



## Intermediate Visual Representations for Attentive Recognition Systems

Antonio J. Rodriguez-Sanchez

Technical Report CSE-2010-06

September 1 2010

Department of Computer Science and Engineering  
4700 Keele Street, Toronto, Ontario M3J 1P3 Canada

INTERMEDIATE VISUAL REPRESENTATIONS  
FOR  
ATTENTIVE RECOGNITION SYSTEMS

Antonio J. Rodríguez-Sánchez

A dissertation submitted to the Faculty of Graduate Studies  
in partial fulfilment of the requirements  
for the degree of

Doctor of Philosophy

Graduate Programme in Computer Science  
York University  
Toronto, Ontario  
July 2010



*A mi madre*

**Intermediate Visual Representations**

**for Attentive Recognition Systems**

by **Antonio J. Rodriguez-Sanchez**

a dissertation submitted to the Faculty of Graduate Studies of York University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

© 2010

Permission has been granted to: a) YORK UNIVERSITY LIBRARIES to lend or sell copies of this dissertation in paper, microform or electronic formats, and b) LIBRARY AND ARCHIVES CANADA to reproduce, lend, distribute, or sell copies of this dissertation anywhere in the world in microform, paper or electronic formats and to authorize or procure the reproduction, loan, distribution or sale of copies of this dissertation anywhere in the world in microform, paper or electronic formats. The author reserves other publication rights, and neither the dissertation nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

# INTERMEDIATE VISUAL REPRESENTATIONS

## FOR ATTENTIVE RECOGNITION SYSTEMS

by **Antonio J. Rodriguez-Sanchez**

By virtue of submitting this document electronically, the author certifies that this is a true electronic equivalent of the copy of the dissertation approved by York University for the award of the degree. No alteration of the content has occurred and if there are any minor variations in formatting, they are as a result of the conversion to Adobe Acrobat format (or similar software application).

Examination Committee Members:

1. John K. Tsotsos
2. Michael Jenkin
3. Robert S. Allison
4. Mazyar Fallah
5. Nick Cercone
6. Kaleem Siddiqi

## Abstract

Computational models of visual processes are of interest in fields such as cybernetics, robotics, computer vision and others. This thesis provides an analysis of a model of attention and of intermediate representation layers in the visual cortex that have direct impact on the next generation of object recognition strategies in computer vision. Biological inspiration - and even biological realism - is currently of great interest in the computer vision community. This thesis includes three major pieces, explained next.

First, I believe that visual attention is a requirement to perform non-detection object recognition tasks. In order to test this hypothesis we compare the Selective Tuning model of attention [Tsotsos et al., 1995] to studies from psychophysics in visual search tasks involving color and 2D shapes. Second, I define a biologically plausible model of Shape Representation which incorporates intermediate layers of visual representation that have not previously been fully explored. I hy-

pothesize that endstopping and curvature cells are of great importance for shape selectivity and show how their combination can lead to shape selective neurons. This Shape Representation model provides a highly accurate fit with neural data from [Pasupathy and Connor, 2001, Pasupathy and Connor, 2002]. Finally, in the same way curvature parts may be configured into shapes, spatial gradients of velocity vectors may be related to optic flow in a hierarchical representation of visual motion analysis. For my last contribution I provide psychophysical evidence of the role of spatial gradients of velocity in optical flow perception as well as neurophysiological evidence for neurons tuned for such gradients.

Following previous authors such as [Zucker, 1981] and [Marr, 1982], I have shown that deeper understanding of visual processes in humans and non-human primates can lead to important advancements in computational perception theories and systems.

## Acknowledgements

I would like to thank all those who have made the present work possible. In particular, my wife Margarita for her unconditional support and encouraging character, my family and her family for their help. Specially my beloved mother Encarna, who I will always have in my heart and mind, my father Cristóbal, my sisters Reme and Olga and my wife's parents Juan and Margarita that came from Spain whenever we needed help. I wish to thank Prof. John K. Tsotsos and his wife Pat for always being there for me and my wife, we will never forget all their help through difficult times.

Finally, a special thank you to Profs. Anitha Pauspathy, Scott Brincat and Ed Connor for providing all the stimuli and data we needed for the work in chapter 3 of this Thesis. Also, thanks to Prof. Julio Martínez-Trujillo for his teachings in neuroscience and the work for chapter 4, and Profs. Alan Dobbins, Michael Jenkin and Rob Allison for their helpful comments.

# Table of Contents

<b>Abstract</b>	<b>iv</b>
<b>Acknowledgements</b>	<b>vi</b>
<b>Table of Contents</b>	<b>vii</b>
<b>List of Figures</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Architecture of the Visual Cortex . . . . .	4
1.2 Attention . . . . .	19
1.3 Shape in Computer Vision . . . . .	30
1.4 Biologically inspired models of Object Recognition . . . . .	36
1.5 Original Contributions . . . . .	41
<b>2 Attention and Visual Search</b>	<b>46</b>

2.1	Introduction . . . . .	47
2.2	A Simple Stimulus Recognition Model . . . . .	50
2.3	Efficiency in Visual Search . . . . .	67
2.4	Discussion . . . . .	75
<b>3</b>	<b>A Model of 2D Shape</b>	<b>77</b>
3.1	A Model of Shape Representation . . . . .	84
3.2	Experiments . . . . .	117
3.3	Discussion . . . . .	147
3.4	Conclusions . . . . .	150
3.5	Appendix . . . . .	152
<b>4</b>	<b>Discrimination of spiral angles in optical flow patterns</b>	<b>154</b>
4.1	Introduction . . . . .	155
4.2	Selectivity for spatial gradient of velocities . . . . .	158
4.3	Spiral discrimination in humans and single neurons . . . . .	167
4.4	Results . . . . .	176
4.5	Discussion . . . . .	184
4.6	Conclusion . . . . .	191
<b>5</b>	<b>Discussion and Future Work</b>	<b>192</b>

5.1 Future work . . . . .	194
<b>Bibliography</b>	<b>201</b>

## List of Figures

1.1	a) Primate Visual Cortex b) Simplified version of the Visual Cortex. Modified from [Felleman and Van Essen, 1991] . . . . .	4
1.2	Columnar organization in IT [Fujita, 2002] . . . . .	14
1.3	a) Feature search. In feature search the target is defined by a single feature (color). b) Conjunction search. In conjunction search the target is the conjunction of two features shared with the distractor (orientation and color in this case) . . . . .	26
2.1	Visual search time course via Selective Tuning . . . . .	51
2.2	The Selective Tuning model of attention . . . . .	53
2.3	Architecture of shape pathway . . . . .	55
2.4	Shape analysis on target stimulus. See text for explanation . . . . .	59
2.5	Analysis of a scene (bottom): find the square. . . . .	64

2.6	Example of silhouette recognition. insided dotted boxes are IT with highest responses corresponding to searched stimulus (next to it). . . . .	66
2.7	Visual Search Results. Example where the target (green circled) and distractors have small color differences for 5 (a), 10 (b), 20 (c) and 25 (d) distractors. Circled in red are distractors attended previous to finding the target, which is under the green circle. . .	70
2.8	a) Number of covert fixations as a function of set size from Selective Tuning. Blank dots are experiment trial results. Filled dots correspond to average from 3 trials. b) Results from [Nagy and Sanchez, 1990]	72
2.9	a) Conjunction search with 18 distractors, target is the red circle, found at the 4th covert fixation. b) Inefficient search: Find the rotated T among 21 Ls, 14 fixations were needed to find the T. . .	73
2.10	a) The number of shifts of attention as a function of set size for feature search (light gray), conjunction search (gray) and inefficient search (black). b) Visual search continuum from [Wolfe, 1998a] . .	75
3.1	Architecture of the Shape Representation model . . . . .	85
3.2	Sample of Difference of Gaussians edge detectors. Four sizes, from left to right: 40, 60, 88 and 120 pixels . . . . .	87

3.3	Comparison between a model simple neuron response (blue plot) and a model complex neuron response (green plot) for orientation selection (a) and location (b) . . . . .	89
3.4	Model endstopped cell with complex components. . . . .	92
3.5	Sigmoid-type Rectification function $\Phi$ . . . . .	98
3.6	Response of the model endstopped cells with complex components to curvatures at preferred (top) and all (bottom) orientations. Simple cell sizes were a) 40, b) 80, c) 100 and d) 120 pixels. $\sigma_y=(10,20,25,30)$ . $AR$ (aspect ratio)=(1.15,2,3,4). $WR$ (width ratio)=2.5 for all cells. Gain $c=(0.7,0.8,1,2)$ . Responses were normalized for the range [0,1] . . . . .	100
3.7	Selectivity of endstopped cells with complex components to curvatures considering 90% of maximum response (blue range) and 80% of maximum response (green range). Highest responses are in red circles. $AR$ (aspect ratio)=(0.7,1.4,2.15,3). $WR$ (width ratio)=2.5 for all cells. Gain $c=(1.5,1.25,1,3)$ . . . . .	102

3.8	Direction selection. Endstopped configuration of cells whose curvature is moving downwards (a), to the right (b), left (c) and upwards (d), e) An example of a circle where only a and d are applied, positive sign selects model endstopped neurons whose direction is downward (blue). Negative sign is curvature upwards (red) . . . . .	104
3.9	a) Shape-selective neurons respond to different curvatures at different positions. The response is maximal when those curvatures are present at their selective positions (red). If they are in nearby positions the neuron provides some response as well (orange and yellow) b) The curvatures at those positions are organized into hypercolumns. . . . .	108
3.10	The hierarchy of 2D shape processing. See text for explanation . .	115
3.11	Shape-selective neuron tuning profile for location and curvature .	116
3.12	a) Pasupathy and Connor's Gaussian shape tuning function describing the response pattern of a neuron b) Pasupathy and Connor's proposal, the object's white line shows the different values of curvature relative to angular position for a squashed raindrop stimulus (icon at the center). . . . .	119

3.13	a) Response to curvature from the model endstopping to the squashed raindrop stimulus (Figure 3.12b), dark red corresponds to very high curvature, red to medium-high curvature, blue to medium curvature and yellow to broad curvature b) Population response from Pasupathy and Connor (colored surface), curvature $\times$ posi- tion veridical function (white line) and the model's function (gray line), arrows mark the difference between the veridical function and our model. . . . .	124
3.14	For every stimulus (left image), Pasupathy and Connor's fit (white plot, center) and neural responses (colored values, center) was com- pared to our model's response (right). . . . .	126
3.15	Euclidean distance from the model to Pasupathy and Connor's data.	129
3.16	Stimuli used in [Pasupathy and Connor, 2001] (Copyright © 2001 The American Physiological Society) . . . . .	130
3.17	How the features for <i>isolating</i> a Shape neuron are obtained for our second set of experiments. See text. . . . .	133

3.18 Comparison to Figure 2 of [Pasupathy and Connor, 2001] (Copyright © 2001 The American Physiological Society). Cells responses are on columns a) and c) and their respective model responses are on columns b) and d). Green shading is for stimuli with high response or stimuli that Pasupathy and Connor considered of interest . . . . .	138
3.19 Comparison to Figure 4 of [Pasupathy and Connor, 2001] (Copyright © 2001 The American Physiological Society). Cells responses are on columns a) and c) and their respective model responses are on columns b) and d). Green shading is for stimuli with high response or stimuli that Pasupathy and Connor considered of interest . . . . .	141
3.20 Comparison to Figure 5 of [Pasupathy and Connor, 2001] (Copyright © 2001 The American Physiological Society). Cells responses are on columns a) and c) and their respective model responses are on columns b) and d). Green shading is for stimuli with high response or stimuli that Pasupathy and Connor considered of interest . . . . .	143

3.21	Comparison to Figure 8 of [Pasupathy and Connor, 2001] (Copyright © 2001 The American Physiological Society). Cells responses are on columns a) and c) and their respective model responses are on columns b) and d). Green shading is for stimuli with high response or stimuli that Pasupathy and Connor considered of interest . . . . .	144
3.22	Difference between the model's Shape-selective neurons and 75 real cells responses from area V4 . . . . .	146
4.1	Spiral space is a coordinate system that interprets expansion ( $0^\circ$ ), contraction ( $180^\circ$ ) and rotations (clockwise: $90^\circ$ , counterclockwise: $270^\circ$ ) as cardinal directions with in- and outward spiraling movement patterns placed in between. Gray arrows show local motion direction. . . . .	156
4.2	Trials consisted on the presentation of an adapting pattern lasting 20 sec followed by a test pattern lasting 0.5 sec. Variation of the adapting pattern gave four different conditions (staircase method): a) Moving and accelerating upwards, b) Moving rightwards and accelerating upwards, c) Moving and accelerating rightwards, and d) Moving upwards and accelerating rightwards. . . . .	160

4.3	Example results from one subject. . . . .	162
4.4	a) Mean and SD values averaged across the eight subjects. b) Ratio $PSE_{SG}/PSE_{control}$ (ordinate) as a function of the adapting stimulus type (abscissa). . . . .	164
4.5	Averaged psychometric functions (upper row, clockwise/counterclockwise rotation; lower row, expansion/contraction) for human subjects. .	175
4.6	Comparison between expansion/contraction and rotations a) Spi- ral discrimination thresholds for every subject b) Point of Subjec- tive Equality corresponding to each of the subjects. . . . .	177
4.7	Gaussian fit for a spiral selective neuron (a) and a neuron tuned to linear motion (b) . . . . .	178
4.8	a) Neuron tuned for spirals b) Neuron tuned for linear motion. . .	181
4.9	Distribution of the population of neurons based on their discrimi- nation thresholds (blue histograms). a, b) Neurons tuned for spi- rals. c) and d) Neurons tuned for linear motion. Human threshold for this kind of motion is $1.1^\circ$ . [Snowden et al., 1992]. Gray his- tograms correspond to the same neurons but after changes in the width (a, c), and height (b, d) (see main text). . . . .	183

4.10	(a) Tuning curve of a medial superior temporal neuron to spiral motion. The abscissa represents spiral direction and the ordinate the response (spikes/s). Error bars represent standard errors. (b) Thresholds, and (c) Neurometric functions (symbols). (d) Minimum threshold (ordinate) of each neuron (black), and for each participant (gray). For the cells, the data are grouped according to the preferred direction of the neuron in spiral space (symbols in abscissa). For the human participants, the data were divided according to the discrimination task. . . . .	186
5.1	Stimuli used in [Brincat and Connor, 2004] . . . . .	197

# 1 Introduction

One of the main difficulties that arises when designing automatic vision systems is developing a mechanism that can recognize, or simply find, an object with the ease of the human visual system. Humans can recognize objects effortlessly under variations in location, lighting and viewpoint. The brain appears to perform a wide variety of complex tasks by means of simple operations. These seemingly simple operations are applied to several layers of neurons, the layers representing increasingly complex, abstract intermediate processing stages. These intermediate representations have not been adequately studied in computer vision systems. Among these, shapes are important object descriptors. Recent experiments have shown that medial areas such as V2, V4 and TEO (Figure 1.1) are involved in analyzing shape features such as corners [Ito and Komatsu, 2004, Boynton and Hegde, 2004] and curvatures [Pasupathy and Connor, 1999, Pasupathy and Connor, 2001,

Pasupathy and Connor, 2002, Brincat and Connor, 2004]. By studying the processes that occur in the human brain we may be able to construct a model that simulates this behaviour.

An important issue for how the human visual system works is attention. The Encyclopedia Britannica defines attention as “in psychology, the concentration of awareness on some phenomenon to the exclusion of other stimuli”. Due to the capacity limitations of the brain, not all the visual information that impinges our retinas can be processed [Tsotsos, 1990]. Attention is applied to this information in order for the visual system to focus on processing salient information, while filtering or inhibiting other parts of the visual scene. A huge amount of psychophysical data has been collected over the years in which the task is to find an element defined by different characteristics among distractors. Many theories have emerged as well to explain the results obtained from these experiments.

Secondly, it is also important to know how that information, the focus of our attention, is analyzed. A hierarchy of neuronal layers exists in the visual cortex, each layer containing neurons that apply different kinds of processing to their input, depending on the layer they belong to. The receptive field of these neurons and the complexity of the processing is largely increased at higher layers in this hierarchy. We still do not know the specifics of how such a system

works, but many research studies in the last decades have shed some light on the behavior of neurons in the visual cortex. Attention is found in almost every layer of the visual cortex, but V4 is the classical area to study the effects of attention [Moran and Desimone, 1985].

A main component of this thesis is concerned with one of those aspects important for object recognition and that the brain performs with astonishing effectiveness, the analysis of shapes. In addition to this, we studied how the Selective Tuning [Tsotsos et al., 1995] top-down model of attention may explain visual search tasks that consider shapes and color.

In addition to analyzing objects, another important source of information for our survival is motion. In the same way that oriented lines can be used to build curvature and shape, we study how simple translational motion builds higher order concepts such as rotation or expansion. Spatial gradients of velocity are an important element. Another aspect of this thesis is the plausibility of selectivity for such gradients of velocity in area MST.

A brief description of the visual cortex will be provided in section 1.1. Then, a review regarding visual search and attention is presented in section 1.2. Section 1.3 summarizes the use of shape and curvature in the computer vision literature. Section 1.4 reviews biological inspired computational models of object recogni-

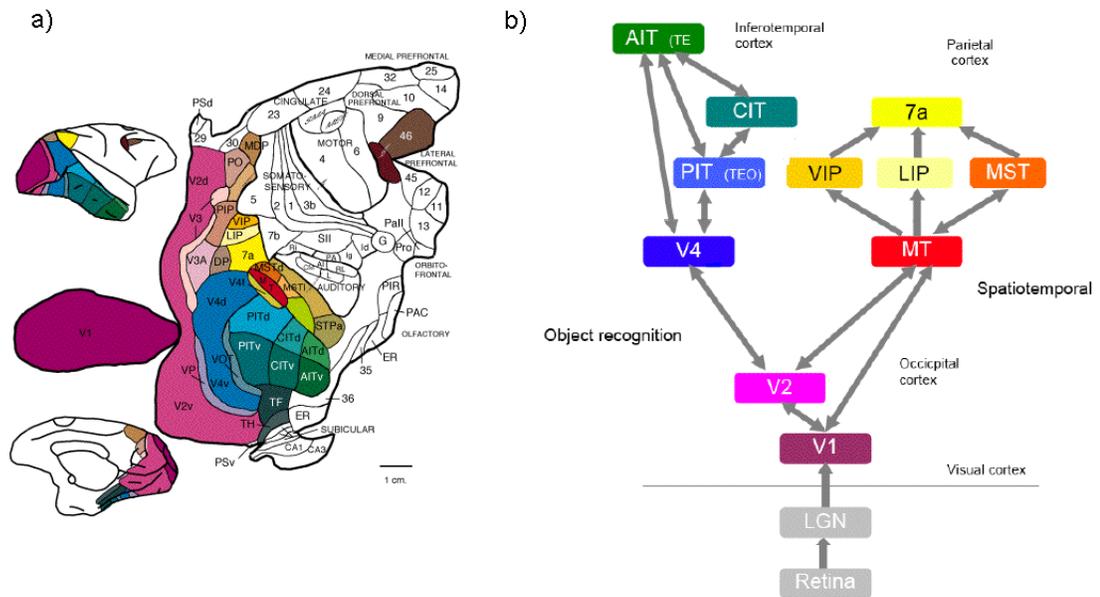


Figure 1.1: a) Primate Visual Cortex b) Simplified version of the Visual Cortex. Modified from [Felleman and Van Essen, 1991]

tion. Finally, section 1.5 summarizes the contributions from this thesis.

## 1.1 Architecture of the Visual Cortex

Since the foundation of modern neuroanatomy by Ramón y Cajal [Ramon-Cajal, 1888, Ramon-Cajal, 1894, Ramon-Cajal, 1904, Jones, 2007], who gave a detailed description of nerve cell organization in the central and peripheral nervous system, great progress has been achieved in understanding the human brain.

The macaque monkey visual cortex occupies 55% of the neocortex (compared with the 11%, 8% and 3% of somatosensory, motor and auditory areas respectively) [Kandel et al., 2000]. There are at least 32 neocortical areas involved in vision [Felleman and Van Essen, 1991] (Figure 1.1a), and inputs can come from visual, auditory, somatosensory, or visuomotor activity. [Felleman and Van Essen, 1991] identified 305 pathways, of which 242 have bidirectional connections, although they can vary in strength (e.g. connection V1-V4 is robust, but V4-V1 is weak). Area connections are organized hierarchically with upwards, downwards and lateral connections.

The visual cortex is organized into different areas (Figure 1.1b). V1 and V2 are the largest, each having an area of approximately 1100 - 1200 mm<sup>2</sup> (11-12% of the macaque neocortex) [Felleman and Van Essen, 1991]. Physiological studies show two different pathways with connections between them: The occipitotemporal pathway (V1, V2, V4, PIT and AIT) is related with object recognition features (color, shape, etc.) [Logothetis and Sheinberg, 1996, Tanaka, 1996], while the occipitoparietal pathway (V1, V2, V3, MT and MST) is associated with spatiotemporal characteristics of the scene (direction of motion, speed gradients, etc.) [Webster and Ungerleider, 1998].

Along this hierarchical architecture, neurons become increasingly selective to

more complex stimuli and less sensitive to stimulus variation as will be explained next. At the bottom of the hierarchy, neurons in V1 are selective for edges (among other features), and at the top, AIT neurons respond to complex objects with significant variation in their orientation, size, illumination and foreshortening. How is early visual information transformed into whole objects? A summary of the most influential work in neurophysiology regarding this aspect is provided here.

### 1.1.1 Object recognition pathway

#### Visual Area 1 (V1)

Neurons in area V1 of the cat respond when bars and edges are present in their receptive field [Hubel and Wiesel, 1959, Hubel and Wiesel, 1962, Hubel and Wiesel, 1965]. This same characteristic was later found in monkeys [Hubel and Wiesel, 1968]. V1 neurons are also selective to other features such as spatial frequency [Valois and De Valois, 1990].

Cytochrome oxidase reveals six layers in V1 and two regions with rich (*blob*) and poor (*interblob*) concentrations of that component, and these regions span several layers. Among these layers, Layer 4B is associated with stereopsis and motion (magnocellular pathway), the blob regions analyze color (parvocellular

neurons) and the interblob region neurons are specialized for lines, some cells are end-stopped (parvocellular pathway) [Livingstone and Hubel, 1988]. In area V1 (as in all retinotopic areas), the retinal fovea has a much larger area of representation than the periphery. There is also ocular dominance [LeVay et al., 1978], which means that there are neurons whose response is dominated either by input from the left eye or the right eye.

Concerning orientation selection, V1 neurons can be classified into three types: simple cells, complex cells, and endstopped cells [Hubel and Wiesel, 1959, Hubel and Wiesel, 1962, Hubel and Wiesel, 1968].

- Simple neurons have small receptive fields ( $0.1^\circ$ - $1^\circ$  [Hubel and Wiesel, 1965, Schiller et al., 1976]), are close to the fovea, and their response is based on small areas relative to the background [Hubel and Wiesel, 1968]. They respond to bars and edges with different orientations as well as to spatial frequency [DeValois et al., 1979]. Their behavior may be modeled by Gabor filters [Marcelja, 1980, Rolls and Deco, 2002] or Difference of Gaussians [Hawken and Parker, 1987].
- Complex neurons are also sensitive to bars and orientations, but with a lower sensitivity [Hubel and Wiesel, 1959, Hubel and Wiesel, 1965, Hubel and Wiesel, 1968]. Their receptive fields are larger than those of

simple neurons ( $0.2^\circ$ - $2^\circ$  [DeValois et al., 1982]). Like simple cells, complex cells are selective for bars presented at a preferred orientation and tuned for spatial frequency [Hubel and Wiesel, 1968]. In contrast to simple cells, a complex neuron responds irrespective of the particular position at which a bar is flashed inside its receptive field [Hubel and Wiesel, 1968].

- Endstopped neurons are sensitive to the termination of an edge or a bar [Hubel and Wiesel, 1965, Hubel and Wiesel, 1968]. Endstopped neurons are also known as hypercomplex and have been described in great detail for the cat [Kato et al., 1978, Orban et al., 1979b, Orban et al., 1979a, Bishop et al., 1980].

[Lee et al., 1998] propose that V1 is a high resolution available to cortex for calculations which requires high resolution image details and spatial precision. V1 is arranged into hypercolumns. Columns are tuned to different orientations in one dimension and alternating ocular dominance in a second dimension and run perpendicular to the surface of the cortex. One third of V1 neurons can be activated by the magnocellular or parvocellular pathways alone. V1 projects mainly to V2, but also to area MT and V3 [Lennie, 1998]. Columnar units are linked through horizontal connections [Kandel et al., 2000].

## Visual area 2 (V2)

V2 receives its input mainly from V1 [Felleman and Van Essen, 1991]. Color, form and stereopsis/motion V1 pathways continue into V2 [Livingstone and Hubel, 1988]:

- V1 blobs project into V2 thin stripes (color), they are not orientation selective and half of them are color sensitive with centre-surround antagonism. Their receptive fields are larger than their corresponding V1 color pathway neurons.
- Interblob regions follow into V2 interstripes (orientation and end-stopping), they have selectivity to orientation but not to direction and half of them respond to terminations of edges or bars (a much larger number of end-stopped neurons compared to V1) .
- Layer 4B of area V1 projects into V2 thick stripes (stereopsis substrate). They are orientation selective and respond occasionally to bars or edge terminations. They respond poorly when only one eye is stimulated, but strongly when there is information coming from both eyes. Stereopsis neurons are sensitive to large disparities [Poggio and Fischer, 1977], and for this reason, these neurons seem to be mainly selective for stereoscopic depth and

motion.

V2 neurons respond to contours, both real and illusory [von der Heydt et al., 1984]. More recent studies [Ito and Komatsu, 2004, Boynton and Hegde, 2004], have found that V2 neurons are mainly selective for angles and corners, and that these neurons also showed submaximal responses for bars.

### **Visual area 4 (V4)**

V4 neurons have receptive fields that range from 2° to 4° [Felleman and Van Essen, 1991]. Some early studies suggested that V4 is the centre of color processing due to the large number of neurons selective to color [Essen and Zeki, 1978]. Later studies suggested that this number may not be so high [Schein et al., 1982]. Some of the qualities of V4 neurons may be that they show selectivity to luminance [Schein and Desimone, 1990, Heywood et al., 1992, Motter, 1994] and color constancy [Zeki, 1983]. They also manifest sensitivity to length, width, orientation, direction of motion and spatial frequency [Desimone et al., 1985].

Later experiments in monkeys where area V4 was ablated showed that V4 is important for the perception of form and pattern/shape discrimination [Merigan and Pham, 1998]. Neurons in V4 respond to

those simple shapes and their responses can be fit by a curvature-position function [Pasupathy and Connor, 1999, Pasupathy and Connor, 2001, Pasupathy and Connor, 2002]. In this representation, the object's curvature is attached to a certain angular position relative to the object center of mass. Most V4 neurons represent individual parts or contour fragments [Connor et al., 2007].

Area V4 has also been studied with glass patterns [Wilson et al., 1997, Wilson and Wilkinson, 1998]. Glass patterns are random dot patterns in which the orientation of each dot pair is tangent to the contours of a global pattern. Subjects were more sensitive to concentric glass patterns than to radial, hyperbolic or parallel ones, meaning that neurons may be organized in a concentric way. These results have been later supported by a functional magnetic resonance imaging (fMRI) experiment [Wilkinson et al., 2000].

### **Inferotemporal Cortex (IT)**

The macaque monkey's inferotemporal cortex (IT) receives inputs mainly from V2 and V4 [Felleman and Van Essen, 1991]. IT is the main area involved in object recognition and discrimination [Gross et al., 1972, Tanaka, 1996, Logothetis and Sheinberg, 1996, Brincat and Connor, 2004]. As with V4, the first studies supporting this claim were based on research on monkeys in which

area IT was ablated [Dean, 1976, Gross, 1992]. Ablation resulted in deficits on tasks that involved visual discrimination or object recognition. IT neurons are view-independent, translation, space and size invariant and respond mainly to objects and faces [Rolls and Deco, 2002].

There is more evidence that IT is involved in object recognition. IT is divided in two main parts [Tanaka et al., 1991]:

- In Posterior IT (PIT or TEO) most of the neurons are activated maximally by a simple combination of features such as bars or disks varying in size, orientation and color [Tanaka et al., 1991]. These neurons are called primary cells. Although in a smaller quantity, another type of cells called elaborate cells are present in TEO that respond to more complex features. TEO is responsible for medium complexity features and integrate information about the shapes and relative positions of multiple contour elements. Recent experiments [Brincat and Connor, 2004] support parts-based shape theories. TEO neuron receptive fields are around  $4^\circ$  average for primary cells and  $10^\circ$  for elaborate cells [Tanaka et al., 1991].
- Anterior IT (AIT or TE) neurons comprise two-thirds to three-quarters of the IT area and they require more complex features for maximal activation. These neurons are called elaborate cells. TE is responsible for high

complexity features [Tanaka, 1996, Kobatake et al., 1998] including faces, hands and other body parts. TE neurons have larger receptive fields (average of around  $13^\circ$ ) [Tanaka et al., 1991]. TE receives inputs from V4 and TEO neurons at different retinal positions [Tanaka, 1996], which may explain its scale, position and view invariance [Booth and Rolls, 1998].

TE neurons not only respond to fairly complex objects, but also have a columnar organization [Wang et al., 1998, Tsunoda et al., 2001, Kiani et al., 2007] (Figure 1.2) in a similar way as areas V1 [Fujita et al., 1992] and MT [Albright, 1984]. TE contains 1,300-2,000 columns [Fujita, 2002]. A column may encode a feature common to similar shapes and the activity of individual neurons may account for small differences between those objects. Related features overlap creating a continuum of features [Tanaka, 1996]. This columnar organization has several properties [Wang et al., 1998]. First, single features activate different groups of neurons. Secondly, different parts have different selectivities, some regions are activated by only one stimulus, while other regions are activated by more than three stimuli. Due to the columnar organization of TE, many TE neurons in a column may represent a complex feature and slight changes in features may be the result of differences of neurons with similar but subtle changes in selectivity in those columns [Tanaka, 1996].



Figure 1.2: Columnar organization in IT [Fujita, 2002]

By recording IT neurons from anesthetized monkeys we know that primary neurons are selective for orientation, bars, disks and texture; surprisingly color was not a very relevant feature [Tanaka et al., 1991]. Elaborate cells only respond to different shapes (e.g. stars, combination of a disk and a bar, T shape, rounded tongue, etc.) or to a combination of different features (shape and texture, shape and color, texture and color, texture and color and shape) [Tanaka et al., 1991].

Among elaborate cells are neurons that respond only to faces or hands [Gross et al., 1972, Logothetis and Sheinberg, 1996]. These elaborate neurons are also sensitive to the orientation and size of the stimulus. An important observation about elaborate cells is that objects were coded by combinations of active neurons, each representing a particular feature [Tanaka, 1996]. But a single neuron was enough for a face or a hand; this type of neuron did not respond to

other objects and had different tunings for different faces. Face neurons encode different faces in a distributed way [Rolls, 1987]. Faces in different orientations activate partially overlapping regions, and these regions are arranged in a continuous map of the view angle along the cortical surface. Face recognition was composed of one component common to all the orientations of the face and a second component that depends on the view of the face.

### **Time-course of object recognition**

Even though the analysis of visual objects may seem instantaneous, in reality, traversing the object recognition pathway requires time. Detection and categorization seem to take about 150ms [Thorpe et al., 1996]. This amount of time is the time required for a single feed-forward pass through the visual system [Bullier, 2001] and allows us to know if an object is present or not on a first view as long as stimuli are well separated, are simple and can be easily segregated from the background. In the visual search literature, this is known as *pop-out*, in which response time is the same regardless of the number of distractors [Treisman and Gelade, 1980].

Identification consists of determining aspects such as size, color and shape. This task needs extra time (65ms, [Grill-Spector and Kanwisher, 2005,

Evans and Treisman, 2005]). If we are required to localize the object (through a saccade or pointing), it takes 100ms to 300ms extra time [Becker, 1991]. If the task is under conditions of clutter or difficult conjunctions, it will take even longer. For within-category-identification and localization, a top-down traversal must be completed. Identification information is available at intermediate layers in the hierarchy. But localization details are available in earlier layers and requires a deeper traversal of the hierarchy to reach those layers [Mehta et al., 2000, Rothenstein et al., 2008].

### **1.1.2 Motion analysis pathway**

Neurons in areas V1 and V2 are not only involved in object recognition but also in motion analysis (Figure 1.1). There are two main paths: magnocellular and parvocellular pathways [Hubel and Wiesel, 1968]. An oversimplified interpretation of the magnocellular and parvocellular pathways is that the former is involved in the perception of motion while the second one is involved in color and edges. In area V1 cells respond to motion in one direction, but motion in the opposite direction has no effect on them [Kandel et al., 2000]. Layer 4B in V1 and the pale stripes in V2 transmit the information regarding motion to upper layers, such as MT and MST which are described next.

### **Middle Temporal Area (MT or V5)**

MT receives inputs from visual cortical areas V1, V2, and dorsal V3 [Felleman and Van Essen, 1991, Movshon and Newsome, 1996]. Electrophysiological studies in area MT showed that a large portion of the cells were tuned to the speed and direction of moving visual stimuli [Dubner and Zeki, 1971, Maunsell and Van Essen, 1983]. Cortical lesions in area MT impair the perception of dots moving in the same direction [Newsome and Paré, 1988], damage to area MT then reduces the ability to detect motion. MT computes motion direction for complex patterns [Movshon et al., 1989] and responds to the actual direction of movement, solving the aperture problem [Pack and Born, 2001]. Cells with similar direction selectivity are organized into columns [Kandel et al., 2000].

### **Medial Superior Temporal Area (MST)**

Studies in primates have found neurons in the dorsal division of the medial superior temporal area that are tuned for spiral direction. MST neurons are not only tuned to linear motion, but are also tuned to radial or circular motion [Tanaka and Saito, 1989, Graziano et al., 1994]. These neurons play an important role in optic flow perception since they are activated by complex configurations of motion [Britten and van Wezel, 1998]. They can encode expanding

and contracting spirals [Heuer and Britten, 2004]. In humans, imaging studies have isolated a region adjacent to the middle temporal area that is selectively activated by spirals [Morrone et al., 1995, Martinez-Trujillo et al., 2005, Holliday and Meese, 2008], and where lesions produce deficits in complex motion perception [Vaina, 1998], which would be the equivalent to the behavior of monkey area MST.

## **Summary**

Along the object recognition pathway, neurons become increasingly selective to more complex stimuli and more invariant to location, rotation and scale. At the bottom of the hierarchy, neurons in V1 are selective for edges, and at the top, AIT neurons respond to complex objects at any orientation. How are edges and other features combined into whole objects? This is what many studies in neuroscience have been trying to explain, and a summary of the most influential work has been provided here. IT cells are not selective to random features, instead, the extracted features are the ones that are useful for the purposes of the monkey's behavior [Tanaka, 1996].

In the motion pathway, the early neurons respond to direction of motion (V1), then to motion regions (MT) and then to optic flow (MST). The same pattern

in both paths is seen for analyzing the visual world:

1. Neuron receptive fields become larger, more specific and more abstract in higher order areas of the hierarchy. Neurons from one area mainly feed into the area immediately above it, although there are also back-connections and connections to other areas.
2. Neurons within one area are usually organized into columns or hyper-columns, and this structure is present in both pathways, object recognition and motion analysis.

It is important to note, as a subsequent section will show, that computational models have to this point only exploited the earliest layers of such processes.

## **1.2 Attention**

The visual system can only analyze a limited amount of information that impinges in the retina [Tsotsos, 1988, Tsotsos, 1990, Tsotsos, 1992]. In order to cope with this computational limitation, attention for non-detection visual tasks (task priming can also affect detection tasks) is the only solution [Tsotsos et al., 2008]. Attention is directed usually to where we foveate. To bring attended objects or locations to the fovea, attention usually interacts with eye movements, al-

though attention can also be activated without the involvement of eye movements [Bushnell et al., 1981, Hoffman, 1998b].

The nature of attentional influence has been debated for a long time. [Sutherland, 1998] writes *after many thousands of experiments, we know only marginally more about attention than about the interior of a black hole*. Attention has been viewed as early selection [Broadbent, 1958], using attenuator theory [Treisman, 1964], as a late selection process [Norman, 1968, Deutsch and Deutsch, 1963], as a result of neural synchrony [Milner, 1974], using the metaphor of a spotlight [Shulman et al., 1979], within the feature integration theory [Treisman and Gelade, 1980], as an object-based phenomenon [Duncan, 1984], using the zoom lens metaphor [Eriksen and St James, 1986], as a premotor theory subserving eye movements [Rizzolatti et al., 1987], as biased competition [Desimone and Duncan, 1995], as feature similarity gain [Treue and Martínez-Trujillo, 1999], and more.

Presently, the main discussion involves bottom-up and top-down attention. Some authors have argued that attention is mainly a bottom-up process [Koch and Ullman, 1985, Itti et al., 1998], others propose that attention operates in two steps [Desimone and Duncan, 1995, Tsotsos et al., 1995]: In a first step, a top-down bias primes the neurons that encode the object to attend, the second

step comprises a competition in which the primed neurons inhibit the effect of the rest of neurons.

In the top-down attention theories, there is cooperation among different neurons that respond to the same stimuli, while the ones that respond to different stimuli compete [Desimone and Duncan, 1995, Tsotsos et al., 1995]. The final response will depend on the strength of the interactions of neurons that respond to the same stimuli, and the strength of the neurons competing against them. This cooperation-competition occurs across multiple brain systems [Duncan et al., 1997], meaning that there is no specific location for attention in the brain. Attentional effects have been found in every area of the visual cortex and other areas of the frontal and parietal cortex [Bushnell et al., 1981, Kastner and Ungerleider, 2000].

A large number of models have an abstract base, being more a theory than a model [Broadbent, 1958, Broadbent, 1982, Deutsch and Deutsch, 1963, Treisman and Gelade, 1980, von der Heydt et al., 1984, Koch and Ullman, 1985, Wolfe et al., 1989, Phaf et al., 1990, Niebur et al., 1993, Grossberg et al., 1994, Desimone and Duncan, 1995, Schneider, 1995]. Only some of them have a strong computational base, and even less have been tested computationally [Sandon, 1990, Olshausen et al., 1993, Tsotsos et al., 1995, Postma et al., 1997,

Itti et al., 1998, Mozer and Sitton, 1998, Deco and Zhil, 2001, Heinke and Humphreys, 2003, Lee et al., 2007] and even a smaller number have been tested in the recognition of basic objects or visual search tasks [Grossberg et al., 1994, Deco and Zhil, 2001, Rodríguez-Sánchez et al., 2006, Rodríguez-Sánchez et al., 2007].

### 1.2.1 Neurobiological evidence for attention

Evidence for attention in physiology was first found in the Superior Colliculus [Goldberg and Wurtz, 1972, Wurtz and Mohler, 1976], and later in other parts such as the posterior parietal cortex [Bushnell et al., 1981], V2 [Wurtz and Mohler, 1976, Motter, 1993, Luck et al., 1997], V4 [Fischer and Boch, 1981, Moran and Desimone, 1985, Motter, 1993, Luck et al., 1997, Reynolds et al., 2000] and IT [Moran and Desimone, 1985, Chelazzi et al., 1993, Chelazzi et al., 1998]. While several studies failed to find attention effects in V1 [Moran and Desimone, 1985, Motter, 1993, Luck et al., 1997], others have reported some modulation in this visual area [Motter, 1993, Press et al., 1994, Kastner et al., 1998]. The large amount of feedback connections existing in the cerebral cortex would also support the idea of visual attention from a neurophysiological perspective

[Moran and Desimone, 1985, Felleman and Van Essen, 1991].

The effects of attention are directly related to the neuron's receptive field size [Kastner and Ungerleider, 2001]. That is, smaller effects are found in V1 neurons than in V4 neurons. In the case of V4, when attention is directed to the ineffective stimulus, the neurons response is greatly attenuated compared to when attention is directed to the effective stimulus, even when both - effective and ineffective - stimuli are in the neuron's receptive field [Moran and Desimone, 1985]. Single unit recordings from IT neurons with objects present in their receptive field have demonstrated that attention is the result of a competition among neurons [Chelazzi et al., 1993, Chelazzi et al., 1998].

These studies support top-down theories of attention, such as Selective Tuning [Tsotsos et al., 1995], Biased Competition [Desimone and Duncan, 1995] and others [Ferrera and Lisberger, 1995]. These models have in common that a neuron is inhibited through competition when an ineffective stimulus is attended even in the presence of an unattended effective stimulus in the neuron's receptive field. Competition is directed by a top-down bias from working memory, where the information regarding objects are stored [Desimone and Duncan, 1995]. Evidence for attention has also been found in the visual cortex areas involved in motion [Treue and Martínez-Trujillo, 1999].

Attentional effects are also found in lower visual areas. Recent studies with fMRI (functional magnetic resonance imaging) in humans found effects of attention in human visual areas V1, V2, V4, IT [Kastner and Ungerleider, 2000, Pessoa et al., 2003] and even as early as LGN [O'Connor et al., 2002]. Also, using a combination of fMRI and ERP, the modulation of attention was found to appear in V1 with a delay of 150-250 ms [Martinez et al., 1999], which may explain why previous studies using single unit recording failed to find such modulation [Moran and Desimone, 1985, Luck et al., 1997] in this area.

### 1.2.2 Visual Search

In a visual search experiment, the task is to look for a target defined by one or more features among a set of distractors that are different from the target, although it may share one or more features with it. When the feature space is one dimensional and the distance between target and distractors in that feature space is large, it is commonly known as feature visual search and the target seems to *pop-out* (e.g., a red vertical bar among a set of green vertical bars).

When there are two different kinds of distractors and the target shares a feature with each one of the two types of distractors, this search is referred to as conjunction search and it requires more time to find the target (e.g., Look for a

red vertical bar among red horizontal bars and green vertical ones). Decades of psychophysical experimentation have analyzed response-time (RT) as a function of the number of distractors for most of the different features under thousands of different situations [Wolfe, 1998b, Wolfe, 1998a].

The analysis of  $RT \times$  set size slopes has been widely used to propose different theories on how the brain works for such tasks. In feature searches, (Figure 1.3a) the target is defined by a single feature, and  $RT \times$  set size slopes are near zero msec/item. In conjunction searches (Figure 1.3b) the target is defined by the conjunction of two different features, and slopes are greater than zero msec/item (e.g. 20-30 msec/item in the case of searching an S among mirror-S's or L among T's) [Wolfe, 1998a].

