Linear Network Models of the Oculomotor Integrators

H. S. Seung
AT&T Bell Laboratories
Murray Hill, NJ 07974

July 31, 1995

Abstract

The oculomotor system uses integrators to transform velocity signals to control eye position. We model such an integrator as a linear network with a marginal mode separated from stable modes by a large spectral gap. The model neurons carry a superposition of position and velocity signals, just as observed in single unit recordings from hindbrain oculomotor areas. The single unit properties of tonic activity, position sensitivity, velocity sensitivity, and position susceptibility are shaped by feedforward input and local feedback. Single unit response can also be given a geometric interpretation in terms of a line attractor in the state space of the network. Unlike nonlinear models, the linear model predicts uniformity of linear response along the line attractor. Both the transverse relaxation time under normal conditions and the integrator time constant after lesion or inactivation suggest that local feedback in the integrator is mediated by synapses with persistence times on the order of 100 ms.

Control theoretic models of the oculomotor system rely on integrators to transform velocity signals into eye position signals. While there has been great progress in localizing these integrators to hindbrain areas, the neural basis of integration is still obscure. Robinson and colleagues have advanced a series of linear network models that are constructed from neurons with short persistence times, but can integrate over long time scales by virtue of their local feedback connections[1, 2, 3, 4]. Because they found a diversity of integrator networks, robust predictions could only be made on the basis of characteristics common to many networks, rather than any particular one. This required an *understanding* of network function, which seemed impossible given the uninterpretability of single unit behavior: "the behavior of hidden units becomes more and more divorced from intuitive behavior as their tasks become more and more complicated"[5].

The present work solves Robinson's problems of nonuniqueness and uninterpretability by precisely characterizing the set of *all* linear network integrators.

Normal mode analysis shows that the capability of a linear network to integrate depends on the existence of a marginal mode. It also reveals how single unit response is determined by the interplay between feedforward input and local feedback. The normal modes of a linear system are the simplest example of collective dynamics encountered in mathematics and the physical sciences. Here we argue that normal modes are also the basis of collective computation in the oculomotor integrators.

The dynamics of a linear network integrator can be qualitatively described as possessing an attractive line of fixed points, or line attractor. These fixed points correspond to the persistent patterns of activity observed in hindbrain oculomotor areas during gaze-holding. Eye position is coded as the coordinate along the line attractor. Hence the oculomotor system provides a confirmation of the attractor hypothesis, that some collective computations in the brain are based on the attractor structure of state space[6, 7, 8].

Linear network models

Much of our analysis is applicable to any linear network that is used to integrate a single scalar variable. Relevance to the oculomotor system will be illustrated using specific comparisons to the horizontal integrator, which deals with rotations in the horizontal plane and is thought to be localized in the medial vestibular nucleus (MVN), nucleus prepositus hypoglossi (NPH), and cerebellum[9]. For simplicity, we will treat the horizontal integrator as separate from the integrators controlling the vertical and torsional degrees of freedom. Although this approximation does not capture the full three-dimensional functionality of the oculomotor system[10], it is a useful first step. Furthermore, we consider only the velocity to position integrators, not the velocity storage integrators that are also important at low frequencies.

We model the dynamics of neurons in the integrator as

$$\tau_m \frac{dv_i}{dt} + v_i = \sum_j W_{ij} v_j + h_i \ . \tag{1}$$

Each neuron i with firing rate v_i is modeled as a leaky integrator of its synaptic inputs. The microscopic time constant τ_m is a short biophysical time scale, like a membrane or synaptic time constant. For the network to properly integrate the feedforward input h_i over a long time scale, the local feedback connections W_{ij} must be chosen so as to balance the intrinsic leakiness of the neurons. Unless otherwise specified, all summations run from 1 to N, the number of neurons in the network.

The integrator receives feedforward input from vestibular afferents, each of which has a tonic firing rate and a phasic modulation that is linear in the head's angular velocity $\dot{H}(t)$[11]. The total feedforward input h_i to neuron i is assumed to be the linear summation of signals from several vestibular afferents, and hence is also linear in head velocity,

$$h_i = h_i^0 + \dot{H}(t)\zeta_i \ . \tag{2}$$

Here h_i^0 is the tonic feedforward input to neuron i, and ζ_i is the velocity sensitivity.

