

Saccadic eye movements: oculomotor control in the superior colliculus

Christopher Burdess < *dog@dog.net.uk* >

February, 1996

Contents

1	Introduction	1
2	Method	2
3	Implementation	4
4	Discussion	7

1 Introduction

An important part of human perceptual activity is the numerous eye movements that must be made during even the shortest task. The organisation of photoreceptors in the eye is such that the vast majority are located in a small, shallowly depressed region in the centre of the retina called the *fovea*, a window that receives light stimuli subtending approximately 1° of visual field. In trichromatic humans, most of the colour-sensitive cone photoreceptors are located in this area, and the proportion of contrast-sensitive rod receptors dominates in this region as well. It is therefore often necessary to relocate the position of the eye so that the image of an object that requires detailed perceptual analysis projects onto this high-resolution area. The method by which this is done is known as a *saccade* - an oculomotor reaction that moves the eyeball such that the image of interest is fovealised.

A great deal of neurophysiological research has been carried out within the field of oculomotor control. Sparks & Nelson[4] have shown that saccades are triggered in the superior colliculus at the dorsal side of the mesencephalon. A topographic mapping is in effect between retinal photoreceptors and groups of neurons in the upper layer of the superior colliculus, i.e. the neighbourhood of retinal neurons corresponds proportionally to the neighbourhood of collicular neurons in such a way that a *retinotopic map* is created in the superior colliculus. In a similar way, the topographic organisation of neurons in the lower layer of the superior colliculus corresponds to saccades triggered by neuronal stimulation in this layer. Experimentation has shown that the intensity of artificial stimulation of collicular neurons does not significantly affect the direction of the resultant saccades: only their topographic location is important in this respect.

This paper is a continuation of the work done by Ritter et al[3] on the implementation of a model of the superior colliculus that can learn saccadic behaviour, based on Robinson's (1972) *fovealisation hypothesis*, the hypothesis that the connections between the upper and lower layers of the superior colliculus serve to learn and represent saccadic motor movements to centre the object of interest in the visual field.

The method by which the model learns the saccades is in accordance with current knowledge of the process of saccadic correction occurring in the oculomotor system. When a saccade

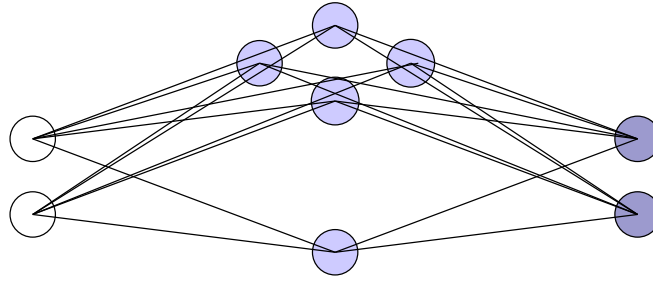


Figure 1: Schematic of the architecture of the model. White nodes represent the stimulus, dark grey nodes represent the direction of the saccade. In the centre, the light grey nodes represent the lattice units. Not all lattice units are shown.

is triggered, the image may or may not be fovealised as a result. If it is not, a corrective saccade is triggered, which also may or may not cause fovealisation of the desired image. However, if the position of the eye after the corrective saccade is closer to the fovea than after the first, the saccade weight for the first is modified to approach the vector addition of the two saccade weights. Ritter et al describe this process as a form of supervised learning. However, in the light of the fact that no learning steps are made if the first saccade correctly fovealises the image or if the corrective saccade leads the image further from the fovea than it was after the first, and that learning occurs without resort to any kind of global error measure, I prefer to consider this algorithm to be a form of reinforcement learning.

2 Method

As in Ritter et al’s[3] model, I have used a single layer of formal neurons to represent the two layers of efferent and afferent neurons in the superior colliculus, where the activation or excitation of these neurons represents a receptive field of variable width centred on an internal representation of the retina with respect to the object of interest. Thus, there are two sets of connection weight values: firstly, those leading to the formal neuron layer, or *lattice*, which are termed *lattice weights*, and which are responsible for determining the receptive fields of the lattice; and secondly, those values representing the saccade that is triggered when the corresponding neuron in the lattice is excited.

Ritter et al use a simplification of the neurophysiological reality to determine the saccade:

“...we describe the correspondence between the visual stimulus and the resultant saccade simply by a pair of values $(w_r, w_r^{(out)})$ of the centrally localized neuron. In reality, the resultant saccade is determined by a group of excited neurons localized at r .”