One of the most influential theories was the Feature Integration Theory [Treisman and Gelade, 1980], which proposed that feature search was the result of a parallel process while a conjunction search was the result of a serial search. More recent models [Wolfe et al., 1989, Duncan and Humphreys, 1989] have rejected that hypothesis, proposing a visual search continuum directly related to the similarity among target and distractors.

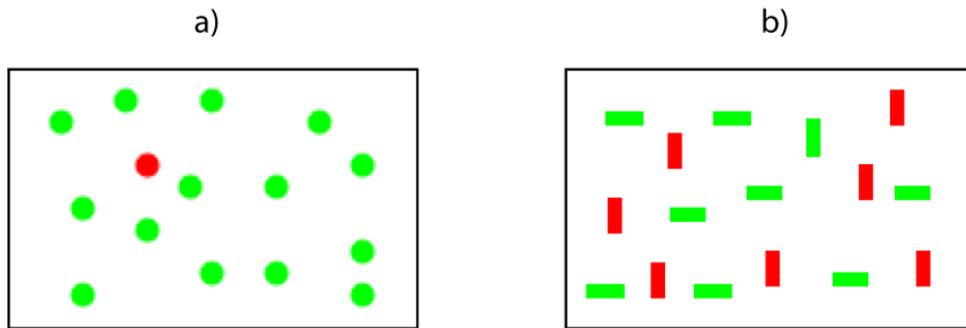


Figure 1.3: a) Feature search. In feature search the target is defined by a single feature (color). b) Conjunction search. In conjunction search the target is the conjunction of two features shared with the distractor (orientation and color in this case)

### 1.2.3 Basic features in visual search

**Color** Color has always been used as one of the basic features in visual search experiments as well as in the different theories that try to explain attention [Treisman and Gelade, 1980, Quinlan and Humphreys, 1987, Humphreys et al., 1989, Wolfe et al., 1989]. Efficient search can be performed with heterogeneous colors (up to nine distractors) as soon as they are widely separated in color space [Wolfe, 1998a].

There exist asymmetries in color search [Treisman and Gormican, 1988]: to search for a magenta target among red distractors is easier than finding a red target among magenta distractors. It is then argued that it is easier to find a variation of red than a red target among variations of red. Since magenta

contains blue, an explanation is that in the first case the target contains blue among distractors that do not contain this color. On the other hand, finding a red item among magenta distractors implies finding the reddest item. This hypothesis has been later studied in greater detail [Nagy and Sanchez, 1990].

V4 seems to be the main area where color is processed [Essen and Zeki, 1978]. But there are other areas where the processing of color occurs, namely V1, V2 and IT. V1 and V2 are involved in correcting changes in luminance due to variations in chromaticity as well [Yoshioka and Dow, 1996]. V4 may be a second stage, where color constancy is computed, the final step is IT where there exist an association of color with form [Zeki et al., 1999].

**Orientation** Orientation is another broadly accepted basic feature. V1 neurons are selective to different orientations of bars and edges [Hubel and Wiesel, 1959, Hubel and Wiesel, 1962, Hubel and Wiesel, 1968]. Subjects are capable of discriminating lines with as little as  $1^\circ$  or  $2^\circ$  difference in orientation. Although, for efficient visual search, a minimum of  $15^\circ$  is required. When distractors are heterogeneous, search becomes inefficient [Wolfe, 1998a]. Search asymmetries were also found in orientation: it is easier to find a vertical target among distractors that are tilted  $20^\circ$  than to find a  $20^\circ$  tilted target among vertical distractors

[Wolfe et al., 1992]. In that same study, it was shown that search is efficient when the target is the only vertical, horizontal, left or right-tilted element, even in the presence of heterogeneous distractors.

**Size** If the size difference is enough, then a target specified by one size can be found quite efficiently among distractors of another size [Treisman and Gelade, 1980, Quinlan and Humphreys, 1987, Duncan and Humphreys, 1992]. Asymmetries are found here as well: it was harder to find a small target among big distractors than a big target among small distractors [Treisman and Gormican, 1988].

**Shape** Shape has been widely used as another feature for visual search. Most of the time, letters are used for studying shape [Treisman and Gelade, 1980, Quinlan and Humphreys, 1987, Humphreys et al., 1989]. Some authors support the idea of letters as basic features [Malinowski and Hubner, 2001], which might be true if we subtract the meaning of a letter and consider it only as shape. This hypothesis is further supported by the fact that junctions (as in letters) can be combined hierarchically [Humphreys et al., 1989].

But shape is still a controversial basic feature, since we tend to think that a basic feature is a low-level feature (like edges and colors). Some studies have

shown that it does not have to be case, that basic features are not necessarily low-level features [Levin et al., 2001]. Curvature and depth are considered basic features as well [Wolfe, 1998a]. Letters, depth and shape may be related with object-based attention, or they may be considered high-level features.

**Motion** Another widely accepted basic feature is motion. It is easy to find a moving target among static distractors. Spatiotemporal features have been demonstrated to be the first features developmentally present in humans for recognizing objects (even sooner than color and orientation) [Xu and Carey, 1996]. Asymmetries exist as well in motion; it is easier to find a moving object among stationary objects than the opposite [Wolfe, 2001].

## Summary

An enormous number of visual search experiments have shown the following:

1. The more the target is similar to distractors, the greater the difficulty for the visual system to localize it.
2. There is a number of basic features that are important for the visual system when analyzing a scene, those are color, orientation, shape, motion and many others. For a complete list refer to [Wolfe, 1998a].

3. For each of these basic features, there exist asymmetries when its role is changed from target to distractor or viceversa.

### 1.3 Shape in Computer Vision

As mentioned previously in section 1.1, it has been recently discovered that areas V4 and TEO are involved in curvature and shape analysis.

Shape and curvature have been popular in computer science and vision sciences for object recognition for decades. Early studies approximated curves using line segments or *chain encoding* [Freeman, 1974, McKee and Aggarwal, 1977]. Polygonal approximations [Pavlidis, 1972] were used early as well, where each segment is considered a graph with branches connecting the nodes corresponding to segments with overlapping boundaries, various criteria determine the breakpoints that determine the best polygon.

Other early approximations to shape and curvature are through skeletons [Blum, 1973], shape descriptors such as area and perimeters [Danielsson, 1978] or angles and sides [Davis, 1977]. Fourier descriptors have been widely used since [Persoon and Fu, 1974] for boundary representation; a curve is represented by the Fourier expansion of a parametric representation of a curve. A similar approach is the use of principal components which uses the strongest set of

eigenvectors [Darroch and Mosimann, 1985]. As early as 1967, [Blum, 1967] already points out the difference between biological and physical representations regarding shape. In this work he presents the Medial Axis Function (MAF) that can describe any pattern by the envelope of circles of proper radius associated with each point. In this paper, Blum discusses about psychological and physiological implications regarding illusions and cortical cells responses reported by [Hubel and Wiesel, 1962, Hubel and Wiesel, 1965]. In a recent book [Siddiqi and Pizer, 2008] summarizes many of the later findings that point to a role of the medial loci (also known as skeletons) in shape perception.

[Ballard, 1981] proposed a method to detect shapes using the Hough transform. The Hough transform is a method for detecting curves by exploiting the duality between points on a curve and parameters of that curve. A mapping between image space and Hough transform space through the boundaries of a shape can be performed, this mapping can be used to detect instances of a particular shape.

The Curvature Primal Sketch [Asada and Brady, 1984] followed the primal sketch proposed by [Marr and Nishihara, 1978] for curves. A curve is approximated using a library of analytic curves, the curvature function is then computed and convolved with a Gaussian of varying standard deviation. Darboux vectors

have been used for contour descriptions [Kehtarnavaz and deFigueiredo, 1988] since they contain information regarding curvature and torsion properties. Hype [Ayache and Faugeras, 1986] used segmented descriptions of the object contours to generate and recursively evaluate a number of selected hypotheses. [Mehrotra and Grosky, 1989] also presented a system of hypothesis generation and testing based on dynamic programming.

Curvature extrema were used by [Richards et al., 1988] which is based on the *codons* from human perception [Hoffman and Richards, 1984]. Codons are sequences of curvature extrema separated by curvature minima and zeroes, whose combinations provide the type of codon. A list of curvature extrema provides a “vocabulary” for curves. Not all combinations of codons are possible in a smooth curve, the possible combinations of codons further constrains the vocabulary [Richards and Hoffman, 1984]. [Leyton, 1987] exploits curvature extrema and symmetry structure as descriptors of shape and presents the *Symmetry-Curvature Duality Theorem* that proposes a relationship between symmetry and curvature extrema.

[Horn and Weldon, 1986] proposed the use of the extended Gaussian image in their *Extended Circular Image*. The extended Gaussian image of a polyhedron is obtained by placing a mass at each point equal to the surface area of the

corresponding face. The circular image of a convex segment encodes information about the absolute orientation of the segment. In the extended circular image, a planar contour is parametrized in terms of its tangent with respect to the x-axis making it invariant to size changes as well as stretching or shrinking. *Landmarks* [Ansari and Delp, 1990] are points of interest with important shape attributes such as points of high curvature. [Ansari and Delp, 1990] used sphericity for matching. Sphericity is a measure of similarity between triangles.

A popular strategy for curve-based representations in object recognition are splines. Splines are piecewise polynomials in an interval [a,b] connected smoothly by knots (joining points). Smoothed splines have been proposed to parametrize the curve [Shahraray and Anderson, 1989]. To obtain the optimal degree of regularization (the smoothing term), cross-validation was used. Cross-validation is a method where the goodness of a set of parameters is measured in terms of the ability of the model to predict some of the observations. [Cohen and Wang, 1994] estimated the best B-spline based on a minimum mean square error estimation and a Bayesian model for deciding the best order ( $k+1$ ) and control points (knots,  $k$ ) of the spline. B-splines are invariant to affine transformations and they globally define a curve through a set of knots [Cohen and Wang, 1994]. But splines fail in what is called “the knot problem”, where splines have more degrees of

freedom than the objects they represent.

The work of [Shahraray and Anderson, 1989] first presented the idea of analyzing a shape by a curvature-position histogram, a representation later proposed for V4 neurons [Pasupathy and Connor, 1999, Pasupathy and Connor, 2001, Pasupathy and Connor, 2002]. [Marimont, 1984] presented a framework that extracted curvatures at different scales. CSS [Mokhtarian and Mackworth, 1984, Mokhtarian, 1995] is a multiscale representation of shape curvatures, a curve is described at varying levels of detail through features invariant to transformations. The features used for matching are the maxima of CSS contours. A three-stage, coarse-to-fine matching algorithm prunes the search space in stage one by applying the CSS aspect ratio test. The maxima of contours in CSS representations of the surviving models are used for fast CSS matching in stage two. Finally, stage three verifies the best match and resolves any ambiguities by determining the distance between the image and model curves.

A similar approach was presented by [Dudek and Tsotsos, 1997], in which curvature is represented at multiple spatial scales. Curved objects are described by a group of segments that encode length, position and curvature. Matches consist of a sequence of segments along the curve being similar to a sequence from another curve. Dynamic programming is used to evaluate the quality of

matching.

Kimia and colleagues [Kimia et al., 1992, Kimia and Siddiqi, 1994, Kimia et al., 1995] proposed a theory of shape based on deformations, parts, bends and seeds. Contour deformations are characterized by a deformation along the normal and a deformation that varies in proportion to curvature. A space of shapes is constructed in which similar shapes are arranged according to the different axes of deformation. A combination of smoothing and erosion to extract decompositions of objects show utility for recognition. An infinite sequence of moments can determine a shape and has been used for shape representation, moment invariants are invariant under certain transformations such as rotation, scale and translation [Chen, 1993].

[Belongie et al., 2002] captured the subset of points from object shapes sampled from the internal or external contours. Shape samples are performed through uniform spacing. A descriptor (shape context) is attached to each point that captures the distribution of the remaining points relative to it such that similar shapes have similar descriptors. [Elder et al., 2003] used prior knowledge of objects for contour grouping. An approximate constructive search technique computes the candidate object boundaries. They tested their approach with lake boundaries from satellite images and human skin boundaries.

This chapter shows how shape has been considered for object recognition. Even though this review may look somewhat extensive, there are many other methods and improvements over the exposed ones that have not been considered here.

## 1.4 Biologically inspired models of Object Recognition

The history of computer vision now spans more than half a century and still there is not a complete satisfactory solution to the problem of object recognition. In object recognition, the goal is to locate an object in a scene. Classical object recognition methods generally apply algorithms based on geometrical or other mathematical methods for finding and recognizing an object in a scene. There are many successful methods for analyzing a scene and finding an object as soon as some -or many- restrictions are considered (see [Besl and Jain, 1985, Bennamoun and Mamic, 2002] for a review). Lately, with the advent of new breakthroughs in neuroscience, some computational systems are trying to emulate the human visual system for object recognition and motion detection. However the ability of these biologically plausible models in real world scenarios is still limited. Although this thesis does not address object recognition directly, it may provide important contributions to elements that may advance the state-of-

the-art.

The last two decades have seen the resurgence of neural networks, which originated in 1962 with the Perceptron [Rosenblatt, 1962]. Fukushima's neocognitron [Fukushima, 1980] is a self-organizing neural network model that achieves position invariance and was later proven to perform well on digit recognition [Fukushima et al., 1983]. Later models that included backpropagation [Rumelhart and McClelland, 1986] were also successful at this task [Lecun et al., 1989, Lecun et al., 1998]. Over the last fifteen years, many models inspired by advances in the anatomy of the visual cortex have been presented.

One of the first such models was proposed by Olshausen and colleagues [Olshausen et al., 1993]. They developed a model that is based on [Anderson and Van Essen, 1987], is position and scale invariant and performs a transformation from the retinal reference frame to an object-centred frame. To accomplish this, shifting circuits and control neurons are used. The control neurons dynamically conduct information from lower levels of a hierarchical network to higher levels of the network. By means of the shifting circuits and the control neurons, the window of attention changes in size for scale invariance and position. This model has been extended on the later SIAM [Heinke and Humphreys, 2003].

Visnet [Wallis and Rolls, 1997] consists of a four layer network that achieves

invariant object recognition. The most crucial part of such a method is a trace learning rule that is Hebbian based. Lateral inhibition is performed following [von der Malsburg, 1973], and competition is applied by means of a soft winner take-all. To achieve translation invariance, the network is trained with inputs at different positions. A later review includes a top-down attentional strategy [Rolls and Deco, 2002].

Probably, the most popular model is the one proposed by Riesenhuber and Poggio [Riesenhuber and Poggio, 1999, Riesenhuber and Poggio, 2000, Riesenhuber and Poggio, 2002]. The model consists of five hierarchical levels of neurons that are connected through linear and non-linear MAX operations (the strongest units determine the response of the system). The first level receives input from the retina and is composed of simple neuron receptive fields that analyze orientations. The next levels account for more complex features (e.g. junctions). The last level is composed of view-tuned neurons that achieve position and scale invariance. The original model included a Radial Basis Function (RBF) [Riesenhuber and Poggio, 1999] for classification, and a later update included a Support Vector Machine (SVM) [Serre et al., 2005]. The model has been recently extended by the inclusion of an unsupervised learning stage [Serre et al., 2007].

[Amit, 2000] presents a parallel neural network for visual selection. This net-

work is trained to detect candidate locations for object recognition and it contains layers similar to those found in the visual cortex. Objects are represented as composed of features localized at different locations with respect to an object centre. Simple features (edges and conjunctions) are detected in lower levels, while higher levels carry out disjunctions over regions. Detection is accomplished by first constructing a graph of features and finding the candidate regions on the image through a Hough transform. The Hough transform also accounts for size and rotation invariance. Visual attention is accomplished by priming the locations containing the object features.

Suzuki and colleagues [Suzuki et al., 2004] construct a model of the form pathway based on predictive coding [Rao and Ballard, 1997, Rao and Ballard, 1999]. Predictive coding hypothesizes that feedback connections from high to lower-order cortical areas carry predictions of lower-level neural activities. Feedforward connections carry residual errors between predictions and the actual lower-level activities. In the model, a fast coarse processing precedes and constrains finely detailed processing.

[Murphy and Finkel, 2007] implement a set of feature vectors of contours: mean polar angle, mean curvature of region, mean curvature of adjacent clockwise region, mean curvature of counter-clockwise region, mean direction of curvature

region, mean distance from center of mass and indication of inner or outer contour. Recognition is then performed through matching using comparison of segments, histogram cross-correlations, minimum sum of squared differences, image cross-correlation, parametric eigenspaces and support vector machines (SVM). They report their results using Earth Mover's distance which is motivated by human perceptual studies [Kahana and Sekuler, 2002].

### **Summary.**

Many biologically inspired object recognition systems have been proposed over the years. The performance of biologically plausible models is not yet comparable to other, more computationally "pure", systems of object recognition, such as [Lowe, 1999, Lowe, 2004, Fidler et al., 2006, Mekuz and Tsotsos, 2007] and others. On the other hand, the ultimate object recognition system may be implemented in each of us, no computer system to date can match the primate visual system.

There is starting to be a large amount of evidence on how the visual system works, which allows us to establish the basis of future systems that emulate the human visual system. In addition to this, top-down attention operating on a finer grain set of intermediate level representation will lead to more powerful computer

vision recognition systems.

## 1.5 Original Contributions

Previous work on object recognition has focussed on biological plausibility for early processing layers but much less so for later processing. In this thesis, the biological emphasis is maintained, but we consider intermediate layers of representation and analysis. We provide several studies concerning those middle areas in the object recognition pathway (V4/TEO) and the motion pathway (MT/MST). These middle areas are also well known in studies of attention [Desimone and Duncan, 1995, Treue and Martínez-Trujillo, 1999].

First, we compare the Selective Tuning model of attention [Tsotsos et al., 1995] with psychophysical experiments, and the selected features to test the model were shape and color. We then study in more detail how shape is analyzed in the visual cortex and we provide a biologically plausible proposal of how shape analysis may be achieved along a hierarchy of neurons that mimics areas V1, V2, V4 and TEO. Finally, since Selective Tuning was tested on a representation of motion [Tsotsos et al., 2005] which included spatial gradients of velocity vectors, we provide a study that analyzes the processing of gradients of velocity in area MST.

### 1.5.1 Testing the Selective Tuning model in visual search

Visual search is closely related to object recognition. For [Besl and Jain, 1985], the problem of object recognition comprises the following steps:

1. *Given a set of objects, examine each object and label it.*
2. *Given an array of pixels from a sensor and a list of objects, those questions arise:*
  - (a) *Is the object present in the scene?*
  - (b) *If so, how many times does it appear?*
  - (c) *For each occurrence find its location in the scene and determine its translation and rotation parameters referred to a known coordinate system.*

A third step would be learning the new unrecognized objects.

It is important to note step 2(a), Is the object present in the scene? this is exactly what visual search experiments consist of, as explained in subsection 1.2.2. Our first contribution provides a test of an attentional model [Tsotsos et al., 1995] with the type of visual search tasks found in psychophysics. The results of several studies in visual search are replicated through the use of the Selective Tuning

model proposed by Tsotsos and colleagues [Tsotsos et al., 1995]. Our experiments illustrate the biological plausibility of the Selective Tuning model and its relevance to object recognition.

### 1.5.2 A biologically plausible model of Shape

Chapter 3 describes a model of shape with a strong biological background. At the earliest layers of computation in the model there is a set of edge extractors, corresponding to V1 which includes simple and complex neurons [Hubel and Wiesel, 1968]. On top of it is a layer composed of endstopped neurons which feeds into local curvature selective neurons, ending into neurons that respond to shapes. Endstopping is achieved following the work of Dobbins and Zucker [Dobbins et al., 1987, Dobbins et al., 1989] and is considered by some authors [Pasupathy and Connor, 1999] as a plausible player for the analysis of curvatures in area V4.

Shapes are an important feature for recognizing objects [Wolfe, 1998a]. Shapes are composed of straight lines, corners, junctions and curves. Curvature has been present in many methods and algorithms that achieve object recognition (e.g., [Mokhtarian, 1995]). Recent studies in area V4 have shed some light into the role of curvature in

shape recognition [Pasupathy and Connor, 1999, Pasupathy and Connor, 2001, Pasupathy and Connor, 2002, Brincat and Connor, 2004]. These studies are taken into account for the model in its later areas.

In this thesis, a complete architecture that mimics shape analysis in the visual cortex is presented as well. In particular, a full hierarchy of neurons strongly derived from biology in its architecture, neuron function and most importantly, interactions. The model is successfully compared with neural data from [Pasupathy and Connor, 2001, Pasupathy and Connor, 2002].

In contrast to existing recognition systems described in section 1.4, this work adds several new layers of shape representation.

### **1.5.3 The role of spatial gradients of velocity in human behavior and primate neurons**

Finally, chapter 4 describes a set of experiments with humans and monkey regarding visual motion. The original presentation of the model of Selective Tuning [Tsotsos et al., 1995] was tested with motion analysis, making use of spatial gradients of velocity [Tsotsos et al., 2005] in the intermediate layers. Here we study the role and the presence of sensitivity to spatial gradients of velocity.

In order to do so we constructed a set of psychophysical experiments that

provides discrimination threshold measurements for the contracting/expanding and radial (rotating) components. We found psychophysical evidence of a role of spatial gradients of velocity in optical flow perception.

Through a collaboration with McGill University and the University of Tuebingen (Germany), similar experiments were performed, recording neurons from a macaque monkey in area MST. After performing the corresponding data analysis from the monkey neurons and the human subjects, when comparing both, in our study, the single neurons did not perform as well as human subjects, and we provide several hypothesis that may help understand neural integration of motion signals in the visual cortex.

## 2 Attention and Visual Search

Selective Tuning (ST) [Tsotsos et al., 1995] presents a framework for modeling attention and in this work we show how it performs in covert visual search tasks by comparing its performance to human performance. The ST Simple Stimulus Recognition Model attends to and detects simple stimuli formed by the conjunction of various features such as color and shape. The validity of the ST Simple Stimulus Recognition Model was first tested by successfully duplicating the results of [Nagy and Sanchez, 1990]. A second experiment was aimed at an evaluation of the model's performance against the observed continuum of search slopes for feature-conjunction searches of varying difficulty. The results from the implementation agreed with the psychophysical data from the simulated experiments. We conclude that ST provides a valid explanatory mechanism for human covert visual search performance, an explanation going far beyond the conventional saliency map based explanations. Parts of this work was presented at

ICANN 2006<sup>1</sup> and completely published in the International Journal of Neural Systems<sup>2</sup>.

## 2.1 Introduction

The breadth of functionality associated with attentional processing can easily be seen in several overviews (e.g., [Hoffman, 1998a, Itti et al., 2005]). One of the most studied topics and with a very significant literature is that of visual search. Visual search experiments formed the basis and motivation for the earliest of the influential models (e.g. [Treisman and Gelade, 1980, Koch and Ullman, 1985]). Yet, no satisfactory explanation of how the network of neurons that comprise the visual cortex performs this task exists. Certainly, no computational explanation or model exists either.

In a visual search experiment, the task is to look for a target defined by one or more features among a set of distractors that are different from the target but may share one or more features with it. When target and distractors are the same except for one feature, it is commonly known as feature visual search and

---

<sup>1</sup>Feature Conjunctions in Serial Visual Search. Rodriguez-Sanchez AJ, Simine E, Tsotsos JK. Talk at the International Conference on Artificial Neural Networks, 10-14 September 2006, Athens (Greece). Work of E. Simine is not reported here and consisted in the analysis of the Selective Tuning for motion visual search.

<sup>2</sup>Attention and visual search. Rodriguez-Sanchez AJ, Simine E, Tsotsos JK. Int J Neural Syst. 2007 Aug;17(4):275-88. PMID: 17696292. Work of E. Simine as above

the target seems to pop-out (e.g. a red vertical bar among a set of green vertical bars). When there are two different kinds of distractors and the target shares a feature with each one of the two types of distractors, this search is referred to as conjunction search and it requires more time to find the target (e.g. Look for a red vertical bar among red horizontal bars and green vertical ones). Decades of psychophysical experimentation have analyzed response-time (RT) as a function of the number of distractors for most of the different features under thousands of different situations [Wolfe, 1998a].

The analysis of  $RT \times \text{size}$  slopes has been widely used to propose different theories on how the brain works for such tasks. One of the most influential was the Feature Integration Theory [Treisman and Gelade, 1980], which proposed that feature search was the result of a parallel process while a conjunction search was the result of a serial search. More recent models [Duncan and Humphreys, 1989, Wolfe et al., 1989] have rejected that hypothesis, proposing a visual search continuum directly related to the similarity among target and distractors.

Some recent models of attention [Itti et al., 1998] have been compared to human eye movement tracks -overt attention- as validation; but this is not the same as visual search data which is almost exclusively covert, with no eye movement. Visual attention involves much more than simply the selection of the next location

to fixate the eyes or camera system, regardless of the fact that the vast majority of all computational approaches to attention focus on this issue exclusively. That humans are able to attend to different locations in their visual field without eye movements has been known since [von Helmholtz, 1924]. Further, eye movements require a shift of visual attention to precede them to their goal ([Hoffman, 1998a] surveys relevant experimental work).

Attentional models have matured sufficiently so that this broader problem of attention can now be confronted. This chapter makes several steps towards the development of such an explanation expanding the Selective Tuning model [Tsotsos et al., 1995, Tsotsos et al., 2005] and comparing performance with existing visual search psychophysical performance. This is done with simple colored shape stimuli. The rest of the chapter is organized as follows: Section 2.2 describes the ST Simple Stimulus Recognition Model with its two main pathways, shape analysis and color analysis. Section 2.3 provides the results of comparing Selective Tuning with psychophysical experiments extracted from the literature, obtaining comparable results to those. Conclusions are presented in section 2.4.

## 2.2 A Simple Stimulus Recognition Model

Given a scene with several stimuli, the model's purpose is to find a particular stimulus that has been presented previously. The model structure is a two-pathway pyramid with information flowing from the input to the top of the pyramid and from the top to the bottom providing feedback. Each one of the two pathways analyze the visual input in a different way, one extracts color information, while the other extracts information about the shape of the stimuli. The model mimics aspects of the human visual pathway for object recognition, simulating four visual areas: LGN, V1, V4 and IT. Each area is organized into feature maps and each feature map encodes the visual field in a unique way.

The model comprises a total of 22 feature maps. Information first flows from the input to area LGN and V1. LGN extracts three color feature maps (red, green and blue). V1 is composed of edge detectors organized in 8 feature planes (each containing neurons tuned to one of 8 directions). Two additional feature maps in V1 compute center-surround color differences from the LGN color feature maps. Information from V1 flows to V4, which comprises 8 feature maps for curvature. Finally, IT neurons encode a representation of the whole object based on curvature and color differences.

Our testing strategy follows the sequence of events in a human visual search

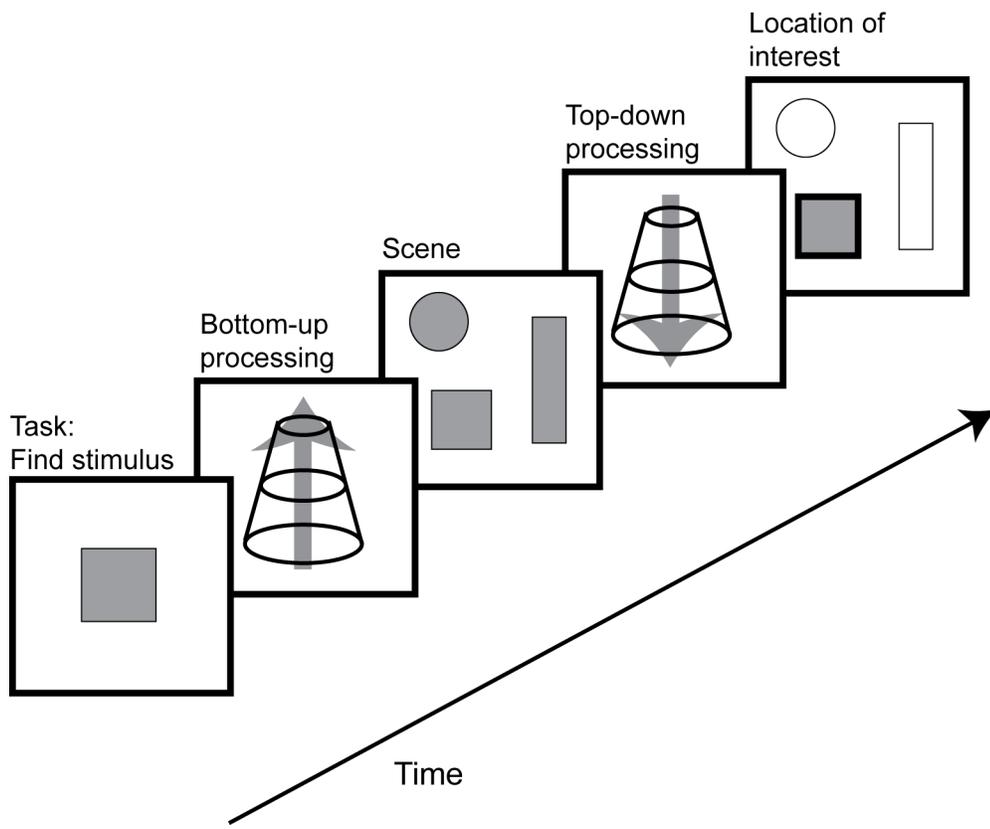


Figure 2.1: Visual search time course via Selective Tuning

experiment, (Figure 2.1), that is, a subject is first shown the target on a blank display, then is shown the test display to be searched. Similarly, the system is first shown the target and extracts a representation of it. This representation is used to bias the subsequent search when the test display is presented. When the test display is presented, biased shape and color analysis proceed in parallel, then the Selective Tuning [Tsotsos et al., 1995] feedback attentive process is applied. The different stages of processing are explained in more detail in the following sections.

### **2.2.1 Attention via Selective Tuning**

The Selective Tuning Model [Tsotsos et al., 1995, Tsotsos et al., 2005] is a hierarchical system with bottom-up and top-down attention (Figure 2.2). Top-down attention is performed by inhibiting the response of neurons at lower levels through gates at every level. The model works in three main steps:

1. The first set of computations to be performed is related to priming of the hierarchy of processing areas, by presenting the system with its task. Task knowledge, such as fixation point, target/cue location, task success criteria, is applied to *tune* the hierarchy.
2. Upon presentation of stimuli, a feed-forward, diverging, cone of neural path-

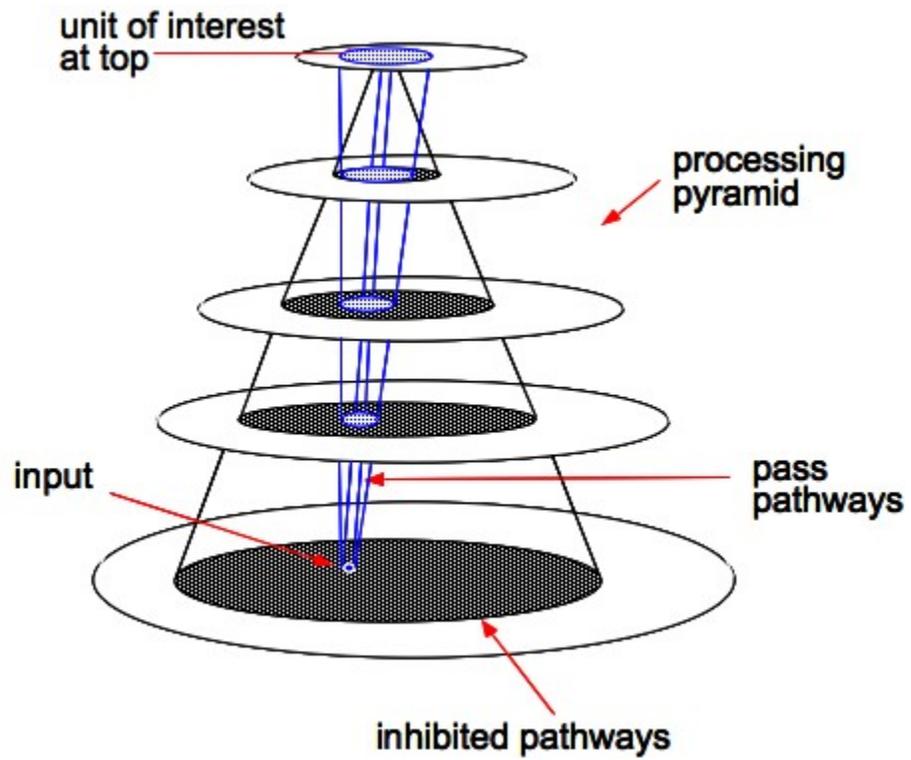


Figure 2.2: The Selective Tuning model of attention

ways is activated within the visual hierarchy. At the end of this feed-forward traversal, sufficient processing has been completed for Detection or Categorization Tasks.

3. Recurrent (or feedback) traversals through the visual processing hierarchy that ‘trace’ the pathways of neural activity that lead to the strongest responding neurons at the top of the hierarchy that result from the feed-forward traversal. as the top-down trace proceeds, connections that do not contribute to the strongest responses are inhibited, resulting in a suppressive surround around the attended stimulus; this suppression is both spatial and featural.

Some feature types may co-exist while others compete with one another, e.g. In area V4 neurons are selective for curvature; different curvatures can co-exist at different locations to represent an object, but they compete if at the same location.

A detailed description of the model is beyond the scope of this thesis and is provided in [Tsotsos et al., 1995, Tsotsos et al., 2005].

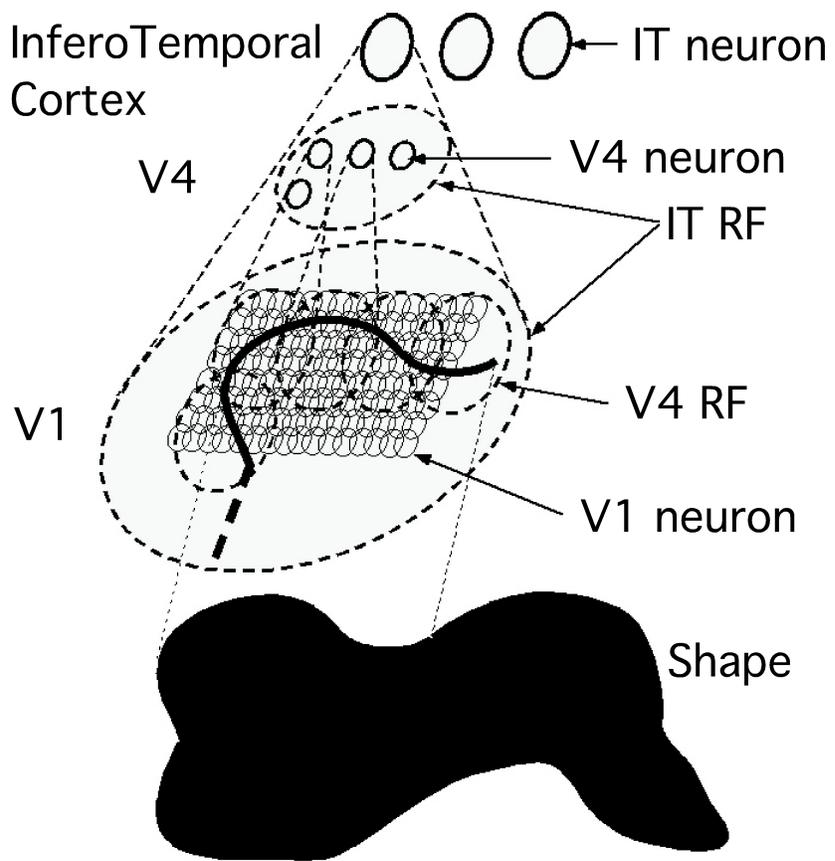


Figure 2.3: Architecture of shape pathway

### 2.2.2 Shape analysis

The shape processing pathway (Figure 2.3) is inspired by work on neurophysiology [Pasupathy and Connor, 1999, Pasupathy and Connor, 2001, Pasupathy and Connor, 2002]. Visual Area V1 contains neurons that perform edge analysis. Gabor filters [Marcelja, 1980] are used with 8 different orientations:

$$G(x, y) = e^{-(\alpha^2 x'^2 + \beta^2 y'^2)} e^{-j2\pi f x'} \quad (2.1)$$

$$x' = x \cos(\theta) + y \sin(\theta)$$

$$y' = -x \sin(\theta) + y \cos(\theta)$$

where  $\alpha$  and  $\beta$  are the sharpness of the Gaussian major and minor axis, with values of 1 and 0.25 in our case;  $f$  is the frequency and  $\theta$  is the orientation. The size of the neuron's receptive field is  $16 \times 16$  pixels. The output of V1 neurons is a set of 8 feature planes, representing edges at 8 orientations. The energy from each Gabor filter is obtained as:

$$V1(x, y) = \sqrt{\text{Re}(G(x, y))^2 + \text{Im}(G(x, y))^2} \quad (2.2)$$

Non-maximal suppression [Canny, 1986] is applied in order to reduce the Ga-

bor filter output to a 1-2 pixel wide images as a pre-processing for the next visual layer. The output from V1 neurons feeds into V4. V4 neurons compute curvature values based on orientation differences within groups of adjacent V1 neurons. Their receptive fields comprise  $4 \times 4$  V1 neurons at 8 orientations. For example, if a set V1 neurons in a V4 receptive field had their highest response for  $\theta_1 = 0$  and another adjacent set had their highest responses for  $\theta_2 = \frac{\pi}{4}$ , we would have a corner. If both orientations were equal, it would correspond to a straight line. Curvature for V4 is then defined as:

$$curv_i = \min(|\theta_1 - \theta_2|, 2\pi - |\theta_1 - \theta_2|) \quad curv \in (0, \pi] \quad i \in \{1 \dots 8\} \quad (2.3)$$

where  $\theta_1$  and  $\theta_2$  are the orientations of two V1 cell sets. A value of  $\pi$  can be added to  $\theta_1$  and/or  $\theta_2$  depending on the neurons' relative positions inside the V4 receptive field due to the fact that the same Gabor filter orientation can account for two different angles. The activation value of the V4 neuron is the summed activations from the V1 neurons used to obtain the curvature:

$$V4_{curv_i} = \sum_{j \in \theta_1} \sum_{l \in \theta_2} V1(j, l)$$

V4 neurons' output is 8 2D feature maps that encode for the difference of

curvature among groups of V1 neurons. This output feeds into IT at the very top of the hierarchy (Figure 2.3). The receptive fields of IT neurons comprise an area of  $32 \times 32$  V4 neurons (that is,  $128 \times 128$  pixels). The center of mass is calculated for every group of V4 neurons as the mean of the V4 neuron coordinates where responses are different from zero. Then, at each angular position (in 10 deg bins), its curvature is computed as in [Pasupathy and Connor, 2001], obtaining a histogram-like representation for IT neurons where one axis correspond to the angular position ( $\lambda$ ) and the other coordinate is the curvature *curv* for that position (Figure 2.4):

$$\lambda = \text{round} \left[ \frac{n\text{bins} * \tan^{-1} \left( \frac{y\text{-centroid}_y}{x\text{-centroid}_x} \right)}{2\pi} \right] \quad IT(\lambda) = \text{curv} \quad (2.4)$$

The term  $18/\pi$  is for the angular position to be in 10 deg bins. All neuron relative sizes were chosen to correspond closely to the neurophysiological measured sizes [Felleman and Van Essen, 1991] considering a distance of 30 cm (usual psychophysical distance) to a  $1280 \times 1024$  display. Neurons' receptive fields are overlapped.

Figure 2.4 shows how shape is analyzed. On the left of figure 2.4 is the bottom-up feedforward analysis. Edges are extracted in V1 at each different orientation, then in V4 curvatures are calculated, finally IT computes the curvature  $\times$  position

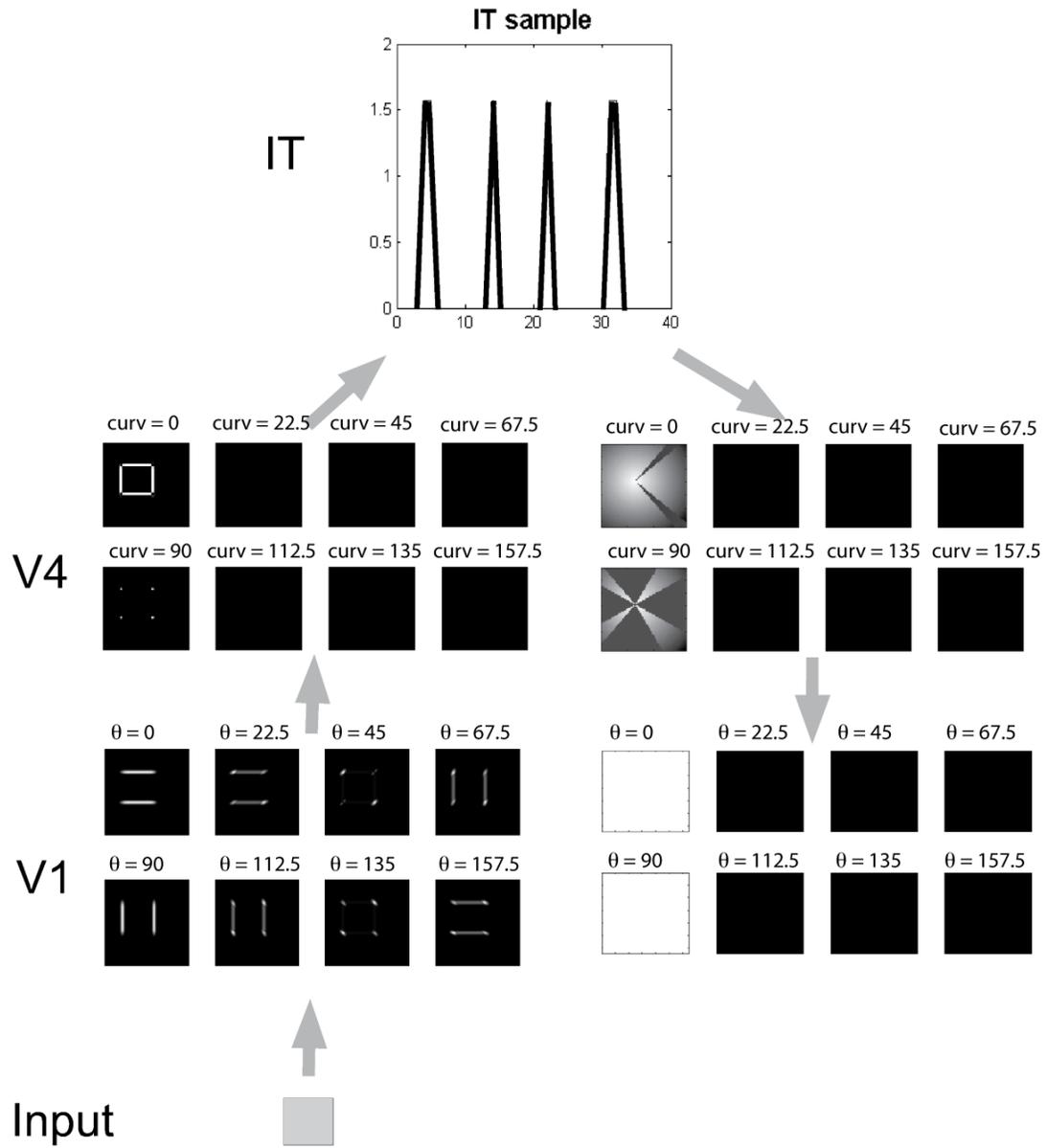


Figure 2.4: Shape analysis on target stimulus. See text for explanation

representation [Pasupathy and Connor, 2001]. On the right of figure 2.4, shows the top-down bias stage: From such a representation, in V4 feature planes that do not have values of curvature corresponding to the stimulus are inhibited (black), and in these V4 feature planes, neurons that are not at the proper location are inhibited as well. In V1, neurons that do not contribute to those V4 feature planes are also inhibited, only allowing the Gabor filters corresponding to the orientations that feed into the non-inhibited V4 neurons.

