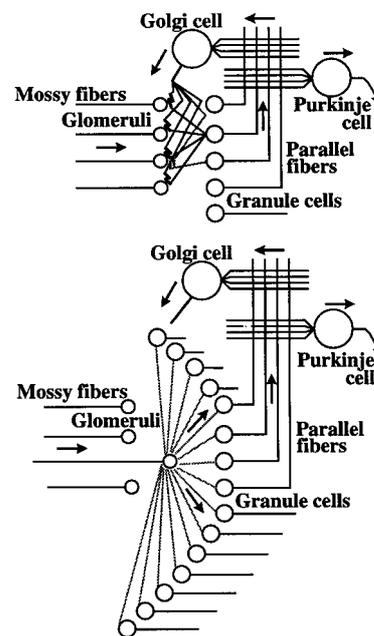

Figure 1.   Granular layer of the cerebellum considered in the model. The mossy fibers (Mfs) end in glomeruli (Gl) which make contact with granule cells (Gcs) dendrites. The granule cells send their parallel fiber axons to the Purkinje cells and also make contact with a Golgi cell (Go) which inhibits every granule cell dendrites at the glomeruli. Golgi cells also receive a number of inputs from mossy fiber glomeruli directly (not shown). A granule cell (Gc) receives 4–7 mossy fiber (Mf) inputs (top) whereas a mossy fiber glomerulus contacts 20–50 granule cells (bottom). The Gl-Gc and Go-Gc synapses are represented respectively by matrices **W** and **V** of identical dimensions, where v_{ji} is the weight of the inhibitory Go-Gc synapse above the excitatory Gl_i-Gc_j synapse with weight w_{ji}. Note that the Purkinje cells, output of the cerebellar cortex, are not part of the granular layer, but are shown to complete the cerebellar network. The arrows indicate the propagation of activity in the network.

focused on the output cells of the cerebellar cortex, the Purkinje cells. We investigate here the computational advantages of learning at the input, the granule cells, of the cerebellum. These cells are the most numerous in the brain and we suggest that they constitute a vast repository of knowledge for learning appropriately for the prevailing context.

Recently, probabilistic and information theoretic approaches to learning in neural networks have made new interesting connections with neural functions (Olshausen and Field, 1996). However, most of the results have been obtained for early sensory systems and have been difficult to apply to higher cortical levels. Here these principles are applied to the cerebellum and are shown to be useful for sensorimotor and
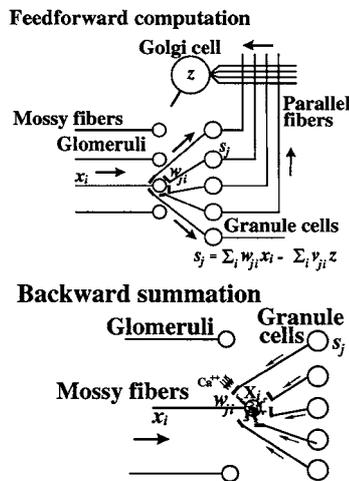
Figure 2. Feedforward computation in the model for computing cell responses and backward summation used for directing plasticity of the synaptic weights. The vectors x, s and $z \geqslant 0$ represent the firing rate activity of the mossy fibers, granule cells and Golgi cell respectively (top). The backward sum X_i directs learning at the glomerulus i (bottom) and is computed using the activity of granule cells propagated back along their dendrites (granule cells are electrotonically compact).

cognitive signals integration and context multiplexing for control.

2. Methods and Results

The role of the granular layer (Fig. 1), which consists of mossy fiber (Mf) glomeruli (Gl), granule (Gc) and Golgi (Go) cells, is two fold. One is to transmit to the Purkinje cells through the parallel fibers a complete contextual account of mossy fibers (Mfs) activity, and the other is to provide it in a form which facilitates adaptation in the inhibitory interneurons and Purkinje cells of the cerebellar cortex. We maintain that a sparse and distributed representation in the parallel fibers that maximizes the mutual information between the mossy fibers (Mfs) and the parallel fibers and minimizes the statistical dependencies among parallel fibers fulfills these two roles.

