



# A model of visual–spatial memory across saccades

Jude Mitchell \*, David Zipser

*Cognitive Science, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0515, USA*

Received 23 June 1999; received in revised form 20 September 2000

---

## Abstract

This paper describes a neural network model that directs saccades back to targets after they disappear and other saccades intervene. This is a simple example of knowing where something is after it is no longer visible and the observer has moved. These tasks require a short-term memory that can store continuous values of spatial location. The model was generated by training a neural network with a recurrently connected hidden layer to specify memory-guided saccades. The trained network maintains stored locations accurately for a few seconds. It uses a leaky integrator mechanism in which there is a slow decay of the stored value to a small number of fixed point attractors. Similar mechanisms have been used to model oculomotor integration (Cannon, S., Robinson, D., & Shamma, S. (1983). A proposed neural network for the integrator of the oculomotor system. *Biological Cybernetics*, 49, 127–136; Seung, H. (1998). Continuous attractors and oculomotor control. *Neural Networks*, 11, 1253–1258). The mechanism is robust to parameters such as the input and output format and the constraints in training. However, the receptive field properties of the hidden units do depend on these parameters. It was possible to find biologically plausible parameters that produced hidden unit behavior similar to that of real neurons involved in saccade memory. In particular, training the model to simultaneously represent the target location in both eye- and head-based reference frames produces units similar to neurons in parietal saccade areas. © 2001 Elsevier Science Ltd. All rights reserved.

*Keywords:* Neural network; Double-step paradigm; Saccade; Working memory; Parietal cortex

---

## 1. Introduction

Spatial working memory stores information that remains valid when observers move and objects disappear from view. The full problem of spatial working memory is very complex, but some of its important features can be studied using simple tasks. Here we model a task requiring the return of gaze to an out of sight target after saccades to other secondary locations intervene. The model addresses the issues of whether eye- or head-based reference frames represent the target location, the mechanism that remembers locations, and the way these locations are updated to correct for intervening eye movements. It demonstrates that the required computations can be distributed among all the units in a recurrent network. Individual units participate in both the eye- and head-based representations of the target location. The close similarity between the behav-

ior of model units and real neurons is consistent with the brain using mechanisms like those of the model.

The computations needed to accomplish the multiple saccade task are quite simple. This makes it straightforward to design a neural network that can implement the task. However, it is very difficult to explicitly design a network that can actually account for the experimental observations in detail. This is because the available data indicate that the brain may use several computational strategies in a complex, distributed way (Sparks, 1989).

Neural network models with realistically distributed computations can be generated using optimization rather than explicit design. This technique requires specifying the temporal sequence of inputs and outputs for the task rather than the detailed connectivity of the network. The optimization, or ‘learning’, procedure finds connection weights that configure a network to implement the task. Analysis of these models shows that, while they often use the same basic computational strategies as explicitly designed networks, the computations are distributed in a realistic way that closely

---

\* Corresponding author. Tel.: +1-858-5344135; fax: +1-858-5341128.

*E-mail address:* jfmitche@cogsci.ucsd.edu (J. Mitchell).

approximates experimental data. This approach has proven useful in accounting for the often obscure behaviors exhibited by neurons in distributed dynamical systems. The optimization paradigm is called ‘neural system identification’ (Zipser, 1992). It was used to configure the model described here. With neural system identification, finding good input and output representations becomes critical for getting models that fit the data.

The input representation we use was chosen to supply the information known to be required for the task. Its format is a rough approximation of what we believe the input to parietal areas is like.

The outputs were chosen to account for the existing experimental data. Both eye- and head-based strategies have been proposed for the multiple saccade task (Sparks, 1989). These strategies differ in the way corrections are made for intervening eye movements. For the eye-based strategy, target locations are stored dynamically. Eye-based memory has to be updated during each intervening saccade, but it can be used to saccade back to the target without reference to the current eye position (Moschovakis, 1987). For the head-based strategy, target locations are stored statically relative to the head. The remembered location does not need to be updated during intervening saccades, but the current eye position has to be used to correct it when a saccade is made back to the target (Robinson, 1975; Sparks & Mays, 1983a). These eye- and head-based correction mechanisms seem mutually exclusive. However, preliminary modeling showed that both mechanisms can be combined in a single network. This is done by including both reference frames in the output representation that the network is trained to produce. The mixed frame model, which we describe here, performed better than either eye- or head-based models alone, and gave more realistic results.

The details of how location is coded in the output determines the shape and size of model neuron receptive fields. We searched for and found codings, described below, for eye- and head-based frames that give a reasonably good fit to experimental observations.

It is known from previous optimization modeling of dynamic short-term memory that different memory mechanisms are generated by training with discrete or continuous values (Zipser, 1991; Moody, Wise, Pellegrino, & Zipser, 1998; Seung, 1998). We found that only training on an unlimited set of continuous valued locations generated models with neurons having realistic receptive fields.

## 2. Neural network model

The task we modeled begins with a briefly flashed target, followed by saccades to secondary, randomly

located, targets, and ending with a saccade back to the original target. The final saccade must take into account the change in eye position due to the intervening saccades. An experimentally studied version of this is the double-saccade task (Hallett & Lightstone, 1976; Mays & Sparks, 1980; Sparks & Mays, 1983b). The target space we used is the *frontal depth plane*. This differs from much of the experimental work which is in the fronto-parallel plane. The results, however, were much the same.

A typical trial of the task is presented in Fig. 1. At the start of each trial a target is illuminated at a randomly chosen location in the horizontal depth plane in front of a model observer. It is visible through the one-dimensional retinal arrays of a left and a right eye. After a brief interval, the target disappears and saccades are executed to randomly chosen intervening locations in complete darkness. Details of the task are given in Appendix A.

To accomplish the task, the network must do three computations — map from sensory inputs to representations of target location, store this information, and remap it at appropriate times.

The model architecture is shown in Fig. 2. It has an input, hidden, and output layer. The input layer consists of two one-dimensional retinas together with eye position and velocity channels. The output layer consists of two  $10 \times 10$  grids of units that give a distributed representation of the target’s location in an eye- and a head-based frame. Details of the input and the output encoding are provided in Appendix A.

The hidden layer consists of recurrently connected logistic units. It transforms the inputs into the desired output and, because of the recurrent connections, maintains an active memory of the location of the stimulus when visual input is absent. Equations for updating the activity of the units are given in Appendix A along with how the model is optimized to perform the task.

## 3. Performance of the model

The network was successful in learning the task. To assess its performance, the center of mass of activity in the output array was taken to be the target location (see Appendix A). Here in Fig. 3, the location output in eye- and head-based frames is shown over a typical task trial. The network had 80 hidden units. It updates the retinal direction and disparity during intervening saccades. It also maintains the head-based direction and distance while buffering them against changes during saccades.

The average spatial error over many trials is given in Table 1 for networks with different sized hidden layers. Each network is run through 5000 randomly generated task trials. The squared error between the output loca-

tion and the actual location is averaged over the duration of the trials. The root mean squared error (RMSE)

is given for each coordinate. As expected, error decreases as the size of the network increases.

