

Egocentric Spatial Representation in Early Vision

Alexandre Pouget, Stephen A. Fisher,* and Terrence J. Sejnowski

Howard Hughes Medical Institute, and
The Salk Institute

Abstract

■ Recent physiological experiments have shown that the responses of many neurons in V1 and V3a are modulated by the direction of gaze. We have developed a neural network model of the hierarchy of maps in visual cortex to explore the hypothesis that visual features are encoded in egocentric (spatiotopic) coordinates at early stages of visual processing. Most psychophysical studies that have attempted to examine this question have concluded that features are represented in reti-

nal coordinates, but the interpretation of these experiments does not preclude the type of retinospatiotopic representation that is embodied in our model. The model also explains why electrical stimulation experiments in visual cortex cannot distinguish between retinal and retinospatiotopic coordinates in the early stages of visual processing. Psychophysical predictions are made for testing the existence of retinospatiotopic representations. ■

INTRODUCTION

The three most common types of coordinates used for representing visual objects in biological and computer vision systems are eye-centered (retinal), object-centered, and viewer-centered (egocentric). Eye-centered representations are consistent with many physiological and psychophysical studies of early vision. Neurons in primary visual cortex and most extrastriate areas are organized into retinotopic maps (Felleman & Van Essen, 1991). Most psychophysical experiments designed to determine the nature of spatial representation at early visual stages have reached the same conclusion, as we review in the discussion.

Other types of representation are believed to be used at the highest stages of visual processing. Mishkin, Ungerleider, and Macko (1983) proposed a functional distinction between two main streams of processing, the "what" and "where" pathways leading, respectively, in the temporal and parietal cortex. Object-centered reference frames have been suggested for the representation of objects in the inferior temporal cortex and egocentric reference frames have been proposed for the representation of spatial location in the parietal cortex (Andersen, 1989).

Goodale and Milner (1990) recently proposed that the dorsal pathway to parietal cortex could also be involved

in object manipulation as opposed to just localization. They suggest that the "where" pathway might be better called the "how" pathway. Egocentric coordinates are natural ones for object manipulation since they directly provide the position of an object from the viewer.

In this paper we raise the possibility that another type of representation that combines aspects of retinotopic and egocentric coordinate systems may be used in the early stages of processing in visual cortex. The term "egocentric" is commonly used in the literature for any set of coordinates whose axes and origin are fixed with respect to some part of the body, except the eye. One example is head-centered coordinates, which, as their name indicates, are fixed with respect to the head. The head-centered position of an object is given by (see Fig. 1):

$$\vec{P} = \vec{R} + \vec{E} \quad (1)$$

where \vec{P} is a 2-D vector whose two components are the horizontal and vertical angular position of the object in body-centered coordinates, \vec{E} is a similar vector for the eye position, and \vec{R} is another 2-D vector for retinal location of the image of the object. If head position, \vec{H} , were added to the right side of Eq. (1), the position would be in body-centered coordinates.

Physiological and psychological data support the existence of head-centered representations. However, in none of these experiments was the head or body position varied, so these could be body-centered or even spatio-

*Present address: Department of Psychology, Yale University.

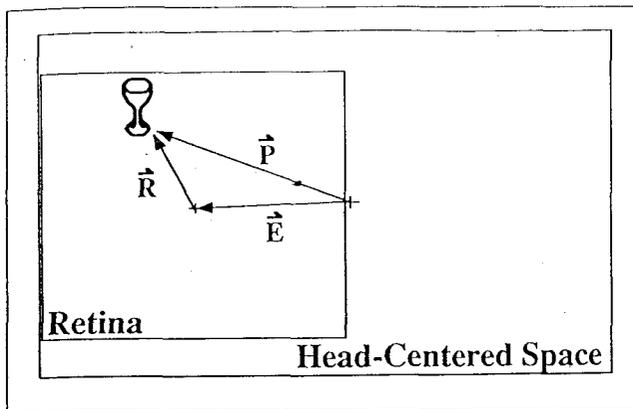


Figure 1. The angular position of an object with respect to an origin fixed in head-centered space, \vec{P} , is the vector sum of the retinal position vector, \vec{R} , and the eye position vector, \vec{E} , which joins the origin in head-centered space to the origin on the retina.

topic. We use the term "egocentric" to cover all of these possibilities. The term "spatiotopic" is used in the psychophysics literature in the same sense as "egocentric," which we will also adopt. Strictly speaking, "spatiotopic" refers to coordinates fixed in space, but if the head and body are kept fixed, "egocentric" and "spatiotopic" coordinates are equivalent.

The most compelling case for representations in egocentric coordinates can be made in the posterior parietal cortex (PPC), which we review here before considering early visual areas. Following ischemic lesions in the PPC, patients often display neglect that is restricted to a particular region of egocentric space, though of variable extent (Heilman, Watson, & Valenstein, 1985). This neglect usually extends to multiple sensory modalities, typically vision, audition, and touch, on the side contralateral to the lesion.

Single cell recording in the PPC has shown that many neurons integrate retinal position with eye and head position, as expected if this area computes body-centered position (Andersen, Essick, & Siegel, 1985; Brotchie & Andersen, 1991). In the simplest type of spatiotopic representation, neurons would have receptive fields fixed in space, independent of eye position. We refer to this kind of representation as a spatiotopic map. The spatial representation found in the PPC is, however, not of this type. Many neurons in the PPC have retinal receptive fields but none has been found to be spatiotopic (Andersen et al., 1985; Brotchie & Andersen, 1991), although the magnitudes of the responses to visual stimuli of many neurons in PPC are modulated by both eye and head position. Zipser, Andersen, and Goodman developed network models showing that these response properties are consistent with a distributed representation of egocentric space (Andersen & Zipser, 1988; Goodman & Andersen, 1989, 1990; Zipser & Andersen, 1988).

The Zipser and Andersen network had three layers of processing units: two sets of input units, a layer of hidden

units with larger receptive fields, and a set of output units representing egocentric position (Andersen & Zipser, 1988; Zipser & Andersen, 1988). One set of input units was organized in a retinotopic map and a second set carried eye position. They trained the network on the task of spatial transformation from retinocentric to egocentric coordinates using backpropagation (Rumelhart, Hinton, & Williams, 1986) and showed that the hidden units of three-layer networks exhibited response properties similar to those of neurons in posterior parietal area 7a and the lateral intraparietal area (LIP). Each hidden unit encoded eye position with a rate code, that is, the rate of firing in the absence of visual stimulation was a monotonic function of eye position. Retinal position, on the other hand, was encoded through the spatial organization of the receptive fields of the hidden units (Zipser & Andersen, 1988).

The main point from the model was that no single hidden unit unambiguously coded the egocentric position, although they all carried partial information about both the eye and retinal position. The hidden layer of the network contained what is called a distributed representation of egocentric position. Given the striking similarities between the response properties of neurons in LIP and 7a and the hidden units in the Zipser and Andersen network, it is possible that the PPC has a distributed representation of egocentric position similar to the one generated within their network.

