

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Computational Models of Spatial Representations

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Biology

by

Alexandre Pouget

Committee in charge:

Professor Terrence J. Sejnowski, Chair
Professor Thomas D. Albright
Professor William A. Harris
Professor William B. Kristan
Professor Vilayanur S. Ramachandran
Professor Simon LeVay

1994

Copyright
Alexandre Pouget, 1994
All rights reserved.

The dissertation of Alexandre Pouget is approved, and
it is acceptable in quality and form for publication on
microfilm:

Chair

University of California, San Diego

1994

*A mes parents, a Alrivi et au Montrose 82
de chez Laperouse.*

TABLE OF CONTENTS

Signature Page	iii
Dedication	iv
Table of Contents	v
List of Figures	viii
Acknowledgements	ix
Vita and Publications	xi
Abstract	xiii
I Introduction	1
A. Spatial representations and sensori-motor coordination	1
B. The posterior parietal cortex	2
C. Neural code for spatial representations	4
1. Dynamic remapping	4
2. Gain modulation	6
3. The Zipser and Andersen Network	6
D. Parallel vectorial representations	9
E. Thesis Outline	10
1. Hierarchy in spatial representations	10
2. A basis function approach for spatial representation	11
II Egocentric spatial representation in early vision	12
A. Introduction	13
B. Results	20
1. Network Architecture	20
2. Comparison Between Hidden Units and Cortical Neurons	22
3. Mimicking Electrical Stimulation Experiments	26
4. Comparison with Electrical Stimulation in V1	30
C. Discussion	32
1. Network output representation	32
2. Explicit vs implicit representation	33
3. Amplitude of saccades evoked by electrical stimulation	34
4. Implications for Psychological Experiments	34
5. Why retinospacial maps are different from a spatiotopic buffer	37
6. Object localization and eye position signal	38
7. New Dimensions : Distance and Size	38
D. Conclusion	39

III	A Neural Model of the Cortical Representation of Egocentric Distance . .	45
	A. Representing egocentric distance with basis functions	52
	1. Basis Functions	52
	2. Methods	53
	3. Results	60
	B. Reliability of Distance Estimation	64
	1. Bias-variance trade-off	64
	2. Methods	64
	3. Results	67
	C. Discussion	70
	1. Posterior parietal cortex	70
	2. Early egocentric representation	71
	3. Do gain-modulated neurons form a basis set?	72
	4. Advantages of basis functions representation over an explicit representation	73
	5. Comparison to backpropagation networks	74
	6. Extraretinal vs retinal cues for viewing distance	75
	7. Distance vs scaling	76
IV	Representing Spatial Transformations in the Parietal Cortex with Basis Functions	87
	A. Introduction	89
	B. Basis Function Representation	91
	1. Sensorimotor Coordination	92
	2. Gain Fields and Basis Functions	95
	3. Predictions of the Basis Function Representation	98
	4. Simulations	103
	C. Comparison with Other Representations	107
	1. Map Representation	107
	2. Vectorial Representation	108
	D. Discussion	112
	1. Basis Function Representation and Frame of Reference	112
	2. Basis functions compared to the Zipser and Andersen network . . .	112
	3. Beyond parietal cortex	114
	4. Other sensory modalities and posture signals	115
	5. Conclusions	116
V	Future Work	117
	A. Development of spatial representations	117
	1. Preliminary work: Predictive learning	118
	2. Optimum representation	119
	B. Modeling Hemineglect	120

VI	Conclusion	123
	A. Explicit and implicit representations	124
	B. Modular architecture and basis functions	125
	Bibliography	127

LIST OF FIGURES

I.1	Line cancellation task	3
I.2	The Zipser and Andersen network	7
II.1	Computing the position of an object in head-centered coordinates . .	14
II.2	Saccadic eye movements evoked by electrical stimulations	17
II.3	Eye movements evoked by microstimulations in V1	19
II.4	Network architecture	21
II.5	Comparison between cortical neurons and hidden units	24
II.6	Stimulation of the input layer	28
II.7	Stimulation of the hidden layer	29
II.8	Stimulation of multiple hidden units	31
III.1	Viewing geometry for stereopsis	47
III.2	Idealized disparity tuning curves	48
III.3	Responses of idealized disparity and distance detectors	50
III.4	Gain modulated neurons in V1	51
III.5	Basis function network for distance approximation	54
III.6	Hidden units tuning curves	57
III.7	Perceived disparity as a function of geometric disparity	59
III.8	Network approximations of geometric distance	61
III.9	Network approximations to perceived distance	63
III.10	Estimation of perceived and geometric disparities	68
IV.1	Typical parietal neuron	90
IV.2	Computing the position of an object in head-centered coordinates . .	92
IV.3	Neural network for map transformation	93
IV.4	Idealized gain modulated unit	97
IV.5	A spatial representation using basis functions	97
IV.6	Four typical visual receptive fields of parietal neurons	98
IV.7	Four typical gain fields of parietal neurons	99
IV.8	Sigmoidal versus linear tuning	100
IV.9	x-intersect	102
IV.10	Approximating head-centered and retinotopic receptive fields	105
IV.11	Computational map	107
IV.12	Vectorial representation	108

ACKNOWLEDGEMENTS

Seven years ago, Terry Sejnowski accepted me in his laboratory for a PhD. He had never met me, he had never spoken to me, he barely knew my advisor, and I did not have any grant to cover the cost of my studies beyond the first year.

I found out later on that this is just the way Terry recruits people. It would be tempted to conclude that his recruiting strategies are totally insane, but I came to believe that it has more to do with generosity. Over the last six years, he has put together and maintained a truly remarkable and eclectic group of people, and being part of it has been a wonderful experience. I am extremely grateful to him for giving me a chance to join his laboratory and for his continuous support and trust.

My gratitude goes also to Rosemary and Hillary not only for their help over the years but, more importantly, for helping me to keep a decent level of sanity throughout my PhD.

As to the rest of the lab, unfortunately, I have to settle for a collective thank. I have met more than 50 members over the last six years and all of them have contributed to this dissertation, whether during lab meetings or through random discussions at tea or in the corridors.

I would like to thank Francis Crick for his support and for sharing with us his contagious enthusiasm for science.

I am deeply indebted to Michael Jordan and Peter Dayan. Both have strongly influenced my way of thinking and many of the ideas that are discussed in the thesis came out of discussions I had with them.

Back in France, I would like to express my gratitude to Michel Imbert, Philippe Ascher and Simon Thorpe who have managed in less than a month, in the fall 1988, to convince me to leave my country, my language, my city, my friends, my girlfriend and my family to go study in the US. Anybody who knew me at the time, knows that this was no small achievement.

I am thankful to Steve Fisher for his help on chapter 2, on which he is co-author, and to Richard Andersen for providing some experimental data, some of

which are shown in figure IV.9.

Lastly, thanks for to my parents and my wife. They have been, and remain, my primary source of inspiration.

VITA

April 17, 1966	Born, Boulogne Billancourt, France
1985–1990	Student at the “Ecole Normale Supérieure de Paris”
1987	B.A., University Paris VI, France
1988	M.S., University Paris VI, France
1990–1994	McDonnell-Pew Fellow, University of California, San Diego
1994	Doctor of Philosophy University of California, San Diego

PUBLICATIONS

Thorpe S.J. and Pouget A. “Coding of Orientation by the Visual Cortex : Neural Network Modelling.” Pfeifer R. (ed), “Connectionism in Perspective”, Amsterdam : Elsevier. 1989.

Thorpe S.J., O’Regan K. and Pouget A. “Human Fails on XOR Pattern Classification Problems.” Personnaz L. and Dreyfus G. (eds), “Neural Networks : From Model to Applications”, Paris, IDSET, 12-25. 1989.

Lehky S.R., Pouget A., Sejnowski T.J. “Neural Models of Binocular Depth Perception.” Cold Spring Harbor Symposium on Quantitative Biology . 55: 765-777. 1990.

Trotter, Y., Beaux, J.C., Pouget, A., Imbert, M. “Temporal limits of the susceptibility of depth perception to proprioceptive deafferentations of extraocular muscles.” Developmental Brain Research, 59: 23-29. 1991.

Pouget, A. and Thorpe, S.J. “Connectionist Model of Orientation Identification.”. Connection Science, 3 (2): 127-142. 1991.

Pouget, A., Fisher, S. and Sejnowski, T.J. “Hierarchical Representation of Space in the Visual System.” In Advances in Neural Information Processing Systems. 4. Moody, J.E.; Hanson, S.J. and Lippmann, R.P (eds), San Mateo, CA: Morgan Kaufmann Publishers. 1992.

Pouget, A., Fisher, S. and Sejnowski, T.J. “Egocentric representation in early vision.”. Journal of Cognitive Neuroscience, 5(2):150-161. 1993.

Montague, P.R., Dayan, P., Nowlan, S.J., Pouget, A. and Sejnowski, T.J. “Using aperiodic reinforcement for directed self-organization.”. Hanson, S.J., Cowan, J.D. and Giles, C.L. (eds), Advances in Neural Information Processing Systems. 5. San Mateo, CA: Morgan Kaufmann Publishers. 1993.

Pouget, A. and Sejnowski, T.J. “The Cortical Representation of Egocentric Distance.” *Cerebral Cortex*, 4(3):314-329. 1994.

Pouget, A. and Sejnowski, T.J. “Dynamical Remapping”. In “The Handbook of Brain Theory and Neural Networks.” Arbib, M.A. (ed). Boston: MIT Press. 1995.

Pouget, A., and Sejnowski, T.J. “Reinforcement Learning Predicts the Site of Plasticity for Auditory Remapping in the Barn Owl” In *Advances in Neural Information Processing Systems*. 7. San Mateo, CA: Morgan Kaufmann Publishers. 1995.

Pouget, A., and Sejnowski, T.J. “Spatial Representations in the Parietal Cortex May Use Basis Functions” In *Advances in Neural Information Processing Systems*. 7. San Mateo, CA: Morgan Kaufmann Publishers. 1995.

Pouget, A., and Sejnowski, T.J. “Representing Spatial Transformations in the Parietal Cortex with Basis Functions” Submitted.

ABSTRACT OF THE DISSERTATION

Computational Models of Spatial Representations

by

Alexandre Pouget

Doctor of Philosophy in Biology

University of California, San Diego, 1994

Professor Terrence J. Sejnowski, Chair

Generating a motor command in response to a sensory stimulus can be regarded as problem of coordinate transformation in which the sensory coordinates of the object are transformed into motor coordinates. It has been suggested that this process involves an intermediate stage in which the object location is encoded in egocentric coordinate, a global frame of reference that can be used for many motor behaviors. Lesions studies and single cell recordings point at the parietal cortex as a primary site for this representation. This view, however, rests on the premise that the position of objects can only be represented with respect to a unique frame of reference. We have explored an alternative possibility in which sensorimotor coordination is viewed as a problem of nonlinear function approximation. The responses of neurons in the parietal cortex appear to be well approximated by a gaussian function of a sensory variable, such as horizontal disparity or retinal location, multiplied with a sigmoid function of a posture signal, such as eye position or vergence angle. A large number of such response functions forms a basis for the space of all possible non-linear functions of the input variables. It follows that any motor command can be generated by a simple linear combination of the responses of parietal neurons, since motor commands are nonlinear functions of the sensory variables. In this representation, the position of objects is encoded in multiple frames of reference simultaneously, a property consistent not only with the responses of single cells, but also with the

neurological deficits resulting from unilateral lesions in the parietal cortex. This theory is applicable to early visual areas as well, such as V1, V3a or MST, in which gain-modulated neurons have been recently found, raising the possibility that spatial transformation may be initiated prior to the parietal cortex, perhaps as early as V1.

Chapter I

Introduction

I.A Spatial representations and sensori-motor coordination

The term *spatial representation* is typically used to refer to the way neurons represent the position of objects in space. This definition suggests that understanding spatial representation amounts to determining the coordinates used by the brain to encode the location of objects, since a position can only be specified with respect to an origin and a set of coordinate axes.

In physics, choosing a frame of reference is often a crucial step toward finding a solution to a particular problem. A good choice can often turn what may appear to be a hard problem into a trivial one. The choice of a set of coordinates is therefore strongly constrained by the type of operations that one intends to carry over the representation. In the case of spatial representations, the computation is likely to be related to the general problem of sensori-motor coordination. Orienting the eyes or head toward an object, or grasping it, requires that we have some knowledge of where objects are located either with respect to each others and with respect to ourselves.

This consideration has motivated one of the dominant approaches in sensori-motor transformation which states the problem in terms of change of coordinates (Soecht-

ing and Flanders, 1992). In the first stages of sensory processing, the location of object is defined in sensory coordinates. For example, visual images are encoded in retinotopic coordinates whereas sounds are encoded in head-centered coordinates. Generating a motor command, on the other hand, involves a recoding of the object position in motor coordinates, such as muscle coordinates for an arm movement, or oculocentric coordinates for moving the eyes. Therefore, sensori-motor transformation can be thought in terms of mapping the sensory coordinates associated with an object into motor coordinates.

Central to this approach is the idea that one may need to decompose the transformation in several steps, each of them using some intermediate frame of reference. Decomposing the transformation is not always required by the computation itself, but is important when integrating multiple signals which are all expressed in different coordinates. For instance, the sound and image generated by a single object cannot be merged directly into a single representation since they use different set of coordinates. A sensible approach would be to recode both sensory signals in a common frame of reference, superseding the ones used by these two modalities. Egocentric coordinates, that is, a coordinate system whose origin is centered on the trunk, would meet this requirement and would allow to combine all sensory signals within a common representation.

I.B The posterior parietal cortex

The posterior parietal cortex is often thought to be the primary brain area where such an egocentric multisensory representation can be found. Its pattern of input and output is certainly consistent with this hypothesis. It is located at the interface between the motor and sensory systems and it receives convergent input from multiple modalities while projecting to cortical prefrontal areas and subcortical motor nuclei (Andersen et al., 1990a; Felleman and Van Essen, 1991).

The most convincing evidence, however, comes from the study of the neu-

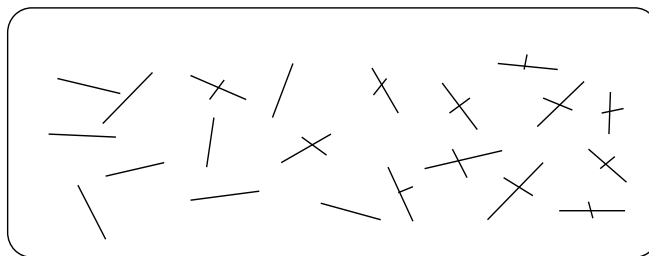


Figure I.1: Typical pattern of response in a line cancellation task from a right parietal patient. Lines on the contralateral side of the page tend to be missed.

rological disorders observed after lesions in the parietal cortex. One of the best documented of these deficit, called unilateral hemineglect, is characterized by a severe neglect to sensory stimulations, whether visual, auditory or somatosensory, toward the contralateral lesional side. Motor exploration is also greatly reduced compared to the intact ipsilateral side, even though it can be shown that there is no primary motor deficit.

The line cancellation task is one of the most common tests used to diagnose this syndrome. Patients are shown a sheet of paper with a group of short line segments with various orientations and they are asked to put a little check mark on each of these line segments. Whereas normal subjects have no trouble crossing out all the line segment, parietal patients invariably miss most of the segments located on the contralateral side of the page, where contralateral is defined with respect to the lesion side (figure I.1).

Several theories have been postulated to explain the nature of this syndrome. They roughly fall into two main categories: the attentional and the representational theories. Whereas the former invokes an inability of the subject to orient spatial attention to the contralateral side, the latter proposes that the syndrome results from deficit in the perception of the spatial position of objects in extrapersonal space (Posner et al., 1984; Rizzolatti and Berti, 1990). These theories are often considered mutually exclusive even though they do not really address and capture the same aspects of the syndrome. Even if the parietal cortex is an attentional center, spa-

tial attention has to operate on spatial representations. Therefore, it is reasonable to believe that the parietal cortex contains not only the machinery required for the guidance of spatial attention but also, the representations themselves. It is not even clear that the machinery for spatial attention and the spatial representations are physically distinguishable. One of the most important contribution of connectionist research has been to demonstrate that the distinction between representation and processor collapse in a neural network, and this principle might very well apply to this particular case.

I.C Neural code for spatial representations

Single cell recording experiments in behaving monkey provide additional support to the notion that the parietal cortex contains a neural code for spatial representations. Several theories have emerged from this studies, one defended by Goldberg and colleagues, and the other one by Andersen and his group. For Goldberg et al., the parietal cortex does not contain an egocentric representation, but it uses a retinotopic representation which behaves like an egocentric one thanks to a dynamic remapping mechanism which compensates for eye movements. Andersen et al. argues for a more explicit egocentric frame of reference.

I.C.1 Dynamic remapping

In order to foveate a visual target, the eyes must move by an amount which is precisely equal to the retinal position of the target. It is therefore possible for the visual system to specify a saccadic eye movement without having to encode the egocentric location of the target. However, there are reasons to believe that the target of saccadic eye movements are encoded in egocentric coordinates. Thus, human subjects can perform spatially accurate saccades toward two remembered visual target which are briefly flashed and extinguished before the onset of the first saccade (Hallett, 1976). Simply keeping track of the retinal location would not be sufficient to drive

the eye to the second target since after the first eye movement, the retinal location of the second target and its actual location in space are no longer in register. Therefore, the subject have to encode the egocentric location of the second target.

Goldberg et al. showed that these results do not entail that there is an explicit code for egocentric location of objects. Instead they propose that the brain remembers only the retinal location of targets, but dynamically remaps, or recomputes, this information, each time the eyes move. In the double saccade experiments described above, this interpretation would predict that subjects compute where the second target would appear on the retina after the first eye movement and then use this information to drive the eyes to the appropriate location. Such a mechanism fully accounts for the observed results without ever using an explicit code in egocentric coordinates.

The outcome of several neurophysiological studies in the lateral intraparietal area (LIP), in the frontal eye field (FEF) and in the superior colliculus, support this dynamic remapping hypothesis. Cells whose receptive fields are located where the second target would appear after the first saccade, start firing during, or sometime slightly before, the first saccade even if the second target has been already extinguished (Mays and Sparks, 1980; Gnadt and Andersen, 1988; Goldberg and Bruce, 1990; Duhamel, Colby and Goldberg, 1992).

This remapping can be easily implemented in a biologically plausible architecture. A recurrent neural network model by Droulez and Berthoz (1991) performs the remapping by moving a hill of activity across a retinotopic map by an amount proportional to the amplitude of the first saccade. The hill motion can be controlled by integrating eye velocity or some approximation of it during the first saccade (Droulez and Berthoz, 1991; Dominey and Arbib, 1992).

It would be tempting to generalize this mechanisms to all sensori-motor transformations, but unfortunately, in its present form, this scheme works only for eye movements. Simply knowing where an object appear on the retina or where it would appear if it was visible, is not sufficient to specify an arm movement toward the

object. Current eye position and other posture signals must be combined with the visual inputs, an integration process which has been extensively studied by Andersen and his group.

I.C.2 Gain modulation

Computing the egocentric position of an object from its visual image requires the retinal location of the object to be added to the current gaze angle. This simple geometrical relationship suggests that cells involved in this type of computation should be responsive to both visual stimulation and eye position. This is what Andersen and his colleagues found in parietal areas LIP and 7a, where typical cells have a large receptive field whose position is fixed on the retina but whose gain, i.e., the amplitude of the response to a visual stimulation, is modulated by eye position (Andersen, Essick and Siegel, 1985).

Although this pattern of response shows that the cells receive the appropriate inputs, it is somewhat surprising. One could have imagined that the parietal cortex contains an egocentric map in which neurons had fixed receptive fields in egocentric coordinates. Recently, Battaglini et al. have claimed that some neurons in parietal area V6 have such egocentric receptive fields (Galletti, Battaglini and Fattori, 1993), but the functional significance of this results is difficult to assess since it concerns only a very small percentage of the cells ($<3\%$).

In contrast, a very large percentage of cells ($\approx 66\%$), are gain modulated in LIP and 7a (Andersen and Brothie, 1992). Moreover, a neural network model by Zipser and Andersen have demonstrated that these gain modulations are sufficient to represent the egocentric location of object (Zipser and Andersen, 1988a).

I.C.3 The Zipser and Andersen Network

The Zipser and Andersen model is a three-layer neural network trained with backpropagation to recover the egocentric location of an object from its retinal lo-

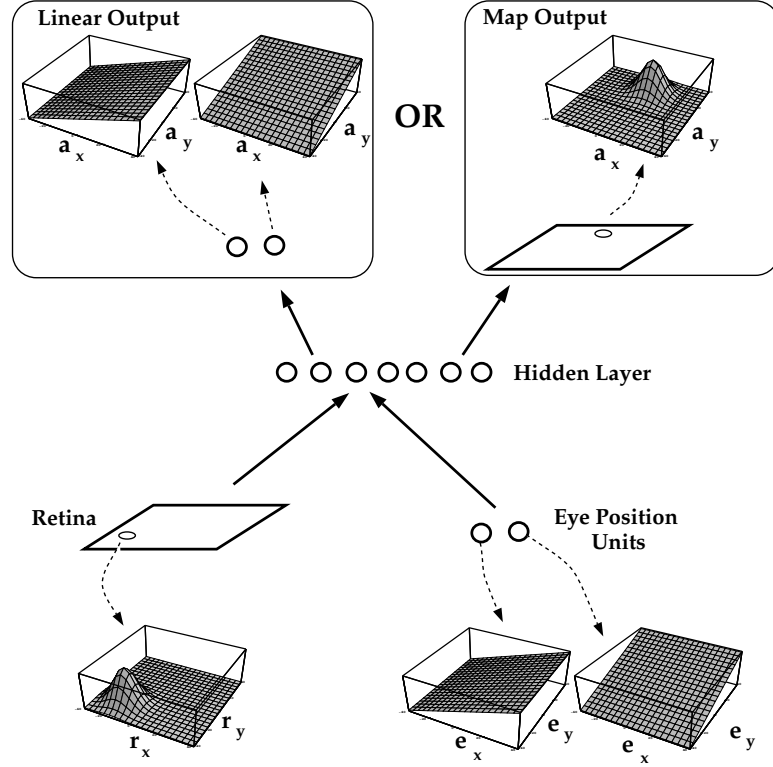


Figure I.2: The Zipser and Andersen network for computing the egocentric position of an object, \vec{A} , from its retinal location, \vec{R} , and current eye position, \vec{E} .

cation and the current eye position. The input layer contained a retina and units encoding eye position with a rate coding, whereas the output was either a map in egocentric coordinates or units encoding the egocentric location of the object with rate coding (figure I.2).

In either case, the hidden units of the trained network were found to have response properties comparable to parietal neurons. In particular, they had fixed retinal receptive field, gain modulated by eye position. This was the first demonstration that gain modulation is sufficient to represent the location of object in egocentric coordinates. Unfortunately, the network model did not provide any additional insight in the nature of the spatial representation used by the network and the parietal cortex. Nonlinear multi-layer networks are often difficult to analyze, and at the time this work was developed, there was no standard method to interpret the hidden representation.

Additional work by Goodman and Andersen (1990) attempted to characterize more precisely the nature of the representation. Their approach consisted in training networks with a minimum number of hidden units, in order to find out what would be the most compact representation. They discovered that two hidden units were sufficient to solve the problem. In this minimal configuration, the units behave like linear units adding the eye and retinal position along a particular direction, called the preferred direction. Since the input vectors, retinal location and eye position, were bidimensional, two independent directions or, equivalently, two hidden units are sufficient to perform the addition (Goodman and Andersen, 1990).

This type of encoding scheme is known as a vectorial representation in the literature (Soechting and Flanders, 1992). In a vectorial representation, each unit computes the projection of the vector to be encoded along its the preferred direction. The best-known case of a vectorial code is the encoding of the direction of hand movement in the primary motor cortex. Schwartz and colleagues (Schwartz, Kettner and Georgopoulos, 1988) have shown that the tuning of single cells in the primary motor cortex to direction of hand movements can be approximated by a cosine function. It can be shown that a cosine tuning is precisely the tuning characteristic expected from linear units (Schwartz, Kettner and Georgopoulos, 1988; Sanger, 1994)

The hidden units in the minimal neural network trained by Goodman and Andersen perform the same operation. In larger networks, the responses of the hidden units becomes more and more nonlinear and do not simply adds the input signals together. However, Goodman and Andersen argue that since the overall task of the network is still the same, the representation in the hidden layer is an approximation of a vectorial representation (Goodman and Andersen, 1990).

Therefore, this line of work suggests that the parietal cortex made indeed contain a vectorial representation of the egocentric position of objects. These conclusions are fully consistent with the classical approach to sensori-motor transformation. It still assumes that representing the spatial characteristic of an object amounts to encode a position vector in a particular frame of reference.

I.D Parallel vectorial representations

The original studies of Andersen et al considered only visual inputs and eye position. Since then, the same group has shown that other posture signals such as head position (Brotchie and Andersen, 1991) or vestibular inputs (Snyder, Brotchie and Andersen, 1993) are also integrated by parietal neurons. These findings indicates that more than one frame of reference might be used by the parietal cortex.

The possibility that the brain might use several frames of references in parallel had been previously suggested by psychophysical experiments (Ladavas, 1987; Soechting and Flanders, 1990). These results have led to the view that there might exist a group of functional modules throughout the brain, each of them specialized for a particular type of sensori-motor transformation, and each of them using their own frame of reference to encode the location of objects (see Stein, 1992, for a review). Hence, single cell response in the frontal eye field (FEF) of the lateral intraparietal area (LIP) indicates that these two centers are mostly devoted to the control of eye movements from visual and possibly auditory inputs (Gnadt and Andersen, 1988; Barash et al., 1991a; Stricanne, Mazzoni and Andersen, 1993). In contrast, parietal area 7b and the hand region of area 6, are more specialized in visually-guided hand movement (Mountcastle et al., 1975; Gallese et al., 1994).

This elaboration on the classical view still assumes that each modules encode vectors, presumably with a vectorial code as the work of Goodman and Andersen (1991) would suggest, but possibly with maps in places like the superior colliculus (Stein and Meredith, 1993).

I.E Thesis Outline

The work presented in this dissertation challenges the classical approach to sensori-motor transformation in two ways.

We first ask whether spatial transformations may be initiated before the parietal cortex, possibly as early as visual area V1. This study, described in the

second chapter, was prompted by recent neurophysiological recordings indicating that a subset of neurons in early visual areas (V1 and V3a) have response properties which are reminiscent of parietal neurons (Galletti and Battaglini, 1989; Trotter et al., 1992).

Second, instead of approaching sensori-motor transformations as a change of coordinates, we formalize the problem in term of nonlinear function approximation. This approach suggests a new type of spatial representation which does not involve coding position vector.

I.E.1 Hierarchy in spatial representations

The retinotopy of the visual maps in early visual areas and the absence of a specific spatial deficit after a stroke in this part of the cortex appears to argue against the possibility that spatial transformations start in the early stages of the visual hierarchy. Interestingly, however, a subset of cells in area V1, V3a and even in the LGN show similar gain modulation to the ones that have been documented in the parietal cortex (Lal and Friedlander, 1989; Galletti and Battaglini, 1989; Trotter et al., 1992). We discuss whether these results can be taken as evidence for early spatial transformations.

We use a network whose architecture is inspired by the hierarchy of retinotopic maps found in the early visual cortex (Felleman and Van Essen, 1991). We train it to perform a spatial transformation and then study the response properties of the hidden units. We show that, just like in the Zipser and Andersen network, the responses of the hidden units resemble the ones recorded from neurons with similar visual receptive field size.