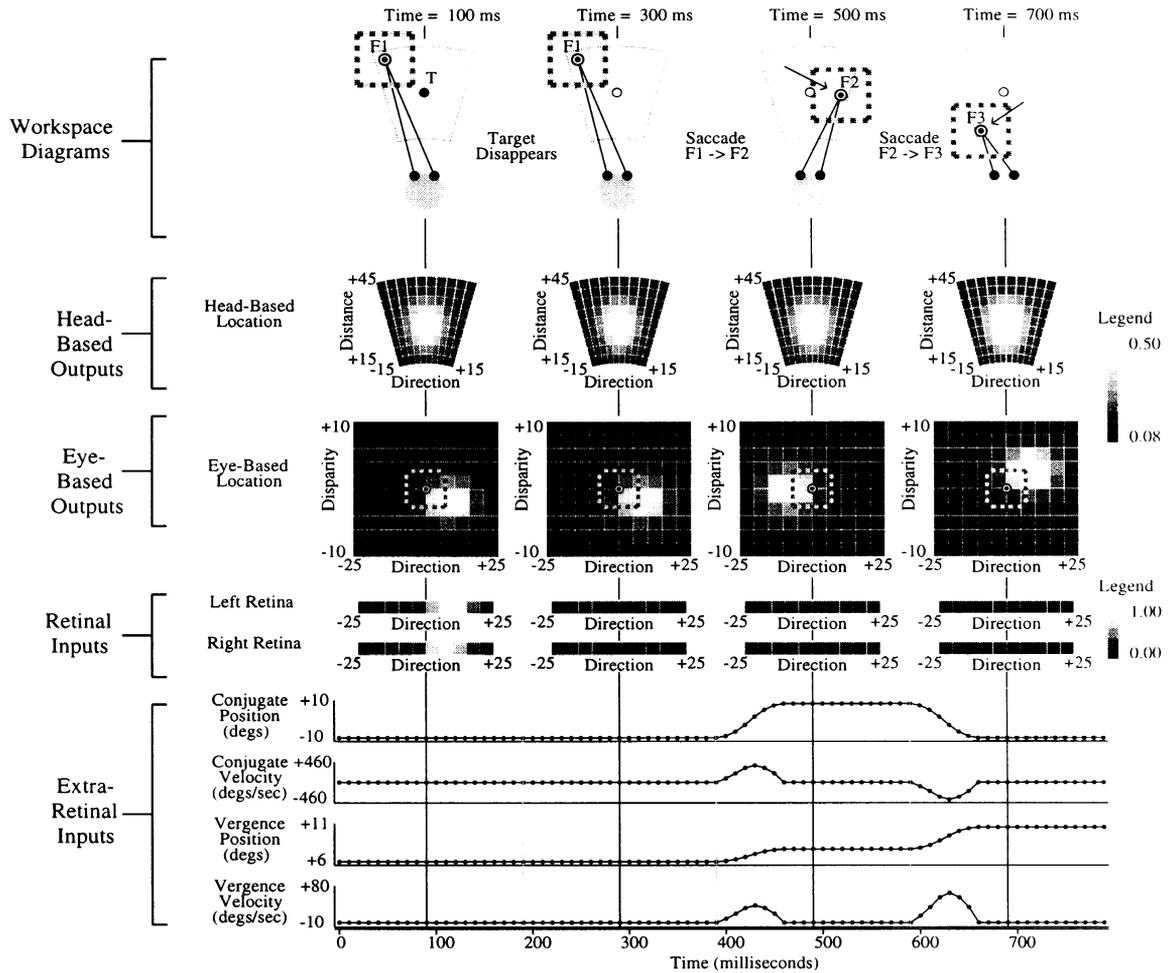


Fig. 1. Diagrams of the target and fixation locations in the task workspace and the corresponding inputs and outputs are presented during a typical trial. Snapshots of the workspace, the input, and the output arrays are taken at four points in time ( $t = 100, 300, 500,$  and  $700$  ms). The inputs and outputs are updated every 10 ms so they actually change in a more continuous fashion than indicated by these four snapshots. *Workspace diagrams:* A diagram of the fixation point and the saccade target is drawn at the four points in time. The trial starts with the eyes fixated on F1 and a visible target at T (left frame). The target disappears (2nd frame) and saccades are made to other secondary fixation points (3rd and 4th frames). The location of the target and fixation points, as well as the timing of events, is chosen at random for each trial. Details are given in Appendix A. *Head-based outputs:* The head-based output format represents the location of the memory target. It consists of a  $10 \times 10$  array of units. The array covers the range from  $-15$  to  $15^\circ$  in direction and  $15$ – $45$  cm in distance. The target location is represented by the center of a Gaussian bump of activity (legend on the far right). Since the target does not move, the location of activity remains stationary over the trial. *Eye-based outputs:* The eye-based output format also consists of a  $10 \times 10$  array of units. The array covers the range from  $-25$  to  $25^\circ$  in retinal direction and from  $-10$  to  $10^\circ$  in retinal disparity. The fixation point is the origin of this coordinate system. It is surrounded by the checkered box that was depicted in the workspace diagrams. Note that the range covered in the eye-based array is larger than the extent of the workspace, so the checkered box appears smaller. Target location is again represented by the location of Gaussian-shaped activity. In contrast to the head-based output, the location in the eye-based outputs is updated with each saccade (3rd and 4th frames). *Retinal inputs:* A left and right retinal array provide visual input to the network. Each retinal array consists of 10 units with Gaussian receptive fields. When a target is visible, the units near the point where light from the target hits the retina are the most active (left frame). After it disappears, both arrays have zero activity (other frames). *Extra-retinal inputs:* The position of the eyes is described by two angles. The first is the average angle of rotation of the eyes, called the conjugate angle, and the second is the difference between the angles of rotation, called the vergence. The position and the velocity of the two angles is shown in four plots. In each plot the value is given every 10 ms during the trial. Saccades at 400 and 600 ms are simulated as continuous movements with bell-shaped velocity profiles. Details are given in Appendix A. The velocity and position values are represented to the network by four input units. Each unit's activity linearly encodes one of the values.

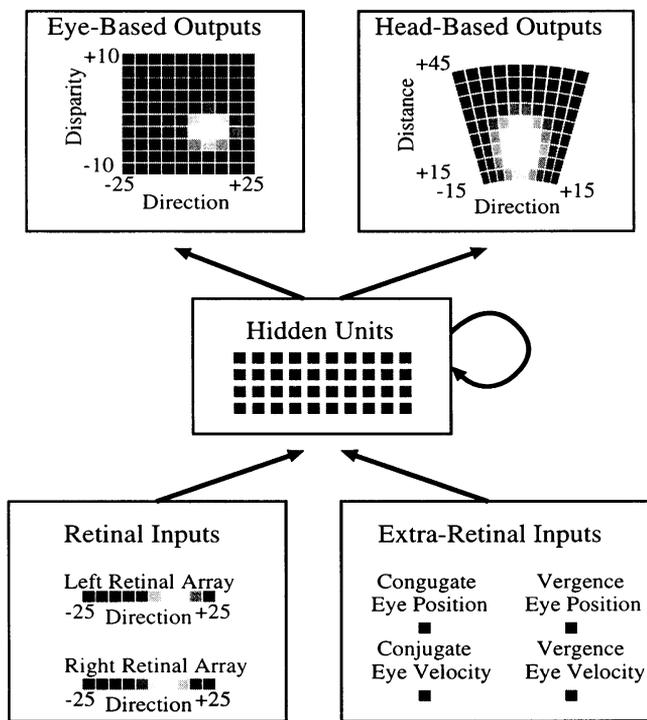


Fig. 2. The neural network architecture includes an input, hidden, and output layer. The input layer consists of the units from the left and right retinal arrays and the four extra-retinal units. Each unit in the hidden layer receives connections from all of the units in the input layer. Hidden units also receive inputs from the rest of the hidden layer through recurrent connections. The hidden layer is fully-connected to the output layer. Details for input and output representations are given in Appendix A.

#### 4. Memory decay

The memory mechanism employed by the neural network is of a leaky integrator type. It accurately maintains the stored value for the time interval trained, but ultimately that value decays to one or more fixed stable points (Zipser, 1991; Seung, 1998). The locations of the fixed points have little relation to the original location. Unlike most attractor neural networks, here the value is not stored by settling to a fixed point. Instead, it is stored by slowing the decay at every continuous location and staying away from the fixed points.

Memory for the location of an extinguished stimulus across saccades decays very slowly. The location of the original stimulus together with the head-based location output by a network with 80 hidden units is shown in Fig. 4. Each black circle reflects the location output by the network at intervals of 50 ms. Initially the memory is close to the target's location (black square) but slowly decays. After 20 s, the location decays to a fixed point attractor. In the example shown, it decays to an attractor near the center of the workspace losing all memory of the original location.

The rate of decay and trajectory of the remembered location are not much affected by intervening saccades. In Fig. 4, four trajectories (in grey) are shown that are produced from repeating the original trial but with randomly selected intervening saccades during the first 2 s. Each saccade introduces a slight perturbation to the remembered location. Also, since changing the position of the eyes alters the input to the network, it can also alter where the final resting points are located. As a result, the direction of decay changes slightly after each intervening saccade. Nonetheless, the decay continues at about the same rate. After 2 s, the eyes are returned to the central fixation point, and the decay continues to the same fixed point as in the original trial.

The network was analyzed to determine the rate of memory decay and the location and number of terminal attractors. Trials similar to that shown in Fig. 4 are repeated with the target being presented at different workspace locations. Within distinct sub-regions of the workspace, the remembered location decays to a unique fixed point. These regions around the fixed point are called its basins of attraction. They are shown in Fig. 5 for a network with 80 units. Within each basin the direction of decay during the first second of the trial is shown. The initial direction does not always point directly to the final resting location that it will eventually reach.

Memory for an extinguished stimulus improves with increasing network size. The number of attractors and the average initial rates of decay are given in Table 2 for different network sizes. Increasing the size does not necessarily increase the number of attractors. For example, a mixed-output network with 40 units has five fixed point attractors while one with 80 units has only two. What does change consistently is that the average rate of decay decreases.

#### 5. Remapping performance

Each time an intervening saccade occurs, the network updates the target's eye-based location. The update requires that the location shift by a vector that is the opposite of the saccade. If the size of the shift is expressed as a percentage of the saccade magnitude, then a shift of 100% is required. Also, if the direction is given relative that opposite of the saccade, then  $0^\circ$  is correct.

Networks are successful in remapping eye-based locations for saccades. In Table 3 the average percentage shifts and average shift directions are given for different sized networks. Averages are computed over the saccades that occurred during 5000 random task trials. Other than the smallest version of the network, there is not much improvement with increasing network size. What does improve is the accuracy of the spatial infor-

mation stored (as reflected in Table 1) and the rate of its decay (as reflected in Table 2).

### 6. Comparison to physiological data

The response properties of the hidden units are compared to those of real neurons to see the degree to which the model accounts for experimental data. The data from the model differs from experimental data in that the target locations tested lie in the *frontal depth plane*. Most experimental work has been in the fronto-parallel plane. Despite the change in coordinates, hidden units exhibit similar response properties.

The preferred saccade direction of each unit remains nearly the same during the visual and memory periods of its response. Units have response curves with a single peak at the preferred direction (Fig. 6B). The peaks during the visual and the memory periods are typically aligned, unless one of the responses is weak (Fig. 6C). Of those units with strong responses in both periods, most prefer either the same direction or an adjacent direction (Fig. 6D). The average difference in direction is 16°. Similar results are found among cells in LIP and area 7a with the median difference in preferred direction of 12° (Barash, Bracewell, Fogassi, & Andersen, 1989).

The magnitude of the directional response is modulated by gain fields. The gain fields are described well as a planar function of the position of the eyes in their orbits (see Appendix A). None of the units had local or peaked gain fields. The effect of modulation can be

quite large. A typical unit is shown in Fig. 7B. Its response is largest for eye positions towards the top right of the workspace. Most of the units have some modulation due to fixation. Across the population, the average response varies by 38% of its peak value over the fixation positions tested. Comparable effects are observed in parietal areas (Andersen, Bracewell, Barash, Gnadt, & Fogassi, 1990). Another property shared by the units and parietal cells is that the gradient direction of their gain fields is typically aligned during the visual and memory response periods (Fig. 7D).

The preferred saccade direction remains nearly the same at different fixation positions. The tuning at different fixations is shown for a typical unit (Fig. 8B). The most common effect from eye position is simply to modulate the magnitude to the directional response. This is consistent with what is found in LIP and area 7a (Andersen et al., 1990).

A small subset of units do change their preferred direction with the fixation position (Fig. 8C). Later analysis shows these units are better described as having a receptive field sensitive to the head-based location. Similar cells have not been identified in area 7a or LIP (Andersen et al., 1990). We address this discrepancy between the data and the model in the discussion.

### 7. Analysis of memory strategies

The neural network model uses a mixture of eye- and head-based strategies to solve the memory-guided sac-

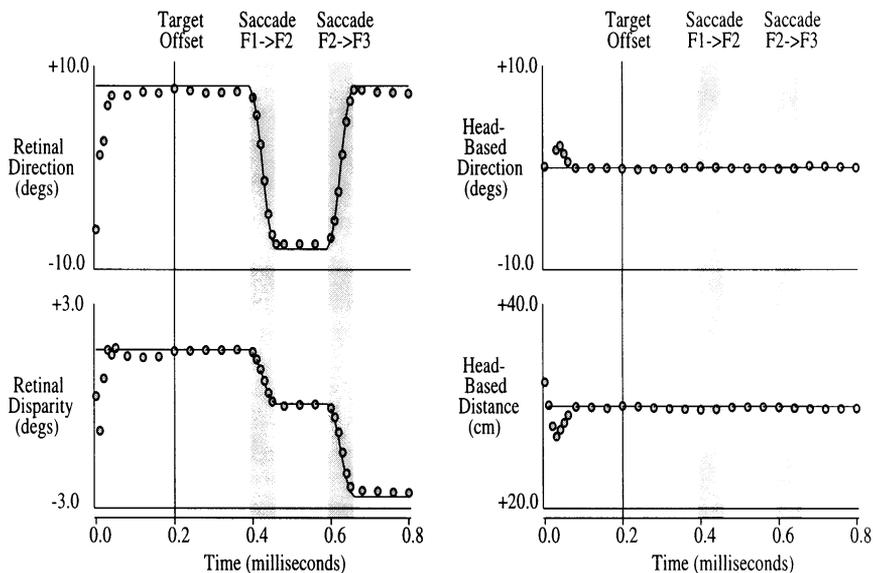


Fig. 3. The location outputs by the network with 80 units are shown for the example trial presented earlier in Fig. 1. Each plot shows the correct target location by the black line along with the location output of the network overlaid as a series of circles. On the left, the plots for the eye-based coordinates, retinal direction and disparity, are given. The eye-based location is updated each time an intervening saccade occurs. On the right, the head-based coordinates, direction and distance, are given. The head-based location remains stable despite intervening saccades.