A Difficult Credit Assignment Problem The Purkinje cells, which receive on the order of 100 000 parallel fibers, axons of the granule cells (Gcs), face a difficult credit assignment problem in identifying which parallel fiber synapse must be modified in connection with climbing fibers activity, which direct learning at the Purkinje cells (Marr, 1969). Cerebellum models often

solve this problem by using thresholding together with a binary code to limit the number of active parallel fibers (Kettner et al., 1997). These models will often have poorer generalization abilities than an analog code would.

Our hypothesis is that powerful unsupervised algorithms direct learning at the granular layer resulting in a very efficient representation of cerebellar inputs by the granule cells. Learning there is suggested to lead to a representation that is sparse, distributed and that minimizes the statistical dependencies among granule cells. A sparse and distributed code tends to minimize the time during which the cells are active, and a statistically-independent representation minimizes the redundancy across active cells. Both properties reduce the complexity of the credit assignment between active parallel fibers and climbing fiber at a Purkinje cell more effectively than previous learning rules that only produced a decorrelated representation at the parallel fibers (Jonker et al., 1998). Therefore, the sparse and statistically-independent (SSI) representation would facilitate learning in the next layers of the cerebellum, including at the Purkinje cells.

This representation would also minimize the destructive interference that is often seen in networks when learning multiple tasks. The algorithm seeks to encode independent contexts using sparse representations that are as stastically independent as possible from each other. With this representation, the different contexts are multiplexed onto the Purkinje cells that are involved in the control of a specific actuator (e.g., muscle, motor). This representation therefore allows the independent control of actuators in different and independent contexts, while generalizing across dependent or overlapping contexts.

2.1. The Granular Layer: Anatomy

Mossy fiber inputs to the cerebellum terminate in glomeruli where granule cell dendrites and Golgi cell (Go) axons converge to make synaptic contacts (Fig. 1). A glomerulus contacts about 20 to 50 granule cells through excitatory receptors and granule cells receive a combination of 4 to 7 mossy fiber inputs (Jakab and Hamori, 1988). The large number of granule cells in the cerebellum may be related to the large number of possible mossy fiber input combinations (Marr, 1969). A Golgi cell integrates the output of approximately 1000–6000 granule cells and receive inputs from a number of mossy fibers.

2.2. The Granular Layer: Computation

The activity of granule cell is the result of two nested projections of the multidimensional mossy fiber space (Coenen et al., 1999, 2000). The first projection is defined by the connectivity of the granule cell, which receives 4 to 7 mossy fiber inputs, whereas the second projection is defined by the synaptic weights of these inputs. The first projection is assumed fixed after development, but the second projection is plastic, as recent evidence of long-term potentiation (LTP)—or increase in activity—at the mossy fiber-granule cell synapses suggests (D'Angelo et al., 1999). The input weights define a direction in the mossy fibers input space along which the granule cell has highest sensitivity. What should determine this direction, hence the weight values? We argue that the direction that gives the granule cell the sparsest probability density function (i.e., one with a high kurtosis) and that makes all neighboring cells as independent as possible, is the direction that leads to a representation with the most interesting and useful properties for subsequent computations in the cerebellum (Coenen et al., 1999).

It has been estimated that it only takes about 50 granule cells firing together to activate a Purkinje cell (Barbour, 1993). Since 50 active inputs out of 200 000 possible inputs to a Purkinje cell is sparse, this suggest that a sparse representation would still activate a Purkinje cell effectively. It also is well known that information storage or capacity is higher for sparse network (Meunier and Nadal, 1995). Therefore, with a sparse code at the parallel fibers, a Purkinje cell can learn the finely tuned control of an actuator in many more different, although slightly, situations.

This approach provides an effective criterion for sensorimotor and other signals integration. In this representation, a granule cell combines signals regardless of their origin such as to have an activity that is both sparse—i.e. inactive most of time—, and as independent from its neighbors as possible. The independence is actually a consequence of the sparseness criterion and ensures that the greatest amount of information from mossy fiber inputs are encoded with the smallest number of active granule cells. This representation is very much dependent on the activities, and their frequencies, experienced by the organism. Different activities or tasks will create different dependencies among sensorimotor and cognitive signals at different times. These dependencies will appear as higher probability densities in the mossy fiber input space. A granule cell will encode these dependencies and their directions in mossy fiber space by modifying its synaptic weights. Since this approach is probabilistic, the most common or most *probable* activities will have the greatest influence upon the granule cell activities. This may explain why *practice makes perfect*.

