13

A Network Model for Learned Spatial Representation in the Posterior Parietal Cortex

RICHARD A. ANDERSEN
DAVID ZIPSER

Anatomists and neurophysiologists who study the cerebral cortex generally believe that they are studying a hard-wired, genetically determined circuitry. It is usually assumed that it is only during brief "critical periods" that the cortex is amenable to changes dictated by exposure to the environment which fine tune its structure. However, recent evidence suggests that the cortex is very plastic and that this plasticity extends into adult life (see Merzenich, Recanzone, Jenkins, Allard, & Nudo, in press, for review). In the experiments described in this chapter, combined neurophysiological and computational approaches were used to investigate how the posterior parietal cortex of macaque monkeys represents the location of visual stimuli in craniotopic coordinates (Andersen & Zipser, 1988; Zipser & Andersen 1988). A computer-generated network model was designed to learn stimulus locations in craniotopic space based on eye and retinal position inputs that were modeled on similar signals derived from the recording data and are assumed to be inputs to the posterior parietal cortex. The units in the network that map from input to output were found to develop the same response properties as a large subset of cells found in area 7a of the posterior parietal cortex. These experiments suggest that the spatial representation found in area 7a is in fact learned by associating eye position with retinal inputs.

SPATIAL REPRESENTATIONS AND THE ROLE OF THE POSTERIOR PARIETAL CORTEX

There are several reasons to believe that the brain uses representations of visual space that are nonretinotopic and are framed in head- or body-centered coordinates. One can reach accurately to the location of visual targets without visual

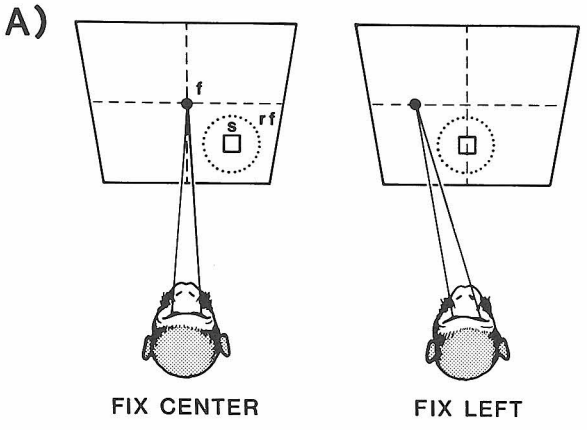
feedback and independent of eye position, head position, or the location of the image on the retinas. Thus the motor system appears to use representations of visual stimuli mapped in body-centered rather than retinal coordinates. There is substantial evidence that the planning of eye movements involves a stage in which the target of the eye movement is represented in craniotopic coordinates (Hallet & Lightstone, 1976; Mays & Sparks, 1980; Robinson, 1975). We know by introspection that the visual world appears perceptually stable in spite of the fact that we are constantly making eye movements and subsequently shifting the location of images on the retinas. All these results suggest that there exist neural representations of space that are head or body centered.

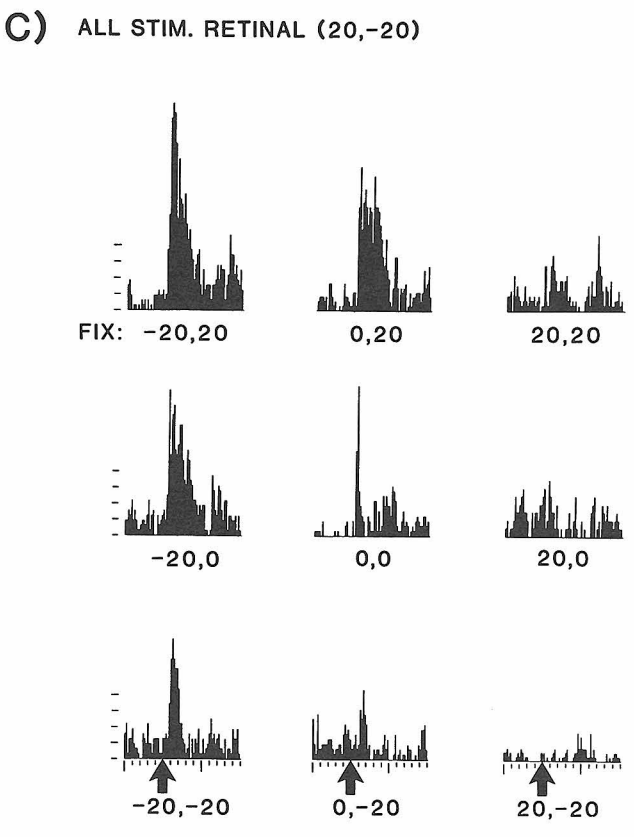
A likely area of the brain to find nonretinotopic representations of visual space is the posterior parietal cortex. Lesions in this area in humans and monkeys produce visual disorientation, a syndrome in which the subjects cannot reach accurately to visual targets and have difficulty navigating around seen obstacles (see Andersen, 1987, for review). The patients are not blind and when tested often have normal visual field functions. However, they appear to be unable to associate what they see with the positions of their bodies.

