From cones to contours: a parallel simulation of neural mechanisms in the primate vision system

Winder, Simon A. J.

Award date:
1995

Awarding institution:
University of Bath

From Cones to Contours

A parallel simulation of neural mechanisms in the primate vision system.

submitted by

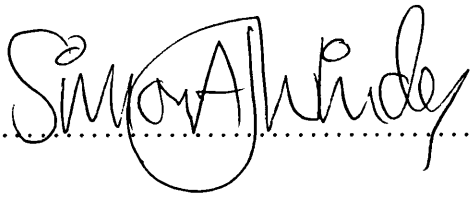
Simon A. J. Winder

for the degree of **Ph.D**

of the **University of Bath**

1995

Copyright

Signature of Author .....  SIMON A. J. WINDER

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Summary

This thesis presents a large-scale computer simulation of neural mechanisms responsible for low-level colour and form vision in the primate. Image processing techniques are employed to simulate the activity of around 40 million neurons in the retina, lateral geniculate nucleus and cortical area V1. This simulation, running on a data-parallel machine, calculates neural responses that should be produced when a variety of natural and artificial images are presented to the animal. The main aim is to investigate and to determine the significance of some of the transformations and representations which comprise the first stages of vision.

Specific areas of study include:

- Development of a realistic simulation scenario for retinal stages, involving a display device model, non-linear receptor model and spatial and chromatic opponency.
- Simulation of contrast gain control mechanisms in the retina that involve M-type ganglion cells. Results show that a compressive contrast response can be obtained, without introducing local non-linearities.
- The coding of chrominance and luminance information in the responses of parvocellular neurons in the geniculate nucleus and the interpretation of these signals by later stages of the vision system. This work clarifies some long-standing issues from the literature with the aid of simulation techniques.
- Simulation of simple cell and complex cell behaviours in area V1, including a novel theory of spatial sampling in the visual cortex.
- Simulation of cortical contrast gain control mechanisms operating across the dimensions of space, orientation and spatial frequency. Results show that many diverse interactions can be accounted for by the presence of gain control, and that this process is effective in reducing the noise sensitivity of neurons in the superficial layers of V1.

- Simulation of long-range divisive end, side and surround-stopping. Apart from introducing specific feature selectivity, these interactions are found to introduce useful texture saliency effects. They also alter orientation perception. Results successfully predict observed surround-induced orientation tilt illusions.

- A theory of convergent mechanisms that implement generalisation over space. Results show that a linear spatial summation process applied to responses from neurons with surround-inhibited receptive fields is able to gather information about features over a large region of space without changing the apparent response selectivity of the underlying centre mechanism. This has a wide application throughout the vision system.

Image processing simulation is found to be an effective way of understanding and visualising the ensemble behaviour of large numbers of neurons in the visual pathway.

Acknowledgements

Thanks are expressed to my supervisor, Dr. Martin Oliver, for advice during the project, for reading the thesis, and for tolerating the changes of direction which finally led to this area of research. I also wish to thank other members of the Computing Group at the *University of Bath*, in particular Dr. Simon Merrall for answering technical questions on LaTeX and the drawing package `tgif`, Dr. Russell Bradford and Dr. John Abbott for checking selected portions of the thesis, and Jerry Leach for many and varied distractions during the life of this research.

I am very grateful to Dr. Andrew Parker for his time during a lengthy discussion at the University Laboratory of Physiology, Oxford in 1994.

I would like to acknowledge the fact that the image of the girl holding the colour test card was scanned from an article by J. J. McCann and K. L. Houston in "Colour Vision" edited by J. D. Mollon and L. T. Sharpe (1983).

I wish to express special thanks to Tim Paget, Steven King, my parents and to my brother Daniel, for putting up with my tendency to talk incessantly about vision.

Finally, financial support provided by the Engineering and Physical Sciences Research Council from October 1991 to September 1994 is gratefully acknowledged.

Contents

Chapter 1

Introduction

1.1 Introduction

The sense of sight is enigmatic, resistant to introspection, and yet it is fundamental to our interaction with the environment and to our understanding and imagination. The purpose of this thesis is to investigate and reproduce some of the hidden processes that underlie our perception of the world.

For a long time, the study of human visual perception lay firmly in the realm of psychology. In the 1930s the Gestalt school made a strong attempt to describe the properties of vision in terms of laws of grouping leading to the formation of "wholes". These laws were thought to be the direct expression of natural physical laws (e.g. diffusion or electrostatics) operating within the brain. Unfortunately, the school suffered because of limited concurrent knowledge of neural principles and a tendency to treat descriptions of phenomenology as explanations.

More recently, understanding of information theory arising together with the development of computers has helped to make us more aware of what constitutes an explanation. At the same time, the perfection of single neuron recording techniques, the development of non-invasive investigation procedures and extensive research into neural physiology and pharmacology, have rapidly improved our knowledge of the organisation of the brain and the types of interactions that take place within it. Inspiration from all these areas has fed back constructively into psychological research.

In this thesis I describe an attempt to simulate activity within the neural pathways that mediate primate colour and form vision. The scope of the simulation is intended to be quite broad in order to tie together many current ideas. In order to achieve this goal, it has

been necessary to draw information from a wide range of sources. Facts about the organisation and functionality of biological vision systems are obtained by the use of comprehensive neurophysiological experimentation, with cats or monkeys as subjects. This information is complemented by data from psychophysical experiments carried out on humans. Further information comes from medical sources—from experiments conducted during brain operations, and especially from case studies of patients who have suffered localised brain damage affecting some aspects of vision without leaving total blindness.

Significant correlation between these sources of information leads us to the conclusion that human and primate vision systems are similarly organised and so results from one system are in many cases applicable to the other.

Transformations and Representations

In the 1960s, Hubel and Wiesel conducted pioneering work which involved recording the electrical activity produced by single neurons within the monkey's visual cortex. They found neurons that seemed to respond to visually useful primitives, such as "edges" or "bars", when presented at specific locations in the scene. This has led to the view that vision is a process by which sense data about the real world is transformed through a series of stages to yield abstract knowledge represented in a form that is useful to the organism.

In his influential book *Vision*, David Marr [154] wrote:

> "Vision is therefore, first and foremost, an information-processing task, but we cannot think of it just as a process. For if we are capable of knowing what is where in the world, our brains must somehow be capable of *representing* this information—in all its profusion of color and form, beauty, motion, and detail. The study of vision must therefore include not only the study of how to extract from images the various aspects of the world that are useful to us, but also an inquiry into the nature of the internal representations by which we capture this information and thus make it available as a basis for decisions about our thoughts and actions."

In this quotation Marr puts forward an essential directive for perceptual research: The requirement for an explanation in terms of transformations and representations.

The starting point for vision is a pattern of light focussed on the retina. This forms an initial representation in terms of point-wise light intensities but does not make explicit any knowledge of scene content. The apocryphal neuron that fires only when one's grandmother

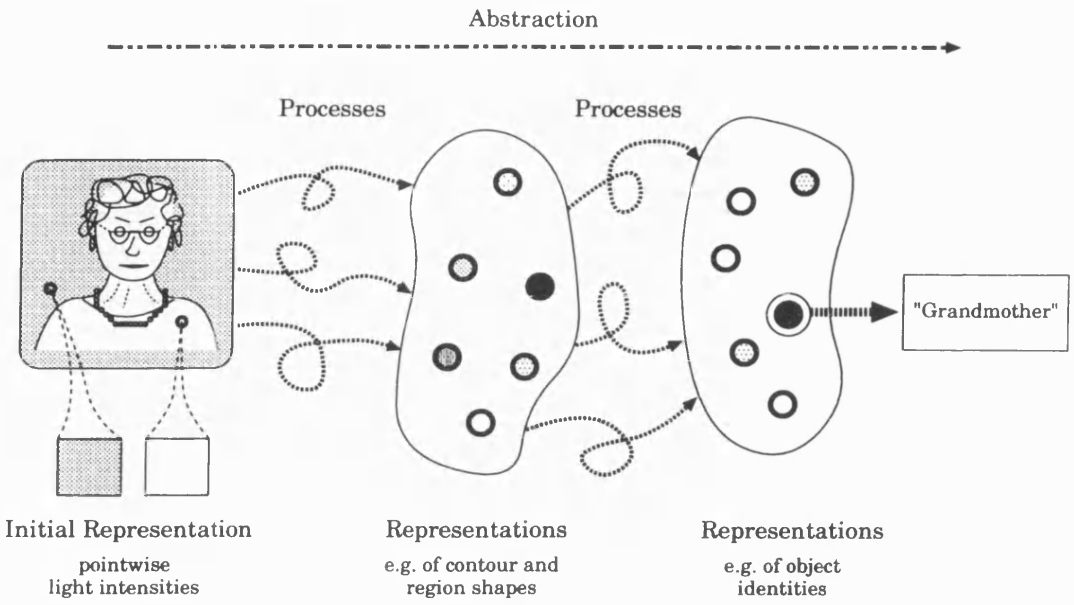


Figure 1-1: Representations and Transformational Processes

is present in the scene is another example of representation. In this case, a certain occurrence (the neuron becoming activated) is coding for some specific property of the outside world.[1] The computations necessary to abstract this information from previous representations of the visual scene are examples of transformational processes (Figure 1-1).

In this thesis I attempt to follow Marr's advice by approaching the following two questions at each stage in the visual hierarchy: How is information transformed at this stage? What is being represented? These questions form a useful basis for enquiry.

1.2 A Large-Scale Simulation

Vision involves significant computation. It is only recently, with the advent of fast computers, large amounts of memory, high quality graphical display devices and parallel processing, that it is becoming feasible to "program vision" as Marr proposed. In particular, I am concerned here with developing a large-scale simulation of neural processes, and as such this work can be identified with large simulations of weather systems, fluid dynamics, gravitating bodies and many others that are making use of the new technology.

The present computer model runs on a data-parallel machine and employs image pro-

[1] This is an example of the "single neuron hypothesis" of cortical representation. See Gross [90] for a discussion of this, and more realistic population coding ideas.

cessing techniques to simulate the activity of around 40 million neurons at stages along the stream of processing from the retina to area V1 in the occipital cortex. The neurons considered are those which receive input from the central 10° of the visual field. The simulation is intended to include all the well-documented spatial and chromatic interactions observed between neurons within the retina and lateral geniculate nucleus (LGN), as well as spatial interactions that take place within the primary visual cortex (V1).

For the purposes of this introduction it is convenient to think of the primate vision system as a hierarchy of stages. At present there are a number of mathematical models which describe various stage-to-stage transformations (e.g. the Gabor model for simple cell receptive fields of Marčelja [157]). Few people have used these models to simulate the actual collective behaviour of large numbers of neurons when the subject views a natural or artificial scene. This is partly because models are often inspired by the results of single cell recording techniques and these lead to a *neuron-centred* rather than an *assembly-centred* viewpoint. Such recording techniques can elucidate response properties with careful experimentation, but cannot easily show how each neural response relates to the ensemble of neurons which are also activated to a greater or lesser extent by test stimuli.

The work described here is an attempt to improve this situation by using some established and some new models as a basis for an *image processing* style approach. The aim is to simulate the response pattern that results at each stage given a certain scene, bearing in mind all of the previous layers of processing.