We also simulate electrical stimulation experiments to show that the pattern of eye movements observed upon stimulation on early visual maps does not reveal the coordinate used by these maps, as it is often assumed. Hence, V1 stimulation lead to fixed-vector saccade, a pattern which is typically associated with retinotopic coordinates. Yet, we find the same results when stimulating our network, even though we know that the coordinates are not retinotopic.

I.E.2 A basis function approach for spatial representation

In chapter 3 and 4, we explore an alternative to the classical view on sensori-motor coordination. Instead of thinking in term of change of coordinates and vector encoding, we approach sensori-motor transformations as a problem of nonlinear function approximation. Generating a motor command in response to a set of sensory inputs involves approximating a nonlinear function of the sensory signals. There are many ways to approximate nonlinear functions, but one of the most robust and powerful method consists in decomposing the function in a linear combination of basis functions (Poggio, 1990). If the brain were to use such a method, this would entail that each neuron computes a basis function of its sensory input, as opposed to a vector projection along its preferred direction.

In chapter 3, we apply this framework to the computation of egocentric distance. A large percentage of disparity selective cells in area V1- cells that are sensitive to the relative depth of objects- are modulated by the distance of fixation (Trotter et al., 1992). We demonstrate that these modulations are consistent with the basis function hypothesis but not with more explicit types of coding. We also show that the way human perceive distance is a compromise between a biased, yet reliable, estimate of how far an object is located from the viewer.

In chapter 4, we consider the position of object within the fronto-parallel plane. We show the relevance of the basis function approach for interpreting the results of Andersen et al and we compare vectorial representation with our own approach. We also discuss the notion of representation and frame of reference within the basis function framework.

Chapter II

Egocentric spatial representation in early vision

Abstract

Neurons encoding simple visual features in area V1 such as orientation, direction of motion and color are organized in retinotopic maps. However, 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 retinal 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.

II.A Introduction

Three main types of coordinates for representing visual objects have been proposed by researchers in neurobiology, computer vision and psychophysics: eye-centered (retinal), object-centered and viewer-centered (egocentric). Eye-centered representations have been proposed for early vision mainly because the 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.

The other two types of representation are believed to be used at the highest stages of visual processing. Mishkin et al. (Mishkin, Ungerleider, & 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 (Goodale & Milner, 1990) have 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 which 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

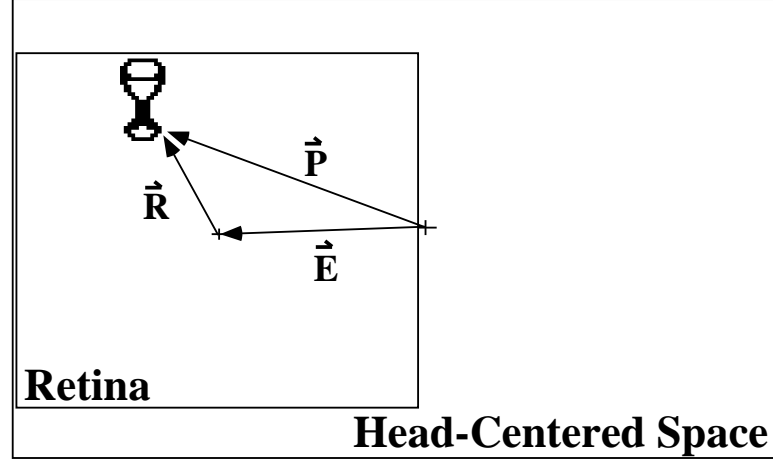


Figure II.1: The angular position of an object with respect to an origin fixed in head-centered space, \vec{P} , is the vector sum of the retinal position vector, \vec{R} , and the eye position vector, \vec{E} , which joins the origin in head-centered space to the origin on the retina.

by (see Fig. 1):

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

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 equation 1, the position would be in body-centered coordinates.

Physiological and psychological data support only the existence of head-centered representations. However, in none of these experiments were the head or body position varied, which leaves open the possibility that these representation are in fact body-centered, or even spatiotopic. Consequently in this paper, we use the term 'egocentric' in a non-specific way. 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 egocentric coordinates is in the posterior parietal cortex (PPC), which we review here before considering the case for early visual areas. Following ischemic lesions in the PPC, patients often display egocentric neglect, though of variable extent (Heilman, Watson, & Valenstein, 1985). This neglect usually extend to multiple sensory modalities, typically vision, audition and touch, in the contralateral side of the lesion.

Single cell recording in the PPC have 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; Brochier & Andersen, 1991). In the simplest type of spatiotopic representation, neurons 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 have been found with spatiotopic receptive fields (Andersen, et al., 1985; Brochier & Andersen, 1991). However, the magnitude of the responses to visual stimuli of many neurons in PPC are modulated by both eye and head position. Zipser and Andersen have developed a network showing that these response properties are consistent with a distributed representation of egocentric space (Andersen & Zipser, 1988; Goodman & Andersen, 1989; Goodman & Andersen, 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 showed that the hidden units of three-layer networks trained to compute the egocentric position of an object exhibited response properties similar to those of neurons in posterior parietal area 7a and the lateral intraparietal area (LIP). Each hidden units 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 organization

of the receptive field 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 found in their network.

The study of saccadic eye movements elicited by electrical stimulation experiments also support the hypothesis that the PPC contains an egocentric representation of space. In LIP, the amplitude, and to a certain extent the direction, of the evoked saccade appears to be sensitive to the initial eye position (Fig. 2-B)(Kurylo & Skavenski, 1991; Shibutani, Sakata, & Hyvarinen, 1986; Thier & Andersen, 1992). As shown by Goodman and Andersen these saccades are consistent with the distributed representation found in the Zipser and Andersen model of the PPC (Goodman & Andersen, 1989). Even more compelling evidence come from stimulation of the ventral intraparietal area (VIP) in which the elicited saccades 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. 2-C). However in this case the correspondence with single cell response 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.

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 & Friedlander, 1989), primary visual cortex area V1

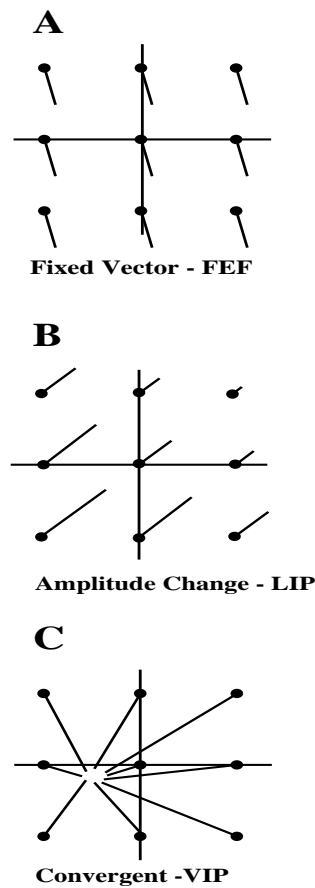


Figure II.2: Schematic of three main patterns of saccadic eye movements evoked by electrical stimulations. The large dots indicates the initial eye position while the lines show the amplitude and direction of the evoked saccades. A.) Fixed vector type. Amplitude and direction are independent of initial eye position. Typical of stimulation in Frontal Eye Field (FEF) (Robinson & Fuchs, 1969) B) Amplitude varying saccades (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).

(Trotter, Celebrini, Thorpe, & Imbert, 1991; Weyand & Malpeli, 1989) and extrastriate area V3a (Galletti & Battaglini, 1989). The static eye position signal seems to mainly control the gain of the neuronal response without changing the selectivity of the cell. As in the PPC, could early cortical visual areas, or even the LGN, be using egocentric representations? The small receptive fields and the retinotopy of those early maps argue against this possibility and psychophysical observations have led to the same conclusion. Furthermore, strong negative evidence comes from a study by McIlwain 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. 2-C).

We attempt to reconcile the conclusions drawn from electrical stimulation experiment 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 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.

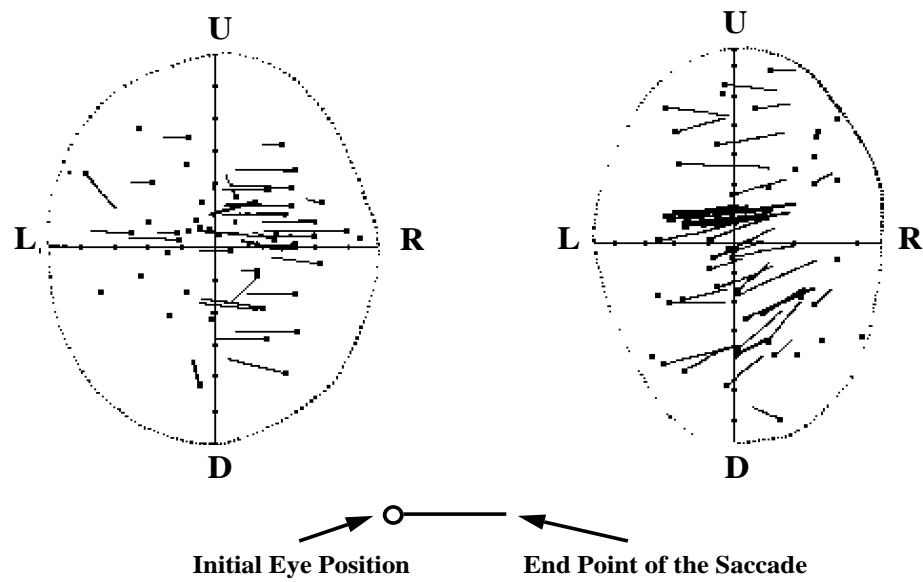


Figure II.3: Saccadic eye movements induced by electrically stimulating cortical area V1. Left plot: Stimulation on a site in V1 representing a position in the visual field 24 left 4 down. Right plot: same for a position 55 right, 12 up. In both cases, the direction of the saccades were primarily determined by the position of the stimulation on V1 and is largely independent of the initial eye position (from (Mc Ilwain, 1988)).

II.B Results

II.B.1 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 equation 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 11x11 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. Within a given layer, the size of all receptive fields were the same 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 3-5 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 et al. (LeCun, Boser, Denker, Henderson, Howard, Hubbard, et al., 1990).

Weights were adjusted using the backpropagation training algorithm (Rumel-

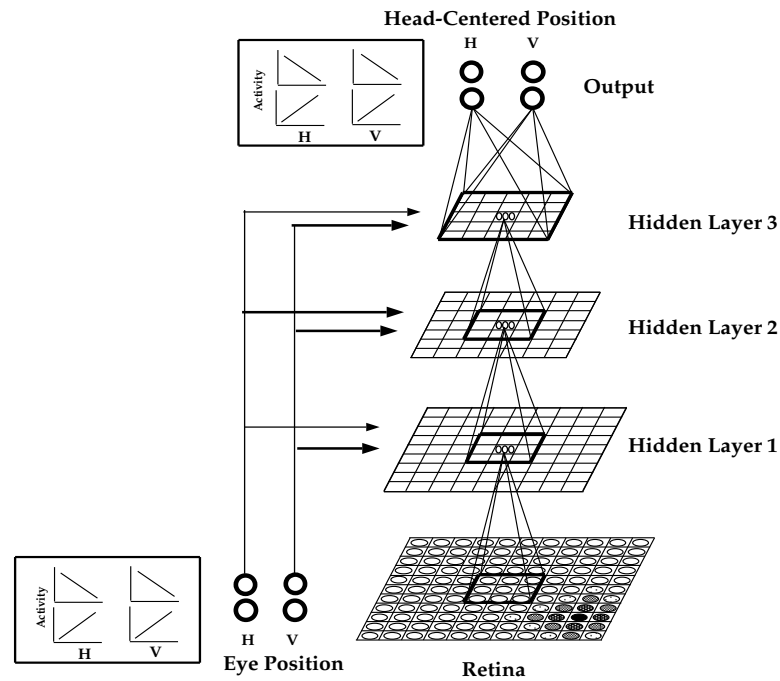


Figure II.4: Network architecture. The input layer is a 2-D retina and four units encoding horizontal (H) and vertical (V) eye position, as shown in the lower left corner. The four units on the output layer compute the egocentric position of a gaussian pattern of light shown on the retina. The hidden layers form a hierarchy of retinotopic maps, each receiving converging inputs from a lower layer and eye position signals. Each hidden layer has 3 to 5 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.

hart, Hinton, & Williams, 1986). All simulations were performed with the SN2 simulator developed by Botou and LeCun (Neuristic). Training examples 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 (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.

II.B.2 Comparison Between Hidden Units and Cortical Neurons

The influence of eye position on the visual response of a cortical neuron is usually 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 real neurons (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 (Zipser & 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 3 to 5 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 (Zipser & Andersen, 1988). The major difference is that the hidden units in our model 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 a 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 2 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

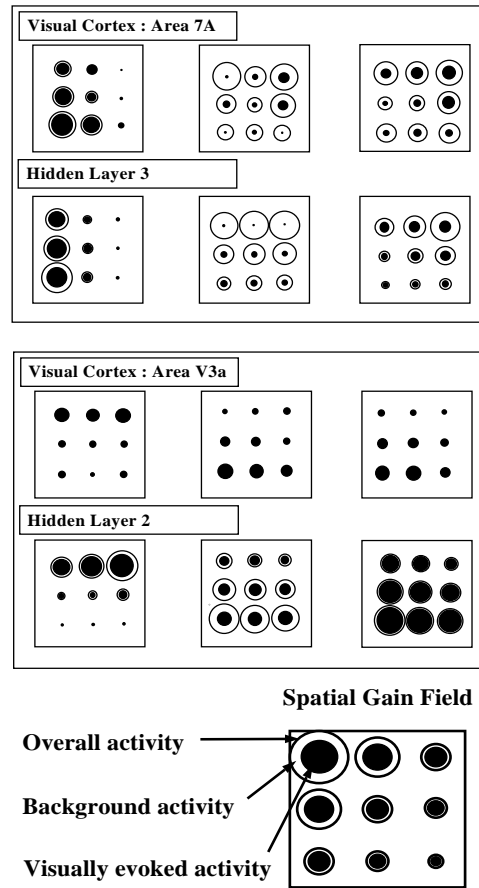


Figure II.5: Comparison between the responses of cortical neurons (Andersen & Zipser, 1988; Galletti & Battaglini, 1989) and hidden units. Each box indicates the response of a single cell to the same retinal stimulation for nine initial eye positions. Activities of the units and the neurons are plotted as 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 those of real neurons.

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.

More elaborate visual selectivities are found in visual cortex. Cortical areas V1 or V3a encode relatively simple visual attributes and at least 40% therefore expect V1 and V3a to encode low-level visual feature 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.

In conclusion, 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).

II.B.3 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 equation 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 an 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 (Mc Ilwain, 1988).

Stimulation in the hidden layers led to different results depending on how many units were stimulated per position. Each location had between 3 to 5 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 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.

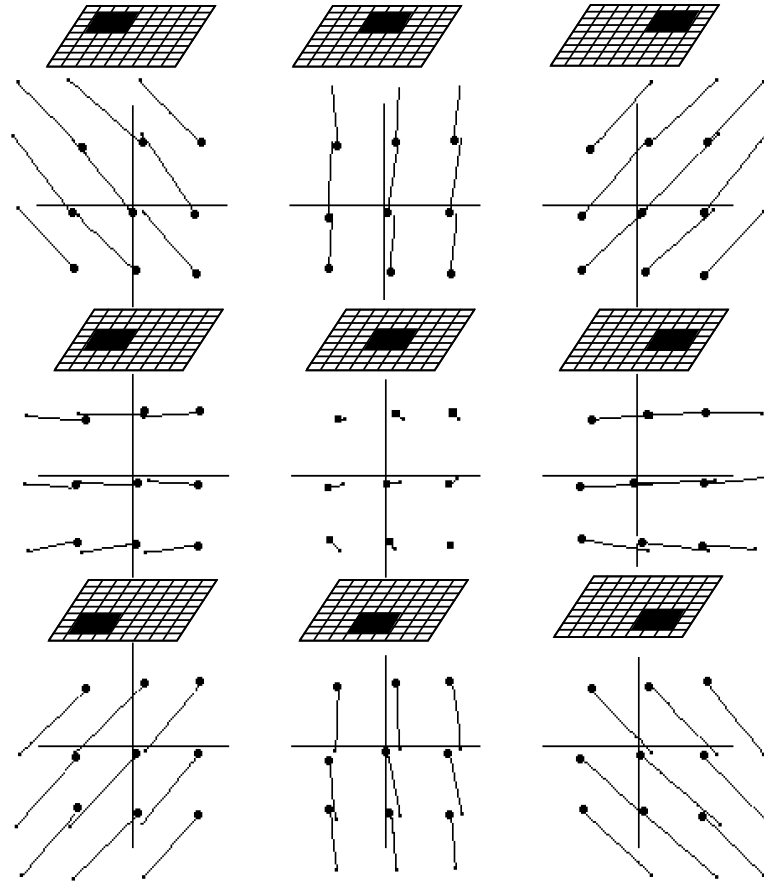


Figure II.6: Eye movements evoked by stimulating the retinal input layer of the network. Each of the nine plots shows the evoked saccades from nine initial eye positions in response to identical stimulation performed at the location illustrated on the grid above each plot. Notice that the evoked eye movements are always in the direction of 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.

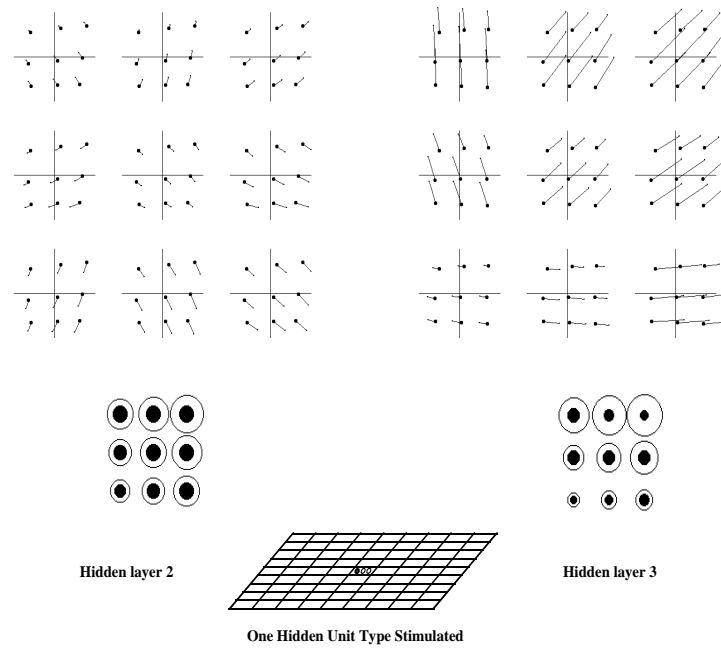


Figure II.7: Eye movements evoked by stimulating one hidden unit type at nine different locations in the intermediate (left) and top (right) hidden layers. Same plotting convention as Figures 2 and 6. The spatial gain field of the unit being stimulated is shown below the corresponding plot. In hidden layer 3, the direction of the saccades depend 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).

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 (Mc Ilwain, 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.

II.B.4 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 shown, in such conditions, even if static eye position modulates the gain of V1 neurons, one would expect to obtain fixed vector saccades which are consistent

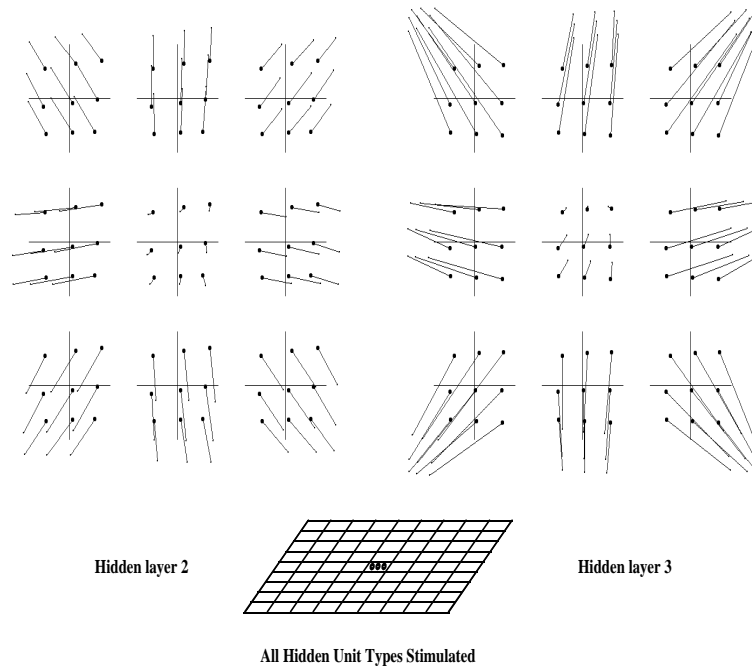


Figure II.8: Eye movements evoked by stimulating all the hidden units sharing the same location. In both layers the pattern of eye movements is very similar to the one induced by stimulating the input layer (Fig. 6), even though the input and hidden layers do not use the same type of spatial representation.

with the results reported by McIlwain. (Mc Ilwain, 1988). Furthermore, many 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. 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.

II.C Discussion

We have demonstrated that the study of eye movements evoked by electrically stimulating early visual areas is unlikely to distinguish between retinal and retinospacial coordinates. In higher visual areas with weaker retinotopy though, it might be possible to obtain patterns closer to those produced by stimulating only one type of hidden unit. If the retinotopy was not as good as it is in our model, and neurons with similar spatial gain fields tended to cluster together, then stimulation in a higher visual area might induce eye movements whose direction would be mainly a function of the preferred eye position direction (PEPD). This pattern of eye movements has already been observed in parietal area LIP (Goodman & Andersen, 1989, Thier91).

II.C.1 Network output representation

The type of units that we used in the output layer of our network, namely units which 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 could really affect our conclusions. We believe that it does not, mainly because we do not make any claim about the development of the representation found in the hidden layer and the actual mechanisms that create these representations in the brain. Once the network has reached its final state, the output layer is just a way to demonstrate the type of information available in the connection sent by the last layer of hidden units. 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 presents similarities with early visual areas and that it contains an implicit representation of egocentric position, as we discuss in the next section.

Similarly, for the electrical stimulation experiments the exact nature of the output representation is not critical. We obtained retinotopic saccades upon stimulation of all the units sharing the same position on one of the hidden layer, through averaging of 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.

II.C.2 Explicit vs implicit 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 and 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 position of objects are needed, such as the

control of head movements. This is a direct corollary of the two criteria defined above.

II.C.3 Amplitude of saccades evoked by electrical stimulation

There is one aspect of McIlwain's result that our model cannot account for: the amplitude of the saccades evoked by electrically stimulating the striate cortex. From the data shown in Figure 3, it is clear that the amplitude of the saccade 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.

There is an alternative interpretation of the data, as discussed by McIlwain. Cats make head saccades as often as eye saccades to acquire a visual target. Electrical stimulation in area V1 might induce head as well as eye movements in combination to acquire a phantom target; but because the head of the cat was held fixed during the stimulation experiments, only the eye movement component was measured. According to this interpretation, the decrease in saccade amplitude would be accompanied by an increase in the head component as the initial eye position moved further into the hemifield ipsilateral to the phantom target. If this interpretation is right, it would suggest that the change in amplitude is not related to the spatial representation used in V1.

II.C.4 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 occurs at a late stage of visual processing (see for example (Otto, Grandguillaume, Boutkhil, Burnod, & Guigon, 1992)). 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 (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 which had been split in two halves presented 50 ms 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 subject could report the words if both halves were flashed on the same retinal location, but not in conditions under which only the spatial location was maintained.

Another example is a study by Irwin et al. (Irwin, Zachs, & 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 ms 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 seems to support the possibility for egocentric representations in early vision (Kohler, 1964; Mayhew, 1973). These experiments looked for after-effects contingent on eye position. After seeing repetitively clockward motion while looking left and anti-clockwise motion while right, subjects reported an anti-clockwise 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-effect with translation motion or other elementary visual attribute such as orientation or disparity. Such result 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 (Mc Cullough, 1965). Color may not be

one of the visual attributes that are combined with eye position since the areas where eye-position modulation have 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.

II.C.5 Why retinospacial 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 map from RSM. Consider, for example, the orientation masking by Irwin et al. (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 V1 it is difficult to imagine how it could be recognized. Even if the orientational spatiotopic maps in V1 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.

II.C.6 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 ((Matin & 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 (ERS) 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 (MacKay, 1970) and O'Regan (O'Regan, 1984), but a recent experiment performed by Gauthier et al. (Gauthier, Nommay, & Vercher, 1990) provides new evidence in favor of the ERS. 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 ERS.

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 RMS 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.

II.C.7 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., 1991)(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 (in this case vergence angle). Such neurons have already been reported in V1 (Trotter, et al., 1991).

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

II.D 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, but may encode egocentric coordinates in the earliest stages of visual processing.

Acknowledgement

This chapter, in full, is a reprint of the material as it appears in *Journal of Cognitive Neuroscience*, 5(2):150-161, 1993. It was written in collaboration with S.A. Fisher and T.J. Sejnowski. It is reprinted with permission from MIT Press and S.A. Fisher.

References

- Andersen, R. A. (1989). Visual and eye movement functions of the posterior parietal cortex. *Ann. Rev. Neurosci.*, 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. *Can. J. Physiol. Pharmacol.*, 66, 488-501.
- Brotchie, P. R., & Andersen, R. A. (1991). A body-centered coordinate system in posterior parietal cortex. *Abst. Soc. Neurosci.*, 17, 1281.
- Feldman, J. A. (1985). Four frames suffice: A provisional model of vision and space. *Behavior. Brain Sci.*, 8, 265-289.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cer. Cortex*, 1, 1-47.
- Galletti, C., & Battaglini, P. P. (1989). Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. *J. Neurosci.*, 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 Neurosci.*, 15, 20-25.
- Goodman, S. J., & Andersen, R. A. (1989). Microstimulation of a neural network model for visually guided saccades. *J. Cog. Neurosci.*, 1, 317-326.
- Goodman, S. J., & Andersen, R. A. (1990). Algorithm programmed by a neural model for coordinate transformation. *Proc. Inter. Joint Conf. Neur. Net.*, 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. *Cog. Psychol.*, 23, 420-456.
- Irwin, D. E., Brown, J. S., & Sun, J. (1988). Visual masking and visual integration across saccadic eye movements. *J. Exp. Psychol.*, 117, 276-287.
- Irwin, D. E., Zachs, J. L., & Brown, J. S. (1990). Visual memory and the perception of a stable environment. *Percep. Psychophy.*, 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 Res.*, 1, 213-215.
- Kohler, I. (1964). The formation and transformation of the perceptual world. *Psychol. Issues*, 3, 62-86.

- Kurylo, D. D., & Skavenski, A. (1991). Eye movements elicited by electrical stimulation of area PG in the monkey. *J. Neurophysiol.*, 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 Networks*, 3, 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*, 227, 731-733.
- Mateef, S. (1978). Saccadic eye movements and localization of visual stimuli. *Percept. Psychophys.*, 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 Res.*, 13, 877-880.
- Mc Cullough, C. (1965). Conditioning of color perception. *Am. J. Psychol.*, 78, 362-378.
- Mc Ilwain, J. T. (1988). Saccadic eye movements evoked by electrical stimulation of the cat visual cortex. *Vis. Neurosci.*, 1, 135-143.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends Neurosci.*, Oct, 414-417.

O'Regan, J. K. (1983). Integrating visual information from successive fixations : Does trans-saccadic fusion exist? *Vision Res.*, 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. *Percept. Psychophys.*, 36, 1-14.