Table 1  
The performance for networks with different sized hidden layers<sup>a</sup>

Training regime	Hidden units	RMSE for retinal direction (degrees)	RMSE for retinal disparity (degrees)	RMSE for spatial direction (degrees)	RMSE for spatial depth (cm)
None	–	6.61	1.83	4.79	4.75
Continuous	20	2.21	0.78	2.36	2.70
Continuous	40	1.47	0.49	1.83	1.82
<b>Continuous*</b>	<b>80</b>	<b>1.06</b>	<b>0.37</b>	<b>1.56</b>	<b>1.74</b>
Continuous	80	1.24	0.37	1.70	1.65
Continuous	160	1.12	0.29	1.55	1.57

<sup>a</sup> The root mean squared error between the location output by the network and the correct location is averaged as 5000 random task trials are simulated. It is reported here for each of the output coordinates. In the top row, it is given for the case when the location is simply set its mean value. This provides a baseline performance. The asterisk and bold print highlight the network whose hidden unit response properties are analysed in detail.

cade task. The two strategies are implemented in a complex distributed way among the hidden units. There is a continuous range of hidden unit types between those with eye-based receptive fields and those with head-based receptive fields. Most units participate in both representations simultaneously. The connectivity in the hidden layer suggests that both representations are involved in maintaining a working memory through recurrent feedback, and that they support each other by sharing information.

The receptive fields during visual and memory periods were mapped in detail at each of nine fixation positions. In Fig. 9, the different types of receptive fields are presented for four hidden units. Only the visual receptive are shown here, but in each case the memory receptive field is similar except for a change in the magnitude of the response. In Fig. 9A, unit 45 has a pure eye-based receptive field that is tuned to the retinal direction and disparity of the target. Note that the location of the receptive field remains nearly constant relative to the fixation point. In Fig. 9B, unit 44 has a pure head-based receptive field that is tuned to the direction and distance of the target in the workspace. The position of its receptive field does not vary with the position of the eyes. A second head-based unit is shown in Fig. 9C. It shows more variation in response with eye position. Also, it has a receptive field that is peaked for locations central in direction. Last, in Fig. 9D a third type of unit is presented. It has an intermediate receptive field. Intermediate receptive fields are characterized by the magnitude of the response being modulated by a gain field that depends on where the eyes are fixating.

To provide a quantitative description of the receptive fields, the data from each unit is fit by four alternative regression models (see Appendix A). The first model is a pure eye-based receptive field. It has six free parameters which define a two-dimensional Gaussian curve that is a function of the retinal direction and retinal disparity of the target. The second model is a pure

head-based receptive field. It has also six free parameters which define a two-dimensional Gaussian curve that is a function of the direction and the distance of the target in space. The first and second models are each augmented with a planar gain field to give the third and fourth models. The planar gain field has three free parameters and is a function of the conjugate and vergence eye position. It multiplicatively modulates the magnitude of the receptive field response for different fixation positions.

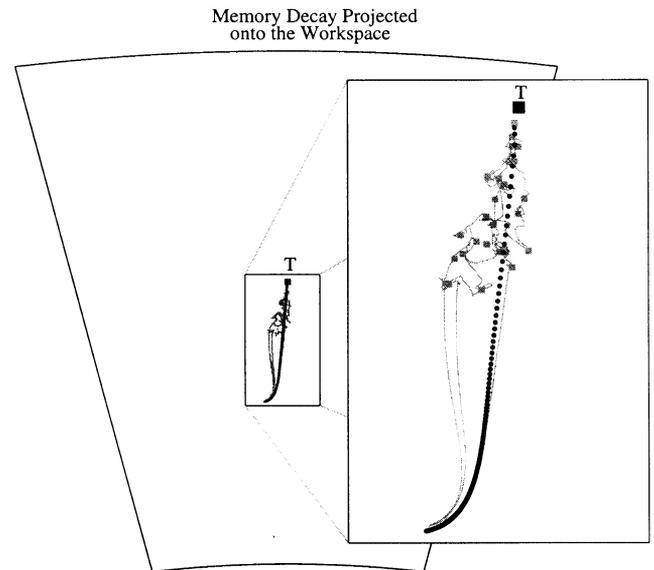


Fig. 4. The trajectory of decay for the head-based location output by a network of 80 units is shown. Each trial begins with the eyes fixated at the center of the workspace and with the target being flashed for 100 ms at the location labeled T. After the target disappears, the location stored by the network is shown over a 20-s period. The trajectory indicated by the series of black circles shows the decay when no intervening saccades occur. Each black circle gives the location every 100 ms. The gray trajectories show the decay on trials where intervening saccades are performed to random locations. The saccades occur every 200 ms during the first 2 s. The gray squares mark the point in the trajectory where a saccade begins. Each saccade introduces a perturbation to the stored location and also to the subsequent direction of decay. On average, the perturbations cancel and the value decays to the same fixed point at a similar rate.

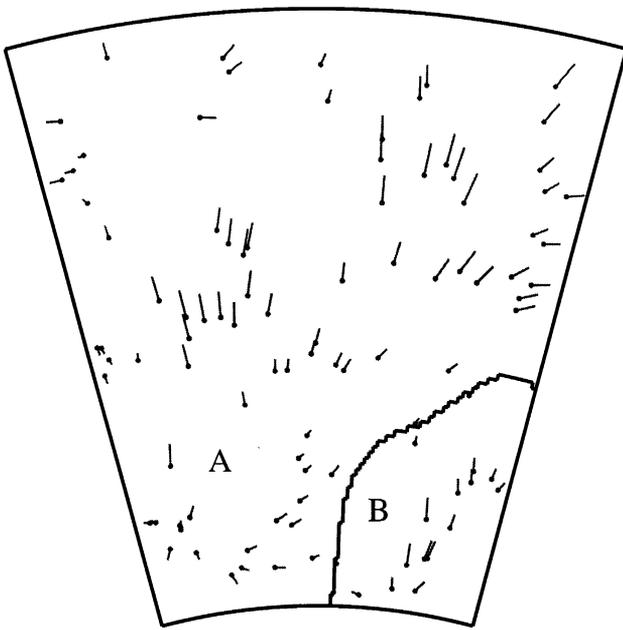


Fig. 5. The attractor basins of the 80 unit network are shown. The final location to which each point in a basin converges is indicated by the letter (A,B). The decay during the first second is shown at several points in the workspace as the vectors with circles on their ends.

Regressions provide estimates for the locations and the sizes of the receptive fields, for the direction and amplitude of the planar gain fields, and also for the amplitude of the visual and memory responses. The locations of receptive fields are spread over the space of each output more or less uniformly (Fig. 10A,B). The directions of gain fields are also more or less uniformly distributed (Fig. 10C). Over the population, there was no significant correlation between where the receptive field was located and the direction of the gain field. In other words, units with similar receptive field locations typically had different gain field directions. The amplitude of visual and memory response in each unit varies widely across the population (Fig. 10D). Some units

respond only to visible targets while others respond only after they disappear. Others respond in both conditions.

The size of the receptive fields in the hidden layer approximately matches that of the receptive fields in the output layer. In fact, it was possible to control the size by adjusting the outputs. These changes do not affect the types of hidden units that form, or the mechanisms used to solve the problem. In the network with 80 units, the average size of the eye-based receptive field is 6.5° in retinal direction and 2.8° in disparity. This is close to what is reported among LIP cells (Gnadt & Mays, 1995; Gnadt & Breznen, 1996). The average size of the head-based receptive field was 6.4° in direction and 6.9 cm in distance. No statistics are available to compare with these head-based sizes. However, the exact size does not appear essential to the network’s behavior.

Among the population there is a continuous range between eye-based, head-based, and intermediate response types. This is illustrated in Fig. 11. Each unit is presented as a single point. Its location on the *x*-axis describes whether or not it has eye- or head-based tuning. Its location on the *y*-axis indicates whether or not it has a strong gain field. Units with pure eye-based tuning, such as unit 45, appear in the lower left corner. Those with pure head-based tuning, such as unit 44, appear in the lower right corner; and those with strong gain fields, like unit 14, appear towards the upper middle. The other units are evenly distributed between these extremes.

Analyzing the strength of the recurrent weights helps to understand how the network solves the problem. The strength of the connections between different types of units indicates how information flows and is transformed in the hidden layer. If a group of units uses information from another group then the weights coming from that group should be non-zero. On the other hand, if the information is ignored then the weights should go to zero to reduce the interference caused by the extraneous activity.

Table 2  
The rate of decay in cm per second are given for each attractor of a different sized networks<sup>a</sup>

Hidden units	Decay within each attractor (cm/s)					Average decay over workspace
	A	B	C	D	E	
20	1.21	1.52	1.51	1.56		1.49
40	0.59	0.81	0.76	0.99	1.16	0.88
<b>80*</b>	<b>0.79</b>	<b>0.72</b>				<b>0.78</b>
80	0.47	0.96	0.65	0.73	0.99	0.75
160	0.43	0.33	0.33	0.42		0.40

<sup>a</sup> In each row, the attractors are arranged in ascending order by their distance from the center of the workspace. Attractors near the center have slower rates of decay because the center is sampled more frequently in training. On the far right the average rate of decay over the entire workspace is given. The row highlighted in bold text or with asterisk corresponds to the network whose units response properties are analysed in detail.