The study of saccadic eye movements elicited by electrical stimulation also supports the hypothesis that the PPC contains an egocentric representation of space. In LIP, the amplitude, and to a certain extent the direction, of the saccade evoked by focal electrical stimulation appears to be sensitive to the initial eye position (Fig. 2B) (Kurylo & Skavenski, 1991; Shibutani, Sakata, & Hyvarinen, 1986; Thier & Andersen, 1992). These saccades are consistent with the distributed representation found in the Zipser and Andersen model of the PPC (Goodman & Andersen, 1989). Stimulation of the ventral intraparietal area (VIP) elicits saccades that tend to converge in a particular region of the visual field regardless of the initial eye position (Kurylo & Skavenski, 1991; Shibutani et al., 1986; Thier & Andersen, 1992) (Fig. 2C). In this case, however, the correspondence with single cell responses is not as clear as in area LIP since static eye position does not seem to exert a gain control on VIP neurons. In contrast to these results in the PPC, eye movements of the fixed vector type have been observed after stimulation of the frontal eye fields (FEF) (Robinson & Fuchs, 1969), suggesting that the PPC and the FEF have different representation of eye movements (Pouget & Sejnowski, 1992).

Recent physiological studies suggest that neurons in the earliest stages of visual processing may be encoding spatial representations similar to those found in parietal cortex. Neurons sensitive to eye position have been reported in the lateral geniculate nucleus (LGN) (Lal &

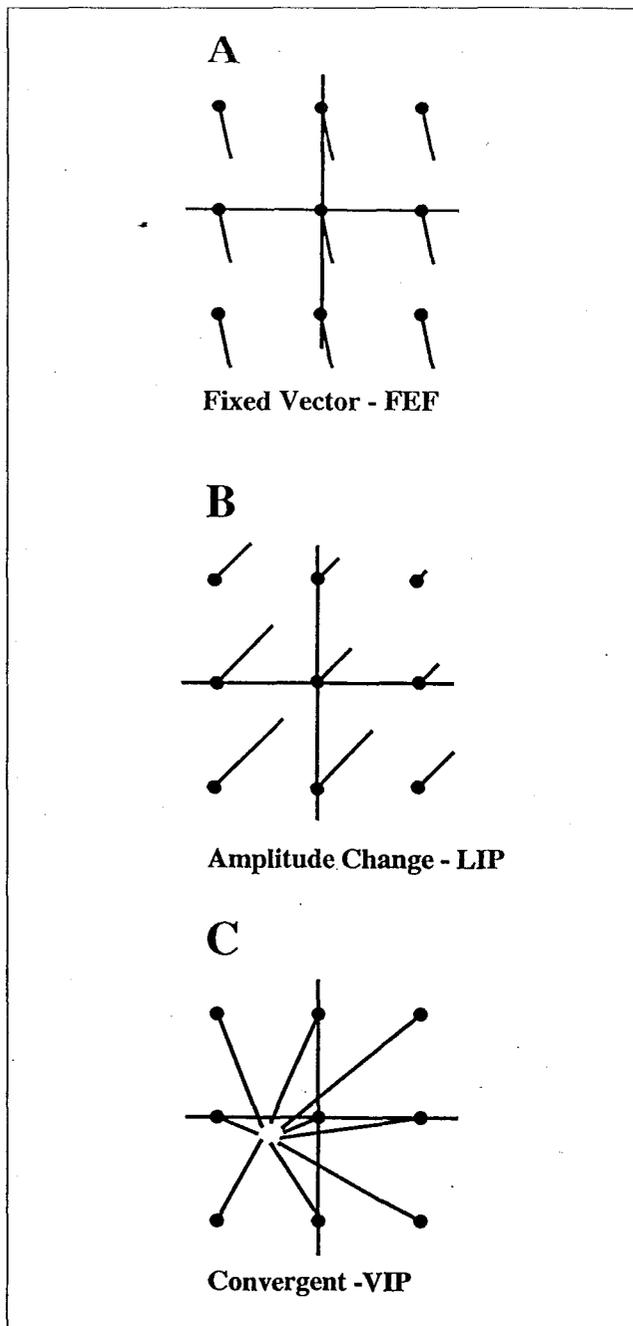


Figure 2. Schematic of three main patterns of saccadic eye movements evoked by electrical stimulation. The large filled circles indicate the initial eye position while the lines show the amplitude and direction of the evoked saccades. (A) Fixed vector type. Amplitude and direction are independent of initial eye position. Typical of stimulation in frontal eye field (FEF) (Robinson & Fuchs, 1969). (B) Amplitude varying saccades found in area LIP (Thier & Andersen, 1992). Amplitude varies with initial position of eye. (C) Convergent saccades. All saccades end in the same zone regardless of initial eye position. Such eye movements have been reported following stimulations in ventral intraparietal area (VIP) (Thier & Andersen, 1992).

Friedlander, 1989), primary visual cortex area V1 (Trotter, Celibrini, Stricanne, Thorpe, & Imbert, 1992; Weyand & Malpeli, 1989), and extrastriate area V3a (Galletti & Battaglini, 1989). Some of these eye position signals prob-

ably arise from proprioceptive afferents from the extraocular muscles (Ashton, Boddy, & Donaldson, 1984; Buisseret & Maffei, 1977), but efference copy of motor signals may also contribute. The eye position signal seems mainly to control the gain of the neuronal response without changing the selectivity of the cell. Could early cortical visual areas, or even the LGN, be using egocentric representations, as in the PPC? The small receptive fields and the strict retinotopy of those early maps argue against this possibility and psychophysical observations, summarized in the discussion, have led to the same conclusion. Furthermore, strong negative evidence comes from a study of saccadic eye movements elicited by electrical stimulation of area V1 (McIlwain, 1988).

McIlwain has shown that the directions of electrically evoked saccades in cats appear to be mainly a function of the position of the stimulation site in V1 and largely independent of the initial eye position. For example, when a position corresponding to the left part of the visual field along the horizontal meridian is stimulated, the eyes move horizontally toward the left (Fig. 3, left), and eye movements in the opposite direction are induced if neurons at an equivalent position on the right side are stimulated (Fig. 3, right). These are the eye movements expected for an attempt to foveate an illusory object whose position was at the site of stimulation as given in eye-centered coordinates. In contrast, electrical stimulation in area 7a produces convergent eye movements (Thier & Andersen, 1992) (see Fig. 2C).

We attempt to reconcile the conclusions drawn from electrical stimulation experiments with the gain modulation of neurons reported in the LGN, V1, and V3a. We have developed a neural network model that computes the egocentric position of objects using an architecture

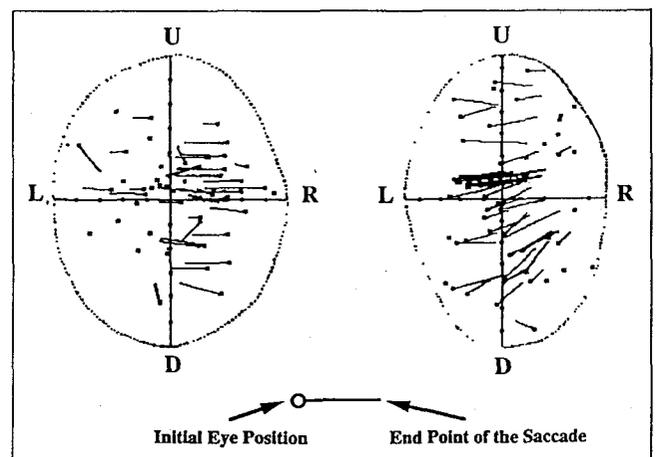


Figure 3. Saccadic eye movements induced by electrically stimulating primary visual cortex in a cat. Left plot: Stimulation on a site in cortex representing a position in the visual field 24° left 4° down. Right plot: same for a position 55° right, 12° up. In both cases, the direction of the saccades was primarily determined by the position of the stimulation on visual cortex and is largely independent of the initial eye position (McIlwain, 1988).

similar in several essential aspects to that of early visual cortex. We conclude that the recent physiological studies that have revealed gain modulation with eye position in the LGN, V1, and V3a are consistent with an egocentric representation despite the fact that those areas are retinotopic. Using the model, we explain why fixed vector eye movements in response to electrical stimulation are to be expected even if V1 uses egocentric coordinates.