The neurons of the integrator synapse onto motor neurons, which in turn drive the oculomotor plant. We assume that this subsequent processing is feedforward and linear, neglecting nonlinearities such as motor neuron recruitment. Due to linearity, we need not model the motor neuron layer explicitly, but instead express the eye's angular position E directly in terms of the activity of neurons in the integrator,

$$\tau_e \dot{E} + E = \sum_i \rho_i v_i . \tag{3}$$

We refer to ρ_i as the "read-out" vector. The plant model is first-order and characterized by a viscoelastic time constant of $\tau_e \approx 200$ ms.

The task of the network is to achieve gaze stabilization–to make eye velocity $\dot{E}$ equal and opposite to head velocity $\dot{H}$, so that gaze velocity $\dot{E} + \dot{H}$ vanishes. In the vestibulo-ocular reflex (VOR), the position signal in (3) is produced by integrating the velocity signal in the vestibular afferents (2). Other types of eye movements also involve the integration of velocity signals, but we will restrict our discussion to the VOR for the sake of simplicity.

Normal Modes

A normal mode analysis requires diagonalization of the synaptic weight matrix W_{ij}. The right eigenvectors ξ_i^μ, left eigenvectors η_i^μ, and eigenvalues $\tilde{W}_\mu$ satisfy $\sum_j W_{ij}\xi_j^\mu = \tilde{W}_\mu \xi_i^\mu$, $\sum_i \eta_i^\mu W_{ij} = \tilde{W}_\mu \eta_j^\mu$, and $\sum_i \xi_i^\mu \eta_i^\nu = \delta_{\mu\nu}$, where $\mu = 1, \ldots, N$. The normal coordinates $\tilde{v}_\mu$ are related to the firing rates v_i via the equations $v_i = \sum_\mu \tilde{v}_\mu \xi_i^\mu$ and $\tilde{v}_\mu = \sum_i \eta_i^\mu v_i$. Analogous equations relate the feedforward input vector h_i to its coordinates $\tilde{h}_\mu$ in the eigenvector basis. In general, the coordinates of a vector with respect to the eigenvector basis will carry a tilde and a Greek index. When written in normal coordinates, the dynamical equations (1) take the form

$$\tau_m \frac{d\tilde{v}_\mu}{dt} = -(1 - \tilde{W}_\mu)\tilde{v}_\mu + \tilde{h}_\mu \tag{4}$$

The advantage of this description over (1) is manifest: N interacting neurons have been transformed into N noninteracting normal modes.

If the eigenvalue $\tilde{W}_\mu$ is real, the mode behaves as an integrator with input $\tilde{h}_\mu$ and time constant

$$\tau_\mu = \frac{\tau_m}{1 - \tilde{W}_\mu} \ . \tag{5}$$

The eigenvalue $\tilde{W}_\mu$ represents the amount of feedback that each mode receives. Positive feedback retards the decay of a mode, and negative feedback accelerates it. The mode is stable if the feedback is less than unity, and unstable if greater. If the feedback exactly balances the decay ($\tilde{W}_\mu = 1$), then the mode is called *marginal*.

For simplicity, we will assume that the eigenvalues $\tilde{W}_\mu$ are real. However, we note in passing that the eigenvalues of a real matrix W_{ij} can also occur in complex conjugate pairs. Because such modes can exhibit resonance behavior at frequency $(1 - \mathrm{Im}\,\tilde{W}_\mu)/\tau_m$, they correspond to the frequency-tuned channels postulated by Lisberger et al. to explain their experiments on VOR gain adaptation[12].