RECORDING DATA FROM AREA 7

We examined the coordinate frame for visual space used by the posterior parietal cortex by mapping visual receptive fields in this area with animals looking in different directions (Andersen, Essick, & Siegel, 1985). The animals' heads were fixed to simplify the coordinate space examined to a head-centered coordinate frame. We reasoned that if the receptive fields moved with the eyes, then the coordinate frame was retinotopic; if they remained static in space, then they were coding in at least craniotopic coordinates. Figure 13.1 shows an example of this experiment. The receptive field is first mapped with a flashed visual stimulus while the animal fixates a small fixation spot located straight ahead at 0,0 in screen coordinates. Figure 13.1B shows a typical receptive field mapped in this way where the axes represent screen coordinates and the contour lines different levels of neural response. Once the receptive field has been mapped, the stimulus is then presented at the retinotopic location that gave the maximum

FIGURE 13.1. (A) Method of determining spatial gain fields of area 7a neurons. The animal fixates point *f* at different locations on the screen with his head fixed. The stimulus, *s*, is always presented in the center of the receptive field, *rf*. (B) Receptive field of a neuron plotted in coordinates of visual angle determined with the animal always fixating straight ahead (screen coordinates 0,0). The contours represent the mean increased response rates in spikes per second. (C) Spatial gain field of the cell in (B). The poststimulus histograms are positioned to correspond to the locations of the fixations on the screen at which the responses were recorded for retinotopically identical stimuli presented in the center of the receptive field (histogram ordinate, 25 spikes per division, and abscissa, 100 msec per division; arrows indicate onset of stimulus flash). (From Andersen et al., 1985)


A)
f
s
rf
FIX CENTER
FIX LEFT
B)
40°
10
20
40


C)
ALL STIM. RETINAL (20,-20)
FIX: -20,20
0,20
20,20
-20,0
0,0
20,0
-20,-20
0,-20
20,-20

response, but with the animal gazing in different directions. If the response changes with eye position for retinotopically identical stimuli, then the cells are not coding in strictly retinotopic coordinates. Figure 13.1A shows how the visual response is tested at different eye positions. The dashed line *rf* delineates the receptive field that was mapped in Figure 13.1B. Stimulus *s* is presented at the most responsive location in the receptive field, in this case the approximate center of the response zone. On the left the animal is first required to fixate straight ahead at 0,0 in screen coordinates. On the right the animal has been required to fixate to the left of straight ahead by 20° at −20,0 in screen coordinates. Since the head is fixed the eyes are now in different positions in the orbits. The target is again flashed in the same retinotopic location; however, since the eyes have moved 20° to the left the stimulus has also been moved 20° to the left such that the stimulus falls on the same retinotopic location.

In these experiments nine eye positions are tested in this manner. Results from this cell are seen in Figure 13.1C. Each histogram is plotted at the corresponding fixation location. The cell was most active for fixations up 20° and left 20°, less active for looking straight ahead, and not active at all when looking down 20° and right 20°. The activity of the cell for retinotopically identical stimuli varied as a function of the angle of gaze. Plots for the nine fixation positions as shown in Figure 13.1C are called spatial gain fields. Notice that the shape of this particular gain field can be described by a plane tilted down and to the right.

We recorded complete spatial gain fields for 86 area 7a cells. The mean evoked responses of these gain fields were further analyzed using a first-order linear model with independent variables of horizontal and vertical eye position to determine how many of these gain fields could be fit with a plane. The evoked activity was obtained by subtracting the background activity before the stimulus flash from the overall activity during and just after the stimulus flash. Three types of gain fields were obtained by this analysis. Thirty-one percent of the gain fields had a significant planar component and no significant lack of fit, indicating that a plane was the best model for the data. Another 32% showed a planar component but also a significant lack of fit, indicating that although a plane could be fit to the data a plane was not the optimal model. Fully 75% of these neurons looked very planar. Finally, another 37% showed no planar component and a significant lack of fit. Thus a majority of cells showed planar or largely planar gain fields (55%), but a significant number of gain fields (45%) are not planar. When the same cells were analyzed for their overall activity rather than just the evoked activity, it was found that a larger proportion of the cells (78%) had planar or largely planar gain fields. Thus the total signal of background activity and evoked activity shows a greater degree of planarity than just the evoked response. The total activity is the most likely signal used by the brain for spatial localization since it is the output of the neurons.