RESULTS

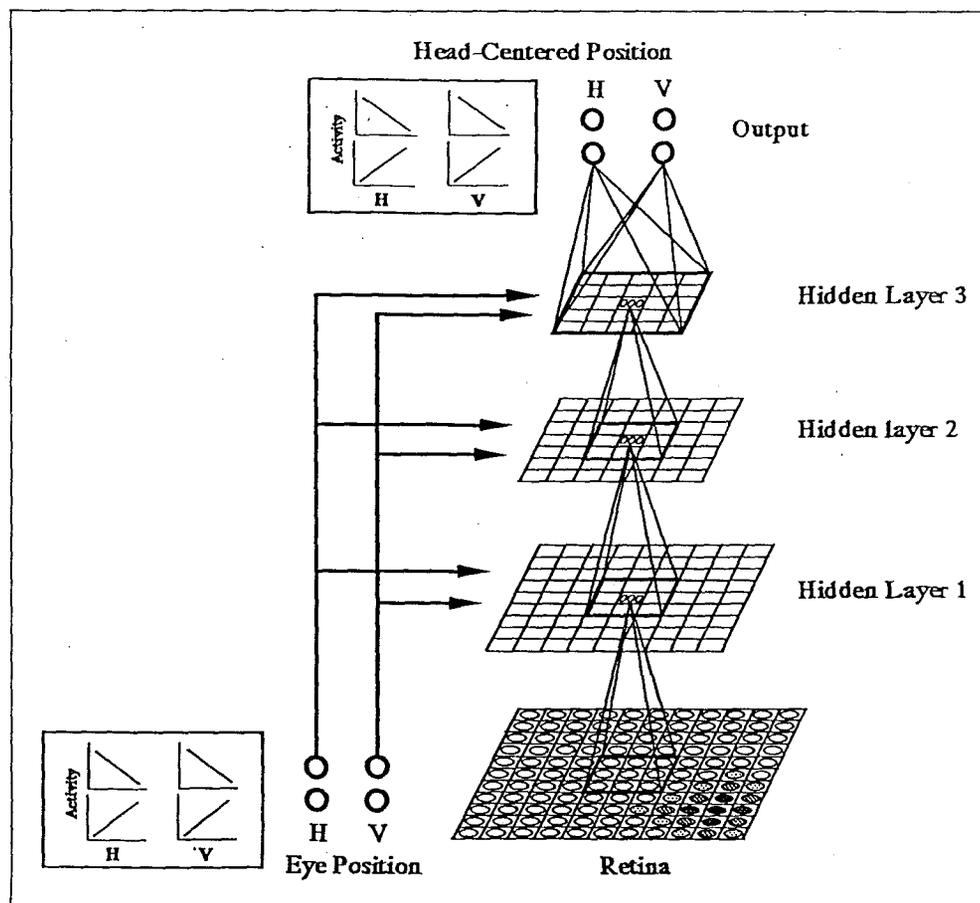
Network Architecture

The task of the network was to compute the egocentric coordinates of a single object from its retinal position and the position of the eye, as indicated in Eq. 1. The input layer had two groups of units (Fig. 4), one for eye position and the other for retinal position. Four units encoded the 2-D eye position, two for the horizontal and two for the vertical. The activity of these units was directly proportional to eye position: each pair had one unit with a positive slope and one with a negative slope. The group of retinal input units was organized in an 11×11 two-dimensional grid. The network architecture incorporated two basic features of cortical architecture: (1) restricted receptive fields organized in retinotopic maps and (2) the sizes of the receptive fields increased as function of the number of synapses away from the retina.

The hidden layers were organized as a feedforward series of retinotopic maps (Fig. 4). Three to five hidden units were located at each position of each map. A hidden unit received connections from the four eye position units and from a limited number of units centered around the corresponding location from the map directly below. Sparse connectivity and limited receptive fields are found in the early stages of processing in visual cortex. Our model is for a small portion of the visual field, so that the cortical representation can be assumed to be isotropic within each layer. Thus, within a given layer, all receptive fields were of the same size and the weights were shared such that the receptive field properties of units of the same type at two different locations on a map were identical. This translational invariance is a hallmark of the columnar organization of visual cortex. It is important to note that the three to five hidden units at each location had different weights, hence different receptive fields. Thus, this small group of units could be considered a highly simplified cortical column. Details about this weight-sharing method can be found in LeCun, Boser, Denker, Henderson, Howard, Hubbard, and Jackel (1990).

Weights were adjusted using the backpropagation training algorithm (Rumelhart et al., 1986). All simulations were performed with the SN2 simulator developed by L. Botou and Y. LeCun (Neuristique). Training exam-

Figure 4. Network architecture. The input layer is a 2-D retina and four units encoding horizontal (H) and vertical (V) eye position, as shown in the lower left corner. The four units on the output layer compute the egocentric position of a gaussian pattern of light shown on the retina. The hidden layers form a hierarchy of retinotopic maps, each receiving converging inputs from a lower layer and eye position signals. Each hidden layer has three to five types of units per location. Each type of hidden unit has the same receptive field throughout the layer. The sizes of the receptive fields on the retina were smallest in hidden layer 1 and increased with distance from the retina.



ples consisted of an eye position vector and a gaussian pattern of activity placed at a particular location on the input layer and these were systematically varied throughout the training (Fig. 4). The network was trained to compute the egocentric position of the peak of the gaussian. For some trials there were no visual inputs and the output layer was trained to reproduce the eye position. The purpose of these trials will become clear when we describe the results of the simulations of electrical stimulations.

The training set had 625 patterns, corresponding to all the possible pairs between 25 retinal and 25 eye positions. We trained several networks with various numbers of hidden units per layer (typically three to five units per location) and found that they all converged to a nearly perfect solution (1% accuracy) in a few thousand sweeps through the training set. The network could also generalize to any new egocentric position, as long as it was within the range of positions used during the training phase.

Comparison between Hidden Units and Cortical Neurons

The influence of eye position on the visual response of a cortical neuron has been assessed by finding the visual stimulus eliciting its best response and measuring the amplitude of the response at nine different eye fixations (Andersen & Zipser, 1988; Galletti & Battaglini, 1989). We performed a similar test on the hidden units of trained networks and the results were compared with the gain fields of neurons recorded from area V3a and parietal cortex (Andersen & Zipser, 1988; Galletti & Battaglini, 1989). Responses were plotted as circles with diameters proportional to activity; the set of nine circles has been termed by Zipser and Andersen (1988) the spatial gain field of a unit because it shows how the amplitude, or gain, of the response varies with spatial position. Only three to five different spatial gain fields per hidden layer could be obtained due to the weight-sharing procedure that we used during training, which forced the properties of equivalent units to be identical. The properties of the units in the model were similar to those observed in cortical neurons regardless of receptive field size (Fig. 5). Despite having restricted receptive fields, the overall activity of most units increased monotonically in one direction in egocentric space. This direction, called the preferred eye position direction (PEPD), will be used in the next section describing the results of modeling electrical stimulation experiments. Note that the inner and outer circles, corresponding to the purely visual activity and the overall activity (visual plus background), do not always increase along the same direction due to the nonlinear sigmoid squashing function of the unit. These gain fields are very similar to those reported by Zipser and Andersen in area 7a (Zipser & Andersen, 1988). The major difference is that the hidden units in our model