We now state the minimal set of five conditions that produce VOR functionality ($\dot{E} = -\dot{H}$) at frequencies less than ω_{max} and are consistent with single unit data. An explanation of these conditions follows. The eigenvalues are labeled in decreasing order, $\tilde{W}_1 \geq \tilde{W}_2 \geq \ldots \geq \tilde{W}_N$. The dominant eigenvalue $\tilde{W}_1$ must be exactly unity, corresponding to a marginal mode,

$$\tilde{W}_1 = 1 \ . \tag{6}$$

The other eigenvalues must be less than one and separated from $\tilde{W}_1$ by a gap large compared to $\omega_{max}\tau_m$,

$$1 - \tilde{W}_2 \gg \omega_{max}\tau_m . \tag{7}$$

The tonic input h_i^0 must be orthogonal to the left eigenvector η_i^1,

$$\sum_i \eta_i^1 h_i^0 = 0 \ . \tag{8}$$

The overlaps of the read-out vector ρ_i and phasic input vector ζ_i on the marginal mode must satisfy

$$\left(\sum_i \rho_i \xi_i^1\right)\left(\sum_j \eta_j^1 \zeta_j\right) = -\tau_m . \tag{9}$$

The overlaps of ρ_i and ξ_i on the stable modes must satisfy

$$\sum_{i,j} \rho_i G_{ij}^{\perp} \zeta_j = -\tau_e , \tag{10}$$

where the transverse response matrix $G_{ij}^{\perp}$ is defined below in (12).

The marginal mode functions as an integrator, $\tau_m d\tilde{v}_1/dt = \tilde{h}_1$, producing a position signal from the velocity input. The stable modes act as followers of the velocity input, $\tilde{v}_\mu \approx \tilde{h}_\mu/(1-\tilde{W}_\mu)$, for frequencies less than ω_{max}. Hence conditions (6) and (7) on the eigenvalue spectrum of the local feedback connections W_{ij} insure that integrator and follower modes exist. Conditions (8)–(10) concern the proper interfacing of these modes to feedforward input and output to produce gaze stabilization.

Condition (8) concerns the tonic input h_i^0. Even when the head is stationary, the vestibular afferents carry a high tonic activity, which the integrator must ignore[1]. The segregation of tonic from phasic input is easily accomplished via orthogonality. If the tonic input vector is orthogonal to the left eigenvector η_i^1, as in Eq. (8), then it does not couple to the integrator mode at all.

Conditions (9) and (10) concern the phasic input ζ_i and the read-out vector ρ_i. Here it is convenient to introduce the neuronal response matrix G_{ij}, defined in the frequency domain by $v_i(\omega) = \sum_j G_{ij}(\omega)h_j(\omega)$. If conditions (6) and (7) are satisfied, then the neuronal response matrix can be approximated by

$$G_{ij}(\omega) \approx \frac{\xi_i^1 \eta_j^1}{i\omega\tau_m} + G_{ij}^{\perp} \tag{11}$$

for frequencies less than ω_{max}, where the transverse response matrix

$$G_{ij}^{\perp} = \sum_{\mu=2}^{N} \frac{\xi_i^\mu \eta_j^\mu}{1-\tilde{W}_\mu} \tag{12}$$

is independent of ω. The form (11) clearly exhibits the distinction between the integrator and follower modes.

The performance of the network at gaze stabilization is quantified by the VOR transfer function $\chi(\omega) \equiv E(\omega)/H(\omega)$, which takes the form

$$\chi(\omega) = \frac{i\omega}{1+i\omega\tau_e} \sum_{i,j} \rho_i G_{ij}(\omega)\zeta_j \tag{13}$$

Perfect gaze stabilization ($\chi(\omega) = -1$) is achieved when conditions (9) and (10) are satisfied.