### 2.2.3 Color analysis

The processing of color follows a centre-surround analysis [Rolls and Deco, 2002]. A first layer (LGN) extract 3 feature maps for red (R), green (G) and blue (B) responses. In the upper layer (V1), surround values for red-green (RG), green-red (RG), blue-yellow (BY) and yellow-blue (YB) are extracted following most models (e.g [Itti et al., 1998]):

$$RG = \frac{R - G}{Luminance} \quad BY = \frac{B - Y}{Luminance} \quad (2.5)$$

$$Luminance = \max(R, G, B)$$

RG feature plane also accounts for GR differences, the same applies to the BY feature plane. As in the shape analysis, color neurons at every level of the

hierarchy are also inhibited if they do not share the values corresponding to center-surround and color activations of the target.

#### **2.2.4 The Bias stage**

An important part of the Selective Tuning Model [Tsotsos et al., 1995] is top-down bias (Fig. 2.2). Given a target stimulus, the features not relevant for target recognition are inhibited by multiplying an inhibitory bias (greater than 0.0 and less than 1.0) . The target representation is first obtained from the responses of the IT level neurons on seeing the target stimulus alone in the visual field.

After the representation of the stimulus shape is obtained, V4 and V1 layers are biased (Figure 2.4 right). In V4, only neurons that are not in the proper angular position and in the desired curvature feature planes are completely inhibited. For those V4 neurons that are not completely inhibited, a partial inhibition will be applied to those ones that are further from the stimulus' center of mass. Inhibition in this case is linearly proportional to the Euclidean distance to the stimulus' center. At a lower level, the neurons inhibited in V1 correspond to those whose orientation values were related to the curvature inhibited V4 neurons.

### 2.2.5 Recognition

Before the presentation of the test display, the network is biased to expect the target stimulus. The point of this bias is to speed up search; it has been shown that advance knowledge of the target indeed speeds up detection in a test display [Williams, 1966, Gould and Dill, 1969, Viviani and Swensson, 1982]. However, erroneous knowledge of the target slows down overall search [Allport, 1989]. The processing is first biased by the presented stimulus or target representation at the different visual layers of the network so that after the first feed-forward pass of processing the test display only locations with the desired target features will be considered. Then, the search begins after a feed-forward activation by considering the best matching IT neuron from the possible candidates containing non-biased features.

To determine how close is the shape to the desired shape, distance to the target IT histogram is computed; for this we used cumulative distance. This distance is very common for computing distances between histograms and it is used here due to the similarity of our representation of IT neurons to a histogram:

$$d(p, q) = \sqrt{\sum_{i=0}^{L-1} \left( \sum_{u=0}^i p_u - \sum_{u=0}^i q_u \right)^2} \quad (2.6)$$

$$IT(act) \propto \frac{1}{d}$$

The activation of the IT model neuron is inversely proportional to  $d$ . Both activation values for color and shape  $\in [0, 1]$  and the activation of the candidate IT neuron is the addition of both values. Even though the target can be in the receptive field of the highest activated IT neuron, due to its large receptive field, it may include other stimulus elements as well that may even disturb the firing values of the IT neuron. Information is further filtered in the lower layers (V4, V1) by computing winner-take-all in a hierarchical fashion [Tsotsos et al., 1995]. The WTA processes in V4 are grouped by curvature angle. There is a separate WTA process for each 10 deg bin (as determined by Equation 2), i.e., a V4 neuron will only compete with neurons in the same bin. In V1 only those neurons connected with the V4 winners are considered, and the same process is applied when going from V1 to the image, finding the contour of the candidate stimulus.

Figure 2.5 shows an example of this process: On the left, layer V1 extracts edges, V4 neurons compute curvature. Here, inside the IT neuron receptive field (RF) lays the square and part of an stimulus of no interest (rectangle). On the right, we have layers V1 and V4 after attention. Layers in the hierarchy are first biased and information is later filtered through a winner take all process (See

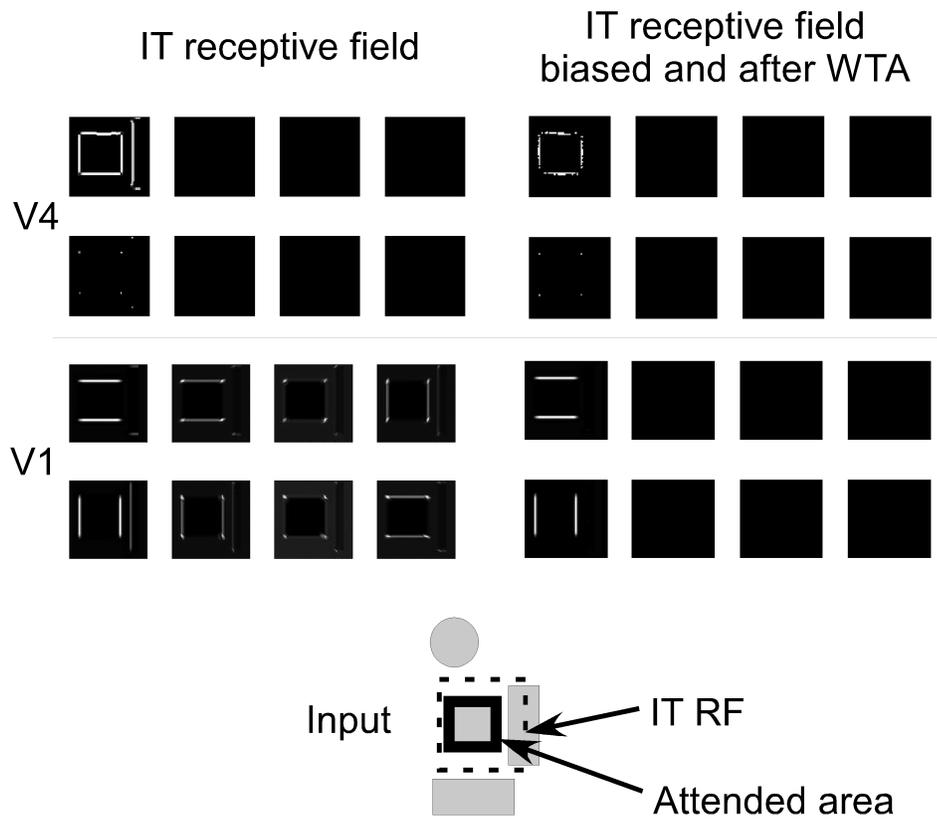


Figure 2.5: Analysis of a scene (bottom): find the square.

[Tsotsos et al., 1995] for a full explanation). Thanks to this process, information is filtered such that the stimulus of interest (square) is the only stimulus that remains inside the IT neuron RF. Inhibition of return was implemented by inhibiting the part of the input image corresponding to the analyzed stimulus.

## 2.2.6 Results

We tested the model’s behavior for different visual search conditions. For these tests, we followed different psychophysical experiments and we compared the results obtained from those works with the results obtained from the model. But before testing the model for visual search, we performed a study on how the shape representation works for a simple recognition of silhouettes.

### 2.2.6.1 Silhouette Search

Simple icons were used to infer how neurons responded to shapes as in [Pasupathy and Connor, 2001]. In a similar manner, we first tested the shape analysis component of our model with 2D silhouettes. To test the model the silhouette database from Sharvit and colleagues was used [Sharvit et al., 1998]. The architecture was fed with different silhouettes, animals, cars, planes, etc. Then, scenes were constructed with such element silhouettes and the response corresponding to the scene IT neurons were evaluated. The IT neuron from the scene with a closest response (in terms of distance) to the neuron’s representation in the database was recognized as containing the silhouette represented.

The test scene images were  $512 \times 512$  pixels. IT neuron receptive fields were  $128 \times 128$  as described previously and there was an IT neuron every ten pixels

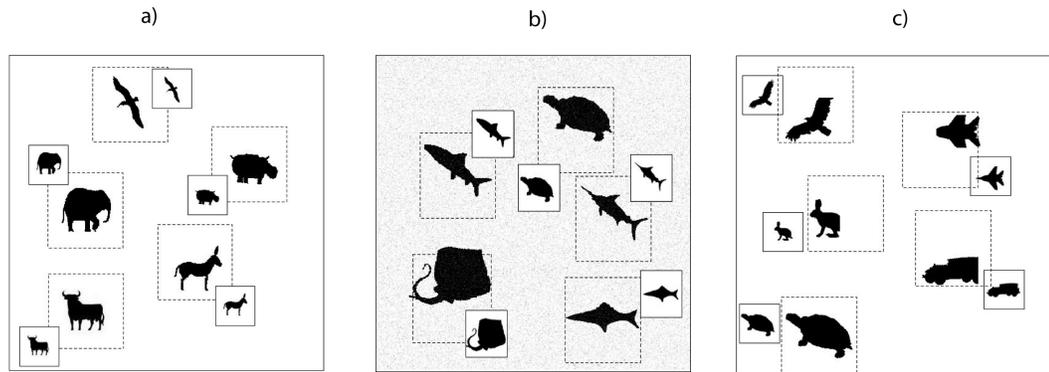


Figure 2.6: Example of silhouette recognition. insided dotted boxes are IT with highest responses corresponding to searched stimulus (next to it).

starting from the coordinates (64,64) until coordinates (448,448). Figure 2.6a show different IT neurons and the silhouettes corresponding to their highest response (inside dotted boxes). We show how the system performs when the whole silhouette is present (Figure 2.6a and b) and in conditions when the silhouettes are partially presented (Figure 2.6c). The system works well in both cases.

Although the silhouette is usually at the center of the neuron when training, the winning IT neuron does not need to have the silhouette exactly at its center, but we can see that this is usually the case. Figure 2.6c shows how the system behaves when there is partial information about the target silhouettes in the scene. We can see that the model correctly finds every silhouette, even when information is quite incomplete (e.g. the plane). Note that in these cases the IT receptive field center is not so close to the silhouette's center, while if it

is in its full shape that is usually the case (turtle). In Figure 2.6b we show how the representation performs for a case with added Gaussian white noise ( $\mu = 0, \sigma_2 = 0.01$ ). The only case where the winning neuron is not the optimal, corresponds to the ray, but a neuron very close to it corresponds to the winning neuron.

### 2.3 Efficiency in Visual Search

We now proceed with testing the model on visual search tasks. Conjunction searches [Wolfe, 1998a] may exhibit shallower slopes than those found by Treisman and Gelade [Treisman and Gelade, 1980], and there seems to exist a continuum from efficient to inefficient visual search. Early theories already postulated that visual search is influenced by the similarity between target and distractors [Duncan and Humphreys, 1989]. They stated that visual search is harder when target and distractors are more similar, but it is easier when this similarity decreases, and this theory has been supported by later experiments [Nagy and Sanchez, 1990, Thompson et al., 2005]. As a result, we decided to test the model first with an experiment concerning the similarity hypothesis. One paper that studies a fundamental basic feature (color) and known in the psychophysical community is that of [Nagy and Sanchez, 1990], and this is the

experiment we first replicate.

In our second experiment we test the search continuum and we compare the performance of the model for feature search, conjunction search and inefficient search. Feature search is a search where a target is distinguished from the distractors by a single feature such as color, shape or orientation. In our second experiment we will use the term feature search to refer to a classical psychophysical feature search experiment defined by its efficiency. As shown in experiment 1, that feature search is efficient is not always the case. In conjunction search a target is defined by a conjunction of two features.

We use the term inefficient search for those visual search experiments that are more difficult than the classical conjunction search [Wolfe, 1998a]. Note that, strictly speaking, inefficient search is also a conjunction search, but we will use a different notation to distinguish it from the classical conjunction search.

The sample was given as input in a  $128 \times 128$  pixel image, and the scenes were  $640 \times 640$  pixels. In our first experiment we will test a known feature: color, and how the model performs under two different similarities of colored stimuli. Summarizing, we first follow a known study [Nagy and Sanchez, 1990] about color similarities and compare our results with those of such study. In a second experiment we study more deeply the continuum efficient-inefficient search with

Selective Tuning. We follow three known experiments and as before, compare our results with those.

### **Experiment 1: Color differences**

**Method** In this experiment we study how the model performs in a color similarity search. We simulate an experiment from [Nagy and Sanchez, 1990], who showed that feature search can be inefficient if the differences in color are small. We used the CIE values from their experiments converted to RGB with a fixed luminance (Y) of 0.25. The task is to find the redder circle among 5, 10, 15, 20 and 25 distractors for two conditions: small and large color differences. The target and distractors were randomly positioned on a black background. The least mean squares method was used to fit the straight line into the set of points.

**Results** Four examples are shown in Figure 2.7, where, when there are small differences between the target and the distractors, a large number of attentional shifts are needed to find the target, this number increases with the number of distractors. 2 (a), 4 (b), 10 (c) and 19 (d) shifts of attention are required prior to finding the target. Note that c) is similar to d) but with an extra 5 distractors to show that the number of *wrong* attentional shifts is incremented and the

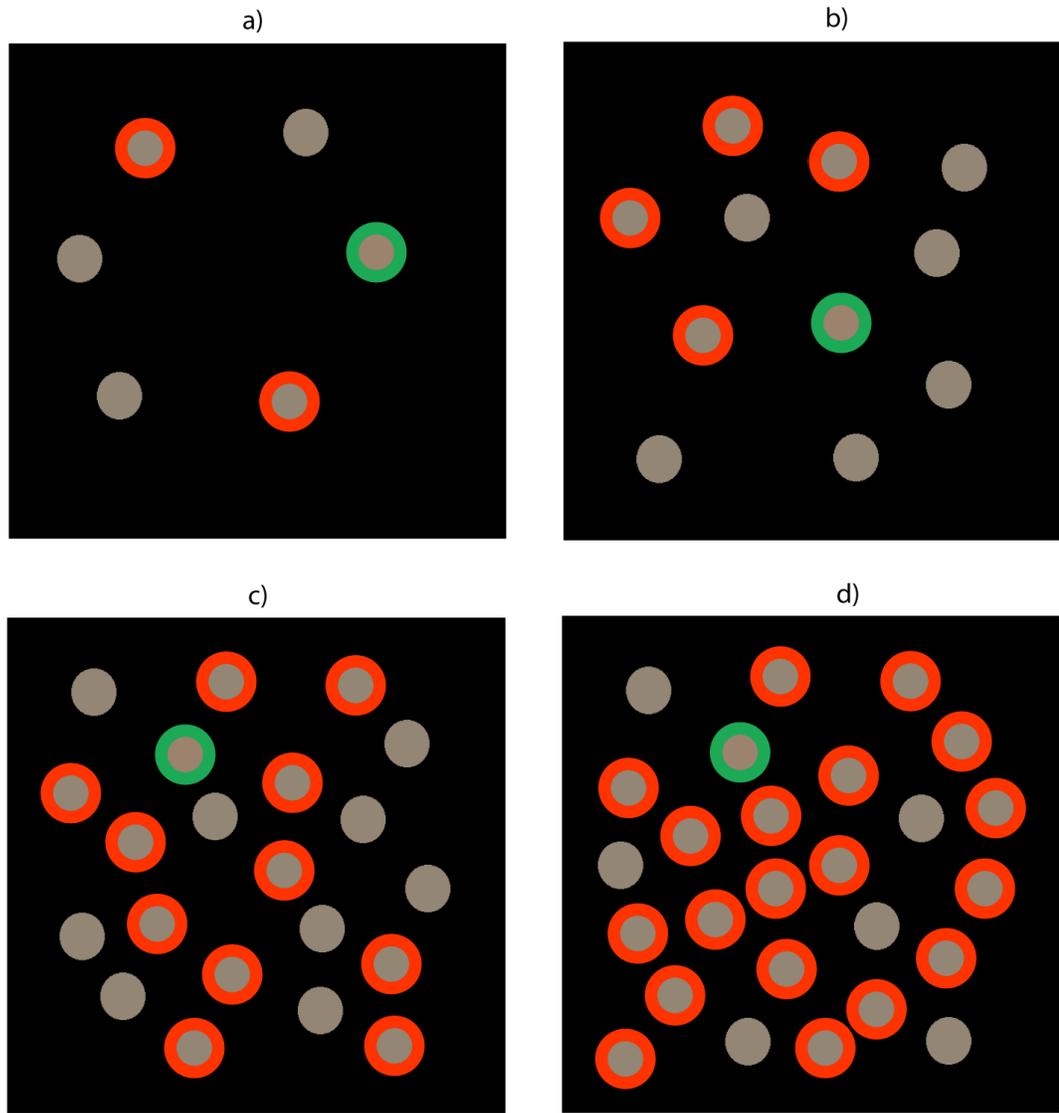


Figure 2.7: Visual Search Results. Example where the target (green circled) and distractors have small color differences for 5 (a), 10 (b), 20 (c) and 25 (d) distractors. Circled in red are distractors attended previous to finding the target, which is under the green circle.

distractors attended may be different.

Figure 2.8a shows the results of running three experiments for each condition and with different number of distractors (5, 10, 15, 20 and 25). Blank dots correspond to the log of the number of covert fixations for each experiment, filled dots is the average. The data for small color difference is in red and in blue is shown the results for the large color differences. Here it is shown how the number of attentional shifts increases as the set size increases for the small color difference condition but remains stable for the large color difference. This experiment reports similar results to [Nagy and Sanchez, 1990] (Figure 2.8b) where color search is inefficient if color difference is small between target and distractors, slope was 0.49 (0.6 in [Nagy and Sanchez, 1990]) and efficient if the difference is large (slope=0.01, same as [Nagy and Sanchez, 1990]). Note that we plot attentional shifts against numbers of distractors and they plot search time with respect to number of distractors. Since this was not a data fitting modeling exercise, only qualitative behavior can be compared. Even so, the comparison is very good.

## **Experiment 2: Feature, conjunction and inefficient search**

[Bichot and Schall, 1999] showed that monkey visual search reaction times are comparable to human [Nagy and Sanchez, 1990], specifically they show

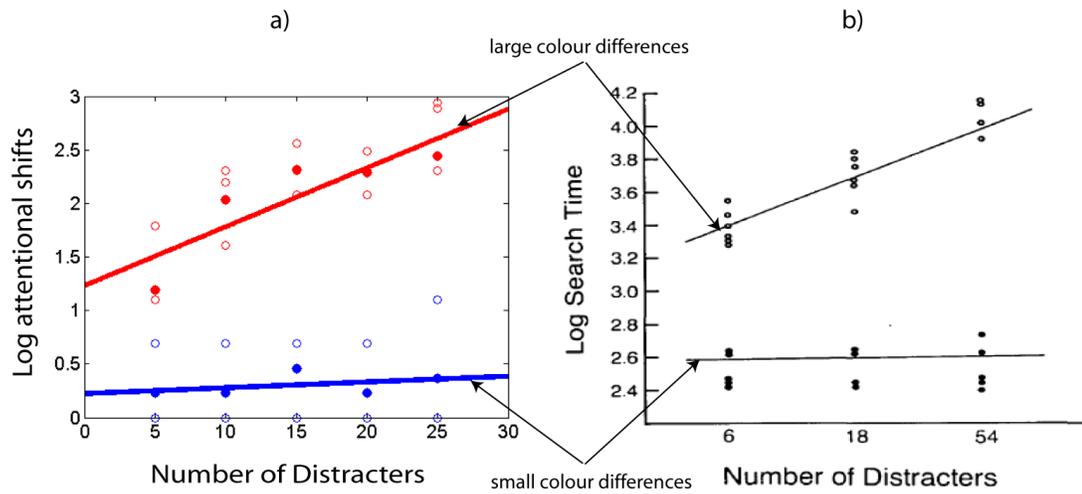


Figure 2.8: a) Number of covert fixations as a function of set size from Selective Tuning. Blank dots are experiment trial results. Filled dots correspond to average from 3 trials. b) Results from [Nagy and Sanchez, 1990]

that the conjunction of two different features (shape and color) is steeper than feature search, but shallower than what was obtained by another study [Treisman and Gelade, 1980]. They report slopes of 3.9 ms/item. On the other hand, searching for a rotated T among rotated Ls, is quite inefficient (20 msec/item) [Egeth and Dagenbach, 1995], and less efficient than conjunction searches. To find a T among Ls is more inefficient than a conjunction search, which is less efficient than a simple feature search.

**Method** In this experiment we study how the model performs in a simple feature search, a conjunction search and an inefficient search. Conjunction search was similar to that of [Bichot and Schall, 1999]. The stimuli were crosses and

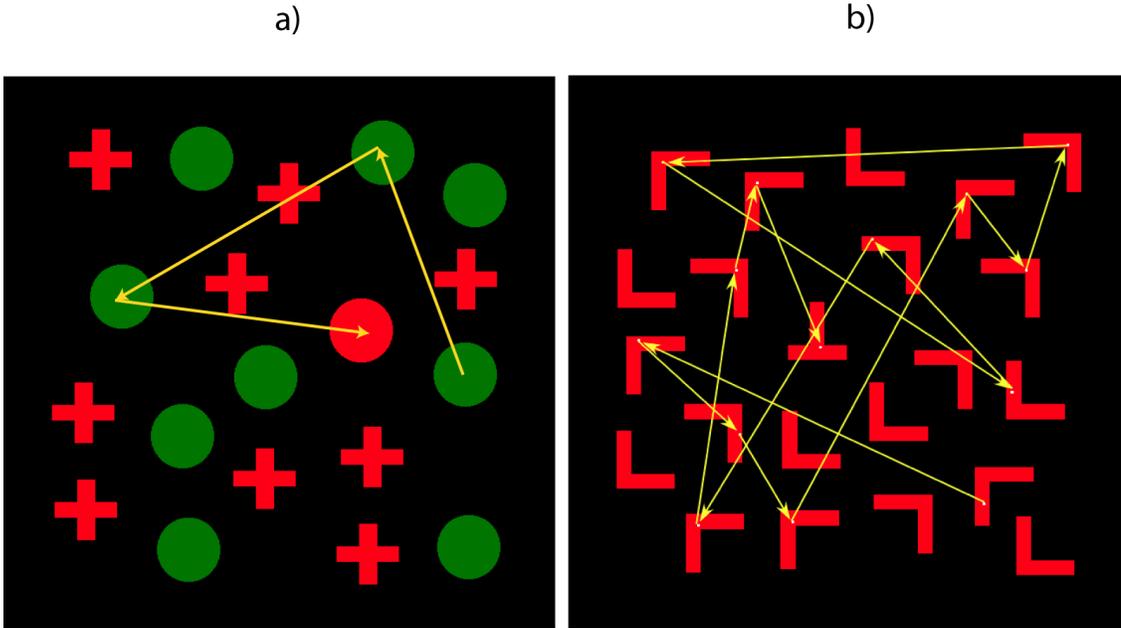


Figure 2.9: a) Conjunction search with 18 distractors, target is the red circle, found at the 4th covert fixation. b) Inefficient search: Find the rotated T among 21 Ls, 14 fixations were needed to find the T.

circles, red or green colored. The task was to find a red circle among green circles and red crosses, here we used 8, 12, 16, 18, 22 and 24 distractors. Feature search was a simplification of the previous conjunction search, that is, to look for a circle among crosses. For inefficient search, a rotated T was to be found among Ls rotated at 0, 90 and 180 degrees, and in this case we used 6, 9, 12, 15, 18 and 21 distractors. Analysis was the same as for previous experiments.

**Results** An example of a conjunction search is in Figure 2.9a and searching for a T among Ls is shown in Figure 2.9b, in this last case more attentional shifts are needed to find the target. Figure 2.10a shows the number of attentional shifts as the set size increases for the feature search (find a circle among crosses), conjunction search (find a red circle among red crosses and green circles) and inefficient search (find a rotated T among Ls). The figure shows how the steepest fitted line is the one corresponding to looking for a T among Ls (inefficient search, slope of 0.49) experiment, followed by conjunction search (slope of 0.36) and feature search is practically flat (slope of 0.01). These results are in accordance with the continuum from efficient to inefficient search psychophysical experiments have shown (Figure 2.10b see [Wolfe, 1998a] for a review).

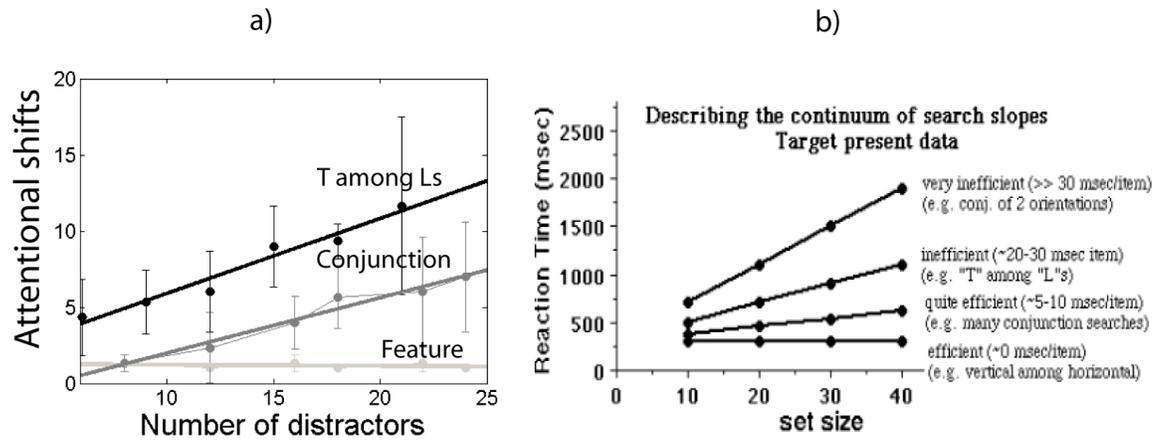


Figure 2.10: a) The number of shifts of attention as a function of set size for feature search (light gray), conjunction search (gray) and inefficient search (black). b) Visual search continuum from [Wolfe, 1998a]

## 2.4 Discussion

The above results show the ability of the Simple Stimulus Recognition Model to perform visual search. The reaction time is shown based on the number of attentional shifts. We performed easy feature search, difficult feature search, conjunction search and inefficient search. The results obtained seem to agree with the increasing degrees of difficulty reported by psychophysical data [Nagy and Sanchez, 1990, Bichot and Schall, 1999, Egeth and Dagenbach, 1995], whose experiments were simulated above. Our experiments seem to agree also with the proposal that search is more efficient when stimuli are more dissimilar [Duncan and Humphreys, 1989] and the continuum efficient-inefficient search

found in the literature [Wolfe, 1998a]. Similar comparable results using the model have been obtained using moving stimuli [Simine, 2006].

The model can differentiate the different types of visual search experiments that have appeared, showing a different efficiency not only between feature and conjunction searches but also more difficult searches (inefficient visual search) as the one described by [Egeth and Dagenbach, 1995]. The behavior of the model agrees with well established models of visual search [Wolfe, 1998a, Duncan and Humphreys, 1989], accounting for a continuum efficient-inefficient search related to the similarity between target and distractors.

The contribution here is of mechanisms that can provide an explanation for visual search performance that have the promise of enhancing performance of recognition algorithms in complex scenes. However this analysis relied on a Simple Stimulus Recognition Model since that was not the main point. The next chapter deals much more deeply into a more sophisticated biologically plausible model of shape representation. Importantly, the result is completely compatible with attentive processing.

### 3 A Model of 2D Shape

In the past decade we have seen a resurgence of interest in shape representation and analysis in the object recognition literature. An example of this is the special section on Shape Analysis in the April 2010 issue of the IEEE transactions on Pattern Analysis and Machine Intelligence. At the same time, an important number of influential studies in neuroscience have shed new light into how the human brain performs the analysis of visual objects. Due to the latter, several models have appeared which are inspired in a higher or lower degree on how neurons achieve object recognition. A good example of this knowledge transfer from neurophysiology to computer science is the compendium of articles edited by [Dickinson et al., 2009]. By understanding how specific simple objects are analyzed by the brain, we may construct subsystems that emulate that behavior. In a not so far away future, by combining those subsystems we may recognize objects “like” a human.

Common to most models is a first step that performs edge-detection in a similar way to some V1 neurons in the brain. But after this first step there is little consensus. First, it is difficult to interpret results from neurophysiology, and even when the behavior of cells in one layer may be determined, it is difficult for modelers to decide how to connect one layer to the next. This is even more true for the later stages, i.e., how do we achieve object recognition in area TE with invariance to translation, rotation, scale, illumination, etc. starting from neurons in V1 that perform edge detection? What if we start with curvatures in V4?

Second, stimuli used in neurophysiology are simple shapes and objects, with no clutter, no variation and under much control to avoid other factors that may influence a cell's response. But from the computer vision community there is some pressure in terms of obtaining results from real-world images, meaning invariance to affine transformations, illumination, clutter, occlusions, etc. As a result some models must resort to complex computational methods such as classifiers, learning and others in order to achieve real-world results. Due to this, even though the starting point is biological plausibility, we find ourselves diverting from it for the sake of results and sometimes reducing the biological plausibility of the model to a pool of Gabor filters.

An interesting study has been recently conducted by [Pinto et al., 2008]. They constructed a simple V1-like model (edge detectors) and showed that such a simple system outperforms several existing object recognition systems for the Caltech 101 database [Fei-Fei et al., 2004]. The model was a neuroscience null model in the sense that it was a simple thresholded Gabor function over 16 different orientations and 6 spatial frequencies. The model did not contain a representation of shape and no mechanisms for recognition under position, size or pose variation. As it was expected, such a simple system performed poorly when tested on a easy task of differentiating just two categories (planes from cars) that introduced real-world variability (position, scale, in-plane rotation and depth rotation). But surprisingly this system performed better than five state of the art systems ([Wang et al., 2006, Grauman and Darrell, 2006, Mutch and Lowe, 2006, Lazebnik et al., 2006, Zhang et al., 2006]) when using the popular Caltech 101 database that included those two categories among the 102 categories in that database. The reason behind this is that even though such a database contains pictures taken from real-world scenarios, it does not include the random variability found in the real world, while the basic two-category test did. The authors warn us about the risks of such biased datasets when performing tests on real-world images.

Among other influential systems, the Scale Invariant Feature Transform (SIFT) [Lowe, 1999, Lowe, 2004] consists of a first stage of difference-of-Gaussian images. After which, distinctive keypoints are constructed from some object of interest such that they are robust to changes in illumination and affine transformations (scale, rotation, position). These keypoints are matched with the keypoints from a scene in order to detect the presence of that object. Hough transforms and least squares were used for object matching in order to achieve a better robustness.

Widely known is the hierarchy of neuronal layers proposed by [Riesenhuber and Poggio, 1999, Riesenhuber and Poggio, 2000, Riesenhuber and Poggio, 2002] with many common aspects to Fukushima's Neocognitron [Fukushima, 1980, Fukushima et al., 1983, Fukushima, 1988]. As described previously, the model consists of five hierarchical levels of neurons. There are two types of layers, one consist of simple units (S layers) and another of complex neurons (C layers), simple and complex layers are interleaved in the hierarchy. At the bottom of the hierarchy there are simple units that achieve edge detection. In successive layers composed of simple units, a simple unit receives inputs from the complex units from the layer below in a Gaussian fashion. A complex unit is fed from simple units at the layer below and at different scales and positions to achieve some level of 2D invariance, from this set of simple

units the strongest is selected. Layers are organized in columns, within a column there are units of the same selectivity but at different scales and positions. The model has been recently extended with the inclusion of an unsupervised learning stage and extended to seven layers [Serre, 2006, Serre et al., 2007]. A summary of the seven layers (S1, C1, S2, C2, S3, C3 and S4) is as follows: S1 contains a pool of edge detectors (Gabor filters) at 4 orientations, 17 sizes and 2 phases. C1 receives inputs from S1 at the same orientation and slightly different positions and sizes. A pool of 10 C1 units at different preferred orientations feed S2 units, the selection of S2 unit parameters is achieved through a learning process from natural images. Then C2 - as before on C1 - select the strongest S2 units at slightly different positions and scales. This process is further iterated in the two new layers (S3 and C3) in order to achieve a higher degree of invariance according to the authors. Finally, S4 is composed of view-tuned cells whose input are C3 units. The authors claim that units in layer S2 become selective to boundary conformations. According to [Pasupathy and Connor, 1999, Pasupathy and Connor, 2001, Pasupathy and Connor, 2002], neurons in area V4 are selective to shapes, and their data from neuronal recording was best fit with a function depending on the curvature and angular position of their boundary conformations. The model of Serre and colleagues does not achieve a shape

representation through any explicit form of curvature computation. Rather, they rely on the repeated convergences layer by layer of approximate straight-line fits to boundaries beginning with edge elements. This strategy does not explicitly include either curvature or end-stopped units, both well-known to exist in the visual cortex. Units that may appear similar may be learned; however, this is not necessarily so and depends on the training data selection.

It is fair to conclude that all these models do not fully explore the possible contributions of intermediate representations as are known in the brain. In this chapter we present a biologically plausible model for Shape Representation. Curvature is considered an important component in order to achieve object recognition in the human brain, along with corners, edges, color, texture and other important features [Connor et al., 2007, Wolfe, 1998a]. A group of studies has shown that neurons in visual area V4 in monkeys are involved with analyzing curvature [Pasupathy and Connor, 1999, Pasupathy and Connor, 2001, Pasupathy and Connor, 2002]. In the hierarchical object recognition pathway, V4 is the area just before the inferotemporal cortex (IT), where object recognition is achieved [Felleman and Van Essen, 1991, Tanaka et al., 1991, Tanaka, 1996].

The model presented here has been built with its biological plausibility in mind, that is, the whole architecture is based on neurons and the computations

that neurons perform at each layer. Our main motivation was to consider biological plausibility at every layer as the starting point. For this reason, we concentrate our efforts on only one part of the object recognition pathway: shape analysis. The focus is on 2D shapes.

Shapes are an important feature for recognizing objects [Wolfe, 1998a]. Shapes are composed of straight lines, corners, junctions and curves. Curvature has been present in many methods and algorithms that achieve object recognition (e.g., [Mokhtarian, 1995, Dudek and Tsotsos, 1997]). But until recently there was no substantial proof that the human visual system analyzes curvature in areas that are involved in the object recognition pathway, such as V4 [Pasupathy and Connor, 1999, Pasupathy and Connor, 2001, Pasupathy and Connor, 2002] and IT [Brincat and Connor, 2004, Tanaka, 1996]. In addition to this, there is evidence that endstopped neurons present in area V2 [von der Heydt et al., 1984] are selective to contours [Dobbins et al., 1989, Parent and Zucker, 1989, Dobbins, 1992]. These works provide the intellectual foundation for the research described in this chapter.

Section 3.1 describes in detail each layer in the model. Section 3.2 shows the results of testing the model with stimuli used in previous single-cell studies. The discussion regarding the characteristics of our Shape Representation model

is presented in section 3.4. Finally, section 3.3 presents the conclusions of the present work.

### **3.1 A Model of Shape Representation**

This section explains the details of the model. A summary of the architecture is presented in Figure 3.1 and briefly described as follows: V1 is composed of simple cells that analyze edges and of complex cells that are the result of the summation of spatially displaced simple cells. V2 contains endstopped cells which respond to variations of straightness. Endstopped neurons are the result of the difference of two simple V1 neurons or one simple and two complex V1 neurons shifted in position. Endstopped cells are the input for local curvature selective neurons. Then, Shape-selective neurons respond to curvature configurations with respect to their position in the neuron's receptive field.

To test the biological plausibility of our model, in the next section the response of our shape-selective neurons are compared with the data from neuronal recordings in area V4 of the macaque monkey's brain.

In what follows whenever a neuron is referred to as model neuron it is one developed for our theory. A neuron referred to without the model adjective is a biological one.

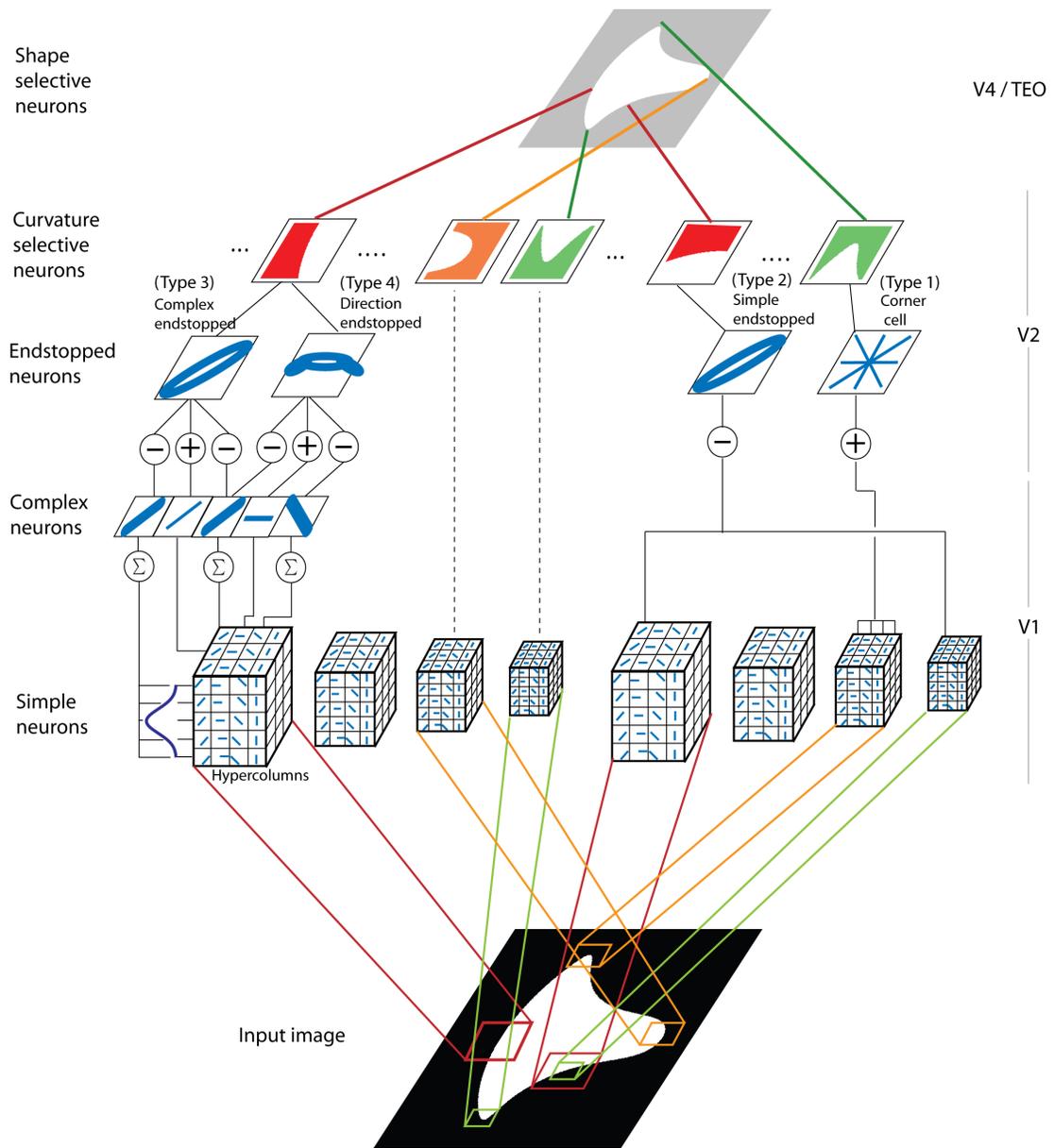


Figure 3.1: Architecture of the Shape Representation model

### 3.1.1 Simple cells (V1)

Simple neurons of visual area V1 are sensitive to bar and edge orientations. Gabor filters [Marcelja, 1980] and Difference of Gaussians have been shown to provide a good fit when modeling simple cells from area V1, although a better fit to neuronal responses has been found with Difference of Gaussians [Hawken and Parker, 1987]. We implemented both types but we obtained better results using Difference of Gaussians as edge detectors in V1 (Figure 3.2), and as a result, we decided to use exclusively the Difference of Gaussians formulation:

$$G(x, y) = \frac{1}{2\pi\sigma_{x_1}\sigma_y} e^{-\frac{1}{2}\left(\left(\frac{x'}{\sigma_{x_1}}\right)^2 + \left(\frac{y'}{\sigma_y}\right)^2\right)} - \frac{1}{2\pi\sigma_{x_2}\sigma_y} e^{-\frac{1}{2}\left(\left(\frac{x'}{\sigma_{x_2}}\right)^2 + \left(\frac{y'}{\sigma_y}\right)^2\right)} \quad (3.1)$$

$$x' = x\cos(\theta) + y\sin(\theta)$$

$$y' = -x\sin(\theta) + y\cos(\theta)$$

where  $\sigma_y$  is the height and  $\sigma_{x_1}$  and  $\sigma_{x_2}$  are the width of each Gaussian function.  $\theta$  is their orientation. The relation between these parameters may be referred to as the aspect ratio  $AR = \frac{\sigma_y}{\sigma_{x_1}}$  and the width ratio  $WR = \frac{\sigma_{x_2}}{\sigma_{x_1}}$ . Size of filters were  $4\sigma_y$ .