Not only is simulation of this nature a useful method of visualising the way in which large collections of neurons respond, it is also an important method for providing information about which models are valid. For example, we may believe that a particular model is appropriate for one transformational stage and we may also think that we know what measure the stage is computing, but a simulation may actually reveal false responses or side effects produced by the model requiring that it should be modified. Such problems would not have been apparent initially, and they are only noticeable in hindsight with the benefit of a simulation.

Simulation allows us to visualise many responses and see how the activity of each cell relates to that of its neighbours. This leads to a more informed idea about how information is represented. In addition, to carry out a simulation requires us to specify everything in the model that we are using, and therefore we are forced to quantify interactions and gather the appropriate data. A simulation can then highlight difficulties with the model which would otherwise not have been noticeable. It can also produce new, interesting, and sometimes

unexpected results, and through these, we can begin to see what problems need to be solved by later stages of visual processing.

Obviously the final success of a large-scale simulation depends on the fidelity of each constituent model—and accurate models rely on accurate descriptions of observed responses at each stage, as well as on an inspired choice of experimental test stimuli to uncover the critical behaviours. With this proviso, the results presented here offer significant insights into commonly encountered interactions among various cell types.

1.3 Objectives of the Research Project

There are a number of primary research aims that will be directly addressed in this thesis. These objectives are:

- To elucidate mechanisms of low-level visual perception in primate vision, in terms of how information is represented and transformed at each stage, particularly with regard to any emergent properties that have not been previously recognised.
- To identify and incorporate a sequence of neural models into a broad ranging but detailed computer simulation pitched at a level appropriate for establishing the main functional significance of each stage in the hierarchy.
- To target colour and form perception as prime areas of study. This leaves out the processing of motion or stereopsis.

As well as these main points, there are three secondary objectives which are relevant: In the first place, I attempt to reinforce the role of simulation in the biological scene by showing that it is a strategic tool for analysis and prediction. Secondly, since brain regions concerned with successively more abstract aspects of form analysis become harder and harder to investigate, it is pertinent to use this tool to make predictions and identify reasonable roles for later stages of neural processing. These can then be verified or rejected by experimental work—which leads to the third aim: suggesting areas for future experimentation.

1.4 Knowing "What is Where" in this Thesis

The rest of the material is arranged in the following way:

In Chapters 2–3 I introduce the subject, present previous simulation work and describe the hardware system, software development and general neural model. In Chapters 4–6 I

describe stages in the simulation from the retina to the striate cortex, introducing important concepts as they are needed. In Chapter 7 I concentrate on remote spatial interactions in V1 and in Chapter 8 I introduce more speculative work relating to general concepts of spatial summation and interations between brain areas. Chapter 9 concludes the thesis.

In Chapter 2 I provide an introduction to visual neurophysiology—introducing neurons, receptive fields, the retina, lateral geniculate nucleus, primary visual cortex, visual area V2 and the motion, form and colour pathways. The architecture of the visual cortex and selected receptive field properties are discussed. I then go on to cover simulation and modelling work that has been previously published by others.

Chapter 3 is split into two. In the first half I describe my simulation strategy and begin by introducing a "situational model" which provides a physical context for work presented here. The level at which neural modelling takes place, the type of models used and assumptions made about what neurons can do, are all included here. In this context, spatial sampling and convolution theory are introduced. In the second half of the chapter I describe the architecture of the data-parallel computer on which simulations were run, together with the software environment that I have developed to make this work possible.

In Chapter 4 I describe the TV model, receptor model and spatial opponency stages which are essential parts of this simulation of the retina. Contrast and brightness issues are also discussed. In the latter half of the chapter I describe a novel simulation of the effects of retinal contrast gain control.

In Chapter 5 I introduce the lateral geniculate nucleus and concentrate on the spatial and chromatic information that is carried by parvocellular receptive fields. In particular, I show how neurons with "type I" receptive fields transmit separable information about colour and brightness.

In Chapter 6 I describe the simulation of cortical receptive fields, including the "simple" and "complex" varieties. In addition, I describe three streams of visual processing relating to contour formation, feature detection and colour/brightness perception. I consider the way that information about features and contours is represented in V1. A novel sampling theory is also introduced.

Cortical contrast gain control and the simulation of end, side and surround-stopping is the subject of Chapter 7. For this simulation stage I include the effects of interactions among different spatial frequency and orientation channels. In this chapter I demonstrate that contrast gain control introduces noise insensitivity. I also find that long-range inhibitory

interactions produce a saliency modulated response in V1 and that the surround-induced orientation tilt illusion is a side effect of this.

In Chapter 8 I introduce a theory of spatial summation that has a bearing on processing throughout the vision system which is concerned with the generalisation of stimulus selectivities over space. In advancing this concept, I describe the effect of corticogeniculate feedback and include this in a simulation of the behaviour of "special complex" cells.

In Chapter 9 I conclude the thesis, suggest further work and describe the benefits of an image processing based simulation approach.

I believe that the most significant contributions in this thesis are the simulation of retinal contrast gain control at the end of Chapter 4, the analysis of spatiochromatic properties of LGN receptive fields in Chapter 5, the simulation of end, side and surround-stopping in the second half of Chapter 7 and the ideas on summation control that are expressed in Chapter 8.

Chapter 2

Background

2.1 Introduction

In this chapter I introduce the structure of the primate vision system and describe evidence from neurophysiology and psychophysics. After this, I review previous computational modelling and simulation work.

2.2 Introducing the Vision System

The primate brain (monkey and human) has been studied using a variety of scientific techniques. The structure of the brain and its neural pathways have been mapped out by physiologists and anatomists so that we now have a good idea of its large-scale architecture and operation. At a lower level, new brain areas are being identified and defined all the time as functionally distinct regions are found by direct electrical recording.

Figure 2-1 shows a picture of the brain from the side and underneath and includes the neural pathway that is responsible for visual perception. Nerves from the back of each eye pass under the brain and eventually find their way to the part of the cerebral cortex concerned with vision (area V1) located at the back of the head. From here, information about the visual scene is transmitted to the parietal lobe (towards the top of the head) and to the temporal lobes (around each side of the head). These regions deal with motion/spatial awareness and visual recognition respectively. Figure 2-2 shows some of the pathways and brain areas concerned with vision that have been identified by electrical recording methods. This diagram shows a hierarchy of functional units concerned with processing information about the visual scene and disseminating it to relevant areas of the brain. A good general

introduction to visual pathways can be found in Zeki [254].

Although I am primarily concerned with the neurophysiology of vision in this chapter, I also introduce some results from the area of psychophysics. Workers in this area discover mechanisms of vision by non-invasive experimentation on humans. Various spatial, chromatic or moving test patterns are presented to a subject, who then has to make discriminations based on what they see. For example, contrast sensitivity measurements are made by asking the subject to indicate whether or not they see a pattern which is presented at low contrast. The contrast that gives rise to a 75% correct response is taken as the contrast threshold for that test pattern. Different stimuli result in different thresholds and such data can help to identify the mechanisms by which spatial processing of the visual scene takes place.

2.2.1 Neurons

Neurons are the electrochemical processing elements of the nervous system. They form its wires and logic gates—allowing information to be communicated in a way that is faster and more specific than chemical signalling. A useful introduction to neurons is found in Ganong [84] chapter 2, and also in Diamond *et al.* [68]. The human nervous system contains about 10^{12} neurons occurring in a variety of sizes and configurations. Neurons consist of a number of parts (Figure 2-3): the cell body; bushy fibres called dendrites extending from the cell body; the axon, which is a long fibre originating from the cell and running for some distance; the axon hillock which is a bulge in the cell body and is the source of the axon. The dendrites form the input fibres to the neuron and the axon, the output fibre. The axon may run for some distance and branch a number of times before terminating by forming *synapses* at the dendrites of another neuron.

Neurons are often drawn with only a few inputs, but in actual fact there are up to 10,000 axon fibres synapsing on a single neuron. Also, axons can be very long indeed. The giant motor neurons in the cortex of the brain send off such axons, some of which terminate in the lower body, up to 1m away. If one imagines the cell body of a neuron scaled up to the size of a tennis ball, then the bushy dendrites would fill a medium sized room and the axon would be about 1cm thick and terminate up to a mile away.

Neuron Operation

Membrane potential is an important concept in understanding how neurons work. This is the potential difference existing across the cell membrane i.e. between the cell interior and

