In: D. Heinke, G.W. Humphreys & A. Olsen (Eds), *Connectionist Models in Cognitive Neuroscience*. London: Springer, 1999.

Modelling Adaptation in the Human Oculomotor System

Patricia M. Riddell, John A. Bullinaria, Simon K. Rushton
Department of Psychology, The University of Reading
Reading, RG6 6AL

Abstract

In order to maintain clear, single binocular vision developmentally and in different visual environments, it is necessary for the oculomotor system to be plastic. During development, plasticity is required to compensate for changes in the size of the eyeball and the distance between the eyes. Adults also show plasticity when wearing optical devices (such as spectacle lenses and head mounted displays) which require local adaptation of the oculomotor system. It is therefore of interest to study the details of this plasticity since it will have consequences for the types of optical devices that can be used with comfort and safety. We have begun this study by simulating the human adaptive response to vertical disparities at different points in the visual field, and using comparisons with empirical data to explore the possible biological limitations of this adaptation process.

1 Introduction

Clear, single binocular vision of an object of interest requires the accommodation system to bring the object into focus and the vergence system to locate the object on the fovea of each eye. As an object approaches it requires progressively more accommodation and vergence. Since accommodation and vergence covary in a predictable manner, this invariant relationship becomes embodied in the visual system as cross-links between the accommodation and vergence systems – a response by one system will drive a response in the other.

Classically four components combine to give the state of vergence: "proximal" information ('pictorial' depth cues such as perspective, texture gradient, etc.); disparity (differences in position of the image on the retina of each eye); accommodation (the cross-link component); and a tonic component (an offset or bias). Accommodation may be conceptualised as having similar components. There are currently several systems models of accommodation and vergence in use (e.g. [3, 4, 9, 10, 13]). However, no one model has yet been universally accepted since there is still some dispute over the details, such as the positioning of the inter-connections between the accommodation and vergence systems (cf. [4, 13]).

Figures 1 shows a typical systems model. The main rival models differ from this in various ways. For instance, tonic components (slow integrators) for both accommodation and vergence have been included in the Schor & Kotulak model [13], whereas the Judge model [4] has a tonic component only for the vergence system. Additionally, there are differences between the models in the positioning of

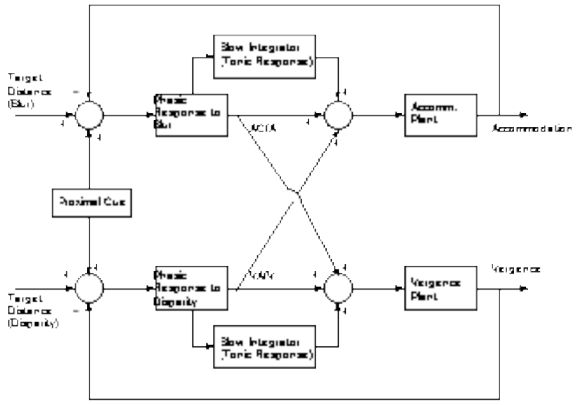

Figure 1: A typical systems model of the accommodation and vergence systems.

the cross-links between accommodation and vergence. In the Schor & Kotulak model, the contribution from the cross-links is added before the tonic components, allowing e.g. the vergence tonic component to be altered by an accommodative signal. In contrast, in the Judge model, the cross links are added after the slow integrator so that accommodation cannot affect the tonic vergence component.

As can be seen, there are many outstanding issues when considering how the accommodation and vergence systems interact. One of our major interests is in how this system develops and how plasticity is used developmentally, and maintained in the adult system. Vergence and accommodation are known to be plastic during early development since it is necessary to adapt to the changes in output required as the visual system grows. When this plasticity ends (if it ever does) is less sure. Most adults can adapt to new viewing situations throughout life. This is reflected in the manufacture of head mounted displays (HMD) for virtual reality (VR) equipment in which vergence cues are changed to give the appearance of a three-dimensional world on a fixed distance screen. As yet, there is no information to determine the effects of frequent use of such systems on the more plastic vergence and accommodation systems of children. Collecting data for this poses ethical problems since it would be unwise to expose children to differing visual environments if this could disrupt the normal development of their visual systems. Our answer to this is to formulate realistic neural network models of this system and to use the models to learn about the effects of frequent VR viewing on the developing system.