Cells in area V1 are heterogeneous, i.e. they are not all uniform. In the

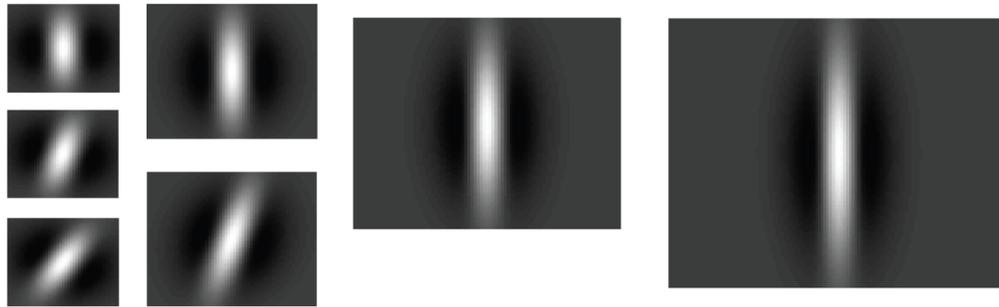


Figure 3.2: Sample of Difference of Gaussians edge detectors. Four sizes, from left to right: 40, 60, 88 and 120 pixels

model, four different groups of simple cells were designed, varying sizes and values of width and length. V1 model cells are organized into hypercolumns. Inside a hypercolumn, cells are spatially displaced and combined into model complex cells as described next (Figure 3.1) . For our experiments, we used 12 orientations and 4 different sizes, this gives a total of 48 types of V1 model simple neurons.

### 3.1.2 Complex cells (V1)

Complex cells have a sensitivity for bars and orientations as well, but their receptive fields are larger than the ones of simple neurons. Hubel and Wiesel [Hubel and Wiesel, 1959, Hubel and Wiesel, 1962, Hubel and Wiesel, 1968] found that simple cells have one or more subfields in which the response is either *on* or *off* while complex cells yield both *on* and *off* responses, which suggest that complex cells integrate the responses of simple

cells.

In addition to this, [Spitzer and Hochstein, 1985] showed that complex cells may be the result of the addition of simple cells along the axis perpendicular to their orientation. Following these biological studies, in our model, a complex cell is the sum of 5 laterally displaced model simple cells within a column. The model complex cell response is given by [Dobbins, 1992]:

$$R_{CX} = \sum_{i=1}^n c_i \phi(R_i) \quad (3.2)$$

$R_i$  is the response of the  $i$ th cell and  $c_i$  is its weight. Model cells are Gaussian weighted with position, with weight inversely proportional to distance to the center.  $\phi$  is a rectification function, where any value less than 0 is set to 0. Cells are proportionally shifted by their Difference of Gaussians parameters, so that the cell separation is related to the size and aspect ratio of the component model simple cells along their preferred orientation:

$$separation = \frac{size}{2AR} \quad (3.3)$$

We compare the response of a model simple neuron with that of a model complex neuron in Figure 3.3 for a stimulus of a 1-pixel wide straight line. Regarding

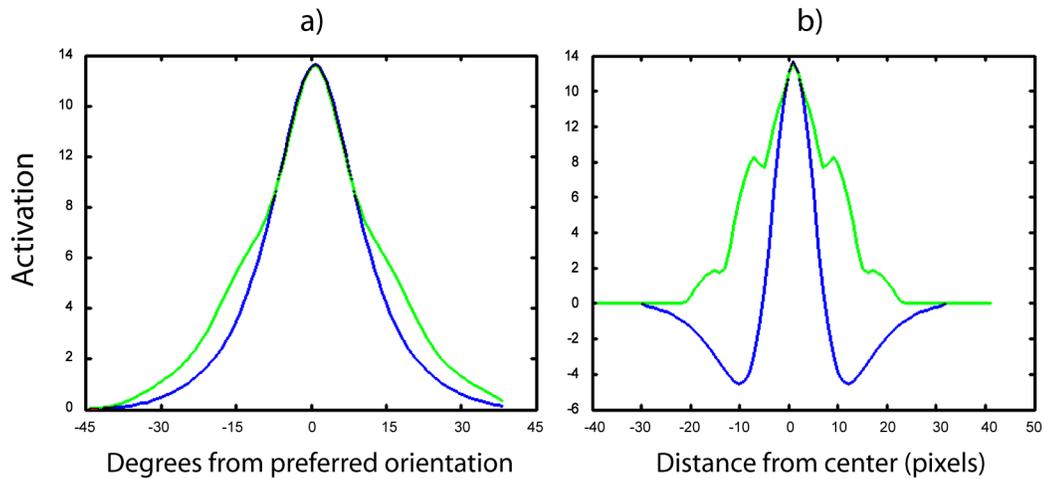


Figure 3.3: Comparison between a model simple neuron response (blue plot) and a model complex neuron response (green plot) for orientation selection (a) and location (b)

orientation selectivity (Figure 3.3a), the model complex neuron (green plot) Gaussian function is wider than that of model simple neurons (blue plot). This results in slightly less sensitivity for orientations and since each integrates five model simple cells. Their receptive fields are larger as well, these two characteristics follow [Hubel and Wiesel, 1959, Hubel and Wiesel, 1962, Hubel and Wiesel, 1968]. When we compare the receptive field of a model complex neuron with the one of a model simple neuron (Figure 3.3b), a model complex neuron yields a positive response for stimuli at more locations inside its receptive field. For both, the highest response is at the center, but at 5 pixels away, for the model complex cell we have a higher than mid-response while for the model simple cell is zero

or close to zero. At ten pixels away, the model complex cell responds with a half-response, while the response from the model simple cell is negative.

The construction of a model complex neuron is depicted in figure 3.4 (zoomed top cell). The orientation of its model simple neuronal components in this case is for  $90^\circ$  (vertical), while the 5 model simple cells are organized perpendicularly (spatially displaced but overlapping) to this preferred orientation, that is  $0^\circ$ . Due to this composition, the output from such model complex cells are a smoothed version of that of model simple cells.

### 3.1.3 Endstopped cells (V2)

V2 neurons respond to contours, both real and illusory [von der Heydt et al., 1984]. A more recent study [Ito and Komatsu, 2004] has found that although V2 neurons are mainly selective for angles and corners, these neurons also showed submaximal responses for bars. Four types of V2 neurons have been implemented in the model, which respond to local curvatures and corners (Figure 3.1):

**Type 1:** Model corner neurons are the result of the sum of the responses of model simple neurons at the same location but with different orientations [Boynton and Hegde, 2004]:  $R_c = \sum_{i=1}^n R_i(\theta)$ .

**Type 2:** Model endstopped cells with simple components integrate the differ-

ence in response of two simple cells of different size at the same location [Kato et al., 1978].

**Type 3:** Model endstopped cells with complex components result from the difference between a simple cell and two displaced complex cells [Kato et al., 1978] (Figure 3.4).

**Type 4:** Model direction endstopped cells, add a rotated component to Type 3 such that they can distinguish between curvature directions, extending the work of [Dobbins, 1992].

We would like to note here that even though neurons of Type 1 and 2 are present in the model we will only provide a brief description since they will not be used on the experiments due to their redundancy for the stimuli used in our analysis. Model corner neurons may be useful for detecting corners but in our experiments they showed support for model endstopped cells with complex components (Type 3). The model endstopped cells with simple components responded similarly to model neurons of Type 3 but were not so selective, so we decided to use the latter exclusively. In summary, the stimuli used in our experiments were best fitted to Type 3 and Type 4 neurons.

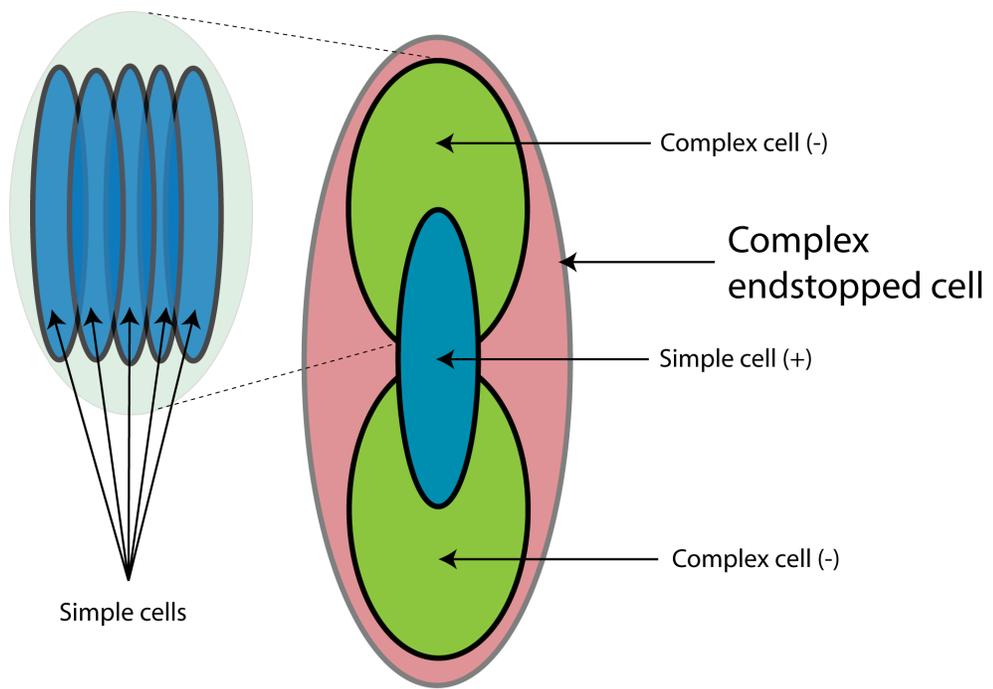


Figure 3.4: Model endstopped cell with complex components.

Endstopped cells - also known as hypercomplex - were fully described by Orban and colleagues in a series of publications [Kato et al., 1978, Orban et al., 1979b, Orban et al., 1979a, Bishop et al., 1980]. In that series of papers, they showed that endstopped cells had different properties from orientation-selective cells and provided a description of those properties as well as a detailed study on the end-zone inhibitory areas that were part of such cells.

Model endstopped cells provide us with a coarse curvature estimation so that we can divide contours into curvature classes. For the design of the model endstopped cells we followed the work of [Dobbins et al., 1987, Dobbins et al., 1989, Dobbins, 1992] which followed Orban and colleagues description and findings about endstopped cells. The aforementioned authors propose an integration of simple [Dobbins et al., 1987] and later of simple and complex cells [Dobbins, 1992] into endstopped cells that are suited for curvature selectivity as well as binocular disparity and motion analysis. We have extended that work with more experimentation into curvature discrimination and found it to be a valid and biologically plausible method of analyzing curvatures. We used Difference of Gaussians since as previously noted we found better fit for the task rather than the original use of Gabor filters.

We had to include extensive testing in order to select the best type of model neurons (combination of model's simple cells or simple and complex) as well as parameter selection. Our tests helped us not only for the task of proper parameter selection but we could also compare the selectivity of the type of model neurons for the job of curvature selectivity. We included a method that would help us on the task of testing the parameter selection (*size*, *AR*, *WR*, Subsection 3.1.1). That method will be called *Calibration* and provided a way of choosing parameters such that their curvature ranges for discrimination do not overlap and do not include any undesired behavior. This calibration consisted of testing model endstopped cells with contours from different curve sizes in order to provide the response of that model endstopped neuron to a range of curvature values.

From a curvature and object detection point of view, it is usually of interest to know if a certain part of a curve is convex or concave. The model endstopped cell receptive fields are small, and thus capture only local information, while in order to determine if a section of a curve is convex or concave more global information concerning the whole of the shape is needed.

Through the local information available to endstopping we may know if a curve is tending to the left, right, upwards or downwards, which may be useful information to be used for later layers in the hierarchy to compute if a contour

is convex or concave with respect to the object. Later in this chapter we provide a means for how this curve orientation may be extracted through endstopping.

We provide next a detailed explanation of the model endstopped cells with simple components, the model endstopped cells with complex components and the model endstopped cells for the direction of curvatures.

### **Endstopped cells with simple components**

These cells are the result of the difference in response of two model simple cells of different size at the same position. Their response to curvature depends on the relation between the size of the two components and the parameters of the Difference of Gaussians as well. A model simple endstopped neuron's response is given by [Dobbins et al., 1987]:

$$R_{ES} = \Phi(c_S\phi(R_S) - c_L\phi(R_L)) \quad (3.4)$$

$\phi$  is a rectification function, where any value less than 0 is set to 0.  $\Phi$  is another rectification function of the sigmoid type, and is given by:

$$\Phi = \frac{1 - e^{-R/\rho}}{1 + 1/\Gamma e^{-R/\rho}} \quad (3.5)$$

where  $c_S$  and  $c_L$  are the gain for the small and large cells respectively. These positive constants normalize the area difference between the receptive fields.  $R_S$  is the response of the small cell and  $R_L$  is the response of the large cell. Sensitivity to curvatures will depend on the sizes of both large and small cells as well as their gain ( $c_S$  and  $c_L$ ). The response of these model neurons as with all the model neurons is band-passed in gain, size and standard deviation (Equations 3.1 and 3.4).

During simulations we found that this type of model endstopped neuron may be well suited for curvatures of small radius, but its selectivity is smaller than that of endstopped cells with complex components, which are explained next. As mentioned previously we had 4 sizes of model simple neurons, we used the smaller cell with combinations of the other three larger sizes in our tests, which considering the 12 different orientations, provided a total 36 types of model endstopped neurons with simple components.

### **Endstopped cells with complex components**

These cells are the result of a difference of neurons like in the previous case. Here, the difference is between a model simple cell and two displaced model complex cells of the same size along the preferred orientation (Figure 3.4). This was

first proposed by [Dobbins, 1992] as a potential curvature estimator, and here it is extended for contour analysis and used for curvature discrimination. The response of such a cell is:

$$R_{ESC} = \Phi[c_c\phi(R_c) - (c_{d1}\phi(R_{d1}) + c_{d2}\phi(R_{d2}))] \quad (3.6)$$

$c_c$ ,  $c_{d1}$  and  $c_{d2}$  are the gains for the center and displaced cells.  $R_c$ ,  $R_{d1}$  and  $R_{d2}$  are the responses of the center and the two displaced cells.  $\phi$  and  $\Phi$  are rectification functions as before. In the experiments, for equation 3.5  $\Gamma=0.01$  and  $\rho$  is the maximum response of the set of neurons for a given scale divided by 8.5, a factor that provided a good normalization approximation for this rectification (Figure 3.5). Displaced cells are shifted 1/2 of their receptive field size. [Dobbins, 1992] provided a study on the best displacement for the centered model, and 3/4 and 1/2 displacement provided good results on that study. We chose 1/2 because it gave the best results in our case.

The center cell is a model simple cell and the displaced cells are model complex cells. An example of this type of model cell is shown in Figure 3.4. The receptive field of an endstopped complex cell includes a simple and two complex cells along their preferred orientation. Its output is given by equation 3.6. The center simple cell has an excitatory effect while the two complex cells (at the top and bottom

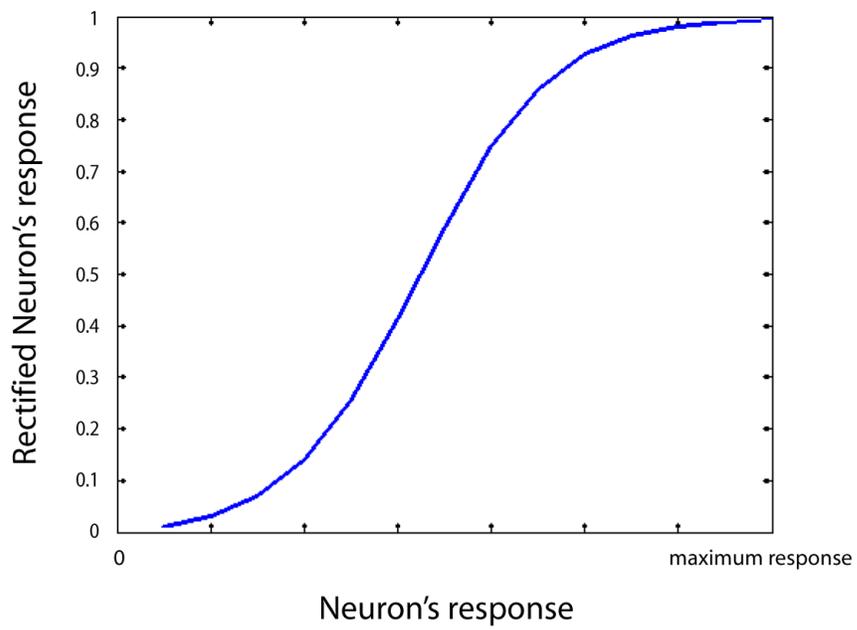


Figure 3.5: Sigmoid-type Rectification function  $\Phi$

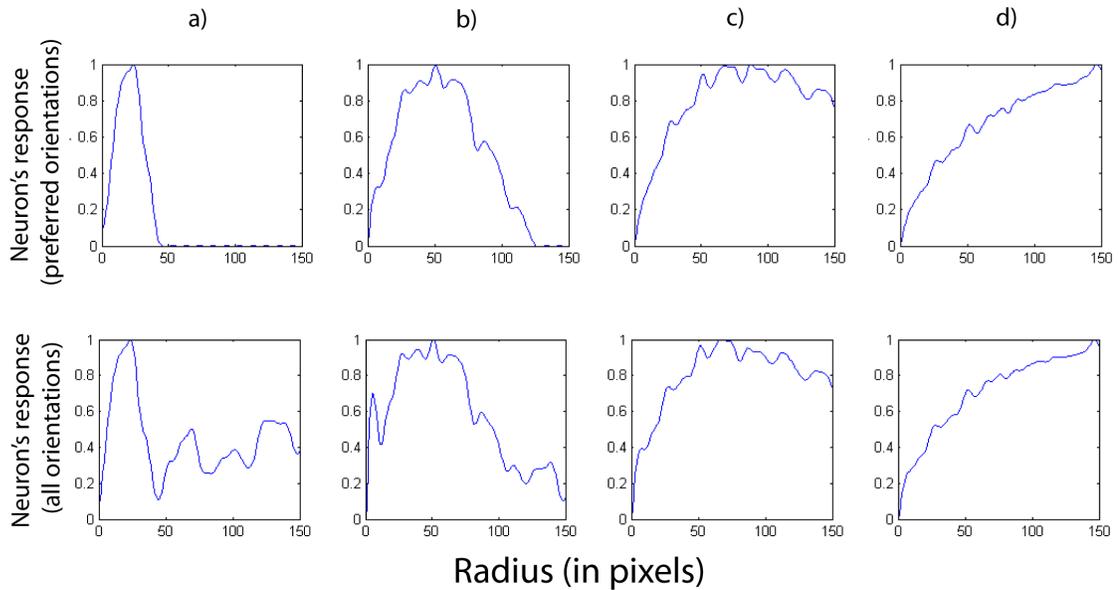


Figure 3.6: Response of the model endstopped cells with complex components to curvatures at preferred (top) and all (bottom) orientations. Simple cell sizes were a) 40, b) 80, c) 100 and d) 120 pixels.  $\sigma_y=(10,20,25,30)$ .  $AR$  (aspect ratio)=(1.15,2,3,4).  $WR$  (width ratio)=2.5 for all cells. Gain  $c=(0.7,0.8,1,2)$ . Responses were normalized for the range [0,1]

in this case) have an inhibitory effect. The response of these cells is similar to the previous type but they show a better selectivity to curvature since their response is not so high at other orientations to the preferred one. This aspect is due to the inhibitory effect of the displaced cells, which we may note are wider than the center cell, following [Orban et al., 1979a]. For this reason, we decided to exclusively use this kind of cell in our experiments.

An example of this is shown in Figure 3.6, where we have four different model neurons (columns) and we can see how they respond to curves depending on their

radius (in pixels) for the preferred (top row) and non-preferred (bottom row) orientations. In our tests we found that the combination from smaller model endstopped neurons is selective for sharper curvatures and the combination of larger cells responds strongly to broader curvatures. The model endstopped smallest neuron (Figure 3.6a) is selective for very high curvatures, the largest model endstopped neuron (Figure 3.6d) was selective to very broad curvatures. We found also that the size ratio between model neurons for a proper curvature discrimination must be 1.25-2 (high ratio is better suited for smaller neurons, a small ratio for larger neurons). The relation between height and width ( $AR$ ) of the Gaussian is large for larger cells and small for smaller cells. Figure 3.6 bottom shows the response of the model cells to all orientations. There is some spurious response to non-preferred orientations at other curvatures, and this is especially true for model cells a and b. The smaller the cell the greater the responses to other curvatures different to the preferred ones, which is no surprise due to the use of smaller size edge detectors providing more spurious responses.

Figure 3.7 shows the response of model complex displaced cells whose components are the model simple neurons from Figure 3.2. V2 model neurons are twice the size of V1 model neurons; in this example they are 80, 120, 166 and 240 pixels. In order to compute the range of selected curvatures, we measured

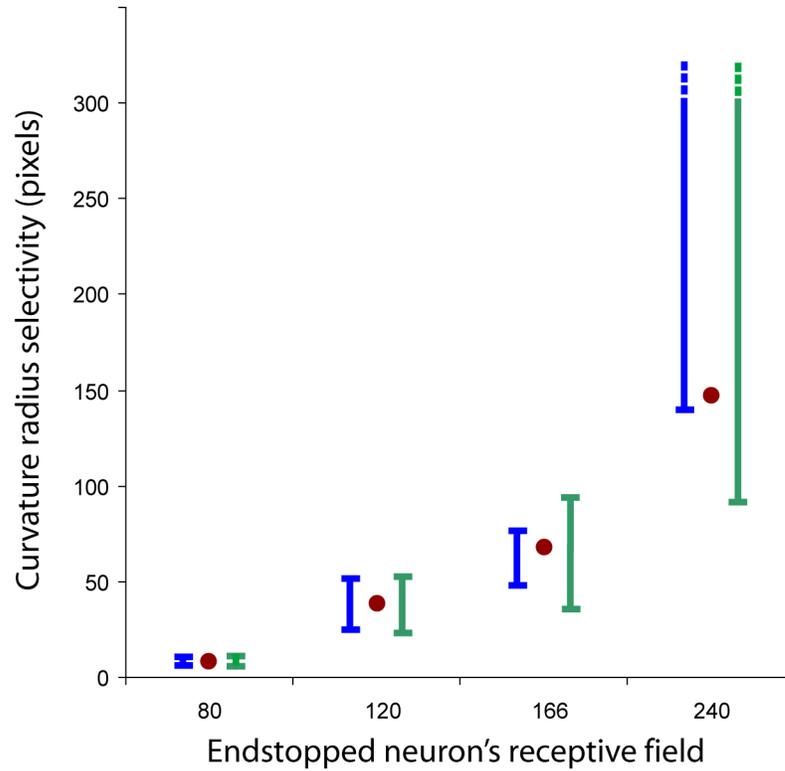


Figure 3.7: Selectivity of endstopped cells with complex components to curvatures considering 90% of maximum response (blue range) and 80% of maximum response (green range). Highest responses are in red circles.  $AR$  (aspect ratio)=(0.7,1.4,2.15,3).  $WR$  (width ratio)=2.5 for all cells. Gain  $c$ =(1.5,1.25,1,3)

the cell response over different circles ranging from radius of 0 to 300 pixels, and we obtained the model cell response at its 80 and 90 highest percentile values. The ranges of selective radius of curvature considering the 90% of the maximum response (red circles in figure 3.7) for these neurons were: 6-11, 25-52, 48-77, 140-300+ pixels (blue range lines in figure 3.7), the 80% maximum responses radius were 5-12, 24-54, 36-95, 96-300+ pixels (green range lines). We can see that there is a high degree of separation regarding curvature responses. If we consider the 90% percentile there is no overlap; some overlap occurs at the 80% between b and c model neurons.

Even though the above analysis is based on the percentile of highest responses from model endstopped neurons, we did not use any threshold or maximum selection regarding the responses of our neurons. The responses from all model endstopped neurons are considered in our model. For example consider a radius of 80 pixels. The response from the model neuron 3.6c would be 80-90% of its maximum response, the response from neuron 3.6d, 70-80% its maximum value, and the responses from neuron 3.6a would be close to zero and from neuron 3.6b, around 50% its maximum response. By using all this information we can have a pattern of responses from our model endstopped neurons that is characteristic of that radius of curvature.

## Direction endstopped neurons

We refer to direction as the path of a curve that is steering upwards vs downward or left vs right respect to the center of the curve. To evaluate the direction, the same model endstopped cells with complex components are used adding a rotated component on each displaced complex cell with opposite directions (e.g.  $45^\circ$  and  $135^\circ$  for the  $0^\circ$  model endstopped neurons). A hint regarding this concept was first proposed by [Dobbins, 1992], which is extended here to all orientations and used on curvatures (Figure 3.8).

Two types of model direction cells are used, and we will use the term *sign* to specify if the curvature is in one direction or the opposite, positive or negative. These different directions are obtained by changing the order of the displaced subtracted neurons.

$$R_+ = \phi[c_c\phi(R_c) - (c_{d145}\phi(R_{d145}) + c_{d2135}\phi(R_{d2135}))] \quad (3.7)$$

$$R_- = \phi[c_c\phi(R_c) - (c_{d1135}\phi(R_{d1135}) + c_{d245}\phi(R_{d245}))] \quad (3.8)$$

where  $c_c$ ,  $c_{d1}$  and  $c_{d2}$  are the gains for the center and displaced cells as before.  $R_c$ ,  $R_{d1}$  and  $R_{d2}$  are the responses of center and displaced cells. The difference here is that the displaced cells are at different orientations of the preferred center simple cell, for the positive sign model endstopped neuron, the displaced model

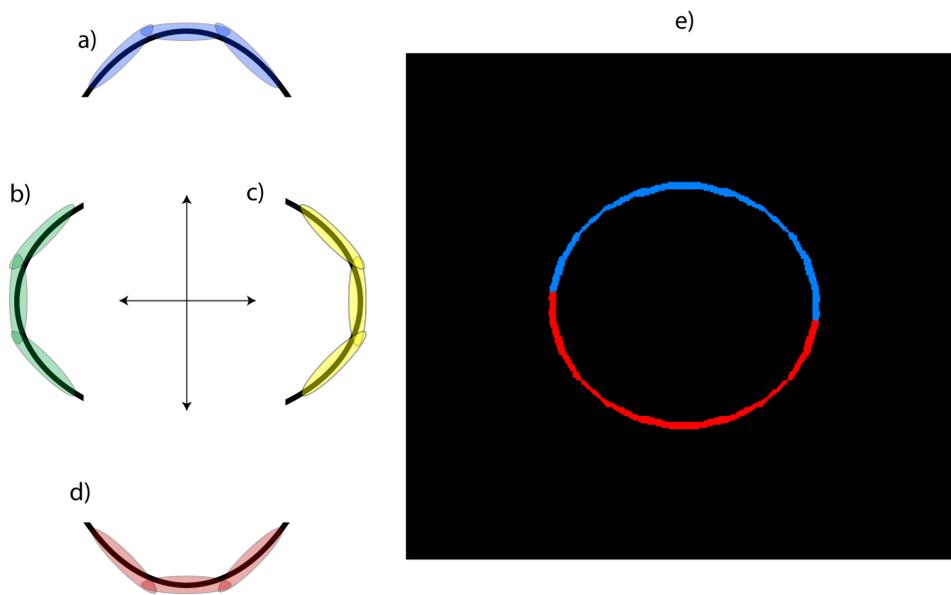


Figure 3.8: Direction selection. Endstopped configuration of cells whose curvature is moving downwards (a), to the right (b), left (c) and upwards (d), e) An example of a circle where only a and d are applied, positive sign selects model endstopped neurons whose direction is downward (blue). Negative sign is curvature upwards (red)

complex neuron  $d1$  is at  $45^\circ$ , while the model complex component  $d2$  is at  $135^\circ$ . For the negative sign model endstopped cell, the order is the opposite. Figure 3.8e shows the response of direction selective cells over a circular shape for direction upwards vs downwards.

### 3.1.4 Local curvature cells (V2)

Curvature cells are obtained due to the neural convergence of model endstopped, corner and orientation-selective cells:

By combining model endstopped cells with complex components and model direction endstopped cells responses, we obtain twice the number of curvature classes than the number of model endstopped cells. For example, if we have four types of model endstopped cells, through the use of the direction of curvature of those cells we obtain eight curvature classes.

$$\begin{aligned} R_{curv_i} &= R_{ESC_i} \cap (R_+ > R_-) \\ R_{curv_{i+n}} &= R_{ESC_i} \cap (R_- > R_+) \end{aligned} \tag{3.9}$$

where  $n$  is the number of model endstopped cell types,  $R_{ESC_i}$  is the response of the model endstopped cell  $i$  and  $R_+$ ,  $R_-$  are the responses of the model direction selective endstopped neurons.

For the case where the response from endstopped cells and corner cells was

small, a high response from a model orientation simple cell meant the contour was a straight line, so its curvature was set to 0.

### 3.1.5 Shape-selective neurons (V4/TEO)

Recent experiments in area V4 [Pasupathy and Connor, 2002] and TEO [Tanaka, 1996, Brincat and Connor, 2004] (Posterior inferotemporal cortex) of the macaque monkey seem to agree in a recognition of objects by parts. In the case of V4 and TEO, those parts would be local curvatures [Pasupathy and Connor, 1999, Pasupathy and Connor, 2001, Pasupathy and Connor, 2002, Brincat and Connor, 2004]. The response to a shape would correspond to the response of the local curvatures of the object. In TEO, some components of local curvatures excite the neuron, and others inhibit its response [Brincat and Connor, 2004].

We consider here a Shape neuron to be in area V4 or TEO. Neurons in areas V4 and TEO share similar characteristics regarding shape analysis [Pasupathy and Connor, 2001, Brincat and Connor, 2004] and selectivity [Boussaoud et al., 1991]. Although similar, TEO neurons show a higher degree of complexity than V4 neurons [Brincat and Connor, 2004]. Our Shape neurons are slightly more complex than just the curvature  $\times$  angular position coding proposed

by [Pasupathy and Connor, 2001] for V4 neurons since they are not only selective to curvatures at angular positions but also to the distance to the center of the neuron or shape. The reason for this added dimension is that the curvature  $\times$  angular position representation alone was too simple for our shape representation. On the other hand they do not show the full complexity of a TEO neuron since they do not include inhibition from local curvatures and nonlinearity regarding curvature combination.

Our Shape cells integrate the responses from model endstopped neurons in area V2, that is, local curvatures. The proposed response of a model's Shape neuron is:

$$R_{shape} = \sum_{i=1}^n c_i R_{curv_i}(\lambda) \quad \lambda = \max_{j=1}^m (\lambda_j) \quad (3.10)$$

$$c_i = \frac{1}{2\pi} e^{-(x^2+y^2)}$$

where  $R_{curv_i}$  is the response of the  $i$ th model curvature neuron from the set of all possible  $n$  model curvature neurons at the preferred curvature direction ( $\lambda$ ) inside the Shape neuron receptive field, and  $c_i$  is a Gaussian weight that would account for partial excitation depending on the selective curvature in distance - angular position (Figure 3.9a).

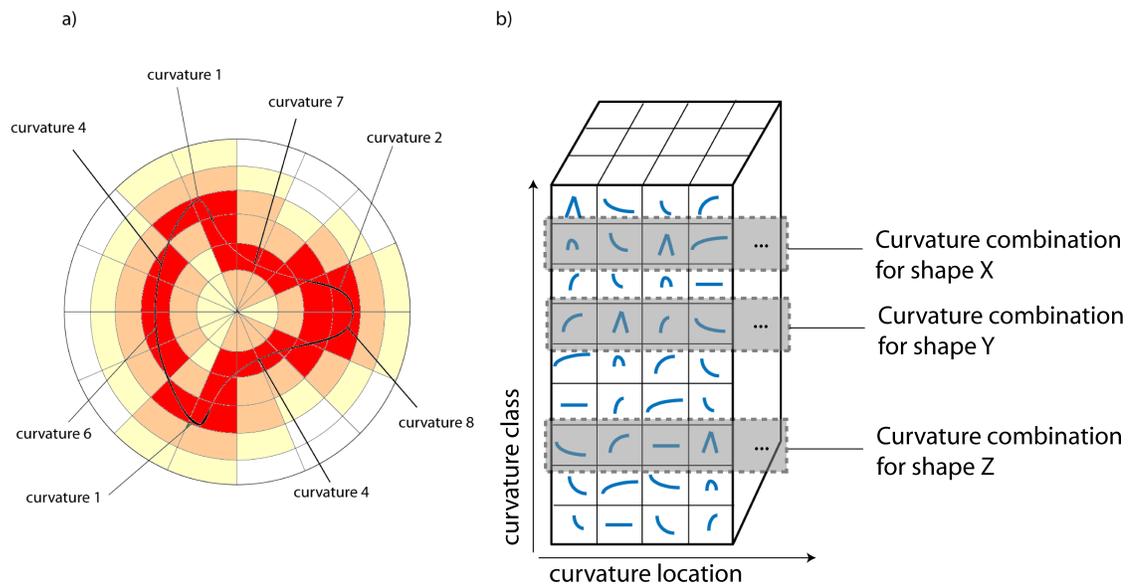


Figure 3.9: a) Shape-selective neurons respond to different curvatures at different positions. The response is maximal when those curvatures are present at their selective positions (red). If they are in nearby positions the neuron provides some response as well (orange and yellow) b) The curvatures at those positions are organized into hypercolumns.

A model Shape neuron will respond to a shape, and depending on how close the stimulus is to its selectivity, its response will be stronger or weaker. The features used here are simple curvatures at different positions and model Shape neuron responses depend on their combination.

### **Shape neurons are organized into hypercolumns**

We propose that Shape cells are organized into a hypercolumn as seems to be common in many other areas of the visual cortex (see [Ts'o et al., 2009] for a review), such as V1 [Hubel and Wiesel, 1968], V2 [Ts'o et al., 2009], IT [Tanaka, 1996] or MT [Deangelis and Newsome, 1999]. [Tanaka, 1996] already proposed an organization into columns in area TE, even though it has not been demonstrated for areas V4 and TEO (the areas preceding TE), it would be no surprise that these areas were organized into columns as well. TE has 1,300-2,000 columns which puts a limit on the number of combinations that can be represented [Fujita, 2002].

We propose that in one dimension of the shape hypercolumn, curvature classes are represented, while the other dimension may encode a shape as the result of the location of their curvature components (Figure 3.9b) since these are the main features to which V4 neurons are selective [Pasupathy and Connor, 2001, Pasupathy and Connor, 2002]. Next, we describe how a Shape neuron respond

to those features: curvatures and their locations.

**Response of a Shape neuron in Curvature space** In addition to being sensitive to location inside the Shape's neuron receptive field, a Shape neuron has a response depending on how close the stimulus is to its curvature selectivity. For example, consider a model neuron selective to a sharp curvature at the top left. This neuron would respond maximally when that feature is present at that specific location, but it would respond also to a broader curvature at that location with a lower value and would have a small response to a very broad curvature or a straight line.

Shape neurons exhibit band-pass tuning for curvature information. Their responses achieve a peak at a specific curvature, then decay providing a decreasing response for curvature values of increasing distance. No response is provided for curvatures very far from the optimal. The Shape neuron in our example is then selective for those model endstopped neurons that respond strongly to sharp curvatures at that position. Since a model endstopped neuron with a high response to a sharp curvature has also some response to a slightly broader type of curvature (Figures 3.6 and 3.7), Shape neurons will not provide a binary response but a range of responses depending on the distance between curvatures

in curvature space.

**Response of a Shape neuron based on Curvature locations** Features (curvatures) inside the Shape neuron are weighted with respect to a factor  $c_i$  (equation 3.10) depending on how close the desired curvature is to the desired position. Let's continue with our example of a Shape neuron selective for a sharp curvature at the top left. This model neuron will have a high response to any stimuli that contain such sharp curvature at that position, but some response will still be elicited in a nearby position, e.g. a sharp curvature at the top mid-left, but no response will be obtained for a sharp curvature present at far away positions (e.g. the sharp curvature is at the bottom).

The curvatures that fall into the preferred cell's positions are considered in their full value (red in Figure 3.9a), but if they fall close, they are weighted in a Gaussian manner depending on how far from the preferred position they are (orange and yellow in Figure 3.9a). For this task we used polar coordinates [Pasupathy and Connor, 2002], that is, the radial distance to the center of the Shape neuron and its angular position. In order to reduce the complexity of computing pixel by pixel and degree by degree, we created bins that are variable in size, depending on the step radial distance and the step angular position. For

our experiments we found 10 pixels for the former and  $4^\circ$  for the latter to provide good results.

In the words of Pasupathy and Connor [Pasupathy and Connor, 2002]: *The population code for shape has to accomodate the virtual infinity of possible objects as well as the variability of a given object's retinal image.* Our model Shape neuron has the capability of representing that virtual infinity of objects: If we consider that our stimuli are inside  $400 \times 400$  pixel images, for the bin size selection given above this gives a total of 1,800 possible curvature parts inside a Shape neuron receptive field. Since we use 8 curvature classes, when we consider any possible combination of curvature/location, our model can represent a maximum of 14,400! possible configurations of stimuli. If we take into account Gestalt properties such as continuity, closeness and others, that number can be reduced to reflect only realizable configurations. The point here is that this representation is sufficiently rich to enable coding of a wide variety of shapes and task knowledge or experience will help determine the relevant subset.

### **3.1.6 Size of neurons**

Here we comment about the relative sizes of our model neurons and compare them to neurobiological measures. In a similar way to the visual cortex hierarchy, in

the model as we go higher in the hierarchy, model neuron receptive field sizes are larger. We consider that an absolute value may be not very useful since they are always dependent on the screen resolution. For a classical psychophysical distance of 30 cm at a typical resolution of  $1280 \times 1024$  pixels, a  $0.1^\circ$  V1 neuron would correspond to 3 pixels, a Gabor filter of size 3 pixels will not return any useable values. Studies with monkeys usually make use of larger distances, which would change the number of pixels.

We think that the pixel-degree equivalence is important for psychophysical and neurobiological studies in order to make sure a stimulus shown on a digital screen covers a specific size of receptive field. But from a modelling point of view, it is more interesting to study the sizes of model neurons relative to each other.

In our case, receptive fields for the Shape neurons (V4/TEO) are  $400 \times 400$  pixels. V4 neurons comprise a receptive field of 2 to  $4^\circ$  in diameter [Felleman and Van Essen, 1991]. TEO neurons receptive field sizes range from  $4^\circ$  for primary cells to  $10^\circ$  for elaborate cells [Tanaka et al., 1991]. Our shape neurons may be considered primary cells or regular V4 cells. As a result, 400 pixels would correspond to  $4^\circ$  in our case.

V2 neurons receptive fields are 0.2 to  $2^\circ$ . Our V2 model neurons are  $80 \times 80$  pixels to  $240 \times 240$  pixels, which in relation to our Shape-selective cells,

would correspond to 0.8 to 2.4°. V1 neurons receptive fields are 0.1 to 1° [Hubel and Wiesel, 1965, Schiller et al., 1976]. In our case, our model V1 neurons would expand 0.4 to 1.2° ( $40 \times 40$  pixels to  $120 \times 120$  pixels).

We can summarize that the relative sizes in our hierarchy are quite similar to the ones in the visual cortex.

### **3.1.7 Summary of the architecture**

A summary of the architecture is presented in Figure 3.10: Model simple neurons perform edge analysis. In the experiments, 4 sizes and 12 orientations are used. Responses from model simple neurons are integrated into model complex neurons receptive fields. By combining model simple and complex neurons we achieve endstopping. Responses of model endstopped cells are used to get different curvature classes (8 in the experiments).

The main element in this architecture is that of Shape-selective neurons, they represent curvature parts in a curvature  $\times$  position (radial and angular) domain. The possible number of shapes that may be represented by our model Shape neurons is very large, given the limited type of neurons at each level of the architecture. A Shape neuron has a response depending on the position and curvature of the stimulus component parts as explained in the previous section.

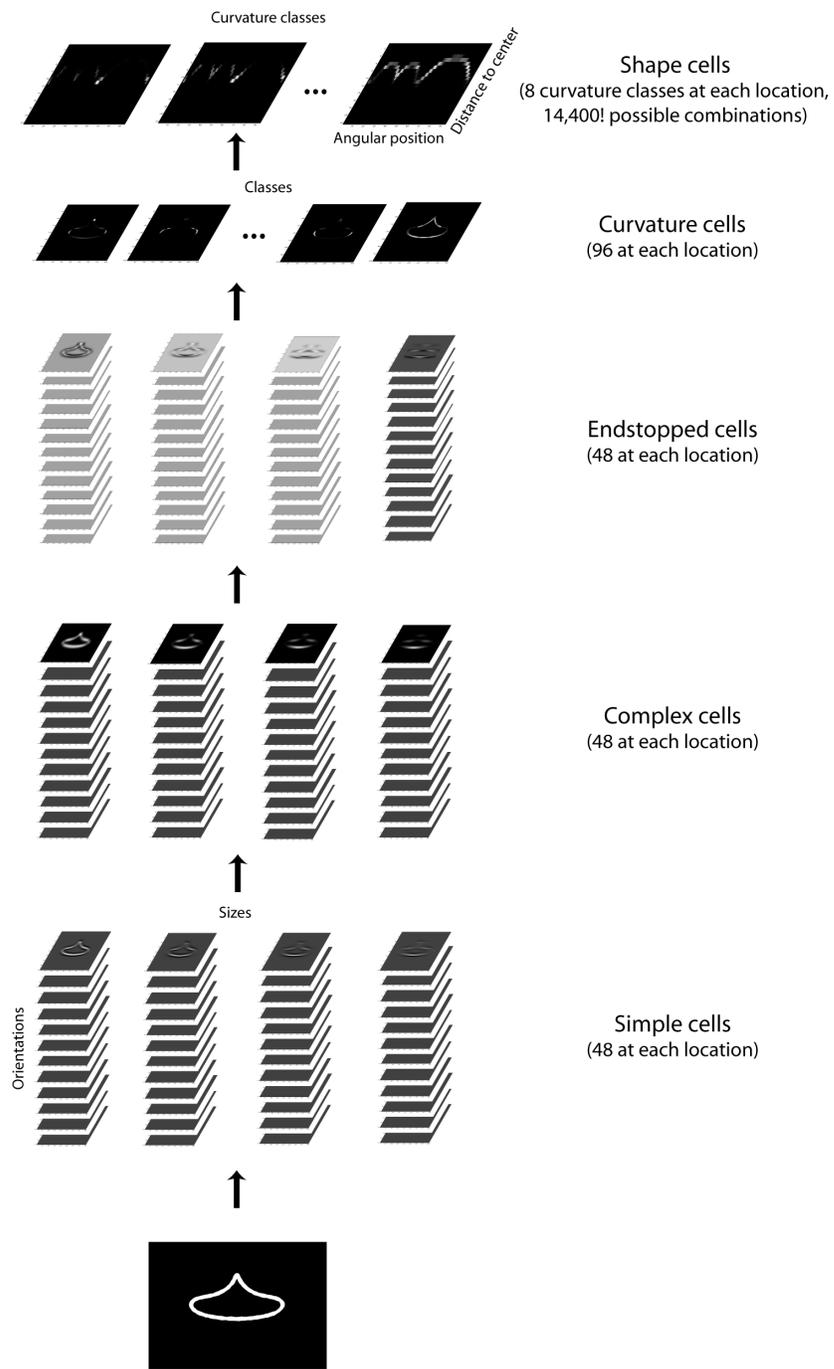


Figure 3.10: The hierarchy of 2D shape processing. See text for explanation

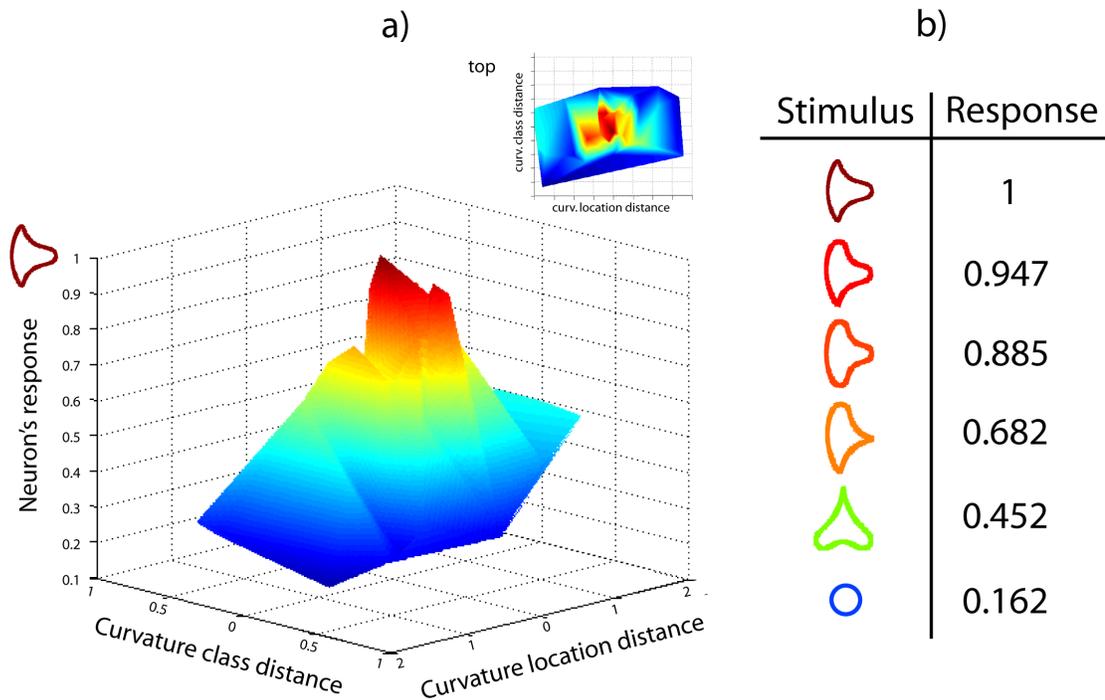


Figure 3.11: Shape-selective neuron tuning profile for location and curvature

The response to a shape from a Shape-selective neuron depends on how similar a stimulus is to a preferred stimulus. An example of this is shown in figure 3.11a that shows the tuning profile of a neuron selective for the stimulus on the left. The responses to other stimuli is plotted relative to their curvature and location distance. Figure 3.11b shows several examples for this model neuron, the closer the stimulus is to its preferred stimulus, the higher its response is.