2.3. The Granular Layer: Model

A linear relationship between the activity of the mossy fiber inputs $\mathbf{x}$ and the granule cells activity $\mathbf{s}$ is assumed. Furthermore, we assumed that a Golgi cell contacts every granule cell dendrites at every glomerulus in its arbor so that the Golgi cell-granule cell synapses can be represented by a weight matrix V. The number of glomerulus-granule cell and Golgi cell-granule cell synapses at a glomerulus are therefore equal, with respective matrices W and V of identical dimensions (top panel in Fig. 1). The Golgi cell is modeled as having a subtracting effect on the mossy fiber input x so that the granule cell activity is modeled as $s = Wx - V\underline{1}z \geqslant 0$, where z is the Golgi cell activity, and $\underline{1}$ is a column vector of 1's. We also write $\mathbf{z} = \underline{1}z$, to represent a column vector of identical z values. The granule cell activities are therefore the responses of the projection of x into s performed by the weights W and activity Vz (Fig. 2).

Role of Golgi Cells   The mossy fiber afferents have the form $x = \hat{x} + x_o$ where x_o is the mean, always positive, of x and where $\hat{x}$ has zero mean. For the granule cell probability distribution to remain sparse, i.e. centered at zero, the bias term w_0 in the granule cell activity $s = Wx + w_0$ must be $w_0 = -Wx_0$ and negative. The bias w_o is approximated by $E[-Vz]$, where $E[\cdot]$ is the expectation over time, giving to the Golgi cell the role of setting the threshold of granule cells so that their activities $\mathbf{s}$ remain sparsely distributed at $\mathbf{0}$.

2.3.1. Probability Density of Granule Cells.   The granule cell activities are modeled as $s = Wx - Vz$ with $s, x, \mathbf{z} \geqslant \mathbf{0}$ (D'Angelo et al., 1995) with a sparse (high kurtosis) and statistically independent prior probability density $\mathbf{f_s}(\mathbf{s}) = \prod_i f_s(s_i)$ where $f_s(s_i)$ was chosen to be the same exponential density for all granule cells, $f_s(s_i) = \alpha \exp(-\alpha s_i) = \gamma_1(s_i)$, and where $\gamma_1(\cdot)$ is the gamma density of order 1.[1] x, s and $\mathbf{z}$ represent the firing rate of the mossy fiber, granule cell and Golgi cell respectively. Due to the positive constraint on the granule cells activity, the prior is more precisely $f_s(s_i) = \alpha e^{-\alpha s_i} U(s_i)$, where $U(\cdot)$ is

the step function. To simplify the derivation of the learning rules, this prior was approximated by two exponential priors, $f_s(s_i) = \alpha \exp(-\alpha s_i)$ for $s_i > 0$ and $f_s(s_i) = \alpha \exp(\beta s_i)$ for $s_i \leq 0$, where $\beta \gg \alpha$. Taking the limit as $\beta \rightarrow \infty$, the original prior with the step function is recovered.

The Golgi cell activity is given by the sum of granule cells activity: $z = \sum_{i=1}^{N} s_i$, where the number of granule cells N is about 1000–6000 (Palay and Chan-Palay, 1974). The Golgi cell density is given by a gamma density of order N, $f_z(z) = \gamma_N(z)$, which approaches a Gaussian density with a mean $\mu(z) = N/\alpha$ and a variance $\sigma^2 = N/\alpha^2$. In the following, the Golgi cell activity is assumed to be independent of the mossy fiber afferents x and is assumed constant at its mean value. The more realistic dependent case has been derived recently and will be presented elsewhere.

2.3.2. Maximum Likelihood. The objective is to maximize the probability density of the input data (X) given the model. The likelihood function in terms of M observations x_k of x is $f_X(X \mid W, V) = \prod_{k=1}^{M} f_x(x_k \mid W, V)$. Assuming a complete representation where 4 granule cells receive the same 4 mossy fiber inputs, the 4-dimensional mossy fiber input can be written as $x = W^{-1}(s + Vz)$ in the linear regime of $s = Wx - Vz$ by inverting the network. Dropping the index k, the density of a single data point is obtained by marginalizing over the states of the network, $f_x(x \mid W, V) = \int f_x(x \mid s, z, W, V) f_{sz}(s, z) \, ds \, dz$ where $f_x(x \mid s, z, W, V) = \delta(x - W^{-1}s - W^{-1}Vz)$ and where $\delta(\cdot)$ is the n-dimensional delta function. The joint density distribution $f_{sz}(s, z) = f_s(s) f_z(z)$, since s and z are assumed independent here, and $f_z(z)$ is a delta function, since z is assumed fixed at its mean value.