Otto, I., Grandguillaume, P., Boutkhil, L., Burnod, Y., & Guigon, E. (1992). Direct and indirect cooperation between temporal and parietal networks for invariant visual recognition. *J. Cog. Neurosci.*, 4, 35-57.

Pollatsek, A., Rayner, K., & Henderson, J. M. (1990). Role of spatial location in integration of pictorial information across saccades. *J. Exp. Psychol. : Hum. Percept. Perf.*, 16, 199-210.

Rayner, K., & Pollatsek, A. (1983). Is visual information integrated across saccades? *Percept. Psychophys.*, 34, 39-48.

Richards, W. (1968). Spatial remapping in the primate visual system. *Bio. Cyber.*, 4, 146-156.

Robinson, D. A., & Fuchs, A. F. (1969). Eye movements evoked by stimulation of the frontal eye field. *J. Neurophysiol.*, 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. *Exp. Brain Res.*, 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. *Exp. Brain Res.*, 55, 1-8.

Sun, J. S., & Irwin, D. E. (1987). Retinal masking during pursuit eye movements: Implications for spatiotopic visual persistence. *J. Exp. Psychol.: Hum. Percept. Perf.*, 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. *Abst. Soc. Neurosci.*, 17, 1281.

Trotter, Y., Celebrini, S., Thorpe, S.J., & Imbert, M. (1991). Modulation of stereoscopic processing in primate visual cortex V1 by the distance fixation. *Abst. Soc. Neurosci.*, 17, 1016.

Weyand, T. G., & Malpeli, J.G. (1989). Responses of neurons in primary visual cortex are influenced by eye position. *Abst. Soc. Neurosci.*, 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*, 331, 679-684.

Chapter III

A Neural Model of the Cortical Representation of Egocentric Distance

Abstract

Neurons in the visual cortex of monkeys respond selectively to the disparity between the images in the two eyes. Recent recordings have shown that some of the disparity-selective neurons in the primary visual cortex and the posterior parietal cortex are modulated by the distance of fixation. A population of such gain-modulated, disparity-selective neurons forms a set of basis functions of horizontal disparity and distance of fixation that can be used as an intermediate representation for computing egocentric distance. This distributed representation is consistent with psychophysical studies of human depth perception; in contrast, neurons explicitly tuned to distance are not consistent with how we perceive distance. In a population model that includes noise in the firing rates of neurons, the perceived distance is shown to be the estimate of geometrical distance that minimizes the variance of the estimation.

Introduction

There are many visual cues to estimate the depth of an object, including stereopsis, motion parallax, shape from shading and occlusion. These are relative depth cues and do not provide sufficient information to recover the position of an object with respect to the viewer (egocentric distance). For stereopsis, the depth of an object is measured relative to the fixation point of the two eyes, as illustrated on figure 1. Because each eye sees an object from a slightly different angles, the images are slightly displaced. The horizontal displacement, or horizontal disparity, is proportional to the distance between the object and the fixation point: zero disparity corresponds to an object at fixation, whereas positive and negative disparities correspond, respectively, to locations in front of or behind the fixation point (Fig. 1).

If the distance to the fixation point, also called the viewing distance, can be recovered, then the position of an object in egocentric coordinates can be estimated by combining this information with horizontal disparity. Psychophysical experiments indicate that at least two cues are used by the visual system to recover the viewing distance: the vergence angle and vertical disparities (Cumming et al. 1991; Ritter 1977; Rogers and Bradshaw 1993; von Hofsten 1976; von Hofsten 1977). For simplicity, we consider only vergence in this paper, but our approach can be easily extended to include vertical disparity. For an object directly in front of the viewer, the distance, D , as a function of disparity, δ , vergence, ν , and interocular distance, I , is given by:

$$D = \frac{I}{2\tan(\frac{\nu-\delta}{2})} \quad (\text{III.1})$$

A three-dimensional plot of this function, with respect to ν and δ , is shown in figure 8A. Although psychophysical experiments have demonstrated that the humans can estimate absolute distance, this problem has only recently been studied with physiological methods (Gnadt and Mays 1991; Trotter et al. 1992; Trotter et al. 1991). pioneering work on the cat visual cortex (Nikara et al. 1968; Pettigrew et al. 1968) and later studies in a variety of species revealed that a large percentage

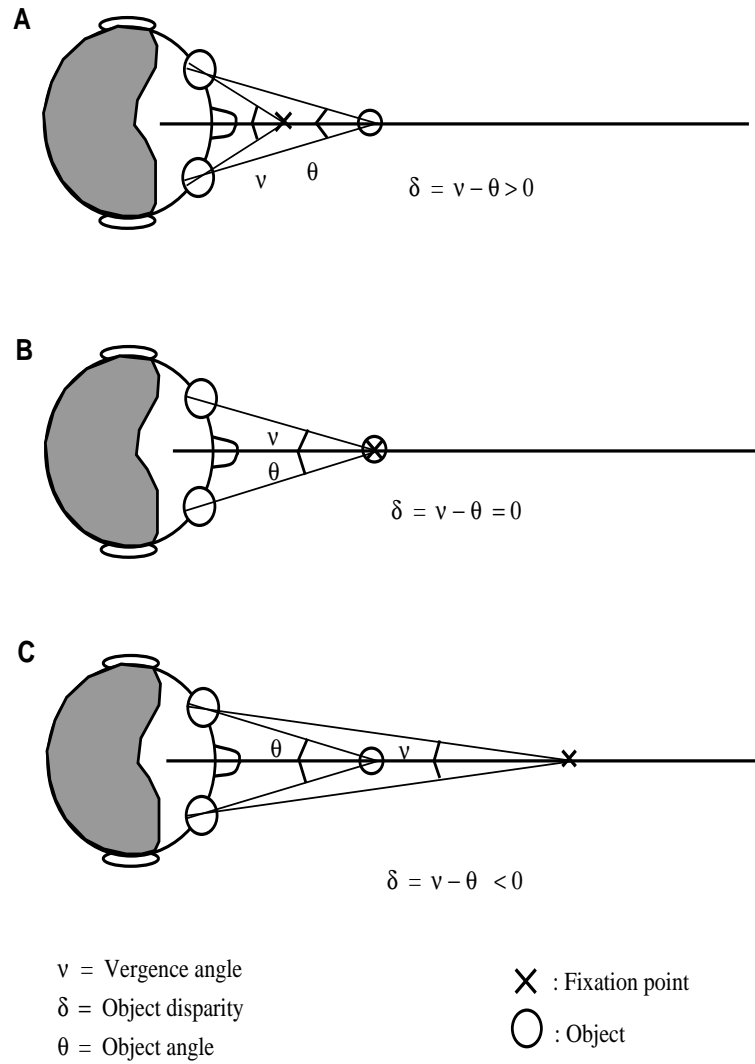


Figure III.1: Viewing geometry for stereopsis. The horizontal disparity, δ , of a stimulus is equal to the difference between the vergence angle, v , and the object angle θ (sometimes called the absolute disparity of the object). The disparity δ depends on the position of the object with respect to the fixation point, which is shown in three positions: A) beyond fixation, B) at fixation and C) in front of fixation. The egocentric distance of an object can be determined by combining a disparity measurement with vergence angle or any cue related to distance of fixation.

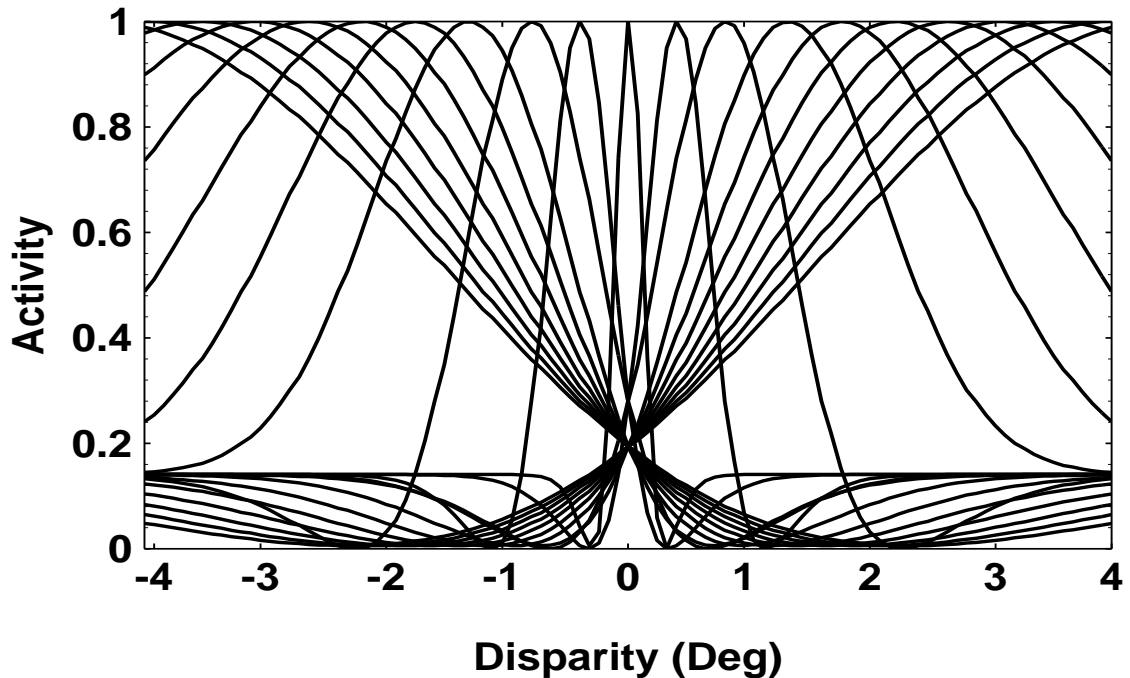


Figure III.2: Idealized disparity tuning curves of cortical neurons. Each curve corresponds to the response of a single cell to disparity (see Eqs. 3-5). The curves illustrated here were sampled from a continuum of tuning curves. The widths of the tuning curves increase as the best disparities at the response peaks increase.

of cells in the primary visual cortex and extrastriate areas are selective to horizontal disparity (Ferster 1981; Hubel and Wiesel 1970; LeVay and Voigt 1988; Maunsell and van Essen 1983; Poggio 1984; Poggio and Fischer 1977; Poggio et al. 1988; Poggio et al. 1985). Poggio classified disparity-selective neurons into three groups, the near, tuned and far cells. Tuned cells were characterized by narrow tuning with a peak close to zero disparity whereas near and far neurons showed broad tuning with peak responses at large values of disparity, positive for near cells and negative for far cells. LeVay and Voigt (1988) and Lehky and Sejnowski (1990) suggested these classes were only the extremes in a continuum of selectivities, and recent psychophysical studies provide confirming evidence (Cormack et al. 1993; Stevenson et al. 1992). Figure 2 shows the full spectrum of disparity-tuned responses that are typically found in the visual cortex.

In the above studies, the disparity selectivity of neurons was tested at a fixed distance, which confounded disparity and distance. The influence of these two variables can be distinguished by measuring the disparity selectivity of cells over a range of fixation distances. Figure 3 illustrates the two types of idealized responses that one might expect to find in such an experiment. One possibility is that the disparity tuning curve would be invariant under changes in distance of fixation: this would imply that the peak of the distance tuning curve should vary with the distance of fixation (Fig. 3A). The opposite extreme would be a cell that was tuned to distance, but with a disparity tuning that varied with distance to fixation (Fig. 3B).

This experiment has been performed by Trotter et al. (1991, 1992) in V1, and Gnadt and Mays (1991) in the parietal cortex. They recorded the disparity tuning curves of cortical neurons for various viewing distances and found selectivities intermediate between these two extremes. The magnitudes of the responses to disparity were modulated by the distance to fixation but the shape and position of the peaks of the tuning curves were unchanged. Figure 4 shows four examples of these neurons (Trotter et al. 1992). Since the position of the peak of the disparity selectivity is invariant with changes in viewing distance, these neurons were not tuned to distance, but neither were they unambiguously tuned to disparity. This type of response was not unexpected since a previous neural network model trained to compute distance from pairs of vergence angles and disparity-tuned units exhibited such "gain" fields for distance (Lehky et al. 1990). However, this network model only provided limited insight into the nature of the representation used by the brain (see discussion). The goal of this paper is formally to characterize the representation found in the brain (Gnadt and Mays 1991; Trotter et al. 1992) and in neural network models (Lehky et al. 1990) and to explore the computational advantages of this representation.

Our analysis relies on the theory of basis functions. This theory takes its root in 19th century mathematics, but it has only recently been applied to neural networks (Casdagli 1989; Moody and Darken 1989; Poggio 1990). Specifically, we show that the gain-modulated neurons reported in the visual cortex forms a set of basis functions for

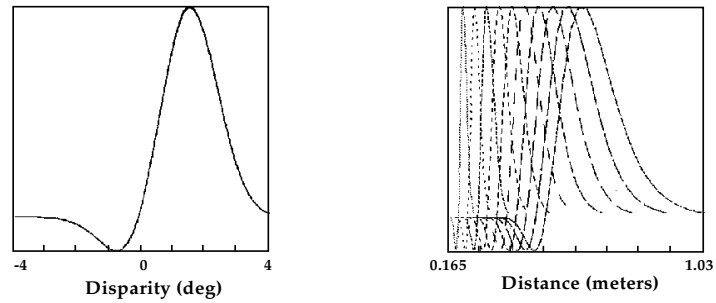
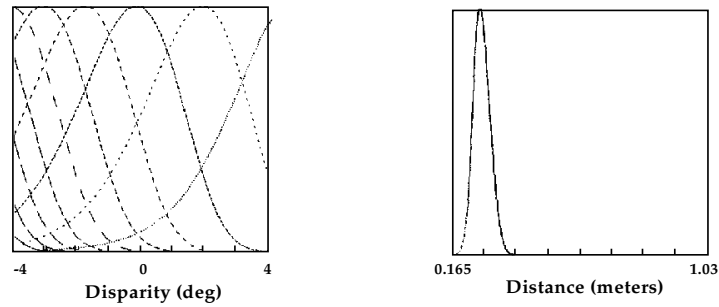
A- Disparity Detector**B- Distance Detector**

Figure III.3: Responses of idealized disparity and distance detectors. A) (left panel) Response of a pure disparity detector as a function of disparity. and (right panel) as a function of distance, for 10 fixation distances (dashed lines). The response of a disparity detector to horizontal disparity would be unaffected by changes in fixation distance. B) (left panel) Response of a pure distance detector is shown as a function of disparity and (right panel) as a function of distance, for 10 fixation distances (dashed lines). The response of a distance detector should only depend on egocentric distance.

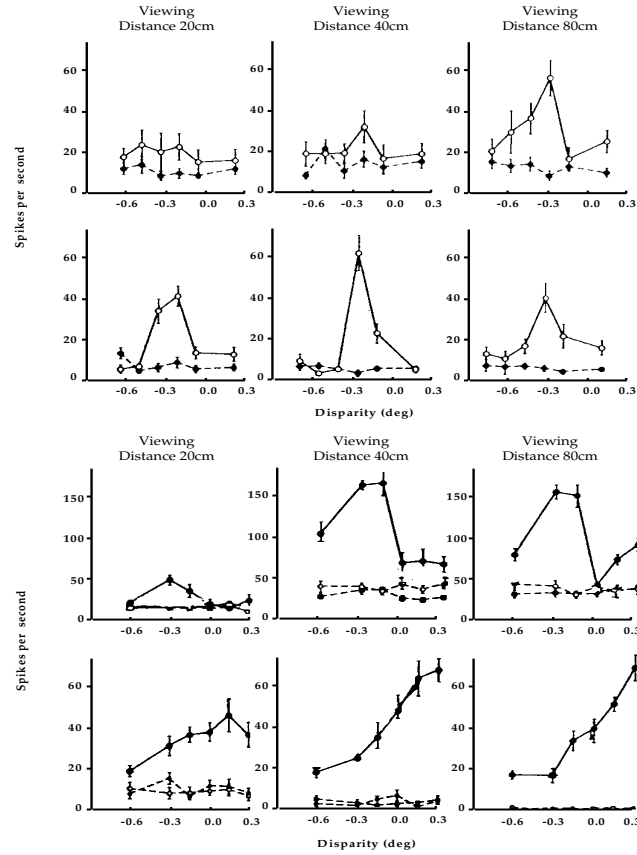


Figure III.4: Experimental recordings showing that the disparity selectivity of V1 neurons is modulated by distance of fixation. Each row of panels is from one neuron and each panel gives the tuning curve for disparity at a different viewing distance. In all cases, only the amplitude of the response varies with fixation distance while the general shape of the disparity tuning curve and the position of response peak remain the same. These neurons are neither pure disparity nor pure distance detectors (see Fig. 3) (Adapted from Trotter et al, 1992).

the representation of distance as well as other functions of disparity and vergence that are needed for guiding actions. The way humans perceive distance is consistent with this representation and corresponds to a low variance estimate of distance. Although gain-modulated neurons may indeed be involved in an intermediate representation of distance, our theoretical analysis shows that they may also be used to represent other functions of disparity and eye position.

III.A Representing egocentric distance with basis functions

III.A.1 Basis Functions

A basic problem in the theory of function approximation is to approximate any function in a class of functions by a linear combination of fixed basis functions. For example, any well behaved function can be approximated arbitrarily closely by a linear combination of sines and cosines summed over all possible frequencies and phases. The Fourier series is only one of many possible basis sets.

Sigmoid functions, commonly used in neural networks, and many radially symmetric functions can also be used as basis set and have various advantages (Hornik et al. 1989; Moody and Darken 1989; Poggio 1990). The familiar gaussian function is radially symmetric and a family of gaussians with all possible centers and variances forms a basis set. Sigmoid functions with all possible discriminant surfaces in the input space also form a basis set, but one that breaks up the input space into extended regions rather than the compact regions of support formed by gaussian basis functions.

There is a close connection between approximation theory and neural networks (Lapedes and Farber 1988; Poggio 1990). In a three-layer feedforward neural network, the hidden layer of processing units represents input patterns by a population code that serves as an intermediate transformation between the inputs and outputs. There are many possible choices of output functions for the hidden units

depending on the nature of the transformation the network needs to perform and the number of hidden units available. One way to determine an appropriate set of hidden units is to use an optimization procedure such as backpropagation (Rumelhart et al. 1986). An alternative is to handcraft a good set of hidden units, such as a set of basis functions. Choosing the hidden units in advance greatly simplifies optimization since the input weights are fixed and only the weights from the hidden to the output units need to be determined.

Whether the brain makes use of basis functions is not yet clear from physiological data, but the problem of spatial localization may be one area where it can be tested (Poggio 1990). We show in this section that the observed cortical representation of distance can be considered a new type of basis function that combines some of the best features of radial basis functions and sigmoids. Each basis function corresponds to the response of one gain-modulated neuron of the type reported by Trotter et al. and Gnadt et al. Figure 5 shows a neural network diagram of our model.

III.A.2 Methods

Our goal is to show that gain-modulated neurons can be used to approximate distance. We first specify a set of basis functions, $B_i(\delta, \nu)$, of disparity, δ , and vergence, ν , that are consistent with the observed responses of gain-modulated neurons in the visual cortex. The next step is to show that a linear combination of these basis functions can be used to estimate the distance, $D(\delta, \nu)$:

$$D(\delta, \nu) = \sum_{i=1}^n w_i B_i(\delta, \nu) \quad (\text{III.2})$$

where the w_i are the weights of the network depicted in figure 5.

Gain-modulated units.

We specified 41 tuning curves for disparity shown in figure 2 (only every other curve is shown). These profiles were idealizations of real disparity tuning curves

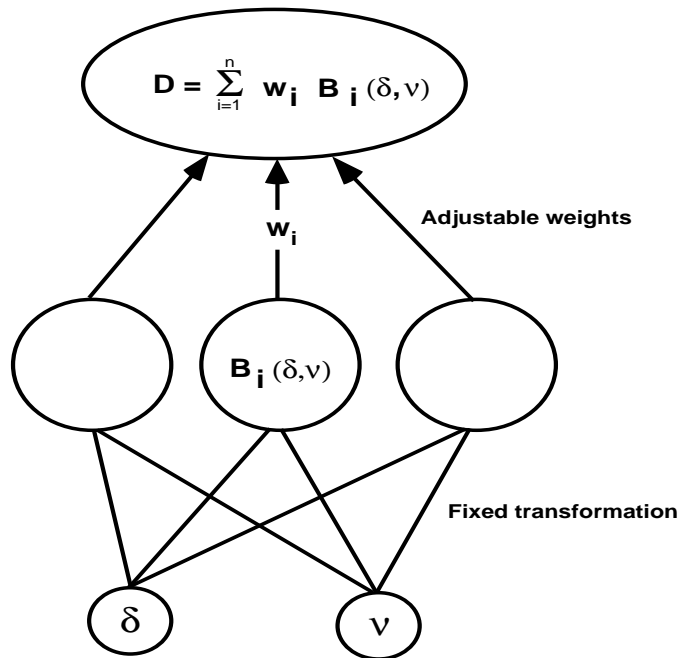


Figure III.5: Neural network representation of a basis function model for computing egocentric distance from vergence and disparity. Each hidden unit receives as inputs the vergence angle, ν , and the disparity, δ , and computes a function of these two variables. The basis functions $B_i(\delta, \nu)$ on the hidden layer represent the gain-modulated neurons found in V1. In a backpropagation network, the hidden unit is usually a sigmoid function of a weighted sum of the inputs. The hidden units project to the output layer, which consists of a single unit that represents distance by its continuous output, D .

from monkeys and cats and were similar to the ones previously used by Lehky and Sejnowski (1990):

Near units (units with preferred disparity, δ_i , less than -1°):

$$a_i(\delta) = 1.12 \exp -\frac{(\delta - \delta_i)^2}{\sigma_i^2} - 0.35 \exp -\frac{(\delta - (\delta_i + \sigma_i^2))^2}{\sigma_i^2} \quad (\text{III.3})$$

Far units (units with preferred disparity, δ_i , greater than 1°):

$$a_i(\delta) = 1.12 \exp -\frac{(\delta - \delta_i)^2}{\sigma_i^2} - 0.35 \exp -\frac{(\delta - (\delta_i - \sigma_i^2))^2}{\sigma_i^2} \quad (\text{III.4})$$

Excitatory tuned units (units with preferred disparity, δ_i , within $[-1^\circ, 1^\circ]$):

$$a_i(\delta) = 1.12 \exp -\frac{(\delta - \delta_i)^2}{\sigma_i^2} - 0.35 \exp -\frac{(\delta - \delta_i)^2}{2\sigma_i^2} \quad (\text{III.5})$$

where δ_i is the peak of response for a given curve, and σ_i^2 is related to the width of the tuning curve. The response peaks were evenly spread over the disparity range $[-4^\circ, 4^\circ]$ and the σ_i^2 were chosen to equal the absolute value of the disparity corresponding to the peak response, except for the curves whose peak were within the disparity range $[-10', 10']$, for which the variance was set to 10 minutes of arc. We did not include tuned inhibitory cells since they could be modeled by tuned excitatory cells with negative output weights.

Disparity tuning curves were not available for neurons in humans, so these values were chosen to reflect data from psychophysical studies and recordings from other species. In cats, LeVay and Voigt recorded neurons which had peaks of disparity tuning that were distributed within this interval (LeVay and Voigt 1988). Most recordings from monkeys have been from neurons in the foveal representation of primary visual cortex, $[-1^\circ, 1^\circ]$. Humans are above chance in perceiving disparity in the range $[-4^\circ, 4^\circ]$ (Richards 1971; Westheimer and Tanzman 1956), which suggests that the range of disparities over which cells respond in human must extend well beyond $[-1^\circ, 1^\circ]$. In any case, the results presented in this paper do not depend critically upon this range.

Vergence tuning curves were modeled as sigmoids and 10 different curves were chosen. The gain modulation of cortical neurons has not yet been fully characterized, but the sigmoid shape of the input-output function of most neurons is at least consistent with our choice. The expression used for the vergence selectivity, $z_i(\nu)$, was:

$$z_i(\nu) = \frac{1}{1 + \exp - \frac{\nu - \nu_i}{T_i}} \quad (\text{III.6})$$

where ν is the vergence angle, and ν_i and T_i are, respectively, the thresholds and the slopes, or *temperatures* of the sigmoids. The threshold controls the position of the sigmoid on the vergence axes, whereas the temperature controls the steepness the sigmoid. We used 10 combinations of 5 thresholds (7.78° , 11.26° , 14.76° , 18.24° , 21.73°) and 2 temperatures (3.45 and 1.15), which correspond to shallow modulation of the disparity responses with viewing distance. Trotter et al. were nonmonotonic and would be better fit by a gaussian (see the second cell in figure 4). Using gaussian modulation in addition to sigmoid modulation would not affect the results presented in this paper and for simplicity we only used sigmoids. We did not include monotonically decreasing sigmoids for they are equivalent to increasing ones multiplied by a negative weights.

Finally, by multiplying combinations of disparity and vergence selectivities, we obtained 10x41 gain-modulated responses; that is, a total of 410 basis functions:

$$B_i^G(\delta, \nu) = a_i(\delta)z_i(\nu) \quad (\text{III.7})$$

Figure 6, A to C shows three typical examples of these functions, plotting the responses as functions of vergence and disparity. These plots are idealizations of the neuronal responses shown in figure 4. Each basis function models the response of a single cell to disparity and vergence.

Distance-tuned units

For some of our simulations, we used a second set of functions tuned to distance rather than disparity in order to compare their properties with those of the

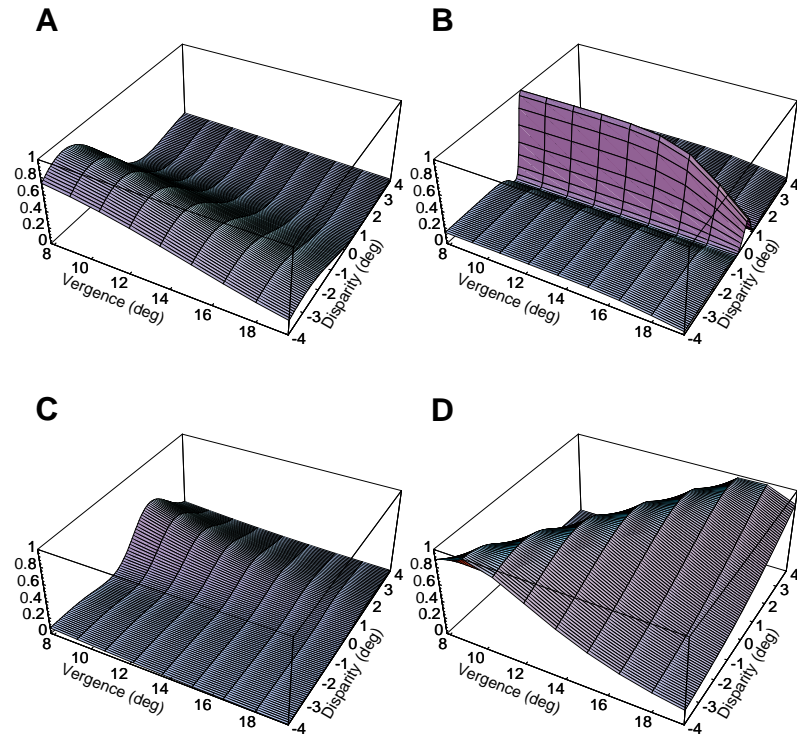


Figure III.6: Typical tuning curves for the hidden units as functions of vergence and disparity. A-C) Three examples of idealized gain-modulated units. The tuning with disparity is a difference of gaussians and the vergence tuning is sigmoidal (see Eqs. 6-7). D) A distance-tuned unit. Isodistance lines run along the diagonal in the vergence-disparity plane (see Eq. 8).

above basis functions. The responses of these units were a gaussian functions of distance similar to the ones shown in figure 3B:

$$B_i^D(\delta, \nu) = e^{-\frac{(D(\delta, \nu) - D_i)^2}{2\sigma^2}} \quad (\text{III.8})$$

where $D(\delta, \nu)$ is the function given in equation 1, and D_i is the preferred distance for basis function i . The σ were set to 0.07 in all $B_i^D(\delta, \nu)$.