## 3.2 Experiments

We performed two sets of experiments in which we compare the performance of the model Shape neurons with neurons in area V4 of the macaque’s visual cortex. In the first set we study the capability of our model Shape-selective neurons to represent complete shapes as aggregates of boundary fragments in curvature  $\times$  angular position domain.

Our second set of experiments compares the normalized responses of the model Shape neurons with those of neurons in area V4 over a set of 366 stimuli. For most cells in area V4 of the macaque, shapes evoking strongest responses are characterized by a consistent type of boundary configuration at a specific position within the stimulus [Pasupathy and Connor, 2001]. We show that this behavior is compatible with the model Shape-selective neurons constructed as explained in the previous chapter.

### 3.2.1 The curvature $\times$ angular position representation

#### Motivation

We wanted to first test the capability of our model to encode curvature in the Curvature  $\times$  angular position representation proposed by Dr. Pasupathy and

colleagues. From the results of their experiments, Pasupathy and Connor proposed that the representation of shapes in V4 are curvatures as parts of boundary patterns relative to their position in the object.

Pasupathy and Connor recorded from neurons in area V4 of the macaque monkey. The response pattern was quantified by using a 2D Gaussian tuning function. Figure 3.12 shows an example extracted from their work. Figure 3.12a shows a Gaussian shape tuning function for a cell whose selectivity is a sharp convex curvature at the top ( $90^\circ$ ). The horizontal axis represents angular position of the boundary fragment relative to the object's center of mass, the vertical axis represents boundary curvature. Negative values correspond to concavity and positive values to convexities. Each stimulus was decomposed into four to eight contour fragments of relatively constant curvature. This is exemplified in Figure 3.12b, the stimulus is in the center, the white line surrounding it shows the curvature values relative to angular position.

Here we compare our model's performance against the results from [Pasupathy and Connor, 2002]. In order to construct the stimuli, a Matlab program was kindly provided by Dr. Pasupathy. The stimuli were constructed combining convex and concave boundary elements to form closed shapes. Boundary elements include sharp convex angles, and medium and high convex and concave

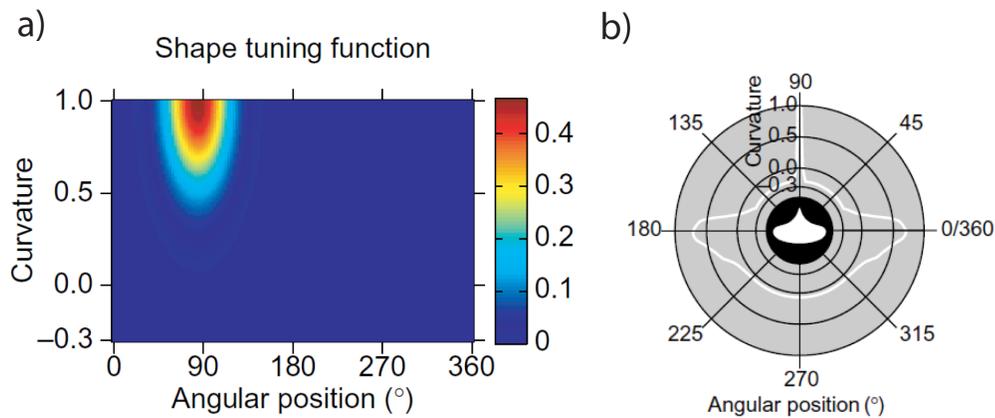


Figure 3.12: a) Pasupathy and Connor’s Gaussian shape tuning function describing the response pattern of a neuron b) Pasupathy and Connor’s proposal, the object’s white line shows the different values of curvature relative to angular position for a squashed raindrop stimulus (icon at the center).

curvatures. The combination of these boundary elements gave rise to 49 different stimuli. Stimuli were composed of white edges against a black background, the inside was black as well but it is shown from now on in our figures as white-filled for illustration purposes.

The data corresponding to V4 neurons for our comparison was provided as well by Dr. Anitha Pasupathy to perform this study.

## Methods

Experiments were run in Matlab in a Mac G5 PowerPC computer with a screen resolution of  $1280 \times 1024$  pixels. The input to the model is a gray-value image.

Images used are  $400 \times 400$  pixels, a shape would span  $300 \times 300$  pixels and correspond to the stimuli used in the aforementioned study. Size of V1 model simple neurons are 40, 60, 88 and 120 pixels, their corresponding values for  $AR$  are 0.7, 1.4, 2.15 and 3 respectively,  $WR$  is 2.5 for all model neurons. 12 orientations were implemented.

For the integration into model endstopped neurons, the values of gain  $c$  were from the smaller to the larger cell: 1.5, 1.25, 1, 3. For the chosen parameters, cells respond (90% of their maximum value) to the following ranges of curvature radius: 6 to 11, 25 to 52, 48 to 77 and 140 to 301 pixels.

**Obtaining convexity and concavity** As explained in Section 3.2 we provide model endstopped neurons that can distinguish the direction of curvature but these neurons cannot specify if a contour belongs to the convex or concave part of a shape. On the other hand, responses to the stimuli from [Pasupathy and Connor, 2002] are reported using convex and concave curvatures (see Motivation). In order to perform our comparison, we must transform the output from our direction endstopped cells into convex and concave responses.

Using local detectors only there is not enough information to know when a contour is concave or convex, and we need some global information to achieve this

task. [Pasupathy and Connor, 2002] use the center of the object as an important quality in the representation of curvature in V4 neurons and we will follow their lead. We only use that information together with our model direction selective endstopped neurons to obtain a value of concavity and convexity.

The method to compute convexity/concavity uses our model endstopped direction neurons and the centroid of the object i.e., the center of mass of the region. To obtain convexity, the centroid of the shape is first obtained for the reasons explained above. Even though local curvature detectors have no information regarding the shape, a V4 model Shape neuron contains that information and local curvatures can then be transformed into actual curvatures with respect to the center of the object. This solution is similar to how MT neurons solve the aperture problem in order to respond to the actual direction of motion [Pack and Born, 2001].

Following this strategy, 8 polar regions are considered in increments of  $45^\circ$ . The first area ( $0^\circ$  to  $45^\circ$ ) is convex provided model direction neurons are to the left, any other condition means it is concave. The second area ( $45^\circ$  to  $90^\circ$ ) is convex, provided model direction neurons are downwards, for any other condition, is concave. For each area, the direction of its model neuron is considered in order to be convex or concave and a map of convexity and concavity is obtained based

on the model neurons directions and the polar area they belong relative to the centroid of the object.

**Curvature classes** The 4 types of model endstopped neurons and the curvature direction selective neurons lead to eight curvatures. In order to compare with [Pasupathy and Connor, 2002] we transformed the curvature radius pixels to which our model endstopped neurons were selective into the squashed curvature values used by Pasupathy and colleagues.

Pasupathy and Connor’s stimuli use higher values for convexities than concavities (their range of curvatures is 1 (convex) to -0.3 (concave)), for this reason we took the higher values of the ranges for the convex contours provided before (*radius*=11, 52, 77, 301 pixels), and the lower values for the concave contours (*radius*=6, 25, 48, 140 pixels). We transformed the pixel values into curvature values considering the resolution of the screen and that the definition of curvature is  $1/radius$ . We finally used the *squashed* curvature equation found in [Pasupathy and Connor, 2001]:

$$c' = \frac{2.0}{1 + e^{-ac}} - 1.0 \quad (3.11)$$

After performing the previous transformations, eight curvature classes were

considered:

1. Very high curvatures/corners,  $c' = 1$
2. High convex curvatures,  $c' = 0.74$
3. Medium convex curvatures,  $c' = 0.45$
4. Low convex curvatures,  $c' = 0.17$
5. Straight lines,  $c' = 0.0$
6. Low concave curvatures,  $c' = -0.1$
7. Medium concave curvatures,  $c' = -0.3$ , and
8. High concave curvatures,  $c' = -0.4$ .

To consider a curvature as class 5 (Straight lines), its response to any of curvature or corner cell is to be less than 20% the maximum response for any curvature/corner cell and have a high response to an edge detector, that is, a V1 model simple neuron.

The maximum curvature responses from model enstopped neurons was integrated at different angular positions (in  $12^\circ$  steps) with respect to the center of the object and data was fitted using a spline in order to obtain a function to compare with [Pasupathy and Connor, 2002].

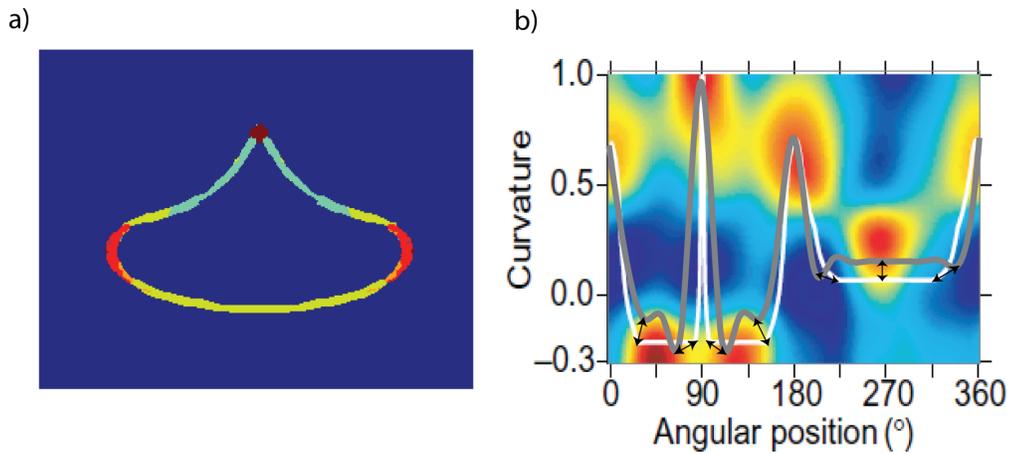


Figure 3.13: a) Response to curvature from the model endstopping to the squashed raindrop stimulus (Figure 3.12b), dark red corresponds to very high curvature, red to medium-high curvature, blue to medium curvature and yellow to broad curvature b) Population response from Pasupathy and Connor (colored surface), curvature  $\times$  position veridical function (white line) and the model's function (gray line), arrows mark the difference between the veridical function and our model.

## Results

Figure 3.13a provides an example of curvature discrimination with the outputs from endstopped cells integrated into a Shape neuron. The stimulus is a squashed raindrop shape (Figure 3.12b). Figure 3.13b shows the a comparison between Pasupathy and colleagues work and the output of the model for that stimulus. The figure includes the Gaussian shape-tuning function describing the response pattern (colour surface). The vertical axis represents boundary curvature and the horizontal axis represents angular position of boundary fragments with respect to

the shape's center of mass. The white line plot corresponds to veridical curvature function, which fits with high accuracy the neural population data. Our results are shown in the grey line plot. Note that it fits the neural data well and it is very close to the veridical plot (in white).

Figure 3.14 shows the results to all the stimuli from [Pasupathy and Connor, 2002]. For each stimulus (left column of each panel), Pasupathy and Connor's results (colored graphs with white plots, center column of each panel) are compared against our results (white background plots, right column of each panel). It can be seen that the results are quite similar to those obtained for neurons in area V4 of the visual cortex.

There are some slight differences between our and Pasupathy and Connor's curves and those will be discussed here. The first is that peaks for very high curvatures are narrower on their fits than ours (e.g. stimulus 39). The reason for this is that although our model small endstopped neurons respond very strongly to very high curvatures and corners, the spline fit may smooth the responses. Another difference may be due to convexity and concavity for some of the stimuli (e.g. stimulus 1). As mentioned before, we perform an easy transformation from our direction-selective neurons to convex/concave selection by just using the centroid. Even though this transformation works for most cases, additional global

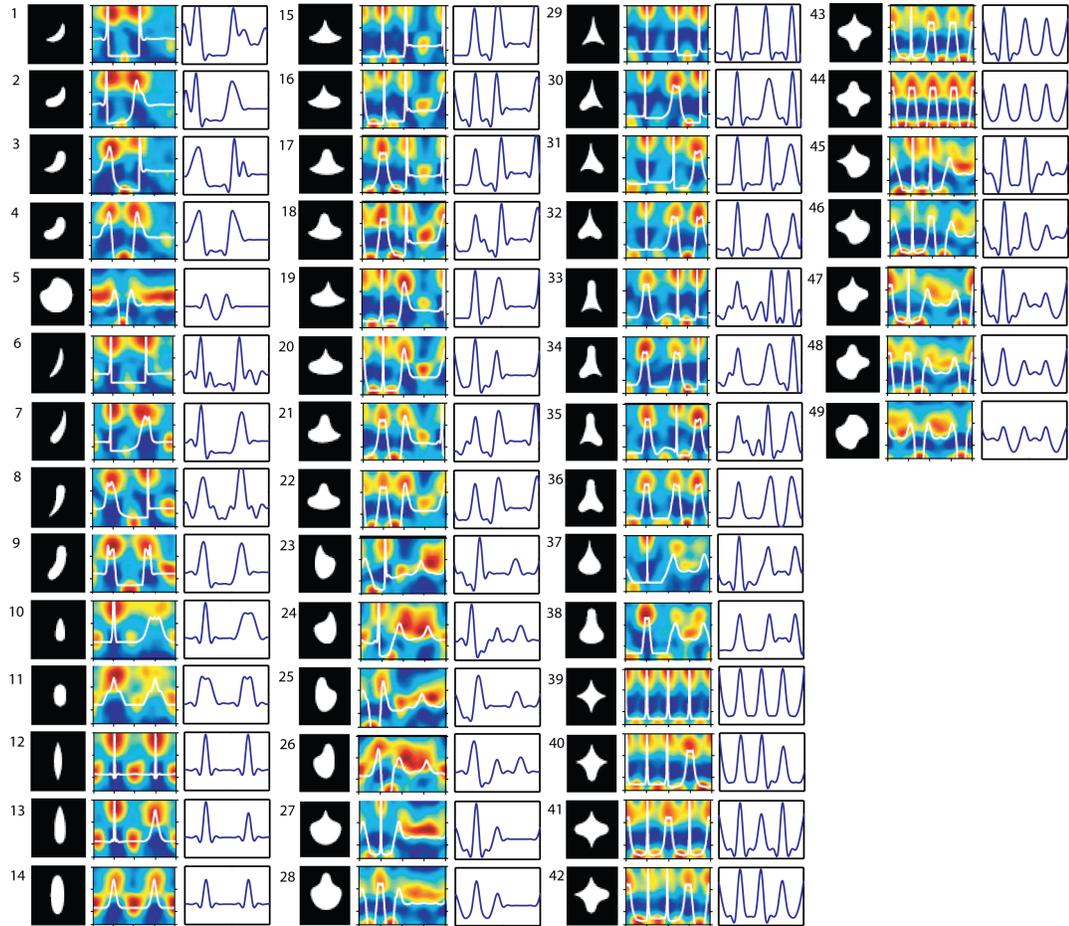


Figure 3.14: For every stimulus (left image), Pasupathy and Connor's fit (white plot, center) and summed population neural responses (colored values, center) was compared to our model's response (right).

information may be needed.

Some concavities are higher than the ones reported from our model (e.g. stimuli 4 and 5). Consider that our model neurons are not curvature detectors in the sense of providing a numeric value but discriminators among sharp, medium-sharp, medium and broad curvatures. These abnormalities may be due to the transformation into specific curvature values, and as such, is prone to slight differences (see methods). Finally, we can see that our plots are not as smooth as the real plots corresponding to the stimuli. At some angular positions our plots show small bumps this is an effect of sampling (we use only 12 orientations), in fact those bumps occur at those orientations and provides a limitation on shape representation.

Nevertheless, we observe here that the model performs closely to neuronal data, fitting very well in almost every case the Gaussian shape response pattern (colored graph). The peaks of our model are highly correlated to the cells response patterns. Another important observation of these results is that the shapes of the model plots are very close to the ones from [Pasupathy and Connor, 2002]. This fact is further supported by the quantitative analysis provided next, where we measured the difference between the results from the model and the data from neurons that Pasupathy and colleagues used to draw their plots.

The original data provided included the angular position and curvature for four to eight different angular positions. For such a task we averaged the Euclidean distance between these four to eight angular position values (Figure 3.12b arrows) and the results from our model in curvature and angular position. Both terms were normalized to 1. That is, for those position values provided in Dr. Pasupathy's data we measured how far our values were in terms of curvature and position (arrows in Figure 3.13b), these distances were averaged over the total number of values. The worst case scenario would be one where the plot provided by the model would be completely the opposite from the plot in [Pasupathy and Connor, 2002] (distance=1). The case where our results completely overlap Pasupathy's plot would mean that Shape neurons have the same exact response pattern as real cells (distance=0).

Normalized distances for each one of the 49 stimuli in figure 3.14 (left column) are shown in the blue bars of figure 3.15. The maximum error is 1 and a perfect match is 0. Total average error for all stimuli was 0.074 (stdev = 0.037). In order to further test the validity of the analysis, we compared the results from our model to each stimulus with those data corresponding to all the other stimuli different to the selected one (e.g., stimulus 1 model values vs data from the other 2-49 stimuli), if the analysis is correct, we should obtain much higher error values

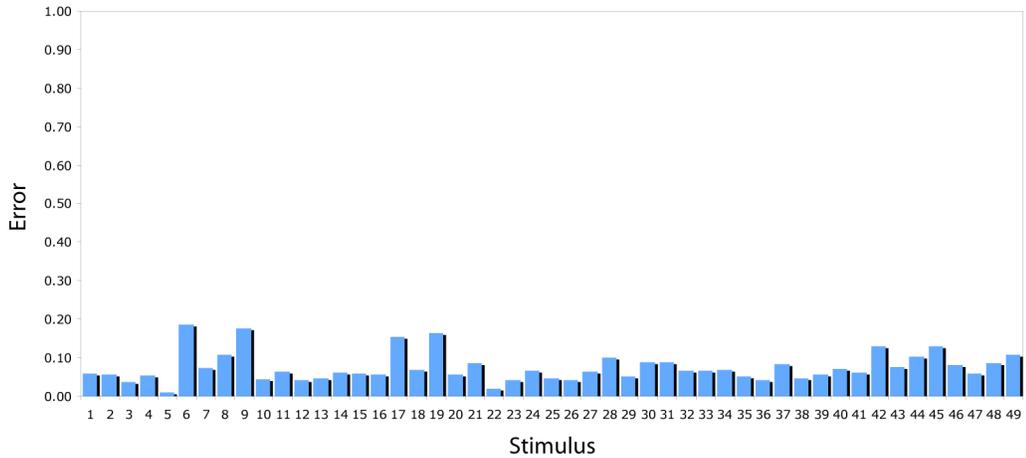


Figure 3.15: Euclidean distance from the model to Pasupathy and Connor's data.

than the ones reported. When we performed this analysis between each stimulus and the other 48 different stimuli, values were much higher (average = 0.218, stdev = 0.087). When comparing both populations (stimulus paired correctly vs stimuli paired wrongly) they are statistically different ( $t\text{-test}=1$ ,  $p<0.0001$ ).

These results show that the proposed hierarchical representation based on endstopping to achieve curvature provides very similar data to data from neurons, only 5 out of the 49 stimulus distances were over 10% error (or 0.10 distance), and all of them were below 19%. Most errors are on the range 0.02-0.08.

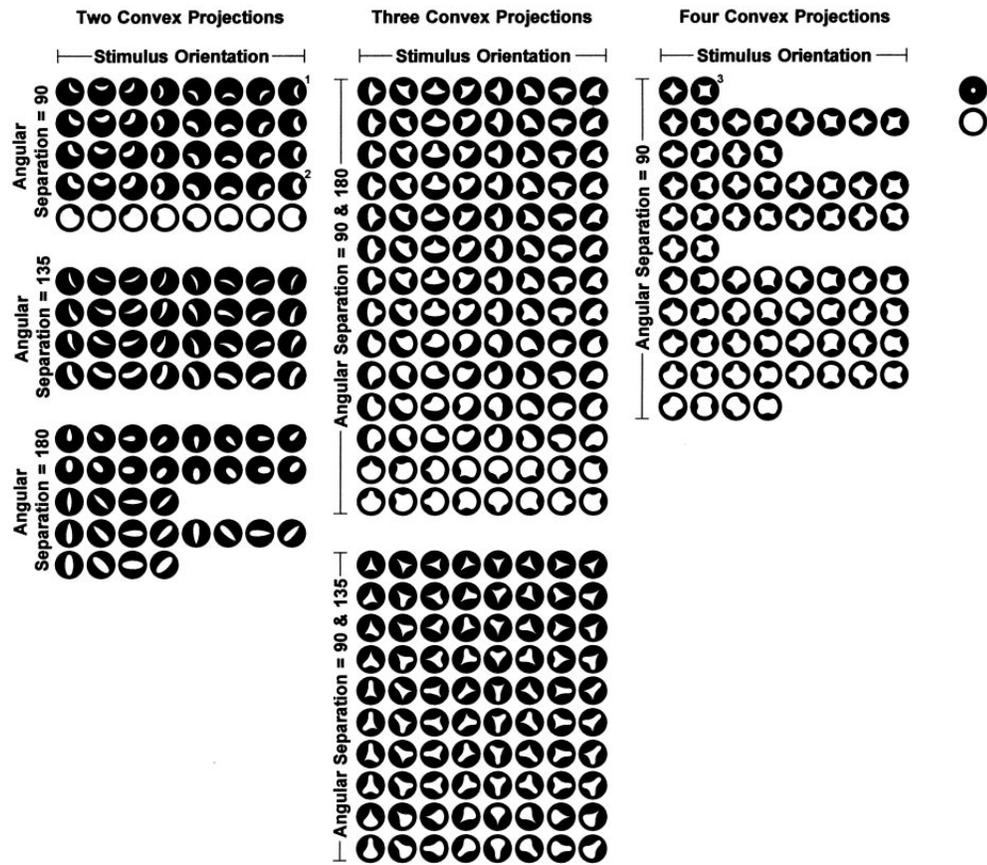


Figure 3.16: Stimuli used in [Pasupathy and Connor, 2001] (Copyright © 2001 The American Physiological Society)

### 3.2.2 Comparison with V4 neuron's responses to shape stimuli

#### Motivation

Now that we have shown how endstopping may be used for curvature discrimination we want to study how similar the behavior of our Shape neurons are to V4 neurons in the brain.

In a detailed study, [Pasupathy and Connor, 2001] recorded the responses of 109 neurons to 366 different shapes. The stimuli were the same 49 shapes used in the previous experiment but rotated 8 orientations (some only 2 or 4 due to redundancies) in 45° increments (Figure 3.16). Each cell in the sample responded to a variety of very different shapes. No cell displayed a response pattern that could be characterized in terms of a single type of global shape. However, for most cells the effective stimuli showed some degree of shape consistency at one position. In other words, these cells were tuned for boundary configuration in one part of the shape.

In order to prove the plausibility of our Shape neurons, we study the behavior of the model Shape neurons by comparing their responses against real neuron responses. If model Shape neurons provide similar responses to those of neurons in area V4, we can summarize that they faithfully mimic the behavior of real

cells.

## Method

**Correspondence between simulated and real neurons** The experiments we will conduct involve the response of our Shape neurons to different types of stimuli used in neurophysiological experiments. We use the database provided by [Pasupathy and Connor, 2001], who recorded from 109 neurons in area V4 of the macaque’s visual cortex for the stimuli shown in Figure 3.16. Based on the responses of those neurons to that stimulus set they could infer the selectivity of specific neurons, e.g. a neuron that responds strongly to any stimuli containing a sharp curvature on the top left would be selective for that feature.

We perform two sets of experiments, a first set performed a comparison with the four figures from [Pasupathy and Connor, 2001]. For comparing with Pasupathy and Connor’s figures we followed their descriptions regarding the selectivity of their neurons (more detail in Results), such as “the most effective stimuli contained not just a convexity at the bottom left but also a concavity at the bottom” (page 2510 from [Pasupathy and Connor, 2001]).

Since the results from the first set of experiments were satisfactory, in a second set of experiments we compared with the responses from 75 of the 109 neurons

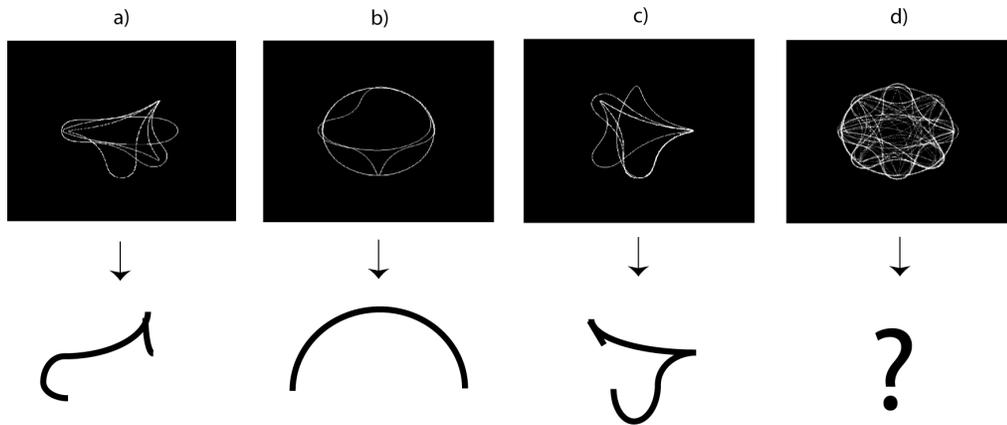


Figure 3.17: How the features for *isolating* a Shape neuron are obtained for our second set of experiments. See text.

recorded by Pasupathy and Connor’s group. Data from real neurons to achieve these set of experiments was kindly provided by Dr. Anitha Pasupathy. To perform this more extensive comparison, we needed to extract the curvature-position features for which those 75 neurons were selective. In order to do this for each real neuron from Pasupathy’s data we considered stimuli with neuronal responses on the 70% maximum percentile. These stimuli eliciting high responses from a real neuron are added into an image. The resulting image gives us an idea of that neuron’s curvature selectivities. An example of this is shown in Figure 3.17: Neuron a) is selective to a sharp convexity on the top left, a broad concavity adjacent to it and a medium-sharp convexity on the right, neuron b) is selective to a broad convexity that spans from left to right, neuron c) is selective for a

sharp convexity on the right, adjacent to a broad concavity leading to a sharp-medium convexity on the top left and/or a medium convexity at the bottom. On the other hand, neuron d) shows a high response to too many different stimuli to find the appropriate curvature features for that neuron. The reason why we used 75 neurons out of the 109 provided is because 34 of them fell into this last case.

As explained in the previous section, the possible representation of shapes with the model Shape-selective neurons is enormous. But in order to perform the comparison between our model Shape neurons and real neurons we have to select a subset of four neurons corresponding to the four figures from [Pasupathy and Connor, 2001] for the first set and then 75 model neurons to match the real 75 neurons for the second set of experiments.

In order to establish an equivalence between our Shape neurons and those real neurons in area V4 we follow a process we call *model neuron isolation*, since it would be the equivalent to recording from a specific neuron in the visual cortex. The process of model neuron isolation consists of the presentation of specific curvature features - e.g. sharp curvature at the top left - to the hierarchy presented here and then storing this representation in working memory. In more detail, the process of isolation is composed of two steps. A first step consists of creating an image with the selective curvature features corresponding to a real

neuron. To do this we follow [Pasupathy and Connor, 2001] descriptions for the first set of experiments (e.g. “the most effective stimuli contained not just a convexity at the bottom left but also a concavity at the bottom”) and the extracted curvature-position selective features for the second set of 75 neurons as has been described earlier (Figure 3.17). Then, we extract the representation corresponding to that set of curvature features. A feed-forward pass through the hierarchy of layers is performed and then its Shape-selective neuron representation is stored in working memory. The weights  $c_i$  (Equation 3.10) are derived from the responses from the eight curvature classes model neurons at their different positions or bins as explained before (Figure 3.9a). Now, we have isolated a model Shape-neuron and we are ready to study how it responds to the 366 stimuli used by [Pasupathy and Connor, 2001] as explained next.

**Response of an *isolated* Shape neuron to a presented stimulus** Real neurons provide a response which is measured in spikes per second, the higher this value the higher the selectivity of the real cell to the presented stimulus. In order to compare our Shape neurons to real cells we are interested in obtaining a value from the Shape neuron that reflects its excitation when presented that same stimulus. The process is as follows:

1. *Isolate* a model Shape neuron selective for a set of curvature feature(s) as explained earlier.
2. Once a model Shape neuron is *isolated*, its response to the most selective stimulus is computed. This value will be matched to the real cell value for that stimulus and used to normalize the responses of our Shape-selective neurons to each stimulus.
3. The model Shape neuron response to each one of the different 366 stimuli is recorded with responses normalized by the factor obtained in step 2.
4. The absolute difference in response between the model's Shape-selective neuron and the real V4 cell is computed for each stimulus. Mean and standard deviation are obtained for each case.

The input and setup are the same as before as well as the model neurons parameters. Shape neuron's receptive fields were organized into angular-radial *bins* (Figure 3.9a) of 10 pixels for radial values and  $\frac{\pi}{45}$  for angular values.

## Results

**Comparison with Figures from [Pasupathy and Connor, 2001]** In each Figure (Figures 3.18 to 3.21), columns a and c (stimuli within circles) correspond

to the four figures extracted from their work (their Figures 2, 4, 5 and 8), which represent the responses from four neurons. The model equivalences are in columns b and d (stimuli within squares) respectively. Each row and column contains up to 8 stimuli, which totals the 49 shapes rotated in steps of  $45^\circ$ . For example, the first row in column a is the response of a neuron from Pasupathy and Connor's work, its equivalent from our model Shape neuron is in that same row, column b. Green shading identifies a stimulus with high response or that is of interest to the study.

For real neurons (columns a and c), each stimulus is represented by a white icon drawn within a circle representing the unit receptive field, the darker the background behind the icon, the higher the response rate of the neuron is to that shape. This is also true for our Shape cells (columns b and d), with the distinction that the icon is within a square. For the model to perform correctly for a shape-stimulus response, the squared background must be similar to real cells circled background, that is, the background at columns b and d (model's Shape neurons) must be similar to the ones from real cells at columns a and c respectively.

We first compared the responses from our Shape-selective neurons with Figure 2 from [Pasupathy and Connor, 2001] (Figure 3.18). For this cell, stimuli with a

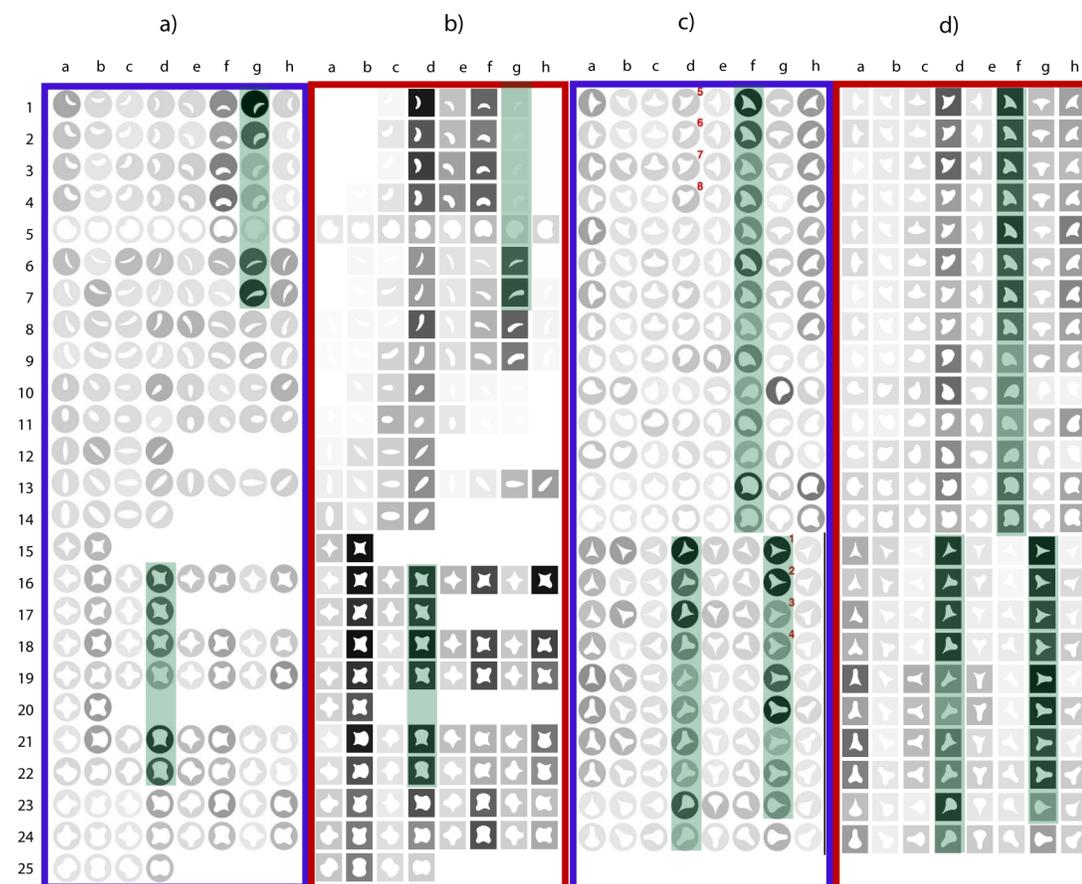


Figure 3.18: Comparison to Figure 2 of [Pasupathy and Connor, 2001] (Copyright © 2001 The American Physiological Society). Cells responses are on columns a) and c) and their respective model responses are on columns b) and d). Green shading is for stimuli with high response or stimuli that Pasupathy and Connor considered of interest

sharp convex angle at this position were particularly effective (e.g. stimuli 1 and 2 in the middle column, bottom block; these stimuli are labeled with superscript numbers). Stimuli with a medium convex curve evoked moderate responses (e.g., stimuli 3 and 4). Thus this cell appears to encode information about the bottom left boundary region, responding well to sharp convexity at this location and poorly to broad convexity or concavity. Based on the response of this cell to the stimuli, this neuron was selective to a sharp convexity at the bottom left and a concavity adjacent to it. The results from the model are in columns b and d to their corresponding cell responses in columns a and c.

A first examination shows that the responses of the model's Shape neurons are very similar to those of real cells. Our Shape-selective neurons respond strongly to a sharp convexity at the bottom left and a concavity at the bottom as well. If the curvature adjacent to the sharp convexity at the bottom left is convex, real cells responses are much weaker, our Shape-selective neurons show also weaker responses as well but not as weak as for real cells. We think that such feature may have an inhibitory effect in a similar way as in TEO [Brincat and Connor, 2004]. Our model as it is presently does not consider inhibitory responses.

There are some other differences between our model and the responses from real cells, e.g. elements g1 and g2 in column a and b. The reason behind this

difference is that the stimulus used as a base to isolate our Shape neuron was stimulus 1, which apart from the concave curvature at the bottom has a concave curvature on the left. It can be seen that such stimuli may be closer to the stimuli in subcolumn d than g, so our model responds stronger for subcolumn d than g, while for real cells, the opposite happens. There are other small differences which may be accounted for with the same reasoning. But in general it appears that the model performs well in comparison to neuronal data.

Another example provided by Pasupathy and Connor is on Figure 4 of their article (Replicated in Figure 3.19a and c). This cell was sensitive to boundary configuration on the right side of the object, responding best to concave curvature at that position. This is exemplified by stimuli 1 and 2; stimulus 1, with a concavity at the right, evoked a stronger response. Stimulus 2 is almost identical, but with a convexity at the right, and it evoked no response. The cell also appears to be tuned for sharper convexities at the counter-clockwise-adjacent position and medium convexities at the clockwise-adjacent position. Pasupathy and Connor note that this is exemplified by stimulus 3 providing a strong response, while for stimulus 4, its response is weak (opposite combination: sharp curvature clockwise and medium curvature counter-clockwise). The results for the model in this case is almost equal for these stimuli as well as the other cases mentioned in Dr.

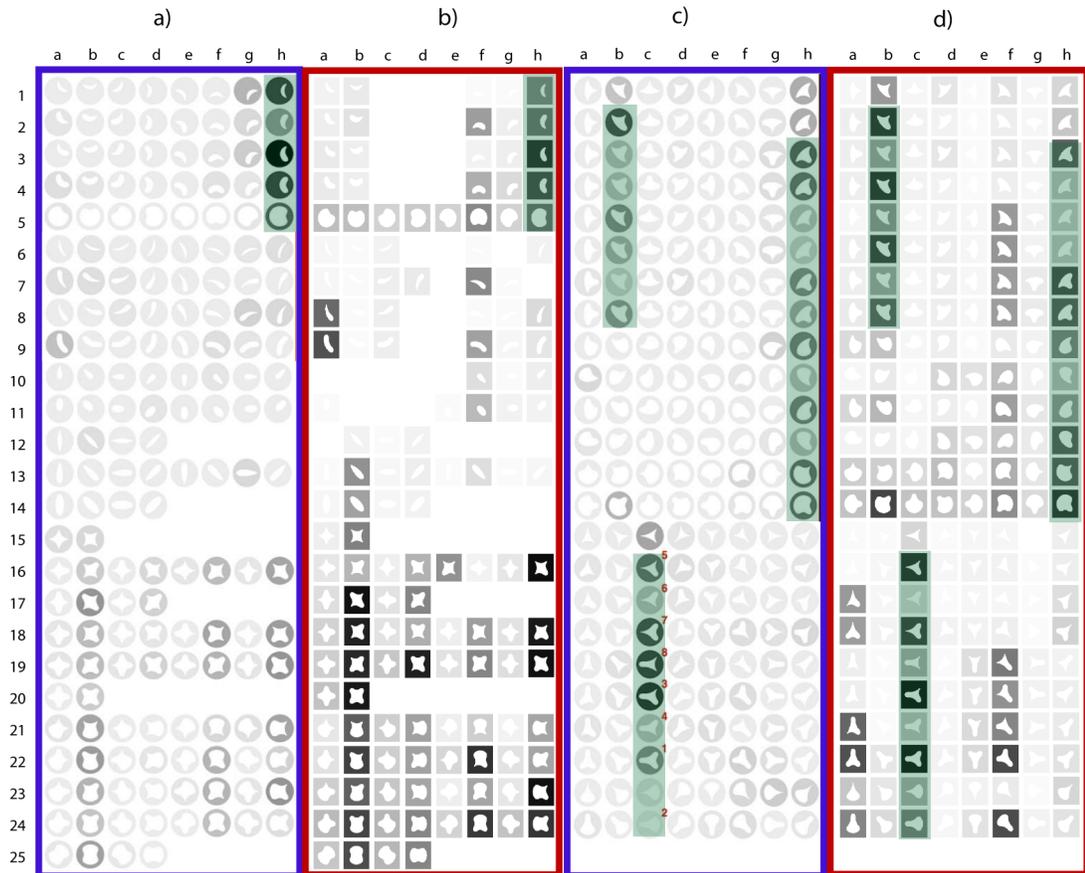


Figure 3.19: Comparison to Figure 4 of [Pasupathy and Connor, 2001] (Copyright © 2001 The American Physiological Society). Cells responses are on columns a) and c) and their respective model responses are on columns b) and d). Green shading is for stimuli with high response or stimuli that Pasupathy and Connor considered of interest

Pasupathy's work: compare shapes 5 and 6, and 7 and 8. As previously, there are some small differences, the model providing stronger responses than the cells for a few stimuli.