Table 3

The mean and S.D. of the percentage shift and direction of the shift are given for different sizes of the network<sup>a</sup>

Training regime	Hidden units	Percentage shift (mean $\pm$ S.D.)	Shift direction (mean $\pm$ S.D.)
Continuous	20	78.3 $\pm$ 20.5	-0.2 $\pm$ 13.4
Continuous	40	91.1 $\pm$ 19.8	-0.3 $\pm$ 12.0
<b>Continuous*</b>	<b>80</b>	<b>93.6 <math>\pm</math> 18.5</b>	<b>-0.2 <math>\pm</math> 10.9</b>
Continuous	80	91.0 $\pm$ 16.8	-0.2 $\pm$ 11.0
Continuous	160	92.4 $\pm$ 17.1	-0.1 $\pm$ 11.0

<sup>a</sup> Statistics are computed from saccades with magnitude greater than 5° in 5000 random task trials. The vector shift in the output location,  $\delta$ , was computed as the difference between the location one timestep after and one timestep before the saccade. The percentage shift is  $|\delta|/|\bar{s}|$  where  $\bar{s}$  is the saccade vector, and the shift direction is the angle between  $\delta$  and  $-\bar{s}$ .

Eye- and head-based strategies can be differentiated by the pattern of recurrent connectivity between different types of units. For the eye-based strategy, the eye-based units alone should maintain the memory of the target. Given the memory provided by them, the intermediate units could transform the location into a head-based frame for the outputs. This transformation could be done in a feed-forward manner without recurrent feedback among either the intermediate or head-based units. In short, only the eye-based units would require strong recurrent feedback to themselves in order for them to maintain the memory. For a pure head-based strategy the reverse is true, only the head-based units would require strong recurrent feedback. The head-based memory can be transformed back into the eye-based frame in a feed-forward manner.

The connectivity in the network suggests that both strategies are active simultaneously, and that information is shared between them. Units are divided into three nearly equal groups: eye-based, intermediate, and head-based as depicted in Fig. 11. The average magnitude of recurrent weights from each group to the others are calculated and presented in Table 4. Both the eye- and head-based groups have strong recurrent feedback. They also make connections of comparable size to the intermediate types. This suggests that besides maintaining memory among themselves, that information is shared in both directions going from eye- to head-based types and vice versa. Further, although the feedback for the intermediate units is smaller, it is still of similar size. Thus intermediate types are unlikely to be limited to relaying information back and forth, but are also involved in maintaining a memory among themselves.

The mixture of strategies employed by the network is also revealed when either its eye velocity or eye position inputs are lesioned. Both of these lesions impair its ability to update locations during intervening saccades. If a pure eye-based strategy was employed, then it would only require eye velocity to keep the location updated (Droulez & Berthoz, 1991; Moschovakis, 1996; Zhang, 1996). In this case, lesions of eye position would be expected to have no effect on the network's updating ability. On the other hand, for a pure head-based

strategy only the eye position should be required in order to update the location from the stored value (Robinson, 1975; Sparks & Mays, 1983a). Here lesioning eye velocity should have no effect. In Table 5, the performance is presented for both types of lesion. In either case there is a slight impairment. The network makes a partial update with about half the desired shift in the correct direction. The two strategies make nearly equal contributions to the update.

## 8. Discussion

The memory mechanism learned here is of the 'analogue' type that works by slowing the ultimate decay to one of a few attractors. This differs in behavior from other possible solutions such as the dynamic memory map proposed by Droulez and Berthoz (1991) and the 'line attractor' proposed by Zhang (1996). In our solution, decay always follows a fixed path for a given starting location. In their models memory decay is noise driven in a random walk motion that does not go to any particular fixed point. Another possible way to implement active memory is to have a large number of attractors spread out over space. This would lead to decay to one of many fixed points, but have the advantage that the terminal states would be near the remembered targets. These three kinds of decay could potentially be distinguished experimentally by observing short-term spatial memory decay.

Networks adopt solutions with a much larger number of attractors when they are trained on a discrete set of spatial locations. A network of 40 units trained with 16 locations in training develops a separate attractor for each location. The rate of decay, however, is much faster than it is in the continuous case (2.13 vs. 1.10 cm/s). Despite the fast initial decay, it performs about as well as the continuous counterpart because its final resting points are closer to the target. The shortcoming of this discrete model is that its hidden units have unrealistic receptive fields. They are square-shaped, often non-local, and have abrupt step-like changes in response moving from one spatial location to another.

Krommenhoek, Opstal, Gielen, and Gisbergen (1993), and Krommenhoek, Opstal, and Gisbergen (1996) present another network model that computes updated eye-based locations for intervening saccades. Their model consists of a feed-forward network with four layers. The first layer contains three signals sufficient for computing an updated eye-based location. The first signal is a memory of the target’s original eye-based location,  $RE$ . The second is a memory of the eye position at the time the target was observed,  $E_T$ . And the third is the current eye position,  $E_A$ . Given these signals, the updated eye-based location, or motor error, is  $ME = (RE + E_T) - E_A$ . Inputs from  $RE$  and  $E_T$  feed into the second layer of the network. Then information from the second layer is combined with input from  $E_A$  in the third layer which then projects to the fourth layer representation  $ME$ .

The four-layered Krommenhoek model is consistent with a head-based mechanism in line with that proposed by Robinson (1975). The second layer of their model combines information about the retinal location of the target,  $RE$ , and eye position,  $E_T$ , to compute its

location in a head-based frame. Interestingly, units in the second layer do not explicitly code for the head-based location with head-based receptive fields. They have eye-based receptive fields that are modulated by gain fields for the eye position similar to the model of Zipser and Andersen (1988). This provides a distributed representation of the head-based location. In the third layer the information from the second layer is combined with the current eye position in order to transform it back into an updated eye-based location.

Our model differs from the Krommenhoek model in two important respects. First, it actively models spatial working memory through the recurrent dynamics in the hidden layer. Thus, it is possible to test how spatial memory decays over time. The Krommenhoek model does not directly address this issue since it assumes that a memory is fixed and provided at the level of the input layer. Second, while their model uses a pure head-based mechanism, ours uses a mixture of both eye- and head-based mechanisms.

It is interesting that the network developed a mixed strategy even though no constraints prevented it from

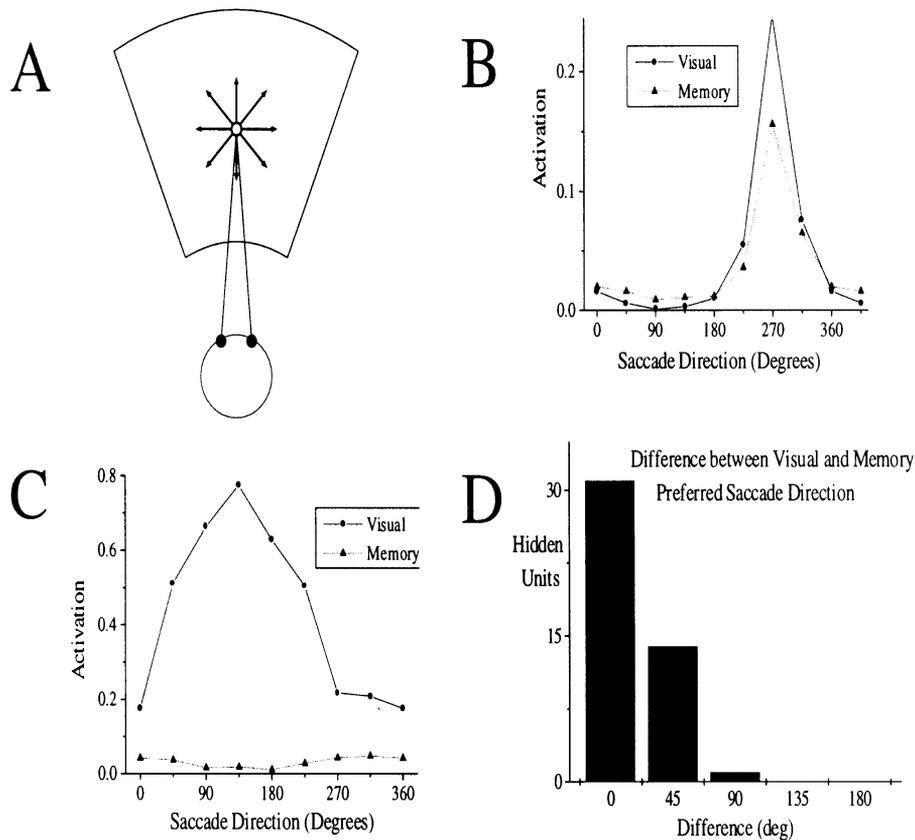


Fig. 6. Testing the preferred saccade direction. (A) A target is flashed 5 cm from the center of the workspace at one of eight directions. Each unit’s response is measured at 100 ms while the target is still visible and then again 100 ms after it disappears. (B) The activation of unit 44 is plotted over the eight tested directions during visual and memory periods. The preferred direction is aligned. (C) Unit 12 has a weak response during the memory period. Its preferred direction does not match in the two periods. (D) The difference between each unit’s preferred direction in the visual and memory periods is computed and accumulated into a histogram. Units with a weak response in one of the two periods are not included. A response is considered weak if the range from maximum to minimum activation over the eight directions is less than 0.05.

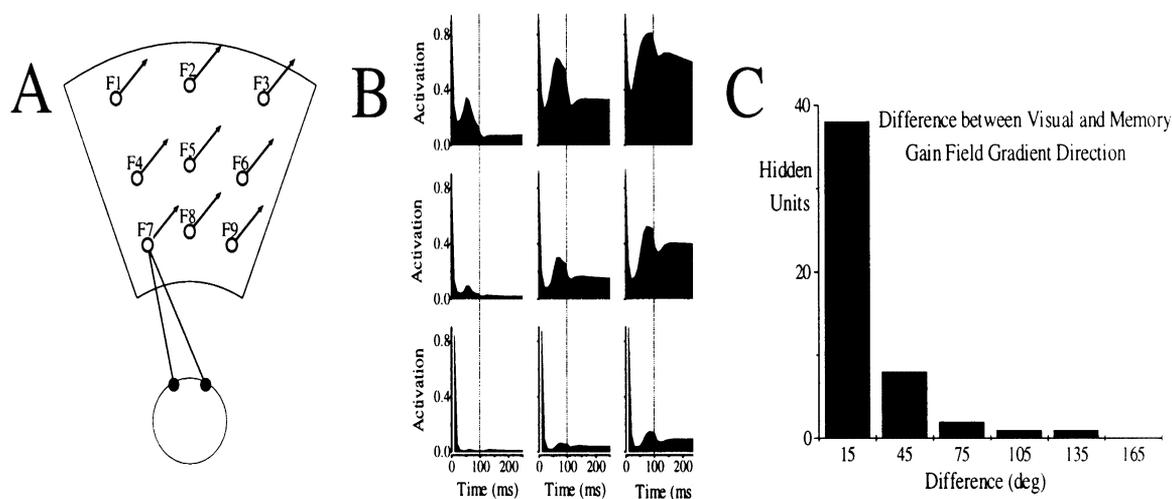


Fig. 7. Testing for gain fields. (A) The response of each unit is measured for a target flashed at its preferred direction from nine different fixation positions (labeled F1 through F9). (B) The response of unit 44 at each of the nine fixation locations is shown. In each plot, the activity is shown over the first 250 ms of the trial. The target disappears at 100 ms. The response is greatest for fixation positions to the top right. (C) Planar gain fields are fit to the response in both the visual and memory periods (see Appendix A). Each fit gives an estimate of the gradient direction of the gain field. The difference between the direction in the two periods is computed for each unit and accumulated into a histogram.

using either the pure eye- or pure head-based strategies. For a pure head-based strategy, the network could have developed a subset of units in the hidden layer that implemented a head-based memory, and then other units could have used the current eye position with that memory in a feed-forward manner to compute the updated eye-based locations. That solution would be equivalent to the Krommenhoek model, but collapsed into a single recurrent hidden layer. Alternatively, a subset of units could have formed a pure eye-based memory, and then other units could have used eye position in a feed-forward manner to compute the head-based location. Instead the network uses both mechanisms in a complex distributed fashion.