A graph showing the response for one of these units is shown in figure 6D. There were 410 distance-tuned units and the peaks of their responses were evenly distributed from 0.16 to 1.08 meters.

Output representations geometrical versus perceived distance

Two different functions of distance were used for the output, one in which the output was proportional to geometric distance (Eq. 1), and a second function in which the output was proportional to perceived distance.

The need for this second function was motivated by experiments showing that the quantitative percept of depth is not a linear function of the stimulus disparity, as one might have expected (Richards 1971; Richards and Kaye 1974). The linear relationship between perceived and geometrical disparity holds only in the range of $[-40', 40']$, outside of which the estimated disparity tends to decrease as the stimulus disparity increases (Fig. 7A). Consequently, human cannot possibly perceive the actual distance of an object for all possible pairs of vergence and disparity. The perceived distance function can be obtained by replacing the disparity, d , with the perceived disparity, d_{perc} , in Eq. 1. The perceived disparity was obtained by parameterizing the psychometric curves (Richards 1971) (Fig. 7, A-B):

$$\begin{aligned} \delta' &= 1.8\delta e^{-|1.5\delta|^{0.9}} \\ \delta_{perc} &= \begin{cases} \delta & \text{if } \delta' > \delta \\ \delta' & \text{otherwise} \end{cases} \end{aligned} \quad (\text{III.9})$$

Figures 8A and 9A show respectively geometrical distance and perceived distance as a function of disparity and vergence.

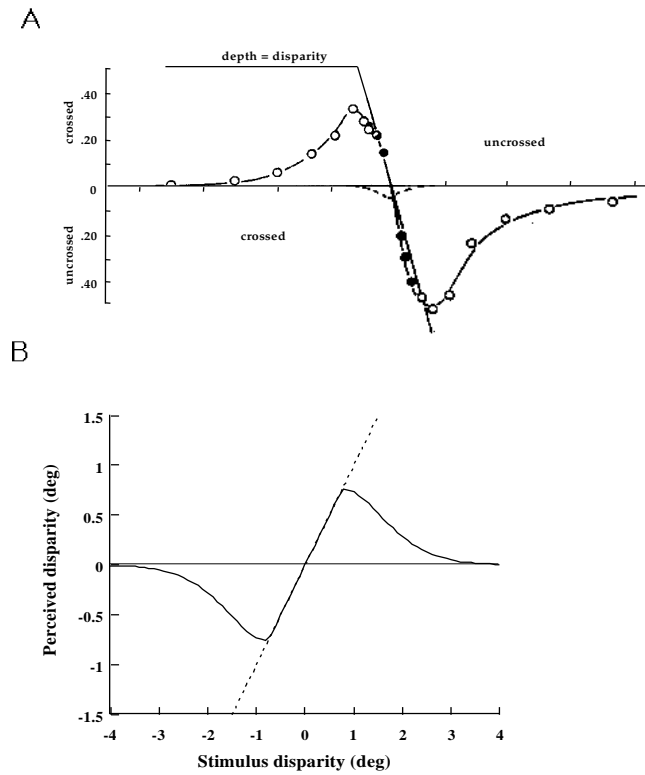


Figure III.7: Perceived disparity as a function of geometric disparity of the stimulus. A) Experimental data from (Richards 1971). Perception is accurate (straight line) only within $[-40, 40]$. B) Parameterized curve in Eq. 9. The dashed line shows the curve geometrical disparity. The curve we used is the mirror image of the one obtained by Richards because he used negative values for uncrossed disparities and positive values for crossed disparities.

The optimal set of weights were determined for all four combinations of the two input representations and the two output functions. The delta rule (Widrow and Hoff 1960), an iterative optimization technique, was used to find the optimal set of weights for the network shown in figure 5 (see appendix). Because the second layer of the network is effectively a linear transformation, the delta rule is not subject to local minima and the weights are guaranteed to converge to the optimal network (Widrow and Stearns 1985).

III.A.3 Results

Geometrical distance

We first determined the set of weights that provided the best approximation to geometrical distance as a function of vergence and disparity using gain-modulated units as a basis set. Figure 8, A and B, show the actual function and the approximated function. The approximation had only 0.63% error, demonstrating that gain-modulated neurons form a set of basis functions that can be used for accurately approximating geometric distance.

We then repeated the procedure with the set of functions made of distance-tuned units. As shown in figure 8C, the best approximation had an overall error of only 0.76%, so distance-tuned units were as good for estimating distance as the basis functions.

Perceived distance

Distance-tuned units have not been reported in the visual cortex even though they are adequate for representing distance and are conceptually simpler than a distributed representation. It is possible that such neurons do exist, but have not yet been found; alternatively, they may not be necessary or might even be a disadvantage. We provide a computational explanation for their absence.

It has been implicitly assumed that humans accurately perceive geometrical distance when only disparity and vergence are available as cues. In fact, humans tend to overestimate close viewing distances and underestimate far distances and the ratios

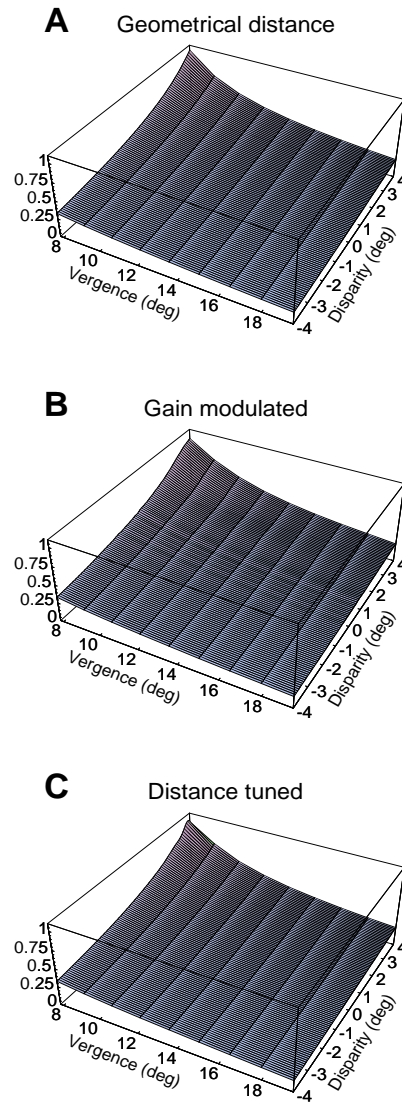


Figure III.8: Network approximations of geometric distance as a function of vergence and disparity. A) Geometric distance. B) Approximation by a set of gain-modulated units. C) Approximation by a set of distance-tuned units.

are a function of the experimental paradigm (Foley 1980; Foley and Held 1972). This would not affect our previous results since both sets of functions can approximate this new distance function by simply multiplying the distance function by a constant (the gain).

A more serious problem arises from nonlinearities in the psychometric curve of perceived disparity as a function of geometric disparity (Richards 1971) (Fig. 7A). The perception of disparity appears to be accurate within the range of $[-40', 40']$ but reverses and returns to zero as the disparity increases. As a consequence, a stimulus with 4 degrees of disparity is perceived at nearly the plane of fixation. As explained in the method section, we derived the function giving the perceived distance as a function of disparity and vergence by replacing disparity in Eq. 1 by perceived disparity (Fig. 9A). Using the same optimization technique as before, we found the best possible approximation of perceived distance for the disparity-tuned base units and the distance-tuned units.

Although the gain-modulated basis functions achieved 0.47% error (Fig. 9B), it was not possible to get a reasonable approximation of perceived distance when using the distance-tuned units (23.43% error) (Fig. 9C). This was because the distance-tuned units (Fig. 6D) were all oriented along the diagonals in the (ν, δ) plane, so they could only approximate functions whose gradients were perpendicular to this diagonal. In particular, they cannot be used to approximate perceived distance, which does not have this property. So even if distance-tuned units were present in the cortex, they could not be used for estimating perceived distance for some ranges of parameters.

If gain-modulated neurons could be used to approximate geometrical distance as well as perceived distance, and such neurons have been found in the cortex, then why does the brain not compute the actual geometrical distance? An answer to this question, given below, depends on noise. Cortical neurons fire action potentials with a Poisson distribution and the variance in the spike rate in response to a stimulus is proportional to the mean value of the response (Tolhurst et al. 1982; Vogels

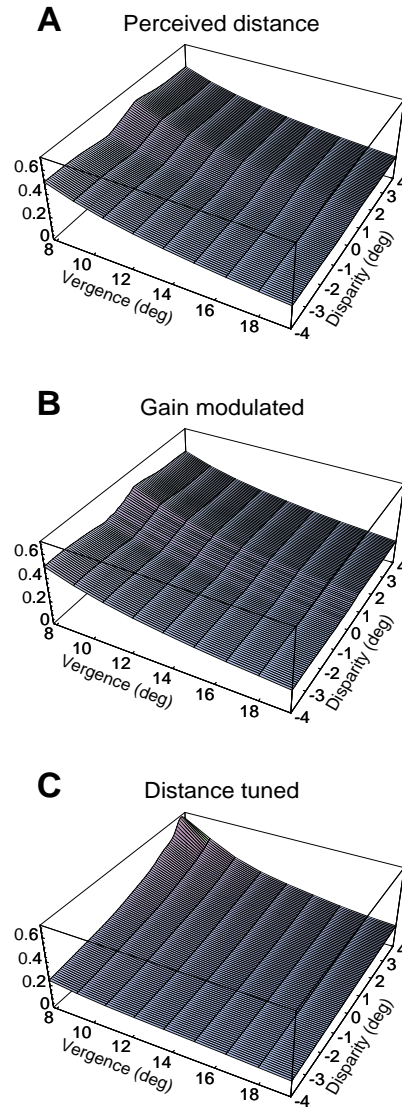


Figure III.9: Network approximations to perceived distance as a function of vergence and disparity. A) Perceived distance. B) Approximation by a network of gain-modulated units C) Approximation by a network of distance tuned units.

et al. 1989). If distance is estimated by pooling the responses of a neuronal population, then it is important to characterize the variance of the estimation; that is, the reliability of the estimated distance. We show in the next section that the variance depends on whether geometrical or perceived distance is being estimated from the response of the gain-modulated neurons.

III.B Reliability of Distance Estimation

III.B.1 Bias-variance trade-off

In estimating the value of a quantity by averaging measurements from noisy data, there is a trade-off between the bias of the estimate – the difference between the mean of the estimator and the true value of the quantity – and the variance of the estimate. Estimators with smaller bias have a higher variance and low variance estimators have greater bias. There is a compromise in choosing an estimator between minimizing the bias and reducing the variance (Scott 1992). Estimating distance from the responses of noisy neurons is subject to this bias-variance trade-off. The bias corresponds to the difference between the estimated and the true geometrical distance. Recovery of the true geometrical distance without bias leads to a high variance. We show here that the perceived distance, a highly biased estimate of geometrical distance especially for large positive and negative disparities, produces a smaller average variance. disparity is the only important variable for the bias-variance trade-off, the analysis described in this section is limited to disparity estimation. The reasons for which vergence is excluded from the analysis are discussed in the results section.

III.B.2 Methods

We assume that there are pools of noisy neurons whose disparity selectivities are of the type shown in figure 2. The goal is to use the information in this neural

population to estimate the disparity.

Noise model

The variance of an estimator depends on the type of noise present in the data. Single cell recordings from the cortex of cats and monkeys (Tolhurst et al. 1982; Vogels et al. 1989) suggest that the variance, σ^2 , of the firing rate of a typical neuron is,

$$\sigma^2 = K a^b \quad (\text{III.10})$$

where K and b are a constant and a is the mean activity (firing rate or total number of spikes) of a neuron for a given stimulus. Measurements of K vary from 1 to 3 and b is usually close to 1. We used $K = 2$ and $b=1$ for most of our analysis, which are the average values for these parameters reported in awake animals (Vogels et al. 1989). Our results were not sensitive to the exact values of these constants. It is essential for our analysis only that σ^2 be proportional to the activity.

For a pool of neurons, i , with the same disparity tuning and independent noise for each neuron, the variance of the average firing rate in the pool, σ_i^2 , is inversely proportional to the number of cells, N_i , in the pool:

$$\sigma_i^2(\delta) = \frac{\sigma_i^2(\delta)}{N_i} = \frac{K}{N_i} a_i(\delta) \quad (\text{III.11})$$

Note that σ_i^2 is a function of horizontal disparity, δ , because the activities of the neurons in the pool, a_i , are functions of disparity. We assume that the number of cells with a given selectivity for horizontal disparity is a gaussian function of the peak of the tuning curve, centered at zero disparity (LeVay and Voigt 1988):

$$N_i = N_0 e^{-\frac{\delta_i^2}{2\alpha^2}} \quad (\text{III.12})$$

Where N_i is the number of cells in the pool with preferred disparity δ_i , N_0 is the number of cells in the pool with a preferred disparity of zero, and α^2 is the variance of the distribution. The influence of N_0 and α^2 on the results are discussed in the next section.

It is important for our results that the disparity tuning curves shown in figure 2 correspond to pools containing a variable number of neurons. In particular, there were many fewer neurons tuned to large disparities than to small disparities.

Our noise model does not capture all aspects of neuronal noise. It describes only how spike rates vary around their mean value while ignoring variations in inter-spike intervals. These variations would have been relevant if we had considered neural codes involving the temporal distribution of the spike train. Evidence for such code in the cortex exists (McClurkin et al. 1991), but are still tentative. Consequently, we assumed that the idealized neurons in our model encode information through their mean firing rates.

Computing the estimator and the associated variance

Our goal was to find a way to combine the responses of disparity selective neurons to recover either true disparity or perceived disparity. This was accomplished by creating networks from weighted linear combination of the curves shown in figure 2 that give a mean output either to the solid line shown in figure 7B, corresponding to perceived disparity, or to the dotted line in figure 7B, corresponding to geometrical disparity. The estimated disparity is:

$$\Delta(\delta) = \sum_{i=1}^n w_i a_i(\delta) \quad (\text{III.13})$$

where w_i is the set of weights for a given estimator (geometrical or perceived), n the number of disparity pools (41 in all cases) and $a_i(\delta)$ is the responses of the disparity selective pool i to the disparity δ .

Instead of applying gradient descent, as before, to find the optimum set of weights in the network, a more efficient technique based on matrix inversion was used (see appendix) (Casdagli 1989); this was feasible because the network had only 41 weights, compared to 410 weights in the previous network. If all the pools of neurons were independent, the variance of the estimate is the weighted sum of variances:

$$\sigma_{\Delta(\delta)}^2 = \sum_{i=1}^n w_i^2 \sigma_i^2(\delta) = \sum_{i=1}^n w_i^2 \frac{K}{N_i} a_i(\delta) \quad (\text{III.14})$$

where $\sigma_{\Delta(\delta)}^2$ is the variance of the estimator for disparity δ .

Consequently, if a pool of neurons contributes strongly to the estimation of disparity, $\Delta(\delta)$, (that is, its weight w_i is high), it will also strongly influence the total variance, $\sigma_{\Delta(\delta)}^2$. Notice too that $\sigma_{\Delta(\delta)}^2$ is a function of δ since σ_i^2 are functions of δ .

III.B.3 Results

Variance of estimation for geometrical and perceived disparity

Figure 10 shows the standard deviation (square root of the variance) as a function of disparity for perceived (Fig. 10A) and geometrical (Fig. 10B) disparity for $N_0=1000$ and $\alpha^2 = 1.15$. With these values for the distribution in Eq. 5, there is only a single neuron in the pools tuned to $+4^\circ$ and -4° , compared to 1000 neurons for the pool centered at 0° . On average the standard deviation is about 20 times smaller ($20.97/0.92 = 22.74$) when perceived disparity is being computed as for the geometrical disparity. For perceived disparity, the large bias for large disparities is compensated by a low standard deviation, particularly in the central region where the bias is also quite small. For geometric disparity, the bias of the estimation is zero but the standard error is high for all disparities.

An explanation for why the variance increases as the bias goes down derives from the sizes of the pools and the weights in the network. For large disparities, for example 4° , only 5 to 6 neuronal pools will be significantly active, the ones with peaks of disparity selectivity between 2° and 4° . The weights on these pools must be quite large to compute the true disparity and they will therefore make a large contribution to the total variance. To reduce the total variance, the weights from the noisiest pools must be reduced, and these are from the pools with the largest preferred disparities. The consequence is, however, a poor estimate for large values of disparity.

The trade-off between bias and variance arises from noise that is not homogeneous over the representation of disparity. More cells are tuned to zero disparity

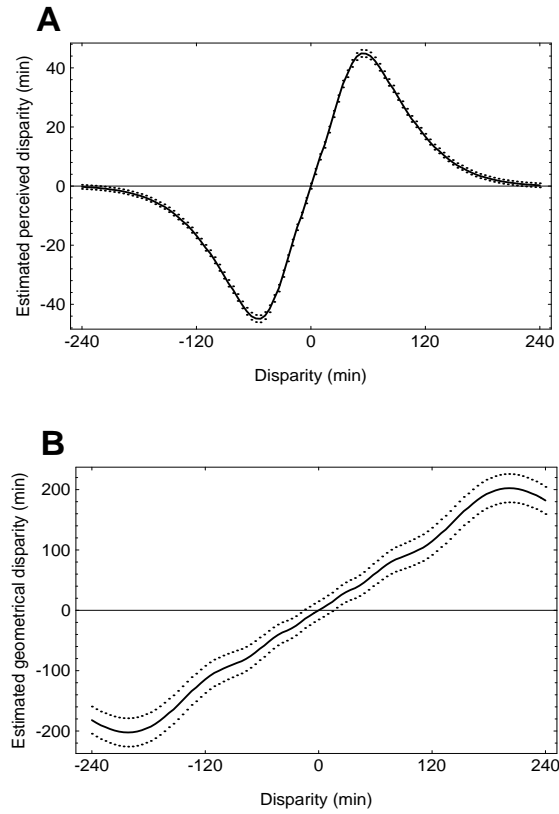


Figure III.10: Estimation of the perceived (A) and geometric (B) disparities from a distributed representation of disparity. The solid lines are the means of the estimations and the dotted lines are one standard deviation away. Note different scales on the two graphs: The standard deviation when estimating geometrical disparity is twenty times greater than the standard deviation for estimating perceived disparity.

than to large values of disparity. We do not have enough data to decide whether vergence is similarly inhomogeneous which is why we have restricted our analysis to disparity. If the noise is homogeneous over the whole vergence axes then this would only scale the variance but would not change our conclusions. If more neurons were selective for some range of vergence, then an analysis similar to that of disparity would apply. For example, if grasping distances were overrepresented in the cortex, then estimates of distance within this range could be optimized at the expense of accuracy for more distant locations.

This result provides a tentative answer to the question of why we perceive distance the way we do when it would be theoretically possible to recover geometric distance. Estimating geometric distance would lead to an extremely high variance over all disparities. Fortunately, many other cues can be used to assess the distance of objects when disparity values are large, so our visual system can afford to optimize distance estimation for small disparities.

Sensitivity to distribution of neurons per disparity pool

Changing the values of N_0 and α^2 affects the quantitative predictions but not the conclusions. Fitting a gaussian to the data in LeVay et al. (LeVay and Voigt 1988) for anesthetized cats leads to a values of $\alpha^2 = 2.0$. Recordings from behaving monkeys rarely extend beyond the disparity range $[-1^\circ, 1^\circ]$, but the value obtained for α^2 in the cat probably provides an upper bound. For $\alpha^2 = 2.0$, the ratio of the standard deviation of the geometric disparity over the standard deviation of the perceived disparity was around 15, and around 30 for $\alpha^2 = 0.6$. There is also uncertainty about the sizes of the disparity pools. For $N_0 = 100$, the ratio was 15. Over all feasible values of N_0 , $[50, 10000]$, and α^2 , $[0.5, 4.0]$, the ratios of the variances was at least 10.

III.C Discussion

This study of how distance is coded in populations of cortical neurons has three conclusions: First, even though none of the gain-modulated units in the model were explicitly tuned to distance, together they formed a distributed representation from which egocentric distance could be estimated. This distributed representation can be described as a set of basis functions with the tuning curve of each neuron providing a single basis function. Second, a representation based on distance-tuned units is not consistent with the way humans perceive distance. It is therefore not surprising that distance-tuned neurons have not been reported in the cortex. Finally, perceived disparity corresponds to a low variance, though biased, estimate of disparity based on the known disparity-selective neurons observed in the visual cortex. This could account for the fact that we do not perceive geometrical distance even though, in principle, we should be able to compute it.

III.C.1 Posterior parietal cortex

Lesions in posterior parietal cortex lead to spatial deficits that include problems with using depth cues (Godwin-Austen 1965; Holmes 1918; Holmes and Horrax 1919). Single-unit studies in LIP and area 7a of the posterior parietal cortex have revealed a distributed representation of egocentric space (see Andersen (1989) and Stein (1992) for reviews). These studies have focused on horizontal and vertical eye position and have shown that retinotopic receptive fields are modulated by eye position. Recent results suggest that the parietal cortex might also encode distance in a similar way. Gnadt has shown that many neurons in LIP respond to accommodation, disconjugate eye movements (Gnadt and Mays 1991) and disparity. Among those neurons tuned to disparity, some are modulated by the distance to fixation in a manner similar to that found in area V1 (Gnadt and Mays 1991). This finding is consistent with our model and strongly supports the hypothesis that the parietal cortex represents egocentric distance as part of a spatial representation. The dis-

tributed representation of depth in this study is similar to the way that position is represented in head-centered space along the vertical and horizontal axes (Zipser and Andersen 1988). Tuning for disparity is eye-centered and the responses are modulated by eye position, though not by conjugate eye movements as in previous models but by vergence eye movements. Neurons with response fields fixed in egocentric space, responding to a fixed distance or a fixed location in head-centered space independent of eye position, have not been reported in LIP or area 7a.

III.C.2 Early egocentric representation

Neurons in the foveal region of area V1 have small retinal receptive fields and are widely believed to encode the positions of objects in eye-centered coordinates. Our model, however, along with the data from Trotter et al., challenges this belief and suggests instead that neurons in V1 encode the egocentric distance of objects with a distributed representation. The limited size of the receptive fields of V1 neurons might seem a serious problem, but a recent hierarchical model demonstrates that a retino-spatiotopic representation of spatial location is feasible (Pouget et al. 1993). By retino-spatiotopic we simply mean that cells encode the spatial location of an object falling on a limited region of the retina, just as many cells code orientation or color.

Further evidence for retino-spatiotopic maps was found in area 17 in cats (Weyand and Malpeli 1989) and area V3A in monkeys (Galletti and Battaglini 1989), where the visual response amplitudes of neurons are modulated by the vertical and horizontal eye position. This suggests that the egocentric position of an object is already encoded in the early visual cortex in a way similar to that found in the parietal cortex (Andersen et al. 1985; Zipser and Andersen 1988). Such egocentric representations would be useful reference frames for planning actions. physiological and psychophysical experiments claiming the contrary (Pouget et al. 1993). Lesion experiments in area V1 cannot be used as a guide since such lesions typically result in blindness, a non-specific impairment which would certainly mask other more specific

disorders. For example, specific deficits for orientation discrimination have been found following cortical damage in higher visual areas (Goodale and Milner 1990) but not following lesions of V1, even though V1 represents orientation.

III.C.3 Do gain-modulated neurons form a basis set?

A set of N vectors is called a basis for a vector space V of dimension N if i) any arbitrary vector in V can be expressed as a linear combination of the vectors in the set and ii) the vectors in the set are linearly independent. The vector space used for function approximation has an infinite number of dimensions so an infinite number of basis functions is required to span this space. Obviously, our basis set does not meet this requirement and therefore cannot be called a basis in the strict sense.

The functions we used were formed from products of gaussian functions of disparity and sigmoid functions of vergence. A set of gaussians with all possible widths and peak positions is known to form a basis. This is also the case for a set of sigmoids including all possible gains (temperature parameters) and threshold positions. If two sets individually form bases, the set made by the product of all possible pairs of functions from these two sets is also a basis. Consequently, the set of functions used here, products of gaussians and sigmoids, is a subset from a larger set which does form a basis

In contrast, the set formed by all distance-tuned units does not form a basis. The response function of a distance-tuned unit is intrinsically a function of only one variable distance. Consequently this set cannot form a basis for all possible functions of vergence and disparity which depend on two variables. This is the reason why it was not possible to find a set of weights to approximate perceived distance using distance-tuned units.

III.C.4 Advantages of basis functions representation over an explicit representation

Basis functions greatly simplify learning because only one layer of weights need to be determined, those from the basis units to the output units (Moody and Darken 1989; Poggio 1990). However, a serious problem with using basis functions is that many are required to accurately estimate an arbitrary function. In our simulations we used 410 basis functions, which is a large but manageable number. This number could have been reduced using an appropriate optimization procedure, but in a nonsystematic attempt to do so we found that at least 100 basis functions were needed to get reasonably accurate estimates of perceived distance.

In the brain, each basis function can be identified with a single neuron or a group of related neurons. We have assumed that each basis function contributes independently to the approximation. However, recent measurements from pairs of nearby cortical neurons in the visual system have shown that the noise in their responses typically has a correlation of 0.15 (Gawne and Richmond 1993; Zohary et al. 1992). This correlation might seem low but in fact it severely limits the information that can be obtained by averaging over many neurons (Britten et al. 1992). A thousand correlated neurons carry little more information than around 50 neurons. A quantitative application of our model to the cortex awaits measurements of the correlations in the relevant neural populations carrying information about disparity and vergence.

We have shown that distance can be recovered from the responses of gain-modulated neurons: The same basis functions can be used to approximate other functions of vergence and disparity – perceived distance is only one function. Take, for example, the visually-guided control of an arm. The angles between the arm segments required to reach a particular object is a different function of the object disparity and viewing distance and could be approximated by the same set of gain-modulated neurons used for estimating distance. Gain-modulated neurons could be used for many different purposes through multiple projections to different areas.

Even though we have shown that gain-modulated neurons in a distributed

representation of disparity can represent egocentric distance, it would be misleading to call this a representation of egocentric distance since other functions could be equally well approximated by the same neural population. The actual functions explicitly represented are determined by the outputs of the neurons, or their projective fields, in addition to their receptive fields (Lehky and Sejnowski 1990). The same argument applies as well to motor cortex, which could encode many functions of reaching in addition to the direction of hand movement (Georgopoulos et al. 1986). In this view, the particular function attributed to the representation depends as much on the task selected by the experimenter as by the intrinsic parameters coded by the cortex (Poggio 1990; Sanger 1993).

III.C.5 Comparison to backpropagation networks

In a previous study, we trained a feedforward neural network to estimate distance by backpropagation of errors using the same input coding schemes for disparity and vergence studied here (Lehky et al. 1990). The weights between the inputs and the hidden layer of units were optimized along with the weights from the hidden to the output layer of units. Thus, the hidden units were not prespecified, as in our basis function network, but were optimized for the desired goal. With this approach it was possible to produce networks with nearly perfect performance using a minimal number of hidden units. In the case of geometrical distance from disparity and vergence, only 20 hidden units were required.