A major concern is that the visual background, which is imaged at different locations on the retinas at different angles of gaze, is influencing the responsiveness of the cells to the test flash. Two controls were performed to eliminate this possibility. Many of the recordings were made in complete darkness except for

the stimulus and the fixation point so that there was no visual background. The second control was to change the angle of gaze using prisms; this requires the animal to change eye position without changing the retinal locations of the imaged background or visual stimulus. The cells showed the same gain fields whether the eye positions were changed with prisms or without prisms by moving the fixation point. Thus we can conclude that the effect of eye position on visual responses was not a result of background shifts.

EYE POSITION–DEPENDENT SPATIAL TUNING

The change in visual responses for retinotopically identical stimuli with eye position could be a result of two mechanisms. First, the cells could have been coding locations of targets in space independent of eye position. In this situation the cells' receptive fields remain static in space and the retinal addresses of the receptive fields change with eye movements to remain constant for spatial location. Second, the receptive fields could remain retinotopic, with only the responsiveness of the cells varying as a function of eye position. In other words, eye position gates the activity of the retinal receptive fields. To distinguish between these two possibilities we mapped entire axes through the center of the receptive fields. It was found that the responsiveness of neurons varies as a function of eye position, but the peaks and symmetry of their receptive fields do not change. Thus the receptive fields remain retinotopic and it is only the responsiveness of the cells that is modulated by eye position.

The activity of parietal neurons was modeled as a multiplicative interaction of eye position and retinal position using the equation $A = G(e_x,e_y) \times R(r_x,r_y)$ where A is the cells' firing rate, G is a gain factor that is a function of horizontal (e_x) and vertical (e_y) eye position, and R is the visual stimulus response profile, which is a function of horizontal (r_x) and vertical (r_y) retinal locations. This multiplicative interaction produced a tuning for the location of visual stimuli in craniotopic space, but it was dependent on eye position. A simple example of this would be a cell that has a gain of 0 for all eye positions except looking 10° to the right, in which case the gain is 1. This cell also has a narrow receptive field centered at 10° to the right of the fovea. This cell will then be tuned to a location 20° to the right in craniotopic coordinates, but only when the animal is looking 10° to the right.

PROBLEMS TO BE ADDRESSED BY MODELS OF SPATIAL REPRESENTATION IN AREA 7a

Cells have never been found that are spatially tuned in an eye position–independent manner. From these experiments it must be concluded that the neural representation of space in an eye position–independent manner is distributed. This distributed coding presents a problem: How do you determine eye posi-

tion–independent spatial location from the population response? One way would be to map spatial location tuning systematically across the tangential dimension of area 7a. Recording experiments have so far not revealed any obvious topography for spatial tuning, indicating that if such a map for craniotopic space does exist, it will likely be crude. Moreover, the very large receptive fields and spatial tuning fields would tend to mitigate against high-resolution mapping like that found for retinotopy in V1.

Another unusual feature of the receptive fields of area 7a neurons is their complexity. The fields are large, have approximately equal weighting to the fovea and periphery, and can have multiple peaks. The fields generally have smoothly varying levels for response for nearby sample points. No topographic organization for retinal location of the receptive fields has been found. Any model for spatial representation in area 7a should reproduce these unusual visual receptive fields.

Another interesting aspect of the area 7a neurons is that the background activity of the cells also often varies as a function of eye position. In many cases the background activity varies in the same direction as the gain on the visual response. This result is not unexpected since the eye position input could increase the visual response by depolarizing the membrane and in some cases this depolarization would not only lower the cell's threshold to visual stimulation but may also fire the cell, leading to an increase in its background activity. However, other cells showed background effects that went in the opposite direction to the gain of the visual response. Again, any model of spatial coding in area 7a would need to explain this behavior.

NETWORK MODEL FOR SPATIAL REPRESENTATION

Zipser and Andersen created a parallel network model that learns to map inputs of retinotopic position and eye position to an output of location in head-centered space (Andersen & Zipser, 1988; Zipser & Andersen, 1988). This network consists of three layers and uses the backpropagation learning algorithm (Rumelhart, Hinton, & Williams, 1986). The units in the middle layer that accomplish the spatial transformation show the same eye position–dependent spatial tuning properties that are found for area 7a neurons. This model also generates retinal receptive fields similar to those found for area 7a neurons and reproduces similar background activities. The remarkable correspondence between the model and experiment suggests that the distributed spatial coding discovered in area 7a neurons is indicative of spatial transformations carried out using the same computational algorithm discovered by the backpropagation learning technique.