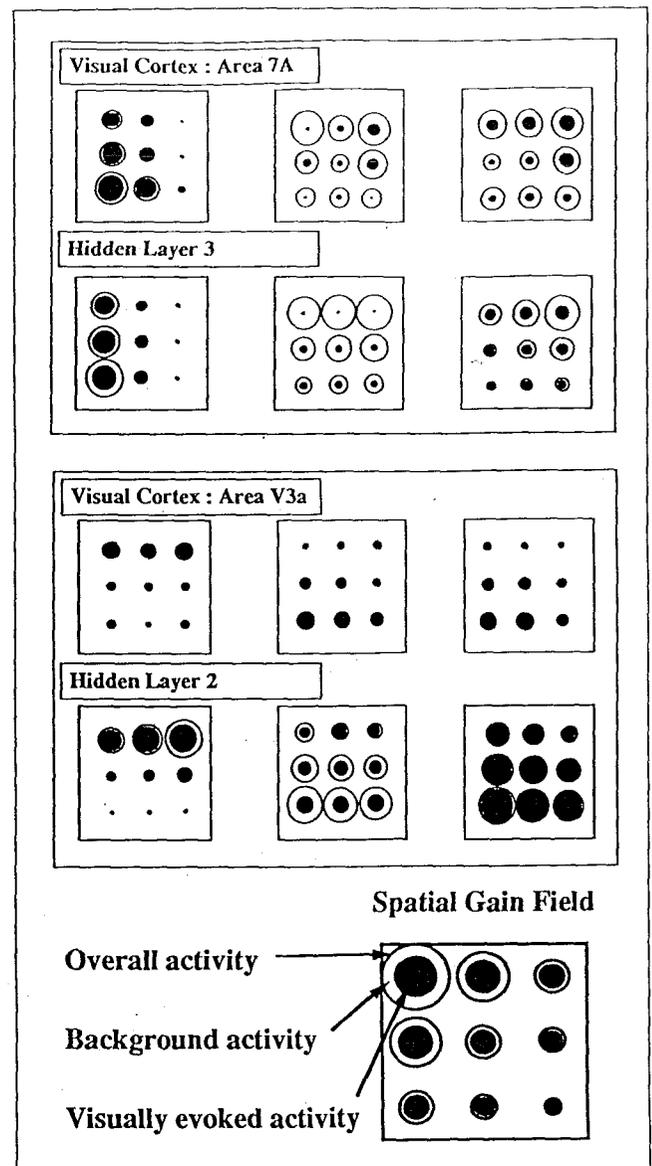


Figure 5. Comparison between the responses of cortical neurons (Andersen & Zipser, 1988; Galletti & Battaglini, 1989) and hidden units. Each box indicates the response of a single cell to the same retinal stimulation for nine initial eye positions. Activities of the units and the neurons are plotted with circles as indicated by the legend on the right. Notice that background activity is not shown for V3a neurons for which only overall activity is plotted. Hidden units are matched with real neurons on the basis of their receptive field sizes. Regardless of receptive field size, the response properties of the hidden units appear to be very similar to those of real neurons.

had restricted receptive fields covering only a small portion of the retina, whereas theirs covered the whole retina.

The modulation of the hidden units with eye position was not unexpected since each unit received inputs from eye position. Is this input enough to provide an egocentric coordinate transformation in the hidden layers? Consider the population of units that was found at any location of the hidden layer (i.e., the local population of units whose receptive field superimpose on the retina).

By virtue of the position of this population on the hidden layer, whose topology is retinotopic, those units encoded one particular value of retinal position, \vec{R} . Furthermore, since the activity of each unit in this local population increased monotonically with eye position, as indicated by the gradient of activity in the spatial gain field, there was also a rate coding of eye position, \vec{E} , at the population level (Since \vec{E} is a 2-D vector, at least two units per location with different PEDP are required). Therefore, at the population level the encoding is very similar to that found in the parietal cortex. Each unit carried information about \vec{R} and \vec{E} and yet none of them encoded the egocentric position, \vec{P} , explicitly.

There is, however, enough information at the population level to recover the egocentric position \vec{P} . After training a network with the same architecture as above, we removed the upper hidden layer (hidden layer 3 in Fig. 4) and connected the intermediate hidden layer (hidden layer 2) to the output layer. We then retrained the network, modifying only the new connections to the output layer. The motivation was to test whether the intermediate representation developed in the hidden layer 2 during the initial training was enough to recover egocentric position. The simulations showed that the network had no trouble relearning the task, thus supporting our claim that egocentric position was available at all levels of the network. The main difference between the representation in the layers was how wide a population had to be sampled to recover the egocentric position since lower hidden layers had more units than upper hidden layers.

It is important to notice that, in any hidden layer, the distributed encoding of egocentric position, \vec{P} , does not involve the visual connections coming from the hidden layer below. The encoding of eye position, \vec{E} , is due to the input from eye position units and the retinal location, \vec{R} , is encoded by virtue of the map topology. Thus, the visual receptive fields (i.e., the connections coming from the retina) are free to develop specific selectivity to any visual feature dimension such as motion, orientation, or disparity. In our model, the network had been exposed to only one type of visual stimulus, namely, a gaussian profile of luminance, so the visual selectivity developed during training was extremely limited.

A variety of visual selectivities is found in visual cortex. Cortical areas V1 and V3a encode relatively simple visual attributes and at least 40% of the neurons are modulated by static eye position. We would therefore expect V1 and V3a to encode low-level visual features in egocentric coordinates. Since a large percentage of cells in these areas are insensitive to eye position, egocentric coordinates appear to be used along with pure retinotopic ones. It will be important to determine where neurons carrying eye position signals project, including subcortical as well as cortical targets, and where the eye position signals originate and what form they take.

In summary, each hidden layer of the network has a

retinotopic map but also contains spatiotopic (i.e., egocentric) information through the spatial gain fields. To distinguish this type of representation from spatiotopic maps (as defined in the introduction), we call these retinospatiotopic maps (RSM).

Mimicking Electrical Stimulation Experiments

Determining the head-centered position \vec{P} of an object is equivalent to computing the position of the eye required to foveate the object [i.e., for a foveated object $\vec{R}=0$, which, according to Eq. (1), implies that $\vec{P}=\vec{E}$]. Therefore, although our model was originally trained to compute the egocentric (i.e., head-centered) position of an object based on its retinal position and the eye position, the output of the network can alternatively be interpreted as the eye position that would result in foveating the object. If the activities of the output units are interpreted as eye position, changes in this activity pattern correspond to eye movements. In this framework, our output units are actually similar to ocular motoneurons since eye position is assumed to be directly proportional to their level of activity. This interpretation allowed us also to mimic electrical stimulation experiments like the ones performed by McIlwain in V1, without any additional training. The change in the output activity was monitored while activating a set of selected hidden units in the network and was interpreted as the evoked eye movement.

First, the output activities were computed for an input pattern consisting of an eye position vector and a blank image (all retinal inputs were set to zero). The network had been previously trained to simply reproduce the eye position when presented with this kind of input. Next, we clamped the activity of a set of hidden units at a particular location in one of the layers to their maximum value, 1. The new output activity pattern was then computed and the change in this pattern was interpreted as an intended saccade. This procedure was repeated for various initial eye positions in order to explore the dependency of the evoked eye movement on this variable. Typically, nine initial eye positions were used, evenly spread in the 2-D visual field. This procedure was first proposed by Goodman and Andersen who applied it to the Zipser and Andersen model of the parietal cortex (Goodman & Andersen, 1989).