There are three disadvantages with this earlier model of distance estimation. First, each hidden unit took on a combination of properties whose function was difficult to decipher. This is a consequence of the compactness of the transformation produced by backpropagation and the complete connectivity between layers of units. Second, the backpropagation network took longer to optimize because two layers of weights were being simultaneously determined. Third, the representation in the hidden layer was highly specialized for the particular task the network was trained to perform and would not necessarily be useful for other purposes. If several functions

of the same input units were required, it would be necessary to create a separate network for each function.

There are several advantages to having compact, dedicated representations, but it is an open question whether the cortex uses them. We suspect that cortical learning mechanisms produce more versatile basis functions that span the dimensions of the space within which functions are synthesized for each cortical area. In summary, a representation using basis functions requires more hidden units than a backpropagation network for any single task, but is more versatile and easier to train.

III.C.6 Extraretinal vs retinal cues for viewing distance

Vertical disparity could also be used to recover viewing distance (Longuet-Higgins 1982; Mayhew and Longuet-Higgins 1982) and recent evidence from humans show conditions under which it can be used for absolute distance judgments and disparity scaling (Rogers and Bradshaw 1993). However, vertical disparity is not used by the visual system for objects smaller than 20° of visual angle (Cumming et al. 1991; Rogers and Bradshaw 1993), probably because the vertical disparity of smaller objects is too small for the visual system to detect. Subjects can, nonetheless, judge the distance or size of objects well below the 20 required for the vertical disparity system. In others experiments, manipulation of the vergence angle shows that it can influence the assessment of viewing distance (Cumming et al. 1991; Ritter 1977; von Hofsten 1976; von Hofsten 1979). Therefore, both vergence and vertical disparity can be used when they provide reliable information.

This raises the issue of whether experimental evidence for vergence modulation could be attributed to vertical disparity instead (Gnadt and Mays 1991; Trotter et al. 1992; Trotter et al. 1991). The receptive fields of the neurons recorded by Trotter et al., were within the central 5° of the visual field (Trotter, personal communication). Vertical disparity differences across small objects in this central region were, at most, $3'$. It would seem highly unlikely that the gain-modulation of the disparity-sensitive neurons observed by Trotter et al. could be fully accounted for by

such small changes in vertical disparity. This remains, however, an empirical question whose resolution would require independent manipulations of vertical disparity and vergence.

Our model could be expanded to include vertical disparity in the input representation. In addition to vergence and disparity, we would expand the basis set to represent arbitrary functions of three variables. If the tuning of neurons in the visual cortex to vertical disparity were found to be a gaussian function or a sigmoid function, we would then be able to approximate any function of these three variables. The problem with continuing to increase the number of input variables in this way is that the number of required basis functions increases exponentially with the number of dimensions.

III.C.7 Distance vs scaling

Vergence provides information for solving a problem that is closely related to distance estimation. When a rigid object moves away from a viewer, its size is usually perceived as constant even though its angular extent on the retina and the relative disparities between the parts of the object change with distance. The invariance of the perceived size of the object with viewing distance is called size constancy.

Several investigators have suggested that size constancy could be achieved by scaling the retinal size to compensate for viewing distance and have reported experimental results supporting this hypothesis (Cumming et al. 1991; Rogers and Bradshaw 1993) (W. Richards, personal communication). The extent of an object in depth, S , as a function of the disparity of its front and back edges, respectively δ_{front} and δ_{back} , and vergence angle, ν , is given by:

$$S = \frac{I}{2\tan(\frac{\nu-\delta_{back}}{2})} - \frac{I}{2\tan(\frac{\nu-\delta_{front}}{2})} \quad (\text{III.15})$$

where I is the interocular distance.

This equation is similar in form to Eq. 1, which gives egocentric distance as a function of vergence and disparity. Thus, the neural mechanisms underlying shape

and size constancy may be similar to the ones described in this paper. In particular, the modulation of disparity-selective neurons by vergence could be responsible for several perceptual constancies.

Acknowledgement

This chapter, in full, is a reprint of the material as it appears in *Cerebral Cortex*, 4(3):314-329, 1994. It was written in collaboration with T.J. Sejnowski. It is reprinted with permission from Oxford University Press.

Appendix

We describe here two methods that were used to compute the optimal weights given a set of input units and the function to be approximated.

Estimating distance from basis functions

Estimating geometrical or perceived distance with a linear combination of basis functions is equivalent to adjusting the hidden to output weights of a three-layer network whose hidden units correspond to the basis functions (Fig. 5). The delta rule provides a way of finding the optimum set of weights by iteratively presenting a set of examples of input-output pairs. The optimum weights are the ones that minimize the mean square error:

$$e = \frac{1}{2} \sum_{p=1}^m (y_p^* - \sum_{i=1}^n w_i B_i(\delta_p, \nu_p))^2 \quad (\text{III.16})$$

where m is the number of examples, y_p^* is the value of the target for an input vergence and disparity pair, (δ_p, ν_p) , w_i is the weight or coefficient associated with basis function i and $B_i(\delta_p, \nu_p)$ is the value of the basis function i for example p . The corresponding values of the basis function, $B_i(\delta_p, \nu_p)$, were described in the

first method section and the targets associated were either geometrical or perceived distance depending on the simulation.

The delta rule changes the weights by a small amount in direction opposite to the gradient of increasing error:

$$\Delta w_i = -\alpha \frac{\partial e}{\partial w_i} \quad (\text{III.17})$$

$$\Delta w_i = \alpha \sum_{p=1}^m (y_p^* - \sum_{i=1}^n w_i B_i(\delta_p, \nu_p)) B_i(\delta_p, \nu_p) \quad (\text{III.18})$$

where α is a parameter that controls the rate at which the weights change. The delta rule is guaranteed to converge to the minimum of the error after a sufficient number of presentations of the training set if α is sufficiently small (Widrow and Stearns 1985).

Estimating geometrical and perceived disparity

In the presence of noise, it is more convenient to treat the output of the network and the activity of the neuronal pools as random variables. Let Y^* be a random variable corresponding to the target disparities (either perceived or geometrical). Let Z be a vector of random variables whose components are the activities of the neuronal pools. Z is the sum of two vectors, $Z=A+N$, where A is the vector of the mean activities for a given disparity and N is the noise vector. Our goal is to find the linear estimation of Y^* from Z , $W_T Z$, which minimizes the variance of the error ($Y^* - W_T Z$):

$$W^* = \arg_W \min [E[(Y^* - W_T Z)(Y^* - W_T Z)^T]] \quad (\text{III.19})$$

The solution of this estimation problem is well-known (Anderson and Moore 1979) and has the form:

$$W^* = C_{ZZ}^{-1} C_{ZY^*} \quad (\text{III.20})$$

where C_{ZZ} and C_{ZY^*} are respectively the covariance matrices of Z with itself and Z and Y^* . Substituting $Z=A+N$, we have:

$$C_{ZZ} = E[(A + N)(A + N)^T] \quad (\text{III.21})$$

$$C_{ZZ} = E[AA^T + NN^T + AN^T + NA^T] \quad (\text{III.22})$$

$$C_{ZZ} = C_{AA} + C_{NN} \quad (\text{III.23})$$

$C_{AN} = C_{NA} = 0$ because the noise around the mean is independent of the mean activity. At first, this statement may sound counterintuitive since the variance of the noise was chosen to be proportional to the mean activity (see Equ. (11)). However, it can be easily shown that these two random variables are nevertheless independent. Similarly:

$$C_{ZY^*} = E[(A + N)Y^{*T}] = C_{AY^*} \quad (\text{III.24})$$

$C_{NY^*} = 0$ because N and Y^* are independent random variables. Substituting 22 and 23 in 19 leads to:

$$W^* = [C_{AA} + C_{NN}]^{-1} C_{AY^*} \quad (\text{III.25})$$

The entries of the covariance matrices can be obtained from, for C_{AA} ,

$$c_{ij} = \frac{1}{m} \sum_{p=1}^m a_i(\delta_p) a_j(\delta_p) \quad (\text{III.26})$$

for C_{AY^*}

$$c_i = \frac{1}{m} \sum_{p=1}^m a_i(\delta_p) y(\delta_p) \quad (\text{III.27})$$

where $a_i(\delta_p)$ is the activity of pool i for input p and $y(\delta_p)$ is the disparity, either perceived or geometrical, for input p .

The covariance matrix of the noise, C_{NN} , can be directly computed from the probability distribution of the noise. Nondiagonal terms are all zero since we assumed independence between neuronal pools. The diagonal terms of are obtained from:

$$\sigma_i^2 = \int_{-\infty}^{\infty} n_i^2 p(n_i) dn_i \quad (\text{III.28})$$

Although we do not know the noise distribution, $p(n_i)$, directly, we have assumed that, given a mean level of activity of the pool, the noise has a normal distribution with a variance proportional to the mean (equation 11). Consequently, we need to introduce conditional distribution in 27. In general, for any random variables, X and Y, we can write:

$$p(x) = \int_{-\infty}^{\infty} p_Y(x)p(y)dy \quad (\text{III.29})$$

or equivalently when Y is a discrete random variable:

$$p(x) = \sum_{p=1}^m p_Y(x)p(y_p) \quad (\text{III.30})$$

where $p(x)$ and $p(y)$ are the probability density functions of X and Y and $p_Y(x)$ is the probability density function of X conditioned on Y.

Applying 29 to 27 leads to:

$$\sigma_i^2 = \sum_{p=1}^m \left(\int_{-\infty}^{\infty} n_i^2 p_{a_i(\delta_p)}(n_i) dn_i \right) \frac{1}{m} \quad (\text{III.31})$$

where $p_{a_i(\delta_p)}(n_i)$ is the probability density of the noise conditional on $a_i(\delta_p)$, the mean activity of pool i for disparity δ_p .

The term inside the parenthesis in equation 30 is the conditional variance of the noise which we know is proportional to the mean activity (Equ. 11):

$$\sigma_i^2(\delta_p) = \int_{-\infty}^{\infty} n_i^2 p_{a_i(\delta_p)}(n_i) dn_i = \frac{K}{N_i} a_i(\delta_p) \quad (\text{III.32})$$

therefore

$$\sigma_i^2 = \frac{K}{N_i} \frac{1}{m} \sum_{p=1}^m a_i(\delta_p) = \frac{K}{N_i} \bar{a} \quad (\text{III.33})$$

Equation 24 provides a one-shot method for computing the optimal set of weights and can be easily implemented in standard symbolic mathematics programs such as Mathematica, the one we used here. However, this method is impractical when the matrices have more than 100 dimensions. Matrix inversion was not used, for example, in the optimization problem posed in the first section of this appendix because the size of the matrix C_{AA} was 410x410.

References

- Andersen, R. A., Essick, G. K., and Siegel, R. M. Encoding of spatial location by posterior parietal neurons. *Science*, 230: 456-458, 1985.
- Anderson, B. D. O., and Moore, J. B. *Optimal Filtering*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1979.
- Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. The analysis of visual motion: A comparison of neuronal and psychophysical performance. *J. Neurosci.* 12: 4745-4765, 1992.
- Casdagli, M. Nonlinear prediction of time series. *Physica*, 35 D: 335-356, 1989.
- Cormack, L. K., Stevenson, S. B., and Schor, C. M. Disparity-tuned channels of the human visual system. *Visual Neurosci.*, 10: 585-596, 1993.
- Cumming, B. G., Johnston, E. B., and Parker, A. J. Vertical disparity and perception of three dimensional shape. *Nature*, 349: 411-413, 1991.
- Ferster, D. A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex. *J. Physiol.*, 311: 623-655, 1981.

- Foley, J. M. Binocular distance perception. *Psychological Rev.*, 87: 411-434, 1980.
- Foley, J. M., and Held, R. Visually directed pointing as a function of target distance, direction and available cues. *Percep. Psychophys.*, 12: 263-268, 1972.
- Galletti, C., and Battaglini, P. P. Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. *J. Neurosci.*, 9: 1112-1125, 1989.
- Gawne, T. J., and Richmond, B. J. How independent are the messages carried by adjacent inferior temporal cortical neurons? *J Neurosci.*, 13: 2758-2771, 1993.
- Georgopoulos, A., Schwartz, A., and Kettner, R. Neuronal population coding of movement direction. *Science*, 233: 1416-1419, 1986.
- Gnadt, J. W., and Mays, L. E. Depth tuning in area LIP by disparity and accommodative cues. *Abst. Soc. Neurosci.*, 17: 1113, 1991.
- Godwin-Austen, R. B. A case of visual disorientation. *J. Neurol. Neurosurg. Psychia.*, 28: 453-48, 1965.
- Goodale, M. A., and Milner, A. D. Separate visual pathways for perception and action. *Trends Neurosci.*, 15: 20-25, 1990.
- Holmes, G. Disturbances of visual orientation. *British J. Ophtalmol.*, 2: 449-68, 1918.
- Holmes, G., and Horrax, G. Disturbances of spatial orientation and visual attention. *Arch. Neurol. Psychia.*, 1: 385-407, 1919.
- Hornik, K., Stinchcombe, M., and White, H. Multilayer feedforward network are universal approximators. *Neural Networks*, 2: 359-366, 1989.
- Hubel, D., and Wiesel, T. Cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Nature*, 225: 41-42, 1970.

Lapedes, A., and Farber, R. How neural nets work. In: *Neural Information Processing System D. N. Anderson* (Ed.), New York: American Institute of Physics, 1988. p. 442-456.

Lehky, S. R., Pouget, A., and Sejnowski, T. J. Neural models of binocular depth perception. *Cold Spring Harbor Symposium on Quantitative Biology: The Brain*, 55: 765-777, 1990.

Lehky, S. R., and Sejnowski, T. J. Neural network model of visual cortex for determining surface curvature from images of shaded surfaces. *Proc. R. Soc. Lond. B*, 240: 251-78, 1990.

LeVay, S., and Voigt, T. Ocular dominance and disparity coding in cat visual cortex. *Visual Neurosci.*, 1: 395-414, 1988.

Longuet-Higgins, H. C. The role of vertical dimension in stereoscopic vision. 11: 377-386, 1982.

Maunsell, J. H. R., and van Essen, D. C. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J. Neurophysiol.*, 49: 1148-1167, 1983.

Mayhew, J. E. W., and Longuet-Higgins, H. C. A computational model of binocular depth perception. *Nature*, 297: 377-78, 1982.

McClurkin, J. W., Optican, L. M., Richmond, B. J., and Gawne, T. J. Concurrent processing and complexity of temporally encoded neuronal messages in visual perception. *Science*, 253: 675-677, 1991.

Moody, J., and Darken, C. J. Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1: 281-294, 1989.

Nikara, T., Bishop, P. O., and Pettigrew, J. D. Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Exp. Brain Res.*, 6: 353-372, 1968.

Pettigrew, J. D., Nikara, T., and Bishop, P. O. Binocular interaction on single units in cat striate cortex : simultaneous stimulation by single moving slit with receptive fields in correspondence. *Exp. Brain Res.*, 6: 391-410, 1968.

Poggio, G. Processing of stereoscopic information in primate visual cortex. In: *Dynamic aspects of neocortical function* G. M. Edelman, W. E. Gall, and W. M. Cowan (Eds.), New York, NY: John Wiley & sons, 1984. p.

Poggio, G. F., and Fischer, B. Binocular interactions and depth sensitivity in striate and prestriate cortex of behaving rhesus monkeys. *J. Neurophysiol.*, 40: 1392-1405, 1977.

Poggio, G. F., Gonzalez, F., and Krause, F. Stereoscopic mechanism in monkey visual cortex : binocular correlation and disparity selectivity. *J. Neurosci.*, 8: 4531-4550, 1988.

Poggio, G. F., Motter, B. C., Squatrito, S., and Trotter, Y. Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Res.*, 25: 397-406, 1985.

Poggio, T. A theory of how the brain might work. *Cold Spring Harbor Symposium on Quantitative Biology: The Brain*, 55: 899-910, 1990.

Pouget, A., Fisher, S. A., and Sejnowski, T. J. Egocentric spatial representation in early vision. *J. Cog. Neurosci.*, 5(2): 150-161, 1993.

Richards, W. Anamalous stereoscopic perception of depth. *J. Opt. Soc. Am.*, 61: 410-414, 1971.

- Richards, W., and Kaye, M. G. Local versus global stereopsis. *Vision Res.*, 14: 1345-1347, 1974.
- Ritter, M. Effect of disparity and viewing distance on perceived depth. *Percept. Psychophys.*, 22: 400-407, 1977.
- Rogers, B. J., and Bradshaw, M. F. Vertical disparities, differential perspective and binocular stereopsis. *Nature*, 361: 253-255, 1993.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning internal representations by error propagation. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations* Cambridge, MA: MIT Press, 1986. p.318-362
- Sanger, T. Theoretical consideration for the analysis of population coding in motor cortex. *Neural Computation*, 6:29-37. 1994.
- Scott, D. W. *Multivariate Density Estimation*. Wiley Interscience, 1992.
- Stevenson, S. B., Cormack, L. K., Schor, C. M., and Tyler, C. M. Disparity-tuned mechanisms of human stereopsis. *Vision Res.*, 32: 1685-1689, 1992.
- Tolhurst, D. J., Movshon, J. A., and Dean, A. D. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.*, 23: 775-785, 1982.
- Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S. J., and Imbert, M. Modulation of neural stereoscopic processing in primate area V1 by the viewing distance. *Science*, 257: 1279-1281, 1992.
- Trotter, Y., Celebrini, S., Thorpe, S. J., and Imbert, M. Modulation of stereoscopic processing in primate visual cortex V1 by the distance fixation. *Abst. Soc. Neurosci.*, 17: 1016, 1991.

Vogels, R., Spilleers, W., and Orban, G. A. The response variability of striate cortical neurons in the behaving monkey. *Exp. Brain Res.*, 77: 432-436, 1989.

von Hofsten, C. The role of convergence in space perception. *Vision Res.*, 16: 193-198, 1976.

von Hofsten, C. Binocular convergence as a determinant of reaching behavior in infancy. *Perception*, 6: 139-144, 1977.

von Hofsten, C. Recalibration of the convergence system. *Perception*, 8: 37-42, 1979.

Westheimer, G., and Tanzman, I. Qualitative depth localization with diplopic images. *J. Opt. Soc. Am.*, 46: 116-117, 1956.

Weyand, T. G., and Malpeli, J. G. Responses of neurons in primary visual cortex are influenced by eye position. *Abst. Soc. Neurosci.*, 15: 1989.

Widrow, B., and Stearns, S. D. Adaptive signal filtering. Englewoods Cliffs, N.J.: Prentice Hall, Inc., 1985.

Widrow, G., and Hoff, M. E. Adaptive switching circuits. *IRE WESCON Convention Record*, 4: 96-104, 1960.

Zipser, D., and Andersen, R. Back propagation learning simulates response properties of a subset of posterior parietal neurons. *Nature*, 331: 679-684, 1988.

Zohary, E., Shadlen, M. N., and Newsome, W. T. Correlated activity of neuron in area MT. *Abst. Soc. Neurosci.*, 18: 1101, 1992.

Chapter IV

Representing Spatial Transformations in the Parietal Cortex with Basis Functions

Summary and Conclusions

1. Sensorimotor transformations can be characterized as nonlinear functions of sensory inputs. For example, reaching for an object in space can be represented as transformations from visual, auditory, somatosensory, vestibular and proprioceptive inputs to the appropriate motor outputs. We explore here the possibility that the parietal cortex uses basis functions to represent these transformations. Basis function decomposition is a general method for approximating nonlinear functions that is computationally efficient and well-suited for adaptive modification.
2. The response properties of neurons in the parietal cortex are consistent with the basis function hypothesis. Most parietal neurons can be modeled as a gaussian function of retinal location multiplied by a sigmoid function of eye position. A large set of such functions can be shown to form a basis set.

3. Simulations demonstrate that basis functions can be used to generate head-centered and retinotopic receptive fields by direct projection from parietal neurons. These two types of receptive fields are believed to be involved in generating motor commands for eye and arm movements.
4. The basis function hypothesis is compared with other approaches that are commonly used to model population codes, such as computational maps and vectorial representations. Neither of these alternatives can fully account for the responses of parietal neurons and both have severe computational limitations.
5. Basis functions have the advantage over other approaches of not depending on any coordinate system or reference frame. As a consequence, the position of an object can be represented in multiple reference frames simultaneously. This is consistent with recent studies of humans with lesions of parietal cortex who exhibit hemineglect.

IV.A Introduction

Located at the crossroads of four sensory systems, visual, auditory, vestibular and somatosensory, the posterior parietal cortex projects to several frontal and premotor areas (Felleman and Van Essen, 1991; Andersen et al., 1990a; Blatt, Andersen and Stoner, 1990). In humans, lesions of the parietal cortex often result in hemineglect, a syndrome characterized by reduced exploration of the hemispace contralateral to the site of the lesion. Patients with hemineglect have difficulties generating eye or arm movements toward visual, auditory or somatosensory stimuli (Heilman, Watson and Valenstein, 1985). This deficit is particularly clear in line cancellation experiments, in which the subject is asked to cross out short line segments uniformly spread over a page. Although this task is easy for normal subjects, parietal patients typically fail to cross the lines on the side of the page contralateral to the lesion.

The deficits in spatial perception following parietal lesions suggest that the parietal cortex supports an egocentric representation, in which the positions of objects in space, with respect to the viewer, are represented in a way that is independent of where the eyes are pointed. Such a representation could compensate for saccadic eyes movements which occur several times per second on average, resulting in large shifts in the retinal location of objects on the retina (Feldman, 1985). It could also be used to control arm movements toward an object since egocentric coordinates are intermediate between visual and motor coordinates.

The simplest form of egocentric representation is one in which the responses of neurons are independent of eye position and the orientation of the body; that is, the neurons code for absolute spatial location. This would entail that neurons have a receptive field whose position on the retina vary with the eye position, but such neurons have not been observed in the parietal cortex. In contrast, the visual receptive fields of parietal neurons have a fixed position on the retina and the amplitude of the response to visual stimuli is modulated on eye position (Andersen, Essick and Siegel, 1985). Figure IV.1A show how the gain of the retinotopic receptive field of a cell

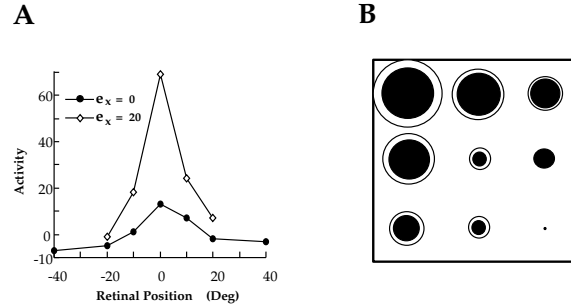


Figure IV.1: A) Typical visual receptive field of a parietal neuron shown for two different gaze angles, e_x . The retinal positions of the receptive fields do not vary with eye position, only the gain of the response changes. B) Typical gain field of a parietal neuron. The circles show the responses of a single cell to visual stimulation in the center of the receptive field for eye fixations at nine different locations. The diameter of each outer circle is proportional to overall activity, while the inner circle corresponds to visually evoked activity (overall activity minus spontaneous activity). Eye fixation positions were sampled on a grid with a 20° spacing, such that the fixation point was straight ahead for the central circle, and 20° up and 20° left for the upper left circle. The activity of this cell increased monotonically for eye positions located upward and to the left. This preferred direction is specific to each cell (adapted from Andersen et al., 1985, 1988).

changes with eye position. In figure IV.1B, the circles indicate the response of a single cell to a stimulus flashed in the middle of the receptive field while the monkey fixated nine different locations. The peak response appears to vary linearly with eye position along a particular direction, left and upward for the cell shown in figure IV.1. This gain field is the receptive field of a single cell for eye position.

These results are inconsistent with the hypothesis of a head-centered map in which single neurons respond to spatial location in a way that is independent of eye position. Psychophysical experiments also contradict this view (O'Regan and Levy-Schoen, 1983; Irwin, 1983; Irwin, Brown and Sun, 1988; Irwin, Zacks and Brown, 1990). An alternative hypothesis, presented here, is that neurons in the parietal

cortex do not represent the positions of objects *per se*, but rather represent inputs from the retina and the eye position in an especially efficient way called basis functions. Any nonlinear transformation can be synthesized from the population of neurons representing basis functions by a weighted linear summation. There are many types of basis functions and we show that a particular type appears to account quite well for the responses of single cells in the parietal cortex.

IV.B Basis Function Representation

The parietal cortex contains a large population of cells with a wide range of retinal receptive field locations and gain fields. Despite the distributed nature of this representation, modeling studies have shown that there is nonetheless sufficient information to recover the position of the object by a direct projection (Zipser and Andersen, 1988a; Pouget, Fisher and Sejnowski, 1993). However, the same representation might be useful for other tasks that involve transformation of the sensory information to motor commands suitable for controlling arm movements for reaching and other motor tasks.

In the general case, recovering the egocentric position of an object from its multiple sensory images is a difficult problem. In this paper, only visual input is considered and all body parts are fixed except the eyes, a condition that has been extensively used for neurophysiological studies in monkeys. The position of an object in the visual field can be specified in retinotopic coordinates, with respect to the point of fixation, or in head-centered coordinates. If we ignore distance and restrict the spatial location to two-dimensions, then the position of an object can be represented by a two-dimensional vector with horizontal and vertical components. Let \vec{R} and \vec{A} be the vectors for, respectively, the retinotopic and head-centered position of an object, and \vec{E} the vector for eye position.

As illustrated in figure IV.2, these three vectors are related by a simple linear relationship:

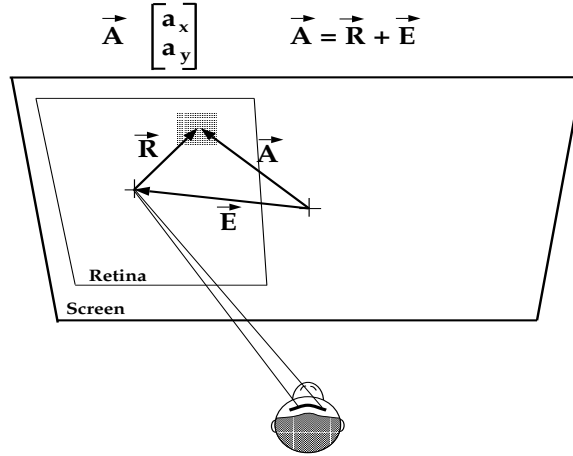


Figure IV.2: The position of an object in head-centered coordinates, \vec{A} , can be computed by adding its retinal location, \vec{R} , with the current eye position, \vec{E} .

$$\vec{A} = \vec{R} + \vec{E} \quad (\text{IV.1})$$

With the head and body fixed, the egocentric location of an object is identical to its head-centered location.

It is often assumed that in order to orient toward an object and interact with it, a representation of the egocentric position of the object is needed in some coordinate system. It is, however, possible to form a representation of the position of the object relative to the body without a fixed coordinate system. Basis functions provide an alternative approach in which motor commands are treated as functions of the sensory inputs.

IV.B.1 Sensorimotor Coordination

The function which specified the pattern of muscle activity required to move a limb, or the body, to a specific spatial location is a highly nonlinear function of the sensory inputs. Although the cortex is not believed to specify patterns of muscle activation, it often uses nonlinear representations in the intermediate stages, even if the underlying transformations are actually linear.