One relatively simple starting point for constraining such models is to look at how the adult visual system adapts to a situation in which the vergence cue is altered independently of accommodation. An example of this would be to place targets with different retinal disparities at one or more points in the visual field and to measure how viewers adapt to these targets. The initial response to such a stimulus is for the vergence system to signal an error which drives the *phasic* vergence system to alter the rotation of the eyes so that the targets become fused. If no adaptation took place, this system would respond equivalently every time the eyes moved, from a point in the visual field with no retinal disparity, back to the targets with retinal disparity. The tonic system will, however, attempt to adapt so that the phasic system does not have

to produce a response each time the targets are fixated. In this way, the visuomotor system can respond to changes in the visual field without visual feedback. Visuomotor cues or extraretinal cues are sufficient to produce these adaptive responses. The adaptation can take several forms. The system could try to adapt globally so that each point in the visual field is treated equally. This is what happens after continuous nearwork – both the tonic vergence and accommodation systems take up some of the effort required to focus on the near position, leaving the phasic system free to respond to greater local challenges. However, when localised targets are presented, which require a response from only the vergence system, a global solution is not sufficient. A solution is required which is local (i.e. limited only to the point in space at which the disparity is present) and therefore takes into account the direction of gaze within the visual field. Thus, the tonic system responds only when the eyes are pointing at the retinal disparity targets. This allows the visual system to produce a suitable vergence position for the disparity targets without compromising the vergence position at other points in the visual field. One problem remains within this solution: the tonic vergence system is no longer available for other local adaptations. To see that this problem is solved, consider the case of someone who is prescribed spectacles for the first time. When they are first given glasses, wearing them produces some uncomfortable visual artefacts due to the increase in prismatic effect towards the edges of the lenses. The world appears to move when it is obvious that it should be still. This effect wears off with use until the spectacle wearer can switch between wearing glasses and no glasses without discomfort. This suggests that there is a higher level of adaptation which goes beyond that accomplished by the tonic system. The visuomotor system programs a more permanent solution which will last, for example, until the spectacle wearer has to be given new glasses and the situation changes again. The advantage of this solution is that the tonic system is freed to respond in the way it would if no glasses were being worn.

Since these levels of adaptation are relevant to our interest in development and HMD use, and provide strong modelling constraints, the response to localised disparity was a convenient problem with which to begin our development of realistic neural network models of the plasticity in the visuomotor system. There is also good, empirical data available which shows how subjects adapt when exposed to targets which vary in their vertical disparity at set positions in the visual field [8, 9, 12, 14] and a first attempt has already been made at modelling this problem with limited success [9]. Our aim is to extend this model to account for all the available data describing adaptation to local disparity targets and gain a better understanding of the underlying biological processes.

2 The Experimental Data

The starting point for our models was data from Schor *et al*. [12] describing the spatial aspects of adaptation, and Maxwell & Schor's data [8] on the temporal aspects of two point adaptation. In both of these studies, subjects were presented with a stimulus which contained two local disparity targets with equal but opposite disparity (Figure 2). Subjects viewed one target for 10 seconds, and then switched to the second target which was again displayed for ten seconds. This procedure was

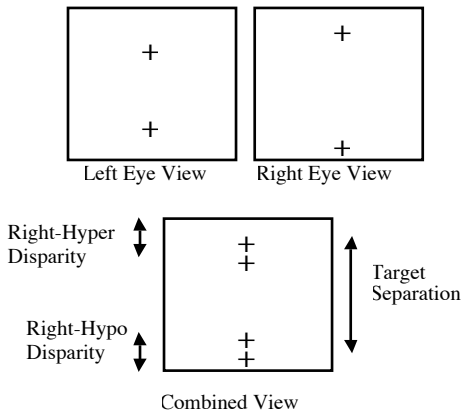

Figure 2: The experimental vertical disparity stimuli.