As a control, we first stimulated retinal inputs. The retina is clearly using eye-centered coordinates so that stimulation at a given retinal location should induce an eye movement to that location. Recall that before stimulation the network output, \vec{O} , was a copy of the input eye position signal \vec{E} . After stimulation in the input layer the network should encode the egocentric position $\vec{R}+\vec{E}$ of the pattern of activity that had been induced on the retina. Therefore the change in \vec{O} , corresponding to the eye movement, is equal to \vec{R} . The results, shown in Figure

6, agreed with these expectations. For example, stimulation of units in the upper left corner of the map produced a saccade in the upper left direction, regardless of initial eye position. The eye movement pattern obtained by stimulating a position at the middle left or middle right of the retina can be directly compared with the experimental results reported by McIlwain shown in Figure 3 (McIlwain, 1988).

Stimulation in the hidden layers led to different results depending on how many units were stimulated per position. Each location had between three and five units, depending on the hidden layer. We examined two situations in which either all the units sharing the same location were stimulated or only one of them. These results were then compared with those obtained by stimulating visual cortex.

When only one hidden unit type was stimulated per location, the pattern of induced eye movements was no longer a function solely of the retinal location of the stimulation (Fig. 7). Other factors, such as the preferred egocentric direction of the stimulated unit, were also important. This was particularly clear when the stimulation was performed in the upper hidden layer, one synapse away from the output (right side of Fig. 7). The directions of the evoked eye movements appeared to be

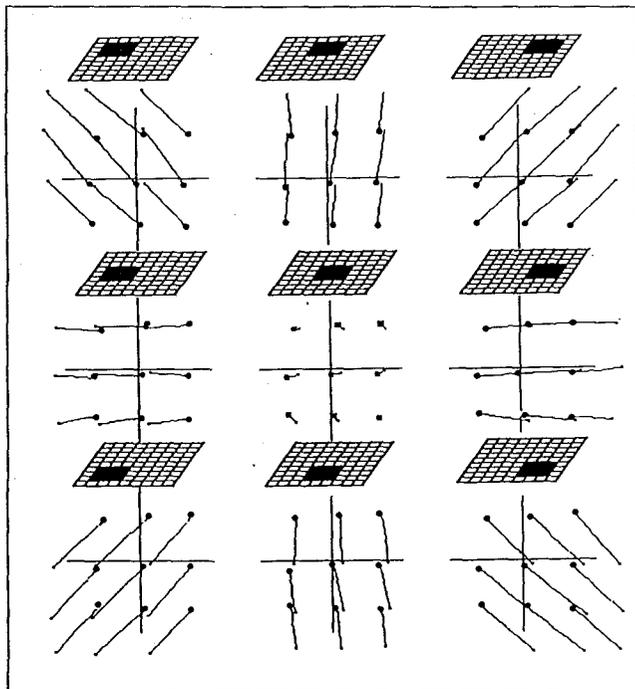


Figure 6. Eye movements evoked by stimulating the retinal input layer of the network. Each of the nine plots shows the evoked saccades from nine initial eye positions in response to identical stimulation performed at the location illustrated on the grid above each plot. Notice that the evoked eye movements are always in the direction of the stimulation site, which is typical of fixed vector saccades. The right and left middle plots can be directly compared to the results found from stimulating visual cortex (Fig. 3). See Figure 2 for plotting convention.

a function of both the position of the stimulation site on the map and the preferred eye position direction (PEPD) of the unit. We refer to these two components as, respectively, the retinal and the PEPD components. Hence, when the stimulation was in the center of the map, that is to say when the retinal component was zero, the evoked eye movement was parallel to the gradient of the spatial gain field, which, as we have seen, is an indication of the PEPD for this unit. When other parts of the map were stimulated, the direction of the eye movements was approximately a linear combination of retinal and the PEPD components.

When a hidden unit located in the intermediate layer was stimulated (left side of Fig. 7), the influence of the PEPD was not as clear as for units in the last hidden layer. In particular, stimulation at the center of the map did not evoke eye movement strictly parallel to the PEPD. This was probably related to the fact that a unit in the intermediate layer projected to units in the upper hidden layer whose PEPD's did not coincide with the PEPD of the stimulated unit. Nevertheless, the pattern of eye movement could still be easily distinguished from the pattern obtained after stimulating the input layer or from the experimental pattern reported by McIlwain (1988).

When all the units sharing the same position were activated together, the output pattern resembled the one obtained by stimulating the input layer (Fig. 8). Even though each hidden unit had a different PEPD, when simultaneously activated these balanced out and the dominant factor became the location of the stimulation. The same results were obtained for all the hidden layers. A careful examination of Figure 8 reveals a slight tendency for converging eye movements. This was particularly apparent for stimulation in any of the corners of the upper hidden layer. This apparent convergence could be accounted for by saturation of the output units whose activities were limited to the range [0,1] by their input/output squashing function. We stimulated the hidden units by setting the activity of those units to 1, their maximum value, which is significantly greater than that obtained by simply showing a stimulus on the retina. This in turn drove the output units into their saturation region; thus, the observed convergence had nothing to do with the spatial representation in the upper hidden layer.

Comparison with Electrical Stimulation in V1

Strong electrical stimulation in area V1 of the visual cortex is likely to recruit many neurons whose receptive fields share the same retinal location. As our simulation shows, in such conditions, even if static eye position modulates the gain of V1 neurons, one would expect to obtain fixed vector saccades that are consistent with the results reported by McIlwain (1988). Furthermore, many

Figure 7. Eye movements evoked by stimulating one hidden unit type at nine different locations in the intermediate (left) and top (right) hidden layers. Same plotting convention as Figures 2 and 6. The spatial gain field of the unit being stimulated is shown below the corresponding plot. In hidden layer 3, the direction of the saccades depends on both the position of the stimulation on the map and the preferred eye position direction of the unit, as indicated by the direction of the gradient of the spatial gain field. Results are more difficult to interpret for stimulation in hidden layer 2, but in both layers, the pattern of eye movements is different from the one obtained from stimulating the input layer (see Fig. 6).