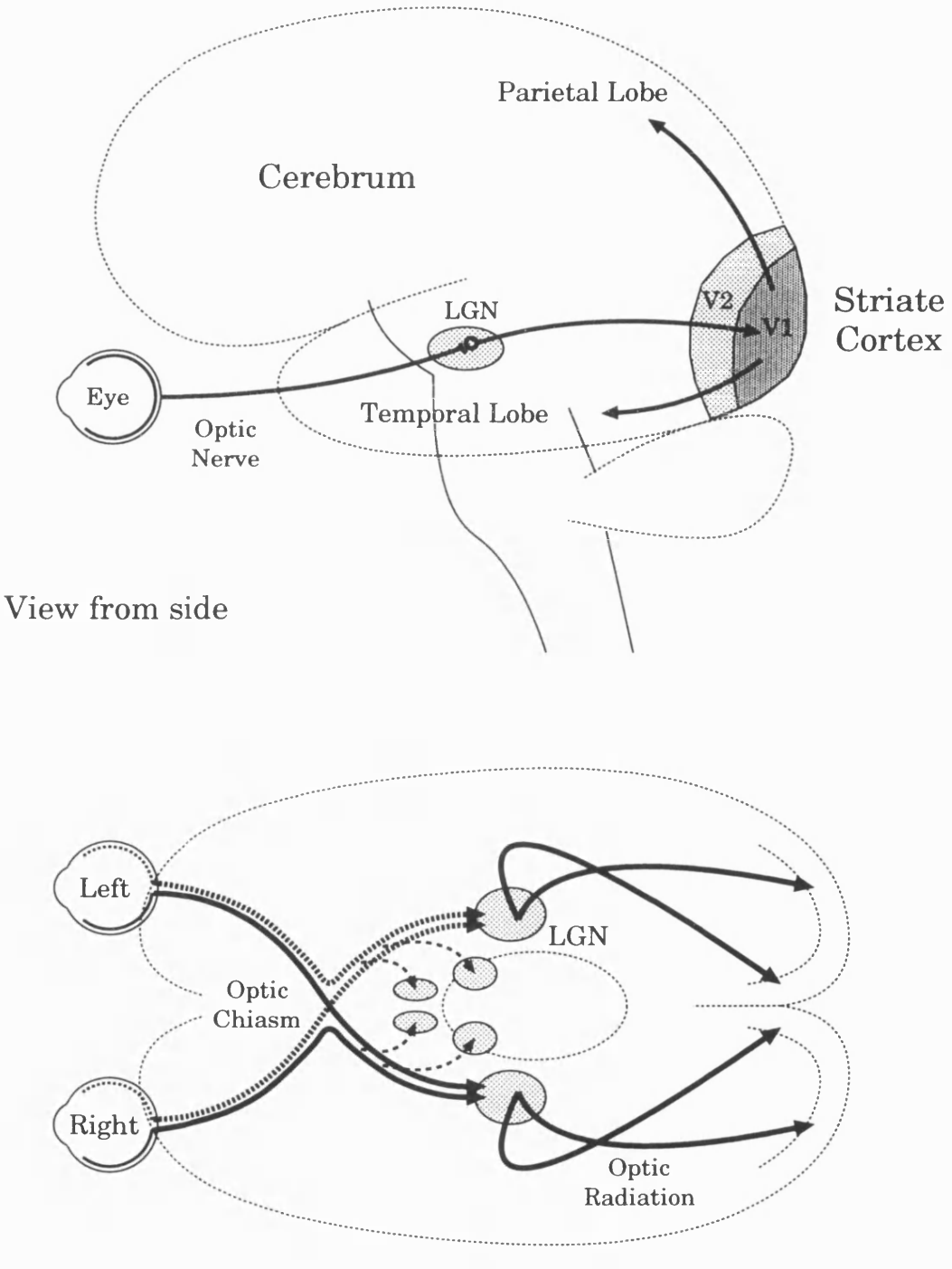


Figure 2-1: The Visual Pathway

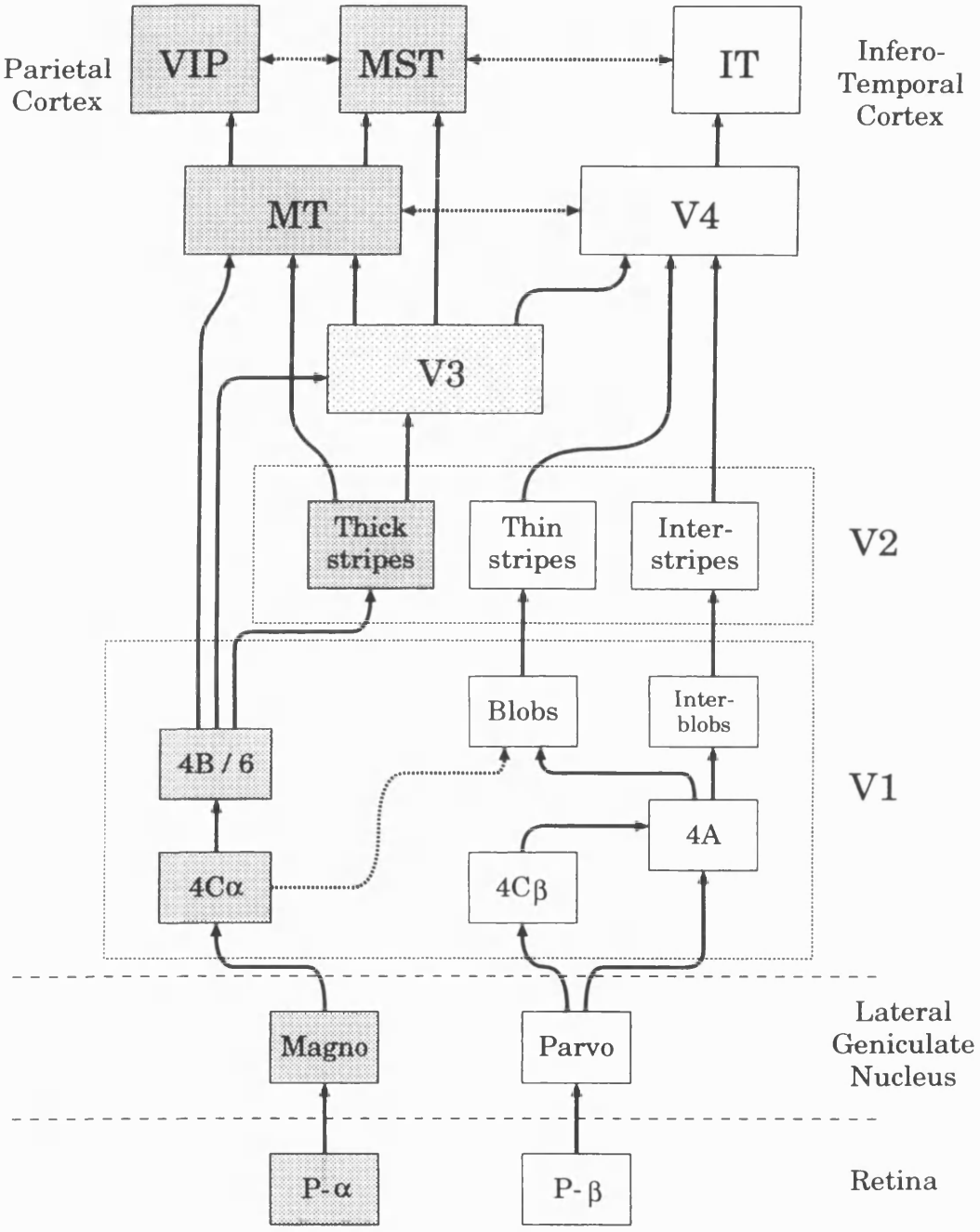


Figure 2-2: Functional Streams in the Visual Cortex. The shaded areas are concerned with motion and stereopsis. In this thesis, I am primarily interested in the non-shaded areas which are responsible for colour and structure vision. MT, middle temporal; VIP, ventral intraparietal; MST, medial superior temporal; IT, inferior temporal. (Adapted from Felleman and Van Essen [76].)

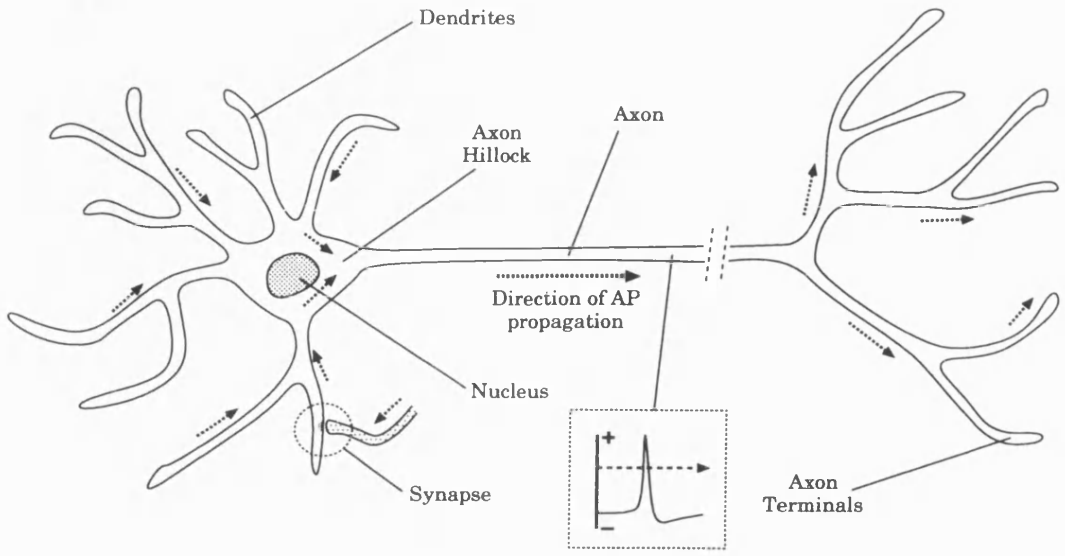


Figure 2-3: Parts of a Neuron

the exterior electrolyte. When the neuron is inactive, this potential sits at −70mV. The membrane potential can be influenced by inputs from synapses on the dendrites. The cell body acts as a summing junction for potentials present on different dendrites so that the body potential can be made more positive (depolarised) or more negative (hyperpolarised) by a combination of influences from its inputs. The axon is effectively insulated from the cell body by the axon hillock, so changes in membrane potential do not propagate down the axon.

When the combined dendritic potentials are large enough to depolarise the cell body above about −55mV, the cell fires and the membrane potential rises to +35mV within a millisecond. Following this rise, the potential decays more slowly back to just below its resting potential (after-hyperpolarisation). There then follows an insensitive period lasting about 40ms, within which, the cell's firing threshold is much increased. The voltage waveform present during the firing period crosses the axon hillock and is propagated down the axon in the form of a short pulse of depolarisation known as an *action potential.* This "all or none" behaviour is a feature of the majority of neurons.

In the retina and central nervous system, there are neurons with dendrites but no axons. These cells serve to transmit action potentials or spread post-synaptic potentials from one neuron to another without a propagated action potential. Such local action appears to be common in various parts of the brain, including the intermediate layers of the retina.

Nerve Fibres (Axons)

Nerves are not "telephone wires" that transmit impulses passively. Conduction of nerve impulses is an active process. The energy for transmission along the axon is provided locally by the axon membrane as the pulse propogates, but the release of energy takes time. It is not done with the speed of an electronic conductor (Freeman [83]). Axon conduction velocities range from $0.5ms^{-1}$ (dorsal root fibres for temperature sensing) to $120ms^{-1}$ (motor fibres mediating muscular control) and increase with fibre diameter. Nerve fibres are frequently myelinated, that is, covered with a sheath of myelin that serves as an insulator and increases conduction velocity.

Axons conduct action potentials equally well in both directions. If an axon is stimulated half way along its length, two action potentials will travel opposite ways, one in the forward direction (orthodromic conduction) terminating at synapses with another neurons, and the other in the reverse direction (antidromic conduction) terminating at the cell body. In a living animal, impulses normally pass in the forward direction only. This is ensured by the action of the synapse that only allows conduction from axon to dendrites.

Synapses

Synapses occur where axon fibres terminate on the dendrites or cell body of another neuron. Fibres arriving at a cell (afferent fibres) have enlarged ends (synaptic knobs) that make contact with the dendritic membrane forming a synapse. Action potentials occurring in the afferent fibres cause the release of chemical neurotransmitters from the surface of the synaptic knobs. These diffuse across the synapse and bind to receptor sites in the dendritic membrane, causing either hyperpolarisation or depolarisation. Hyperpolarisation leads to inhibition of the neuron firing rate and is mediated by inhibitory neurotransmitters while depolarisation facilitates firing and is mediated by excitatory neurotransmitters.

There are a vast number of different neurotransmitters in the vertebrate nervous system. Each has a subtly different effect and determines the way that the neuron responds to incoming action potentials. The type of neurotransmitter specifies the intensity and time course of polarisation at the dendrites, as well as whether the interaction is excitatory or inhibitory.

Representation

Neurons are often thought of as representing quantities by the frequency of action potential production. Each neuron sums inputs in a way that is dictated by a variety of factors

including the types of neurotransmitter present. When this sum is above a certain threshold level, the neuron fires an action potential. Action potentials are more likely to be produced if facilitatory inputs arrive simultaneously from a number of places (summation in space), or if any particular input is stimulating the cell repeatedly (summation in time). A cell that is being strongly excited fires a stream of action potentials up to a maximum rate of about 300 impulses per second.