repeated for 40 minutes until adaptation was thought to have reached a steady state. Data was collected for a range of different disparities at a number of different target separations. The degree of adaptation in each case was assessed by measuring changes in vergence position when there was no fusion target present (dissociated phoria). This measure should change as tonic adaptation takes place. By measuring the phoria at different positions in the visual field, it was possible to determine the extent to which adaptation is occurring locally or globally.

The time course of adaptation can also be assessed by measuring phoria at different times during the adaptation process. The first adaptive change that occurs is a global shift in the offset of the phoria towards the vergence position appropriate for the target in the lower visual field [8]. This global change is followed by a more local change as the phoria adapts to the vergence position at each of the target positions. Thus, empirical measurements suggest that the time course of adaptation is faster for global than for local responses, and that there is a preference to adapt to the target in the lower visual field.

Figure 3 shows the empirical data, taken from Schor *et al.* [12], for stimuli of different spatial configurations. The data show that the size of the response is dependent on both the target separation and the size of the disparity gradient between targets. As the stimulus separation is increased at any one disparity gradient, the response gain increases, and as the disparity gradient of the targets is increased for a given separation, the response gain decreases. This suggests that a local solution becomes more difficult as the changes that have to be made are at increasingly smaller separations in the visual field, or as the size of the disparity stimulus required is increased. One of the main aims of our modelling is to understand what biological processes might be at the root of these difficulties.

3 Previous Models

Previous attempts to simulate adaptation have used “spatial spread models” (Henson & Dharamshi [2]) in which the phoria adaptation spreads locally with a Gaussian

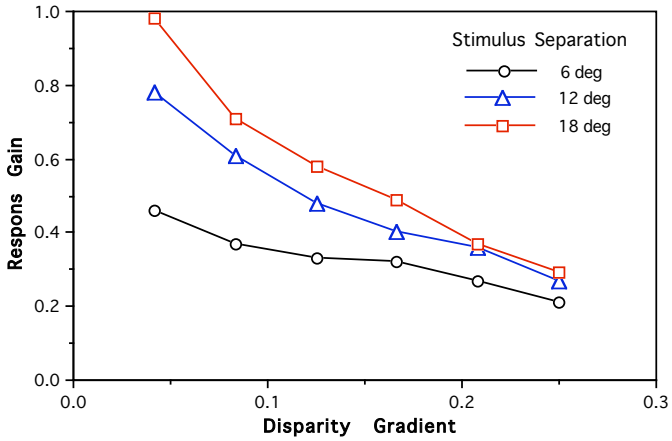

Figure 3: The experimental pattern of response gains (from Schor *et al*. [12]).

decay. As the disparity increases, the height of the Gaussian function increases. If more than one target is present in the visual field, the local response at points between the targets is given by the sum of the adaptive fields for each target at that particular point in space. This model produces the required increase in gain with increasing spatial separation of the targets, but the changes are larger than found empirically. Moreover, this model not only fails to produce the decrease in gain with increasing disparity, but actually shows an increase in gain with increasing disparity. These models also suffer from having rather little biological basis.

In an attempt to produce a more realistic simulation, Schor *et al*. [12] added two further parameters: In order to limit the simulated decrease in gain with decreasing spatial separation, the height of the Gaussian function was reduced proportionally with increasing stimulus separation. The decrease in gain with increasing disparity was modelled using a "disparity gradient scaling factor" which was determined from the empirical data. Adding these components resulted in a simulation which gave appropriate fits to the data. This confirmed that the spatial spread model of Henson & Dharamshi [2] was insufficient on its own to account for local adaptive changes in the vergence system. However, it was still unclear how the Schor *et al*. model [12] could be implemented physiologically.