Figure 13.2 presents a schematic diagram of the network. The input layer consists of a 10 by 10 retinal array and four eye position units. The retinal receptive fields are Gaussian in shape with $1/e$ widths of 15°. This input is designed to be similar to the receptive fields that are found in area 7a that do not show eye

position effects and are assumed to be the retinal inputs to area 7a. The centers of the 100 receptive fields are equally spaced over the 10 by 10 grid with 8° spacings. The four eye position units consist of two units coding vertical position and two horizontal position using opposite, symmetrical slopes. Each unit used either a linear or, in later simulations, a squared function to approximate the signal coming from eye position cells. The rationale for using a squared function is to approximate the cumulative response of a group of eye position cells. Eye position inputs are assumed to be those cells in area 7a that have only eye position signals and no visual response. These cells generally code horizontal and/or vertical position in a linear fashion (Andersen & Zipser, 1988). We ran simulations with both square functions and simple linear functions and got indistinguishable results, indicating that the exact representation of the eye position does not appear to be crucial as long as it is a monotonically increasing function.

The intermediate layer receives inputs from all 104 input units and in turn projects to two or four output units. The output units code position in head-

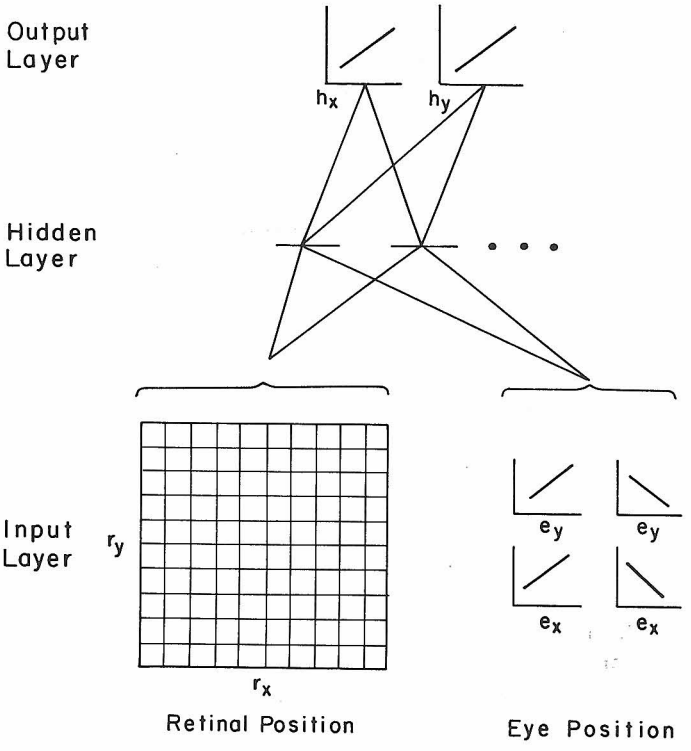


FIGURE 13.2. Backpropagation network used to model area 7a. The input to the network consists of retinal position and eye position information. The activity of the output units is a monotonic function of the location of the visual stimulus in craniotopic coordinates. The middle or "hidden" layer units map input to output. The details of the network are explained in the text.

centered space as a linear function of spike rate. There are four output units with pairs of opposite slope for horizontal and vertical position. As in the case with the eye position inputs, the exact form of the function did not appear to matter as long as it was monotonically increasing; also, whether only one unit of positive slope or two units of opposite slope were used did not seem to be important. The rationale for using a monotonic function for the output was that the eye position cells could be used as a teaching signal if the animal saccaded to fixate the stimulus. In later experiments we also tried mapping to Gaussian representations of head-centered location; the interesting results are listed below.

The output of each cell in the network is calculated by first summing all inputs, both inhibitory and excitatory, and then calculating the output as a sigmoidal function of the input. A sigmoid is chosen as an output function since it is similar to the operation performed by actual nerve cells that sum inputs, have a threshold, and saturate at high levels of activity. There is also a threshold term that can be either trained or set, and simulations using both of these options will be discussed.

The network begins training with all the connections set to random weights and completes training when the output units accurately indicate positions in head-centered space for any pair of arbitrary retinal and eye position inputs. The network learns by subtracting the output vector from the desired output vector for each input pattern to generate an error. This error signal is then propagated back through the network to change the weights in the network. The back-propagation algorithm ensures that the weights will change to reduce error in the performance of the network. The actual equations and derivation of the back-propagation procedure are discussed in Rumelhart et al. (1986). This cycle is repeated until the network reduces error to desired levels. The spatial transformation network learns quickly and always settles to very low error values. Within 1000 trials the network shows accuracy that is better than the spacing of the distance between the centers of the retinal receptive fields. For large numbers of trials the network continues to show improvement to vanishingly small errors.