The network produced internal representations that are similar to those found in parietal areas. The main characteristic shared by parietal cells and the hidden units is that their receptive field responses are modulated by the position of the eyes. The modulation takes the form of a *planar gain field* in which the response increases monotonically as the eyes move towards a peripheral fixation point. Similar planar gain fields have been observed for the memory responses of LIP and area 7a saccade cells (Andersen et al., 1990). However, these tests were performed only in the fronto-parallel plane. Planar gain fields have also been reported in depth, but only for visual responses of these cells (Gnadt & Mays, 1989). Our model predicts that gain fields in depth are also present during memory periods.

One discrepancy between the model and physiological data is that head-based cells have not been found in area 7a or LIP. In the study of Andersen et al. (1990), over 50 cells in area 7a and LIP were mapped for their preferred saccade direction from two or more fixation

locations. If a cell had a head-based receptive field, then the preferred direction is expected to change with fixation in a manner that is similar to hidden unit 12 in the model (Fig. 8C). None of the 50 cells tested in area 7a or LIP exhibited changes in preferred direction. A later study did find some cells that shifted their preferred direction with fixation, but this only occurred for auditory targets (Stricanne, Andersen, & Mazzoni, 1996).

It should be noted that some head-based units may be overlooked in experiments where only the preferred direction is tested. This is demonstrated by unit 44. Its preferred direction remains the same at different fixations (Fig. 8B). With this data alone, it would appear to be an eye-based receptive field with a strong gain field preferring fixations to the upper right (Fig. 7B). However, if its receptive field is mapped in detail at several fixation locations, then it is clearly a head-based unit (Fig. 9B).

Although area 7a and LIP may lack head-based cells, those cells do appear in nearby parietal areas that may also be involved in memory-guided saccades. A recent study has found that micro-stimulation of cells in a restricted region near LIP results in saccades directed to head-based locations (Thier & Andersen, 1998). This area may overlap with parts of area VIP in which cells with head-based visual receptive fields have been identified (Duhamel, Bremmer, BenHamed, & Graf, 1997). Another parietal area, area PO, also contains cells with head-based receptive fields and has cells with saccade-related activity (Galletti, Battaglini, & Fattori, 1991). Both VIP and PO are reciprocally connected to LIP and area 7a (Blatt & Andersen, 1990). Thus cells in these areas could participate in programming memory-guided saccades through their recurrent connections to

LIP and area 7a. It remains untested whether or not these cells maintain memory activity during saccade tasks.

**Appendix A**

*A.1. The simulated saccade task*

The simulated task is intended to be similar to the double-saccade paradigm (Hallett & Lightstone, 1976; Mays & Sparks, 1980; Sparks & Mays, 1983b). A visual target appears for a short period. After it disappears several saccades to other secondary locations are performed in complete darkness. The network is trained to remember how to saccade back to the target.

The network is trained on a continuous range of locations. The locations for both the target and the initial and secondary fixation points are chosen at random from the horizontal depth plane. The boundaries of the space sampled extends from  $-10$  to  $10^\circ$  in direction and from 20 to 40 cm in depth. Target and fixation locations are chosen randomly from a Gaussian distribution positioned at the center of the space and with a S.D. of  $5^\circ$  in direction and 5 cm in depth. Sampling the center preferentially in this manner yields hidden units with more realistic local receptive fields. Otherwise, receptive fields tend to be broad and located only on the periphery of the space.

The timing of events in the task must be chosen at random. Otherwise, the network may develop solutions that rely on its specifics. The duration of the visual target's presentation is chosen from an exponential distribution with a mean time of 100 ms. Once the target disappears, saccades to secondary locations are performed at random intervals. The interval between saccades is chosen from an exponential distribution with a mean time of 300 ms. The total duration of the

trial is also chosen from an exponential distribution with a mean time of 1000 ms. There are typically 2–3 intervening saccades on each trial. Time is divided into steps of 10 ms. In the average trial there are 100 time steps.

Intervening saccades are simulated by gradually moving the fixation point in a straight line to its new location. Saccades occur over several time steps. The size of the increment on each step is chosen to give a bell-shaped velocity profile over the duration of the movement. If the movement occurs over  $N$  time steps, then the magnitude of the increment at time step  $i$  is given by

$$\text{Inc}(i) = \frac{1}{Z} e^{-\frac{(i - (N/2))^2}{2(N/4)^2}} \tag{1}$$

where  $Z$  normalizes the increments so they sum to the magnitude of the saccade.

The duration of intervening saccades is longer for larger saccades. In humans, conjugate movements have been estimated to take  $20 + 2.5|\Delta\theta_C|$  ms where  $|\Delta\theta_C|$  is the angular change in horizontal visual degrees (Baloh, Sills, Kumley, & Hornibia, 1975). Joint movements in direction and depth are known to be slower (Collewyn, Casper, & Steinman, 1995). We model the slower duration as  $30 + 3.0|\Delta\theta|$  where  $|\Delta\theta|$  is the magnitude of the 2-d angular change in conjugate and vergence eye position (which are described in the next section). The duration is rounded off to the nearest multiple of the time step size.

*A.2. Inputs and output encodings of the network*

*A.2.1. Extra-retinal inputs*

The conjugate and vergence angles are used to describe the position of the two eyes. The conjugate angle is the average angle of rotation of the two eyes in their orbits. The vergence angle is the difference in the

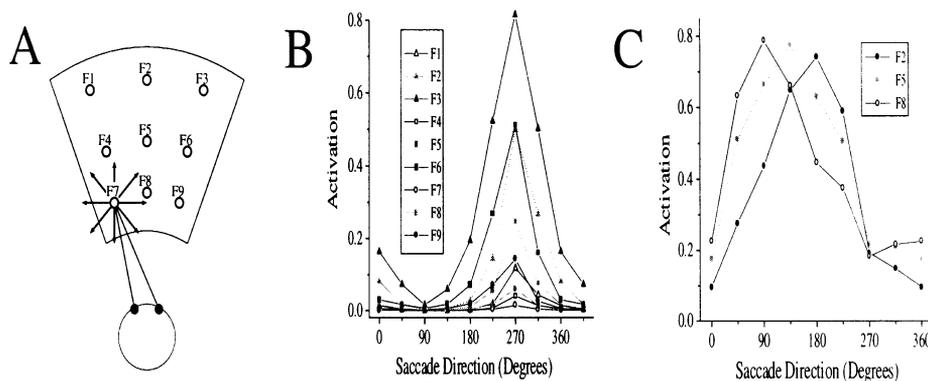


Fig. 8. Testing the preferred direction at different fixations. (A) The preferred direction test performed earlier at the central fixation is now repeated at the other eight fixation locations. (B) Unit 44 prefers the same saccade direction for all of the fixations tested. However, the magnitude of its response changes with each fixation. (C) A small set of units shift their preferred direction for different fixations. Unit 12 exhibits this type of behavior.

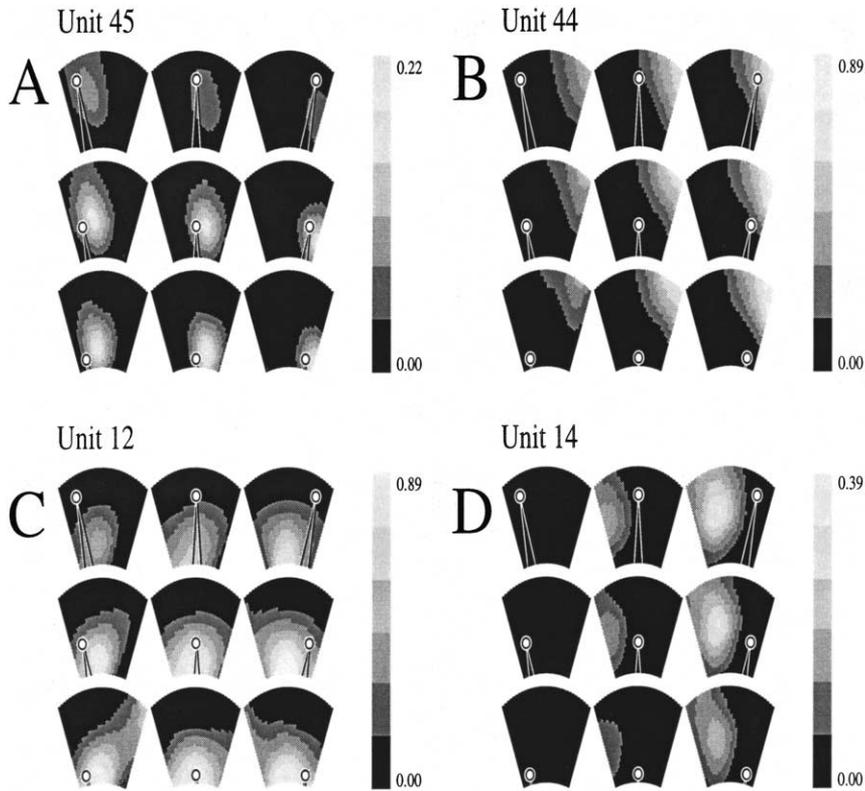


Fig. 9. The receptive fields of four different hidden units from a 80 unit network are presented. The activation of each unit’s response is shown as a function of the visual target’s location in the workspace. This plot is repeated nine times, each time with the target presented while the eyes are at a different fixation position as indicated by the circle. The activity of the unit is measured during the visual response period 100 ms after the target has first appeared. The receptive fields during the memory period (not shown) are similar. (A) Unit 45 has an eye-based receptive field. Its maximal response occurs in roughly the same place relative to the fixation point. This unit prefers locations just right of the fixation point. (B) Unit 44 has a head-based receptive field. The location of its maximal response occurs at the same place in the workspace regardless of where the eyes are fixating. This unit prefers locations in the upper right of the workspace. (C) Unit 12 also has a head-based receptive field but with more variation with fixation. It prefers locations that are central in direction and towards the bottom of the workspace. (D) Unit 14 has an intermediate receptive field. The response is modulated by where the eyes are fixating. It is stronger for fixation positions toward the right of the workspace.

angle of rotation of the left and right eyes. Given a fixation point in space, the corresponding conjugate angle,  $C_F$ , and the vergence angle,  $V_F$ , are computed as

$$C_F = \theta \tag{2}$$

$$V_F = \frac{180}{\pi} \frac{I \times \cos(\theta)}{D} \tag{3}$$

where  $\theta$  and  $D$  are, respectively, the direction and distance of the fixation point, and  $I$  is the inter-ocular distance between the eyes. The inter-ocular distance of the model is 4 cm, which is comparable to a monkey.