2.2.2 The Retina

The retina is a multi-layered structure consisting of photoreceptors and neurons. An introduction can be found in Ganong [84] chapter 8, and also in Masland [158]. The photoreceptors form a layer which is in contact with the pigment epithelium at the back of the eye. In front of the receptors are four main classes of neurons which transform information from the receptor layer before transmitting it to the brain via the optic nerve. These are the horizontal, bipolar, amacrine and ganglion cells which are organised into three nuclear layers separated by the inner and outer synaptic layers (Figure 2-4). Ganglion cells form the output layer and have long axons which cross the surface of the retina, exit the back of the eye at the *blind spot* and together make up the optic nerve. Light must pass through all these layers before being absorbed in the receptors.

Photoreceptors

Photoreceptors are present in two types: rods are responsible for achromatic night-time vision and cones are responsible for colour vision when there is sufficient light. Both rods and cones contain pigment molecules which are isomerised by light. When a rod pigment molecule absorbs a photon it initiates a biochemical chain reaction which results in measurable electrical activity at the cell membrane (the details of this process can be found in Schnapf and Baylor [207]). Rods are therefore very sensitive, but become insensitive at higher light levels as gain reduction mechanisms come into play. Cones work in a similar way but are responsible for a higher intensity range.

Both rods and cones *adapt* to light levels with an exponential time constant of a few minutes. This gives them a larger dynamic range to signal intensity modulations above and below the level of prevailing illumination. Bipolar cells can synapse with many rods whereas cones usually have a one-to-one relationship with these cells. This arrangement results in a lower spatial resolution but higher light sensitivity for night-time vision.

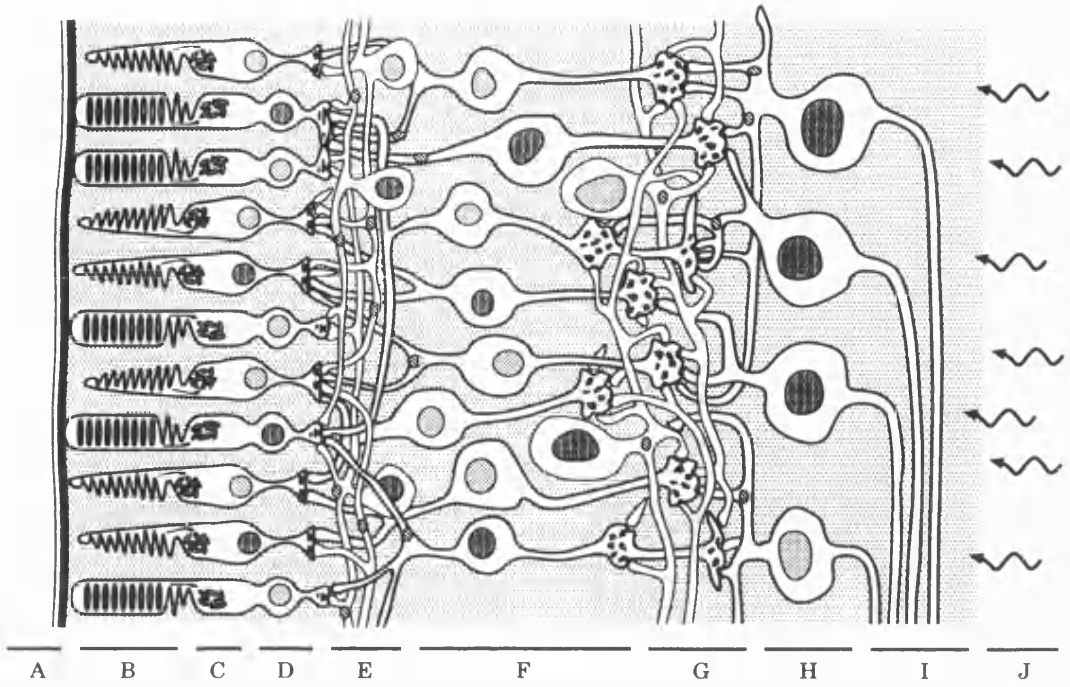


Figure 2-4: Cross-section through the retina. A, pigment epithelium; B, rod and cone outer segments; C, inner segments; D, outer nuclear layer; E, outer plexiform layer; F, inner nuclear layer (horizontal, bipolar and amacrine cells); G, inner plexiform layer; H, ganglion cell layer; I, optic nerve fibres; J, incident light rays.

The human eye is foveated. This means that different parts of the retina have different acuities and on average, the receptor density reduces with visual angle. At the fovea, in central vision, there are no rods and the cones are very densely packed. As eccentricity (distance from the fovea) increases, cone density reduces while rod density increases to a peak at 20 degrees. After this point, there are very few cones. Further increases in eccentricity find a reducing number of rods.

Receptive Fields

There are about 6 million cones and 120 million rods in each eye. Since about 1.2 million nerve fibres exit the eye in each optic nerve, this represents a significant convergence of signals. Intermediate neural layers in the retina perform summation or differencing among receptor outputs and thereby implement a stage of "image processing". The details of these operations will be explored later.

For any one neuron in the retina, or occipital cortex, there will be a small region of the visual field over which light can indirectly affect the activity of that cell: either by

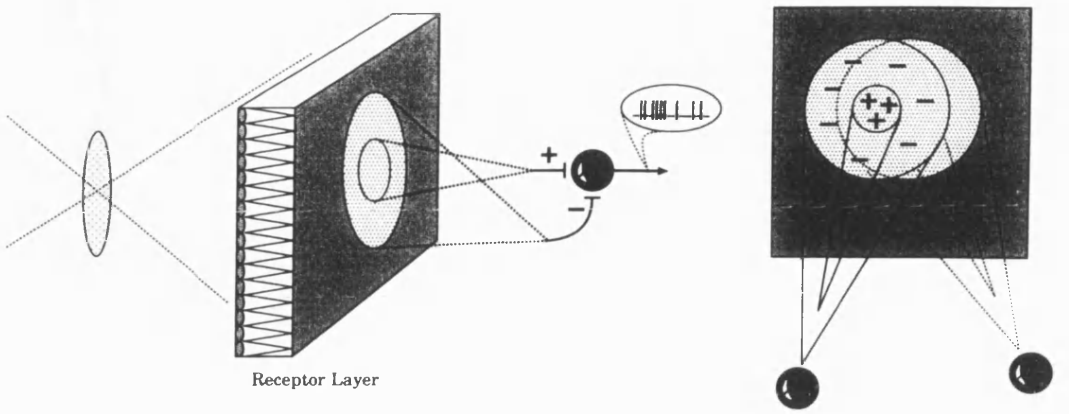


Figure 2-5: Centre-surround receptive field arrangement. Signals from the photoreceptors are summed together over two overlapping regions having different sizes. The summed signals are then subtracted and the resultant produces action potentials at the ganglion cell stage. Receptive fields of neighbouring cells overlap considerably and in the central retina there is a one-to-one mapping with receptors.

suppression or by excitation. This region is known as the neuron's receptive field—a concept that is central to visual neurophysiology (Hartline [97]). At higher levels of the vision system, neurons have larger receptive fields because they combine together responses that come from earlier neurons. Each of these has a small receptive field, but when combined, they may cover a large area. This convergence of neural signals is fundamental to the process of abstracting useful visual concepts from a scene.

Single-cell electrical recordings from retinal ganglion cells in the primate have been made from cell bodies, fibres in the optic nerve or terminations in the lateral geniculate nucleus. When the retina is stimulated with small spots of light, these cells are found to increase their level of activity when light falls on a small patch of receptors, but stimulation of the locally surrounding area results in suppression. This means that when viewing a scene, small spots of light elicit a strong response in the optic nerve, while large patches of light show a reduced response because the surrounding receptors are also stimulated.

This form of spatial antagonism is known as a *centre-surround* arrangement (as shown in Figure 2-5). A number of roles for centre-surround receptive fields have been suggested, including edge enhancement, forming contrast measures and removing the effects of variable illumination. Receptive fields of this type have been discovered in the optic nerves of a variety of animals, including the frog (Barlow [8]), horseshoe crab (Ratliff and Hartline [195]), cat and monkey (Wiesel [245]; Hubel and Wiesel [112]).

Centre-surround neurons in the eye exist in two forms. Those with excitatory centres and

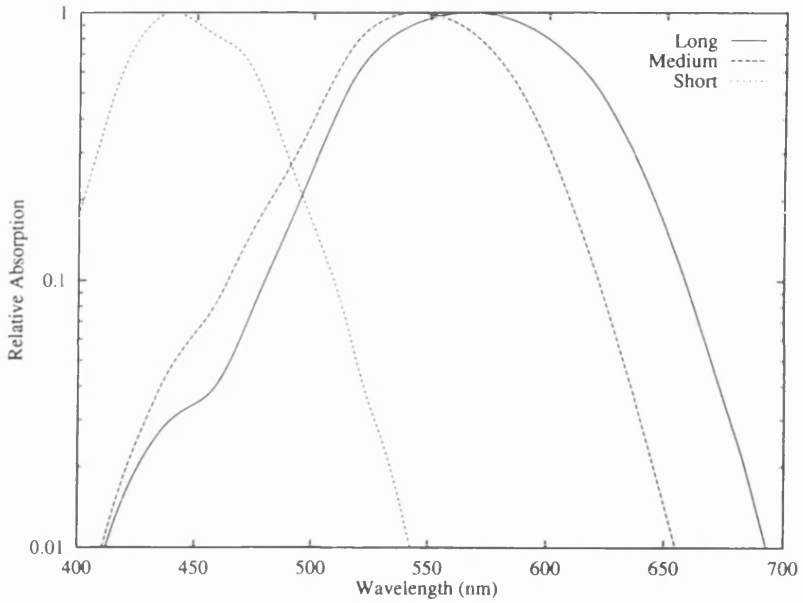


Figure 2-6: Receptor absorption characteristics. Wavelengths of maximum absorption for long, medium and short wavelength cones are 566, 543 and 440nm, respectively.

inhibitory surrounds are *on-centre* neurons. The reverse arrangement is also present, with neurons being excited by a dark spot on a light background. These are *off-centre* receptive fields because spots of light suppress their spontaneous activity and they often show rebound excitation when such a spot is turned off. In this way, the visual system uses separate channels to differentially code for light-on-dark and dark-on-light (brightness increments and decrements).

Colour Opponency

Three types of cones mediate primate colour vision. These have peak sensitivities in the long, medium and short wavelength regions of the visible spectrum, although there is considerable overlap (Figure 2-6). A significant cone interaction that occurs in the retina is the formation of colour opponent channels. Responses from cones with different spectral sensitivities are subtracted to yield a channel that differentially codes for colour. There are two frequently encountered colour differences: the red-green system and the blue-yellow system. These are evident in the primate where spectral sensitivity curves for retinal ganglion cell responses often have narrower bandwidths than individual cone curves.

This behaviour fits quite well with theories of colour perception that arose at the turn of the century. The Young-Helmholtz theory of colour vision postulated the existence of three

receptors or pigments, which accounted for trichromacy and adaptation effects (Helmholtz [237]). Hering's Opponent Process Theory hypothesised the existence of three optic nerve processes that were capable of functioning in opposition to give sensations of red *vs* green, blue *vs* yellow and black *vs* white (Hering [103]). This theory was able to explain the stability of the *unique hues*: yellow (~578nm) and green (~502nm) because they are neutral points where these sensations are in balance.[1] The three opponent processes are often referred to as the *tritanopic, deuteranopic* and *achromatic* mechanisms. More recent formulations cascade both models as stages—sometimes together with a third stage (Judd [125]; Guth *et al.* [93]).