Single unit response

In the minimal model described above, it can be shown that single unit response is a linear function of eye position and velocity for frequencies less than ω_{max},

$$v_i = v_i^0 + k_i E + r_i \dot{E} \ . \tag{14}$$

The position sensitivity k_i, velocity sensitivity r_i, and tonic activity v_i^0 take the form

$$\begin{aligned} k_i &= \xi_i^1 \ , & (15)\\ r_i &= -\sum_j G_{ij}^{\perp}\zeta_j \ , & (16)\\ v_i^0 &= \sum_{jk} P_{ij}G_{jk}^{\perp}h_k^0 \ . & (17) \end{aligned}$$

Here we have normalized ξ_i^1 to satisfy $\sum_i \rho_i\xi_i^1 = 1$. Since the normal coordinate $\tilde{v}_1$ codes for eye position, its eigenvector ξ_i^1 is the position sensitivity. The velocity sensitivity is given by the transverse response (that of the follower modes) to the phasic input ζ_i. The tonic activity v_i^0 is the transverse response to the tonic input h_i^0, with the application of a projection operator $P_{ij} = \delta_{ij} - \xi_i^1\rho_j$ that adds a multiple of ξ_i^1 to satisfy the constraint of $\sum_i \rho_i v_i^0 = 0$ at $E = 0$.

Single unit recordings in the MVN and NPH[13, 14] have in fact revealed a population of cells with responses of the form (14). The directions of increasing and decreasing firing rates are known as the on- and off-directions respectively. Since the firing rate is never negative, there is a threshold nonlinearity not described by the linear relationship (14). When the eyes are stationary ($\dot{E} = 0$), the threshold position is at $E_T = -v^0/k$, which is typically about 20° in the off-direction.

Note that the position sensitivity is purely a property of the local feedback connections W_{ij}. However, the tonic activity and velocity sensitivity depend on both local feedback and feedforward input. This explains the empirical fact that the position sensitivity of a neuron is the same regardless of the type of eye movement, but the velocity sensitivity can vary[2]. Although all types of eye movement appear to be velocity coded, the phasic input ζ_i varies according to the kind of eye movement, causing variation in velocity sensitivities r_i.

Our normal mode analysis suggests another single unit response property hitherto unmeasured. The components of the left eigenvector η_i^1 are the *position susceptibilities* of the neurons. Since the input to the integrator mode is $\tilde{h}_1 = \sum_i \eta_i^1 h_i$, the position susceptibility is a measure of the rate at which the integrator will move in response to intracellular injection of current into neuron i. Although this movement may be down by a factor of $1/N$ compared to natural movements, it could be observable in a small integrator network, like that of the goldfish, which is purported to have only 25 to 40 neurons[15].

Robinson has emphasized the apparent uninterpretability of single unit response[5], noting that in both models and experiments, one finds "rogue cells" with the nonintuitive property that their preferred directions for position and velocity sensitivity are opposite, $k_i r_i < 0$. In fact, due to the existence of three directional single unit properties, position sensitivity k_i, velocity sensitivity r_i, and position susceptibility η_i^1, there are three types of "rogue" behavior: $k_i r_i < 0$, $k_i\eta_i^1 < 0$, and $r_i\eta_i^1 < 0$. Because of the sum rule $\sum_i r_i\eta_i^1 = 0$, any integrating network must contain some rogue cells. None of these rogue behaviors are surprising, given that signals are actually carried by collective modes rather than single neurons.

Though we have not given a formal proof that conditions (6)–(10) are indeed minimal, it should be clear from the informal discussion above that these conditions are sufficient to produce VOR functionality and the single unit response (14). Some of the conditions can be relaxed without losing VOR functionality. For example, there could be more than one marginal mode for a single degree of ocular freedom, or invisible modes (with zero coupling to the input or to the read-out) occupying the gap between the integrator and follower modes. However, these possibilities so far seem inconsistent with single unit recordings, so conditions (6)–(10) are necessary as well as sufficient.

Relaxation time to the line attractor

Single unit response also has geometric meaning in the state space of the network, the N-dimensional space whose coordinates are v_i. The tonic activity vector v_i^0 is a point in state space, and the position sensitivity vector k_i is a direction. Together they specify a line $v_i^0 + k_i E$ parametrized by eye position. In the presence of purely tonic input, every point on this line is a fixed point of the dynamics. The firing rates $v_i^0 + k_i E$ constitute a persistent pattern of activity corresponding to persistent gaze at eye position E. The line is an attractor of the dynamics, because any initial condition off the line flows to a point on the line. This line attractor is the basis of the memory of eye position.