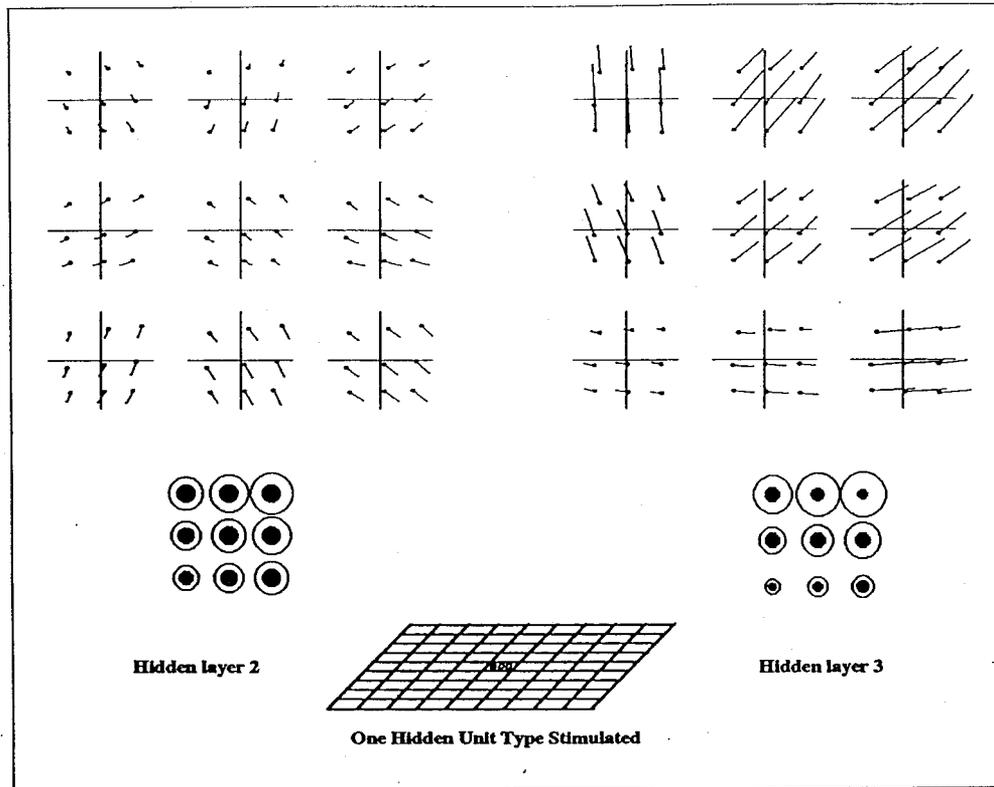
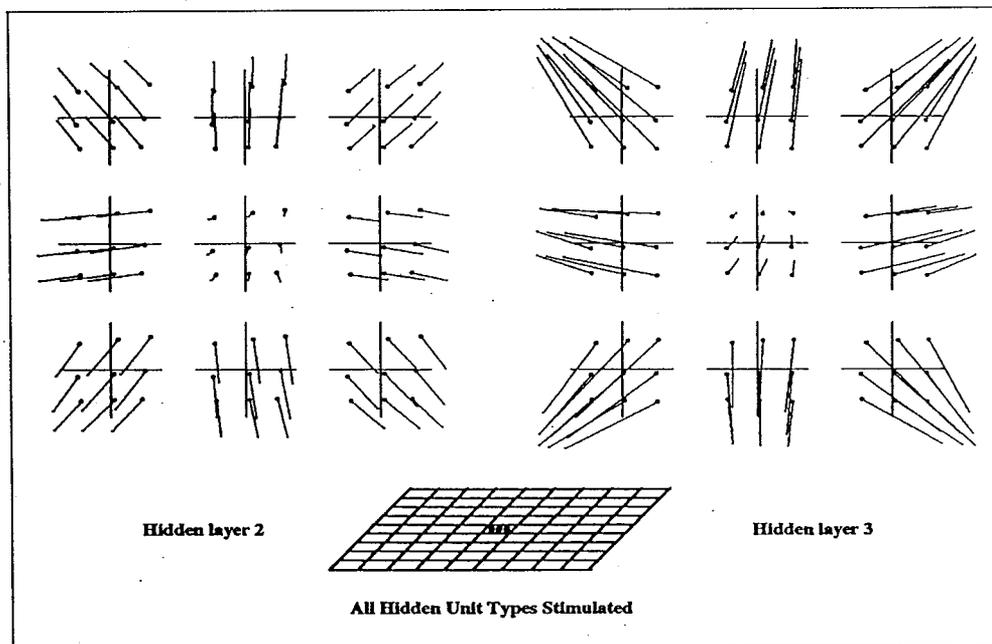


Figure 8. Eye movements evoked by stimulating all the hidden units sharing the same location. In both layers the pattern of eye movements is very similar to the one induced by stimulating the input layer (Fig. 6), even though the input and hidden layers do not use the same type of spatial representation.



cells in V1 do not show any eye position modulation of their response, and when they do, the modulation appears to be weaker than that reported in the posterior parietal cortex. This would make the retinal component even more dominant in the determination of the direction of the induced eye movement. Also, in V3a, there is a correlation between the retinotopic position of a cell and its preferred eye position direction (PEPD) such that

neurons in the right cortex, encoding the left visual field, tend to fire more when the eyes fixate also in the left part of the visual field (Galletti & Battaglini, 1989). Thus, the two components which influence the direction of the evoked eye movement, the retinal and the PEPD components, would tend to line up such that the influence of the PEPD component would be even more difficult to detect.

DISCUSSION

The model that we have presented demonstrates that the study of eye movements evoked by electrically stimulating early visual areas may not be able to distinguish between retinal and retinospacial coordinates. In higher visual areas with weaker retinotopy, it might be possible to obtain patterns closer to those produced by stimulating only one type of hidden unit. If in such an area neurons with similar spatial gain fields tended to cluster together in a columnar organization, then stimulation might induce eye movements whose direction would depend on the preferred eye position direction (PEPD). This pattern of eye movements has already been observed in parietal area LIP (Goodman & Battaglini, 1989; Thier & Andersen, 1992).

One aspect of McIlwain's result cannot be accounted for by our model. From the data shown in Figure 3, it is clear that the amplitude of the saccade evoked by electrically stimulating the striate cortex was not independent of initial eye position. In fact, the saccade amplitude decreased systematically as the initial position was moved in the direction of the saccade. This decreased amplitude was accompanied by a lower probability of evoking a saccade. One contributing factor for this tendency is mechanical: The cat can make saccades only to an eccentricity of around 20°, so if the starting eye position is close to the limit, strong mechanical damping can anisotropically reduce the amplitude of some saccades. A contribution from the stimulation of head movements is not likely since head movements influence the amplitude of eye movements only through the vestibulo-ocular reflex (Blakemore & Donaghy, 1980).

Network Representations: Explicit vs. Implicit

The type of units that we used in the output layer of our network, namely units that encode explicitly head-centered position, have never been found in the cortex. One might, therefore, wonder whether this important difference between the cortex and our model would invalidate our conclusions. This would be a serious problem if the model were intended to account for the development of the representation found in the hidden layer and the actual mechanisms that create these representations in the brain. Our intention is less ambitious. We want to compare the properties of units in a mature network with the properties of neurons in an adult visual cortex. Once the network has reached its final state, the output layer can be taken away. An explicit output does, however, demonstrate that the type of information available implicitly in the last layer of hidden units could be extracted through a projection. Whether a network with a different output representation could have developed a different hidden representation is irrelevant for our purpose. What matters is that the representation we report

in the hidden layer is similar to what is found in early visual areas and that it contains an implicit representation of egocentric position.

Neither is the exact nature of the output representation critical for the electrical stimulation experiments. We obtained retinotopic saccades on stimulating all the units sharing the same position on one of the hidden layers by averaging the elementary saccades that are obtained after stimulating one unit at a time. Whether we are dealing with units or neurons, averaging many vectors evenly distributed in all possible directions leads to a null vector. This result is independent of the output representation.

The output representation we used in our model is an explicit representation of head-centered position of objects. In contrast, the egocentric representation that we have found in the hidden layers of our model, which seems to exist in early visual areas, is implicit in the following sense: (1) the representation is self-contained; that is, no additional signals are necessary to recover the variable encoded, and (2) only one stage of processing (one layer of weights) is required to obtain an explicit representation of the variable. This definition of an implicit representation captures the notion that early visual areas are not just retinotopic, but also contain enough information to guide behaviors in which the head-centered positions of objects are needed, such as the control of head movements. This is a direct corollary of the two criteria defined above.

Implications for Psychological Experiments

Our conclusions are based primarily on physiological findings and are at odds with many psychophysical results, which we review here. With few exceptions, attempts to find evidence for an early spatiotopic buffer have failed. The consensus of opinion is that spatial transformations occur at a late stage of visual processing. How can we account for this major discrepancy?