After training is complete the middle-layer units have receptive fields that remain retinotopic, but their activity becomes modulated by eye position in a manner similar to that seen in the recording data from area 7a neurons. The receptive fields remain retinotopic but the responsiveness changes with eye position. The change in responsiveness is roughly planar and similar to a majority of the gain fields recorded from area 7a neurons. Moreover, the receptive fields are large and can have peculiar shapes, not unlike the cells in area 7a.

In Figure 13.3 the retinotopic visual receptive fields from recording experiments are compared with retinotopic receptive fields generated by the model. It should be emphasized that these comparisons are intended to be qualitative and show only that they are similar; obviously they will not be exactly the same, just as no two area 7a neuron receptive fields will ever be identical. Surfaces have been fit to the recording and model data using a Gaussian interpolation algorithm. Each receptive field is 80° in diameter. The fields have been categorized

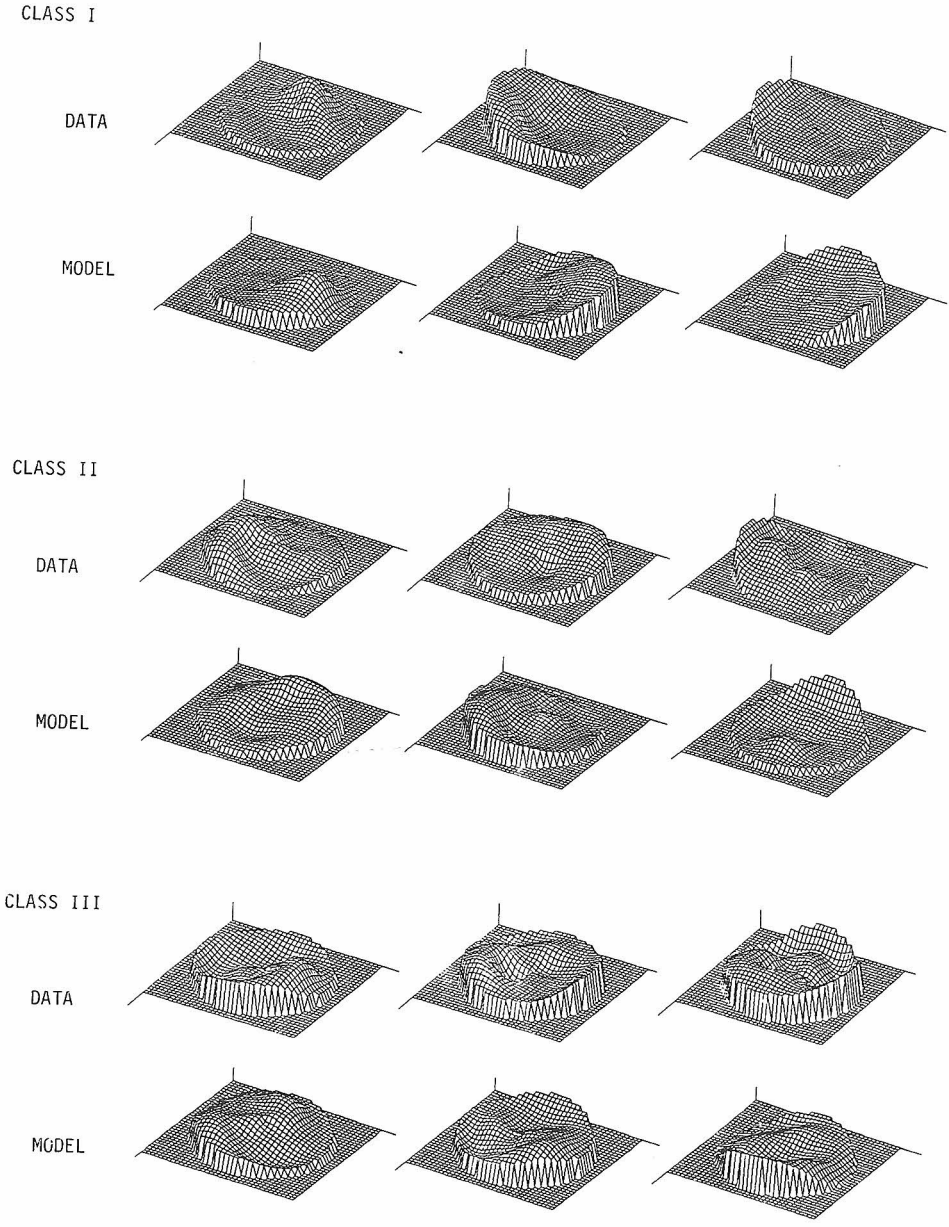


FIGURE 13.3. Visual receptive fields from the data and the model compared. The receptive fields were divided into three classes: class I cells have a single, smooth peak of activity; class II cells have one peak of activity but also other smaller peaks or depressions in the receptive field; class III cells have multiple peaks of activity. Note the close correspondence between model- and data-receptive fields.

by complexity into three classes: class I has the simplest receptive fields with each having a single peak of activity; class II fields are of intermediate complexity with each field having a single greatest peak of activity and one or more smaller peaks of activity; class III receptive fields are the most complex with each having multiple peaks of greatest activity. The most complex fields are similar to fields of the untrained model. The trained model seldom produces such complex fields.