A line attractor is well-suited for the storage of a continuously varying quantity like eye position. However, it is not the only state space structure that can be used to construct an integrator. A digital integrator, in contrast, would have discrete fixed points corresponding to quantized values of eye position. This sort of state space structure is also found in attractor-based models of associative memory[6, 7, 8]. However, neurons in such models have strongly nonlinear behavior, contrary to the linear single unit responses (14) that have been observed in the oculomotor integrator. Network integrators that are intermediate between the analog and digital extremes can also be constructed, as will be discussed elsewhere.

Phasic modulation of the input causes the network to move in state space. The movement consists of two components: a longitudinal one along the line attractor, and a transverse one away it. Movement along the line is persistent; after the phasic input is over, any changes in eye position persist. Movement off the line is transient; after the phasic input is over, the network relaxes back to the line.

The stable mode with the longest time constant, $\tau_2 = \tau_m(1 - \tilde{W}_2)^{-1}$, determines the transverse relaxation time of the model. This time can be measured experimentally from the response of the integrator to a delta-function input.

Such a brief, transient stimulation occurs naturally with every saccade, which is driven by a short pulse of activity from burst neurons. Although there has been no systematic study of the saccadic responses of integrator neurons, they seem to exhibit a pulse-slide-step response similar to that observed in motor neurons, a brief pulse of high frequency activity followed by an order 100 ms relaxation (slide) to a new steady level of activity (step)[15]. So the transverse relaxation time τ_2 appears to be of order 100 ms. By condition (7), our integrator model can give proper VOR functionality only for frequencies less than $\omega_{max} \approx 1/\tau_2 = 10$ Hz. Whether or not VOR function truly deteriorates at such frequencies has not yet been resolved experimentally[16].

The transverse relaxation time τ_2 is an upper bound for the microscopic time scale τ_m. This is because negative feedback is required to make a mode relax more quickly than the microscopic time scale τ_m. But if the eigenvalues of the stable modes are all negative, then each neuron must inhibit itself very strongly, since the eigenvalues $\tilde{W}_\mu$ and the self-couplings W_{ii} are related via $\sum_\mu \tilde{W}_\mu = \sum_i W_{ii}$. Such a scenario of strong self-inhibition is implausible. In particular, the upper bound $\tau_m \leq \tau_2$ rules out the extreme hypothesis that each neuron by itself is an integrator.

Robustness

Up to now we discussed the case of a perfect integrator with neither leak, instability, nor bias. Of course, this is only a mathematical idealization. Under normal conditions, the integrator is slightly leaky in humans, with a time constant τ_1 of typically 20–40 seconds[17]. This is measured by asking subjects to hold eccentric eye positions in the dark. Since $1 - \tilde{W}_1 = \tau_m/\tau_1$, the ratio τ_m/τ_1 gives the precision to which $\tilde{W}_1$ must be tuned to unity. If $\tau_m = 5$ ms, as proposed by Robinson and colleagues, and $\tau_1 = 20$ sec, the tuning of $\tilde{W}_1$ must be accurate to within one part in 4000. Is this delicate tuning biologically plausible?

We can only guess at the types of perturbations that the integrator is subject to under normal conditions. Two plausible types are independent random fluctuations of synaptic strengths, and loss of a single neuron. The effects of these two types of perturbations on $\tilde{W}_1$ will be analyzed in detail elsewhere; here we briefly summarize the results. The resulting shift depends on the structure of W_{ij}. In the most robust case, and given a large gap, the shift is of order $1/N$, where N is the number of neurons in the network. Intuitively, this is because the effects of random fluctuations at different synapses tend to cancel each other, and loss of a single neuron has less effect when there are many neurons. So the required precision of one in 4000 can be met with something like 4000 neurons. While this is barely plausible in monkey and cat, recent work in the goldfish indicates that the integrator is localized to nuclei containing no more than 100 neurons[15]. This suggests that τ_m must be larger than 5 ms.