Sustained and Transient Channels

When the retina is stimulated by spots of light, some ganglion cells are found to respond transiently to the onset or offset of illumination, but not to steady light. These *phasic* cells in the primate are known as **M**-cells or P_α ganglion cells after their morphological classification. Others respond in a more sustained manner to steady illumination and are known as *tonic*, **P** or P_β ganglion cells. Both have a centre-surround receptive field arrangement but there are a number of significant morphological and functional differences between the two types: **M**-cells respond transiently to light, have a high contrast sensitivity, are almost colour-blind, exhibit response saturation with high contrasts, have thick axons with high conduction velocity that terminate in the magnocellular layers of the lateral geniculate nucleus. **P**-cells respond in a sustained fashion, have lower contrast sensitivity, are clearly colour-opponent, respond linearly with little saturation at high contrasts, have thin axons with a low conduction velocity that terminate in the parvocellular geniculate layers (De Monasterio and Gouras [52]; Lennie [143]; Kaplan and Shapley [131]; Livingstone and Hubel [149]). **M**-type neurons have a lower acuity than **P**-type neurons at the same retinal eccentricity and they are common in the periphery while **P** ganglion cells form the majority in central vision.

The emergence of two distinctive groups of visually responsive neurons at the retina marks the beginning of two functional streams specialised for motion and colour vision. The **M** neurons, which are very sensitive to moving stimuli, are able to communicate motion information rapidly to the brain via their fast-conducting axons. **P** neurons appear to be specialised for sustained inspection of colour and fine detail, since they have good spatial and chromatic discrimination and are present in the fovea where resolution is high and image

[1] There are two other unique hues: red and blue, but their behaviour is not so easily explained by interactions at the retinal level.

velocities are kept low by tracking eye movements.

As with colour vision, the spatial and temporal properties of neurons result in effects that are measurable by psychophysical experimentation. Schade [204], and later, Campbell and Robson [43] pioneered the characterisation of spatial vision when they investigated the visibility of sine wave gratings. These researchers both obtained a contrast sensitivity curve that showed attenuation at low spatial frequencies partly as a result of the surround inhibition present in retinal receptive fields. They found that the visibility and discrimination of various patterns could be predicted from the visual system's sensitivity to the spatial frequencies present in the Fourier spectrum of a stimulus.

Wilson and Bergen [247] showed that patterns that were briefly flashed excited *channels* with different spatial properties than those excited by patterns that were displayed for longer. They proposed a four mechanism model for spatial vision in which two centre-surround mechanisms respond to low spatial frequencies in a transient manner and two sustained mechanisms respond to higher spatial frequencies. These four mechanisms were assumed to be present at every retinal locus. Temporal and spatial properties were obtained for these channels that are qualitatively similar to those of tonic and phasic units in the primate retina.

Other Animals

The previous discussion relates primarily to the retina of the primate. In lower-order animals, the retina is often used to carry out more specialised scene analysis. In the rabbit, some ganglion cells have been shown to be selective for the direction of motion of spots of light (Barlow and Levick [11]). Maturana and Frenk [161] reported that there are ganglion cells in the pigeon's retina that are sensitive to motion direction. About 5% of their sample were only sensitive to horizontal edges moving vertically. In the cat, which shares many spatial receptive field properties with the primate, Cleland and Levick [47] found that about 8% of cells could not be classed as centre-surround units. This group included local edge detectors, directionally selective cells, colour units and some that were suppressed by visual stimulation. It seems that for higher animals requiring general purpose vision, specialisation for complex stimulus properties tends to be deferred to later stages of the visual pathway.

In the previous section, we encountered tonic and phasic responses in the monkey's retina. In the cat, a similar cell classification has become apparent, although there are very few colour interactions. Enroth-Cugell and Robson [72] described sustained and transient ganglion cells in the cat's retina in terms of linear and nonlinear summation and called these **X** and **Y**

cells. Cleland and Levick [46] related these to the β (**P**) and α (**M**) morphological types of ganglion cells. This XY terminology is also used when talking about the primate. A complication arises, however, because this classification is often used to group cells according to temporal and spatial properties simultaneously; but some transient cells do not show the increased levels of mean firing rate with moving stimuli that is a mark of nonlinearity. This is especially true for primates (Kaplan and Shapley [130]; Crook *et al.* [50]). The **Y** classification, therefore, is best used to describe transient cells with significant nonlinearity.

Projections from the Retina

At the optic chiasm, fibres from each eye cross over so that those from the right hand half of each retina which receive visual stimuli from the left hemifield are sent to the right hemisphere of the brain, while stimuli which appear to the right of the centre of gaze are sent to the left hemisphere (Figure 2-1). In this way, regardless of eye, the visual field is split down the middle—the two halves being analysed by opposite halves of the brain (Diamond *et al.* [68]).

From the retina, some optic nerve fibres project to the pretectal nucleus which is concerned with the accommodation (focussing) reflex and constriction of the iris in response to bright light. Others project to the superior colliculus, which is controls eye and head movements associated with visual tracking and startling reflexes (Diamond *et al.* [68]). Fibres also project to the hypothalamus to regulate the light-dark cycle of activity (Ganong [84], page 216). In the cat, there is a greater emphasis on subcortical targets for optic tract fibres. The vast majority in the primate, however, project to the lateral geniculate nucleus from where signals are relayed to the visual cortex.

2.2.3 Lateral Geniculate Nucleus

The lateral geniculate nucleus (LGN) is part of the thalamus which, being at the core of the brain, acts as a way-station for sensory input of many types. Fibres exiting from the thalamus fan out to their destinations in the crinkled canopy of the cortical surface (Figure 2-1).

There are two geniculate nuclei and, in the primate, they each consist of six cellular layer; forming a threefold representation of the opposite binocular visual hemifield in exact anatomic registration (Livingstone and Hubel [149]).

The four dorsal "parvocellular" layers receive input from **P** ganglion cells with sustained colour opponent properties, while the more ventral "magnocellular" layers receive projections from **M** retinal units with transient responses. Signals from the two eyes remain segregated

in this subcortical area. In the cat, the situation is slightly less clear, but there are generally thought to be two dorsal layers and up to three ventral layers, receiving mixed visual input from both **X** and **Y**-type ganglion cells. Lennie [143] presents a review.

Receptive Field Properties

In general, there is very little difference between neural response properties in the retina and LGN. Wiesel and Hubel [246] introduced a classification system for primate LGN receptive fields. This system is described more fully in Chapter 5, but essentially, there are four types: type I cells have centre-surround receptive fields which show both spatial and chromatic opponency by virtue of having spatially segregated inputs from different cone types; type II cells show only chromatic opponency, drawing their input from a diffuse region of the retina rather than showing a centre-surround arrangement; type III cells are colour-blind, but do show spatial opponency; the more obscure type IV cells are similar to type III cells, but are strongly suppressed by red light. Cells of types I, II and sometimes type III are found in the parvocellular layers, while types III and IV cells form the magnocellular layers.

Chromatic properties of LGN neurons were studied by De Valois *et al.* [56], Schiller and Colby [205] and Derrington *et al.* [63]. Spatial properties were investigated by Kaplan and Shapley [130] who found that the second harmonic response, a mark of nonlinearity that is common in cat **Y**-cells, was not so prevalent in the primate geniculate, indicating a greater degree of linearity.[2]

Derrington and Lennie [64] studied spatial and temporal contrast sensitivities of neurons in the macaque monkey's LGN. They found that magnocellular units had higher sensitivities and temporal frequency preferences, but poorer spatial resolution than parvocellular units—in line with the **P** and **M** retinal cell properties already discussed.

Integrative Action

It is common to think of the LGN as just relaying visual information from the retina to the cortex. There are, however, significant modulating influences, since this area receives feedback from the cortex, and there is scope for interactions within the LGN and with other subcortical sites. Hubel and Wiesel [113] found that neurons in the cat's LGN were less responsive to diffuse light, and therefore had stronger receptive field surrounds, than those of optic nerve fibre terminations recorded at the same site. Lennie [143] included reports of response

[2] Although some **Y**-like response are still present in the magnocellular layers.

suppression by contrasting patterns when these were presented outside of the conventional receptive field. Such evidences point to the existence of lateral inhibitory interactions within the LGN.[3]

Hubel [110] reported a variation in the spiking patterns of LGN neurons in the cat with levels of arousal. Erulkar and Fillenz [75] found that, following a strong stimulus, spikes with long latency were abolished with deeper anaesthesia. These two observations suggest that corticogeniculate feedback and/or arousal signals from the brain stem could have an important modulatory role.[4]

More specifically, Murphy and Sillito [178] find that LGN neurons are far less sensitive to long contrasting bars of light than their centre-surround receptive field arrangement would suggest. Removal of corticofugal feedback increases the sensitivity to these stimuli. This leads to the conclusion that neurons in the cortex, sensitive to long visual contours, feed back a signal that acts to inhibit some LGN neurons. These nevertheless show a centre-surround arrangement when tested with spots of light because the feedback path is not activated by such stimuli.

Projections from the LGN

Fibres from the geniculate nucleus fan out to form the *optic radiation* and terminate in the visual cortex (Figure 2-1). In the cat, **X**-cells from the dorsal geniculate layers project to Brodmann's area 17 (equivalent to V1). **Y**-cells and neurons in the ventral layers, project primarily to areas 18 (V2) and 19, although some also enter area 17 (Lennie, 1980). In the primate, fibres from the LGN project almost exclusively to visual area V1.

2.2.4 Primary Visual Cortex (V1)

The visual cortex is a folded sheet of grey neural matter about 3mm thick situated at the back of the brain. Nerve fibres entering and leaving this surface form the white matter beneath. The plane of the cortex is divided into a number of areas based on functional and anatomical differences. In visual area V1, there is a retinotopic mapping of the entire visual field onto the cortical surface by way of fibres arising from the lateral geniculate. Neurons in the lateral portions of V1 in each hemisphere of the brain, receive input from the centre of vision and those near the calcarine fissure at the centre receive input from the periphery.

[3]This may relate to the lower spontaneous activity and lower incidence of **Y**-cell nonlinearity in the LGN when compared to the retina.

[4]Steriade *et al.* [222] describes burst oscillations in the thalamocortical loop during sleep.