2.3.3. Learning Rules. The learning rules for $\{w_{ji}\}$, $\{v_{ji}\}$ are derived by taking the gradient of the log likelihood and multiplying the results by $\mathbf{W^T W}$ (Mackay, 1996): $\Delta w_{ji} \propto w_{ji} + \phi(s_j) \sum_j s_j w_{ji}$, $\Delta v_{ji} \propto -\phi(s_j)$, for active mossy fiber, $x_i > 0$, where $\phi(s_j) = -\alpha$ for $s_j > 0$ and $\phi(s_j) = \beta$ for $s_j = 0$ (Coenen et al., 1999). The backward connectivity is shown in Fig. 2. Notice that a backward summation $\sum_j s_j w_{ji}$ from granule cell activity is required at the ith glomerulus and that the particular connectivity at the glomerulus makes its computation possible (Fig. 2 and below). This summation is unique to the ith glomerulus, and is the same for all weight changes Δw_{ji}

at that glomerulus, but the sum $w_{ji} + \phi(s_j) \sum_j s_j w_{ji}$ is unique to each synapse at that glomerulus. Note that the learning rules hold whether **s** forms a (local) complete or undercomplete representation of **x** (Girolami et al., 1998), i.e., in our case, whether the number of granule cells with the same mossy fiber inputs is equal or smaller than the number of mossy fiber inputs to a granule cell. Since the number of granule cells exceeds the number of mossy fibers by a factor of more than 10 000, the global representation at the granular layer may itself be overcomplete or not.

The backward summation $\sum_j s_j w_{ji}$ is biological plausible in the granular layer of the cerebellum due to the unique convergence of information at the glomeruli (Fig. 2). Because the granule cells are electrotonically compact, the spiking activity at the soma is assumed to be reflected at the dendrites (D'Angelo et al., 1995). The biophysical mechanisms for computing the backward summation and its distribution at every synapse at the glomerulus are currently being investigated (Coenen et al., 2001).

2.4. Smooth Pursuit Model

A spiking neural network model based on the anatomy and physiology of the cerebellum was simulated using The Alchera Simulation Environment (M.P. Arnold, Univ. of Sydney). The model learns to generate predictive smooth pursuit eye movements to follow target trajectories, and respond to large tracking error by producing corrective saccades.

The granular layer contained 195 granule cells that received 21 mossy fiber inputs encoding for different aspects of eye movements and target: 4 mossy fibers encoded the retinal slip target position, velocity and acceleration for each, 4 mossy fibers encoded the position and velocity of the eye relative to the head for each, and finally 1 mossy fiber encoded the occurrence of a saccade. Each granule cell received afferents from a single Golgi cell and a unique set of 4 mossy fibers. The granule cells projected to 10 Golgi cells and a single Purkinje cell. The Purkinje cell projected to a cerebellar nucleus neuron and received a climbing fiber from the inferior olive that encoded retinal velocity slip of the target during smooth pursuit. The nucleus neuron encoded the output eye velocity to follow a target that oscillated in one dimension with angular position given by $\theta = 3\sin(2\pi t/4)$ where t is the time. The parameters of the two granule cell exponential priors were $\alpha = 0.2$ and $\beta = 10$ (Section 2.3.1). A saccade
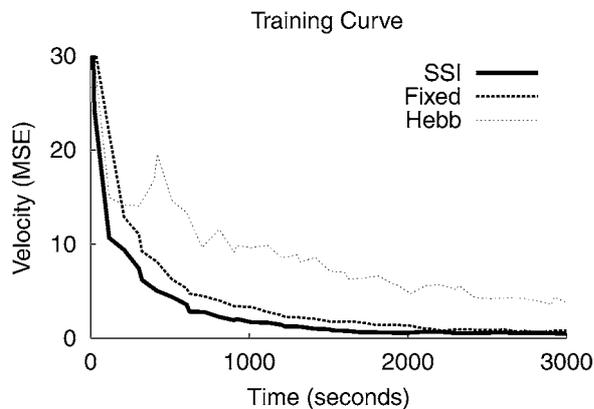
Training Curve

Figure 3. Performance comparison of the spiking cerebellar model for two different granular layer representations. The smooth pursuit velocity mean squared error (MSE) is plotted as function of time during learning. The curves show the results for 1) the sparse and statistically-independent (SSI) representation obtained from maximum likelihood, 2) the decorrelated representation (Hebb) that used Hebbian and anti-Hebbian learning rules for the mossy fiber-granule cell and Golgi cell-granule cell synapses, respectively, and 3) a fixed random representation (Fixed) with no learning.