The extra-retinal inputs to the network consist of 4 units. The first two units encode the eye position by the conjugate and vergence angles. The other two units encode the velocity of the conjugate and vergence angles. The velocity is approximated as the difference between the current and the previous time step.

Each extra-retinal input has an activation that is a linear function of the angle it encodes. Cells with linear coding of the conjugate and vergence position are found

in parietal areas (Sakata, Shibutani, & Kawano, 1980; Squatrito & Maioli, 1996). Linear coding of conjugate and vergence velocity is also found at the level of the brainstem (Mays & Gamlin, 1995; Moschovakis, Scudder, & Highstein, 1996). Since cells in the brain typically share similar dynamic ranges, we scale the activation of each input so its minimum and maximum values will range from  $-0.5$  to  $0.5$  during the task.

#### A.2.2. Visual inputs

The visual input to the network consists of a retinal array for the left and right eyes. Although parietal cells do not receive direct retinal inputs, we assume that the same information is retained in their extra-striate inputs.

The retinal arrays each have 10 units. Each unit has a Gaussian receptive field that is sensitive to where light from the target hits the retina. The centers of the receptive fields are evenly spaced to cover the range from  $-25$  to  $25^\circ$ . The activation of a unit at the receptive field location  $R_x$  is given by

$$Act = e^{-\frac{(R_x - R_T)^2}{2\sigma^2}} \quad (4)$$

where  $R_T$  is the location of light from the target on the retina and  $\sigma$  is the receptive field width. The width is  $7^\circ$  similar to visual cells in LIP (Gnadt & Breznen, 1996; Platt & Glimcher, 1998).

The point where light hits each retina is computed from the target and fixation locations in space. First, the location of the target is expressed in terms of the conjugate and vergence angles that would be required to fixate it,  $C_T$  and  $V_T$  using Eqs. (2) and (3). Then the retinal direction of the target from the cyclopedian eye,  $R_\theta$ , and the retinal disparity between the eyes,  $R_d$ , are given by

$$R_\theta = C_T - C_F \quad (5)$$

$$R_d = V_T - V_F \quad (6)$$

where  $C_F$  and  $V_F$  are the conjugate and vergence angles of the fixation point. Then the location on the left retina is  $R_\theta - R_d/2$  and on the right it is  $R_\theta + R_d/2$ .

### A.2.3. Eye-based outputs

A  $10 \times 10$  array of units encodes the eye-based location. The location is given by the cyclopedian retinal direction and retinal disparity of the target,  $R_\theta$  and  $R_d$  in Eqs. (5) and (6). Each unit has a Gaussian receptive field centered over a different location. Receptive field

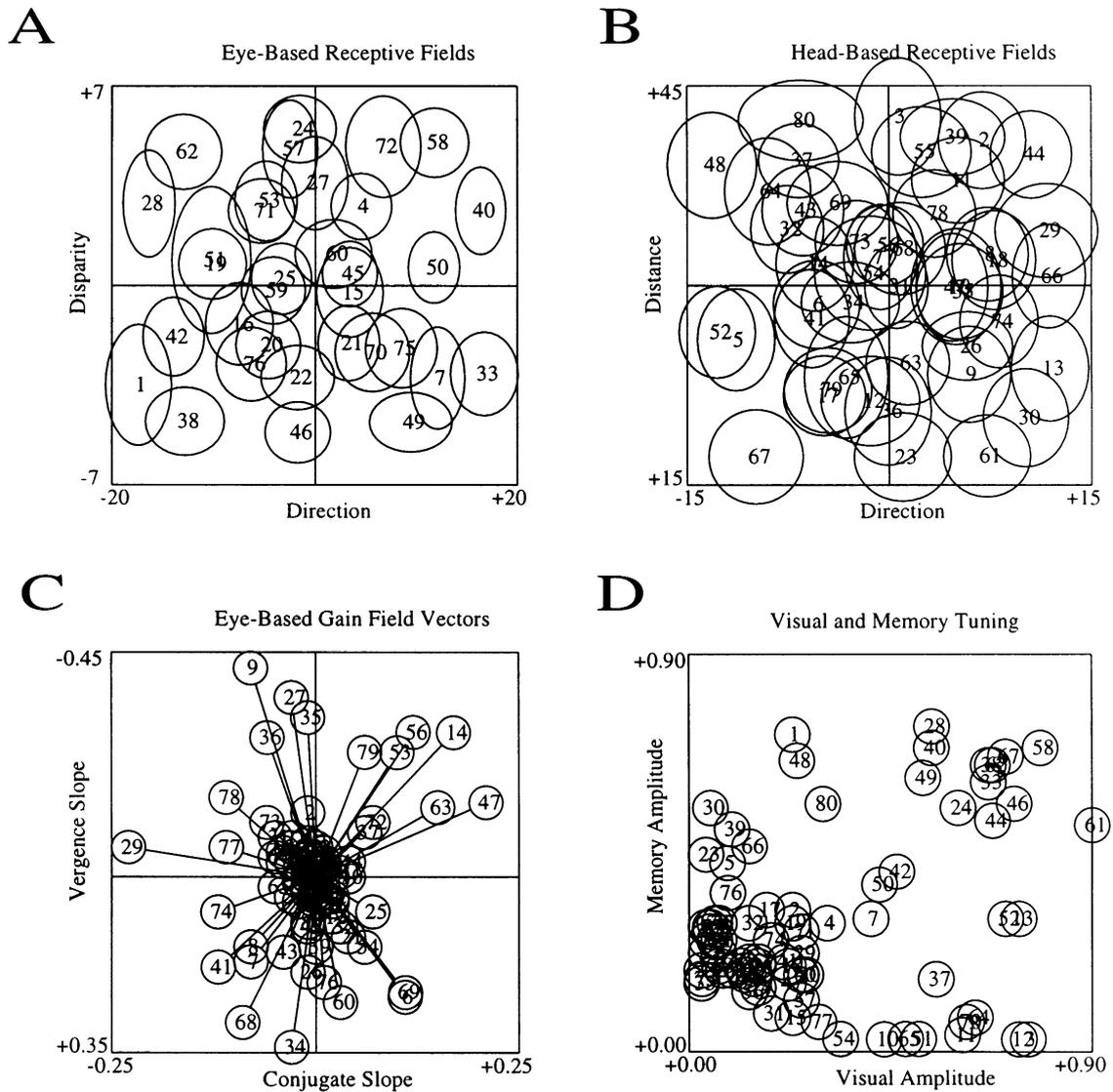


Fig. 10. This figure summarizes the receptive field, the planar gain field, and the visual and memory response magnitudes in a network with 80 hidden units. (A) The location and size of the receptive fields of those units best fit with eye-based models are depicted. Each unit's receptive field is drawn as a circle centered at its location and with a width and height up to  $1\sigma$  or away from the center. (B) The location and size of the receptive fields of those units best fit with head-based models are depicted. (C) The vector of the gain field's gradient is drawn for each hidden unit. There was no difference between the gain fields of eye- and head-based units. Both are included here. (D) The magnitude of each unit's visual and memory response is plotted.

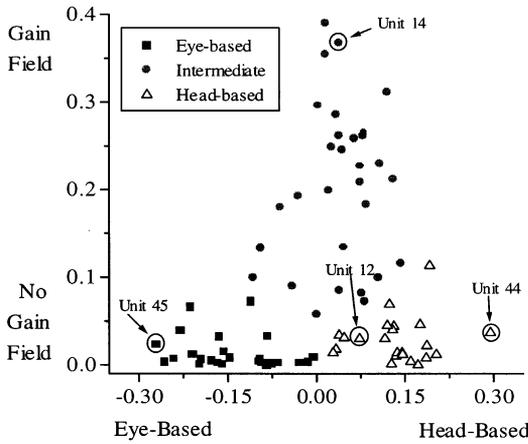


Fig. 11. The distribution of eye, head, and intermediate types. The tuning of each unit is represented as a point in the plot. The location on the  $x$ -axis indicates the degree to which the unit is eye- or head-based. The location on the  $y$ -axis indicates the strength of the gain field (see Appendix A for axis locations computed from the fit models). In the plot, the eye-based unit (number 45), the head-based units (numbers 44 and 12) and the intermediate unit (number 14) are labeled. The units are divided into three nearly equal groups of head-, eye-based, and intermediate types as indicated by the symbols. Although there is clearly a continuous range of types, this classification is useful later for analyzing the connectivity between types.

locations are evenly sampled from  $-25$  to  $25^\circ$  in direction and from  $-10$  to  $10^\circ$  in disparity. This range is sufficient to insure that it falls inside the array.

The activation of a unit at location  $(R_x, R_y)$  is given by

$$\text{Act} = B + (P - B)e^{-\frac{(R_x - R_\theta)^2}{2\sigma_\theta^2}} e^{-\frac{(R_y - R_d)^2}{2\sigma_d^2}} \quad (7)$$

where  $\sigma_\theta$  is the receptive field width in direction,  $\sigma_d$  is the width in disparity,  $B$  is the baseline activation, and  $P$  is the peak activation. The baseline and peak activation were set to be 0.10 and 0.50, respectively. This gives a low baseline firing rate and a peak rate that remains below the saturation of the sigmoid activation function of the output units. Likewise, cortical cells have low baseline rates and their peak firing typically is well below saturation.