The model outlined in this paper uses just such a group of neurons to determine the actual saccade output. Specifically, a Gaussian distribution over all the neurons in the lattice, centred on the neuron with the receptive field centred on the location of the stimulus vector in retinal space, is used to calculate the resultant saccadic movement. It is hoped that this endows the model with a greater degree of neurobiological plausibility.

The lattice architecture is ring-shaped, as in Ritter et al’s model, with 30 circumferential neurons in a layer and 20 radial layers. The rotational symmetry of this architecture is suggested by the rotational symmetry of the input space. Ritter et al additionally mention that

“Every neural unit has two radial and two circumferential neighbours.”

This is obviously false, as what they are describing is a torus, rather than a cobweb. Obviously, in a ring-shaped lattice, the innermost (foveal) and outermost (peripheral) neurons only

have three neighbours. We will return to another problem with their description of the model later.

The choice of stimulus vector, as in Ritter et al's model, is random yet determined by a fixed probability density $P(v^L)$ corresponding to a Gaussian distribution with a width of 40° from the centre of the retina (maximum 90°), which function corresponds closely to the distribution of receptors on the retina from the centre, as noted by Korn[2] with stimuli absent within the radius of the fovea (1°). The initial state of the lattice and saccade weights is also randomised.

The algorithm for the development of this model of the superior colliculus is as follows:

1. Present a vector v^L in accordance with the probability distribution $P(v^L)$ above.
2. Find the centre of excitation c^L in the lattice according to the "winner-takes-all" condition

$$\|v^L - w_{o^L}^L\| \leq \|v^L - w_i^L\| \forall i$$
3. Perform the learning step for the lattice weights to form a Kohonen-like topology-conserving map onto the lattice

$$\Delta w_i^L = \eta^L h_{i,c^L}^L (v^L - w_i^L) \forall i$$
4. Trigger the saccade centred on c^L so that the new position v^S of the image occurs at

$$v^S = v^L + \frac{\sum_i h_{i,c^L}^S w_i^S}{\sum_i h_{i,c^L}^S}$$
5. If the image is now in the fovea, i.e. $\|v^S\| < r_{fovea}$, go to step (1)
6. Find the new centre of excitation c^S in the lattice for the stimulus v^S according to

$$\|v^S - w_{c^S}^L\| \leq \|v^S - w_i^L\| \forall i$$
7. Trigger the saccade centred on c^S so that the new position $v^{S'}$ of the image occurs at

$$v^{S'} = v^S + \frac{\sum_i h_{i,c^S}^S w_i^S}{\sum_i h_{i,c^S}^S}$$
8. If this corrective saccade is an improvement, i.e. $\|v^{S'}\| < \|v^S\|$, perform the learning step for the saccade weights according to

$$\Delta w_i^S = \eta^S h_{i,c^L}^S ((w_{c^L}^S + w_{c^S}^S) - w_i^S) \forall i$$
9. Go to step (1)

The terms $h_{i,u}^L$ and $h_{i,u}^S$ are Gaussian functions of the magnitude of the distance $\|i - u\|$ contingent on σ^L and σ^S , respectively, which in turn decrease over time, like the learning rates η^L and η^S , according to a standard exponential decay. The parameters in this model were chosen to be

- $\eta^L(t) = 0.3 \cdot \exp(-0.0002t)$
- $\sigma^L(t) = 10 \cdot \exp(-0.0003t)$
- $\sigma^L(t) = 10 \cdot \exp(-0.0003t)$
- $\sigma^S(t) = 3 \cdot \exp(-0.0003t)$

where t is an integer value indexing the time, or stimulus presentation. The actual determination of the distance between two neurons i and u is a complex matter since there are a number of ways in which this can be calculated. Ritter et al describe the situation as follows:

“The precise form of the distance measure between two neural units in the lattice is inconsequential for the organizational process to converge; however, sometimes a certain metric may fit a problem better than other distance measures. ...we used the ‘Manhattan’ rather than the Euclidean metric. ... it is the distances in the lattice which determine the spatial interaction between the neurons themselves and, thereby, determine the distance-dependent adaptation steps in the model. ... (I)n the vicinity of the fovea ... receptors that are directly opposite to each other lie close together but have to learn saccades that differ as much as saccades that belong to receptive fields directly opposite ... Therefore, it makes sense to use the ‘Manhattan’ metric which yields, for the foveal and peripheral pair, the same lattice distance between the diametrically opposite neural units.”