Next we compared gain fields generated by the model with data gain fields. Figure 13.4 shows gain fields from recording data and Figure 13.5 gain fields

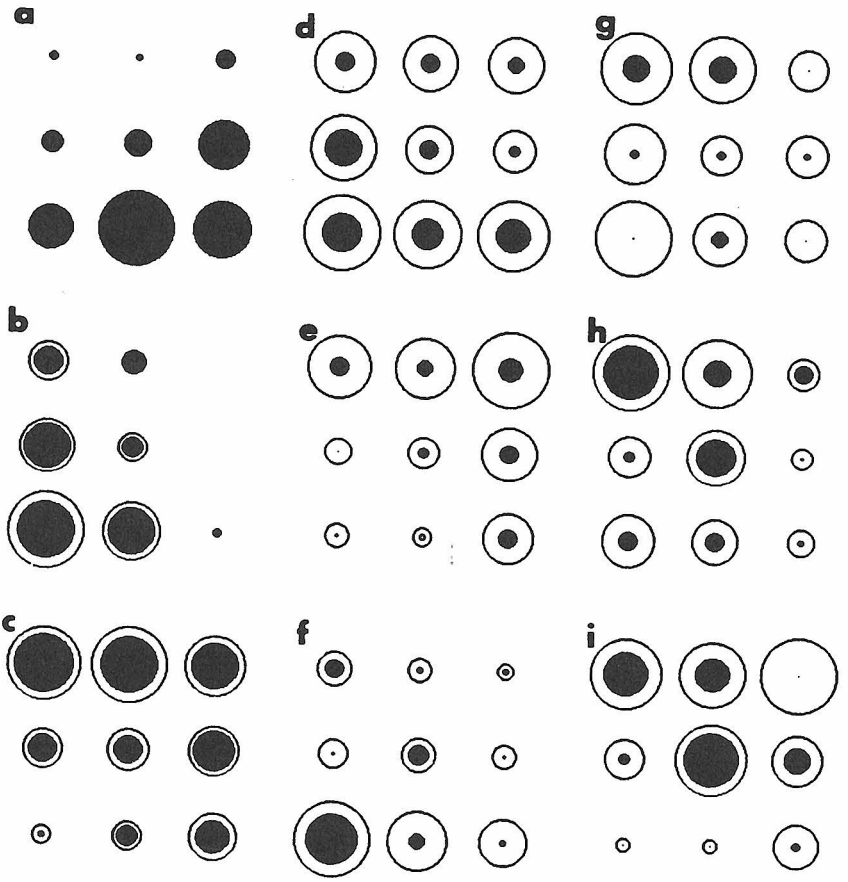


FIGURE 13.4. The spatial gain fields of nine cells recorded from area 7a using the technique shown in Figure 13.1. The diameter of the inner, dark circle is proportional to the magnitude of the visually evoked response. The outer circle diameter is proportional to the total rate (visual response and background activity). The white annulus represents the firing rate in the absence of visual stimulation.