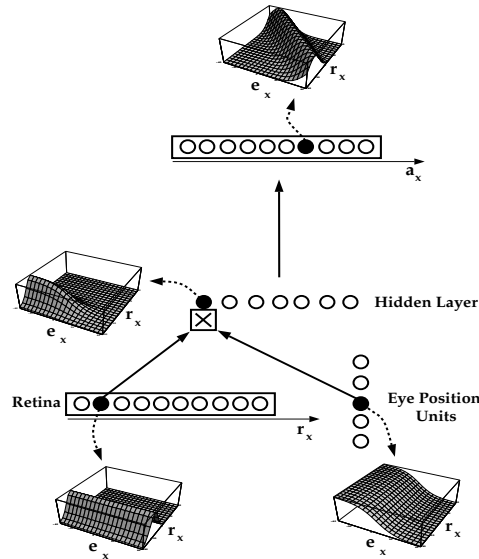


Figure IV.3: Neural network for transforming a retinotopic map to a head-centered map. The input contains a retinotopic map of the visual input and the output represents a head-centered map. The eye position units have a sigmoidal tuning to eye position and a range of thresholds. The function represented by the network is non-linear, as illustrated by the fact that the response to r_x and e_x of the units in the output layer is clearly not a plane. This mapping could be implemented by hidden units that compute the product of a gaussian of r_x multiplied by the sigmoid of e_x . Such units would provide basis functions of the input variables and would respond like gain-modulated neurons found in the parietal cortex.

Consider a concrete example to illustrate this point. Recent results have shown that cells in the premotor cortex and the putamen have head-centered visual receptive fields, suggesting that a head-centered map might exist beyond the parietal cortex (Graziano and Gross, 1993). What type of computation is involved in generating this receptive field from the retinotopic visual fields found in the early stages of the visual system? It might appear that the transformation is linear since, according to equation IV.1, the head-centered location of an object can be computed with a vector addition. However, visual receptive fields, whether in the visual system or premotor cortex, are typically limited in size and are approximately bell-shaped, which are nonlinear functions of retinal position. Vector addition therefore becomes nonlinear when mapping from retinotopic to head-centered maps.

This point is illustrated in figure IV.3 which shows a three-layer network model for performing this mapping in one dimension.

The input layer contains a one-dimensional retinotopic map similar to that found in the early stages of the visual system, where neurons respond to visual stimuli in a limited region of the visual field. In addition, several input units encode the horizontal position of the eye, e_x . In the output layer, units are organized in a one-dimensional head-centered map. They respond as a gaussian function of a_x , or, equivalently, a gaussian function of $r_x + e_x$.

In this one-dimensional case, the activity of all the units in the network can be plotted with respect to the input variables, namely, the retinal position of targets in space, r_x , and eye position, e_x . Examination of the plot for a typical output unit reveals that a head-centered receptive field is a nonlinear function of the input variables; that is, none of the plots are planar.

Other sensory remapping problems are formally identical to this case. The generation of eye movements toward auditory and somatosensory targets requires the same type of transformation (Jay and Sparks, 1987; Groh and Sparks, 1993). A neural network model for how signals in the primary auditory cortex can be converted into motor commands for saccadic eye movements in the superior colliculus would look

identical to the one shown in figure IV.3, with the label a_x and r_x interchanged (Pouget et al., 1993; Groh and Sparks, 1993).

In addition to sensory remapping with gaussian-shaped receptive fields, the brain is faced with various nonlinear estimation problems. Specifying the direction of an arm movement from the retinal position of a target and proprioceptive input requires a bi-linear product, which, despite its name, is actually nonlinear (Burnod et al., 1992). Estimating the egocentric distance of objects, an essential information for reaching, from horizontal disparity and vergence angle is also a nonlinear transformation (Pouget and Sejnowski, 1994).

Finally, nonlinear representations are also essential for sensory integration. For instance, in order to localize an object, one may want to integrate location estimates obtained by analyzing its image and sound. If these estimates are added together, the mean of the final estimate will be twice as large when the object is seen and heard, versus seen only. One solution to this problem is to use a nonlinear integration mechanism, such as a bayesian estimator.

These examples illustrate the nonlinearities inherent to typical sensorimotor transformations. If the parietal cortex computes these transformations, the spatial representations must reflect their computational constraints.

IV.B.2 Gain Fields and Basis Functions

A nonlinear function, such as e^x , is typically represented in a computer by a Taylor series, a polynomial expansion that is simpler to compute. This is not the only way to approximate a nonlinear function. An alternative method is to express the function as a linear combination of sines and cosines weighted by numbers called Fourier coefficients. Because sines and cosines can be used to approximate a very large ensemble of nonlinear functions, they are called basis functions. There are many other types of basis functions.

Two classes of basis functions, Gaussians and sigmoids, are especially promising candidates for fitting physiological data. Gaussians or, more generally, radially

symmetric basis functions (often called RBFs) have been extensively used for nonlinear function approximation (Casdagli, 1989; Moody and Darken, 1989; Poggio and Girosi, 1990). The set of all possible sigmoid functions, used in backpropagation models, also forms a basis set (Hornik, Stinchcombe and White, 1989; Baldi, 1991).

Radial basis functions of sensory inputs can also be used for object recognition (Poggio, 1990; Poggio and Edelman, 1990). Neurophysiological experiments by Logothetis et al. have provided some evidence that the inferotemporal cortex might be using this kind of strategy (Logothetis et al., 1993). We have also used this approach to model the perception of egocentric distance and have shown that the response of some neurons in V1 and area LIP (Gnadt and Mays, 1991; Trotter et al., 1992) approximate basis functions of vergence and disparity (Pouget and Sejnowski, 1994).

A visuomotor adaptation study by Ghahramani et al (1995) suggests that similar basis functions might be used also for sensorimotor coordination. Their experiments involved two phases. They first trained human subjects to point accurately to a visual target despite distortions to the visual location of the finger. They found that when subjects are trained to adapt at one particular location in space, generalization is spatially limited around the training point, a pattern consistent with basis functions such as sigmoids or gaussian.

We believe that this basis function framework can also be used to interpret the response of gain modulated neurons in the parietal cortex. The response of a single cell like the one illustrated in figure IV.1, can be modeled by a product of a gaussian function of retinal location with a sigmoid function of eye position. Figure IV.4 shows the correspondence between the gain field and these idealized response functions. Both gaussians and sigmoids are basis functions and it can be shown that the product of two basis functions form a basis function, but all combinations must be represented.

If the neurons in the parietal cortex are considered as a hidden layer of units in a three-layer network (figure IV.3, for instance, shows a simulated parietal neuron that computes the product of a gaussian with a sigmoid), the parietal units encode

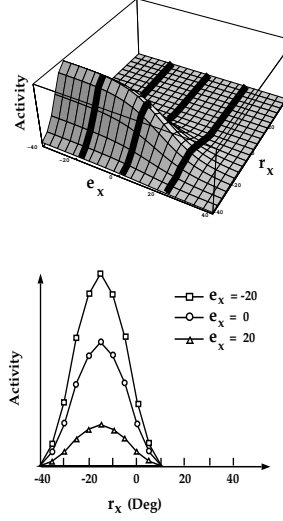


Figure IV.4: Response function obtained by multiplying a gaussian of retinal location with a sigmoid of eye position (top). When sampled at three different gaze angles (thick lines on top graph), the visual receptive field (bottom) shows the same gain modulation as found in the parietal cortex (figure IV.1-A).

the input variables \vec{R} and \vec{E} so that any function of these variables can be recovered by a linear combination of the hidden units activities (figure IV.5). As our simulations show, this is a more general and flexible way to encode the location of an object, since this representation contains simultaneously the head-centered, \vec{A} , and retinotopic, \vec{R} , location of the object.

Few neurons in the parietal cortex have response functions that can be perfectly fitted by a product of a gaussian and a sigmoid. These idealized responses are not strictly required, however, and other types of responses would also work as long as they meet certain general requirements.

IV.B.3 Predictions of the Basis Function Representation

1. The visual receptive field as well as the gain fields should be smooth nonlinear functions of \vec{R} and \vec{E} .
2. The selectivity to \vec{R} and \vec{E} should interact nonlinearly

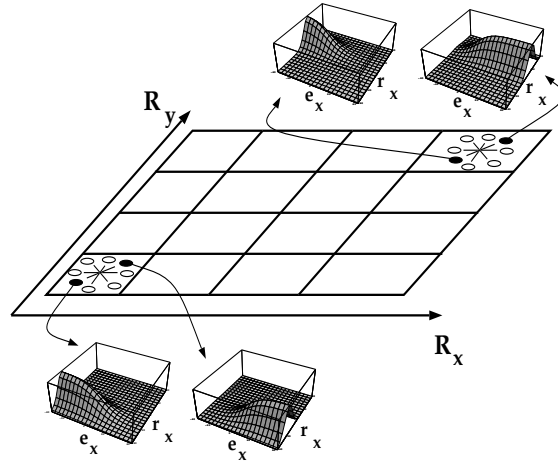


Figure IV.5: A spatial representation using basis functions. Units have a gaussian retinal receptive field multiplied by a sigmoid of eye position. For each retinal location, a small population of units represents all possible gaze angle. the spatial location of an object is represented by patterns of activity in this map.

These two predictions impose some important constraints on the class of neuronal responses that would be compatible with our basis function hypothesis. For example, certain functions would not work, such as linear functions, nor would the sums of nonlinear functions of \vec{R} and \vec{E} . Discontinuous or very irregular functions would work in principle but would be extremely inefficient for learning.

Visual receptive fields of parietal neurons are typically smooth and nonlinear, which is consistent with our first prediction. Gaussian or sum of gaussians provide good models of their profile, as seen in figure IV.6. The eye position selectivity, however, called the gain field (figure IV.1B), appears quite linear.

Andersen et al.(1988,1990b) performed a linear regression analysis on a large sample of parietal neurons. This analysis revealed that about 40% of the cells had a planar gain field, another 40% had a planar component in their gain field (they were not purely planar but were monotonically increasing in one direction of space), and the final 20% had non-planar gain fields (Andersen and Zipser, 1988; Andersen et al., 1990b). (see figure IV.7 for four examples). Although linear regression analysis revealed that about 80% of the gain fields are either planar or have a planar compo-

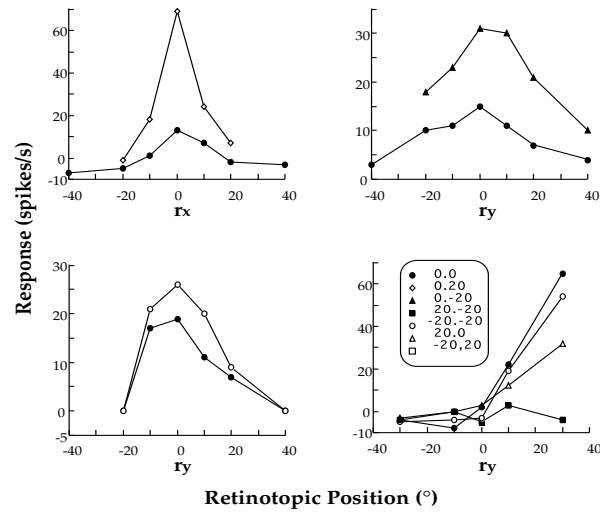


Figure IV.6: Four typical visual receptive fields of parietal neurons. Each receptive field is shown for several gaze angles (adapted from Andersen et al., 1985).

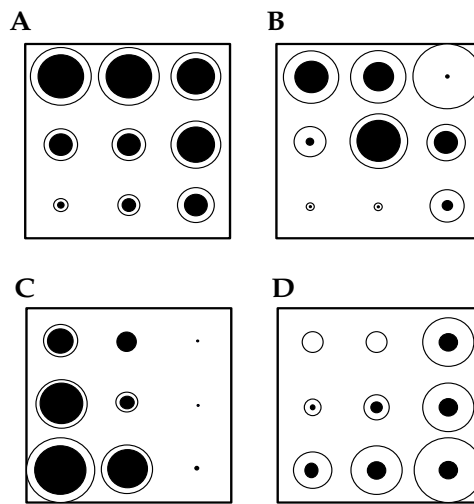


Figure IV.7: Four typical gain fields of parietal neurons. Notice that gain fields B and C show clear signs of saturation. See figure IV.1B for explanation of circles (adapted from Andersen et al., 1988).

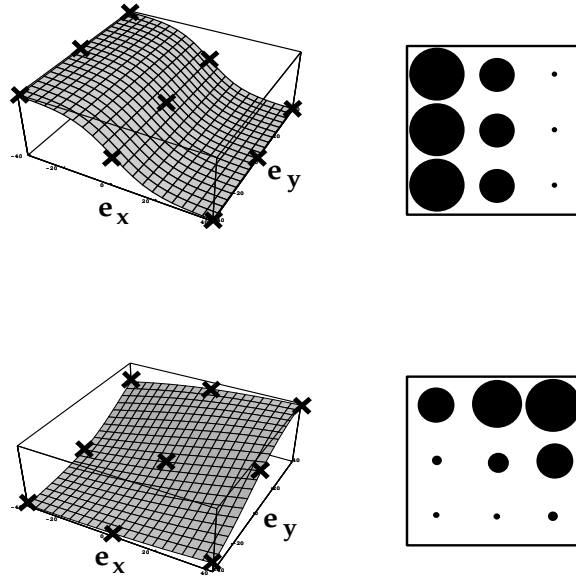


Figure IV.8: Examples of two sigmoid functions sampled at nine eye positions (crosses) showing that they are nearly planar gain fields.

ment, a closer analysis shows that a planar component will be overestimated by the procedure used to sample the gain fields.

Figure IV.8 illustrates that sampling a sigmoid at nine symmetrical positions results in a gain field that would appear to be planar if tested with a linear regression analysis. If the inflection point of the sigmoid is not exactly at $e_x = e_y = 0$, the resulting gain field will look less and less linear, but they would still be monotonic and a linear regression analysis would find that there is a statistically significant linear components to the tuning. Therefore, the 80% of neurons whose gain fields are either planar or have a planar component are consistent with sigmoid gain fields. The remaining 20% classified as non-planar had a peak of activity at one of the 9 sampling positions (Andersen and Zipser, 1988; Andersen et al., 1990b); these might be just as useful as the others and are consistent with the first prediction above.

To distinguish a genuine planar tuning to eye position and a sigmoidal tuning, one would have to show that the activity of neurons saturate either at zero or at maximum firing rate within working range of eye position and retinal locations. A linear code (also called vectorial code as we will show later) would require that such

saturation takes place only when the eye position reaches its physical limits (beyond ± 50 degrees). Sigmoidal tuning on the other hand would predict that neurons should saturate within the working range of eye position.

Saturation of responses in Parietal Cortex

The predicted saturation was tested by examining data from neurons in parietal cortex. Saturation at maximum firing rate is difficult to demonstrate because it requires a large number of measurements. Saturation at zero, on the other hand, can be deduced by linearly interpolating the gain field of each cell beyond the experimental sample points to determine which eye position would silence the cell. We have performed this analysis on the gain fields of 174 cells recorded in the parietal area 7a by Andersen et al. (Andersen, Essick and Siegel, 1985). An example of one of these gain field is shown in figure IV.1-B. Gain fields were fitted with planes such that the activity, a , of each cell was approximated with:

$$a = \alpha e_x + \beta e_y + \gamma \quad (\text{IV.2})$$

where α , β and γ were obtained with a linear regression analysis (see Andersen and Zipser, 1988, for more details on this procedure). The x-intersect, θ , that is to say, the minimum eye deviation for which the cell would stop firing, is given by:

$$\theta = \frac{\gamma}{\beta} \cos(\text{atan}(\frac{\alpha}{\beta})) \quad (\text{IV.3})$$

Figure IV.9 shows the distribution of the x-intersect. Although, the distribution is not uniform, many cells saturate within the $\pm 50^\circ$, the working range of normal saccadic eye movements.

The two large peaks on the sides are for cells having intersects beyond $\pm 80^\circ$. These cells might be either true linear cells or cells with sigmoidal tuning that are near saturation at maximum firing rate around ± 20 degrees. Additional measurements are needed to distinguish between these two possibilities.

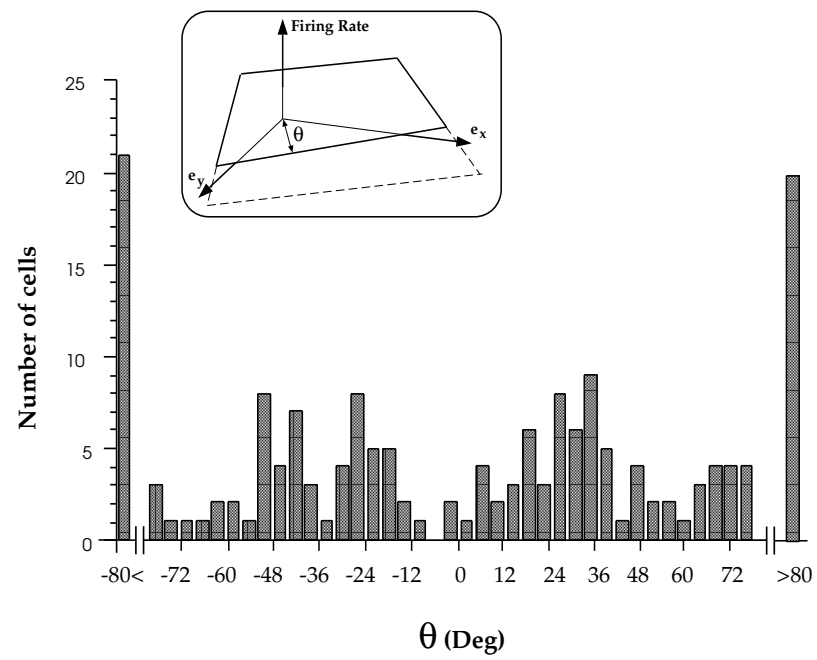


Figure IV.9: Histogram of the gain field x-intersects, θ , of 174 cells recorded in area 7a (data from Andersen et al., 1988). As illustrated in the inset, the x-intersect was defined as being the smallest eye deviation that would silence the cell.

The data in figure IV.9 provide clear evidence of saturation at zero firing rate within working range of eye positions. We cannot conclude yet that the tuning to eye position is sigmoidal since we do not have enough data to demonstrate saturation at maximum firing rate. We can conclude, however, that the tuning is nonlinear within the normal range of eye positions. As we demonstrate in the next section, the saturation at zero by itself is actually sufficient to estimate nonlinear functions.

Integration of retinal and eye position

It is not enough that the responses of parietal neurons are nonlinear functions of \vec{R} and \vec{E} for each variable individually; it is also essential that the two inputs are combined nonlinearly. According to the second prediction, the responses to \vec{R} and \vec{E} should not be a linear combination of the two responses. A simple nonlinear interaction would be the product of the separate functions, as we have assumed in our model, but there are other possibilities, such as a weighted sum of the inputs squashed through a sigmoid activation function (Zipser and Andersen, 1988a). There is evidence for both multiplicative and sigmoidal nonlinearities in the way that parietal neurons integrate eye position and visual signals (Andersen, Essick and Siegel, 1985).

The second prediction precludes representations in which separate populations of cells are dedicated to \vec{E} and \vec{R} , as in the input layer of the network in figure IV.3. An intermediate representation is needed in which units respond jointly to both inputs. These hidden units can be thought of as pre-computed "building blocks" from which an arbitrary nonlinear transformation can be synthesized. The receptive field properties of parietal neurons are broadly consistent with this prediction. One of the earliest reports of gain fields in parietal neurons suggested gaussian visual receptive field multiplied by an eye position signal as an adequate model and argued that a population of these cells would be sufficient to create head-centered visual receptive field (Andersen, Essick and Siegel, 1985). Basis functions formalize this insight, emphasize the importance of the nonlinearities for the eye position signals, and provide a general framework for sensorimotor coordination.

IV.B.4 Simulations

The accuracy with which a sum of basis functions can approximate a transformation depends on the number of basis functions used. Perfect accuracy is only possible in the limit as the number of basis function becomes infinite. Good approximations can be obtained to many functions with a reasonable number of basis function. We illustrate this point by showing how the same set of gain-modulated units can be used to generate two output functions that are computed in the premotor cortex and the superior colliculus: a head-centered and a retinotopic receptive field.

The model used 121 gain-modulated units, corresponding to the hidden units in figure IV.3, whose response functions were computed by multiplying a gaussian retinal receptive field with a sigmoid of eye position:

$$h_i = \frac{e^{-\frac{(r_x - r_{x_i})^2}{2\sigma^2}}}{1 + e^{-\frac{e_x - e_{x_i}}{T}}} \quad (\text{IV.4})$$

where h_i is the activity of unit i . The peaks of the gaussians, r_{x_i} , were spread uniformly between -60° and 60° in increments of 12° . The standard deviation of the gaussian, σ , was fixed at 18° . This corresponds to a radius of 25° , defined as the distance from the peak corresponding to 37% of maximum activity. An average radius of 22° has been reported in area 7a (Andersen, Essick and Siegel, 1985). The inflection points of the sigmoids, e_{x_i} , were also uniformly spread between -40° and 40° in steps of 8° . The temperature factor, T , was set at 8° . Four typical units used in the simulations are shown at the bottom of figure IV.10.

In a second series of simulation, we used a different set of functions which were obtained by multiplying a gaussian of r_x by a piecewise linear function of e_x :

$$h_i = e^{-\frac{(r_x - r_{x_i})^2}{2\sigma^2}} f_i(e_x) \quad (\text{IV.5})$$

$$f_i(e_x) = \begin{cases} e_x & \text{if } e_x > e_{x_i} \\ 0 & \text{otherwise} \end{cases} \quad (\text{IV.6})$$

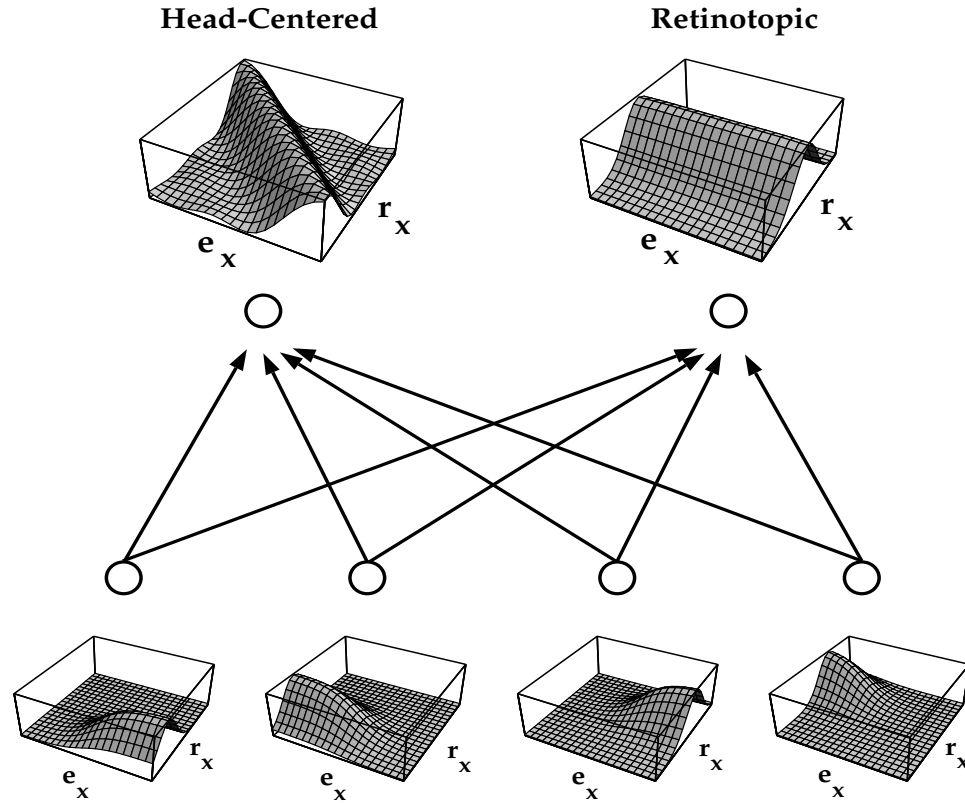


Figure IV.10: Approximating a head-centered and a retinotopic gaussian receptive field by the use of the same gain-modulated input units. Only four units are shown, but the approximations were obtained with 121 units.

These functions look similar to the previous ones except that activity does not saturate at a maximum level. The goal of this second simulation was to demonstrate that saturation at zero is sufficient as long as the e_{x_i} are spread over the range of all possible eye positions.

The two output functions were gaussian functions of a_x and r_x respectively, with a standard deviation, σ , of 18° . The peak in both case were at $a_x = r_x = 0$. A wide range of peak position could have been approximated equally well by the set of basis functions.

We used a supervised optimization procedure for determining the weights w_i between the basis functions and output unit. This procedure was used only to find such a set of weights and not to model the actual process that might be used

in the brain to learn these weights. The optimization procedure, called the delta rule (Widrow and Hoff, 1960), minimized the square error between our estimation, o , and the actual function, o^* , over all possible examples, p :

$$E = \sum_{p=1}^M (o_p^* - o_p)^2 \quad (\text{IV.7})$$

where

$$o_p = \sum_{i=1}^{121} w_i h_{ip} \quad (\text{IV.8})$$

The training set was composed of 441 pairs of retinal position, r_x and eye position, e_x , selected from 21 different retinal location within the range -40° and 40° , and as many eye positions between -20° and 20° . Weights were adjusted until the approximation was on average within 3% of the actual values. Figure IV.10 shows the resulting approximation for a head-centered and a retinotopic receptive field when using product of gaussians and sigmoids. Identical results (not shown) were obtained when using the second type of basis functions.

The retinotopic response shown in Fig. IV.10 could correspond to a presaccadic motor field found in the superior colliculus and the frontal eye field, which are known to play an important role in the generation of saccadic eye movements. The head-centered gaussian tuning curve is similar to the receptive properties of some of the neurons in the premotor cortex and putamen, which are believed to be involved in reaching. That both functions can be approximated from the same pool of units suggests that the parietal cortex could be involved in coordinating different types of movement toward a visual stimulus.

The ability to generate any nonlinear function in the output stage is a defining characteristic of a basis function representation. Had the hidden units in the network been linear, these functions could not have been approximated by a linear recombination of hidden units activity.

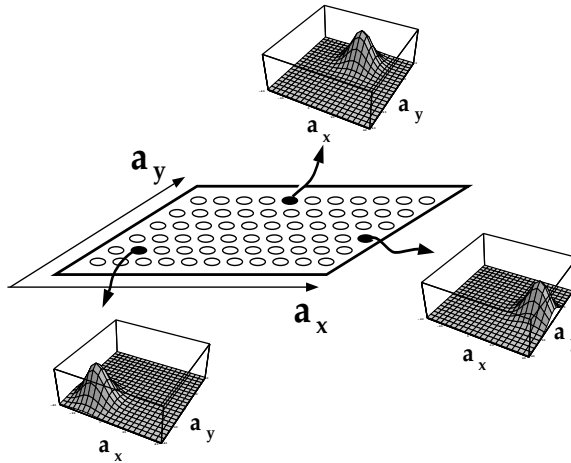


Figure IV.11: Computational map for the head-centered location of an object, \vec{A} . Each unit respond to a limited range of a_x and a_y with a gaussian tuning.

IV.C Comparison with Other Representations

In this paper we have considered a new type of spatial representation using basis functions. Here we compare basis functions with alternative approaches. The two most common types of distributed representation for space are computational maps (Knudsen, du Lac and Esterly, 1987) and vectorial schemes (Soechting and Flanders, 1992; Goodman and Andersen, 1990; Touretzky, Redish and Wan, 1993).