More recently, McCandless & Schor have attempted to provide a more solid biological basis for these models [9]. They presented neural network simulations that demonstrated how physiological outputs from brainstem neurons sensitive to eye position could be associated with a vergence response to produce appropriate adaptation. As above, the stimuli consisted of two (or three) local disparities placed in the upper and lower visual fields, so that local adaptation depended on knowing the eye position for each disparity target. In effect, all the model had to do was to develop associations between the available representations of eye position and the motor control signal required to produce the appropriate disparity. Nothing more complicated than a simple feed-forward network with direct connections between the inputs and outputs is required to perform such a simple association. It is well known that neurons near the interstitial nucleus of Cajal respond almost linearly over a wide

range of eye positions [5, 6], and so representations of these were taken as the inputs of their network. In fact, they took 200 input units with a similar distribution of eye position responses to the real neurons. For their output they took a single linear unit to represent the motor output of the vertical vergence system. This reflects the differences in the known physiology of these systems. They then used a standard gradient descent learning algorithm to learn the required association by updating the connection weights to reduce the error on the stimulus for each eye position. Since this model is capable of learning to produce an error of zero in all cases, and since the empirical data show that subjects cannot fully adapt to some stimuli, training was stopped when the computed adaptation matched the observed adaptation. This was justified on the grounds that subjects are known to adapt to different stimuli at different rates, for example, they are faster for small disparities than for larger ones. However, the training times were chosen on the basis of the required response gains rather than the empirical adaptation rates.

As with human subjects, although the model is trained on only two stimuli locations, i.e. the two input positions, the non-zero output adaptation extends to other locations. In effect, the network is generalizing to produce a non-local response. The model was tested by comparing the vertical vergence responses at un-trained eye positions with the corresponding observed human responses. The simulation demonstrated that the output from eye position sensitive brainstem neurons could indeed be used to drive appropriate vertical vergence changes. However, the simulation results deviated from the observed data in that the model response was linear unlike the empirical data which showed a saturation in phoria as eye position moved away from the adaptation points (e.g. Figure 2 of McCandless & Schor [9]). It was suggested that a sigmoidal output function could be used to deal with this difference, but the details were not presented. In summary, McCandless & Schor [9] have begun to provide a biological basis for connectionist models of oculomotor adaptation, but we feel that there is room for improvement and refinement.

4 The New Models

Our aim now is to formulate biologically plausible neural network models which simulate the empirical data for adaptation in a more accurate manner, and check the extent to which specific aspects of those models are crucial to account for the data. Specifically, we want to model the time course of adaptation more completely, and to account for the detailed dependence of the adaptation response gain on both the spatial separation of the stimuli and the magnitude of the target disparity.

The simplest first attempt at modelling is usually to see how well a standard back-propagation network with one hidden layer maps between the required inputs and outputs. This is actually a very simple problem – we have just one input unit to represent the eye position, one output unit to represent the corresponding disparity, and a layer of hidden units in between them. Then for each two point adaptation network we have just two training patterns (e.g. –6, –1.2 and +6, +1.2, for which the stimulus separation is 12 and the disparity gradient is 0.2), but test the response of the trained network over the whole range of eye positions. The idea is to train the same network architecture on a number of different training sets, corresponding to a

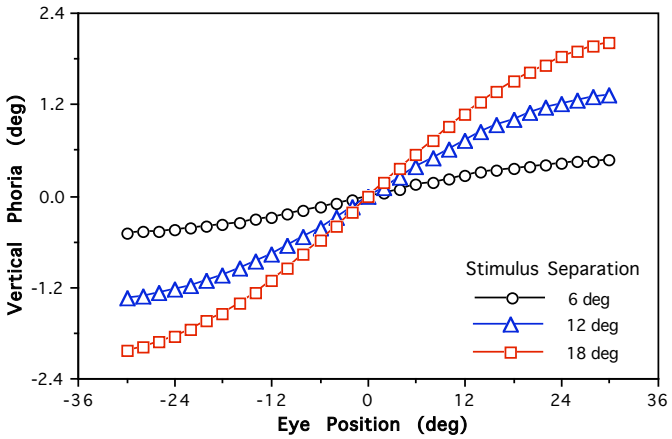

Figure 4: The output response of a typical network model based on sigmoids.