The two final examples are even more similar and appear in Figures 5 and 8 of [Pasupathy and Connor, 2001]. The first one is replicated in figure 3.20, the neuron was sensitive to a sharp convexity at the top right flanked by a concavity on one side or the other. In this case too, the behavior of the Shape-selective neurons is very similar. As it is the case for figure 8 of that same article, that cell was selective for broad convex curvature at the top. Their results are replicated here in Figure 3.21.

**Quantitative comparison with [Pasupathy and Connor, 2001]** After successfully comparing our model with the examples provided in Pasupathy and Connor's paper, we decided to test our model against raw neuronal data. For this, Dr. Pasupathy kindly provided the data corresponding to the 109 cells they recorded in area V4. Cells responses were normalized to 1 (although 1 was not always the highest value), the responses to the mentioned 366 stimuli (Figure 3.16) were recorded. We followed the steps explained before in the Method subsection and we used 75 out the 109 of neurons (recall Figure 3.17d).

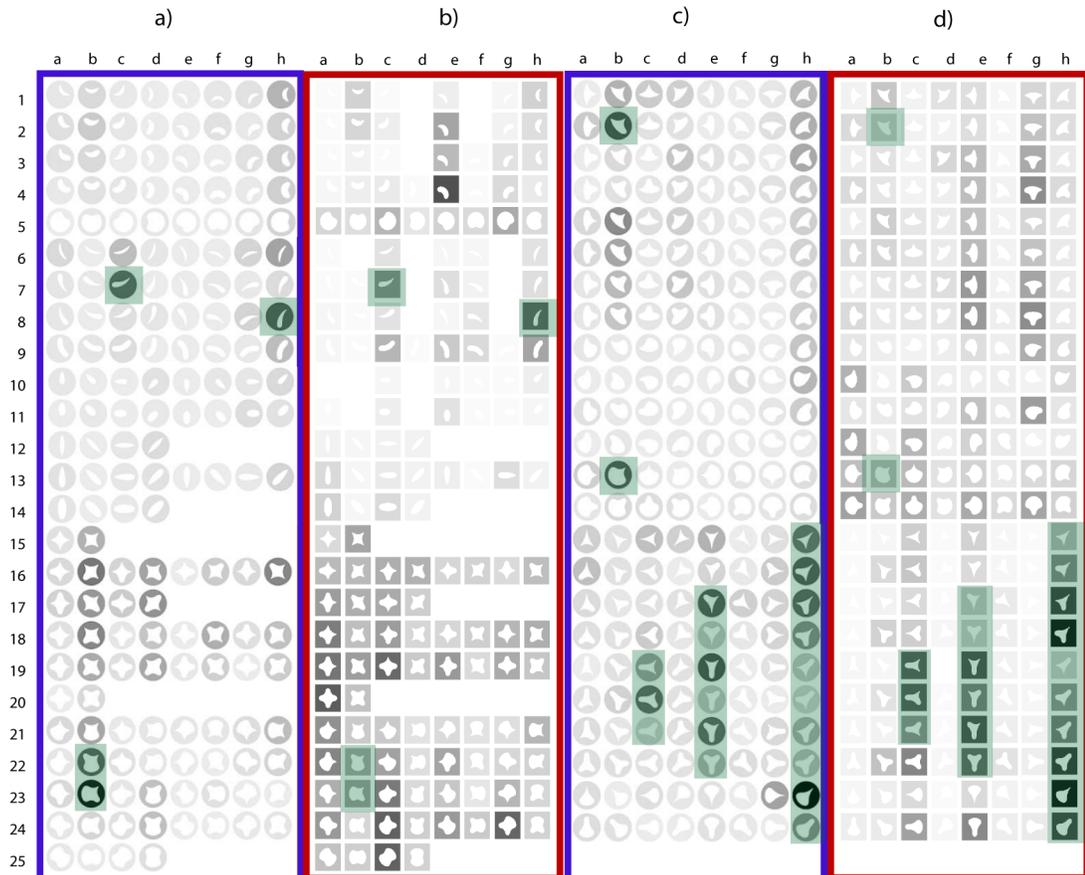


Figure 3.20: Comparison to Figure 5 of [Pasupathy and Connor, 2001] (Copyright © 2001 The American Physiological Society). Cells responses are on columns a) and c) and their respective model responses are on columns b) and d). Green shading is for stimuli with high response or stimuli that Pasupathy and Connor considered of interest



The difference between our model responses and V4 cells responses is computed as explained before, that is, the absolute difference between the responses of our Shape neurons and that of real cells is first obtained. For each cell, mean and standard deviation were computed and results will be provided next as error percentages, meaning mean difference between our Shape neurons and real cells.

The results for all the 75 cells considered in this study are shown in Figure 3.22. We did two subsets of experiments, the first one, model neuron responses were recorded using the curvature parts with respect to the center of the neuron (blue bars). For the second subset, model neuron responses were with respect the centroid of the shape (green bars). Note that the stimuli from [Pasupathy and Connor, 2001] are not always at the receptive field center (Figure 3.16). We did not find difference from using curvature parts with respect to the center of the model neurons or the centroid of the object.

For both cases we can see that there are only a few Shape neurons with over 20% error, most of the differences between the model and that of real cells fall in the range 10-20%. Average error for all Shape neurons was 16.95% for the center of the model neuron ( $stdev=12.61$ ) and almost the same when using the centroid of the shape ( $error=16.98\%$ ,  $stdev=12.25$ ). This shows that even for such a large number of neurons the model performs successfully and the difference between

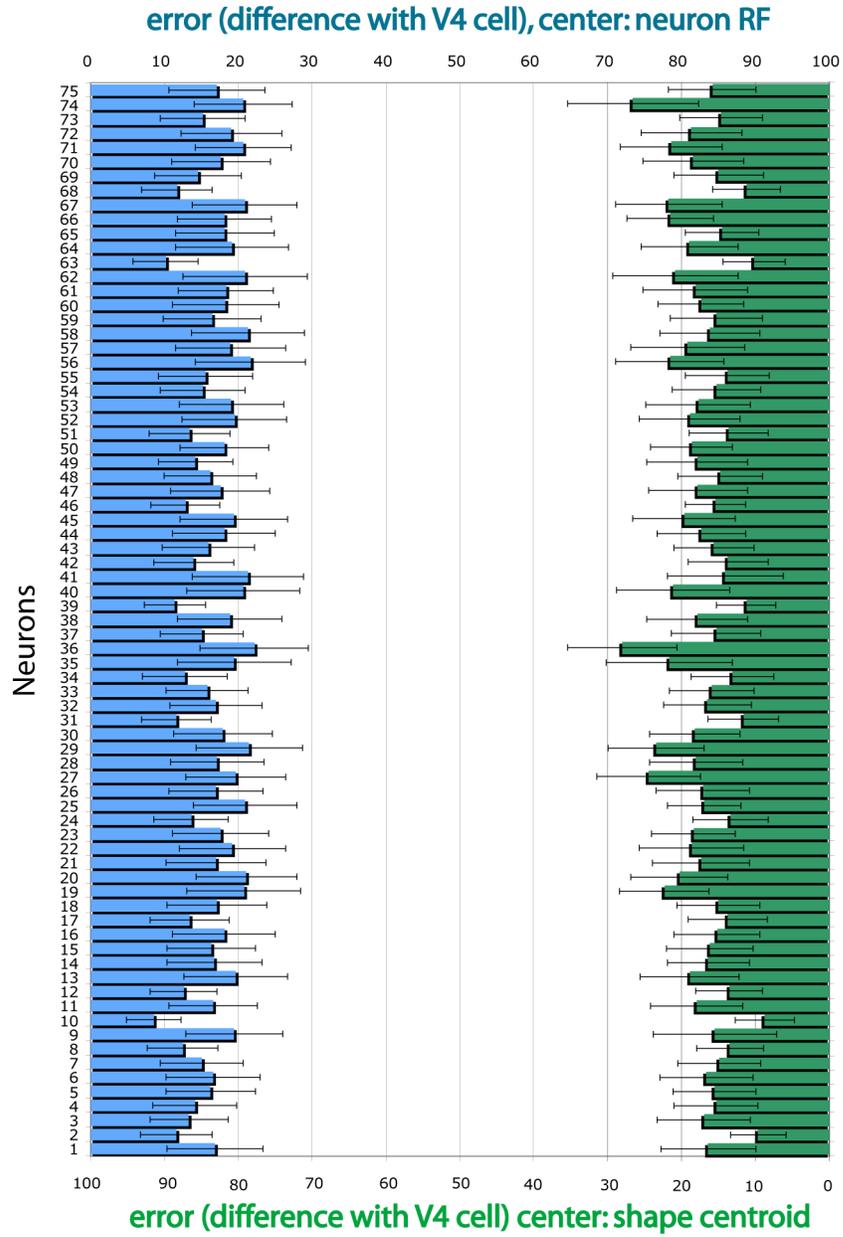


Figure 3.22: Difference between the model's Shape-selective neurons and 75 real cells responses from area V4

the response of the Shape-selective neurons and that of real cells is small.

### 3.3 Discussion

We first have shown the capabilities of our model at curvature selection, our first experiment successfully showed the validity of endstopping in a curvature discrimination task. Before there was a clear proof that V4 and TEO neurons performed curvature analysis, some authors considered that endstopping could be important for such analysis [Dobbins et al., 1987, Dobbins et al., 1989]. Pasupathy and Connor also consider that - among other hypothesis - endstopping may be a way to achieve curvature selection in the brain [Pasupathy and Connor, 1999]. Here, we tested this hypothesis and have shown how a hierarchy starting from basic simple edge detectors, which combine into complex neurons and further endstopped neurons provide local curvature neurons that are selective for shape stimuli.

Our model local curvature neurons do not provide an exact value of curvature but can discriminate between degrees of curvature. This was done using a starting point where V1 is composed of neurons of different sizes. Through the use of different neuronal sizes and the integration of model simple neurons into model complex neurons we obtained model endstopped neurons able to bandpass

between degrees of curvature, from very sharp to very broad (Figure 3.6).

It is important to note as well that these neurons do not provide a binary response for a given curve, that is, a 1 for the curvature to which is selective and 0 to the others. Model local curvature neurons provide a band-pass curvature filtering, with the highest response to the selective curvature and a decaying response that is inversely proportional to the curvature distances in curvature space. The response of model endstopped and curvature neurons over a range of curvatures have a Gaussian shape (Figure 3.6), as well as a model Shape neuron (Figure 3.11), which is the classical fit for neuron responses.

These characteristics of the model added to the relative sizes between neurons in V1, V2 and V4/TEO makes the proposed architecture biological plausible in every layer of the hierarchy. We compared the selectivity of our model neurons for curvatures with the selectivity of neurons in area V4 from a known study [Pasupathy and Connor, 2002] and we showed that their behavior for curvature selection was very similar to those of neurons in the visual cortex.

We have gone even further and compared the response of our Shape neurons with real neurons from [Pasupathy and Connor, 2001] with the examples provided in that article and 75 more neurons used for that work (data kindly provided by Dr. Pasupathy). The results obtained by the model are very simi-

lar to those of the neurons, and accomplished without any learning or classifier method and no maximum selection.

It is important when comparing to real neurons to extract the proper features from the stimuli in order to obtain similar responses to those neurons. Even though the way in which the selective features were obtained was somewhat primitive (Figure 3.17), the results are very promising.

All responses from model neurons are summed for the response of a Shape-selective neuron, but cells in area TEO of the brain have shown to integrate information in a more complex way [Brincat and Connor, 2004], in which some curvatures have a subtractive effect and there can be a nonlinear interaction between the elements as well. The way the brain works is still more complex than what we have proposed here and it would be interesting to study also inhibitory features and provide a more heterogeneous model of the Shape-selective neurons whose behavior would be even closer to real neurons than the one proposed here. There are a number of factors that could be included into the present model for a more faithful and similar performance to neurons in areas V4 and TEO of the brain.

Finally, our Shape neurons can represent a very large number (14,400!) of possible shapes required by neuroscientists [Pasupathy and Connor, 2002], but

this poses the problem of combinatorial explosion as a computer vision system. Even though the primate visual system and our model has the capability to represent a virtual infinity of shapes, the way to handle the large but finite number of shapes in our world may be thanks to experience achieved through learning, selecting those configurations of curvatures and corners relevant to recognize the shapes around us. Since the representation has the capability to represent any shape, a new shape can be easily incorporated into the system.

### **3.4 Conclusions**

We have presented a model of 2D Shape Representation that follows the structure and behavior of the visual cortex. In our aim was to implement a model with a high degree of biological plausibility. We would like to stress next the differences with other existing models.

Most models use a set of Gabor filters or Difference of Gaussians at the lowest level and then, incorporate a learned classifier at the top. In those models, the in-between layers are non-existent or their biological plausibility is questionable. In our case, from a starting point of edge detectors, we provide a set of endstopped neurons to obtain sets of neurons that are sensitive to different curvatures and their direction to finally obtain Shape-selective neurons.

Due to the nature of the representation there is a transformation from a pure retinotopic representation in V1 and V2 to a non-cartesian representation in V4 as proposed in the literature [Gallant et al., 1996], which may be the intermediate stage to the less retinotopic representation found in later areas [Tanaka, 1996]. In V4, representation is based on neurons depending on distance from the center of the neuron and angular position following recent studies in that area [Pasupathy and Connor, 2001]. The model supports a recognition by parts strategy, in which the parts are curvature values at different positions, whose support comes from Connor’s group as well [Brincat and Connor, 2004].

The differences between our model and other recent models, e.g. [Serre, 2006, Serre et al., 2007] are several. Whereas Serre and colleagues define their cell types as combinations of edge centric responses successively over 7 hierarchical layers, here our neurons in each layer compute quite different quantities. The goal was to include curvature computations directly, and not indirectly as Serre does through the conjunctions of edges. How the visual cortex might accomplish this has been extensively investigated, endstopped cells play a major role. However, except for the notable exception of [Dobbins et al., 1987], they have not been adequately investigated computationally. This is where our approach and that of Serre diverges. This is also what enables our true representation of curvature and

2D shape. The success of the approach in modeling the neural levels involved is evident in the matches to neural recordings which surpasses those shown in [Serre, 2006, Serre et al., 2007].

### **3.5 Appendix**

Here the parameters used for the reported experimentation are provided.

Simple cells	Parameter	Value		Endstopped	Parameter	Value
Simple 1	$\sigma_y$	10		ESC 1	$c_c$	1.5
	AR	1.75			$c_{d1}$	1
	WR	2.5			$c_{d2}$	1
Simple 2	$\sigma_y$	15		ESC 2	$c_c$	1.25
	AR	3.5			$c_{d1}$	1
	WR	2.5			$c_{d2}$	1
Simple 3	$\sigma_y$	22		ESC 3	$c_c$	1
	AR	5.325			$c_{d1}$	1
	WR	2.5			$c_{d2}$	1
Simple 4	$\sigma_y$	30		ESC 4	$c_c$	3
	AR	7.5			$c_{d1}$	1
	WR	2.5			$c_{d2}$	1

Complex cells	Parameter	Value		Direction	Parameter	Value
CX 1	separation	11		ESC+/- 1	displacement	16
CX 2	separation	8		ESC+/- 2	displacement	15
CX 3	separation	8		ESC+/- 3	displacement	22
CX 4	separation	8		ESC+/- 4	displacement	24

## 4 Discrimination of spiral angles in optical flow patterns

As we move, the projection of moving objects on our retinas generates an array of velocity vectors known as optic flow. Several action recognition systems use such flow information. However, here as for object recognition, other intermediate motion representations can be investigated. One class of optic flow is spiral motion, defined by the angle between a local vector direction and the direction of the steepest increase in local speed. By discriminating amongst such angles, an organism could discern between different flow patterns and effectively interact with the environment. In primates, spiral selective neurons in medial superior temporal (MST) area are thought to provide the substrate for this ability. We found that these cells show higher discrimination thresholds than found behaviorally in humans, suggesting that when discriminating spiral motions the primate brain integrates information across many of these neurons to achieve its

high perceptual performance. Parts of this work (Chapters 4.1 and 4.3 to 4.6) have been published in *Neuroreport*<sup>3</sup>.

## 4.1 Introduction

When we move through the environment, a spatial pattern of velocity vectors, commonly known as optic flow, is projected onto our retinas [Gibson, 1950]. One family of optic flow patterns, the spirals, can be described by a single parameter: the angle between the local motion direction and the direction of the steepest increase in local speed [Graziano et al., 1994]. This angle, or spiral direction, is 0 degrees in an expanding spiral, 180 degrees in a contracting spiral, and 90 and 270 degrees in clockwise and counterclockwise rotating spirals respectively (Figure 4.1).

Previous studies in primates have found neurons in the dorsal division of the medial superior temporal area that are tuned for spiral direction [Tanaka and Saito, 1989, Graziano et al., 1994], in a similar manner as middle temporal neurons are tuned for the direction of linear motion

---

<sup>3</sup>Comparing neuronal and behavioral thresholds for spiral motion discrimination. Rodríguez-Sánchez AJ, Tsotsos JK, Treue S, Martínez-Trujillo JC. *Neuroreport*. 2009 Dec 9;20(18):1619-1624. PMID: 19957382. Psychophysical experiments and neuronal data analysis were conducted by the author of this thesis. Neural recordings are thanks to JC Martínez-Trujillo performed at the University of Tübingen (Germany).

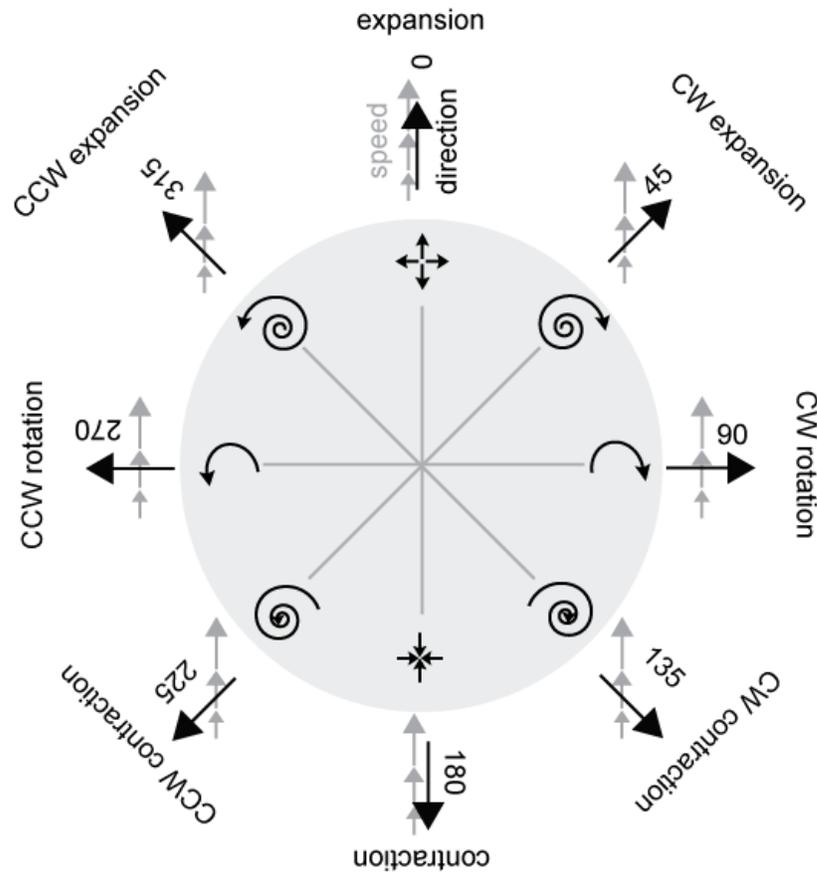


Figure 4.1: Spiral space is a coordinate system that interprets expansion ( $0^\circ$ ), contraction ( $180^\circ$ ) and rotations (clockwise:  $90^\circ$ , counterclockwise:  $270^\circ$ ) as cardinal directions with in- and outward spiraling movement patterns placed in between. Gray arrows show local motion direction.

[Snowden et al., 1992]. These neurons may play an important role in optic flow perception [Britten and van Wezel, 1998]. They can encode expanding and contracting spirals with similar accuracy as the animals [Heuer and Britten, 2004]. However, it is unclear whether this ability is restricted to coarse differences between these spiral types (Figure 4.1), or if it also generalizes to fine discrimination between spiral directions.

In humans, imaging studies have isolated a region adjacent to the middle temporal area that is selectively activated by spirals [Morrone et al., 1995, Martinez-Trujillo et al., 2005, Holliday and Meese, 2008], and where lesions produce deficits in complex motion perception [Vaina, 1998]. It is reasonable to assume that this region is the human homologue to the medial superior temporal (MST) area in macaques, and that in both primate species spiral-selective neurons within the region have similar physiological properties. Based on this assumption we asked the question of how the ability of humans to discriminate spirals compares to the one of medial superior temporal neurons in the monkey. The goal of this study was to inform our computational model of attentive visual motion processing [Tsotsos et al., 2005].

## 4.2 Selectivity for spatial gradient of velocities

First we investigated the hypothesis that the visual system possesses mechanisms sensitive to the spatial gradient of velocity direction relative to the motion direction in optical flow patterns which has also been previously investigated by others (e.g. [Graziano et al., 1994]). We used different combinations of spatial gradient of velocity directions and motion directions as adapting stimuli and examined the effect of adaptation on the perception of a set of test patterns containing fixed combinations of spatial gradient of velocity direction and motion directions. We hypothesized that if there exist a mechanism that is sensitive to the spatial gradient of velocity direction relative to motion direction, the effect of adaptation on the test patterns should differ for the different combinations of adapting patterns.

### 4.2.1 Material and Methods

Seven healthy human males (age 28-40), with normal or corrected to normal vision participated in the experiments, conducted at York University (Toronto, Canada) and preapproved by the University Institutional Ethics Review Board. All participants were trained in the task for two to three sessions before the experiments. Subjects sat in front of a 22" CRT computer monitor (LaCie Inc, Oregon, USA), using a chin rest at a viewing distance of 57cm. Stimuli were

generated using an Apple Power PC and custom made software.

**Stimuli and task** Stimuli consisted of 100%-coherence moving Random Dot Patterns containing only one motion direction. The shape of the patterns was square (side =  $9.01^\circ$ ) and they were shown at  $7.57^\circ$  eccentricity to the right of a central fixation point during 0.5 sec (Figure 4.2). Dots were black on a white background. A given trial consisted of the presentation of an adapting Random Dot Pattern (RDP, e.g., accelerating at  $3.03^\circ/sec^2$ ) during 20 seconds followed by a short interval (0.5 sec) in which no stimulus, except the fixation point, appeared on the screen, and finally followed by the presentation of a test pattern during 0.5 sec. The test stimuli consisted of RDPs moving upwards with different spatial gradient of velocity values ranged from  $-3.03^\circ/sec^2$  to  $+3.03^\circ/sec^2$ , where “-“ indicates deceleration and “+” indicates acceleration (in steps of  $0.25^\circ/sec^2$ ). The average speed of all the stimuli was  $2.14^\circ/sec$ .

We performed the experiment under four different conditions differing in the direction of spatial gradient of velocity relative to the motion direction of the adapting pattern. In the first condition, the adapting stimulus was moving upwards and accelerating, i.e., the spatial gradient of velocity direction was parallel to the motion direction ( $0^\circ-0^\circ$ ). In the second condition the adapting stimulus

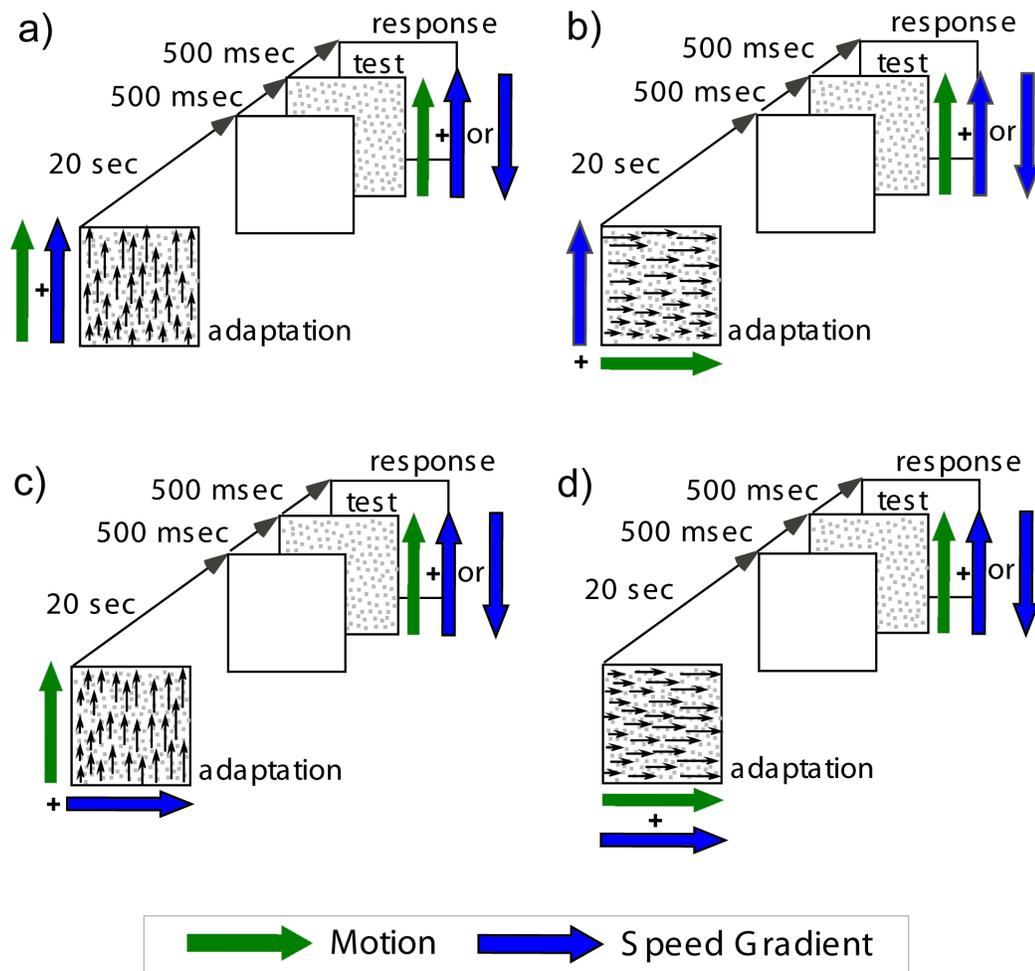


Figure 4.2: Trials consisted on the presentation of an adapting pattern lasting 20 sec followed by a test pattern lasting 0.5 sec. Variation of the adapting pattern gave four different conditions (staircase method): a) Moving and accelerating upwards, b) Moving rightwards and accelerating upwards, c) Moving and accelerating rightwards, and d) Moving upwards and accelerating rightwards.

spatial gradient of velocity direction was upwards, but the direction of motion was to the right ( $0^\circ$ – $90^\circ$ ). In the third condition, the adapting stimulus spatial gradient of velocity was oriented to the right but the stimulus moved upwards ( $90^\circ$ – $0^\circ$  angle). In the fourth and final condition the adapting stimulus spatial gradient of velocity was oriented to the right and the pattern moved to the right ( $90^\circ$ – $90^\circ$ ). The test pattern was always moving upwards, after each trial, subjects decided whether the test pattern was accelerating or decelerating. Eight subjects ran each of the four different blocks. Each block consisted of 33 trials and involved one of the four conditions. Two control blocks were run. Each control condition corresponded to two of the different adapting conditions.

Control blocks had the same design as the corresponding adapting blocks with the difference that spatial gradient of velocity was removed from the adapting stimuli. The adapting stimulus in one block was moving upwards at a constant speed of  $2.14^\circ/sec$ , this would correspond to removing the spatial gradient of velocity from conditions (1) and (3). For the other control block, the adapting stimulus was moving rightwards at the same constant speed, this would correspond to removing the spatial gradient of velocity from conditions (2) and (4). We used a multiple staircase procedure [García-Pérez, 1998] in order to determine the point of subject equality (PSE), i.e., the value of the spatial gradient of

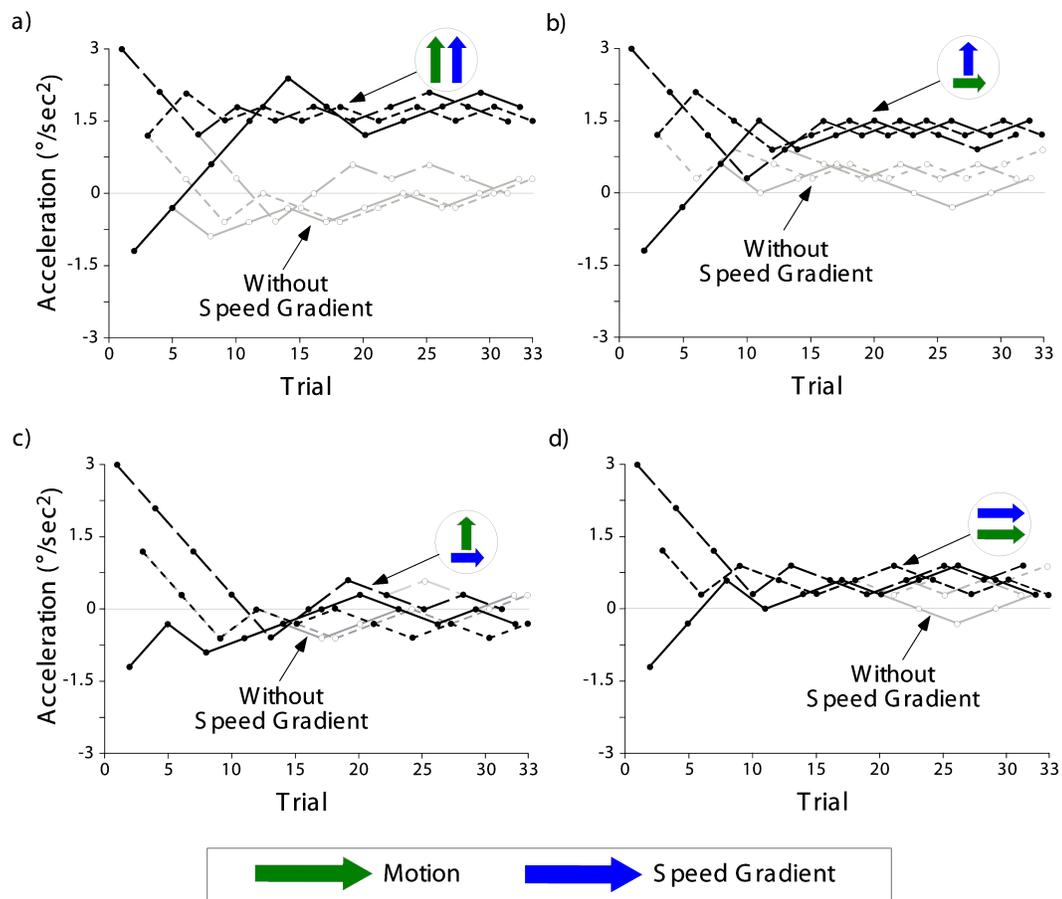


Figure 4.3: Example results from one subject.

velocity at which subjects perceived neither acceleration nor deceleration in the test pattern. The point of physical equality (PPE) corresponds to an acceleration of the test pattern of  $0^{\circ}/\text{sec}^2$ , i.e., no acceleration/deceleration.

## Results and discussion

Figure 4.3 shows data from an example subject. The filled dots and dark lines represent the conditions where the adapting pattern contained spatial gradient of velocity and the blank dots and light lines represents the control conditions in which the adapting pattern did not contain spatial gradient of velocity. For each condition we used three different staircases (solid line, long dashed lines and short dashed lines) with different starting points. The symbols represent the direction of motion and the spatial gradient of velocity direction (Note: Blank dots and light lines sometimes occluded by filled dots and dark lines).

Figure 4.3a is for the case in which the adapting pattern was moving upwards and accelerating ( $0^\circ-0^\circ$ ). Averaged among the three staircases, the PSE was at  $1.72^\circ/sec^2$ , i.e., after adaptation, the subjects perceived as not accelerating/decelerating a pattern that accelerates at a rate  $1.72^\circ/sec^2$ . As a control, we show in the same Figure 4.3a the data corresponding to an experimental session in which the adapting pattern did not contain spatial gradient of velocity and moved upwards ( $0^\circ$ ). In this case, the PSE was  $0.06^\circ/sec^2$ , a value very close to the PPE ( $0^\circ/sec^2$ ).

Figure 4.3b shows the results for the case an adapting pattern moving to the right but with a spatial gradient of velocity oriented upwards ( $0^\circ-90^\circ$ ). The

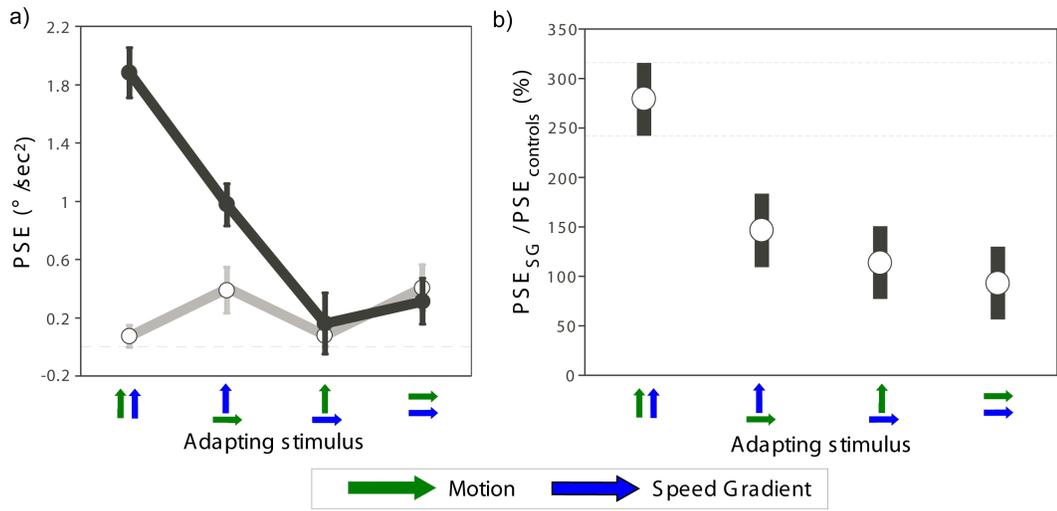


Figure 4.4: a) Mean and SD values averaged across the eight subjects. b) Ratio  $PSE_{SG}/PSE_{control}$  (ordinate) as a function of the adapting stimulus type (abscissa).

PSE was  $1.4^{\circ}/sec^2$ . Again, control data for an adapting pattern without spatial gradient of velocity are shown in the same graph (gray lines). The mean PSE for these control data was very close to  $0^{\circ}/sec^2$  indicating that adaptation shifted the PSE significantly. Figure 4.3c and Figure 4.3d show the data for the two remaining conditions, motion up and spatial gradient of velocity oriented to the right ( $90^{\circ}-0^{\circ}$ ) and motion to the right and spatial gradient of velocity oriented to the right ( $90^{\circ}-90^{\circ}$ ). The PSE in the former case was  $-0.15^{\circ}/sec^2$  compared to  $0.06^{\circ}/sec^2$  for the control and in the latter  $0.62^{\circ}/sec^2$  compared to  $0.35^{\circ}/sec^2$  for the control. Thus, in both cases the PSE was very similar to the control data suggesting that adaptation did not have an effect in these conditions.

In order to examine the results across our sample of different subjects we repeated the previous analysis in each one of the eight subjects and averaged the PSE values across them. Figure 4.4 shows the results. In figure 4.4a, the abscissa indicates the direction of the spatial gradient of velocity and direction of the adapting pattern and the ordinate the acceleration of the test pattern. The lines indicate the 95% confidence interval for the mean. The dark line and filled dots correspond to the with spatial gradient of velocity condition, the light line and blank dots correspond to the controls. When adapting with the stimulus moving upwards and the spatial gradient of velocity oriented upwards ( $0^\circ-0^\circ$ ), the average PSE was  $1.88^\circ/sec^2$  ( $\pm SD = 0.1^\circ/sec^2$ ). When the adapting stimulus moved to the right and the spatial gradient of velocity was oriented upwards ( $0^\circ-90^\circ$ ), the average PSE was  $0.97^\circ/sec^2$  ( $\pm SD = 0.09^\circ/sec^2$ ). When the stimulus moved upwards but the spatial gradient of velocity was oriented to the right ( $90^\circ-0^\circ$ ), the average PSE was  $0.15^\circ/sec^2$  ( $\pm SD = 0.13^\circ/sec^2$ ).

Finally, when the stimulus moved to the right and the spatial gradient of velocity was oriented to the right ( $90^\circ-90^\circ$ ), the average PSE was  $0.31^\circ/sec^2$  ( $\pm SD = 0.1^\circ/sec^2$ ). Control data for adapting stimuli without spatial gradient of velocities are shown on the same graph (gray lines). Since in our four conditions, the adapting pattern moved in two different directions, only two control experi-

ments were conducted (i.e., when the adapting stimulus moved upwards and to the right). Therefore, each control data is plotted two times to illustrate the differences with the spatial gradient of velocity data. The PSE for the first control (motion upwards) was  $0.18^{\circ}/sec^2$  ( $\pm SD = 0.03^{\circ}/sec^2$ ), and  $0.43^{\circ}/sec^2$  ( $\pm SD = 0.09^{\circ}/sec^2$ ) for the second control (motion rightwards).

In order to compare the results among different groups we used the experimental and control data to compute a ratio measurement for each subjects and condition.

$$ratio = \frac{PSE_{SG}}{PSE_{control}} * 100\% \quad (4.1)$$

In the above formula  $PSE_{SG}$  indicates the PSE for the adapting patterns with spatial gradient of velocity and  $PSE_{control}$  indicates the PSE for the control data. If the value is higher or lower than 100% it means that the spatial gradient of velocity in adaptadapting pattern had an effect on the perception of the test stimulus. A value of 100% signifies no effect of the spatial gradient of velocity. The average values for the four conditions were (278.93%, 146.58%, 113.85%, 93.11%). Figure 4.4b shows these values as well as their 95 % confidence intervals. When the spatial gradient of velocity was oriented upwards (first two cases), the mean ratios were significantly different from 100% (i.e., confidence interval - CI

- do not overlap with the horizontal line at 100%). When the spatial gradient of velocities was oriented to the right, the ratios were not different from one (i.e., the CI overlap with the horizontal line at 100).

An analysis of variance showed that the effect of the first adapting stimulus ( $90^{\circ}$ – $90^{\circ}$ ) was significant ( $p < 0.001$ ). Between any other pair of groups we found no significant differences ( $p > 0.05$ ). However it was a clear trend toward ratio values higher than 100 when the adapting stimulus moved to the right with a spatial gradient of velocity oriented upwards ( $0^{\circ}$ – $90^{\circ}$ ). Summarizing, we found a clear effect when the direction of the spatial gradient of velocity in the adapting stimulus was upwards and the direction of the stimulus was parallel to that direction. This effect clearly decreased when the direction of the stimulus changed and disappeared when the direction of the spatial gradient of velocity also changed.

## **4.3 Spiral discrimination in humans and single neurons**

### **4.3.1 Human Subjects**

Seven healthy human males (age 28-40, with normal or corrected to normal vision) participated in the experiments, conducted at York University (Toronto, Canada) and preapproved by the University Institutional Ethics Review Board. All par-

ticipants were trained in the task for two to three sessions before the experiments, some of them participated in the previous experiment. Subjects sat in front of a 22" CRT computer monitor (LaCie Inc, Oregon, USA), using a chin rest at a viewing distance of 57cm. Stimuli were generated using an Apple Power PC and custom made software.

### **Apparatus and task**

Subjects sat in front of a 22" CRT computer monitor (LaCie Inc, Oregon, USA), using a chin rest at a viewing distance of 57cm. Stimuli were generated using an Apple Power PC and custom made software. We measured spiral direction discrimination thresholds near the cardinal directions in spiral motion space (i.e. rotation, expansion and contraction) in seven human subjects using the method of constant stimuli. In one set of trials, the standard stimulus was an expanding spiral ( $0^\circ$ ), and the test stimuli were expanding spirals with clockwise or counterclockwise rotation components ( $350^\circ, 352^\circ, 354^\circ, 356^\circ, 358^\circ, 2^\circ, 4^\circ, 6^\circ, 8^\circ, 10^\circ$ ). In a second set of trials, a contracting stimulus served as the standard ( $180^\circ$ ), and the test stimuli were contracting stimuli with clockwise or counterclockwise rotation components ( $170^\circ, 172^\circ, 174^\circ, 176^\circ, 178^\circ, 182^\circ, 184^\circ, 186^\circ, 188^\circ, 190^\circ$ ). Subjects reported whether the stimulus moved clockwise or counterclockwise.

In a third set of trials clockwise or counterclockwise rotating stimuli served as standard. The test stimuli were: a) clockwise rotating spiral stimuli with different amounts of expansion or contraction (80°, 82°, 84°, 86°, 88°, 92°, 94°, 96°, 98°, 100°) for the clockwise standard stimulus (90°), and b) counterclockwise rotating stimuli with different amounts of expansion and contraction (260°, 262°, 264°, 266°, 268°, 272°, 274°, 276°, 278°, 280°) for the counterclockwise standard stimulus (270°). Subjects reported whether the stimulus contracted or expanded.

Trials started when the subjects foveated the fixation point and pressed the space bar on a computer keyboard. This initiated the appearance of a random dot pattern for 500 ms to the right of the fixation point. In trials with expanding/contracting patterns, subjects pressed the “3” (clockwise), or “1” (counterclockwise) key. In trials with rotating patterns, subjects pressed “3” (expansion) or “1” (contraction). The experiment was run in two blocks of 240 trials, one block of expanding/contracting trials and the other of rotating trials. After two to three training sessions, each subject performed 6 blocks (three of each type) in a randomized order. We instructed subjects to fixate and monitored eye movements using a video-camera. A session was excluded from the analysis if the experimenter detected that a subject broke fixation in at least 10% of trials by visually inspecting the videorecordings. Since our subjects were trained in the

task, no session met this criterion.

## Stimuli

Stimuli were black random dots on a white background, moving coherently behind a circular aperture (diameter:  $10.2^\circ$ ), and centered  $7.6^\circ$  to the right of a central fixation point. The dot density was 5 dots/ $deg^2$ , the monitor resolution was 33 pixels/deg and the monitor refresh rate was 75 Hz. The dot size was 3x3 pixels. The dot speed formed a linear gradient with zero at the center and  $6.9^\circ/sec$  at the edge.

## Data analysis

We computed a psychometric function for each subject and block type by fitting equation 4.2 to the proportion of times  $P(s)$  that the subjects reported: a) the stimulus rotating clockwise, for contracting/expanding spirals with rotating components, and b) the stimulus expanding, for rotating spirals with expanding/contracting components. Data from blocks of the same spiral type were pooled.

$$P(s) = \frac{1}{1 + e^{-(a+bs)}} \quad (4.2)$$

The discrimination threshold was the distance (in degrees) between the point of subjective equality (spiral direction at which  $P(s)=0.5$ ) and the point where  $P(s)=0.25$ .