For the "normal" perturbations discussed above, networks become more robust with increasing size. However, under extreme perturbations this is not the case. For example, injection of neurotransmitter agonists or antagonists presumably leads to systematic changes in synaptic efficacy, rather than random uncorrelated ones. Electrolytic lesions and pharmacological inactivation affect a substantial fraction of the network, not just a single neuron. In both cases, networks do not gain robustness from increased size. Generically, we would expect that extreme perturbations should result in a time constant τ_1 that is an order unity multiple of τ_m. After extreme perturbations of the network, experimental measurements of the integrator time constant generally find a τ_1 on the order of hundreds of milliseconds[14]. This suggests that τ_m is of order 100 ms.

Proof of this hypothesis requires single unit recordings after lesions or inactivation, in contrast to behavioral experiments, which are unable to resolve an integrator time constant that is less than τ_e, the time constant of the oculomotor plant. We predict that the integrator time constant cannot be made to fall below 100 ms, thus ruling out the possibility that $\tau_m = 5$ ms.

The microscopic time scale

Based on comparisons with the transverse relaxation time τ_2 under normal conditions and the integrator time constant τ_1 after extreme perturbations, we have argued that τ_m is of order 100 ms. What is the biophysical basis of this time scale? To answer this question we must discuss the biophysical interpretation of the neural dynamics (1) in more detail

The linear network dynamics (1) is a special case of model equations proposed by Hopfield[18] and others:

$$\tau_m \dot{u}_i + u_i = \sum_j W_{ij} g(u_j) + h_i \tag{18}$$

The state of each neuron is described by its average firing rate $v_i = g(u_i)$, where the average is taken over a short time period. Since this coarse-grained description ignores the "quantal" nature of action potentials, Hopfield has called it "classical neurodynamics." A necessary condition for the validity of classical neurodynamics is that there be no synchronized firing of neurons. This could be empirically verified via cross-correlation analysis.

There are two alternative biophysical interpretations of u_i, either as a membrane potential or as a synaptic current. If u_i is the membrane potential of the soma, then τ_m is the membrane time constant, and g a function relating membrane potential to firing rate. This is the interpretation of Robinson and colleagues, but it is implausible because membrane time constants are too short to account for the observed robustness of the integrator.

Another interpretation is that u_i represents the total synaptic current into neuron i. In this case, τ_m is the temporal integration time of the synapse.

A microscopic time scale of $\tau_m \sim 100$ ms could thus correspond to the time constant of the NMDA[19] or GABA-B receptors, both of which are involved in recurrent pathways in the MVN. The transfer function g relates firing rate to current rather than voltage. In vitro current injection studies in MVN slices have revealed a major class of neurons that possess an unusually wide linear operating range[20]. At threshold, the firing rate rises almost continuously from zero, and thereafter increases linearly up to hundreds of spikes per second. Hence the function g is threshold linear. Since most neurons operate above threshold for eye positions near central gaze, it is a reasonable approximation to neglect the threshold nonlinearity, and the equations (18) reduce to the linear network dynamics (1).

Testing linearity

How can the linear network hypothesis be tested? To test the hypothesis, we must have an alternative, and that alternative is a nonlinear network. For a general nonlinear network with a line attractor, the linear response to an infinitesimal perturbation takes a form similar to (11). However, the linear response matrix may vary along the line attractor, while for a linear network the response matrix is uniform. What is the empirical evidence for uniformity of ξ_i^1, η_i^1, and $G_{ij}^{\perp}$, the three elements of (11)? It is already known that the vector ξ_i^1 is uniform all along the line, because the line attractor is straight. Many models with strong nonlinearity are inconsistent with this simple fact[21]. The uniformity of η_i^1, and $G_{ij}^{\perp}$ remains to be checked. The uniformity of the velocity sensitivity $r_i = \sum_j G_{ij}^{\perp}\zeta_j$ should be checked thoroughly, and also using artificial means of stimulation to produce phasic input vectors ζ_i that are not naturally seen by the integrator.