I would like to take issue with this explanation. By defining a distance metric, one is also defining the architecture of the formal neuron lattice. If the Manhattan metric is used, the lattice will be seen to have the form of a cylinder rather than a ring, since the distance between two circumferential units at foveal-receptive positions and peripheral-receptive positions is the same. In practice, this means that it is unlikely that the lattice weights will ever come to represent the input space effectively: the geometry of the input space and the lattice architecture is such that unless the first few learning steps cause one ring edge (foveal or peripheral) to become completely encircled by the other, the lattice weights will come to represent the input space in the form of a windsock, with one edge representing one half of the visual field and the other the other half. The same problem occurs with the straightforward Euclidean metric (counting only single units). Therefore, a distance metric must be found that preserves the actual geometric structure of the lattice. For the saccades, it is true, the Manhattan metric may be beneficial for the reasons described above, but unless Ritter et al are to explain why they have chosen different neighbourhood metrics for models of groups of neurons that at a neurophysiological level appear to be quite homogeneous, we should have to reject such an explanation, as the above justification looks very much like an ad hoc solution to the problem. How can the system “know” that some neurons will come to represent a particular part of the input space? If, for instance, some part of the visual field had significantly less interesting objects for which fovealisation was required, the network should come to represent that asymmetry naturally. In that case (the “foveal” inner ring of the lattice being off-centre from the fovea), neurons on the *same* side of the ring would have to learn very different saccade vectors. In a biological implementation, the agent might have to cope with this very same problem, in that after the lattice has formed, afferent connections from the retina or from the executive interest-generating system may be lesioned, providing an asymmetric stimulus probability distribution which must nevertheless be properly represented in order for sensible functioning to occur. Ritter et al’s model, in this case, will always be much more severely impaired than one which uses a straightforward metric representative of the true form of the lattice.

In this model, I have chosen a suitable metric for the distance between units in the lattice, namely

$$\| i - u \| = \sqrt{[(r_i \sin(\frac{2\pi c_i}{C})) - (r_u \sin(\frac{2\pi c_u}{C}))]^2 + [(r_i \cos(\frac{2\pi c_i}{C})) - (r_u \cos(\frac{2\pi c_u}{C}))]^2}$$

where r is the radial component of a unit (or how far it is from the centre of the lattice), c is the circumferential component of a unit (some proportion of C), and C is the total number of circumferential units (30 in this model). This metric represents the Euclidean distance between the positions of the units in the lattice, assuming that they are arranged evenly in the ring. In this way, we preserve the geometric structure of the lattice.

3 Implementation

The above model was implemented on an 80486DX2-66 processor over 16,000 training presentations. In the simulation, the two main windows represent the states of the lattice and saccade weights. The outer ring circumscribes the limit of the visual field. Each neuron’s position is

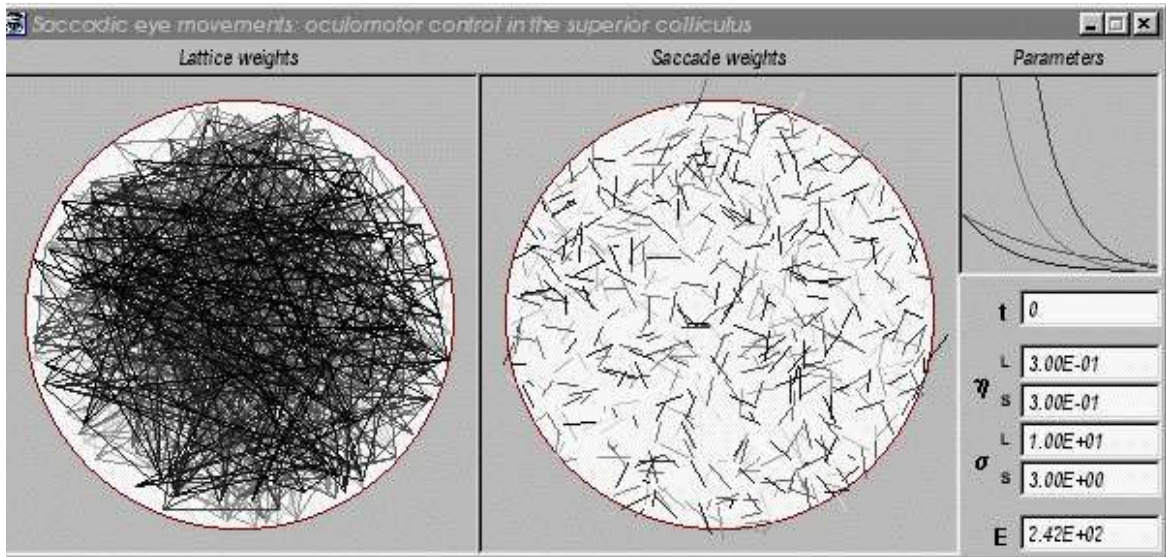


Figure 2: Lattice and saccade weight space before any stimuli.