range of stimulus separations and disparity gradients, and test the pattern of response gains against the experimental data in Figure 3. The kind of output activation patterns we can get for three such trained networks are shown in Figure 4. The problem we find, as was found with the model of McCandless & Schor [9], is that unless we introduce something to restrict the network learning, the network is powerful enough to have no problem in reproducing the training data to arbitrary accuracy. In other words, unlike human subjects, our networks always achieve response gains of 1, independent of the stimulus separation or disparity gradient. Furthermore, a check on the internal representations learnt on the network's hidden layer reveals that very little structure was actually required there.

As is conventional, we started the training process for each network with a set of random initial weights. Normally a network will learn appropriate connection weights between the input and hidden layers to form useful internal representations on the hidden layer. In our case, however, the task was so simple that the learning process actually resulted in very little change of the initial weights into the hidden layers. In particular, if we chose the initial random weights so that the initial pattern of activations on the hidden layer mimicked those position dependent neurons near the interstitial nucleus of Cajal [6] that we discussed above, then they still mimicked those cells *after* the training process. In effect, we gained nothing by having this extra layer of learnable weights – we got exactly the same results as we would by just learning to map between the simulated position dependent neurons and the required outputs. We were thus led back to studying the "no hidden layer" networks of McCandless & Schor [9] where the existing biologically plausible internal representation of eye position is hard wired onto the inputs. Given that the adaptation process is simply an association between the required outputs and the existing representation of the state requiring that output, this makes good sense. The network simplification also has the advantage of rendering feasible a much more systematic investigation of all the possible relevant factors.

Our next set of models thus consisted of a set of input units on which eye position was coded, connected to a single linear output unit that represents the vergence

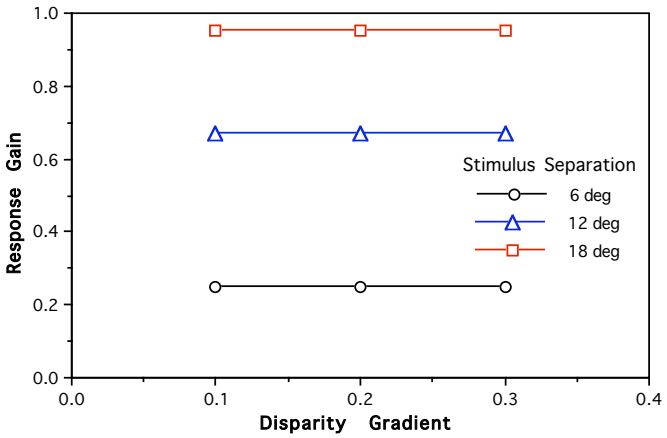

Figure 5: The pattern of response gains for our model with limited epochs.

response. As discussed above, McCandless & Schor [9] have already shown that taking the input coding to be linear, corresponding to physiologically realistic Cajal cells [6], works reasonably well. However, this cannot reliably inform us about the underlying biological processes unless we also test whether other, equally biologically plausible, codings can work just as well, or even better. We therefore also investigated networks with alternative input codings, specifically Sigmoidal (which can also be a good approximation to linear over the range in question), Gaussian and Difference of Gaussians (which might correspond to the topographic maps common in the visual system). Each Sigmoid, Gaussian, or other function, is defined by a set of parameters (e.g. height, width, centre) and a random set of these (within reasonable ranges) were assigned to each input unit to give its activation for each eye position. Typically, between 80 and 800 input units were sufficient to avoid randomization artefacts. Following McCandless & Schor [9], the networks were trained by fixed step-size gradient descent on a sum squared error measure; first on zero disparity to provide a steady pre-adaptive state, and then on our two point disparity stimuli. Naturally, this again resulted in perfect adaptation, i.e. a response gain of 1, in every case. Clearly something must be stopping this perfect adaptation in the human subjects. Again following McCandless & Schor [9], we attempted to simulate this by restricting the number of training epochs.