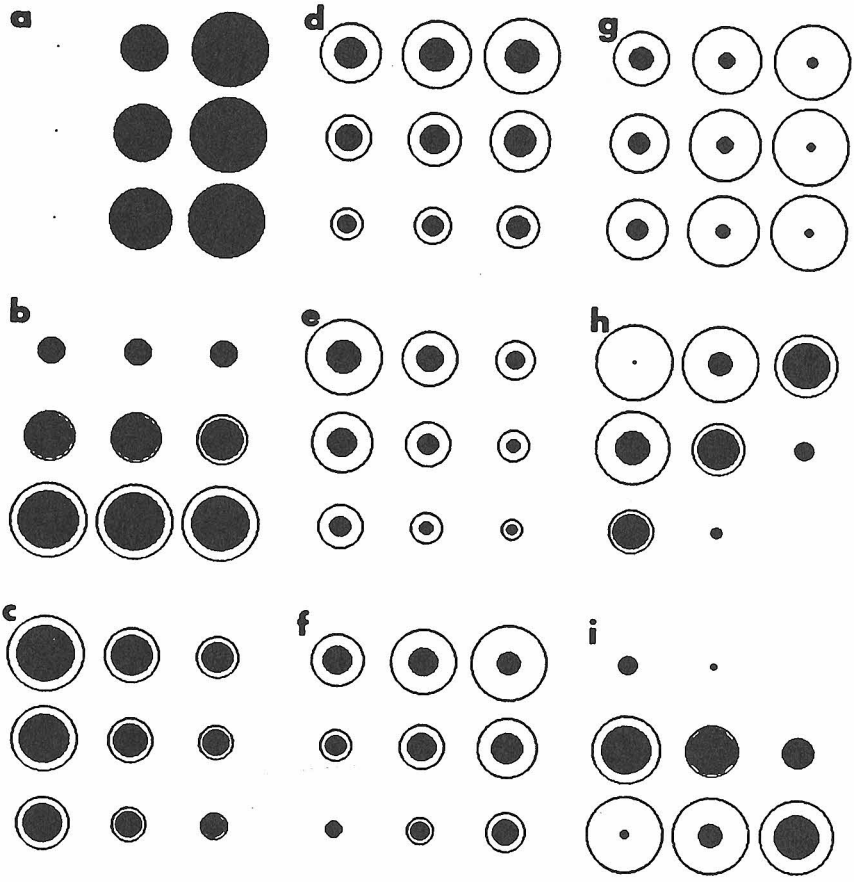


FIGURE 13.5. Hidden unit spatial gain fields generated by the network model, in the same format as Figure 13.4. All the fields illustrated resulted from 10,000 training trials. Units g, h, and i are from runs in which the craniotopic Gaussian format for the output layer was used. The remainder of the cells were using a craniotopic monotonic format.

generated by the model's hidden unit. The nine pairs of circles for each data or model unit represent the activity for the same retinotopic stimulus delivered at nine different eye positions. The dark inner circle's diameter is proportional to the response to the visual stimulus alone, and the outer circle's diameter is proportional to the entire activity, both background and evoked.

Three basic types of planar gain fields were found from the recording data. For 28%, the background and evoked activities changed in a parallel fashion (Figure 13.4b,e,f). In the largest proportion of cells (43%), the evoked activity changed while the background activity, if any, remained constant (Fig. 13.4a,c,d); 75% of these cells had low rates of background activity. For the

remaining 28% of the neurons the background and evoked activities changed in different directions, so the activity of either alone was grossly nonplanar, but the overall activity was planar (Fig. 13.4g,h,i).

Figure 13.5 shows gain fields generated by the model; the arrangement shows the close correspondence between recording and model gain fields. Both monotonic and Gaussian output representations were used in the training that generated these fields. Either output representation generates total response gain fields that are planar. However, when the training is made to monotonic output function the visual response gain fields are generally planar (67%; Fig. 13.5a–f), whereas training to the Gaussian format produces only 13% in this class. These figures compare to the 55% found in the experimental data. If the thresholds are trained along with the synaptic weights for monotonic outputs, then the background and evoked activities almost always change in a parallel fashion. However, if the threshold is held constant and at high values, the background activity is low and does not change with eye position, similar to the largest proportion of parietal cells. Finally, if training is done using a Gaussian format output, then the visual evoked response is usually grossly nonplanar but the combination of background and evoked responses is planar (Fig. 13.5g,h,i).

DISCUSSION OF THE MODEL

The simulation results show that training a parallel network to perform coordinate transformations produces the same type of distributed code that is found in area 7a. This similarity should be pursued to determine if it represents a fundamental outcome of using parallel networks to perform coordinate transformations. One line of research would be to make the model more complex by incorporating features analogous to those found in the brain such as Hebb learning and reciprocal pathways for error feedback. It will be interesting to see if these more complex and brainlike models still produce the same distributed code. Another avenue would be to see if this model generalizes to three-dimensional space- and body-centered coordinates by collecting data under these conditions for parietal neurons and comparing the results to predictions made from the model.

These results suggest that the posterior parietal cortex learns to associate body position with respect to visual space. Thus the parietal lobe appears to form associative memories for performing spatial transformations. A learning theory for parietal spatial functions seems in order, since it would not be practical to hard-wire spatial representations during development when the body is changing size. Moreover, distortions of space with prisms lead to rapid recalibration in adults, suggesting that, unlike ocular dominance, there is no critical period for spatial representations and they remain plastic in adults.

It is important to note that the model by definition does not have a topographic organization. Thus there is no requirement for topographic organization in the brain. The reason topography is not necessary is that the organization of

the network is distributed and the information is contained in the weights in the synapses. It would be interesting to determine whether putting a crude topography into the connections of the network would accelerate learning.