represented in this space by the retinal location of the centre of its receptive field. In the lattice window, each neuron is connected by lines to its immediate neighbours. In the saccade window, each line shows the saccade vector. Since this model uses a group of neighbouring neurons to dictate the eventual position of the image after the saccade, the actual position of the image may not be precisely the same as the saccade vector would appear to show (the context of the centre of excitation in the lattice determines the measure of error of that vector). To make the weight space a little more comprehensible, I have used a greyscale function to display the lattice connections and saccade vectors proportionally to the radial component of their neural units in the lattice. In the parameters window, the values of the parameters are plotted over a 20,000-step time window. Additionally, an error measure was calculated every 200 timesteps and plotted in white in this window. This measure is simply defined as

$$E = \frac{\sum_i \|w_i^L - w_i^S\|}{N}$$

where N is the number of neurons in the lattice, as a proportion of the radius of the visual field.

Figure 3 shows the state of the network before any training has taken place. The magnitude of the randomised saccade vectors was set to a small proportion of the radius of the visual field. This helps the network to settle down quickly, and was also the method used by Ritter et al, but is not strictly necessary, as convergence of the saccade weights will still occur given enough time.

Figure 3 shows the state of the network after 4,000 timesteps. At this point we can see that the lattice has formed a more or less isomorphic cobweb shape, and the saccades have all become directed towards the fovea. The error measure display shows, however, that the network's performance on saccade learning over time is nonmonotonic. This problem occurs because the lattice weights, in the early stages, move around the input space following the stimulus vectors to a great degree (η^L is relatively high). This means that even if a saccade led directly into the fovea at one timestep, the centre of the receptive field of that neuron might change in the next timestep, causing the saccade to miss the fovea. It is therefore important for the lattice to become relatively stable with respect to the saccades at an early stage. This will be discussed further later.

Figure 3 shows the state of the network after 8,000 timesteps. At this point, the error measure is becoming monotonically smaller as the lattice weights are now much more stable with respect to the saccades.

Figure 3 shows the state of the network after 16,000 timesteps. The total error has now reached a very small level of magnitude, so it would seem that all saccades now lead directly into the fovea. The actual effect of the saccades, given that a number of neighbouring neurons

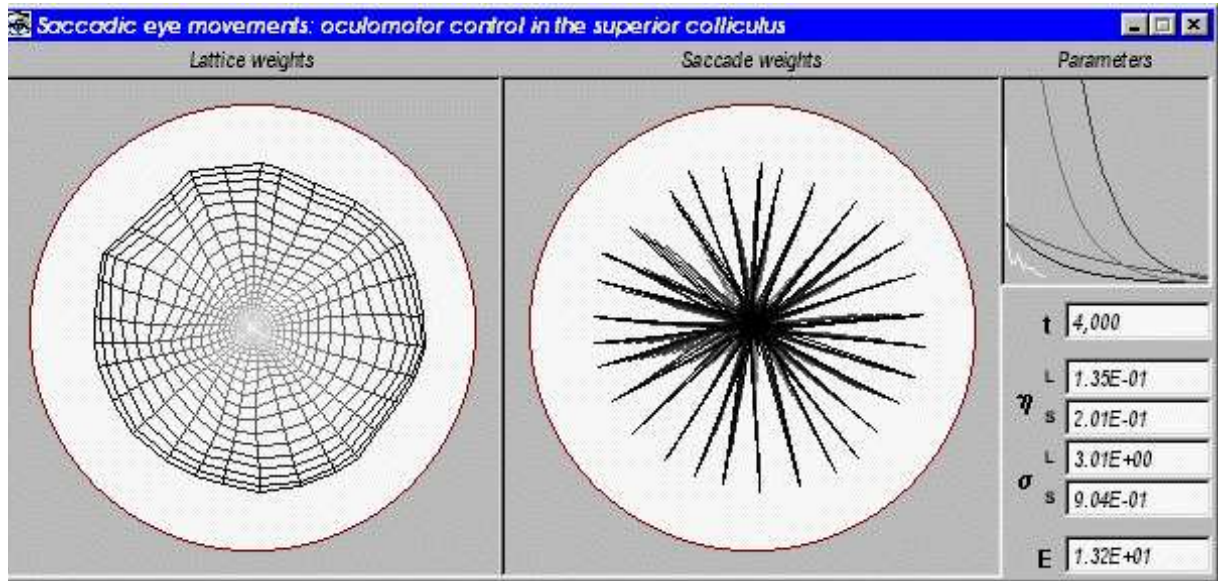


Figure 3: Lattice and saccade weight space after 4,000 stimulus presentations.