IV.C.1 Map Representation

Several brain structures represent vectors by two-dimensional maps of neurons. The representation is used for the retinal position of visual stimuli in area V1 and for encoding the direction and amplitude of the next saccadic eye movement in the superior colliculus (Lee, Rohrer and Sparks, 1988). The parietal cortex may, in a similar manner, represent the egocentric location of object as a two-dimensional map, which would correspond to a head-centered map (See figure IV.11). Each neurons in the map would discharge for a limited range of values of a_x and a_y so that their receptive fields would be fixed in head-centered coordinates. When the head and body were fixed, neurons in such a map would respond to visual stimulation at a particular

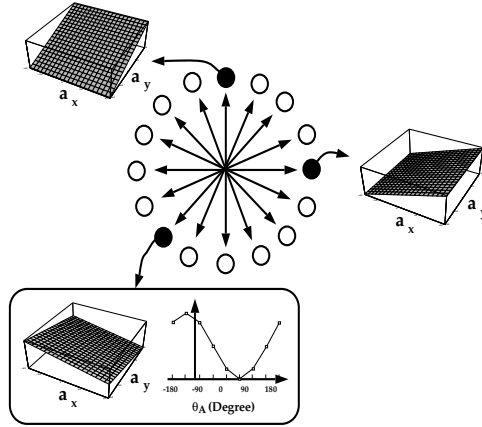


Figure IV.12: Vectorial representation for the head-centered location of an object, \vec{A} . Each neuron computes the projection of \vec{A} along its preferred direction (central arrows). As a consequence, the tuning curve for a_x and a_y is planar, whereas the tuning curve to θ (the angle between \vec{A} and the cells preferred direction) is a sine function.

location in space, regardless of where the eyes are fixating.

There is no evidence for such neurons in the parietal cortex. It would also have limited computational utility since a place code for location in space does not form a set of basis functions and would could not be used, for example, to compute the retinal location of the stimulus.

IV.C.2 Vectorial Representation

Definition

The components of a two-dimensional vector are typically the projections of the vector along the horizontal and vertical axes. The choice of the axes is, however, arbitrary. The same two-dimensional vector can be represented by its projection on any pair of axes as long as they are independent (see figure IV.12). The parietal cortex could encode the egocentric position of object, \vec{A} , by projections along vectors, so that the firing rate of each neuron would report the projection of \vec{A} along a direction specific to each neurons. This entails that the activity, h , of a neuron can be modeled

as :

$$h = \vec{W}_a^T \vec{A} = \|\vec{W}_a\| \|\vec{A}\| \cos(\theta) \quad (\text{IV.9})$$

where, θ is the angle between the egocentric position of the object, \vec{A} and the vector \vec{W}_a . \vec{W}_a is called the preferred direction of the cells because the activity is maximum whenever $\theta = 0$; that is, when \vec{A} points in the same direction as \vec{W}_a . this representation predicts that neurons should have a cosine tuning to the direction of the egocentric location of object. Hence, if an object is moved in the visual field along a circle centered on the point of fixation, the neuron response should follow a cosine tuning function.

Such cosine tuning have been reported in the motor cortex for the direction of hand movement, suggesting that the motor cortex uses a vectorial code for the direction of hand movement in extrapersonal space (Georgopoulos et al., 1989, but see Sanger, 1994). The same scheme has been also used by Goodman and Andersen (1990), and Touretzky et al. (1993) to model the encoding of egocentric position of objects in the parietal cortex. Touretzki et al. (1993) called their representation a *sinusoidal array* to refer to the cosine tuning of the units.

Neurons in the parietal cortex do not receive directly the egocentric position of objects. Instead they receive signals related to the retinotopic position of object, \vec{R} , and the current eye position, \vec{E} . Upon substituting \vec{A} from equation IV.1 in equation IV.9, the response of the unit to \vec{R} and \vec{E} is:

$$h = \vec{W}_a^T (\vec{R} + \vec{E}) = \vec{W}_a^T \vec{R} + \vec{W}_a^T \vec{E} \quad (\text{IV.10})$$

This equation is linear in \vec{R} and \vec{E} , which leads to two important predictions regarding the tuning of parietal neurons to the retinal location of the object and the eye position.

Predictions of the Vectorial Representation

1. The visual and eye position receptive fields of parietal neurons should be planar.

2. The preferred direction for retinal location and eye position should be identical. For example, if a unit responds best to a stimulus on the right of the retina it should also respond maximally when the eyes are fixating on the right.

The visual receptive field of a cells is determined by systematically varying the retinal position of the stimulus while keeping the eye constant. From equation IV.10, the function should be linear in \vec{R} , which defines a plane. The same argument holds for the eye-position receptive field.

The visual receptive fields of neurons in the early visual areas are not linear. In V1, for example, receptive fields are typically gaussian with a half width of about $1/2^\circ$ or less. In contrast, the vectorial representation predicts that parietal neurons visual receptive field should cover the entire visual field and the response to a stimulus should linearly increase in one direction of space. The direction of maximum increase is what is called the preferred direction, and it is therefore equivalent to the vector \vec{W}_a in equation.

The second prediction is a consequence of the fact that \vec{R} and \vec{E} in equation IV.10 are multiplied by the same vector \vec{W}_a which defines the preferred direction.

In the next section, these two predictions are compared to available neurophysiological recordings of parietal neurons. Neurons are intrinsically nonlinear and it would be unreasonable to expect them to have perfectly linear responses to any stimulus. The key question to ask is whether the nonlinearities are large and functionally significant for the cortex or whether they are small irrelevant random deviations from linear responses.

Match Between the Vectorial Representation and Parietal Cortex

The receptive field of a neuron for eye position is formally equivalent to its gain field, 80% of which are linear or contain a linear component in the parietal cortex (Andersen and Zipser, 1988; Andersen et al., 1990b). This would therefore appear to be consistent with the first prediction of the vectorial hypothesis. However,

20% are nonlinear, and, as discussed above, the other 80% may be better describe by sigmoids and some of these show clear sign of saturation.

In contrast, visual receptive fields in the parietal cortex are not even approximately planar, as shown in figure IV.6. They are typically bell-shaped, sometimes with multiple peaks (Andersen et al., 1990a). Only a small percentage of neurons appears to have extremely large receptive fields which cover almost the entire visual field, as predicted for a vectorial code. On average the receptive field diameter are about 44° in diameter, which is large when compared to earlier visual areas, but still well under the full extent of the visual field, (180°) (Andersen et al., 1990a).

The second prediction made by the vectorial hypothesis regards the visual and eye position preferences of single cells. The only two cells for which both the visual receptive field and the gain field are available have opposite preferred directions for retinal and eye position (See figures 1 and 6 in Andersen and Zipser, 1988). Clearly more data are needed on the correlation between the preferred eye and retinal position. Indirect evidence suggests, however, that if such a correlation exists, it is probably not strong. The majority of neurons in area 7a and LIP ($\approx 70\%$) tend to have the peak of their retinal receptive fields (Barash et al., 1991b) or their preferred retinal direction (Andersen et al., 1990a) on the contralateral side, but no significant bias appears to exist for the preferred eye position. Measurement of the distribution of the direction of the gain field indicates that they are uniformly distributed along a circle within each cortical hemisphere (Andersen et al., 1990a; Andersen et al., 1990b). If preferred eye and retinal position were strongly positively correlated one would expect that both should be either uniformly distributed or both biased toward the contralateral side.

In conclusion, the experimental data do not appear to be fully consistent with the predictions of the vectorial code. The visual receptive fields, in particular, are strongly nonlinear. It is still possible, however, that these nonlinearities are averaged out in subsequent stages of processing in the cortex so that the next results is a linear mapping. Most sensori-motor mappings are nonlinear, but, as we argue in

the discussion, there might be particular tasks for which a linear mapping might be required.

IV.D Discussion

IV.D.1 Basis Function Representation and Frame of Reference

In the vectorial and map representation of \vec{A} , the retinal position of the object cannot be recovered directly and these representations could justifiably be called head-centered. In contrast, the basis function representation does not encode the head-centered location of objects. It might be tempting to call it a distributed representation of \vec{A} since it is possible to recover \vec{A} by a simple linear combination and without additional signals; however, many other functions of the retinal position and eye position of the object can be recovered as well. In a basis function representation there is no preferred reference frame, although for a given experiment the responses of neurons could be interpreted as having the frame of reference imposed by the task (Sanger, 1994).

The ability of basis functions to support multiple reference frames can be tested. One strong prediction is that hemineglect resulting from lesions in the parietal cortex should not be confined to a particular frame of reference. Recent studies of parietal patients have reached precisely this conclusion (Driver and Halligan, 1991; Behrmann and Moscovitch, 1994). Further predictions can be made by lesioning the model and comparing the pattern of breakdown with deficits reported in patients with a variety of parietal lesions.

IV.D.2 Basis functions compared to the Zipser and Andersen network

Basis functions are particularly useful when the representation is being used to compute several functions of the input. If a system needs to compute only one type of output it is generally possible to find intermediate representation that is more compact and specialized for that particular task.

The Zipser and Andersen model of the parietal cortex (Zipser and Andersen, 1988a) is an example of such a special-purpose nonlinear transformation. This network was trained with backpropagation to compute the position of an object in egocentric coordinates. The input was similar to that in figure IV.3, but the output was trained to represent the egocentric position of the stimulus with either a vectorial code or a map output – the latter case is illustrated in figure IV.3. When the output unit is chosen to be linear rather than sigmoidal, backpropagation tends to develop a vectorial code in the hidden layer (Goodman and Andersen, 1990). The gain fields, in particular, are planar and rarely saturate within working range of eye position. The representation becomes specialized and is unable to approximate other functions.

With the map output on the other hand, the representation was more nonlinear, with the gain field saturating for common range of eye position (Andersen and Zipser, 1988). Nevertheless, it is likely that the learning procedure had found nonlinearities that were specialized for a head-centered gaussian output and as such may not have been usable for other type of outputs.

Whether the parietal cortex has a specialized representation or a basis function representation is not easy to resolve. It could be that each column in parietal cortex is a different specialized representation. The parietal cortex could support hundreds of such transformations. In a basis function representation, the same neurons would be involved in the coordination of many different types of behavior. The fact that hemineglect is often generalized across sensory modalities and across behaviors is consistent with this possibility (Pouget and Sejnowski, 1992). Alternatively, the specialized modules for different behaviors might all be damaged in a large parietal

lesions (Stein, 1992).

If a module implements a linear transformation it could use a vectorial code. There are many cells in the parietal cortex with very large visual receptive fields ($> 80^\circ$) and linear gain fields, properties that do not fit well with the basis function approach. Goodman and Andersen (1989) note that a linear function of the head-centered position of an object in space corresponds to the position that the eye should adopt to foveate the object seen on the retina. Whether such signals are used directly to control eye position remains controversial (Schlag and Schlag-Rey, 1990; Gnadt, Bracewell and Andersen, 1991; Goldberg and Bruce, 1990; Duhamel, Colby and Goldberg, 1992). Droulez and Berthoz (1991) have argued that, instead, such neurons could be used for dynamic remapping of motor command in double saccade experiments (Mays and Sparks, 1980; Gnadt and Andersen, 1988; Goldberg and Bruce, 1990).

IV.D.3 Beyond parietal cortex

Can the basis function representation that we have explored in the limited domain of visual input with fixed head position be generalized to other sensory inputs and unrestricted motion of the body? The parietal cortex is at an intermediate stage of transformation between sensory and motor representations, as suggested by lesion studies; however, other regions of the brain may also be involved.

Modulation of sensory response by eye position has been reported outside of the parietal cortex. Presaccadic activity of some neurons in the supplementary eye field in the prefrontal cortex are clearly modulated by eye position. Their motor fields are nevertheless retinotopic as suggested by the fact that they discharge maximally before saccades of a particular directions regardless of starting eye position (Schall, 1991). A careful study by Boussaoud et al. has led to similar conclusions for most visual neurons in the ventral premotor cortex (Boussaoud, Barth and Wise, 1993). In all cases, motor or receptive fields are limited in space and typically bell-shaped. The modulation by eye position appears to be smooth and nonlinear, although there

have been no attempt so far to carefully determine their profiles.

Modulation of responses with eye position have also been observed in the striate cortex (Trotter et al., 1992; Weyand and Malpeli, 1993), extrastriate cortex (area V3a (Galletti and Battaglini, 1989) and DP (Andersen et al., 1990a)) and even in the lateral geniculate nucleus (LGN) (Lal and Friedlander, 1989). These areas have a clear retinotopic organization, but we have shown in a previous modeling study that eye position modulation provides evidence for early spatial transformation in the visual system (Pouget, Fisher and Sejnowski, 1993). We called these early spatial representations, a *retinospatiotopic map* (figure IV.5 is a good example of a retinospatiotopic map).

What distinguishes the representation found in area V3a, for example, from that in area 7a, is the resolution of the visual map. The smaller size of the visual receptive fields in V3a could provide a spatial representation of object subparts, whereas area 7a might be more concerned with whole objects. Regardless of the size of the retinal receptive fields, a retinospatiotopic map is a basis set.

Therefore, the basis function hypothesis might be applicable to spatial representations outside of the parietal cortex, from the primary visual cortex to the premotor cortex (Pouget, Fisher and Sejnowski, 1993; Pouget and Sejnowski, 1994).

IV.D.4 Other sensory modalities and posture signals

The visual response of many parietal neurons is also modulated by head position (Brotchie and Andersen, 1991) and vestibular inputs (Snyder, Brotchie and Andersen, 1993). As observed for eye position, sensory evoked activity increases monotonically as the head is moved in the preferred direction of the cell.

Basis functions can include head position signals in a manner that is similar to eye position signals, as a product of a gaussian retinal receptive field with a sigmoidal function head position. A neuron that responded to all three would be characterized by the retinal receptive field and two preferred directions, for the eye and head position inputs. In principle, it is possible to compute any function of

these three variables with basis functions having all possible combinations of preferred directions and receptive field location. This would predict that over the entire population of cells, there should be no correlation between preferred directions for eye and head position. A similar reasoning would apply to vestibular signals.

Preliminary results indicate that, in fact, preferred eye and head direction are positively correlated (Brotchie and Andersen, 1991). If confirmed, this finding puts important constraints on the basis functions present in the parietal cortex. A strong positive correlation between preferred eye and head position suggests that eye and head position might be added together, in which case the resulting basis functions could not be used to compute transforms that depend on the two inputs separately. The cells would thereby encode basis function of the retinal position and gaze angle, defined with respect to the body; that is, eye position added with head position. Such a neuronal population would encode simultaneously the position of objects in retinal and body-centered coordinates.

IV.D.5 Conclusions

The fundamental assumption underlying this paper is that the spatial representation in the parietal cortex can be best understood from the perspective of sensorimotor transformation. We propose that the parietal cortex does not represent the positions of objects in egocentric coordinates or any other coordinate system. Instead, inputs signals, such as retinal location of objects and eye position, are represented in a format suitable for nonlinear function approximation, a critical intermediate step for sensorimotor coordination.

Available neurophysiological data are consistent with this hypothesis. A neuron with a restricted visual receptive fields modulated as a monotonic function of eye position can be modeled by a product of a gaussian and a sigmoid. Since products of gaussians and sigmoids form basis functions, this representation can be used to approximate any nonlinear functions of the input variables.

Previous attempts to characterize spatial representation have emphasized

linear encoding schemes that cannot be used for nonlinear function approximation and as such may not be adequate for sensorimotor coordination. These other representations are computationally interesting for some important operations, such as vector rotation. Some parts of the brain more specialized in navigation like the hippocampus might use such a scheme (Touretzky, Redish and Wan, 1993).

Chapter V

Future Work

V.A Development of spatial representations

Representing the spatial position of object requires integrating sensory inputs with internal posture signals such as eye or head position. The mechanisms responsible for the anatomical convergence of these signals during the first few weeks or months of life are still largely unknown. Genetical determinants might be involved but it seems fairly implausible that they could explain the whole process. Spatial representations are intimately related to sensori-motor integration whose development is clearly dependent on interacting with the real world. Therefore, it is likely that the maturation of spatial representations involves some activity-dependent process during which proper connections, as well as their weights, are set.

Numerous models of such activity-dependent maturation process have used either, the hebb rule (Linsker, 1986a; Linsker, 1986b; Linsker, 1986c; Miller, Keller and Stryker, 1989), or the delta rule (Kuperstein, 1988) for synaptic plasticity. Unfortunately, none of these models can be applied to spatial transformations. The delta rule requires a high dimensional teaching signal which is not believed to be available in the parietal cortex. The hebb rule works by discovering covariances in the input data (Oja, 1982), but there is no guarantee that the first-order statistical properties of the inputs are relevant for the overall task performed by the system. In fact, in

the case of sensori-motor coordination, there is no reason to believe that posture and sensory signals are correlated at all (Zipser and Andersen, 1988b). In such a situation, the hebb rule would keep these signals segregated, which is the opposite of what we need for spatial representations..

There are, however, many alternatives to the hebb rule which have been developed over the past few years. These learning rule can be roughly categorized into unsupervised and self-supervised rules. In the case of unsupervised learning weight are updated in order to satisfy some optimality principle such as maximization of information. In the self-supervised case, the development is the results of interactions in which one input serves as a teacher for the other inputs.

V.A.1 Preliminary work: Predictive learning

We have recently explored, in collaboration with Montague, Dayan and Sejnowski, a self-supervised learning mechanism for the development of multisensory spatial representations for saccadic eye movements (Pouget et al., 1993). Our model learned to generate spatially accurate saccades toward auditory target by using the visual system as an internal teacher. Several neurophysiological experiments have demonstrated that vision plays a critical role in the development of auditory maps in the superior colliculus (Withington, Binns and Keating, 1990; Knudsen and Knudsen, 1991). This would seem quite natural given that eye movements are essential for proper vision, and that visual targets and motor command for saccadic eye movements are both specified in retinotopic coordinates (Sparks, 1991). Therefore, the visual system is in a perfect position to teach other sensory modalities, which do not use retinotopic coordinates, such as audition, to remap in the appropriate frame of reference.

Our model learned by adjusting the weights after successful eye movements toward bimodal -visual and auditory -targets. We used a predictive hebbian learning rule (Montague et al., 1993) which combines the hebb rule with reinforcement learning. Weights were updated according to the covariance between the input and

output activities gated by unexpected reinforcement, that is to say, the difference between reinforcement and predicted reinforcement. Reinforcement was controlled by the foveal representation of the visual system, such that it was delivered only after successful eye movement. The predicted reinforcement, on the other hand, was computed by the auditory system. Since weight update were proportional to the difference between this two terms, learning proceeded until the auditory system could predict away the reinforcement signal controlled by vision, thereby placing the visual system in a teacher position.

Upon completion of the learning phase, we were able to show that the response properties of the hidden units in the network were in agreement with the tuning curves that have been reported in the parietal cortex and the superior colliculus (Jay and Sparks, 1987; Stricanne, Mazzoni and Andersen, 1993). We found however that the gain of the eye movements was slightly different when the network was presented with purely auditory versus bimodal targets. Subsequent analysis revealed that the problem stemmed from the linear representation used in the intermediate layer of the network. In a linear representation, the output of the units is roughly proportional to the number of inputs. Therefore, the activity in the hidden layer was larger for bimodal targets compared to auditory targets, leading to a slight gain difference in the resulting eye movements.

This problem can be readily solved by using basis function units in the hidden layer, but learning would become more complex. The predictive hebbian learning rule in its present form does not work with arbitrary network of nonlinear units. To our knowledge, there is no biologically realistic rule that can deal with such units. There are however some promising possibilities to explore. Hence, the rule used in Kohonen maps (Kohonen, 1982), or the algorithm of Saund (Saund, 1989), or Zemel and Hinton (Zemel and Hinton, 1994), would work for gaussian activation functions. It is not yet clear how this would generalize to our particular problem, or whether biologically plausible implementation of this rule could be found, but this is a path worth exploring.

V.A.2 Optimum representation

More recent work on information maximization might also be relevant. Spatial representations can be thought of as being an optimal combination of sensory and posture signals. Optimal in this case would be defined in terms of information content for which there exists an objective formulation (Shannon and Weaver, 1963).

Bell (Bell and Sejnowski, 1995) has recently derived a learning rule which can successfully perform a gradient ascent in information content for a network containing an identical number of input and output units. It should be possible to generalize this approach to arbitrary network architecture with nonlinear activation functions. In the general case, this approach leads to a non-local learning rule, i.e., changing one weight depends on local pre and post-synaptic activities as well as the values of every other weight in the network. We suspect however that when the derivative of the activation function has a local support, that is, when the derivatives are non-zero only over a finite range of input values, it should be possible to derive semi-local rules in which the change in one weight depends on the change in other weights in a local neighborhood.

This approach could potentially demonstrate that spatial representations found at the highest level of the cortical hierarchy are the results of an unsupervised learning rule which optimizes the information content of the representation. The possibility that self-organizing principles might be responsible for the development of topographic maps in the primary cortices has been raised before (Linsker, 1986a; Miller, Keller and Stryker, 1989), but it has never been applied to higher level representations.

V.B Modeling Hemineglect

Chapter 3 and 4 demonstrated that the basis function hypothesis is a viable alternative to the type of coding schemes that have been previously proposed to account for single cell responses. The next logical step is to show that the same

framework can be applied at the neuropsychological level with little modification. Parietal cortex lesions result in a well-known neurological disorder called hemineglect which has been extensively studied since its initial discovery. This literature provides a valuable source of observations against which competing model can be tested.

Hemineglect is characterized by a conspicuous lack of reaction to contralateral stimulus as well as little if no motor exploration in the same half hemispace. Several tests have revealed that this behavior is not the result of simple sensory deficits like hemianopia. Rather, patients show a lack of interest in sensory stimuli that are nevertheless perceived in the lesioned hemispace. The spatial nature of the deficit, in particular the fact that it seems to be restricted to one hemispace, suggests that the lesion affects either the spatial perception of object, or, the spatial guidance of attention. There is evidence for both theories suggesting that they describe complementary aspects of the deficit (Heilman, Watson and Valenstein, 1985).

There have been multiple attempts to determine which frame of reference is affected by hemineglect. All of these experiments work by dissociating the frames of references across experimental conditions. In the control condition, patient perform a spatial task with their head upright and their eyes fixating straight ahead. Then, they are asked to perform the same task while their head is tilted or rotated, or while their eyes deviate from the straight ahead direction.

Ladavas et al have been among the first to use this paradigm. They measured reaction time to visual stimulation when the patient head was upright versus tilted by 90° to the right or left (Ladavas, 1987; ?). It had been known previously that, with the head upright, patients with right parietal lobe lesions showed longer reaction time for stimulus displayed in the left visual field versus right visual field.

If the deficit were purely retinotopic, one would expect that tilting the head would not influence the reaction time pattern, namely, subjects should be slower for targets located in the left visual field. They found, however, a significative interaction between head position and reaction time, suggesting that the deficit is both retinotopic and body-centered. Several experiments support this conclusion that the

parietal cortex contains a representation in which several frames of reference coexist (?; Farah et al., 1990; Karnath, Schenkel and Fischer, 1991; Driver and Halligan, 1991; Moscovitch and Behrmann, 1994; Behrmann and Moscovitch, 1994).

The coexistence of several frames of reference within the same representation is one of the hallmark of the basis function hypothesis, therefore, it seems likely that a lesion in a basis function representation would affect several frames of reference simultaneously. Imagine, for example, that each hemisphere contains a basis function representation covering all possible retinal locations and eye positions, but with a higher density of neurons for contralateral positions. A unilateral lesion would damage one of these maps and subjects would be left with a biased representation. As a result, objects associated with either ipsilateral retinal location or ipsilateral eye position would activate larger neuronal population than objects associated with more contralateral positions.

Therefore, when multiple stimuli compete for a patient's attention, we predict that the subject is more likely to orient toward ipsilateral versus contralateral stimuli, where contralateral is defined both in retinotopic and eye position terms. Consequently, the deficit would be more pronounced for objects appearing on the contralateral side of the visual field while the eye fixate on the contralateral side. A similar model with head tilt, instead of eye position, could easily account for Ladavas results (1987).

Notice also, that in this model, the deficit is independent of the absolute position of object in space. Rather it will depend on the relative position of objects in the visual scenes. Hence, we postulate that subject do not ignore objects on the contralateral side but rather ignore object that are located on the contralateral side of other objects. Hence, deficit should be observed even if an object appears in the supposedly intact ipsilateral side if it is competing for attention with another object located even more ipsilaterally. Several experiments agree with this prediction (Kinsbourne, 1987).

Thus, the basis function representation appears to be a promising hypoth-

esis to model the behavior of patients with hemineglect. Our approach would be considerably strengthened if we can show that the same theory can be applied to the neurophysiological and neuropsychological levels.

Chapter VI

Conclusion

We have presented a series of models of spatial representations which account for the response characteristics of neurons involved in this process. Like previous studies, we have approached spatial representation as an intermediate stages in sensori-motor transformation. However, our work differs from the previous research in two important ways. First, we have shown that spatial transformations might be a process which is initiated in early stages of visual processing, as suggested by the finding of what we called retinospatiotopic maps, in area V1 and V3a. Second, we formalized the sensori-motor transformation as a problem of nonlinear function approximation, in contrast to the classical approach which cast the problem in term of change of coordinates.

The basis function idea which is developed, in chapter 3 and 4, provides an interesting formalization of the retinospatiotopic map concept of the second chapter. We have shown that neurons involved in spatial representation appear to compute basis function of their inputs variable, such as retinal location of objects or the current gaze angle. Hence, a gaussian function of retinal location multiplied by a sigmoid function of gaze angle matches quite well the response properties of single cells in the parietal cortex. A retinospatiotopic map is simply a collection of such neurons which covers all possible retinal location and eye position. Since our theory does not require any particular visual receptive size, it applies equally well to the representations in

early visual areas than it does to the parietal cortex.

In contrast, vectorial representation which requires large receptive fields, would be only applicable to the parietal cortex, but even there, this theory can account only for the linear component of the responses.

VI.A Explicit and implicit representations

One of the most important aspects of the basis function representation is that it contains multiple frames of reference in parallel. This notion may be able to resolve the long standing issue regarding explicit versus implicit representation. These terms are commonly used in the literature on spatial representation and it is sometimes argued that the parietal cortex may contain an implicit code for the egocentric representation of object. Unfortunately, these terms have never been clearly defined.

We believe that the term explicit should be applied exclusively to representation of vectors, whether this representation uses a map or a vectorial code. An examples of an explicit representation is the encoding of saccades direction and amplitude in the superior colliculus or the orientation of visual contours in V1. Such codes have all in common that they are invariant with respect to the variable they represent. If the input variable stays constant the overall population response stays the same. Hence the preferred orientation of a cell in V1 is the same whether the stimulus is a grating or a bar. Similarly, cells in MT respond maximally to a particular direction of motion across many stimulus configurations (Albright, 1992).

In the basis function representation, activity patterns are not invariant with respect to vectors that can nevertheless be recovered in a simple transformation of the activities. Hence, given a pattern of activity in a retinospatiotopic map, the corresponding retinal and egocentric location can be computed through a simple linear combination of the activity. However, there is an infinity of activity pattern that would lead to the same egocentric location or the same retinal location. Therefore, the representation is not invariant. In that sense, the representation is implicit, although

the term implicit is not very informative. It would be more precise to say that the representation contains an optimum representation of the sensory variables for the purpose of nonlinear function approximation, a statement which avoids the issue of implicit representation altogether.