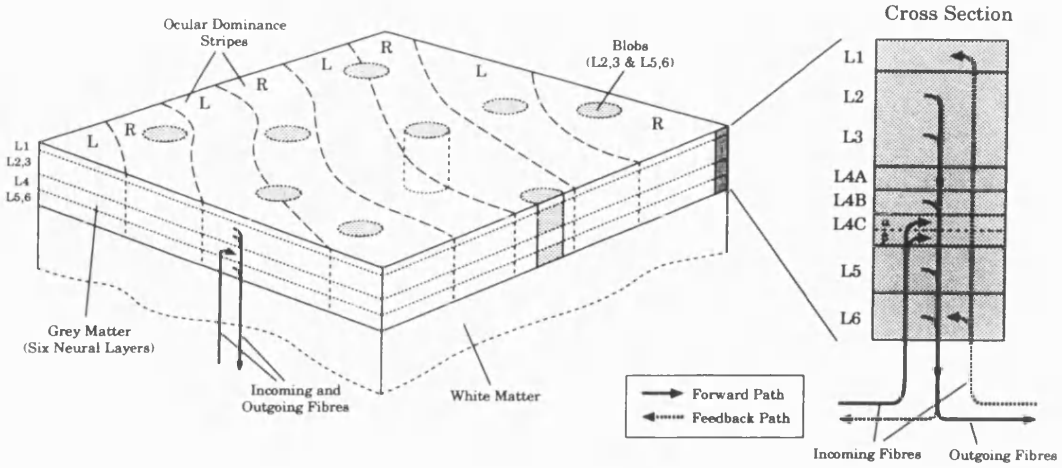


Figure 2-7: A view of the striate cortex (V1) (left) and a cross-section through the six neural layers (right). Incoming nerve fibres rise through the white matter and terminate predominantly in layer 4 of the grey matter. Output fibres descend from most of the other layers. The surface of the cortex is divided up into functionally distinct regions. L and R: regions devoted to the left and right eyes. Blobs: regions devoted to colour processing, present mostly in layers 2 and 3, but also in 5 and 6 (Livingstone and Hubel [150]).

Figure 2-7 shows the physical arrangement of V1. The grey matter is divided into six layers which contain different types of neurons. Incoming fibres project predominantly to the deeper part of layer 4, while outgoing fibres arise from other layers. Within the neural sheet, there are two sorts of connectivity: vertical connections, which tie together neurons in different layers (columns); and horizontal connections, which allow neurons to communicate in two dimensions within the plane of each layer. This structure is common to most areas of the cortex.

Receptive Field Properties

Hubel and Wiesel [114, 115, 116] pioneered the study of receptive field properties in area V1 of both cats and monkeys. They found that, unlike retinal receptive fields, the majority of V1 neurons were selectively sensitive to oriented contrasting borders or lines. For example, a neuron would fire only when an edge positioned within its receptive field was presented at a narrow range of orientations around 30° to the horizontal. Spots of light could often be used to elicit a response, but the optimum stimuli were reported as being edges or bars, preferably moving in a direction orthogonal to their orientation.

In this way, V1 carries out a piecewise analysis of the entire visual scene in terms of the orientation of its structural elements—many different orientations being represented at each

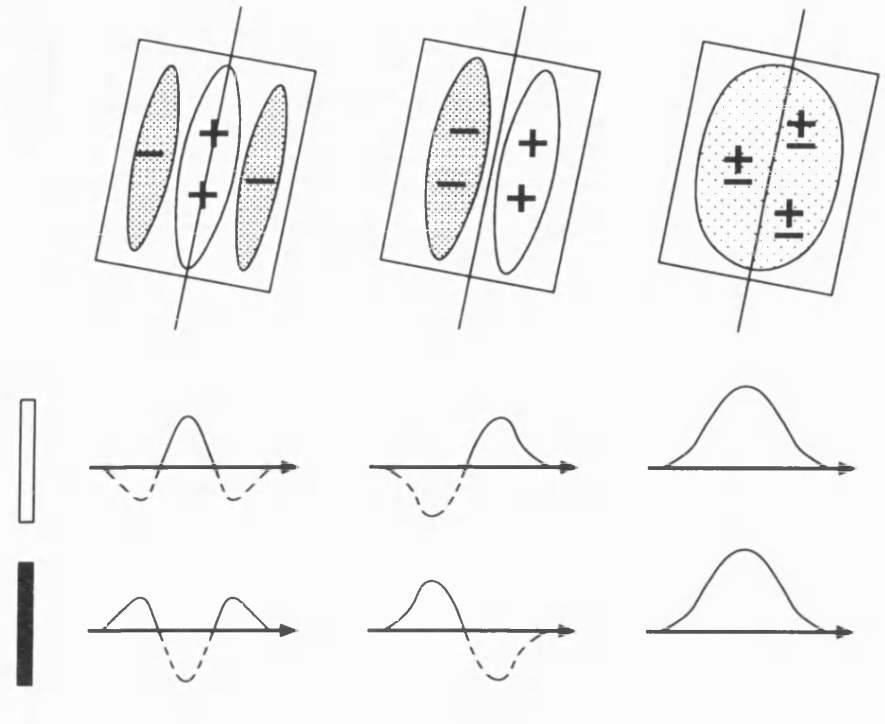


Figure 2-8: Responses of orientation-tuned cells in V1. From left to right: Tripartite simple cells, bipartite simple cells and complex cells. The two rows of response indicate stimulation with light and dark bars flashed at various positions along a line perpendicular to the orientation axis. Responses below the zero level (dotted) indicate suppression which is not transmitted due to rectification.

retinal locus by a collection of specially selective neurons.

Hubel and Wiesel distinguished three types of cortical receptive field:

Simple cells. These are sensitive to oriented bars or edges, and their responses can be predicted from the way they react to small spots of light. They show distinct summing regions which are elongated and are either excitatory or inhibitory (Figure 2-8). Bipartite and tripartite fields are found. A light spot flashed in the middle of a tripartite field with an excitatory centre will elicit a response, but a greater response will be produced by a light bar—the receptive field exhibits "length summation". An excitatory centre is often flanked by a pair of inhibitory regions which also sum signals from the retina, so that light stimuli away from the centre produces inhibition, and no response results from large stimuli which cover the entire receptive field (Bishop *et al.* [22, 21]). Stimuli that are far from the optimum orientation also result in no response and, for primate V1, the orientation bandwidth at half height averages about 40° (Parker and Hawken [187]). Simple cells often show extremely low levels of spontaneous activity, so their responses appear rectified—otherwise, these cells show

quite a linear behaviour (Movshon *et al.* [176]). Simple cells are found to be selective for contrast polarity (light or dark stimuli), orientation and stimulus width.

Complex Cells. Unlike simple cells, the responses from these neurons cannot be predicted from their behaviour when stimulated by spots of light. Distinct excitatory ("on") and inhibitory ("off") regions cannot be identified and such neurons are excited by bar or edge stimuli, regardless of their position in the receptive field along an axis perpendicular to the preferred orientation. Even though their receptive fields appear homogeneous in this way, complex cell selectivity for stimuli is the same as that of simple cells, as if their responses were the result of response pooling from a number of overlapping simple cell sub-units (Movshon *et al.* [174]).

Hypercomplex Cells. Hubel and Wiesel defined a third class of cells which show complex cell behaviour, but do not respond as strongly to long bars as they do to shorter ones. These receptive fields are still selective for orientation and bar width, but as the stimulus length is increased along the axis of preferred orientation, responses first increase, then decrease as an inhibitory region is entered. This type of receptive field property has come to be known as *end-stopping* (Sillito [214]; Rose [198]). Other researchers reported that end-stopping is also present in a sub-population of simple cells, but is not directly observable by mapping the receptive fields with small spots of light (Gilbert [85]).

Apart from stimulus type (edge/bar), length, width and orientation, cortical cells are also found to be selective to motion direction and to the choice of eye through which stimulation is provided.

Simple cells are often preferentially excited by stimuli that pass in a direction which takes them from suppression to excitation. Contrast sign is a factor in determining this optimum direction (Livingstone and Hubel [150]). Some complex cells, on the other hand, are sensitive to motion in one direction only, and this preference remains the same, regardless of edge/bar polarity. Direction selectivity is the start of a stream of processing responsible for motion perception.

Preference to stimuli from one eye over the other, or *ocular dominance*, is a further parameter. Neurons in V1 vary over which eye they prefer, from those that are strongly monocular, to neurons that can be equally excited from either eye. Hubel and Wiesel [116] demonstrated that this selectivity varied across the surface of the primary visual cortex, forming a pattern of stripes, with cells in adjacent stripes taking input from alternating eyes (Figure 2-7). Neurons at the boundaries of these stripes were found to be driven binocularly.

For vertically oriented simple cells in this category, receptive field spatial phase can sometimes be different depending on which eye the stimulus is presented to (Parker [186]; De Angelis *et al.* [62]). These selectivities are the precursors of a system sensing stereoscopic disparity.

Spatial Frequency Selectivity

Neurons in V1 were initially described as being "edge" or "bar" detectors by Hubel and Wiesel, but later researchers tested them with sine wave gratings and found that these were a more optimal stimulus. Furthermore, neurons were found to be selective for a narrow range of spatial frequencies. A controversy arose between proponents of spatial filtering and "feature detector" models. To resolve this issue, Albrecht *et al.* [3] showed that cells in V1 were more selective to the bar width of a sine wave grating stimulus than to that of isolated bars. Similarly, De Valois *et al.* [53] were able to show that simple and complex cells are tuned to the orientation of the Fourier (sine wave grating) components present in a chequer-board pattern rather than the obvious pattern elements.

Albrecht and De Valois [2] studied patterns with and without their fundamental harmonic components and found that there was good evidence for a linearly independent spatial frequency channel model of the primary visual cortex, where neurons respond to specific frequency components within a patterned stimulus. In such a model, simple cells having bipartite and tripartite fields are really examples of linear filters which differ in their spatial phase selectivity. Complex cells respond to the magnitude of the spatial frequency component they are selective for, without regard to its phase. Since these discoveries, various analytical spatial filtering models have been suggested (Marčelja [157]; Daugman [51]; Heitger *et al.* [101]) and evaluated (Webster and De Valois [244]; Hawken and Parker [98]).

The notion of the primary visual cortex as a spatial frequency analyser has been reinforced by De Valois *et al.* [57] who find that spatial frequency tuning in V1 "covers" two dimensional Fourier space with a set of patches (one for each neuron) which are each selective over a small range of orientations and spatial frequencies (e.g. 30° and 1.3 octaves at half bandwidth). Foster *et al.* [82] present a survey of the spatial frequency and temporal frequency selectivity of neurons in V1 and V2 of the macaque. Movshon *et al.* [175] present a similar set of results for the cat. Maffei and Fiorentini [153] compare spatial frequency tuning and contrast responses in the retina, LGN and cortex. Finally, Webster and De Valois [244] find that spatial frequency tuning is nearly invariant with orientation.

Chromatic Properties

Colour selectivity in the cortex has been explored by Gouras [86], Michael [166, 167, 165, 168], Livingstone and Hubel [150], Ts'o and Gilbert [227] and Lennie *et al.* [145].

Initially, all neurons in the primary visual cortex were thought to be orientation selective, but later it was found that, when stained for cytochrome oxidase, blob-like regions of the cortical surface were visible. Electrophysiological recording showed that these "blobs" contained pockets of cells that were not orientation selective, but were strongly colour selective (Livingstone and Hubel [150]). Such cells were found to have circular colour-opponent receptive fields that were differently arranged from the type I geniculate inputs—they were much larger, and often had centres and surrounds that were both colour-opponent (Michael [166]). This classical arrangement came to be known as the double opponent-colour concentric cell. Blue-yellow and red-green varieties were found, and these are thought to be selectively responsive to coloured regions, but not to white stimuli. The blob-cell route has been suggested as the start of a pathway specialised for colour perception.