Figure 4. Probability density function (pdf) of the granule cells activity after learning the sparse and statistically-independent (SSI) representation. The mean pdf of granule cells activity is plotted along with the standard deviation (std dev), the minimum (min) and maximum (max) values of the pdfs of all granule cells. The mean pdf is more sparse than the best fit exponential function (exp), which is the form of the sparse prior pdf used. The pdf for each of the 195 granule cells was computed over the final 1000 seconds of the simulation (9000 to 10000 seconds).

to the target was generated whenever the eye was more than $1.0°$ away from the target.

The performance of the network with the sparse and statistically-independent (SSI) representation in the granule cells was compared with the performance using a decorrelated representation (Hebbian) (Jonker et al., 1998) and a random representation with no learning (Fig. 3). The learning rules for the decorrelated representation were: 1) Hebbian for the mossy fiber-granule cell and granule cell-Golgi cell synapses and 2) anti-Hebbian for the Golgi cell-granule cell synapses. The random representation was obtained using random weight values. With the granule cells SSI representation, the mean squared error (MSE) in the smooth pursuit velocity decreased faster and remained lower than with the Hebbian representation. It also did slightly better than a random representation at the beginning of learning, but lost its advantage over time. The advantage of the SSI representation over a random representation should grow as the number of granule cells increases. An increase of three order of magnitude—from $\sim$200 to $\sim$200 000 granule cells—is needed to compare the different representations in more realistic physiological conditions, and such simulations are currently under way. The degree of independence of the granule cells will be analyzed and compared across all three representations. The storage capacity will also
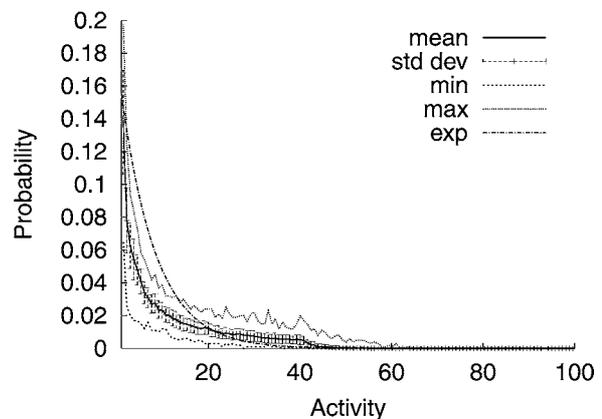
be estimated by estimating the destructive interference after learning different tasks.

Examination of the granule cell activities shows a sparse pattern of activities developing in the SSI model (Fig. 4). With the SSI representation, the granule cell probability densities are sparser than the exponential prior used. For the Hebbian representation or with no learning in the granular layer, the granule cells density is significantly more uniform between 10 and 50 Hz before dropping off to zero (not shown). With SSI, the weights were in general more stable over time and did not show the constant drift that they displayed with the Hebbian learning. This is because the weight change at a synapse with the SSI learning algorithm takes into account the values of all the other synapses at a glomerulus, in contrast to the Hebbian rules.

3. Conclusions and Future Work

This paper suggests that a powerful unsupervised learning algorithm in the granular layer of the cerebellum is physiologically plausible. The current preliminary results bring support to the hypothesis that these powerful unsupervised computations will facilitate learning at later stages in the cerebellum. Although the reduction of destructive interference during learning multiple

tasks remains to be demonstrated experimentally, the approach has great potential for continuous and life-long learning in robotics. Analysis, simulations, as well as robotic experimentations are currently under way to further expand the results presented here. Finally, the biophysical mechanisms supporting our hypothesis will be presented shortly (Coenen et al., 2001).