Most experiments have tested for a spatial representation that is quite different from the retinospacial maps used in our model. In a purely spatiotopic map, neurons would respond to visual features such as orientation or color of an object at a fixed spatial location. Such maps would have receptive fields in egocentric rather than retinal coordinates. Feldman (1985) has advocated such egocentric maps, arguing that they would have important computational advantages for object recognition.

The basic reasoning behind most psychophysical experiments designed to uncover evidence for egocentric representations is the following: When two stimuli are presented in short temporal succession at the same retinal location, the processing of the second one is known to "interact" with the processing of the first one. Depending on the experiments, "interact" can mean "mask," "facilitate," "inhibit," or simply "superimpose." If there were an egocentric buffer, such interference should be

observed when the spatiotopic position is preserved, even if the retinotopic position changes. This situation can be obtained by flashing a prime and a target at the same position on a screen and asking the subject to make a saccadic eye movement during the interstimulus interval. These experiments have failed to show any kind of spatiotopic interaction for low-level visual features, under conditions in which retinotopic interactions are commonly reported.

The first experiment of this kind was performed by O'Regan and Levy-Schoen (O'Regan, 1983). In their experiment, subjects were asked to read a word that had been broken down into component strokes and half of the strokes presented 50 msec apart. The two halves were chosen such that the words were recognizable only when they were superposed. For instance the word "him":

| - | | | \ / |

was displayed as

| - \ /

followed by

| | / |

Their results showed that subjects could report the words if both halves were flashed on the same retinal location, but not in conditions under which the spatial location of the stimulus was maintained and the eyes moved.

Another example is a study by Irwin, Zachs, and Brown (1990) on orientation masking, which showed that, with the eyes fixed, the detection threshold of an oriented test grating was increased by the presentation of a 40 msec prime grating of similar orientation. However, if the subject was asked to make an eye movement during the interstimulus interval, such that a prime and a target appeared at the same spatiotopic location, but not on the same retinal location, the detection threshold was not affected by the prime. Many other experiments have been performed along these lines and all of them have reached the same conclusion (Irwin, 1991; Irwin, Brown, & Sun, 1988; Pollatsek, Rayner, & Henderson, 1990; Rayner & Pollatsek, 1983; Sun & Irwin, 1987).

Only two psychophysical experiments seem to support the possibility for egocentric representations in early vision (Kohler, 1964; Mayhew, 1973). These experiments looked for after-effects contingent on eye position. Mayhew found that after seeing repetitively clockwise motion while looking left and anticlockwise motion while looking right, subjects reported an anticlockwise motion after-effect when looking right and clockwise motion when looking left (Mayhew, 1973). Neurons sensitive to rotational motion have been found in area MST and other relatively late stages of processing (Sakata, Shibutani, & Tsurugai, 1986). We predict, however, that it should be possible to demonstrate contingent after-effects with translational motion or other elementary visual attributes

such as orientation or disparity. Such results would suggest that these features are part of a retinospatiotopic map.

Kohler reported a positive result for color after-effects (Kohler, 1964), but these experiments have not been replicated (McCullough, 1965). Color may not be one of the visual attributes that are combined with eye position since the areas where eye-position modulation has been reported belong mainly to the "dorsal" pathway to the parietal cortex (V3a, 7a, LIP) where color is not a primary feature being represented. Attributes such as color and shape, which are preferentially represented in the ventral stream of the visual system, the "what" pathway, may not be integrated with eye-position signals.

Why Retinospatial Maps Are Different from a Spatiotopic Buffer

The type of egocentric representation we have proposed is quite different from the egocentric representations that have been probed by psychophysical experiments. Retinospatiotopic maps (RSM) differ from a purely spatiotopic buffer in that their topology is retinotopic, not egocentric. Since most psychophysical experiments have been performed under the assumption that spatiotopic representation ought to involve spatiotopic topology, they could not distinguish purely retinotopic maps from RSMs. Consider, for example, the orientation masking by Irwin et al. (1990). When the prime and target gratings are flashed on the same retinal position, a masking is expected whether the maps are retinotopic or retinospatiotopic. Conversely, when the spatiotopic position is preserved but the retinotopic position is changed, no masking can take place in any map with retinotopic topology.

The same explanation could account for the results of the experiments with words (O'Regan, 1983). If the two components of a word do not superimpose in VI it is difficult to imagine how it could be recognized. Even if the orientational spatiotopic maps in VI managed to encode simultaneously all the segments of each letter along with their spatiotopic coordinates, it may be that the visual cortex needs relationships between the subparts to be represented explicitly before an object can be recognized. Since these local spatial relationships are lost when the two halves are separately presented, the subject cannot identify the letters.

Object Localization and Eye Position Signal

Even though egocentric coordinates at early stages in the visual system may not be essential for object recognition, they could still be useful for localization and manipulation of object subparts. Several experiments have addressed the issue of whether eye position signals are used for localizing an object.

Matin and Pearce (1965; see also Mateef, 1978) first

demonstrated that a human subject can accurately localize a point of light briefly flashed on a screen while making a saccadic eye movement. This suggests that the subject had access to an extra retinal signal encoding eye position for use in computing the position of the flash on the screen during the eye movement. This conclusion has been seriously questioned by MacKay (1970) and O'Regan (1984), but a recent experiment performed by Gauthier, Nommay, and Vercher (1990) provides new evidence in favor of the extra retinal signal. This experiment tested the influence of eye position on hand-pointing to visual targets. When one eye was covered and deviated with a suction lens while the other one was fixated on the target, subjects systematically mislocalized in the direction of the deviated eye. This demonstrates not only that eye position is used in target localization but also that proprioception is part of the extraretinal signal (Ashton et al., 1984; Buisseret & Maffei, 1977).

These experiments, however, do not distinguish between the localization of objects, or subparts of objects, or the visual features of the objects, and consequently do not directly support our hypothesis that a RSM could be useful for localizing low-level visual features. Nonetheless, the results are encouraging and motivate additional experiments to measure the influence of eye position on the localization of parts of objects.

New Dimensions: Distance and Size

Although this paper addressed only the egocentric position of visual features in two dimensions, our model can be readily extended to the egocentric position of an object along the third spatial dimension. Models have already been developed for representing distance using disparity selective neurons whose gain is modulated by vergence angle (Lehky, Pouget, & Sejnowski, 1990). Neurons with these properties have been reported in area V1 of behaving monkeys (Trotter et al., 1992) [There is also indirect evidence for vergence modulation of monocular neurons in the LGN (Kawanura & Marchiafava, 1966; Richards, 1968).] This model could be extended to include retinotopic maps, so that a similar retinospacial representation of egocentric distance might also be found in visual cortex. Neurons involved in coding low-level features, such as orientation, in three-dimensional space would have to be selective for disparity as well as for eye position (vergence angle). Such neurons have already been reported in V1 (Trotter et al., 1992).

Eye-position modulation of neurons at early stages of the visual system may not be limited to the representation of egocentric space. Retinospacial maps might also subserve size constancy, our ability to perceive rigid objects as having constant physical size even though their angular size and their disparity vary with the distance of the object to the body. A convergence of visual and eye position signals is necessary to perform the appropriate compensation required for size constancy. The mathe-

matical transformations for computing size constancy are similar to those involved in computing egocentric location.