Detailed investigation of the blobs by Livingstone and Hubel [150] and Ts'o and Gilbert [227] showed that the majority of the cells within these structures (extending mainly through layers 2 and 3, but also found in layers 5 and 6) could be classed as having concentric broad-band (black-white) and opponent (red-green or blue-yellow) receptive fields, but often the surround region produced mixed suppression in a more enigmatic way than the pure double opponent-colour types that were described by Michael [166].

Oriented neurons at the blob borders were also found to show strong colour-opponency and particular sensitivity to chromatic borders (Livingstone and Hubel [150]). Michael [167] also found oriented simple cells that he classed as having double opponent-colour receptive fields; possibly receiving input from nearby non-oriented blob cells. For example, a neuron might have a tripartite receptive field with the central strip excited by red line stimuli and suppressed by green line stimuli. For the two flanking regions, the situation would be reversed, with a green line providing excitation. Such a receptive field is optimally stimulated by a red line on a green background. Michael [165, 168] also found complex cells and hypercomplex cells that showed wavelength specific response properties and identified the cone inputs responsible by the use of strong chromatic adapting fields.

Lennie *et al.* [145] carried out a sophisticated analysis of chromatic opponency in V1 and found that most cells receive cone-opponent input, but for many simple and complex cells, one cone mechanism is very dominant, resulting in a primary sensitivity to achromatic boundaries.

Some simple cells and a large population of cells with concentric receptive fields receive fairly balanced cone-opponent inputs and so could legitimately be thought of as mediating colour sensations.

Interactions

The view of the primary visual cortex presented so far is further complicated by a large number of interactions that occur among cells having simple or complex receptive fields and also between these cells in V1 and those in other areas of the vision system. Such interactions are primarily of a modulatory nature, adjusting the relative responsiveness of groups of cells. The principle areas of influence are:

Inter-spatial-frequency inhibition or facilitation. Neurons sensitive to a particular spatial frequency show phase dependent or phase independent response inhibition or facilitation from a range of frequencies outside their pass-band, but only when stimulated by spatial frequencies that are usually excitatory (De Valois and Tootell [55]; Albrecht and De Valois [2]; Bauman and Bonds [14]).

Surround effects. Neurons often show inhibition when a grating stimulus is increased in size to cover an area larger than the excitatory receptive field (Foster *et al.* [82]; De Valois *et al.* [59]). Increases in width or height can result in partial end or side-stopping. The spatial frequency and orientation bandwidths of this inhibitory influence are often greater than those of the excitatory mechanism (DeAngelis *et al.* [61]). In the same way, textured surrounds are effective at suppressing the response of V1 neurons (Knierim and Van Essen [137]). This implicates pooled inhibition from a large number of spatially separated neurons via long range horizontal interconnections, and/or reentrant input from higher visual areas.

Inter-orientation inhibition. Neurons which analyse the same retinal locus, but have different orientation preferences are found to interact in an inhibitory manner, which may have the effect of sharpening their orientation bandwidth (Sillito [213]; Morrone *et al.* [172]).

Layers

This section includes a brief summary of the inputs, outputs and receptive field properties of various layers in primate V1. Information about interconnections between this area and other areas are provided by Lennie [143], Zeki and Shipp [255], Livingstone and Hubel [149]. Receptive field properties in different layers are covered by Gilbert [85], Livingstone and Hubel [150] and Hawken *et al.* [99]. This information is important in uncovering the connectivity

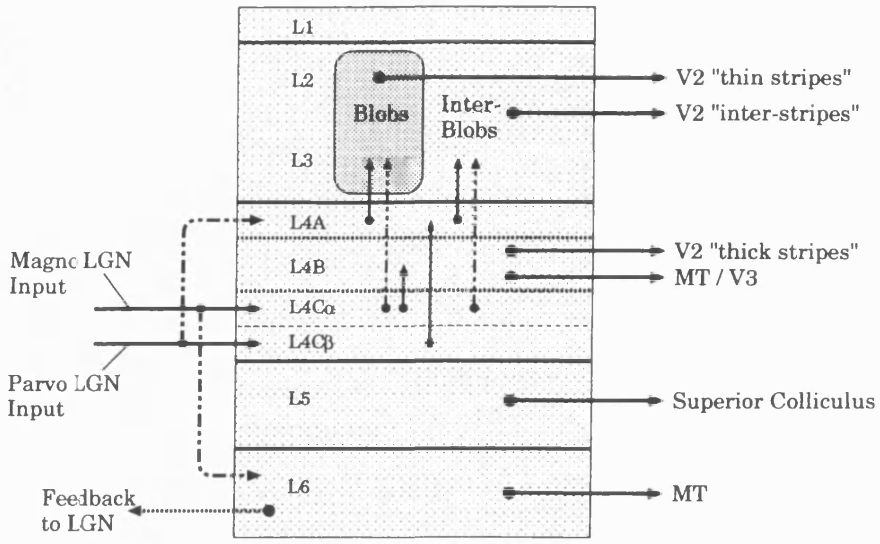


Figure 2-9: Connectivity within V1. Thick arrows indicate forward connections and thin arrows indicate inter-layer connections (Zeki and Shipp [255]).

underlying the functional behaviour of neurons throughout the visual system.

Input from the magnocellular layers of the LGN terminates primarily in layer 4Cα but there is possibly some connection to layer 6. Input from the colour-selective parvocellular LGN layers terminates in layers 4Cβ and 4A. Forward-path output from V1 to V2 arises from layers 2, 3 and 4B, and output to visual areas V3 and MT (Zeki's V5) from layers 4B and 6. Feedback from MT to V1 terminates in layers 4B and 6. Layer 6 is also the source of feedback from V1 to the lateral geniculate nucleus.

Figure 2-9 illustrates this information and indicates the major hypothesised routes of connectivity between the layers. It is useful to think of this cross section of the cortex as a unit of parallel computation, since many "cores" or "columns" with this organisation are stacked together to make up the cortical surface. Figure 2-10 indicates the receptive field properties of cells found in each of the major layers.

Emergent Pathways

The connectivity and receptive field properties encountered in V1 suggest the emergence of three pathways specialised for the analysis of motion, form and colour:

The motion pathway starts with the magnocellular neurons of the LGN, continues via layers 4Cα and 4B to visual area MT and the V2 thick stripes which are both specialised for motion and stereopsis. Neurons within this stream are generally colour-blind, directionally

Layer	RF Type/Size	Colour Selectivity	Orientation Selectivity	Direction Selectivity
2,3	Blobs: Large (0.5°), centre-surround or centre-only RFs.	Broad-band or overtly colour-opponent.	None	None
	Non-blobs: Small, complex RFs, 30% end-stopped.	Mixed	Sharp	None
4B	Medium sized, simple RFs, some end-stopping.	Broad-band	Medium	Good
4Cα	Medium sized, simple RFs, 30% end-stopped.	Broad-band	Medium	Poor
4Cβ	Small, monocular, centre-surround type I RFs.	Type I colour-opponent.	None	None
5	Large, complex RFs, many end-stopped. Respond to moving spots.	Unknown	Sharp	None
6	Large, simple or complex RFs. Often prefer long moving contours.	Unknown	Medium	Good

Figure 2-10: Receptive field properties in V1. Compiled from Livingstone and Hubel [150] and Hawken *et al.* [99]. Insufficient information was available about layer 4A.

selective and can be sensitive to stereoscopic disparity.

The form pathway starts in both magnocellular and parvocellular LGN layers[5] and passes through V1 via layers 4C, 4A and the inter-blob regions in layers 2 and 3. From there, the pathway continues to inter-stripe V2. This stream contains neurons with small receptive fields that are sensitive to achromatic and chromatic edges, but probably do not explicitly code for colour. Many are also end-stopped—a feature that may be necessary for the detection of corners and pattern discontinuities.

Colour and brightness information is probably carried by a stream from the parvocellular (colour/brightness) and magnocellular (brightness) layers of the LGN, via V1 layer 4C the layer 2 and 3 blob regions, to the thin stripes of V2. Receptive fields in this pathway are not selective to orientation but many are very wavelength specific. They are also rather larger than receptive fields which are specialised for pattern vision.

[5]There is debate about the role of the magnocellular pathway in form vision.

2.2.5 Visual Area V2

V2, the second visual area, is a region of cortex with its own retinotopic map that surrounds V1 and is the primary target for fibres which leave there (Zeki [256]). V2 represents a second level in the cortical hierarchy of visual processing and is responsible for further integration of visual information. This integration takes the form of a strategic convergence of signals from previous levels, to generate new receptive fields which have more complex selectivities and cover a greater area of visual space. Increasing the size of the region over which an image is analysed by each neuron allows for a responsiveness to more abstract and less spatially localised features. This process in repeated as we move on to higher cortical areas.

Functional Organisation

V2 has been less intensely studied than V1, but some distinct internal organisation has become apparent (Livingstone and Hubel [149]; Zeki and Shipp [255]). Staining for cytochrome oxidase (a process that reveals the colour-selective blobs in V1) shows up three types of surface feature in V2: thick stripes, thin stripes and inter-stripe regions. These are found to be functionally segregated regions and their properties will now be summarised:

Thin stripes. These receive projections from blob cells in layers 2 and 3 of V1 (Livingstone and Hubel [150]). Receptive fields in the thin stripes are non-oriented and 50% of them are colour selective. Receptive fields are larger than in V1 and double opponent-colour cells are present. A new type of response has been observed: the "spot complex" cell (Hubel and Livingstone [111]). Typically, a cell in this category might show responsiveness to a red spot of 0.5° diameter wherever it is placed within a 3.5° diameter receptive field; green spots would fail to excite the cell and response reduction would result from increasing or decreasing the spot diameter. This type of response appears to flag the emergence of selectivity for the colour of a suitably sized target independently of its spatial position.

Thin stripes in V2 are known to project to visual area V4 (Figure 2-2). This pathway is primarily concerned with colour processing without regard for detailed stimulus structure or position.

Thick stripes. Projections from layer 4B of V1 terminate in the thick stripes of V2. Here, receptive fields are often oriented but not at all wavelength selective. Not many are clearly end-stopped but they are often selective for direction of motion when stimulated by spots or bars. These neurons often require a binocular stimulus to produce a significant response and they are tuned to specific ranges of disparity (stereoscopic depth relative to the

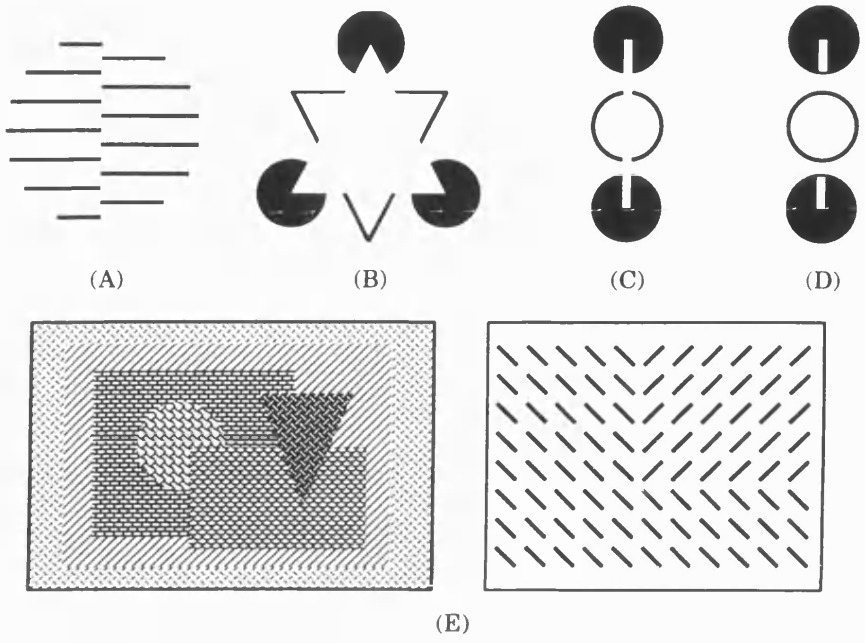


Figure 2-11: Demonstration of Subjective Contours. The vertical boundary and circle in (A), the triangle in (B), the white bar in (C) and the textured regions in (E) are all defined by boundaries that are not made explicit. In (D), adding extra lines destroys the effect.

plain of fixation). The most common sensitivities are to near, far or centre disparity.