Table 4  
The connectivity between different types of units. The units of a network of size 80 are divided into three groups of eye-based, intermediate, and head-based types (see Appendix A). The average magnitude of recurrent weights from one group to another is given in each entry

From	To		
	Eye-based	Intermediate	Head-based
Eye-based	1.36	1.34	0.98
Intermediate	0.84	0.64	0.78
Head-based	1.05	1.08	1.51

Table 5  
The effect of eye velocity and eye position lesions on the percentage shift and direction in remapping the eye-based outputs is given for a network with 80 units<sup>a</sup>

Lesion type	Percentage shift (mean $\pm$ S.D.)	Shift direction (mean $\pm$ S.D.)
None	93.6 $\pm$ 18.5	-0.2 $\pm$ 10.9
Velocity	47.8 $\pm$ 22.1	-0.1 $\pm$ 16.8
Position	54.9 $\pm$ 20.0	-0.6 $\pm$ 16.5

<sup>a</sup> Eye velocity lesions were implemented by setting the velocity inputs to zero. Eye position lesions were implemented by holding the eye position inputs constant. The percentage shift and directional error are defined the same as in Table 3.

The width of the output receptive fields are chosen to match saccade cells in area LIP. In direction the width is  $7^\circ$  (Gnadt & Breznen, 1996; Platt & Glimcher, 1998). In disparity, cells in LIP have widths between 2 and  $5^\circ$  (Gnadt & Mays, 1995). We use  $7^\circ$  for direction and  $2.5^\circ$  for disparity. This choice makes the shape of the receptive field circular in the coordinate frame of the output array.

#### A.2.4. Head-based outputs

A  $10 \times 10$  array of units also encode the head-based location. The location is described by its direction and distance of the target in space. Similar to the eye-based outputs, each unit has a Gaussian receptive field that is centered over a different location. Locations are evenly sampled from  $-15$  to  $15^\circ$  in direction and from 15 to 45 cm in depth. This range covers any target in the workspace.

The activation of a unit at location  $(H_x, H_y)$  in the array is given by

$$\text{Act} = B + (P - B)e^{-\frac{(H_x - T_\theta)^2}{2\sigma_\theta^2}} e^{-\frac{(H_y - T_D)^2}{2\sigma_D^2}} \quad (8)$$

where  $T_\theta$  and  $T_D$  are the target direction and distance, and  $\sigma_\theta$  and  $\sigma_D$  are the width of the receptive field in direction and distance, respectively. Parameters  $B$  and  $P$  are the same as for the eye-based outputs.

Cells that have receptive fields tuned to the direction and distance of targets are found in several parietal areas (Sakata, Shibutani, Kawano, & Harrington, 1985; Galletti, Battaglini, & Fattori, 1995; Stricanne et al., 1996). Estimates of their size are not available. We choose a width of  $7^\circ$  in direction that matches the eye-based visual cells (Gnadt & Breznen, 1996; Platt & Glimcher, 1998). The width in depth is also set at 7 cm in order to make the receptive fields circular-shaped.

#### A.3. The neural network and training

The neural network architecture depicted in Fig. 2 consists of an input, hidden, and output layer of units.

Each hidden unit,  $h_i$ , receives weighted connections,  $V_{ik}$ , from every input unit,  $x_k$ , and recurrent connections,  $W_{ij}$ , from every hidden unit. Each also has a bias,  $B_i$ . The activation of a unit at time  $t + 1$  is given by

$$h_i(t + 1) = f\left(\sum_j W_{ij}h_j(t) + \sum_k V_{ik}x_k(t) + B_i\right) \quad (9)$$

where  $f(x)$  is the logistic function

$$f(x) = \frac{1}{1 + e^{-x}}$$

The activation is intended to represent the average firing rate of a neuron. The hidden units project to the output layer. Each output unit,  $o_m$ , has an activation given by

$$o_m(t + 2) = f\left(\sum_i U_{mi}h_i(t + 1) + B_m\right) \quad (10)$$

where  $U_{mi}$  denote the weighted connections from the hidden units to the outputs and  $B_m$  is a bias.

The connections and biases of the network are optimized through gradient descent using the *Backpropagation Through Time Algorithm* (Rumelhart, Hinton, & Williams, 1986; Williams & Zipser, 1995). Three extra constraints are included in the optimization that produce more realistic receptive field properties among hidden units. The first encourages hidden units to have low resting activities similar to real cortical neurons. The second constrains the weights from the hidden to output units to be positive. This forces the hidden units to develop receptive fields that are similar to those of the output units. This is desirable because the outputs are similar to saccade cells in parietal cortex. The third constraint sets the sum of the weights from each hidden unit to all the output units to be equal to a constant. This forces each hidden unit to make a nearly equal contribution to the computation.

The first constraint is included in the optimization by adding a cost to the error function used in the backpropagation algorithm. The appended error function is

$$E(t) = \sum_{m=1}^M (y_m^*(t) - y_m(t))^2 + \lambda \sum_{i=1}^H C_i(t) \quad (11)$$

where  $M$  is the number of outputs,  $y_m^*(t)$  is the desired target activation for output unit  $m$ , and  $y_m(t)$  is the actual activation.  $C_i(t)$  is a cost function that is defined for each of the units in a hidden layer of size  $H$ . The constant  $\lambda$  was set to be a small fraction (0.01).

The cost function  $C_i$  is computed from several statistics of the hidden unit activity over time. The first is the average activity estimated by  $\bar{h}_i(t) = \alpha h_i(t) + (1 - \alpha)\bar{h}_i(t - 1)$  where  $\alpha$  is a constant set at 0.005. The second is the instantaneous average of the population's activity

$$\bar{h}(t) = \frac{1}{H} \sum_{i=1}^H h_i(t)$$

The third is each unit's variance estimated by  $v_i(t) = \alpha(h_i(t) - \bar{h}_i(t))^2 + (1 - \alpha)v_i(t - 1)$ . The fourth, and last, is the average variance of the population

$$\bar{v}(t) = \frac{1}{H} \sum_{i=1}^H v_i(t)$$

The cost function  $C_i$  is then given by

$$C_i(t) = (\bar{h}_i(t) - \mu)^2 + \alpha H (\bar{h}(t) - \mu)^2 + \frac{1}{2} (\bar{v}_i(t) - \bar{v}(t))^2 \quad (12)$$

where  $\mu$  is the desired resting activity. The first term forces individual units to have an average activity  $\mu$  which is set to a low value of 0.1. The second term prevents the population of units from being active at the same times. If the first two terms are applied alone, low average activations will result but a large fraction of the units will remain inactive at all times and effectively play no role in the computation. The third term corrects for this by forcing them to have approximately the same variance over time.

The other constraints are enforced by resetting the weights to the output units after each time step. First, any negative weights are reset to zero. Then the weights coming out of each hidden unit to the outputs are renormalized. The renormalized weight,  $U'_{mi}$ , is computed as

$$U'_{mi} = \frac{\bar{U} \times U_{mi}}{\bar{U}_i} \quad (13)$$

where  $U_{mi}$  denotes the weight from hidden unit  $i$  to output unit  $m$ ,  $\bar{U}_i$  is the average output weight from hidden unit  $i$ , and  $\bar{U}$  is the average output weight from all hidden units.

During gradient descent, training takes place on each time step with a probability of 0.10. This feature prevents recurrent networks from over-fitting the current trial, enabling them to generalize to all trials. The networks are trained in successive stages in which the learning rate is gradually decreased. They are trained for 25000 trials with a learning rate of 0.05. Then this rate is decreased by half every 2500 trials for the next 10000 trials. This allows a type of annealing in which rough solutions are found quickly and then progressively fine tuned at lower learning rates.

The initial state of activity in the network is reset for each trial. Since the network is recurrent, the initial state influences all subsequent behavior. To control for this, each hidden unit is set to the baseline activity as given by its bias value. The network learns to perform the task starting from this state.

#### A.4. Network performance

In analysis of the network's performance, the center of mass of the distributed activity in the output arrays is taken to be the target's location. The location of the center of mass is given as

$$(x, y) = \frac{\sum_i \sum_j (o_{ij} - B)(x_i, y_j)}{\sum_i \sum_j (o_{ij} - B)} \quad (14)$$

where  $o_{ij}$  is the activation of the output unit at the grid position  $(i, j)$ ,  $B$  is the baseline activity for the outputs, and  $(x_i, y_j)$  gives the location in space corresponding to grid position  $(i, j)$ .

#### A.5. Comparison to physiological data

Each hidden unit is subjected to similar tests as those done on saccade cells in LIP and area 7a (Barash et al., 1989; Andersen et al., 1990). The tests assess the unit's preferred saccade direction, how its response changes for different fixation positions, and how its response changes from visual to memory periods.

The preferred direction is assessed in a series of trials in which a target is flashed at one of eight directions (Fig. 6A). The response is measured at 100 ms while the target is visible and then 100 ms after it has disappeared.

To assess the gain field of each unit, a set of nine trials was performed. On each trial the eyes are positioned at one of nine different locations in a  $3 \times 3$  grid (Fig. 7A). The fixation positions are located at  $-5, 0$ , and  $5^\circ$  in conjugate direction, and at  $7, 8.5$ , and  $10^\circ$  in vergence. The target is always flashed at the unit's best saccade direction. Again the response is recorded at 100 ms while the target is visible and then 100 ms again after it disappears.

To quantify the gain field response, a planar model is fit to the data by least-squares regression. The planar model is given as

$$y = \mu + c_1(\theta_c - \bar{\theta}_c) + c_2(\theta_v - \bar{\theta}_v) \quad (15)$$

where  $\theta_c$  and  $\theta_v$  are the conjugate and vergence eye position, respectively, and  $\bar{\theta}_c$  and  $\bar{\theta}_v$  are the mean values. The gradient direction of the gain field is then computed as the inverse tangent of the conjugate slope,  $c_1$ , divided by the vergence slope,  $c_2$ .

The preferred saccade direction is assessed at several different fixation positions. The original test used to examine the preferred direction at the central fixation location is now repeated at the other eight fixation locations (Fig. 8A).

#### A.6. Quantitative analysis of receptive fields

The receptive field data for each hidden unit is fit to eye- and head-based models and models including planar gain fields. The data consist of samples from a  $10 \times 10$  grid of target locations at each of  $3 \times 3$  different fixation locations giving a total of 900 points (Fig. 9). There are 1800 points in total because a measurement is taken both when the target is visible and 100 ms after it has disappeared. A least squares minimization algorithm in Matlab finds the best model parameters.