VI.B Modular architecture and basis functions

As Stein (1992) argues in a recent review, there is strong evidence that sensori-motor coordination involves a network of parallel modules which are specialized for a restricted number of transformations. Within the parietal cortex itself, several cortical areas have been found, each with its own distinctive set of input and outputs. In area LIP, neurons have visual receptive field and a motor receptive field for saccadic eye movement. In contrast, most neurons in area 7b have somatosensory responses and tend to fire prior to arm movements (Mountcastle et al., 1975; Gallese et al., 1994).

The hypothesis that the brain contains functional modules working in parallel has dominated cognitive science and neuroscience over the last 15 years (Fodor, 1983; Mishkin, Ungerleider and Macko, 1983; Kalaska and Crammond, 1992). Stein's proposal is one of the numerous examples of this approach. Although, this research strategy has led to tremendous progress in the study of the brain, it also raises a very difficult and yet, essential problem: how do the modules interact and coordinate their actions?

This is a particularly important issue for Stein's proposal. If each module uses its own set of coordinates, communication between modules would involve remapping the signals from the sender to the receiver coordinates, a costly transformation.

Several solutions have been offered in the literature to deal with this coordination problem. Oscillations, or synchronization, can be used as a mean to link activities across modules (Singer, 1993; Tononi, Sporns and Edelman, 1992), but it

is unclear how this scheme would be applied when modules use different coordinates. An alternative is to use a global representation which would work as a control tower, by having bidirectional projections communication with all the other modules and act as a coordinator.

This representation would specify where objects are located in space, regardless of the sensory modalities in which they are perceived, and in a format that would be usable for any kind of motor behaviors. Basis function representations are one of the few alternatives that would be general enough to meet these requirements. This hypothesis would predict that there should exist a cortical area in which neurons are responsive to any sensory modality while being gain modulated by all posture signals (Pouget and Sejnowski, 1992).

This solution, however, would run into a combinatorial explosion. With two degrees of freedom for eye position, three for head position, three for trunk position... etc, the number of dimensions, or degrees of freedom, that the representation would have to handle is somewhere between 20 and 30. In such a high dimensional space, the basis function framework is simply not feasible. The number of basis functions needed grows exponentially with the number of dimensions. If we use 10 basis function per axes, a fairly small numbers, we would need 10^{20} to 10^{30} neurons, a number that seems difficult to reconcile with the total number of neurons in the brain which is estimated to be in the order of 10^{11} .

One should therefore consider an intermediate possibility in which the basis function are used within specialized modules dealing only with a restricted numbers of inputs and outputs. It would still be impossible to assign a frame of reference to any of these modules, but each of them might contain only a limited number of coordinates systems, perhaps only 3 or 4. This consistent with the balkanized nature of the areas in the parietal cortex.

In such an architecture, modules sharing at least one frame of reference could easily communicate, without having to constantly change coordinates. Not all modules could talk with each other but they could be organized in a network

of modules in which consensus are reached in parallel. We have not studied yet how networks of basis function modules would behave, but we believe that this is a promising avenue to explore in the future.

Bibliography

- Albright, T. (1992). Form-cue invariant motion processing in primate visual cortex. *Science*, 255:1141–1143.
- Andersen, R. (1989). Visual and eye movement functions of the posterior parietal cortex. *Annual Review in Neuroscience*, 12:377–403.
- Andersen, R., Asanuma, C., Essick, G., and Siegel, R. (1990a). Corticocortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobule. *Journal of Comparative Neurology*, 296(1):65–113.
- Andersen, R., Bracewell, R., Barash, S., Gnadt, J., and Fogassi, L. (1990b). Eye position effect on visual memory and saccade-related activity in areas LIP and 7a of macaque. *Journal of Neuroscience*, 10:1176–1196.
- Andersen, R. and Brotchie, P. (1992). Spatial maps versus distributed representation. *Behavior and Brain Sciences*, 15(4):707–9.
- Andersen, R., Essick, G., and Siegel, R. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, 230:456–458.
- Andersen, R. and 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.
- Anderson, B. and Moore, J. (1979). *Optimal Filtering*. Prentice-Hall, Inc., Englewood Cliffs, N.J.

- Baldi, P. (1991). Computing with arrays of bell-shaped and sigmoid functions. In Lippman, R., Moody, J., and Touretzky, D., editors, *Advances in Neural Information Processing Systems*, volume 3. Morgan-Kaufmann, San Mateo, CA.
- Barash, S., Bracewell, R., Fogassi, L., Gnadt, J., and Andersen, R. (1991a). Saccade-related activity in the lateral intraparietal area. I. temporal properties; comparison with area 7a. *Journal of Neurophysiology*, 66(3):1095–108.
- Barash, S., Bracewell, R., Fogassi, L., Gnadt, J., and Andersen, R. (1991b). Saccade-related activity in the lateral intraparietal area. II. spatial properties. *Journal of Neurophysiology*, 66(3):1109–24.
- Behrmann, M. and Moscovitch, M. (1994). Object-centered neglect in patient with unilateral neglect: effect of left-right coordinates of objects. *Journal of Cognitive Neuroscience*, 6:1–16.
- Bell, A. and Sejnowski, T. (1995). A non-linear information maximisation algorithm that performs blind separation. In *Advances in Neural Information Processing Systems*, volume 7. Morgan-Kaufmann, San Mateo, CA.
- Blatt, G., Andersen, R., and Stoner, G. (1990). Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area LIP) in the macaque. *Journal of Comparative Neurology*, 299(4):421–445.
- Boussaoud, D., Barth, T., and Wise, S. (1993). Effects of gaze on apparent visual responses of frontal cortex neurons. *Experimental Brain Research*, 93(3):423–34.
- Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, 12:4745–4765.
- Brotchie, P. and Andersen, R. (1991). A body-centered coordinate system in posterior parietal cortex. In *Society For Neuroscience Abstracts*, page 1281, New Orleans.

- Burnod, Y., Grandguillaume, P., Otto, I., Ferraina, S., Johnson, P., and Caminiti, R. (1992). Visuomotor transformations underlying arm movements toward visual targets: a neural network model of cerebral cortical operations. *Journal of Neuroscience*, 12(4):1435–53.
- Casdagli, M. (1989). Non linear prediction of chaotic time series. *Physica D*, 35:335–356.
- Cormack, L. K., Stevenson, B., S., , and Schor, C. M. (1993). Disparity-tuned channels of the human visual system. *Visual Neuroscience*, 10:585–596.
- Cumming, B., Johnston, E., and Parker, A. (1991). Vertical disparities and perception of three-dimensional shape. *Nature*, 349:411–413.
- Dominey, P. and Arbib, M. (1992). A cortico-subcortical model for the generation of spatially accurate sequential saccades. *Cerebral Cortex*, 2:153–175.
- Driver, J. and Halligan, P. (1991). Can visual neglect operate in object-centered coordinates? an affirmative single case study. *Cognitive Neuropsychology*, 8(6):475–496.
- Droulez, J. and Berthoz, A. (1991). A neural model of sensoritopic maps with predictive short-term memory properties. *Proc. Nat. Ac. Sc.*, 88:9653–9657.
- Duhamel, J., Colby, C., and Goldberg, M. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, 255(5040):90–92.
- Farah, M., Brunn, J., Wong, A., Wallace, M., and Carpenter, P. (1990). Frames of reference for allocating attention to space: evidence from the neglect syndrome. *Neuropsychologia*, 28(4):335–47.
- Feldman, J. (1985). Four frames suffice: a provisional model of vision and space. *Behavior and Brain Sciences*, 8:265–289.

- Felleman, D. and Van Essen, D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1:1–47.
- Ferster, D. (1981). A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex. *Journal of Physiology*, 311:623–655.
- Fodor, J. (1983). *The modularity of mind : an essay on faculty psychology*. MIT Press, Cambridge.
- Foley, J. (1980). Binocular distance perception. *Psychological Review*, 87:411–434.
- Foley, J. and Held, R. (1972). Visually directed pointing as a function of target distance, direction and available cues. *Perception and Psychophysics*, 12:263–268.
- Gallese, V., Murata, A., Kaseda, M., Niki, N., and Sakata, H. (1994). Deficit in hand preshaping after muscimol injection in monkey parietal cortex. *Neuroreport*, 5(12):1525–29.
- Galletti, C. and Battaglini, P. (1989). Gaze-dependent visual neurons in area V3a of monkey prestriate cortex. *Journal of Neuroscience*, 9:1112–1125.
- Galletti, C., Battaglini, P., and Fattori, P. (1993). Parietal neurons encoding spatial locations in craniotopic coordinates. *Experimental Brain Research*, 96:221–229.
- Gauthier, G., Nommay, D., and Vercher, J. (1990). The role of ocular muscle proprioception in visual localization of targets. *Science*, 249:58–61.
- Gawne, T. and Richmond, B. (1993). How independent are the messages carried by adjacent inferior temporal cortical neurons? *Journal of Neuroscience*, 13:2758–2771.
- Georgopoulos, A., Lurito, J., Petrides, M., Schwartz, A., and Massey, J. (1989). Mental rotation of the neuronal population vector. *Science*, 243:234–6.
- Georgopoulos, A., Schwartz, A., and Kettner, R. (1986). Neuronal population coding of movement direction. *Science*, 233:1416–1419.

- Ghahramani, Z., Wolpert, D., and Jordan, M. (1995). Computational structure of coordinate transformations: a generalization study. In *Advances in Neural Information Processing Systems*, volume 7. Morgan-Kaufmann, San Mateo, CA.
- Gnadt, J. and Andersen, R. (1988). Memory related motor planning activity in posterior parietal cortex of macaque. *Experimental Brain Research*, 70(1):216–20.
- Gnadt, J., Bracewell, R., and Andersen, R. (1991). Sensorimotor transformation during eye movements to remembered visual targets. *Vision Research*, 31(4):693–715.
- Gnadt, J. and Mays, L. (1991). Depth tuning in area LIP by disparity and accommodative cues. In *Abstract Society for Neuroscience*, volume 17, page 1113.
- Godwin-Austen, R. (1965). A case of visual disorientation. *Journal of Neurology, Neurosurgery and Psychiatry*, 28:453–48.
- Goldberg, M. and Bruce, C. (1990). Primate frontal eye fields. III. maintenance of a spatially accurate saccade signal. *Journal of Neurophysiology*, 64(2):489–508.
- Goodale, M. and Milner, A. (1990). Separate visual pathways for perception and action. *Trends in Neuroscience*, 15:20–25.
- Goodman, S. and Andersen, R. (1989). Microstimulations of a neural network model for visually guided saccades. *J. Cog. Neurosc.*, 1:317–326.
- Goodman, S. and Andersen, R. (1990). Algorithm programmed by a neural model for coordinate transformation. In *International Joint Conference on Neural Networks*, San Diego.
- Graziano, M. and Gross, C. (1993). A bimodal map of space - somatosensory receptive fields in the macaque putamen with corresponding visual receptive fields. *Experimental Brain Research*, 97(1):96–109.

- Groh, J. and Sparks, D. (1993). Somatosensory activity in the superior colliculus (SC) influenced by eye position. In *Society For Neuroscience Abstracts*, page 858, Washington.
- Hallett, P.E.; Lightstone, A. (1976). Saccadic eye movements to flashed targets. *Vision Research*, 16:107–114.
- Heilman, K., Watson, R., and Valenstein, E. (1985). Neglect and related disorders. In Heilman, K. and Valenstein, E., editors, *Clinical Neuropsychology*, pages 243–294. Oxford University Press, New York.
- Holmes, G. (1918). Disturbances of visual orientation. *British Journal of Ophthalmology*, 2:449–68.
- Holmes, G. and Horrax, G. (1919). Disturbances of spatial orientation and visual attention. *Archives in Neurology and Psychiatry*, 1:385–407.
- Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward network are universal approximators. *Neural Networks*, 2:359–366.
- Hubel, D. and Wiesel, T. (1970). Cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Nature*, 225:41–42.
- Irwin, D. (1983). Evidence against visual integration across saccadic eye movements. *Perception and Psychophysics*, 34:49–57.
- Irwin, D. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, 23:420–456.
- Irwin, D., Brown, J., and Sun, J. (1988). Visual masking and visual integration across saccadic eye movements. *Journal of Experimental Psychology*, 117:276–287.
- Irwin, D., Zacks, J., and Brown, J. (1990). Visual memory and the perception of a stable visual environment. *Perception and Psychophysics*, 47:35–46.

- Jay, M. and Sparks, D. (1987). Sensorimotor integration in the primate superior colliculus:1. motor convergence. *Journal of Neurophysiology*, 57:22–34.
- Kalaska, J. and Crammond, D. (1992). Cerebral cortical mechanisms of reaching movements. *Science*, 255(5051):1517–1523.
- Karnath, H., Schenkel, P., and Fischer, B. (1991). Trunk orientation as the determining factor of the 'contralateral' deficit in the neglect syndrome and as the physical anchor of the internal representation of body orientation in space. *Brain*, 114:1997–2014.
- Kawanura, H. and Marchiafava, P. (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.
- Kinsbourne, M. (1987). Mechanisms of unilateral neglect. In Jeannerod, M., editor, *Neurophysiological and neuropsychological aspects of spatial neglect*, pages 69–86. North-Holland.
- Knudsen, E., du Lac, S., and Esterly, S. (1987). Computational maps in the brain. *Annual Review of Neuroscience*, 10:41–65.
- Knudsen, E. and Knudsen, P. (1991). Stretched and upside-down maps of auditory space in the optic tectum of blind-reared owls; acoustic basis and behavioral correlates. *Journal of Neuroscience*, 11(6):1727–1747.
- Kohler, I. (1964). The formation and transformation of the perceptual world. *Psychological Issues*, 3:62–86.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69.
- Kuperstein, M. (1988). Neural model of adaptative hand-eye coordination for single postures. *Science*, 239:1308–1311.

- Kurylo, D. and Skavenski, A. (1991). Eye movements elicited by electrical stimulation of area PG in the monkey. *Journal of Neurophysiology*, 65:1243–1253.
- Ladavas, E. (1987). Is the hemispatial deficit produced by right parietal lobe damage associated with retinal or gravitational coordinates? *Brain*, 110:167–180.
- Lal, R. and Friedlander, M. (1989). Gating of retinal transmission by afferent eye position and movement signals. *Science*, 243:93–96.
- Lapedes, A. and Farber, R. (1988). How neural nets work. In Anderson, D. N., editor, *Neural Information Processing System*, volume 1, pages 442–456. American Institute of Physics, New York.
- LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., and Jackel, L. (1990). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1:540–566.
- Lee, C., Rohrer, W., and Sparks, D. (1988). Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature*, 332(6162):357–60.
- Lehky, S., Pouget, A., and Sejnowski, T. (1990). Neural models of binocular depth perception. *Cold Spring Harbor Symposium on Quantitative Biology*, 55:765–777.
- Lehky, S. and Sejnowski, T. (1990). Neural network model of visual cortex for determining surface curvature from images of shaded surfaces. *Proceedings of the Royal Society of London B*, 240:251–78.
- LeVay, S. and Voigt, T. (1988). Ocular dominance and disparity coding in cat visual cortex. *Visual Neuroscience*, 1:395–414.
- Linsker, R. (1986b). From basic neural network principles to neural architecture: emergence of orientation selective cells. *Proceedings of the National Academy of Science. USA.*, 83:8390–8394.

- Linsker, R. (1986c). From basic neural network principles to neural architecture: emergence of orientation columns. *Proceedings of the National Academy of Science. USA.*, 83:8779–83.
- Linsker, R. (1986a). From basic neural network principles to neural architecture: emergence of spatial opponents cells. *Proceedings of the National Academy of Science. USA.*, 83:7508–12.
- Logothetis, N., Pauls, J., Bulthoff, H., and Poggio, T. (1993). Responses of inferotemporal (IT) neurons to novel wire-objects in monkey trained in an object recognition task. In *Society For Neuroscience Abstracts*, page 27.
- MacKay, D. (1970). Mislocalization of test flashes during saccadic image displacements. *Nature*, 227:731–733.
- Mateef, S. (1978). Saccadic eye movements and localization of visual stimuli. *Perception and Psychophysics*, 24:215–224.
- Martin, L. and Pearce, D. (1965). Visual perception of direction for stimuli flashed during voluntary saccadic eye movements. *Science*, 248:1485–1488.
- Maunsell, J. and van Essen, D. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology*, 49:1148–1167.
- Mayhew, J. (1973). After-effects of movement contingent on direction of gaze. *Vision Res.*, 13:877–880.
- Mayhew, J. and Longuet-Higgins, H. (1982). A computational model of binocular depth perception. *Nature*, 297:376–378.
- Mays, L. and Sparks, D. (1980). Dissociation of visual and saccade-related responses in superior colliculus neurons. *Journal of Neurophysiology*, 43(1):207–32.

- McIlwain, J. (1988). Saccadic eye movements evoked by electrical stimulation of the cat visual cortex. *Visual Neurosci.*, 1:135–143.
- McClurkin, J., Optican, L., Richmond, B., and Gawne, T. (1991). Concurrent processing and complexity of temporally encoded neuronal messages in visual perception. *Science*, 253:675–677.
- Miller, K., Keller, J., and Stryker, M. (1989). Ocular dominance column development: analysis and simulation. *Science*, 245(4918):605–15.
- Mishkin, M., Ungerleider, L., and Macko, K. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neuroscience*, Oct:414–417.
- Montague, P., Dayan, P., Nowlan, S., Pouget, A., and Sejnowski, T. (1993). Using aperiodic reinforcement for directed self-organization during development. In Hanson, S., Cowan, J., and Giles, C., editors, *Advances in Neural Information Processing Systems*, volume 5. Morgan-Kaufmann, San Mateo, CA.
- Moody, J. and Darken, C. (1989). Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1:281–294.
- Moscovitch, M. and Behrmann, M. (1994). Coding of spatial information in the somatosensory system: evidence from patients with neglect following parietal lobe damage. *Journal of Cognitive Neuroscience*, 6(2):151–155.
- Mountcastle, V., Lynch, J., Georgopoulos, A., Sakata, H., and Acuna, C. (1975). Posterior parietal association cortex of the monkey: command functions for operation within extrapersonal space. *Journal of Neurophysiology*, 38:871–908.
- Nikara, T., Bishop, P., and Pettigrew, J. (1968). Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Experimental Brain Research*, 6:353–372.
- Oja, E. (1982). A simplified neuron model as a principal components analyzer. *Journal of Mathematical Biology*, 15:267–273.

- O'Regan, J. (1984). Retinal versus extraretinal influences in flash localization during saccadic eye movements in the presence of visible background. *Perception Psychophysics*, 36:1–14.
- O'Regan, J. and Levy-Schoen, A. (1983). Integrating visual information from successive fixations: does trans-saccadic fusion exist? *Vision Research*, 23:765–768.
- Otto, I., Grandguillaume, P., Boutkhil, L., Burnod, Y., and Guigon, E. (1992). Direct and indirect cooperation between temporal and parietal networks for invariant visual recognition. *Journal of Cognitive Neuroscience*, 4:35–57.
- Pettigrew, J., Nikara, T., and Bishop, P. (1968). Binocular interaction on single units in cat striate cortex: simultaneous stimulation by single moving slit with receptive fields in correspondence. *Exp. Brain Res.*, 6:391–410.
- Poggio, G., Motter, B., Squatrito, S., and Trotter, Y. (1985). Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Research*, 25:397–406.
- Poggio, G. F. and Fischer, B. (1977). Binocular interactions and depth sensitivity in striate and prestriate cortex of behaving rhesus monkeys. *Journal of Neurophysiology*, 40:1392–1405.
- Poggio, G. F., Gonzalez, F., and Krause, F. (1988). Stereoscopic mechanism in monkey visual cortex : binocular correlation and disparity selectivity. *Journal of Neuroscience*, 8:4531–4550.
- Poggio, T. (1990). A theory of how the brain might work. *Cold Spring Harbor Symposium on Quantitative Biology*, 55:899–910.
- Poggio, T. and Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, 343(6255):263–266.
- Poggio, T. and Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, 247:978–982.

- Pollatsek, A., Rayner, K., and 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.
- Posner, M., Walker, J., Friedrich, F., and Rafal, R. (1984). Effects of parietal injury on covert orienting of visual attention. *Journal of Neuroscience*, 4:1863–1877.
- Pouget, A., Fisher, S., and Sejnowski, T. (1993). Egocentric spatial representation in early vision. *Journal of Cognitive Neuroscience*, 5:150–161.
- Pouget, A., Montague, P., Dayan, P., and Sejnowski, T. (1993). A developmental model of map registration in the superior colliculus using predictive hebbian learning. In *Society For Neuroscience Abstracts*, volume 19, page 858.
- Pouget, A. and Sejnowski, T. (1992). A distributed reference frame for egocentric space in the posterior parietal cortex. *Behavioral and Brain Sciences*, 15(4):787.
- Pouget, A. and Sejnowski, T. (1994). A neural model for the cortical representation of egocentric distance. *Cerebral Cortex*, 4:314–329.
- Rayner, K. and 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.
- Richards, W. (1971). Anomalous stereoscopic perception of depth. *Journal of Optical Society of America*, 61:410–414.
- Rizzolatti, G. and Berti, A. (1990). Neglect as a neural representation deficit. *Revue Neurologique*, 146(10):626–634.
- Rogers, B. and Bradshaw, M. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature*, 361:253–255.

- Rumelhart, D., Hinton, G., and Williams, R. (1986). Learning internal representations by error propagation. In Rumelhart, D. E., McClelland, J. L., and the PDP Research Group, editors, *Parallel Distributed Processing*, volume 1, chapter 8, pages 318–362. MIT Press, Cambridge, MA.
- Sakata, H., Shibutani, H., and Tsurugai, K. (1986). Parietal cortical neurons responding to rotary movement of visual stimulus in space. *Experimental Brain Research*, 61:658–663.
- Sanger, T. (1994). Theoretical consideration for the analysis of population coding in motor cortex. *Neural Computation*, 6:29–37.
- Saund, E. (1989). Dimensionality-reduction using connectionist networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(3):304–314.
- Schall, J. (1991). Neuronal activity related to visually guided saccadic eye movements in the supplementary motor area of rhesus monkeys. *Journal of Neurophysiology*, 66(2):530–58.
- Schlag, J. and Schlag-Rey, M. (1990). Colliding saccades may reveal the secret of their marching orders. *Trends in Neuroscience*, 13:410–415.
- Schwartz, A., Kettner, R., and Georgopoulos, A. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space. I. relations between single cell discharge and direction of movement. *Journal of Neuroscience*, 8:2913–27.
- Scott, D. (1992). *Multivariate Density Estimation*. Wiley Interscience.
- Shannon, E. and Weaver, W. (1963). *The mathematical theory of communication*. University of Illinois Press, Urbana and Chicago.
- Shibutani, H., Sakata, H., and Hyvarinen, J. (1986). Saccade and blinking evoked by microstimulation of the posterior parietal association cortex of the monkey. *Experimental Brain Research*, 55:1–8.

- Singer, W. (1993). Synchronization of cortical activity and its putative role in information processing and learning. *Annual Review in Physiology*, 55:349–374.
- Snyder, L., Brotchie, P., and Andersen, R. (1993). World-centered encoding of location in posterior parietal cortex of monkey. In *Society For Neuroscience Abstracts*, volume 19, page 770.
- Soechting, J. and Flanders, M. (1990). Transformation from head-to-shoulder centered representation of target direction in arm movement. *Journal of Cognitive Neuroscience*, 2:32–43.
- Soechting, J. and Flanders, M. (1992). Moving in three-dimensional space: frames of reference, vectors and coordinate systems. *Annual Review in Neuroscience*, 15:167–91.
- Sparks, D. (1991). Sensori-motor integration in the primate superior colliculus. *Seminars in the Neurosciences*, 3:39–50.
- Stein, B. and Meredith, M. (1993). *The merging of the senses*. MIT Press, Cambridge.
- Stein, J. (1992). A distributed reference frame for egocentric space in the posterior parietal cortex. *Behavioral and Brain Sciences*, 15(4):691–700.
- Stevenson, S., Cormack, L., Schor, C., and Tyler, C. (1992). Disparity-tuned mechanisms of human stereopsis. *Vision Research*, 32:1685–1689.
- Stricanne, B., Mazzoni, P., and Andersen, R. (1993). Modulation by the eye position of auditory responses of macaque area LIP in an auditory memory saccade task. In *Society For Neuroscience Abstracts*, page 26, Washington, D.C.
- Sun, J. and Irwin, D. (1987). Retinal masking during pursuit eye movements: Implications for spatiotopic visual persistence. *Journal of Experimental Psychology: Human Perception and Performance*, 13:140–145.

- Tolhurst, D., Movshon, J., and Dean, A. (1982). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, 23:775–785.
- Tononi, G., Sporns, O., and Edelman, G. (1992). Reentry and the problem of integrating multiple cortical areas: simulation of dynamic integration in the visual system. *Cerebral Cortex*, 2(4):310–35.
- Touretzky, D., Redish, A., and Wan, H. (1993). Neural representation of space using sinusoidal arrays. *Neural Computation*, 5:869–884.
- Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S., and Imbert, M. (1992). Modulation of neural stereoscopic processing in primate area v1 by the viewing distance. *Science*, 257:1279–81.
- Trotter, Y., Celebrini, S., Thorpe, S., and M., I. (1991). Modulation of stereoscopic processing in primate visual cortex v1 by the distance of fixation. In *Abstract Society for Neuroscience*, New-Orleans.
- Vogels, R., Spilleers, W., and Orban, G. (1989). The response variability of striate cortical neurons in the behaving monkey. *Experimental Brain Research*, 77:432–436.
- von Hofsten, C. (1976). The role of convergence in space perception. *Vision Research*, 16:193–198.
- von Hofsten, C. (1977). Binocular convergence as a determinant of reaching behavior in infancy. *Perception*, 6:139–144.
- von Hofsten, C. (1979). Recalibration of the convergence system. *Perception*, 8:37–42.
- Westheimer, G. and Tanzman, I. (1956). Qualitative depth localization with diplopic images. *Journal of Optical Society of America*, 46:116–117.
- Weyand, T. and Malpeli, J. (1989). Responses of neurons in primary visual cortex are influenced by eye position. In *Abs. Soc. Neurosci.*, volume 15.

- Weyand, T. and Malpeli, J. (1993). Responses of neurons in primary visual cortex are modulated by eye position. *Journal of Neurophysiology*, 69(6):2258–60.
- Widrow, B. and Hoff, M. (1960). Adaptative switching circuits. *IRE WESCON Convention Rec.*, 4:96–104.
- Widrow, B. and Stearns, S. (1985). *Adaptive signal filtering*. Prentice Hall, Inc., Englewoods Cliffs, N.J.
- Withington, D., Binns, K., and Keating, M. (1990). The maturation of the superior colliculus map of the auditory space in the guinea pig is disrupted by developmental visual deprivation. *European Journal of Neuroscience*, 2:693–703.
- Zemel, R. S. and Hinton, G. E. (1994). Developing population codes by minimizing description length. In *Advances in Neural Information Processing Systems 6*, pages 11–18, San Mateo, CA. Morgan Kaufmann.
- Zipser, D. and Andersen, R. (1988a). A back-propagation programmed network that stimulates reponse properties of a subset of posterior parietal neurons. *Nature*, 331:679–684.
- Zipser, D. and Andersen, R. (1988b). The role of the teacher in learning-based models of parietal area 7a. *Brain Research Bulletin*, 21:505–512.
- Zohary, E., Shadlen, M., and Newsome, W. (1992). Correlated activity of neuron in area MT. *Abstract Society for Neuroscience*, 18:1101.