If the spatial representation in area 7a lacks topography, this does not of course mean that there is not topographic organization in this area. We imagine that learning spatial localization can occur within the dimensions of a typical cortical column, 1 mm^2. Recordings made within an area of parietal cortex of this size contain a complete complement of receptive fields, eye position signals, and gain fields necessary for a complete representation of craniotopic space (Andersen, Essick, & Siegel, unpublished observation, 1985). Thus spatial location can be mapped over and over again in many repeating units in the cortex and may overlie some as yet unknown functional repetitive architecture that would need in each of its modules the complete machinery for coordinate transformations.

The rather large receptive fields suggest that every posterior parietal neuron has access to the entire retina and the particular shape of each receptive field is a result of competitive learning. This competition produces in both the parietal neurons and the model units receptive fields that, although they are complex, tend to coalesce so that they are smoothly varying rather than random in structure. The large receptive fields are presumably due to the cascading divergence that occurs in the multistage corticocortical projections from V1 to area 7a.

The fact that many fewer eye position synapses than retinal synapses were requried at the convergence onto the hidden units in the model has interesting parallels to the anatomy of the parietal lobe connections. It is believed that the source of eye position information comes from lower brain stem centers and is relayed through the intralaminar nuclei (Schlag-Rey & Schlag, 1984). However, these nuclei are small compared to the cortical areas relaying visual information to area 7a.

There is the question of where the output units of the model might exist in the brain. Cells showing the expected eye position–independent behavior have not been found in area 7a. Another possibility is that this distributed coding is used throughout all brain regions that need spatial representations. The final spatial output might be seen only in the eventual motor output. The final spatial output may be pointing the eye or a finger accurately to a location in space and no single cell in the brain might be found that codes the location of visual space in an eye position–independent fashion.

Finally, there is the question of whether this form of model of distributed coding in parallel networks, which appears to explain the parietal data rather well, will be useful in other brain regions. Recently the response of area V4 cells was examined in a task in which the monkey must match the orientation of a visual or somatosensory cue grating with the orientation of a visual test grating (Haenny, Maunsell, & Schiller, in press). They find cells that respond to the cue that are orientation tuned, cells that respond to the test stimulus that are orientation tuned, and cells that show facilitated activity for a particular combination of cue and test stimulus. Thus the activity of some V4 neurons shows a

multiplicative interaction between two inputs that is similar to the interaction for eye and retinal inputs for area 7a neurons. It would be useful to construct a similar network in which the inputs were the cue and test stimuli and the output the correct match. Would the hidden units develop properties like those of V4 neurons?

Acknowledgments

We thank Carol Andersen for editorial assistance. R.A.A. was supported by grants EY05522 and EY07492 from the National Institutes of Health, the Sloan Foundation, and the Whitaker Health Sciences Foundation. D.Z. was supported by grants from the System Development Foundation, ONR contract N0014-87K-0671, and the AFOSR.

REFERENCES

Andersen, R. A. (1987). Inferior parietal lobule function in spatial perception and visuomotor integration. In F. Plum (Ed.), *Handbook of physiology* (pp. 483–518). Bethesda, Md.: American Physiological Society.

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science, 230,* 456–458.

Andersen, R. A. & Zipser, D. (1988). The role of the posterior parietal cortex in coordinate transformations for visual-motor integration. *Canadian Journal of Physiology and Pharmacology, 66,* 488–501.

Haenny, P. E., Maunsell, J. H. R., & Schiller, P. H. (in press). State dependent activity in monkey visual cortex. *Experimental Brain Research.*

Hallet, P. E., & Lightstone, A. D. (1976). Saccadic eye movements toward stimuli triggered by prior saccades. *Vision Research, 16,* 99.

Mays, L. E., & Sparks, D. L. (1980). Dissociation of visual and saccade-related responses in superior colliculus neurons. *Journal of Neurophysiology, 43,* 207–232.

Merzenich, M. M., Recanzone, G., Jenkins, W. M., Allard, T. T., & Nudo, R. J. (in press). Cortical representational plasticity. In P. Rakic & W. Singer (Eds.), *Neurobiology of neocortex. Dahlem konferenze.* Chichester: Wiley.

Robinson, D. A. (1975). Oculomotor control signals. In P. Bach-y-Rita & G. Lernerstrand (Eds.), *Basic mechanisms of ocular motility and their clinical implications* (pp. 337–374). London: Pergamon.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 318–362). Cambridge, Mass.: Bradford Books/MIT Press.

Schlag-Rey, M., & Schlag, J. (1984). Visuomotor functions of central thalamus in monkey. I. Unit activity related to spontaneous eye movements. *Journal of Neurophysiology, 51,* 1149–1174.

Zipser, D., & Andersen, R. A. (1988). A back propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature. 331,* 679–684.