Pure eye- and pure head-based receptive field models consist of a two-dimensional Gaussian curve with six free parameters. The equation for the curve is defined as

$$\text{Gaussian}(x, y) = \left( \mu + \left( A \times e^{-\frac{(x - \mu_x)^2}{2\sigma_x^2}} \times e^{-\frac{(y - \mu_y)^2}{2\sigma_y^2}} \right) \right) \quad (16)$$

where  $x$  and  $y$  refer to the location of the target (either in eye- or head-based coordinates),  $\mu$  is the baseline activity,  $A$  is the amplitude of the maximum response above baseline,  $\mu_x$  and  $\mu_y$  give the location of the receptive field, and  $\sigma_x$  and  $\sigma_y$  give the size of the receptive field. In the pure eye-based model, the curve is tuned to retinal direction and disparity. In the pure head-based model, it is tuned to the direction and distance in space.

The pure eye- and head-based models are augmented with a planar gain field. The gain field multiplicatively modulates the receptive field response as a function of eye position. It is defined as

$$\text{GainField}(C_F, V_F) = g(c_1 + c_2 C_F + c_3 V_F) \quad (17)$$

where  $C_F$  and  $V_F$  are the conjugate and vergence eye position, and the parameters  $c_1$ ,  $c_2$ , and  $c_3$  define a plane. The function  $g$  performs a threshold operation that leaves positive values unchanged and sets negative values to zero. Since the regressions use gradient information to find minima in parameter space,  $g$  must be a continuous function. It is approximated here as  $g(x) = 0.01 \times \ln(1 + e^{100x})$ .

Other gain field models did not fit the data as well. Other models included non-planar sigmoidal shapes and an additive instead of multiplicative modulation. For every unit, the simple planar model fit better or as well. Thus, only its results are presented.

Picking good initial values for the parameters is important for avoiding local minima in fitting the models. The initial receptive field location was set to be the location of the unit's peak response. Initial values for the other Gaussian parameters were set as  $\mu = 0$ ,  $A = 0.4$ ,  $\sigma_x = 7$ , and  $\sigma_y = 7$ . The planar gain field was initialized with  $c_1 = 1$  and  $c_2 = c_3 = 0$ . The results from regressions were inspected visually to insure they converged on reasonable values. In the network with 80

units, the average  $R^2$  of the best fit model was 0.945. The unit with the worst fit still had an  $R^2$  of 0.81.

The visual and memory response of each hidden unit had similar receptive field properties. The average absolute difference receptive field location was only  $0.6^\circ$  in retinal direction (over a range from  $-16$  to  $17^\circ$ ) and  $0.3^\circ$  in disparity (over a range from  $-5$  to  $5^\circ$ ). The average absolute difference in head-based direction was 1.2 degrees (range from  $-11$  to  $11^\circ$ ) and 1.7 cm in depth (range from 18 to 40). Further, the direction of the gain field was similar with an average absolute difference of  $12.2^\circ$ . The main difference in the two periods was the magnitude of the response changed. In the final analysis the same parameters were used to fit the receptive field with the exception that the amplitude parameter,  $A$ , was allowed to be different.

#### A.7. Unit classification

Units are divided into three nearly equal groups of eye-based, intermediate, and head-based types. Although there are not distinct boundaries between the types, the classification is still useful for estimating the strength of the recurrent connections between the groups. The classification for each unit is made by how well its receptive field is fit as either eye- or head-based and by the strength of its gain field. In Fig. 11 each unit is shown as a point along these dimensions. The  $x$ -axis location is given by the  $R$ -squared of the full head-based model minus the  $R$ -squared of the eye-based model. The  $y$ -axis location is given by the  $R$ -squared of best full model (either eye- or head-based) minus the  $R$ -squared of the corresponding reduced model with no gain field. The class is indicated by the symbol for each unit.

#### References

- Andersen, R., Bracewell, R., Barash, S., Gnadt, J., & Fogassi, L. (1990). Eye position effects on visual, memory, and saccade-related activity in areas lip and 7a of macaque. *Journal of Neuroscience*, 10(4), 1176–1196.
- Baloh, R., Sills, A., Kumley, W., & Hornibia, V. (1975). Quantitative measurements of saccade amplitude, duration, and velocity. *Neurology*, 25, 1065–1070.
- Barash, S., Bracewell, M., Fogassi, L., & Andersen, R. (1989). Interactions of visual and motor-planning activities in the lateral intra-parietal area (lip). *Society of Neuroscience Abstracts*, 15, 1203.
- Blatt, G. J., & Andersen, R. A. (1990). Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area lip). *Journal of Comparative Neurology*, 299, 421–445.
- Collewijn, H., Casper, J., & Steinman, R. (1995). Voluntary binocular gaze-shifts in the plane of regard: dynamics of version and vergence. *Vision Research*, 35(23/24), 3335–3358.
- Droulez, J., & Berthoz, A. (1991). A neural network model of sensoritopic maps with predictive short-term memory properties. *Proceedings of the National Academy of Science*, 88, 9653–9657.
- Duhamel, J., Bremmer, F., BenHamed, S., & Graf, W. (1997). Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature*, 389, 845–848.
- Galletti, C., Battaglini, P., & Fattori, P. (1991). Function properties of neurons in the anterior bank of the parietal-occipital sulcus of the macaque monkey. *European Journal of Neuroscience*, 3, 452–461.
- Galletti, C., Battaglini, P., & Fattori, P. (1995). Eye position influence on the parietal-occipital area po(v6) of the macaque monkey. *European Journal of Neuroscience*, 7, 2486–2501.
- Gnadt, J., & Breznen, B. (1996). Statistical analysis of the information content in the activity of cortical neurons. *Vision Research*, 36(21), 3525–3537.
- Gnadt, J., & Mays, L. (1989). Posterior parietal cortex, the oculomotor near response and spatial coding in 3-d space. *Society of Neuroscience Abstracts*, 15, 786.
- Gnadt, J., & Mays, L. (1995). Neurons in monkey parietal area lip are tuned for eye-movement parameters in three-dimensional space. *Journal of Neurophysiology*, 73(1), 280–297.
- Hallett, P., & Lightstone, A. (1976). Saccadic eye movements towards stimuli triggered by prior saccades. *Vision Research*, 16(1), 99–106.
- Krommenhoek, K., Opstal, A. V., Gielen, C., & Gisbergen, J. V. (1993). Remapping of neural activity in the motor colliculus: a neural network study. *Vision Research*, 33(9), 1287–1298.
- Krommenhoek, K., Opstal, A. V., & Gisbergen, J. V. (1996). An analysis of craniocentric and oculocentric coding stages in a neural network model of the saccadic system. *Neural Networks*, 9(9), 1497–1511.
- Mays, L., & Gamlin, P. (1995). Neuronal circuitry controlling the near response. *Current Opinion in Neurobiology*, 5, 763–768.
- Mays, L., & Sparks, D. (1980). Saccades are spatially, not retinocentrically, coded. *Science*, 208, 1163–1165.
- Moody, S. L., Wise, S., Pellegrino, G., & Zipser, D. (1998). A model that accounts for activity in primate frontal cortex during a delayed matching-to-sample task. *Journal of Neuroscience*, 18(1), 399–410.
- Moschovakis, A. (1987). *Observations on the appearance and function of neurons in the primate superior colliculus. An intracellular HRP study*. Ph.D. thesis, Washington University, St. Louis.
- Moschovakis, A. (1996). Neural network simulations of the primate oculomotor system. II. Frames of reference. *Brain Research Bulletin*, 40(5-6), 337–343.
- Moschovakis, A., Scudder, C., & Highstein, S. (1996). The microscopic anatomy and physiology of the mammalian saccadic system. *Progress in Neurobiology*, 50, 133–254.
- Platt, M., & Glimcher, P. (1998). Response fields of intraparietal neurons quantified with multiple saccadic targets. *Experimental Brain Research*, 121, 65–75.
- Robinson, D. (1975). Oculomotor control signals. In *Basic mechanisms of ocular motility and their clinical implications* (pp. 337–374). Oxford: Pergamon.
- Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In *Parallel distributed processing: explorations in the microstructures of cognition*, vol. 1 (pp. 316–362). Cambridge, MA: MIT Press.
- Sakata, H., Shibutani, H., Kawano, K., & Harrington, T. (1985). Neural mechanisms of space vision in the parietal association cortex of the monkey. *Vision Research*, 25(3), 453–463.
- Sakata, H., Shibutani, H., & Kawano, K. (1980). Spatial properties of visual fixation neurons in the posterior parietal association cortex of the monkey. *Journal of Neurophysiology*, 43(6), 1654–1672.
- Seung, H. (1998). Continuous attractors and oculomotor control. *Neural Networks*, 11, 1253–1258.
- Sparks, D. (1989). The neural encoding of the location of targets for saccadic eye movements. *Journal of Experimental Biology*, 146, 195–207.

- Sparks, D., & Mays, L. (1983a). Role of the monkey superior colliculus in the spatial localization of saccade targets. In A. Hein, & M. Jeannerod, *Spatially oriented behavior* (pp. 63–85). New York: Springer.
- Sparks, D., & Mays, L. (1983b). The spatial localization of saccade targets. I: Compensation for stimulation-induced perturbations in eye-position. *Journal of Neurophysiology*, *49*, 64–74.
- Squatrito, S., & Maioli, M. (1996). Gaze field properties of eye position neurones in areas mst and 7a of the macaque monkey. *Visual Neuroscience*, *13*, 385–398.
- Stricanne, B., Andersen, R., & Mazzoni, P. (1996). Eye-centered, head-centered, and intermediate coding of remembered sound locations in area lip. *Journal of Neurophysiology*, *76*(3), 2071–2076.
- Thier, P., & Andersen, R. (1998). Electrical microstimulation distinguishes distinct saccade-related areas in the posterior parietal cortex. *Journal of Neurophysiology*, *80*, 1713–1735.
- Williams, R., & Zipser, D. (1995). Gradient-based learning algorithms for recurrent networks and their computational complexity. In *Backpropagation: theory, architecture, and applications* (pp. 433–486). Hillsdale, NJ: Lawrence Erlbaum.
- Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *Journal of Neuroscience*, *16*(6), 2112–2126.
- Zipser, D. (1991). Recurrent network model of the neural mechanism of short-term active memory. *Neural Computation*, *3*, 179–193.
- Zipser, D. (1992). Identification models of the nervous system. *Neuroscience*, *47*, 853–862.
- Zipser, D., & Andersen, R. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, *331*, 679–684.