### **4.3.2 Single cell recordings**

#### **Subject**

We recorded the responses of spiral-selective neurons to moving random dot patterns, in the dorsal subdivision of the medial superior temporal area in the superior temporal sulcus of one rhesus monkey (4-year old, 6.5 kg male *Macaca mulatta*) while the animal performed a detection task. Prior to the recordings and the final training a head holder and recording chamber were implanted under general anesthesia (see [Martinez-Trujillo and Treue, 2004] for more details). Recordings were conducted at the University of Tuebingen, Germany. All the procedures were in agreement with the German local and national rules and regulations, and were approved by the Regierungspraesidium Tuebingen.

#### **Apparatus and task**

The experimental procedures have been described in more detail elsewhere [Martinez-Trujillo and Treue, 2004]. Stimuli were white random dots on a dark

background (luminance:  $55 \text{ cd/m}^2$ , background luminance:  $0.1 \text{ cd/m}^2$ ), with dot size and density similar to the ones used in the human experiments. The dots' average speed was optimized to match the preferred speed of the neurons. The pattern eccentricity varied from  $5^\circ$  to  $12^\circ$  from the fixation point, and the pattern size was approximately equal to that eccentricity. Most neurons had preferred speeds between 4 and  $16^\circ/\text{sec}$ . We adjusted the speeds of the dots at the aperture's border depending on that preferred speed. To successfully complete a trial, the animal had to maintain fixation within  $0.5^\circ$  from the fixation spot.

A trial consisted of the following sequence of events: a) a fixation spot and a static random dot pattern appeared inside the cell's receptive field, b) the monkey fixated the spot and pressed a lever, c) 200ms later the pattern began to move, d) the animal was rewarded with a drop of juice for releasing the lever in response to a transient speed change in the dots (200 ms duration) occurring between 200 and 2000 ms after motion onset. We recorded the responses to 8 different spiral directions ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ,  $180^\circ$ ,  $225^\circ$ ,  $270^\circ$ , and  $315^\circ$ ).

### **Data analysis.**

We determined the spiral tuning of each cell (n=26) by fitting the mean responses in correctly performed trials averaged over the time period from 300 ms to 800 ms

after stimulus onset with a Gaussian function (equation 4.3). We only included neurons in which we recorded at least one correctly performed trial per direction. A cell was considered spiral-selective if the fit provided by the function yielded a correlation coefficient  $\geq 0.9$ , and if responses to spirals were better fit than responses to random dot patterns moving in 8 different directions of linear motion. Usually, spiral-selective neurons had large receptive fields that extended into the hemifield ipsilateral to the recording site [Treue and Martínez-Trujillo, 1999].

$$response = rmin + rmax \bullet e^{-0.5\left(\frac{direction-center}{\sigma}\right)^2} \quad (4.3)$$

$rmin$  represents the cell's response to the anti-preferred spiral direction,  $rmax$  the difference between  $rmin$  and the response to the preferred direction (sensitivity),  $center$  represents the preferred direction, and  $\sigma$  the standard deviation of the Gaussian (selectivity) [Snowden et al., 1992, Graziano et al., 1994, Treue and Martínez-Trujillo, 1999].

Using the fit parameters we simulated responses to 600 trials of each one of 46 different spiral directions (spaced every  $4^\circ$ ). We assumed that the response variance of a spiral-selective neuron follows a Fano factor of 1.0, (i.e. variance = mean firing rate), as a frequently reported property of cortical neurons [Shadlen and Newsome, 1998].

From the responses we computed neurometric functions for 7 different response levels around the flanks of the tuning curves [Snowden et al., 1992], with in the range of  $0.7^*$ (*maximal response*) and  $1.3^*$ (*minimal response*). For each level, we calculated the probability that more than a given number of spikes would be elicited for each stimulus. These data were fitted with equation (4.3) resulting in the criterion level neurometric function. The discrimination threshold for each function was considered as the distance in degrees of spiral direction between the 0.5 and 0.25 probability values.

$$P = \gamma - (\gamma - \vartheta)e^{-\left(\frac{d}{\alpha}\right)^\beta} \quad (4.4)$$

$d$  is the spiral angle,  $\alpha$  the direction at which a criterion probability is reached,  $\beta$  is the parameter governing the slope of the function,  $\vartheta$  is the asymptotic value of  $P$  (when  $d = 0$ ) and  $\gamma$  is the probability of reaching criterion for the least preferred direction. These functions are similar to the psychometric functions in that they describe the cell's ability to encode spiral direction [Snowden et al., 1992].

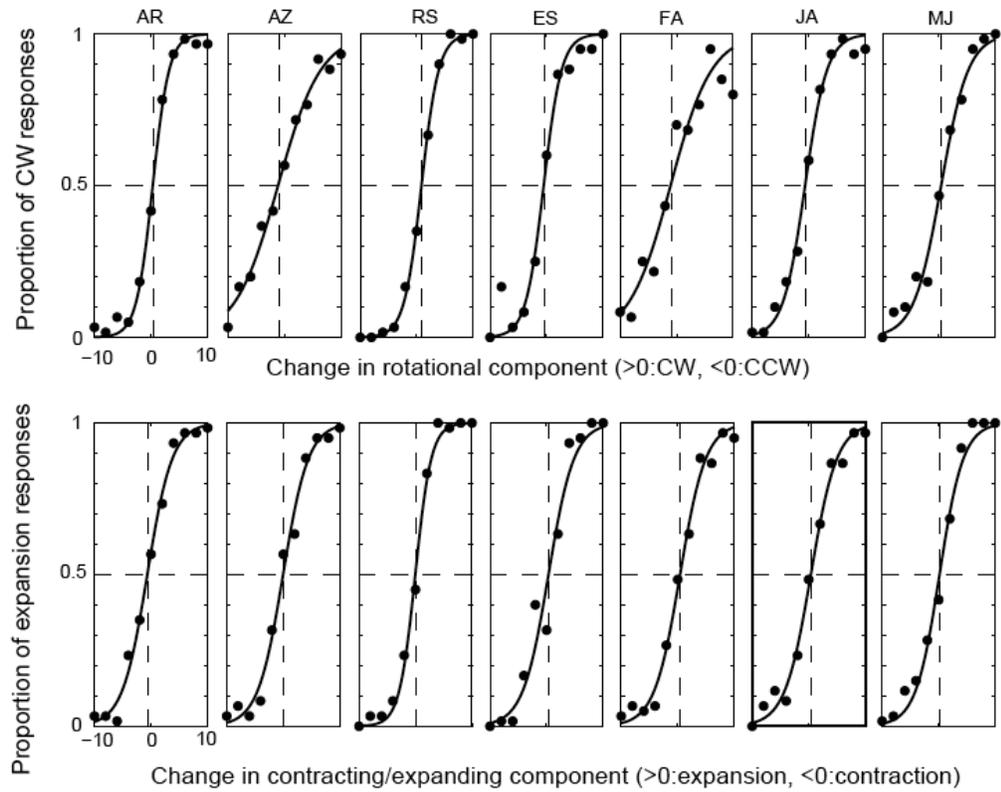


Figure 4.5: Averaged psychometric functions (upper row, clockwise/counterclockwise rotation; lower row, expansion/contraction) for human subjects.

## 4.4 Results

### 4.4.1 Human discrimination thresholds for spiral stimuli

We measured spiral discrimination thresholds in 7 human subjects. Figure 4.5 shows psychometric functions corresponding to each subject. The abscissa represents the amount of variation in the spiral direction and the ordinate the proportion of a given response type. The lines represent the fits through the data. All fits yield correlation coefficients ( $r$ ) larger than 0.9. On the top row for expanding and contracting spirals with changes in the amount of clockwise or counterclockwise rotational components and on the bottom row, for rotating spirals with changes in the expanding or contracting components. In all subjects, a sigmoid function (see methods) provided an excellent fit to the data ( $r > 0.9$ ). For changes in the rotational component, the lowest discrimination threshold was  $1.58^\circ$  and the largest  $4.1^\circ$ . For changes in the expanding-contracting components the lowest threshold was  $1.54^\circ$  and the largest was  $2.49^\circ$ .

In order to test whether the thresholds for rotation were systematically different from the ones for expansion/contraction, we compared the thresholds appearing on the top row of Figure 4.5 against the ones appearing at the bottom. The differences were not statistically significant ( $p > 0.8$ , Wilcoxon sign-rank test

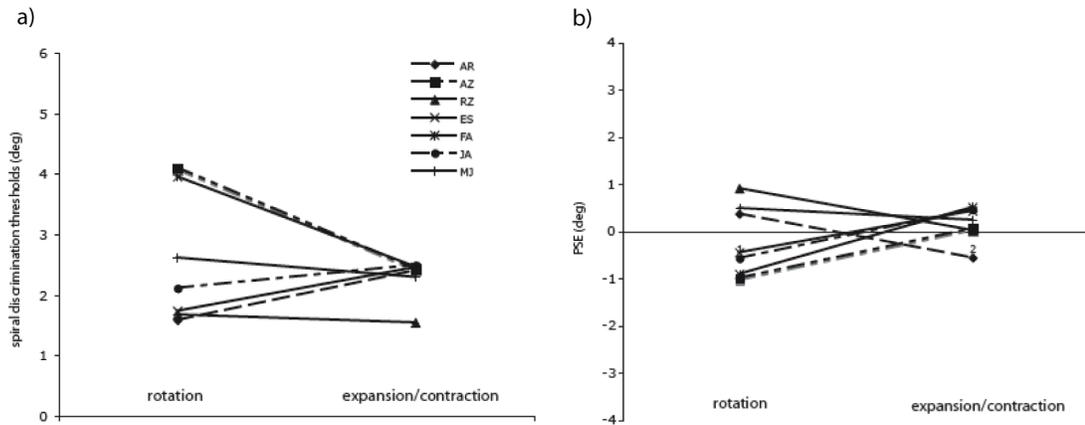


Figure 4.6: Comparison between expansion/contraction and rotations a) Spiral discrimination thresholds for every subject b) Point of Subjective Equality corresponding to each of the subjects.

for paired data) suggesting that in general subjects were equally accurate at discriminating clockwise vs. counterclockwise and contracting vs. expanding spirals. This result are visualized in Figure 4.6 where thresholds for each one of the subjects in the two tasks are plotted and joined by lines. Although in two subjects thresholds for rotation seem to be substantially larger than the expansion/contraction, at the level of the group these differences were not statistically significant ( $p > 0.8$ , Wilcoxon sign-rank test for paired data).

In order to investigate possible biases in the subjects' ability to perceive one type of spiral we compared the point of subjective equality (PSE) of these functions (Figure 4.6b). We found that for the case of rotation they were not statistically different from zero or point of physical equality ( $p = 0.49$ , Wilcoxon rank-sign

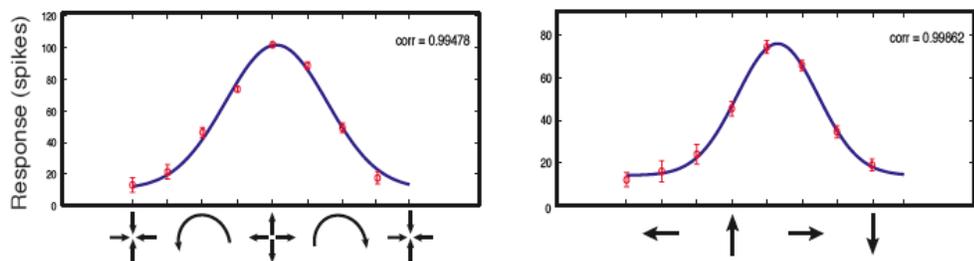


Figure 4.7: Gaussian fit for a spiral selective neuron (a) and a neuron tuned to linear motion (b)

test for a single group). For expansion/contraction, we found a small bias (less than  $1^\circ$ ) for the subjects to report the stimulus as contracting ( $p < 0.03$ , Wilcoxon rank-sign test for a single group).

#### 4.4.2 Neurometric analysis of single cell responses

The distribution of the tuning curves parameters for a spiral selective neuron is displayed in Figure 4.7a. The abscissa represents spiral direction and the ordinate the response (spikes/s). Error bars represent standard errors. In order to compare spiral selectivity of MST neurons against the selectivity of MT neurons for linear motion direction, we computed the distribution of the same parameters for a subsample of 26 MT neurons randomly selected amongst a population of 69 MT recorded neurons (Figure 4.7b shows a linear selective neuron). The mean parameters values for the spiral stimuli in MST neurons were:  $r_{min} = 10.3$ ,

$rmax = 42$ ,  $center = 201$ , and  $\sigma = 50$ . The mean values for the sample of MT neurons tested with linear moving RDPs were:  $rmin = 3.5$ ,  $rmax = 65.8$ ,  $center = 143$ , and  $\sigma = 44.6$ . The mean  $rmin$  was significantly higher while the mean  $rmax$  was significantly lower for the spiral selective neurons ( $p < 0.05$ , unpaired t-test in both cases). The parameter center was widely distributed in both samples demonstrating that we recorded from neurons tuned to different spiral angles and motion directions. The mean of the parameter sigma was somewhat larger for the spiral selective neurons, but this difference did not reach statistical significance ( $p > 0.05$ , unpaired t-test). These parameter values for both samples of neurons are very similar to the ones reported by other authors [Snowden et al., 1992, Graziano et al., 1994, Treue and Martínez-Trujillo, 1999].

We computed neurometric thresholds for the different neurons using the procedure described in the methods section. Neurometric functions were produced by selecting a criterion, e.g. 50 spikes/s and calculating the probability that the cell fired at or greater than this rate for each direction of motion (46 directions). 7 different criteria were used producing 7 neurometric functions. The data were then fitted by the integral of a Weibull function. From these neurometric functions, discrimination thresholds were obtained (Figure 4.8, lower panels).

The left top panel in figure 4.8 shows neurometric functions for spiral discrim-

ination of one MST neuron. The lower panel shows the threshold values for the different criterion levels as a function of spiral type (see methods). This curve has a U-shape and its minimum represents the neuron's best performance at discriminating between two directions (best discriminability or threshold). The spiral selective neuron on the left could discriminate between stimuli that differed by  $5.7^\circ$  of spiral angle along the slope of the tuning curve (i.e., rotating stimuli). For comparison purposes, the right panel in Figure 4.8 shows similar data for the responses of an example MT unit to linear motion. The neurometric functions and the neuron's best discriminability produced similar values as the ones of the spiral selective neuron. The same analysis was repeated for each one of the remaining units in both groups and corresponding samples of neurometrics thresholds were obtained.

Figure 4.9a shows the distribution of the spiral selective neurons performance (lower discriminability or threshold values). The abscissa represents the threshold values and the ordinate the number of units. The circles represent spiral angle discrimination thresholds from the human subjects (see Figure 4.5). Green dots correspond to human data for expansion/contraction, red dots are human data corresponding to rotations. In general, humans show lower threshold values than the neurons ( $p < 0.05$ , Wilcoxon ranksum test). Figure 4.9c shows a similar

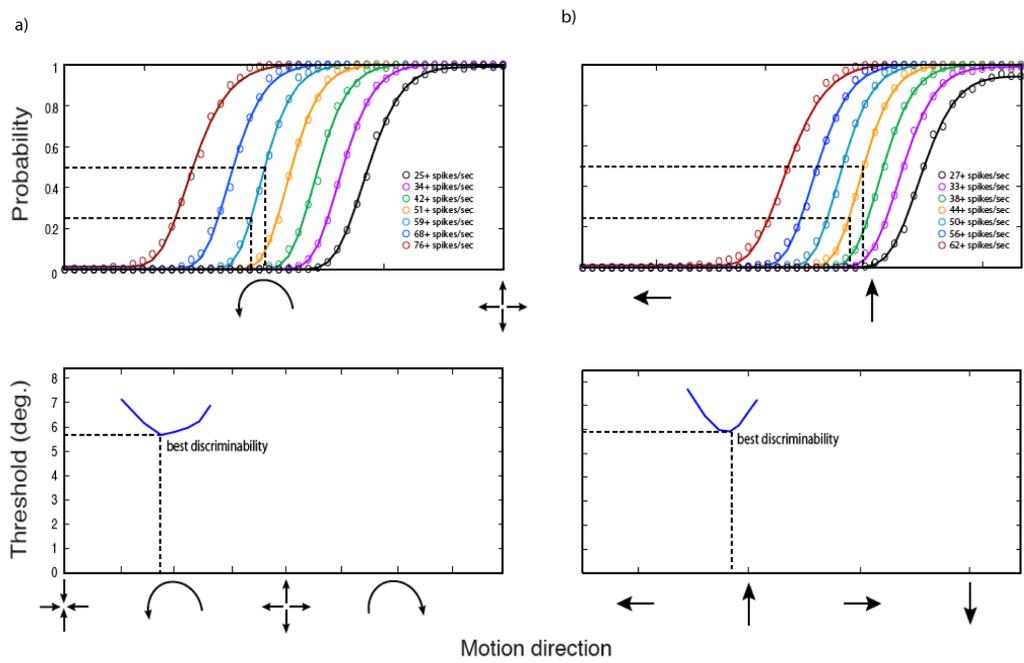


Figure 4.8: a) Neuron tuned for spirals b) Neuron tuned for linear motion.

histogram for the sample of 69 MT neurons selective for the direction of linear motion. In this case we did not measure the human discrimination thresholds but we indicate with the arrow the values reported by Snowden and coworkers [Snowden et al., 1992] ( $1.1^\circ$ ). Again, human thresholds are lower than neuronal ones ( $p < 0.05$ , Wilcoxon ranksum test). In addition, neuronal thresholds for linear motion direction discrimination were lower than the ones for spirals ( $p < 0.05$ , unpaired t-test).

The mismatch between the discrimination thresholds of single neurons and humans may be due to several possibilities. One straightforward interpretation is that human MST neurons may be more tuned for spirals than in the monkeys (higher  $r_{max}$  and lower width). Another possibility is that since the animal was not doing a spiral discrimination task, as human did, the tuning curve profiles do not reflect how these neurons react to the task. During discrimination, tuning curves may become higher (larger  $r_{max}$ ) and narrower (smaller width). A third possibility is that the neurometric performance of MST neurons does not directly determine the subjects' performance. But this is done downstream from MST in areas where neurons are more tuned to spirals.

It is difficult with our data set to distinguish between these possibilities. However, we could quantify the changes in  $r_{max}$  or width needed for the cell's neuro-

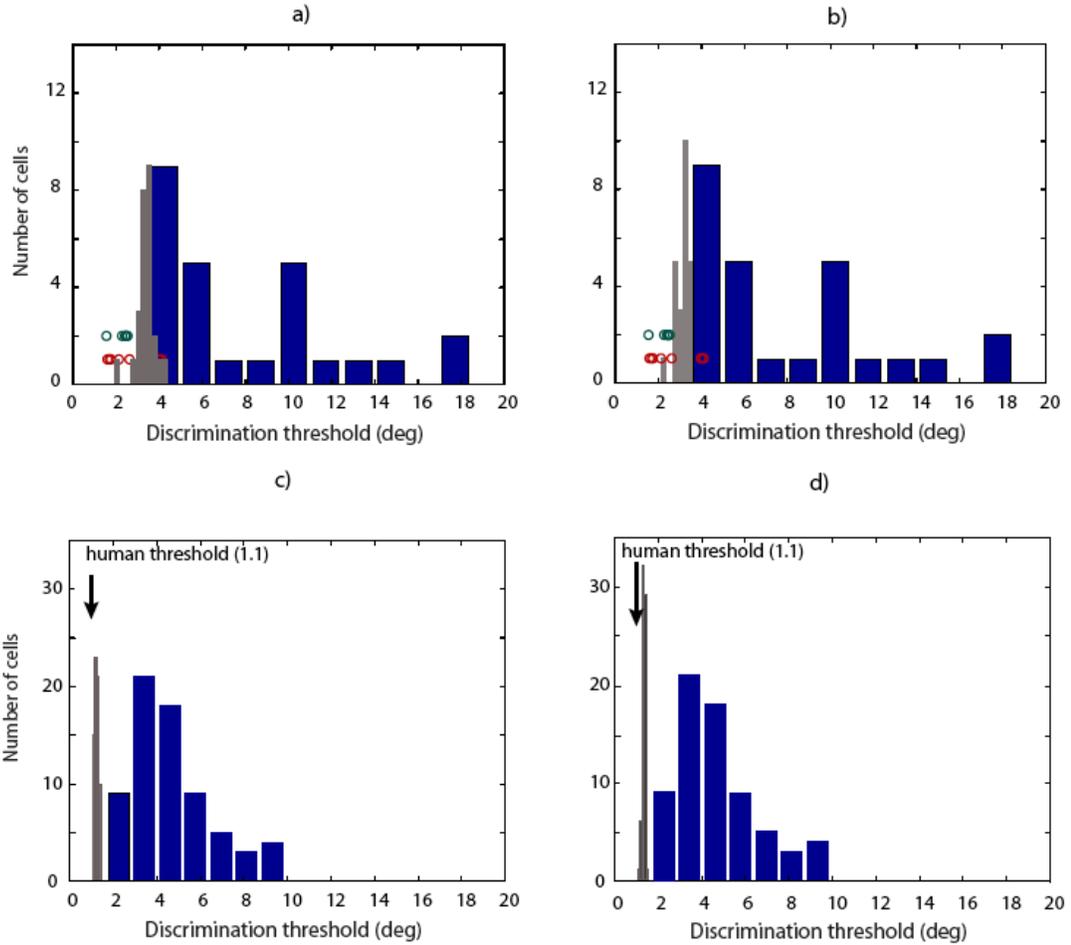


Figure 4.9: Distribution of the population of neurons based on their discrimination thresholds (blue histograms). a, b) Neurons tuned for spirals. c) and d) Neurons tuned for linear motion. Human threshold for this kind of motion is  $1.1^\circ$ . [Snowden et al., 1992]. Gray histograms correspond to the same neurons but after changes in the width (a, c), and height (b, d) (see main text).

metric thresholds to reach human behavioral thresholds. By increasing the gain of the spiral selective neurons by 3.9 the best neuron in our sample achieved a neurometric threshold similar to the average human behavioral threshold (Figure 4.9a, gray histogram). Also, by decreasing the width to 55% of its original value, our best unit performed similar to the average human (Figure 4.9b, gray histogram). We performed a similar computation on the direction data and the needed increase in  $R_{\max}$  and decrease in width were 7.6 and 65% (Figure 4.9c,d, gray histograms). Thus, in both scenarios the histograms have shifted toward the human threshold values.

## 4.5 Discussion

The main contributions of the present study are:

1. To demonstrate that the presence of non-zero spatial gradient of velocity in optical flow patterns influences optical flow perception,
2. To quantify the ability of human individuals and monkey MST spiral selective neurons to discriminate spiral angles and demonstrated that under our experimental conditions, humans outperformed monkey MST units,
3. We have also shown that MT neurons selective for the direction of linear

motion also perform poorly at discriminating the direction of linearly moving RDPs relative to the behavioral thresholds reported by previous studies in humans, and

4. Changes in the gain or width of the neurons tuning curves could decrease neurometric thresholds in both, MT and MST units to a level similar to behavioral thresholds in humans.

A summary of our methods and findings is provided in Figure 4.10.

#### **4.5.1 Selectivity for spatial gradient of velocities**

The first main result of the current study was a selective effect of adaptation to a particular combination of spatial gradient of velocity direction and motion direction on the perception of spatial gradient of velocities. This result suggests that coding of spatial gradient of velocity direction is closely linked to the coding of motion direction in the visual system. This agrees with results from single cells studies in area MT of macaques describing neurons selective for the angle between the spatial gradient of velocity direction and the direction of linear motion [Treue and Andersen, 1996, Xiao et al., 1997]. The former authors reported that about 60% of the neurons they recorded from showed this kind of selectivity to a certain degree.

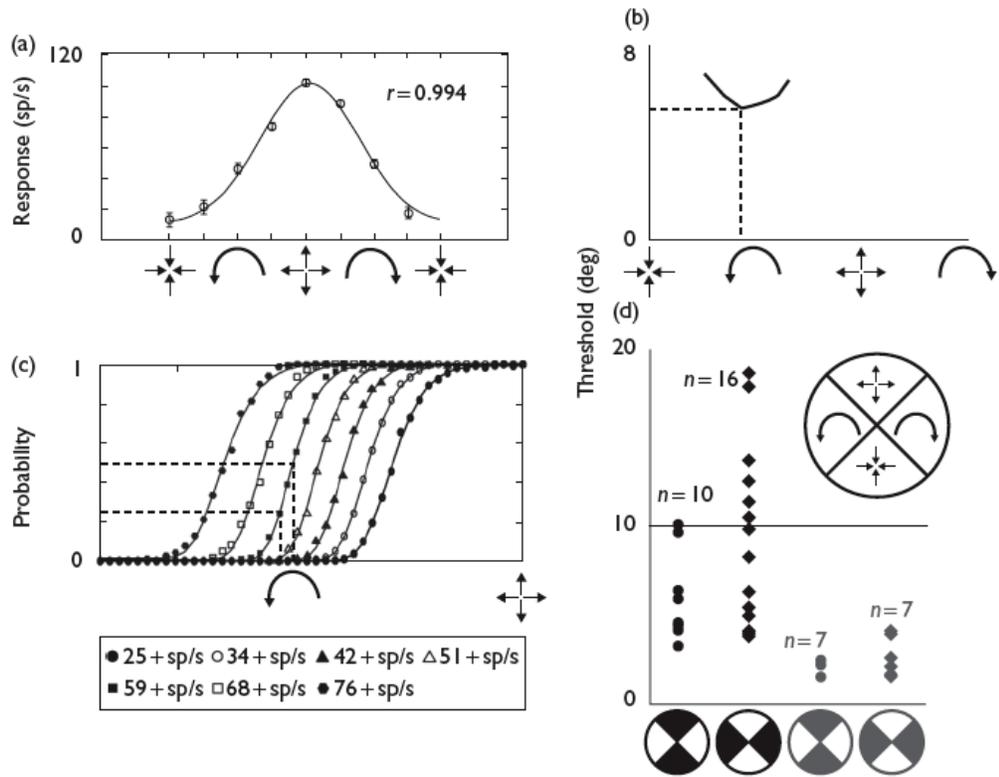


Figure 4.10: (a) Tuning curve of a medial superior temporal neuron to spiral motion. The abscissa represents spiral direction and the ordinate the response (spikes/s). Error bars represent standard errors. (b) Thresholds, and (c) Neurometric functions (symbols). (d) Minimum threshold (ordinate) of each neuron (black), and for each participant (gray). For the cells, the data are grouped according to the preferred direction of the neuron in spiral space (symbols in abscissa). For the human participants, the data were divided according to the discrimination task.

Interestingly, a recent fMRI study in humans has reported a stronger activation within human area MT/V5 by moving patterns containing spatial gradient of velocities oriented parallel to the patterns' direction relative to when the spatial gradient of velocity is removed from the same patterns [Martinez-Trujillo et al., 2005]. The same authors have suggested that similar spatial gradient of velocity direction selective mechanisms, as the ones described in the monkey [Treue and Andersen, 1996, Xiao et al., 1997], are present in human area MT/V5 and therefore they can explain this result. This could also explain the selective effect of adaptation reported in the present study.

We hypothesized that adaptation to a particular combination of spatial gradient of velocity direction and motion direction would affect neurons selective for that combination while favors neurons selective for the opposite combination (i.e., an antiparallel oriented spatial gradient of velocity and the same motion direction). This effect, however, is attenuated when the motion direction of the adapting stimulus does not match the one of the test stimulus. This may suggest that within the population of neurons encoding the same motion direction, interactions between units depend on their selectivity for spatial gradient of velocity direction. Future physiological studies may provide more insight into the mechanisms of these interactions, which likely underlie the behavioral effects

demonstrated here.

The same spatial gradient of velocity direction selective mechanisms could play a role in the perception of expanding, contracting and rotating flow fields. This may explain why when removing the contribution of such mechanisms – by removing spatial gradient of velocities from the optical flow – optical flow perception in tasks such as the one describe in the first experiment of this study is affected.

#### **4.5.2 Neuronal and behavioral thresholds**

The neurometric thresholds in our study were significantly larger than human behavioral thresholds. This result apparently disagrees with a previous report of medial superior temporal neurons having sensitivities equal or superior to monkey's thresholds for expanding and contracting spirals [Heuer and Britten, 2004]. However, in that study the authors used a task in which animals performed a coarse discrimination between expanding and contracting stimuli embedded in noise. In contrast, our task required a fine discrimination judgment, and we used 100% coherence random dot patterns. Given their tuning properties, the neurons in our study should be highly capable of discriminating spiral directions 180° apart. Previous studies in the medial superior and middle tempo-

ral area of monkeys have reported similar results as ours [Osborne et al., 2004, Purushothaman and Bradley, 2005, Gu et al., 2008].

A plausible explanation for our results is that in our study the stimulus feature (spiral angle) was not behaviorally relevant to the animal, as it performed a fixation task, leading to underestimation of the neurons sensitivity and selectivity [Spitzer et al., 1988, Dubin and Duffy, 2007, Treue and Martínez-Trujillo, 1999, Mcadams and Maunsell, 1999]. We, however, estimated that the tuning curve height (sensitivity) would need to increase by at least 400%, or the width (selectivity) decrease by 55% for the mean threshold value across units to reach the average human threshold. These effects are disproportionately larger than the ones reported in previous studies, making this explanation unlikely.

A recent study examined the ability of single neurons in the middle temporal area of monkey to discriminate linear motion direction [Cohen and Newsome, 2009] using a coarse discrimination task and similar response integration times as us (500 ms). They concluded that at such short integration times (but not at longer times) neurometric thresholds were higher than behavioral thresholds [Spitzer et al., 1988, Cohen and Newsome, 2009]. This suggests that the integration time used in the neurometric analysis plays a role in computing neuronal thresholds.

Finally, several studies in monkeys have shown that medial superior temporal neurons receive inputs from different sources, such as vestibular and eye movements signals [Thier and Erickson, 1992, Gu et al., 2008]. Under conditions in which the information from those sources becomes available, the performance of some of these neurons correlates better with behavior [Gu et al., 2008]. Thus, it may be that medial superior temporal units perform best when integrating signals from multiple sources.

To summarize, our results revealed a mismatch between the performance of a sample of individual neurons in area MST of one macaque monkey and the one of human subjects at discriminating different types of spiral stimuli. This suggests that although the activity of spiral selective neurons in this area provide useful information about spiral stimuli, this information is not sufficient to account for the performance of human subjects at discriminating different spiral types. Other areas located higher in the hierarchy of processing such as VIP and 7a may contain neurons with higher selectivity for these types of stimuli that better correlate with the subjects' performance.

## 4.6 Conclusion

We found psychophysical evidence of a role of spatial gradient of velocities in optical flow perception as well as of the existence of mechanisms selective for the relationship between spatial gradient of velocity direction and motion direction in moving stimuli. Our results show that neurometric spiral direction-discrimination thresholds of primate medial superior temporal neurons to discriminate spirals are significantly higher than behavioral thresholds. This suggests that the computations underlying this discrimination are either conducted by these units using population codes [Shadlen and Newsome, 1998], and/or by other neurons located in different brain areas with higher selectivity and/or sensitivity for spirals.

This is important for modeling since the theoretical definition of affine motion involves these derivatives, the velocity vector at each point  $(x, y)$  of an image is given as the temporal derivative of spatial position, i.e.,  $(u, v) = (dx/dt, dy/dt)$ . Spatial derivatives are then taken of each velocity component  $u$  and  $v$  in the  $x$  and  $y$  directions  $(u_x, u_y, v_x, v_y)$ . The inclusion of the derivatives form an important, biologically realistic, intermediate layer of representation for motion -in a manner similar to curvature for shape- which is present in the model of Selective Tuning model for motion [Tsotsos et al., 2005].

## 5 Discussion and Future Work

Recognition of objects and of actions is not an easy task for any computer vision system, and most current systems seem to be good at solving a specific task but less good at solving more general or more complex ones. After decades of research it remains an open problem to design systems that perform well with the efficiency, efficacy and invariance to changes of the human brain. This thesis explored several aspects of visual processes that will have an impact in future generations of object recognition and motion analysis.

This thesis first shows that an attentive recognition model perform visual search tasks. While most of the popular models present currently in the literature seem to neglect attention, e.g. [Riesenhuber and Poggio, 1999], we have shown that Selective Tuning [Tsotsos et al., 1995] performs in covert visual search tasks similarly to human performance. In fact, the feed-forward max-like mechanism prevents any effective top-down traversal since decisions regarding relevance for

elements of a neuron's receptive field are made too early (note Marr's principle of least commitment in this context [Marr, 1982]).

Secondly, we provide a biologically plausible model of 2D Shape. Our Shape Representation model makes several contributions to research in computer vision and computational neuroscience. First, it provides a biologically plausible hypothesis on how to achieve shape representation in a hierarchy of layers of neurons. Second, it demonstrates the importance of endstopping for curvature and shape. We have shown how a hierarchy starting from basic simple edge detectors, which combine into complex neurons and further endstopped and local curvatures neurons can obtain neurons that are selective for shape stimuli. And third, it validates the design of the Shape-selective neurons by matching their response to that of real neurons in area V4 of monkey with high accuracy.

Finally, we provide evidence for sensitivity to spatial gradients of velocity in the visual cortex. This is important in computer vision since the theoretical definition of affine motion involves these derivatives [Longuet-Higgins and Prazdny, 1980]. We found psychophysical evidence of a role of spatial gradients of velocity in optical flow perception. Additionally, such gradient cells in monkey show higher discrimination thresholds than found behaviorally in humans, suggesting that when discriminating spiral motions the brain

integrates information across many of these neurons to achieve its high perceptual performance.

## **5.1 Future work**

### **5.1.1 Extension of the Shape Representation Model**

#### **An improved TEO representation**

A first natural extension for shape representation would be to consider the work of [Brincat and Connor, 2004] and include more variability in a layer above our shape-selective neurons. TEO - the layer above V4 - integrates not only information regarding contours and their positions in the object but their integration may be linear or nonlinear or even have an inhibitory effect. This may be important in order to differentiate objects that having a similar shape are considered as different classes of objects in the brain.

#### **3D and area TE**

Several researchers have have reported selectivity for 3D shape in IT [Janssen et al., 2000, Janssen et al., 2001, Durand et al., 2007, Verhoef et al., 2010]. The lower bank of STS (superior temporal sulcus -

a subarea of TE) was selective to 3D shape, while lateral TE were selective to 2D shape [Janssen et al., 2000]. This is a natural extension of our model, from curves to surfaces and from shapes in a plane to shapes in 3D space. Could this area be modeled using differential geometry? The work of this same group proposes the extraction of 3D shape from disparity [Orban et al., 2006, Theys et al., 2009].

Connor's group have been working on 3D images as well. An interesting study has been published recently [Yamane et al., 2008] that follows the strategy regarding shape in areas V4 and TEO [Pasupathy and Connor, 1999, Pasupathy and Connor, 2001, Pasupathy and Connor, 2002, Brincat and Connor, 2004] but more complex for stimuli generation due to the higher variability that involves an added new dimension. The way to create the stimuli was quite ingenious, first, 50 random 3D stimuli were generated. This stimuli would evolve to different ones over a number of generations (8, 10, ...) through an evolutionary morphing algorithm, selecting the 3D shapes for the next generation based on the neuron's responses.

They measured neuronal responses in the inferotemporal cortex and found neurons selective for surface fragments in a three-dimensional curvature/orientation/position domain in a similar way as in area V4 neurons were selective to curvature at different positions. [Yamane et al., 2008] findings are

consistent with theories of configural representations where parts are complex volumetric components or *geons* [Marr and Nishihara, 1978, Biederman, 1987]. Although, neurons in area IT do not just represent single object parts but multiple parts similarly to what happened with 2D shapes. Another interesting conclusion from their study is that neural representations were different depending on viewpoint, thus partially supporting view-centered approaches.

### **Attention**

A third extension to the model of Shape Representation would be to include it within a model of attention. Since Selective Tuning [Tsotsos et al., 1995] has already been successfully tested in tasks such as visual search with a previous more simple version of the Shape Representation model (chapter 2) [Rodríguez-Sánchez et al., 2007], it is a firm candidate for a final extension to the model. Once the hierarchy is complete, attention would be an elegant strategy to find a specific shape given a cluttered scene.

### **Additional experimentation**

**Additional stimuli for TEO layer** The stimuli used in [Brincat and Connor, 2004] were created in a similar way to Pasupathy's.

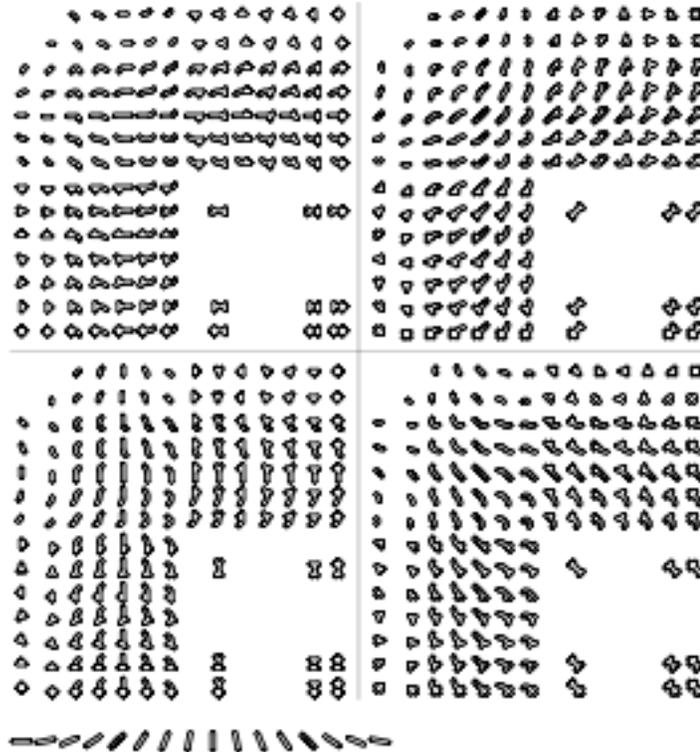


Figure 5.1: Stimuli used in [Brincat and Connor, 2004]

Convex, straight and concave contour elements were created at specific orientations and object-relative positions. These geometric elements were crossed in a large permutation matrix of smooth, closed 2D shapes. These stimuli was slightly more complex than the stimuli used by [Pasupathy and Connor, 1999] and shown in Figure 5.1.

It would be interesting to - once a TEO layer of neurons is completed into the model - evaluate how similar those TEO neurons would

be to real neurons from [Brincat and Connor, 2004] in a similar way as the study was performed comparing V4 neurons and [Pasupathy and Connor, 2001, Pasupathy and Connor, 2002].

**Experiments with real-world images** The final goal of any shape recognition system is to find objects in the real world. After our shape representation has been extended with attention, our system would be ready to accomplish recognition in cluttered scenes, at least those defined by shape. Even with the sort of images with the target partially occluded as soon as part of the main information is available. Once other features are added to the system, the system would be able to perform in more complex situations.

In addition to attention for finding an object in the real world, the system could incorporate a learning method that would extract the configuration of corners, edges and curvatures of those real world objects.

### 5.1.2 Extended testing via Selective Tuning

As mentioned in the previous chapter, a last extension to the Shape model would be the inclusion of Selective Tuning for shape search for cluttered scenes, other features are important as well from a visual search point of view such as colors,

orientation or sizes [Wolfe, 1998a].

The Selective Tuning model of attention should be able to perform a search for objects of a specific color or texture or the largest object. Size is related to shape. Size visual search finds the element that has a unique size. If the size difference is enough, then a target specified by one size can be found quite efficiently among distractors of another size [Treisman and Gelade, 1980, Quinlan and Humphreys, 1987, Duncan and Humphreys, 1992].

Color is a very important feature as visual search experiments have shown, a future addition to the model should be this feature, but not in its simplistic form of just color-opposite cells, but a more complex cells with double-color opponency in V1/V2 and color constancy in V4. Future additions should include texture as well.

As explained in Chapter 1, there exist a number of asymmetries in visual search. [Treisman and Gormican, 1988] found an asymmetry in size: it was harder to find a small target among big distractors than a big target among small distractors. It would be also interesting to see how the Selective Tuning model perform in search asymmetries such as the search for a target magenta among red distractors versus a red target among magenta distractors or find a vertical target among distractors that are tilted  $20^\circ$  than versus a  $20^\circ$  tilted target among vertical distractors

and compare it with psychophysical data in the same way as the experiments performed in chapter 2.

## Bibliography

- [Albright, 1984] Albright, T. (1984). Direction and orientation selectivity of neurons in visual area mt of the macaque. *Journal of Neurophysiology*, 52(6):1106–1130.
- [Allport, 1989] Allport, A. (1989). *Foundations of Cognitive Science*. MIT Press Bradford Books.
- [Amit, 2000] Amit, Y. (2000). A neural network architecture for visual selection. *Neural Computation*, 12:1141–1164.
- [Anderson and Van Essen, 1987] Anderson, C. and Van Essen, D. (1987). Shifter circuits: a computational strategy for dynamic aspects of visual processing. *Proceedings of the National Academy of Sciences of the United States of America*, 84(17):6297–6301.
- [Ansari and Delp, 1990] Ansari, N. and Delp, E. (1990). Partial shape recognition: A landmark-based approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):470–483.
- [Asada and Brady, 1984] Asada, H. and Brady, M. (1984). *The curvature primal sketch*. AIM 758: Massachusetts Institute of Technology (MIT).
- [Ayache and Faugeras, 1986] Ayache, N. and Faugeras, O. (1986). Hyper: A new approach for the recognition and positioning of two-dimensional objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):44–54.
- [Ballard, 1981] Ballard, D. (1981). Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122.
- [Becker, 1991] Becker, W. (1991). Saccades. In Carpenter, R. H. S., editor, *Eye Movements*. CRC Press.