Conclusion

We have argued that positive feedback creates a persistent memory of eye position that can be altered by feedforward input. This kind of explanation dates back to the 1930s, when Lorente de Nó formulated his theory of 'closed "self-reexciting" chains' to account for the "reflex after-discharge," which the three-neuron vestibuloocular reflex arc alone could not explain[22]. We have refined this explanation by advancing an attractor hypothesis: local feedback creates a line attractor in the state space of the network, and eye position is coded by the coordinate along this line. The persistent patterns of activity observed in the integrator network during gaze-holding correspond to fixed points on this line attractor. Feedforward input produces eye movements by pushing the network along the line.

The qualitative notion of a line attractor finds quantitative expression in the linear network model, previously studied by Robinson and colleagues. Our minimal model (6)–(10) gives a general specification of linear network integrators, freeing us from dependence on particular models. Thus we are able to distinguish between those network characteristics that are important for integration, and those that are extraneous. Our analysis also identifies important quantities like the spectral gap, the transverse relaxation time, and the position susceptibility, which are experimentally observable.

Unlike Robinson and colleagues, and Galiana and Outerbridge[23], we do not single out lateral inhibition as the mechanism of integration. Cannon et al.[1] observed that a network with purely excitatory local feedback cannot ignore the tonic input, as can be shown formally using the Perron-Frobenius theory of nonnegative matrices. However, this does not imply that excitation is irrelevant to integration in a network with a mixture of excitation and inhibition. Since both excitatory and inhibitory local feedback are known to exist in the MVN and NPH, it is most likely that both contribute to the operation of the integrator.

We have also argued that the microscopic time constant is a synaptic time scale of order 100 ms. This is in contradiction to Cannon et al.[1], who found that their model was very robust to single neuron lesions, even with a small microscopic time constant τ_m of 5 ms. By the inclusion principle[24], after a symmetric network loses a single neuron, the post-lesion value of $\tilde{W}_1$ can never drop below the pre-lesion value of $\tilde{W}_2$. Since Cannon et al. studied a symmetric weight matrix with a very small gap, the shift in $\tilde{W}_1$ was negligible. But in the relevant case of a large gap, networks are not so robust to the same single neuron lesions, as Arnold and Robinson[3] were later surprised to find. Cannon et al. also reported robustness to random synaptic fluctuations, but this was with the artificial assumption that the fluctuations had no overlap on the integrating mode.

Future work must address the issues of nonlinearity and learning. Modeling should elucidate the relationship between nonlinearities in single unit response and the computational capabilities of the oculomotor system, such as the noncommutativity of a truly three-dimensional VOR[10]. Following Arnold and Robinson[4], it is important to understand how a Hebb-type learning rule can tune a linear network to marginality.

Marginality is a basic feature of the oculomotor system, which abounds in integrators. Will marginality prove to be a basic feature of other motor areas? Postural systems are prime candidates for marginality, due to their similarity to gaze-stabilizing systems. If the equilibrium-point hypothesis of movement as a kind of postural shift is correct[25], marginality could also be important for the control of voluntary movement. Line attractors have also appeared in recent models of the motor[26], primary visual[27], and inferotemporal[28] cortices.

Acknowledgments

I am indebted to B. Shraiman and D. Tank for ideas and encouragement, D. Kleinfeld for a critical reading of the manuscript, and H. Sompolinsky for introducing me to line attractors. I am also grateful to R. Baker, K. Delaney, P. Mitra, D. A. Robinson, and S. Shastry for helpful conversations.

References

[1] S. C. Cannon, D. A. Robinson, and S. Shamma. A proposed neural network for the integrator of the oculomotor system. *Biol. Cybern.*, 49:127–136, 1983.

[2] S. C. Cannon and D. A. Robinson. An improved neural-network model for the neural integrator of the oculomotor system: more realistic neuron behavior. *Biol. Cybern.*, 53:93–108, 1985.

[3] D. B. Arnold and D. A. Robinson. A learning network model of the neural integrator of the oculomotor system. *Biol. Cybern.*, 64:447–454, 1991.

[4] D. B. Arnold and D. A. Robinson. A neural network model of the vestibulo-ocular reflex using a local synaptic learning rule. *Phil. Trans. R. Soc. of London*, B337:327–330, 1992.