Figure 4 shows typical patterns of output activation for our network with Sigmoidal input coding after 7000 epochs with a learning rate of 0.00001 and a stimulus disparity gradient of 0.1. To relate these to the experimental results, the response gradients are defined using the change in output between eye positions of ±6 degrees and the response gains are these as a fraction of the stimulus disparity gradient. Repeating this for each stimulus set, Figure 5 allows a direct comparison of our network's pattern of results with the experimental data shown in Figure 3. We see that the response gains are now less than one, as required, and moreover, they now depend on the stimulus separation in the right way. The effect of the stimulus disparity gradient, however, cancels out in the gain calculation. Using the other input functions (Linear, Gaussian, Difference of Gaussians) changes the form of the

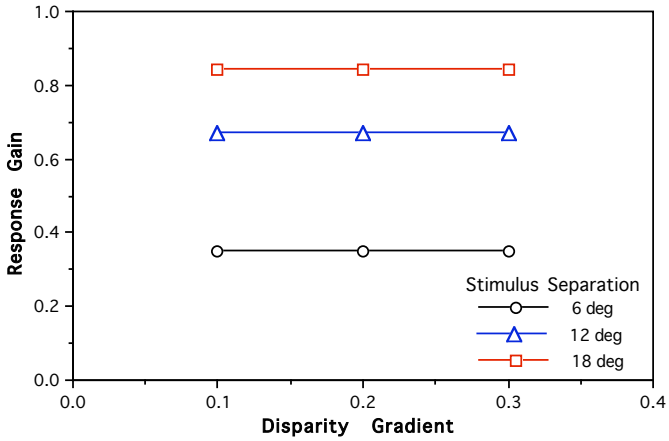

Figure 6: The pattern of response gains for our model with quadratic weight cost.

network's extrapolation to larger eye positions, but has little effect on the pattern of response gains. We see that a simple restriction on the amount of training does not fully account for the data. Clearly, by picking an appropriate number of training epochs for each training set, we will be able to get any pattern of gains we want. Before following that route, however, it makes sense to consider what other, more biological, processes might account for the experimental data.

5 Weight Decay

Another method commonly used to prevent networks from learning their training data exactly is to introduce weight decay into their training algorithm (e.g. as discussed by Bishop [1, p338]). Weight decay, like other forms of regularization, provides a natural procedure for smoothing adaptation to potentially noisy stimuli, and limiting possibly detrimental local changes to a control system which also has to act globally. Poggio *et al.* [11] have reviewed the use of regularization principles in various areas of computational vision, and have outlined how they could be implemented in single biological neurons or small networks of neurons.

Normally, gradient descent learning operates by iteratively adjusting the network weights w to reduce some output activation error function E (e.g. the sum of squared output activation errors). If one adds to this error function a weight cost function, then this results in an extra contribution to the gradient descent weight changes $\Delta w \sim \partial E/\partial w$ proportional to the derivative of the weight cost. If the cost function is quadratic (i.e. the sum of squared weights), we thus get a linear weight decay contribution to the weight changes. This keeps the weights low and smoothes out the network response at the expense of the higher output activation errors. Using such a standard quadratic weight cost in our basic model with Sigmoidal input coding results in the asymptotic pattern of response gains shown in Figure 6. All the gains are now less than 1, and we get the experimentally observed effect of stimulus separation, but we still do not obtain the required effect of disparity gradient. Enforcing a natural weight decay during training has the same result on the network