- [Belongie et al., 2002] Belongie, S., Malik, J., and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(24):509–522.
- [Bennamoun and Mamic, 2002] Bennamoun, M. and Mamic, G. (2002). *Object Recognition, fundamentals and case studies*. Springer.
- [Besl and Jain, 1985] Besl, P. and Jain, R. (1985). Three-dimensional object recognition. *ACM Computational Survey*, 17(1):75–145.
- [Bichot and Schall, 1999] Bichot, N. and Schall, J. (1999). Saccade target selection in macaque during feature and conjunction visual search. *Visual Neuroscience*, 16(1):81–89.
- [Biederman, 1987] Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94(2):115–147.
- [Bishop et al., 1980] Bishop, P., Kato, H., and Orban, G. (1980). Direction-selective cells in complex family in cat striate cortex. *Journal of Neurophysiology*, 43:1266–1283.
- [Blum, 1967] Blum, H. (1967). A transformation for extracting new descriptors of shape. In Whalen-Dunn, W., editor, *Proceedings of the Symposium on Models for the Perception of Speech and Visual Form*, pages 362–380. MIT Press.
- [Blum, 1973] Blum, H. (1973). Biological shape and visual science. i. *Journal of Theoretical Biology*, 38(2):205–287.
- [Booth and Rolls, 1998] Booth, M. and Rolls, E. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral cortex*, 8(6):510–523.
- [Boussaoud et al., 1991] Boussaoud, D., Desimone, R., and Ungerleider, L. (1991). Visual topography of area teo in the macaque. *The Journal of comparative neurology*, 306(4):554–575.
- [Boynton and Hegde, 2004] Boynton, G. and Hegde, J. (2004). Visual cortex: The continuing puzzle of area v2. *Current Biology*, 14(13):R523–R524.
- [Brincat and Connor, 2004] Brincat, S. and Connor, C. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*, 7(8):880–886.

- [Britten and van Wezel, 1998] Britten, K. and van Wezel, R. (1998). Electrical microstimulation of cortical area mst biases heading perception in monkeys. *Nature Neuroscience*, 1(59–63).
- [Broadbent, 1958] Broadbent, D. (1958). *Perception and Communication*. London: Pergamon.
- [Broadbent, 1982] Broadbent, D. (1982). Task combination and selective intake of information. *Acta Psychologica*, 50:253–290.
- [Bullier, 2001] Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, 36(2-3):96–107.
- [Bushnell et al., 1981] Bushnell, M., Goldberg, M., and Robinson, D. (1981). Behavioral enhancement of visual responses in monkey cerebral cortex i. modulation in posterior parietal cortex related to selective visual attention. *Journal of Neurophysiology*, 42:755–772.
- [Canny, 1986] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698.
- [Chelazzi et al., 1998] Chelazzi, L., Duncan, J., Miller, E., and Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, 80(6):2918–2940.
- [Chelazzi et al., 1993] Chelazzi, L., Miller, E., Duncan, J., and Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, 363(6427):345–347.
- [Chen, 1993] Chen, C. (1993). Improved moment invariants for shape discriminations. *Pattern Recognition*, 26(5):683–686.
- [Cohen and Wang, 1994] Cohen, F. S. and Wang, J. (1994). Part i: Modeling image curves using invariant 3-d object curve models/spl minus/a path to 3-d recognition and shape estimation from image contours. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(1):1–12.
- [Cohen and Newsome, 2009] Cohen, M. and Newsome, W. (2009). Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *Journal of Neuroscience*, 29(20):6635–6648.

- [Connor et al., 2007] Connor, C., Brincatt, S., and Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology*, 17(2):140–147.
- [Danielsson, 1978] Danielsson, P. E. (1978). A new shape factor. *Computer Graphics and Image Processing*, 7:292–299.
- [Darroch and Mosimann, 1985] Darroch, J. N. and Mosimann, J. (1985). Canonical and principal components of shape. *Biometrika*, 72(2):241–252.
- [Davis, 1977] Davis, L. S. (1977). Understanding shape: Angles and sides. *IEEE Transactions on Computation*, C(26):236–242.
- [Dean, 1976] Dean, P. (1976). Effects of inferotemporal lesions on the behavior of monkeys. *Psychological bulletin*, 83(1):41–71.
- [Deangelis and Newsome, 1999] Deangelis, G. and Newsome, W. (1999). Organization of disparity-selective neurons in macaque area mt. *Journal of Neuroscience*, 19(4):1398–1415.
- [Deco and Zhil, 2001] Deco, G. and Zhil, J. (2001). A neurodynamical model of visual attention: Feedback enhancement of spatial resolution in a hierarchical system. *Journal of Computational Neuroscience*, 10(3):231–253.
- [Desimone and Duncan, 1995] Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review on Neuroscience*, 18:193–222.
- [Desimone et al., 1985] Desimone, R., Schein, S., Moran, J., and Ungerleider, L. (1985). Contour, color and shape analysis beyond the striate cortex. *Vision Research*, 25(3):441–452.
- [Deutsch and Deutsch, 1963] Deutsch, J. and Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological review*, 70:80–90.
- [DeValois et al., 1979] DeValois, K., DeValois, R., and Yund, E. (1979). Responses of striate cortex cells to grating and checkerboard patterns. *Journal of Physiology*, 291(1):483–505.
- [DeValois et al., 1982] DeValois, R., Albrecht, D., and Thorell, L. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22(5):545–559.

- [Dickinson et al., 2009] Dickinson, S., Leonardis, A., Schiele, B., and Tarr, M. J., editors (2009). *Object categorization, computer and human vision perspectives*. Cambridge University Press.
- [Dobbins, 1992] Dobbins, A. (1992). *Difference models of Visual Cortical neurons*. PhD thesis, Department of Electrical Engineering. McGill University.
- [Dobbins et al., 1987] Dobbins, A., Zucker, S., and Cynader, M. (1987). End-stopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*. 1987 Oct 1-7;329(6138):438-41., 329(6138):438–441.
- [Dobbins et al., 1989] Dobbins, A., Zucker, S., and Cynader, M. (1989). End-stopping and curvature. *Vision Research*, 29(10):1371–1387.
- [Dubin and Duffy, 2007] Dubin, M. and Duffy, C. (2007). Behavioral influences on cortical neuronal responses to optic flow. *Cerebral cortex*, 17:1722–1732.
- [Dubner and Zeki, 1971] Dubner, R. and Zeki, S. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research*, 35(2):528–532.
- [Dudek and Tsotsos, 1997] Dudek, G. and Tsotsos, J. (1997). Shape representation and recognition from multiscale curvature. *Computer Vision and Image Understanding*, 68(2):170–189.
- [Duncan, 1984] Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113(4):501–517.
- [Duncan and Humphreys, 1989] Duncan, J. and Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological review*, 96(3):433–458.
- [Duncan and Humphreys, 1992] Duncan, J. and Humphreys, G. (1992). Beyond the search surface: Visual search and attentional engagements. *Journal of Experimental Psychology*, 18(2):578–588.
- [Duncan et al., 1997] Duncan, J., Humphreys, G., and Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology*, 7(2):255–261.

- [Durand et al., 2007] Durand, J.-B., Nelissen, K., Joly, O., Wardak, C., Todd, J., Norman, F., Janssen, P., Vanduffel, W., and Orban, G. (2007). Anterior regions of monkey parietal cortex process visual 3d shape. *Neuron*, 55(3):493–505.
- [Egeth and Dagenbach, 1995] Egeth, H. and Dagenbach, D. (1995). Parallel versus serial processing in visual search: further evidence from subadditive effects of visual quality. *Journal of Experimental Psychology and Human Perception Performance*, 17(2):551–560.
- [Elder et al., 2003] Elder, J. H., Krupnik, A., and Johnston, L. A. (2003). Contour grouping with prior models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):661–674.
- [Eriksen and St James, 1986] Eriksen, C. W. and St James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Percept Psychophys*, 40(4):225–40.
- [Essen and Zeki, 1978] Essen, D. V. and Zeki, S. (1978). The topographic organization of rhesus monkey prestriate cortex. *Journal of Physiology*, 277:193–226.
- [Evans and Treisman, 2005] Evans, K. K. and Treisman, A. M. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance*, 31(6):1476–1492.
- [Fei-Fei et al., 2004] Fei-Fei, L., Fergus, R., and Perona, P. (2004). Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. *IEEE CVPR*, page 178.
- [Felleman and Van Essen, 1991] Felleman, D. and Van Essen, D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex*, 1(1):1–47.
- [Ferrera and Lisberger, 1995] Ferrera, V. and Lisberger, S. (1995). Attention and target selection for smooth pursuit eye movements. *Journal of Neuroscience*, 15(11):7472–7484.
- [Fidler et al., 2006] Fidler, S., Berginc, G., and Leonardis, A. (2006). Hierarchical statistical learning of generic parts of object structure. *IEEE CVPR*, (182–189).

- [Fischer and Boch, 1981] Fischer, B. and Boch, R. (1981). Selection of visual targets activates prelunate cortical cells in trained rhesus monkey. *Experimental Brain Research*, 41(3):431–433.
- [Freeman, 1974] Freeman, H. (1974). Computer processing of line-drawing images. *Computer Surveys*, 6.
- [Fujita, 2002] Fujita, I. (2002). The inferior temporal cortex: Architecture, computation, and representation. *Journal of Neurocytology*, 31(3):359–371.
- [Fujita et al., 1992] Fujita, I., Tanaka, K., Ito, M., and Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, 360(6402):343–346.
- [Fukushima, 1980] Fukushima, K. (1980). Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202.
- [Fukushima, 1988] Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Networks*, 1:119–130.
- [Fukushima et al., 1983] Fukushima, K., Miyake, S., and Ito, T. (1983). Neocognitron: A neural network model for a mechanism of visual patten recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 13(826–834).
- [Gallant et al., 1996] Gallant, J., Connor, C., Rakshit, S., Lewis, J., and Van Essen, D. (1996). Neural responses to polar, hyperbolic, and cartesian gratings in area v4 of the macaque monkey. *Journal of Neurophysiology*, 76(4):2718–2739.
- [García-Pérez, 1998] García-Pérez, M. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Research*, 38(12):1861–1881.
- [Gibson, 1950] Gibson, J. (1950). The perception of visual surfaces. *American Journal of Psychology*, 63(367–384.).
- [Goldberg and Wurtz, 1972] Goldberg, M. and Wurtz, R. (1972). Activity of superior colliculus in behaving monkey. effect of attention on neuronal responses. *Journal of Neurophysiology*, 35:560–574.
- [Gould and Dill, 1969] Gould, J. and Dill, A. (1969). Eye-movement parameters and pattern discrimination. *Perception and Psychophysics*, 6:311–320.

- [Grauman and Darrell, 2006] Grauman, K. and Darrell, T. (2006). Pyramid match kernels: Discriminative classification with sets of image features. *MIT Technical Report*, CSAIL-TR-2006-20.
- [Graziano et al., 1994] Graziano, M., Andersen, R., and Snowden, R. (1994). Tuning of MST neurons to spiral motions. *Journal of Neuroscience*, 14:54–67.
- [Grill-Spector and Kanwisher, 2005] Grill-Spector, K. and Kanwisher, N. (2005). Visual recognition: as soon as you see it, you know what it is. *Psychological Science*, 16(2):152–160.
- [Gross, 1992] Gross, C. (1992). Representation of visual stimuli in inferior temporal cortex. *Proceedings of the Royal Society of London, series B, Biological Sciences*, 335(1273):3–10.
- [Gross et al., 1972] Gross, C., Rocha-Miranda, C., and Bender, D. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, 35:96–111.
- [Grossberg et al., 1994] Grossberg, S., Mingolla, E., and Ross, W. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological review*, 101(3):470–489.
- [Gu et al., 2008] Gu, Y., Angelaki, D., and Deangelis, G. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, 11:1201–1210.
- [Hawken and Parker, 1987] Hawken, M. and Parker, A. (1987). Spatial properties of neurons in the monkey striate cortex. *Proceedings of the Royal Society of London, series B, Biological Sciences*, 231:251–288.
- [Heinke and Humphreys, 2003] Heinke, D. and Humphreys, G. (2003). Attention, spatial representation, and visual neglect: simulating emergent attention and spatial memory in the selective attention for identification model (saim). *Psychological review*, 110(1):29–87.
- [Heuer and Britten, 2004] Heuer, H. and Britten, K. (2004). Optic flow signals in extrastriate area MST: comparison of perceptual and neuronal sensitivity. *Journal of Neurophysiology*, 91(3):1314–1326.

- [Heywood et al., 1992] Heywood, C., Gadotti, A., and Cowey, A. (1992). Cortical area v4 and its role in the perception of color. *Journal of Neuroscience*, 12(10):4056–4065.
- [Hoffman and Richards, 1984] Hoffman, D. and Richards, W. (1984). Parts of recognition. *Cognition*, 18(1-3):65–96.
- [Hoffman, 1998a] Hoffman, J. (1998a). *Attention*. University College London Press.
- [Hoffman, 1998b] Hoffman, J. (1998b). *Visual Attention and eye movements, in Attention*. 119–154. Psychology Press.
- [Holliday and Meese, 2008] Holliday, I. and Meese, T. (2008). Optic flow in human vision: Meg reveals a foveo-fugal bias in v1, specialization for spiral space in hmsts, and global motion sensitivity in the ips. *Journal of Vision*, 8(10).
- [Horn and Weldon, 1986] Horn, B. K. and Weldon, E. J. (1986). Filtering closed curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:665–668.
- [Hubel and Wiesel, 1959] Hubel, D. and Wiesel, T. (1959). Receptive fields of single neurones in the cat’s striate cortex. *The Journal of physiology*, 148:574–591.
- [Hubel and Wiesel, 1962] Hubel, D. and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *Journal of Physiology*, 160:106–154.
- [Hubel and Wiesel, 1965] Hubel, D. and Wiesel, T. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 28:229–289.
- [Hubel and Wiesel, 1968] Hubel, D. and Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195(1):215–243.
- [Humphreys et al., 1989] Humphreys, G., Quinlan, P., and Riddoch, M. (1989). Grouping processes in visual search: Effects with single and combined-feature targets. *Journal of Experimental Psychology General*, 118:258–279.

- [Ito and Komatsu, 2004] Ito, M. and Komatsu, H. (2004). Representation of angles embedded within contour stimuli in area v2 of macaque monkeys. *Journal of Neuroscience*, 24(13):3313–3324.
- [Itti et al., 1998] Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259.
- [Itti et al., 2005] Itti, L., Rees, G., and Tsotsos, J. (2005). *Neurobiology of Attention*. Elsevier Science.
- [Janssen et al., 2001] Janssen, P., Vogels, R., Liu, Y., and Orban, G. (2001). Macaque inferior temporal neurons are selective for three-dimensional boundaries and surfaces. *The Journal of Neuroscience*, 21(23):9419–9429.
- [Janssen et al., 2000] Janssen, P., Vogels, R., and Orban, G. (2000). Selectivity for 3d shape that reveals distinct areas within macaque inferior temporal cortex. *Science*, 288(5473):2054–2056.
- [Jones, 2007] Jones, E. (2007). Neuroanatomy: Cajal and after cajal. *Brain Research Reviews*, 55(2):248–255.
- [Kahana and Sekuler, 2002] Kahana, M. and Sekuler, R. (2002). Recognizing spatial patterns: a noisy exemplar approach. *Vision Research*, 42(18):2177–2192.
- [Kandel et al., 2000] Kandel, E., Schwartz, J., and Jessell, T. (2000). *Principles of neural science, 4th ed.* McGraw-Hill.
- [Kastner et al., 1998] Kastner, S., De Weerd, P., Desimone, R., and Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional mri. *Science*, 282(5386):108–11. 0036-8075 Journal Article.
- [Kastner and Ungerleider, 2000] Kastner, S. and Ungerleider, L. (2000). Mechanisms of visual attention in the human cortex. *Annual Review on Neuroscience*, 23:315–341.
- [Kastner and Ungerleider, 2001] Kastner, S. and Ungerleider, L. (2001). The neural basis of biased competition in human visual cortex. *Neuropsychologia*, 39:1263–1276.

- [Kato et al., 1978] Kato, H., Bishop, P., and Orban, G. (1978). Hypercomplex and simple/complex cells classifications in cat striate cortex. *Journal of Neurophysiology*, pages 1071–1095.
- [Kehtarnavaz and deFigueiredo, 1988] Kehtarnavaz, N. and deFigueiredo, R. (1988). A 3-d contour segmentation scheme based on curvature and torsion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5):707–713.
- [Kiani et al., 2007] Kiani, R., Esteky, H., Mirpour, K., and Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, 97(6):4296–4309.
- [Kimia and Siddiqi, 1994] Kimia, B. and Siddiqi, K. (1994). Geometric heat equation and nonlinear diffusion of shapes and images. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 113–120.
- [Kimia et al., 1992] Kimia, B., Tannenbaum, A., and Zucker, S. W. (1992). On the evolution of curves via a function of curvature, 1: The classical case. *Journal of Mathematical Analysis Applications*, 163(2):438–458.
- [Kimia et al., 1995] Kimia, B., Tannenbaum, A., and Zucker, S. W. (1995). Shapes, shocks and deformations i: The components of two-dimensional shape and the reaction-diffusion space. *International Journal of Computer Vision*, 15:189–224.
- [Kobatake et al., 1998] Kobatake, E., WANG, G., and Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology*, 80:324–330.
- [Koch and Ullman, 1985] Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4(4):219–227.
- [Lazebnik et al., 2006] Lazebnik, S., Schmid, C., and Ponce, J. (2006). Beyond bags of features: spatial pyramid matching for recognizing natural scenes categories. *IEEE CVPR*, pages 2169–2178.
- [Lecun et al., 1989] Lecun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551.

- [Lecun et al., 1998] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- [Lee et al., 2007] Lee, J., Williford, T., and Maunsell, H. (2007). Spatial attention and the latency of neuronal responses in macaque area v4. *Journal of Neuroscience*, 27(36):9632–9637.
- [Lee et al., 1998] Lee, T., Mumford, D., Romero, R., and Lamme, V. (1998). The role of the primary visual cortex in higher level vision. *Vision Research*, 38(15-16):2429–2454.
- [Lennie, 1998] Lennie, P. (1998). Single units and visual cortical organization. *Perception*, 27(8):889–935.
- [LeVay et al., 1978] LeVay, S., Stryker, M., and Shatz, C. (1978). Ocular dominance columns and their development in layer iv of the cat’s visual cortex: a quantitative study. *The Journal of comparative neurology*, 179(1):223–244.
- [Levin et al., 2001] Levin, D., Takarae, Y., Miner, A., and Keli, F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception and Psychophysics*, 63:676–697.
- [Leyton, 1987] Leyton, M. (1987). Symmetry-curvature duality. *Computer Vision, Graphics and Image Processing*, 38(3):327–341.
- [Livingstone and Hubel, 1988] Livingstone, M. and Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*, 240(4853):740–749.
- [Logothetis and Sheinberg, 1996] Logothetis, N. and Sheinberg, D. (1996). Visual object recognition. *Annual Review on Neuroscience*, 19:577–621.
- [Longuet-Higgins and Prazdny, 1980] Longuet-Higgins, H. and Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 208(1173):385–397.
- [Lowe, 1999] Lowe, D. (1999). Object recognition from local scale-invariant features. *Proceedings of the International Conference on Computer Vision*.

- [Lowe, 2004] Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*.
- [Luck et al., 1997] Luck, S., Chelazzi, L., Hillyard, S., and Desimone, R. (1997). Neural mechanisms of spatial selective attention on areas v1, v2 and v4 of macaque visual cortex. *Journal of Neurophysiology*, 77:24–42.
- [Malinowski and Hubner, 2001] Malinowski, P. and Hubner, R. (2001). The effect of familiarity on visual-search performance: evidence for learned basic features. *Perception and Psychophysics*, 63(3):458–463.
- [Marcelja, 1980] Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of Optical Society of America*, 70(11):1297–1300.
- [Marimont, 1984] Marimont, D. H. (1984). A representation for image curves. *AAAI*, pages 237–242.
- [Marr, 1982] Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W.H. Freeman.
- [Marr and Nishihara, 1978] Marr, D. and Nishihara, H. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, series B, Biological Sciences*, 200(1140):269–294.
- [Martinez et al., 1999] Martinez, A., Anillo-Vento, L., Sereno, M., Frank, L., Buxton, R., Dubowitz, D., Wong, E., Hinrichs, H., Heize, H., and Hillyard, S. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, 15:359–367.
- [Martinez-Trujillo and Treue, 2004] Martinez-Trujillo, J. and Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology*, 14(9):744–751.
- [Martinez-Trujillo et al., 2005] Martinez-Trujillo, J., Tsotsos, J., Simine, E., Pomplun, M., Wildes, R., and Treue, S. (2005). Selectivity for speed gradients in human area mt/v5. *Neuroreport*, 16(435–438).
- [Maunsell and Van Essen, 1983] Maunsell, J. and Van Essen, D. (1983). Functional properties of neurons in middle temporal visual area of the macaque

- monkey. i. selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, 49(5):1127–1147.
- [Mcadams and Maunsell, 1999] Mcadams, C. and Maunsell, J. (1999). Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron*, 23(4):765–773.
- [McKee and Aggarwal, 1977] McKee, J. W. and Aggarwal, J. K. (1977). Computer recognition of partial views of curved objects. *IEEE Transactions on Computation*, C(26):790–800.
- [Mehrotra and Grosky, 1989] Mehrotra, R. and Grosky, W. (1989). Shape matching utilizing indexed hypothesis generation and testing. *IEEE Transactions on Robotics and Automation*, 5(1):70–77.
- [Mehta et al., 2000] Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2000). Intermodal selective attention in monkeys. I: distribution and timing of effects across visual areas. *Cerebral Cortex*, 10(4):343–358.
- [Mekuz and Tsotsos, 2007] Mekuz, N. and Tsotsos, J. (2007). Hierarchical learning of dominant constellations for object class recognition. *ACCV*, (492–501).
- [Merigan and Pham, 1998] Merigan, W. and Pham, H. (1998). 4 lesions in macaques affect both single and multiple-viewpoint shape discriminations. *Visual Neuroscience*, 15:359–367.
- [Milner, 1974] Milner, P. (1974). A model for visual shape recognition. *Psychological review*, 81:521–535.
- [Mokhtarian, 1995] Mokhtarian, F. (1995). Silhouette-based isolated object recognition through curvature scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):539–544.
- [Mokhtarian and Mackworth, 1984] Mokhtarian, F. and Mackworth, A. (1984). Scale-based description and recognition of planar curves and two-dimensional objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(1):34–43.
- [Moran and Desimone, 1985] Moran, J. and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, 229(4715):782–784.

- [Morrone et al., 1995] Morrone, M., Burr, D., and Vaina, L. (1995). Two stages of visual processing for radial and circular motion. *Nature*, 376(6540):507–509.
- [Motter, 1993] Motter, B. (1993). Focal attention produces spatially selective processing in visual cortical areas v1, v2, and v4 in the presence of competing stimuli. *Journal of Neurophysiology*, 70(3):909–919.
- [Motter, 1994] Motter, B. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area v4. *Journal of Neuroscience*, 14(4):2178–2189.
- [Movshon and Newsome, 1996] Movshon, A. and Newsome, W. (1996). Visual response properties of striate cortical neurons projecting to area mt in macaque monkeys. *Journal of Neuroscience*, 16(23):7733–7741.
- [Movshon et al., 1989] Movshon, J., Adelson, E., Gizzi, M., and Newsome, W. (1989). The analysis of moving visual patterns. *Experimental Brain Research*, 75(53–64).
- [Mozer and Sitton, 1998] Mozer, M. and Sitton, M. (1998). *Computational modeling of spatial attention*, pages 341–393. London: UCL Press.
- [Murphy and Finkel, 2007] Murphy, T. and Finkel, L. (2007). Shape representation by a network of v4-like cells. *Neural Networks*, 20:851–867.
- [Mutch and Lowe, 2006] Mutch, J. and Lowe, D. (2006). Multiclass object recognition with sparse, localized features. *IEEE CVPR*, pages 11–18.
- [Nagy and Sanchez, 1990] Nagy, A. and Sanchez, R. (1990). Critical color differences determined with a visual search task. *Journal of Optical Society of America A*, 7(7):1209–1217.
- [Newsome and Paré, 1988] Newsome, W. and Paré, E. (1988). the official journal of the society for neuroscience; a selective impairment of motion perception following lesions of the middle temporal visual area (mt). *The Journal of neuroscience*, 8(6):2201–2211.
- [Niebur et al., 1993] Niebur, E., Koch, C., and Rosin, C. (1993). An oscillation-based model for the neuronal basis of attention. *Vision Research*, 33(18):2789–2802.
- [Norman, 1968] Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review*, 75(6):522–536.

- [O'Connor et al., 2002] O'Connor, D., Fukui, M., Pinsk, M., and Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, 5(11):1203–1209.
- [Olshausen et al., 1993] Olshausen, B., Anderson, C., and Van Essen, D. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, 13(11):4700–4719.
- [Orban et al., 2006] Orban, G., Janssen, P., and Vogels, R. (2006). Extracting 3d structure from disparity. *Trends in Neurosciences*, 29(8):466–473.
- [Orban et al., 1979a] Orban, G., Kato, H., and Bishop, P. (1979a). Dimensions and properties of end-zone inhibitory areas of hypercomplex cells in cat striate cortex. *Journal of Neurophysiology*, 42:833–849.
- [Orban et al., 1979b] Orban, G., Kato, H., and Bishop, P. (1979b). End-zone region in receptive fields of hypercomplex and other striate neurons in the cat. *Journal of Neurophysiology*, 42:818–832.
- [Osborne et al., 2004] Osborne, L., Bialek, W., and Lisberger, S. (2004). Time course of information about motion direction in visual area mt of macaque monkeys. *The Journal of Neuroscience*, 24(13):3210–3222.
- [Pack and Born, 2001] Pack, C. and Born, R. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area mt of macaque brain. *Nature*, 409(6823):1040–1042.
- [Parent and Zucker, 1989] Parent, P. and Zucker, S. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):823–839.
- [Pasupathy and Connor, 1999] Pasupathy, A. and Connor, C. (1999). Responses to contour features in macaque area v4. *Journal of Neurophysiology*, 82(5):2490–2502.
- [Pasupathy and Connor, 2001] Pasupathy, A. and Connor, C. (2001). Shape representation in area v4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology*, 86(5):2505–2519.

- [Pasupathy and Connor, 2002] Pasupathy, A. and Connor, C. (2002). Population coding of shape in area v4. *Nature Neuroscience*, 5(12):1332–1338.
- [Pavlidis, 1972] Pavlidis, T. (1972). Segmentation of pictures and maps through functional approximation. *Computer Graphics and Image Processing*, 1:360–372.
- [Persoon and Fu, 1974] Persoon, E. and Fu, K. S. (1974). Shape discrimination using fourier descriptors. *Proceedings of IJ CPR*, pages 126–130.
- [Pessoa et al., 2003] Pessoa, L., Kastner, S., and Ungerleider, L. (2003). Neuroimaging studies of attention: From modulation of sensory processing to top-down control. *Journal of Neuroscience*, 23(10):3990–3998.
- [Phaf et al., 1990] Phaf, R. H., Van der Heijden, A. H., and Hudson, P. T. (1990). SLAM: a connectionist model for attention in visual selection tasks. *Cognitive Psychology*, 22(3):273–341.
- [Pinto et al., 2008] Pinto, N., Cox, D., and Dicarlo, J. (2008). Why is real-world visual object recognition hard? *PLoS Computational Biology*, 4(1):151–156.
- [Poggio and Fischer, 1977] Poggio, G. and Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *Journal of Neurophysiology*, 40:1392–1405.
- [Postma et al., 1997] Postma, E. O., vandenHerik, H. J., and Hudson, P. T. W. (1997). SCAN: A scalable model of attentional selection. *Neural Networks*, 10(6):993–1015.
- [Press et al., 1994] Press, W., Knierim, J., and Van Essen, D. (1994). Neuronal correlates of attention to texture patterns in macaque striate cortex. *Society for Neuroscience*, 20:838.
- [Purushothaman and Bradley, 2005] Purushothaman, G. and Bradley, D. (2005). Neural population code for fine perceptual decisions in area mt. *Nature neuroscience*, 8(1):99–106.
- [Quinlan and Humphreys, 1987] Quinlan, P. and Humphreys, G. (1987). Visual search for targets defined by combinations of color, shape and size: An examination of the task constraints on feature and conjunction searches. *Perception and Psychophysics*, 41(5):455–472.

- [Ramon-Cajal, 1888] Ramon-Cajal, S. (1888). Sobre las fibras nerviosas de la capa molecular del cerebelo. *Rev Trim Histol Norm Patol*, 1:33–49.
- [Ramon-Cajal, 1894] Ramon-Cajal, S. (1894). The croonian lecture: La fine structure des centres nerveux. *Royal Society of London Proceedings Series I*, 55:444–468.
- [Ramon-Cajal, 1904] Ramon-Cajal, S. (1904). Variaciones morfológicas, normales y patológicas del retículo neurofibrilar. *Trab Lab Investig Biol Madrid*, 3:9–15.
- [Rao and Ballard, 1997] Rao, R. and Ballard, D. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation*, 9(4):721–763.
- [Rao and Ballard, 1999] Rao, R. and Ballard, D. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87.
- [Reynolds et al., 2000] Reynolds, J., Pasternak, T., and Desimone, R. (2000). Attention increases sensitivity of v4 neurons. *Neuron*, 26:703–714.
- [Richards et al., 1988] Richards, W., Dawson, B., and Whittingham, D. (1988). *Encoding contour shape by curvature extrema*, pages 83–98. MIT Press.
- [Richards and Hoffman, 1984] Richards, W. and Hoffman, D. D. (1984). Codon constraints on closed 2d shapes. AI Memo 769, MIT, AI lab.
- [Riesenhuber and Poggio, 1999] Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019–1025.
- [Riesenhuber and Poggio, 2000] Riesenhuber, M. and Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, 3 Suppl:1199–1204.
- [Riesenhuber and Poggio, 2002] Riesenhuber, M. and Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology*, 12(2):162–168.
- [Rizzolatti et al., 1987] Rizzolatti, G., Riggio, L., Dascola, I., and Umiltà, C. (1987). Reorienting attention across the horizontal and vertical meridians -

- evidence in favor of a premotor theory of attention. *Neuropsychologia*, 25:31–40.
- [Rodríguez-Sánchez et al., 2006] Rodríguez-Sánchez, A., Simine, E., and Tsotsos, J. (2006). Feature Conjunctions in Serial Visual Search. In Kollias, S., editor, *Lecture Notes in Computer Science 4132*, volume 2, pages 498–507. Springer-Verlag Berlin Heidelberg.
- [Rodríguez-Sánchez et al., 2007] Rodríguez-Sánchez, A., Simine, E., and Tsotsos, J. (2007). Attention and visual search. *International Journal of Neural Systems*, 17(4).
- [Rolls, 1987] Rolls, E. (1987). *Information representation, processing and storage in the brain: analysis at the single neuron level*, in *The Neural and Molecular bases of learning*. Number 503–540. Wiley-Interscience.
- [Rolls and Deco, 2002] Rolls, E. and Deco, G. (2002). *Computational Neuroscience of Vision*. Oxford.
- [Rosenblatt, 1962] Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Spartan Books.
- [Rothenstein et al., 2008] Rothenstein, A., Rodriguez-Sanchez, A., Simine, E., and Tsotsos, J. (2008). Visual feature binding within the selective tuning attention framework. *Int. J. Pattern Recognition and Artificial Intelligence - Special Issue on Brain, Vision and Artificial Intelligence*, 22(5):861–881.
- [Rumelhart and McClelland, 1986] Rumelhart, D. and McClelland, J. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition : Foundations (Parallel Distributed Processing)*. MIT Press.
- [Sandon, 1990] Sandon, P. (1990). Simulating visual attention. *Journal of Cognitive Neuroscience*, 2(3):213–231.
- [Schein and Desimone, 1990] Schein, S. and Desimone, R. (1990). Spectral properties of v4 neurons in the macaque. *Journal of Neuroscience*, 10(10):3369–3389.
- [Schein et al., 1982] Schein, S., Marrocco, R., and de Monasterio, F. (1982). Is there a high concentration of color-selective cells in area v4 of monkey visual cortex? *Journal of Neurophysiology*, 47(2):193–213.

- [Schiller et al., 1976] Schiller, P., Finlay, B., and Volman, S. (1976). Quantitative studies of single-cell properties in monkey striate cortex. i. spatiotemporal organization of receptive fields. *Journal of Neurophysiology*, 39(6):1288–1319.
- [Schneider, 1995] Schneider, W. (1995). Vam: neuro-cognitive model for visual attention control of segmentation, object recognition and space-based motor action. *Visual Cognition*, 2:331–375.
- [Serre, 2006] Serre, T. (2006). *Learning a Dictionary of Shape-Components in Visual Cortex: Comparison with Neurons, Humans and Machines*. PhD thesis, Massachusetts Institute of Technology.
- [Serre et al., 2005] Serre, T., Wolf, L., , and Poggio, T. (2005). Object recognition with features inspired by visual cortex. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*.
- [Serre et al., 2007] Serre, T., Wolf, L., Bileschi, S., and Riesenhuber, M. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3):411–426.
- [Shadlen and Newsome, 1998] Shadlen, M. and Newsome, W. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18(10):3870–3896.
- [Shahraray and Anderson, 1989] Shahraray, B. and Anderson, D. J. (1989). Optimal estimation of contour properties by cross-validated regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:600–610.
- [Sharvit et al., 1998] Sharvit, D., Chan, H., Tek, H., and Kimia, B. (1998). Symmetry-based indexing of image databases. *Proceedings IEEE Workshop on Content-Based Access of Image and Video Libraries.*, pages 56–62.
- [Shulman et al., 1979] Shulman, G., Remington, R., and McLean, J. (1979). Moving attention through visual space. *Journal of Experimental Psychology*, 92:428–431.
- [Siddiqi and Pizer, 2008] Siddiqi, K. and Pizer, M. (2008). *Medial Representations: Mathematics, Algorithms and Applications*, chapter 1. Springer.
- [Simine, 2006] Simine, E. (2006). Motion Model: Extending and improving performance and providing biological evidence for motion change detectors. Master’s thesis, York University.

- [Snowden et al., 1992] Snowden, R., Treue, S., and Andersen, R. (1992). The response of neurons in areas v1 and mt of the alert rhesus monkey to moving random dot patterns. *Experimental Brain Research*, 88(2):389–400.
- [Spitzer et al., 1988] Spitzer, H., Desimone, R., and Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, 240(4850):338–340.
- [Spitzer and Hochstein, 1985] Spitzer, H. and Hochstein, S. (1985). A complex-cell receptive-field model. *Journal of Neurophysiology*, 53(5):1266–1286.
- [Sutherland, 1998] Sutherland, S. (1998). Feature selection. *Nature*, 392(350).
- [Suzuki et al., 2004] Suzuki, N., Hashimoto, N., Kashimori, Y., Zheng, M., and Kambara, T. (2004). A neural model of predictive recognition in form pathway of visual cortex. *BioSystems*, 76:33–42.
- [Tanaka, 1996] Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review on Neuroscience*, 19:109–139.
- [Tanaka and Saito, 1989] Tanaka, K. and Saito, H. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology*, 62:626–641.
- [Tanaka et al., 1991] Tanaka, K., Saito, H., Fukada, Y., and Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, 66(1):170–189.
- [Theys et al., 2009] Theys, T., van Loon, T., Goffin, J., Srivastava, S., Orban, G., and Janssen, P. (2009). Disparity-defined 3d boundary and surface selectivity in macaque parietal cortex. *Surgical Neurology*, 72(5):523–524.
- [Thier and Erickson, 1992] Thier, P. and Erickson, R. (1992). Responses of visual tracking neurons from cortical area mst-i to visual, eye and head motion. *European Journal of Neuroscience*, 4:539–553.
- [Thompson et al., 2005] Thompson, K., Bichot, N., and Sato, T. (2005). Frontal eye field activity before visual search errors reveals the integration of bottom-up and top-down salience. *Journal of Neurophysiology*, 93(1):337–351.

- [Thorpe et al., 1996] Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582):520–522.
- [Treisman and Gelade, 1980] Treisman, A. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136.
- [Treisman and Gormican, 1988] Treisman, A. and Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological review*, 95:15–48.
- [Treisman, 1964] Treisman, A. M. (1964). The effect of irrelevant material on the efficiency of selective listening. *American Journal of Psychology*, 77:533–546.
- [Treue and Andersen, 1996] Treue, S. and Andersen, R. (1996). Neural responses to velocity gradients in macaque cortical area mt. *Visual Neuroscience*, 13(4):797–804.
- [Treue and Martínez-Trujillo, 1999] Treue, S. and Martínez-Trujillo, J. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399(6736):575–579.
- [Ts'o et al., 2009] Ts'o, D., Zarella, M., and Burkitt, G. (2009). Whither the hypercolumn? *Journal of Physiology*, 587(12):2791–2805.
- [Tsotsos, 1988] Tsotsos, J. (1988). A complexity level analysis of immediate vision. *International Journal of Computer Vision*, 2(1):303–320.
- [Tsotsos, 1990] Tsotsos, J. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, 13(3):423–444.
- [Tsotsos, 1992] Tsotsos, J. (1992). On the relative complexity of passive vs active visual search. *International Journal of Computer Vision*, 7(2):127–141.
- [Tsotsos et al., 1995] Tsotsos, J., Culhane, S., Winky, W., Lai, Y., Davis, N., and Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, 78(1-2):507–545.
- [Tsotsos et al., 2005] Tsotsos, J., Liu, Y., Martineztrujillo, J., Pomplun, M., Simine, E., and Zhou, K. (2005). Attending to visual motion. *Computer Vision and Image Understanding*, 100(1-2):3–40.

- [Tsotsos et al., 2008] Tsotsos, J., Rodríguez-Sánchez, A., Rothenstein, A., and Simine, E. (2008). The different stages of visual recognition need different attentional binding strategies. *Brain research*, 1225:119–132.
- [Tsunoda et al., 2001] Tsunoda, K., Yamane, Y., Nishizaki, M., and Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*, 4(8):832–838.
- [Vaina, 1998] Vaina, L. (1998). Complex motion perception and its deficits. *Current opinion in neurobiology*, 8(4):494–502.
- [Valois and De Valois, 1990] Valois, R. D. and De Valois, K. (1990). *Spatial Vision*. Oxford University Press, New York.
- [Verhoef et al., 2010] Verhoef, B.-E., Vogels, R., and Janssen, P. (2010). Contribution of inferior temporal and posterior parietal activity to three-dimensional shape perception. *Current Biology*.
- [Viviani and Swensson, 1982] Viviani, P. and Swensson, R. (1982). Saccadic eye movements to peripherally discriminated visual targets. *Journal of Experimental Psychology. Human perception and performance*, 8(1):113–126.
- [von der Heydt et al., 1984] von der Heydt, R., Peterhans, E., and Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, 224(4654):1260–1262.
- [von der Malsburg, 1973] von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14(2):85–100.
- [von Helmholtz, 1924] von Helmholtz, H. (1924). Mechanism of accommodation. in j. p. c. southall (ed.), *helmholtz’s treatise on physiological optics*. *The Optical Society of America*, pages 143–173.
- [Wallis and Rolls, 1997] Wallis, G. and Rolls, E. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology*, 51(2):167–194.
- [Wang et al., 1998] Wang, G., Manabu, T., and Tanaka, K. (1998). Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. *Neuroscience Research*, 32:33–46.

- [Wang et al., 2006] Wang, G., Zhang, Y., and Fei-Fei, L. (2006). Using dependent regions for object categorization in a generative framework. *IEEE CVPR*, pages 1597–1604.
- [Webster and Ungerleider, 1998] Webster, M. and Ungerleider, L. (1998). *Neuroanatomy of Visual Attention*, in *The attentive brain*. MIT Press.
- [Wilkinson et al., 2000] Wilkinson, F., James, T., Wilson, H., Gati, J., Menon, R., and Goodale, A. (2000). An fmri study on the selective activation of human extrastriate form vision areas by radial and concentric gratings. *Current Biology*, 10:1455–1458.
- [Williams, 1966] Williams, L. (1966). The effect of target specification on objects fixated during visual search. *Perception and Psychophysics*, 1:315–318.
- [Wilson and Wilkinson, 1998] Wilson, H. and Wilkinson, F. (1998). Detection of global structures in glass patterns: implications for form vision. *Vision Research*, 38:2933–2947.
- [Wilson et al., 1997] Wilson, H., Wilkinson, F., and Assad, W. (1997). Concentric orientation summation in human form vision. *Vision Research*, 37(17):2325–2330.
- [Wolfe, 1998a] Wolfe, J. (1998a). *Visual Search*, in *Attention*. University College London Press.
- [Wolfe, 2001] Wolfe, J. (2001). Asymmetries in visual search: an introduction. *Perception and psychophysics*, 63(3):381–389.
- [Wolfe et al., 1989] Wolfe, J., Cave, K., and Franzel, S. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of experimental psychology. Human perception and performance*, 15(3):419–433.
- [Wolfe et al., 1992] Wolfe, J., Friedman-Hill, S., Stewart, M., and O’Connell, K. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology and Human Perception Performance*, 1(879–892).
- [Wolfe, 1998b] Wolfe, J. M. (1998b). What can 1,000,000 trials tell us about visual search? *Psychological Science*, 9(1):33–39.

- [Wurtz and Mohler, 1976] Wurtz, R. and Mohler, C. (1976). Enhancement of visual responses in monkey striate cortex and frontal eye fields. *Journal of Neurophysiology*, 39(4):766–772.
- [Xiao et al., 1997] Xiao, D., Marcar, V., Raiguel, S., and Orban, G. (1997). Selectivity of macaque mt/v5 neurons for surface orientation in depth specified by motion. *The European journal of neuroscience*, 9(5):956–964.
- [Xu and Carey, 1996] Xu, F. and Carey, S. (1996). Infants’ metaphysics: the case of numerical identity. *Cognitive psychology*, 30:111–153.
- [Yamane et al., 2008] Yamane, Y., Carlson, E., Bowman, K., Wang, Z., and Connor, C. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nature Neuroscience*, 11(11):1352–1360.
- [Yoshioka and Dow, 1996] Yoshioka, T. and Dow, B. (1996). Color, orientation and cytochrome oxidase reactivity in areas v1, v2 and v4 of macaque monkey visual cortex. *Behavioral and Brain Research*, 76(1-2):71–88.
- [Zeki, 1983] Zeki, S. (1983). Colour coding in the cerebral cortex: the responses of wavelength-selective and colour-coded cells in monkey visual cortex to changes in wavelength composition. *Neuroscience*, 9(4):767–781.
- [Zeki et al., 1999] Zeki, S., Aglioti, S., McKeefry, D., and Berlucchi, G. (1999). The neurobiological basis of conscious color perception in a blind patient. *PNAS*, 96:14124–14129.
- [Zhang et al., 2006] Zhang, H., Berg, A., Marie, M., and Malik, J. (2006). Svm-knn: Discriminative nearest neighbor classification for visual category recognition. *IEEE CVPR*, pages 2126–2136.
- [Zucker, 1981] Zucker, S. W. (1981). Computer vision and human perception: An essay on the discovery of constraints. *Proceedings of the International Conference on Artificial Intelligence*, pages 1102–1116.