Acknowledgments

Note

1. The gamma density of order N is $\gamma_N(s_i) = \frac{\alpha^N}{\Gamma(N)} s_i^{N-1} e^{-\alpha s_i}$ where $\Gamma(N)$ is the gamma function. Gamma densities have the property that the density $f_z(z)$ of the sum of two independent random variables $z = s_1 + s_2$ with respective gamma densities of order p and q, $\gamma_p(s_1)$ and $\gamma_q(s_2)$, is a gamma density of order $p + q$, $f_z(z) = \gamma_{p+q}(z)$.

References

Atkeson, C.G., Moore, A.W., and Schaal, S. 1997. Locally weighted learning for control. *Artificial Intelligence Review*, 11(1–5):75–113.

Barbour, B. 1993. Synaptic currents evoked in Purkinje cells by stimulating individual granule cells. *Neuron*, 11(4):759–769.

Blakemore, S.J., Goodbody, S.J., and Wolpert, D.M. 1998. Predicting the consequences of our own actions: The role of sensorimotor context estimation. *J Neurosci*, 18(18):7511–7518.

Blakemore, S.J., Wolpert, D.M., and Frith, C.D. 1999. The cerebellum contributes to somatosensory cortical activity during self-produced tactile stimulation. *Neuroimage*, 10(4):448–459.

Coenen, O.J.-M.D. 1998. Modeling the vestibulo-ocular reflex and the cerebellum: Analytical & computational approaches. Ph.D. thesis, University of California, San Diego. Physics Department. Available at www.cnl.salk.edu/~olivier.

Coenen, O.J.-M.D., Arnold, M., Jabri, M.A., Courchesne, E., and Sejnowski, T.J. 1999. A hypothesis for parallel fiber coding in the cerebellum. In *Society for Neuroscience Abstracts*, Vol. 25.

Coenen, O.J.-M.D., Arnold, M.P., Sejnowski, T.J., and Jabri, M.A. 2000. Bayesian analysis for parallel fiber coding in the cerebellum. In *Proceedings of the 7th International Conference on Neural Information Processing (ICONIP-2000)*, Taejon, Korea, pp. 1301–1306.

Coenen, O.J.-M.D., Eagleman, D.M., Mitsner, V., Bartol, T.M., Bell, A.J., and Sejnowski, T.J. 2001. Cerebellar glomeruli: Does limited extracellular calcium direct a new kind of plasticity?. In *Society for Neuroscience Abstracts*, Vol. 27.

D'Angelo, E., De Filippi, G., Rossi, P., and Taglietti, V. 1995. Synaptic excitation of individual rat cerebellar granule cells in situ: Evidence for the role of NMDA receptors. *J Physiol* (Lond), 484 (Pt 2):397–413.

D'Angelo, E., Rossi, P., Armano, S., and Taglietti, V. 1999. Evidence for NMDA and mGlu receptor-dependent long-term potentiation of mossy fiber-granule cell transmission in rat cerebellum. *J Neurophysiol*, 81(1):277–287.

Girolami, M., Cichocki, A., and Amari, S.-I. 1998. A Common neural network model for unsupervised exploratory data analysis and independent component analysis. *IEEE Transactions on Neural Networks*, 9(6):1495.

Jakab, R.L. and Hamori, J. 1988. Quantitative morphology and synaptology of cerebellar glomeruli in the rat. *Anat Embryol* (Berl), 179(1):81–88.

Jonker, H.J., Coolen, A.C., and Denier van der Gon, J.J. 1998. Autonomous development of decorrelation filters in neural networks with recurrent inhibition. *Network*, 9(3):345–362.

Kettner, R.E., Mahamud, S., Leung, H.C., Sitkoff, N., Houk, J.C., Peterson, B.W., and G, B.A. 1997. Prediction of complex two-dimensional trajectories by a cerebellar model of smooth pursuit eye movement. *Journal of Neurophysiology*, 77(4):2115–2130.

MacKay, D.J.C. 1996. Maximum likelihood and covariant algorithms for independent component analysis. Unpublished manuscript, available at http://wol.ra.phy.cam.ac.uk/mackay/BayesICA.html.

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

Meunier, C. and Nadal, J.-P. 1995. *The Handbook of Brain Theory and Neural Networks.* Cambridge, MA: MIT Press. pp. 899–901.

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.

Olshausen, B.A. and Field, D.J. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images [see comments]. *Nature*, 381(6583):607–609.

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