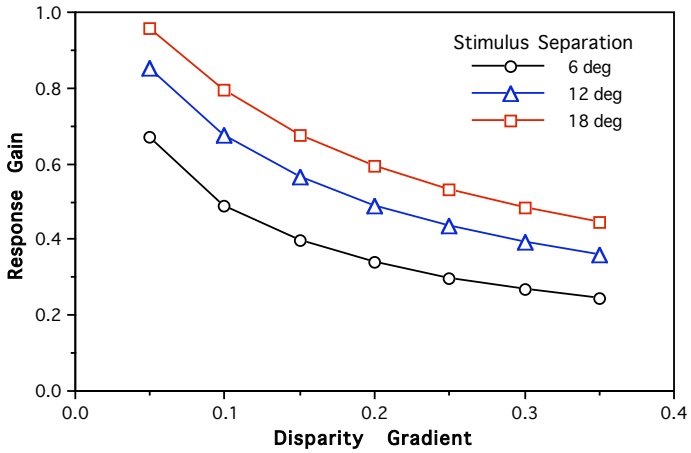

Figure 7: The pattern of response gains for our model with quartic weight cost.

outputs as the reduction in the weights due to stopping the training early. Neither approach accounts for the observed gains after unlimited adaptation time.

One advantage of connectionist models is that we are able to analyze their processing abilities at a much finer level than real biological systems. In particular, we can easily investigate the connection strengths associated with each individual input unit for the same network trained on different stimuli. In this way we were able to confirm our intuition that adaptation to large disparity gradients requires larger weights than small disparity gradients, and understand that this was not having the expected disparity gradient effect on the response gains because the trade-off between the quadratic error term and the quadratic weight cost was leaving no net effect. Once we understood this, it was clear that a higher order weight cost should be able to dominate the quadratic error term and have the desired effect. A quartic weight cost (i.e. the sum of fourth powers of the weights), for example, will result in a stronger cubic weight decay which will preferentially restrict the learning of the large disparity gradients. Using this in our standard model, the outputs again take the form of those in Figure 4, but now the pattern of response gains, shown in Figure 7, is much closer to the experimental pattern of Figure 3.

We have thus shown that a biologically plausible weight decay during adaptation in a simple network that learns to associate between biologically plausible representations of eye position and vergence control, provides a remarkably good fit to the experimental data. Moreover, we have been able to avoid using any of the *ad hoc* training time restrictions or scaling factors that have been used in the past. Unfortunately, the experimental data is too noisy to constrain our model much further at this stage.

6 Future Research

One aspect of the data we have not yet addressed is the differential time courses of the three mechanisms thought to be responsible for oculomotor adaptation. The problem of modelling multiple mechanisms with different time courses is well

known to researchers working on memory, where the different time courses of short and long term memory processes require special treatment. Levy & Bairaktaris [7] provide a review of several solutions suggested for this problem. In particular, dual-component or dual-weight systems have been implemented and shown to be effective in modelling interactions between short term and long term processes. We are currently investigating the possibility of using such architectures to model the interactions between phasic and tonic components of oculomotor control.

We have had insufficient room to discuss the details here, but we have been able to demonstrate that the known hemifield differences in visual response [8] can be accommodated in our models, either by varying the training data frequencies in line with attention, or by manipulating the parameters describing the distribution of eye position representations. Further study of these effects is underway.

Another avenue for future research is to look again at adjusting the training times. It was noted by McCandless & Schor [9] that their subjects reported that whilst small disparities were easily fused, the larger disparities often remained diplopic. This means that subjects were not able to produce the required eye rotations to foveate the targets in each eye. This would result in there being be less – or even no – training time with the eyes in the correct position. Although McCandless & Schor [9] used this justification to fix their network training times, they did not adjust the training times in line with empirical evidence. We would like to use direct empirical data to determine appropriate training times for the different target configurations and see if this really can provide an alternative account of the empirical data

Yet another direction for future research is to investigate how the use of different learning algorithms can affect the results. We intend to implement models with more biologically plausible algorithms to determine whether these can further improve the realism of the model. We also intend to extend the model to determine its efficacy in predicting the data for more complex target configurations. Eventually, of course, all this is to be extended to provide a complete connectionist account of the full accommodation and vergence system represented in Figure 1.