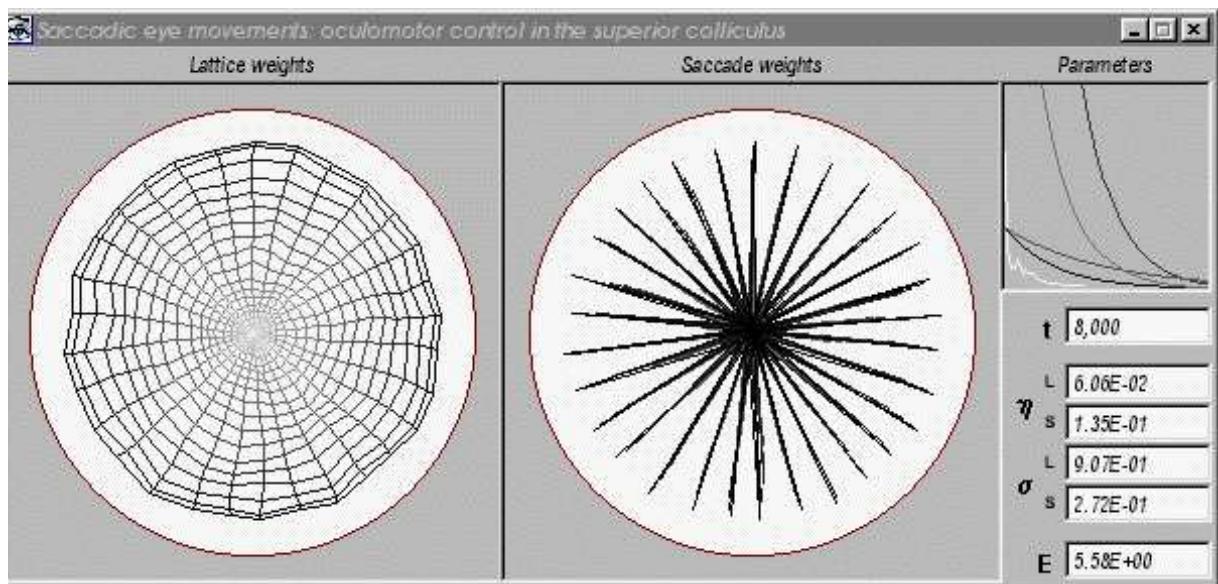


Figure 4: Lattice and saccade weight space after 8,000 stimulus presentations.