[5] D. A. Robinson. Implications of neural networks for how we think about brain function. *Behav. Brain Sci.*, 15:644–655, 1992.

[6] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Nat. Acad. Sci. USA*, 79:2554–2558, 1982.

[7] J. J. Hopfield and D. W. Tank. Computing with neural circuits: a model. *Science*, 233:625–633, 1986.

[8] D. Amit. *Modeling Brain Function.* Cambridge University Press, Cambridge, 1989.

[9] R. J. Leigh and D. S. Zee. *The neurology of eye movements.* Davis, Philadelphia, 2nd edition, 1991.

[10] D. B. Tweed and T. Vilis. Implications of rotational kinematics for the oculomotor system in three dimensions. *J. Neurophysiol.*, 58:832–849, 1987.

[11] D. A. Robinson. The use of control systems analysis in the neurophysiology of eye movements. *Ann. Rev. Neurosci.*, 4:463–503, 1981.

[12] S. G. Lisberger, F. A. Miles, and L. M. Optican. Frequency-selective adaptation: evidence for channels in the vestibulo-ocular reflex? *J. Neurosci.*, 3:1234–1244, 1983.

[13] J. Lopez-Barneo, C. Darlot, A. Berthoz, and R. Baker. Neuronal activity in prepositus nucleus correlated with eye movement in the alert cat. *J. Neurophysiol.*, 47:329–352, 1982.

[14] K. Fukushima, C. R. S. Kaneko, and A. F. Fuchs. The neuronal substrate of integration in the oculomotor system. *Prog. in Neurobiol.*, 39:609–639, 1992.

[15] A. M. Pastor, R. R. De La Cruz, and R. Baker. Eye position and eye velocity integrators reside in separate brainstem nuclei. *Proc. Natl. Acad. Sci. USA*, 91:807–811, 1994.

[16] G. E. Grossman, R. J. Leigh, E. N. Bruce, W. P. Huebner, and D. J. Lanska. Performance of the human vestibuloocular reflex during locomotion. *J. Neurophysiol.*, 62, 1989.

[17] K. Hess, H. Reisine, and M. Dursteler. Normal eye drift and saccadic drift correction in darkness. *Neuro-ophthalmol.*, 5:247–252, 1985.

[18] J. J. Hopfield. Neurons, dynamics and computation. *Physics Today*, 47:40–46, 1994.

[19] G. A. Kinney, B. W. Peterson, and N. T. Slater. The synaptic activation of N-methyl-D-aspartate receptors in the rat medial vestibular nucleus. *J. Neurophysiol.*, 72:1588–1595, 1994.

[20] M. Serafin, C. de Waele, A. Khateb, P. P. Vidal, and M. Muhlethaler. Medial vestibular nucleus in the guinea-pig. i. intrinsic membrane properties in brainstem slices. *Exp. Brain Res.*, 84:417–425, 1991.

[21] M. J. Rosen. A theoretical neural integrator. *IEEE Trans. Biomed. Engin.*, BME-19:362–367, 1972.

[22] F. Lorente de Nó. Vestibulo-ocular reflex arc. *Archives of Neurology and Psychiatry*, 30:245–291, 1933.

[23] H. L. Galiana and J. S. Outerbridge. A bilateral model for central neural pathways in vestibuloocular reflex. *J. Neurophysiol.*, 51:210–241, 1984.

[24] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge Univ., 1985.

[25] E. Bizzi, N. Hogan, F. A. Mussa-Ivaldi, and S. Giszter. Does the nervous system use equilibrium-point control to guide single and multiple joint movements? *Behav. and Brain Sci.*, 15:603–613, 1992.

[26] A. P. Georgopoulos, M. Taira, and A. Lukashin. Cognitive neurophysiology of the motor cortex. *Science*, 260:47–52, 1993.

[27] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky. Theory of orientation tuning in visual cortex. *Proc. Nat. Acad. Sci. USA*, 1995. to appear.

[28] D. Zipser, B. Kehoe, G. Littlewort, and J. Fuster. A spiking network model of short-term active memory. *J. Neurosci.*, 13:3406–3420, 1993.