7 Conclusions

In this paper, we have presented an investigation into possible improvements of the physiologically plausible neural network model of oculomotor adaptation of McCandless & Schor [9]. We have argued that, rather than imposing limits on the network training times, a better account of the empirical data can be made by the incorporation of weight decay into the learning process to prevent total adaptation. We feel this provides a more accurate reflection of the adaptation process since it reflects the compromise between local and global solutions and does not depend on assumptions about available training times. However, we intend to study the effect of incorporating empirically measured variations in training times to see just how well this approach could do. We have also shown that the details of the biologically inspired eye position coding are not crucial to the models' performance, and that the observed hemifield difference may arise in a physiologically plausible manner.

The motivation behind our modelling work is to understand the biological mechanisms involved in local adaptation and to gain insight into the plasticity of the

oculomotor system both in adults and during development. In this way, we can hope to study the development of oculomotor disorders and to resolve, ethically, issues of safety when altering visual environments (e.g. by the use of virtual reality headsets) both in adults and in children with their more plastic visual systems.

Acknowledgement

This work is funded by the EPSRC Research Grant GR/L82274.

References

1. Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*. Oxford, UK: Oxford University Press.
2. Henson, D.B. & Dharamshi, B.G. (1982) Oculomotor adaptation to induced heterophoria and anisometropia. *Investigative Ophthalmology and Visual Science*, **22**, 234-240.
3. Hung, G., Semmlow, J, & Ciuffreda, K.J. (1982) Accommodative oscillation can enhance average accommodation. *IEEE Transactions*, **SMC-**12, 594-598.
4. Judge, S.J. (1996) How is binocularity maintained during convergence and divergence? *Eye*, **10**, 172-176.
5. Judge, S.J. & Cumming, B.G. (1986) Neurons in the monkey midbrain with activity related to vergence eye movement and accommodation. *Journal of Neurophysiology*, **55**, 915-930.
6. King, M.W., Fuchs, A.F. & Magnin, M. (1981) Vertical eye movement-related responses of neurons in midbrain near interstitial nucleus of Cajal. *Journal of Neurophysiology*, **46**, 549-562.
7. Levy, J.P. & Bairaktaris, D. (1995) Interactions between short- and long-term weights: applications for cognitive modelling. In: Levy, J.P., Bairaktaris, D., Bullinaria, J.A. & Cairns, P. (Eds) *Connectionist Models of Memory and Language*. London, UK: UCL Press.
8. Maxwell, J.S. & Schor, C.M. (1994) Mechanisms of vertical phoria adaptation revealed by time-course and two dimensional spatiotopic maps. *Vision Research*, **34**, 241-251.
9. McCandless, J.W. & Schor, C.M. (1997) A neural net model of the adaptation of binocular vertical eye alignment. *Network*, **8,** 55-70.
10. Patel, S.S., Ogmen, H., White, J.M. & Jiang, B.C. (1997) Neural network model of short-term horizontal disparity vergence dynamics. *Vision Research*, **37**, 1383-1400.
11. Poggio, T., Torre, V. & Koch, C. (1985). Computational Vision and Regularization Theory. *Nature*, **317**, 314-319.
12. Schor, C., Gleason, G., Maxwell, J. & Lunn, R. (1993) Spatial aspects of vertical phoria adaptation. *Vision Research*, **33**, 73-84.
13. Schor, C.M. & Kotulak, J.C. (1986) Dynamic interactions between accommodation and convergence are velocity sensitive. *Vision Research*, **26**, 927-942.
14. Schor, C.M. & McCandless, J.W. (1997) Context-specific adaptation of vertical vergence to correlates of eye position. *Vision Research*, **37**, 1929-1937.