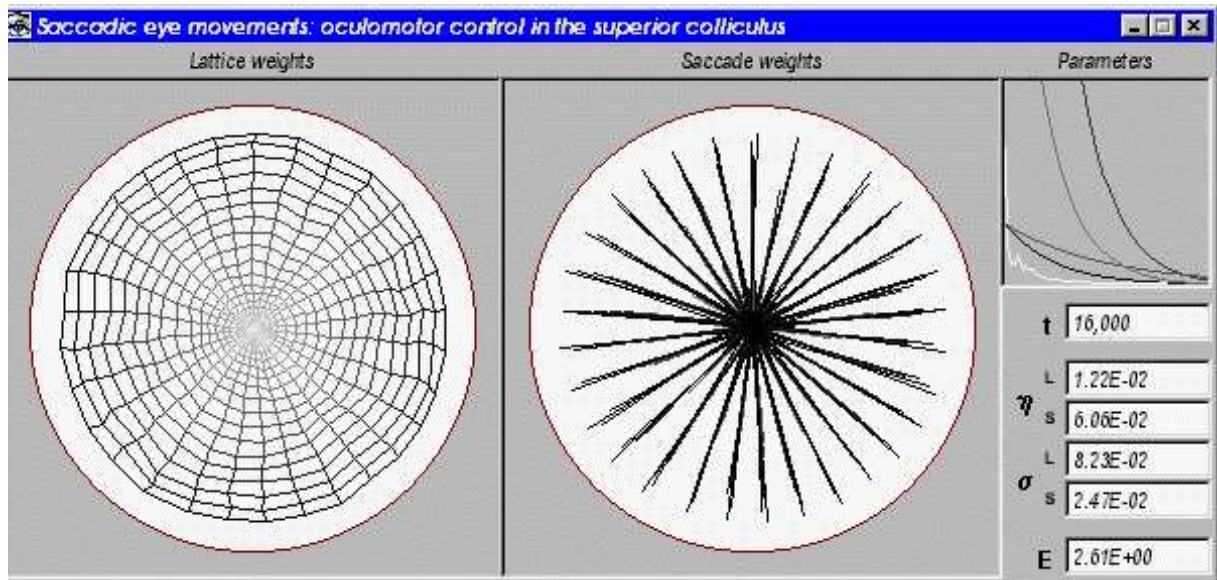


Figure 5: Lattice and saccade weight space after 16,000 stimulus presentations.

contribute to the overall effect, may still cause some saccades to miss the fovea proportional to the radial component of their location. This effect, given that the neighbourhood function h^S of the saccades is monotonically decreasing to 0 in the limit, will eventually be corrected. I will examine the implications of this effect, however, below.

4 Discussion

The effect of implementing a saccade strategy by which a number of neurons cooperate to determine the eventual position of the image leads to some interesting findings. The purpose of this model, as revealed above, is to endow the simpler saccade-learning model of Ritter et al[3] with this particular feature. However, the phenomenology of saccadic learning in humans, as documented by Becker and Fuchs[1], shows another interesting feature: in humans, saccades almost never lead directly into the fovea but usually close by, whereupon a corrective saccade takes over and fovealises the image. Ritter et al note this succinctly:

“Compared to these observations, our model learns its saccadic eye motions much too ‘well’ because at the end of the learning process all of our saccades precisely lead in the fovea.”

They further hypothesise that the consistent error in saccadic learning in humans may be related to the fact that humans are mobile agents that plan ahead to be able to track relatively mobile objects. This sounds quite plausible. On the other hand, let us examine the model above. When the neighbourhood function h^S of the saccade weights is still significant, any saccade still has an error proportional to the difference between the saccade weight of the centre of excitation in the lattice and the combined weights of other neurons in its vicinity. Due to the geometry of the lattice once the lattice weights have been learned, any particular saccade will tend to undershoot the centre of the visual field to some extent. This is in accordance with the observations made on human subjects by Becker and Fuchs[1].

Although this model has used a monotonically decreasing neighbourhood and learning rate for the saccades, let us now conceive of a model in which only the neighbourhood and learning rates for the lattice weights decrease, or in which the decay rates for h^S and η^S over time are much smaller than for the lattice weights. In such a scenario, the corrective power of the above model is much reduced, and the saccades will consistently undershoot the fovea by some small margin.

The other issue still left latent is the early freezing of the receptive fields in the lattice with respect to the saccades. The introduction of contextual information from neurons in the vicinity of the centre of excitation for a particular stimulus can be extremely important in the initial ordering of the saccades. If no contextual information is present, it would appear to be unjustified to introduce a learning rule for the saccades into the algorithm that updates neighbouring weights for a particular saccade, since those weights did not contribute to the saccade in the first place. Additionally, it might also be noted that in a physical implementation of the model, there is always the risk that some saccade weight fails to respond to learning commands, or interference from some external source affects the plasticity of the connection. In such a situation, if a group of neurons are responsible for the saccade, the network can learn to overcome the effect of the recalcitrant connection to some extent. This is not possible in Ritter et al's model.

In conclusion, we can see that this model of the oculomotor system offers numerous benefits in terms of neurobiological plausibility and graceful degradation, and hopefully serves to elucidate yet more clearly the issues surrounding simple visual reflexes in humans and other mammals. It is also yet another example of how a very simple learning principle, embodied in this model by the process of corrective saccades, can come to be a powerful organisational tool within a neural architecture.

References

- [1] Becker, W, & Fuchs, A F [1969] 'Further Properties of the Human Saccadic System: Eye Movements and Correction Saccades with and without Visual Fixation Points', *Vision Research* 9: 1247-1258
- [2] Korn, A [1982] *Bildverarbeitung durch das visuelle System. Fachberichte Messen, Steuern, Regeln* 8, Springer-Verlag
- [3] Ritter, H, Martinetz, T, & Shulten, K [1992] *Neural Computation and Self-Organizing Maps*, Addison-Wesley
- [4] Sparks, D L, & Nelson, J S [1987] 'Sensory and Motor Maps in the Mammalian Superior Colliculus', *Trends in Neuroscience* 10: 312-317