CONCLUSION

Our model has shown how the gain modulation of visual responses in single neurons observed at early stages of visual processing are consistent with the hypothesis that low-level visual features are encoded in egocentric coordinates. These results suggest that subparts of objects, which in early cortical areas are represented by different neurons, are also encoded in this set of coordinates. Thus the visual cortex may not be purely visual and, as a result, *may* encode egocentric coordinates in the earliest stages of visual processing.

Acknowledgments

We would like to thank Drs. Yves Trotter, Simon Thorpe, Jean Rene Duhamel, Colin Blakemore, Richard Andersen, John Mayhew, and Sidney Lehky for helpful discussions and suggestions. This research was supported by a grant from the Office of Naval Research and by the Howard Hughes Medical Institute.

Reprint requests should be sent to Alexandre Pouget, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037.

REFERENCES

- Andersen, R. A. (1989). Visual and eye movement functions of the posterior parietal cortex. *Annual Review of Neuroscience*, *12*, 377-403.
- 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 visuo-motor coordination. *Canadian Journal of Physiology and Pharmacology*, *66*, 488-501.
- Ashton, J. A., Boddy, A., & Donaldson, I. M. L. (1984). Directional selectivity in the responses of units in the cat primary visual cortex to passive eye movement. *Neuroscience*, *13*, 653-662.
- Blakemore, C., & Donaghy, M. J. (1980). Co-ordination of head and eyes in the gaze changing behavior of cats. *Journal of Physiology*, *300*, 317-335.
- Brotchie, P. R., & Andersen, R. A. (1991). A body-centered coordinate system in posterior parietal cortex. *Abstract of Society for Neuroscience*, *17*, 1281.
- Buisseret, P., & Maffei, L. (1977). Extraocular proprioceptive projections to the visual cortex. *Experimental Brain Research*, *28*, 421-425.
- Feldman, J. A. (1985). Four frames suffice: A provisional model of vision and space. *Behavioral and Brain Science*, *8*, 265-289.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1-47.
- Galletti, C., & Battaglini, P. P. (1989). Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. *Journal of Neuroscience*, *9*, 1112-1125.
- Gauthier, G. M., Nommay, D., & Vercher, J. L. (1990). The role

- of proprioception in visual localization of targets. *Science*, 249, 58–61.
- Goodale, M. A., & Milner, A. D. (1990). Separate visual pathways for proprioception and action. *Trends in Neuroscience*, 15, 20–25.
- Goodman, S. J., & Andersen, R. A. (1989). Microstimulation of a neural network model for visually guided saccades. *Journal of Cognitive Neuroscience*, 1, 317–326.
- Goodman, S. J., & Andersen, R. A. (1990). Algorithm programmed by a neural model for coordinate transformation. *Proceedings of the International Joint Conference on Neural Networks*, 2, 381.
- Heilman, K. M., Watson, R. T., & Valenstein, E. (1985). Neglect and related disorders. In K. M. Heilman & E. Valenstein (Eds.), *Clinical Neuropsychology* (pp. 243–294). New York: Oxford University Press.
- Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, 23, 420–456.
- Irwin, D. E., Brown, J. S., & Sun, J. (1988). Visual masking and visual integration across saccadic eye movements. *Journal of Experimental Psychology*, 117, 276–287.
- Irwin, D. E., Zachs, J. L., & Brown, J. S. (1990). Visual memory and the perception of a stable environment. *Perception and Psychophysics*, 47, 35–46.
- Kawanura, H., & Marchiafava, P. L. (1966). Modulation of transmission of optic nerve impulses in alert cat: Evidence of presynaptic inhibition of primary afferents during ocular movements. *Brain Research*, 1, 213–215.
- Kohler, I. (1964). The formation and transformation of the perceptual world. *Psychological Issues*, 3, 62–86.
- Kurylo, D. D., & Skavenski, A. (1991). Eye movements elicited by electrical stimulation of area PG in the monkey. *Journal of Neurophysiology*, 65, 1243–1253.
- Lal, R., & Friedlander, M. J. (1989). Gating of the retinal transmission by afferent eye position and movement signals. *Science*, 243, 93–96.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1990). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1, 540–566.
- Lehky, S. R., Pouget, A., & Sejnowski, T. J. (1990). Neural models of binocular depth perception. In E. R. Kandel, T. J. Sejnowski, C. F. Stevens, & J. D. Watson (Eds.), *Cold Spring Harbor symposium on quantitative biology: The Brain*. New York: Cold Spring Harbor Press.
- MacKay, D. M. (1970). Mislocalization of test flashes during saccadic image displacements. *Nature (London)*, 227, 731–733.
- Mateef, S. (1978). Saccadic eye movements and localization of visual stimuli. *Perception and Psychophysics*, 24, 215–224.
- Matin, L., & Pearce, D. G. (1965). Visual perception of direction for stimuli flashed during voluntary saccadic eye movements. *Science*, 248, 1485–1488.
- Mayhew, J. E. W. (1973). After-effects of movement contingent on direction of gaze. *Vision Research*, 13, 877–880.
- McCullough, C. (1965). Conditioning of color perception. *American Journal of Psychology*, 78, 362–378.
- McIlwain, J. T. (1988). Saccadic eye movements evoked by electrical stimulation of the cat visual cortex. *Visual Neuroscience*, 1, 135–143.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neuroscience*, 6, 414–417.
- O'Regan, J. K. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research*, 23, 765–768.
- O'Regan, J. K. (1984). Retinal versus extraretinal influences in flash localization during saccadic eye movements in the presence of visible background. *Perception and Psychophysics*, 36, 1–14.
- Pollatsek, A., Rayner, K., & Henderson, J. M. (1990). Role of spatial location in integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 199–210.
- Pouget, A., & Sejnowski, T. J. (1992). A distributed common reference frame for egocentric space in the posterior parietal cortex. *Behavioral and Brain Sciences*, 15, 787–788.
- Rayner, K., & Pollatsek, A. (1983). Is visual information integrated across saccades? *Perception and Psychophysics*, 34, 39–48.
- Richards, W. (1968). Spatial remapping in the primate visual system. *Biological Cybernetics*, 4, 146–156.
- Robinson, D. A., & Fuchs, A. F. (1969). Eye movements evoked by stimulation of the frontal eye field. *Journal of Neurophysiology*, 32, 637–648.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations*. Cambridge, MA: MIT Press.
- Sakata, H., Shibutani, H., & Tsurugai, K. (1986). Parietal cortical neurons responding to rotary movement of visual stimulus in space. *Experimental Brain Research*, 61, 658–663.
- Shibutani, H., Sakata, H., & Hyvarinen, J. (1986). Saccade and blinking evoked by microstimulation of the posterior parietal association cortex of the monkey. *Experimental Brain Research*, 55, 1–8.
- Sun, J. S., & Irwin, D. E. (1987). Retinal masking during pursuit eye movements: Implications for spatiotopic visual persistence. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 140–145.
- Thier, P., & Andersen, R. A. (1992). Electrical microstimulation delineates 3 distinct eye-movement related areas in the posterior parietal cortex of the Rhesus monkey. *Abstract for Society for Neuroscience*, 17, 1281.
- Trotter, Y., Celibrini, S., Stricanne, B., Thorpe, S., & Imbert, M. (1992). Modulation of neural stereoscopic processing in primate area V1 by the viewing distance. *Science*, 257, 1279–1281.
- Weyand, T. G., & Malpeli, J. G. (1989). Responses of neurons in primary visual cortex are influenced by eye position. *Abstracts of the Society of Neuroscience*, 15, 1016.
- Zipser, D., & Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature (London)*, 331, 679–684.