Projections from the thick stripes are sent to V3 and MT, and form part of a pathway specialised for the analysis of motion, stereoscopic depth and dynamic form.

Inter-stripes. This region of V2 receives its input from the layer 2 and 3 inter-blob regions of V1, which have small oriented receptive fields. Those in V2 are larger and the most obvious difference from V1 is an increase in the percentage of neurons that are end-stopped, with 50% fitting this category. Such V2 cells project to visual area V4 (and perhaps to V3) and form a pathway which is concerned with detailed shape analysis.

Illusory Contours

It has long been known that the human vision system can determine three dimensional form or outlines from visual clues that are very sparse indeed (e.g. see Marr [154]). Since we live in a world where objects are textured and backgrounds are textured and lighting is variable, object occlusion boundaries are rarely completely defined by light-dark transitions. In spite of this, we have the ability to appreciate these boundaries.

Figure 2-11 shows some examples of "subjective" or "illusory" contours—defined only by sets of discontinuities, such as might result from occlusion. Our ability to perceive a virtual

boundary, triangle or bar implies that there should be neurons in the brain that can respond to stimuli like these.

Neurons with this capacity have been found in primate V2 by von der Heydt *et al.* [236], Peterhans and von der Heydt [190] and von der Heydt and Peterhans [235]. In the latter reference, 44% of V2 neurons tested were shown to respond to the orientation of the illusory boundary in Figure 2-11A as well as responding to conventional line stimuli at this orientation. V1 neurons were not found to possess this property. Von der Heydt and Peterhans explained this behaviour in terms of a convergence of signals from V1 contour and end-stopped cells—the latter being responsive to line ends.

In Peterhans and von der Heydt [190], 32% of V2 neurons are described as being able to bridge gaps in bars or edges that are defined by surrounding stimuli, e.g. Figure 2-11B and C. As well as being responsive to a continuous white bar, these neurons respond to stimuli like Figure 2-11C where a white bar appears to occlude other elements, even with the end elements outside the excitatory receptive field. Making minor changes to the figure (Figure 2-11D) obliterates the perception and also the V2 neural response.

Figure 2-11E shows that boundaries can be seen between regions having dissimilar textures and there may be a role for V2 in this. Long range horizontal interactions in V1 have been shown to contribute to texture "popout" (Knierim and Van Essen [137]) and it can only be expected that such processing will be elaborated in V2—heading towards the kind of perceptual popout and pre-attentive texture grouping effects described by Julesz and co-workers (Caelli and Julesz [41]; Julesz [126]; Bergen and Julesz [19]; Julesz and Kröse [129]).

Contour length integration and curvature analysis is another task for which V2 seems suited. End-stopped cells form a large proportion and it is known that V1 cells of this type can show selectivity for curves (Versavel *et al.* [234]). Dobbins *et al.* [69, 70] have given a mathematical treatment of end-stopped cell curvature selectivity, and Zucker *et al.* [259] have devised a model of contour formation based on this hypothesis which involves interactions between V1 and V2.

Features

Comparison between the spatial frequency preferences of V1 and V2 show that V2 covers a frequency range that is about four times lower when tested with sine wave gratings (Foster *et al.* [82]). Some of these cells even appear to be simple cells because the optimum stimulus is a half cycle of grating that fully covers the receptive field and there is summation within this

area. In this respect, they are similar to some of the fields that are found in V4 (Desimone *et al.* [66]).

Peterhans and von der Heydt [190] found that some complex cells in V2 were clearly selective for feature type. For example, a neuron was excited by a dark bar, but not an edge or a light bar. Similar selectivities for edges of one polarity over bars or edges of the other contrast polarity are known, and some neurons respond better to square wave patterns than to sine wave gratings.

Whereas neurons in V1 are solely selective for Fourier components in the stimulus, it appears that V2 may begin to synthesise a description in terms of edge and bar features, with bar detection neurons that are not only stimulated by the fundamental sine wave component in a square wave grating, but also by the sharp edges, whose high frequency components are sensed by excitatory convergent inputs from V1 (this is my interpretation). Further investigation of these properties is needed.

2.2.6 Higher Visual Areas

As we move on to areas of the cortex that are concerned with more abstract aspects of the visual experience, neurons are more likely to show highly complex properties and this presents a problem for experimenters. Simple localised two dimensional pattern elements, such as spots or bars, are inadequate stimuli for characterising receptive fields at this stage. I suggest that a classification in terms of orientation or directional selectivity becomes less of a key to the computational processes that take place within higher visual areas like V3, V4 and IT, and this should be borne in mind.

The current view about visual processing beyond V1 and V2 is that there are two pathways, the first of which is responsible for perception of motion, spatial relations, and coordination of visually-guided behaviour, and the second is involved with object recognition, memory and discrimination (Figure 2-2). The first uses areas V3, MT, MST, VIP and other areas in parietal cortex; the second, V3, V4, IT and the temporal lobe. (Ungerleider and Mishkin [230]; Desimone *et al.* [66]).

Visual Area V3

The role of V3 is unclear. Receptive fields in this area are around 2° in diameter at 3° eccentricity—similar to V4 and twice as large as those in V2 at comparable eccentricities. Zeki [254] describes V3 as being responsible for the perception of dynamic form. It receives its

primary input from layer 4B of V1 (lower visual field only) and the thick stripes of V2—areas which are both concerned with the orientation, motion and disparity of stimuli.

Significant investigation into receptive field properties within V3 has been conducted by Baizer [7] and Felleman and Van Essen [76]. The main features are: No colour opponent cells, with only a slight colour selectivity preventing V3 cells from being blind to isoluminant stimuli; a large range of orientation tuning, but cells are generally less sharply tuned than in lower visual areas; some directional selectivity (which can be independent of stimulus orientation), with preferred velocities distributed around 16°/sec; around half of the cells are much more responsive to binocular stimulation, especially when stimuli are centred around the fixation plane; some cells respond to multiple discrete orientations or directions of motion.

V3 projects to V4—an area concerned with structural recognition and discrimination—and MT which is concerned with the analysis of motion.

The Motion Pathway and Space Vision

Early psychophysical research into motion showed that there were two processes responsible for its recognition: one for fast short range displacements and another which could associate together long range events providing that the jump repetition rate was not too high (Braddick [26]; Ullman [229]). Later, research into psychophysical motion perception (Kelly [133]; Burr and Ross [37]; Burr *et al.* [38]) and the spatiotemporal selectivities of V1 neurons complemented *energy models* of motion perception in accounting for the first process (Adelson and Bergen [1]; Watson and Ahumada [240]). Nakayama [180] and Movshon [173] present reviews of biological motion processing and outline some of the problems that must be solved by the vision system.

Visual area MT is primarily concerned with the analysis of motion and stereopsis. This area receives inputs from parts of V1, V2 and V3 that are driven by the transiently responding magnocellular LGN layers. Neurons in this stream have thick myelinated axons which conduct action potentials much faster than other pathways.

Recordings have been made from MT neurons by Maunsell and Van Essen [162, 163] and Albright [4]. All cells were found to be directionally selective when stimulated with moving random dot patterns; many were also selective for stimulus orientation, but optimum orientations and directional preferences were not necessarily related. There was also a wide range of speed selectivity—the most common being 32°/sec. Most units in MT received balanced input from both eyes and were disparity selective: either near, far, tuned excitatory

or tuned inhibitory types (Maunsell and Van Essen [163]).

The main result is that the ensemble of MT neurons sense the motion and disparity of stimuli at a multitude of localised regions over the visual scene without a strong regard for the type of stimulus. Separation of stimulus form and motion can be experienced in visual after-effects when motion becomes attributed to stationary objects. For example, looking at a waterfall for some time, then looking at the static river bank can result in the illusory impression of motion flow in the opposite direction. Similar effects are experienced when a train stops in a station: one can have the visual sensation of sliding backwards.

Two perceptual capabilities that may be attributed to MT and higher parietal areas are *structure from motion* and *structure from stereopsis.* Structure from motion (Ullman [228]) is the capacity to see structural form in a pattern that is in motion, even though instantaneous views are unrecognisable. Structure from stereopsis allows one to build a shape description from disparity information alone, and is responsible for the familiar random-dot stereogram effects.

MT projects to areas MST, VIP and hence to the parietal cortex. MST neurons have been found to respond to full-field motions, particularly those which are relevant to visual locomotion,[6] or to complex optical flow patterns other than translation, such as rotation or expansion (Saito *et al.* [202]). The posterior parietal association cortex (area 7a) has been shown to contain neurons involved with visual reaching or hand manipulation, as well as those that are able to remove the effects of eye-movements and distinguish self-induced motion from external motion (Sakata *et al.* [203]). Such capabilities are indicative of a system that is not just specialised for motion, but is also able to relate the positions of objects in three dimensions without involving recognition.

Form, Colour and Recognition

The pathway involving V4 and the inferior temporal cortex (IT) has begun to be regarded as the location of visual recognition and learning (Desimone *et al.* [66]; Tanaka [223]).

V4 receives direct input from V2 and V3 and was originally thought to be an area solely concerned with colour processing (Zeki [253]) because of the large numbers of colour selective cells present, and due to the fact that damage to a similar brain area in humans results in achromatopsia (lack of colour sensations). Later work has proved that this area is responsible for many aspects of form vision, rather than just colour discrimination (Heywood *et al.* [104]).

[6] Perrone [189] has produced an interesting biologically inspired model relating to these abilities.

In particular, monkeys that were trained to differentiate between stimuli were unable to generalise the specific task to other stimulus configurations or spatial locations following V4 lesions (Schiller and Kyoungmin [206]).

Receptive fields in V4 are large[7] and frequently cover the centre of vision, extending across the mid-line. Neurons are broadly selective for a wide variety of stimuli: colour, light and dark, edges, bars, oriented or non-oriented, moving or stationary, square wave and sine wave gratings of various spatial frequencies, and particularly the length and width of bar stimuli (Desimone and Schein, [65]). One consistent feature is that they have centre-surround receptive fields, with centres that are excited by some stimulus form, and very large silent suppressive surrounds that are similarly sensitive. Maximum response is produced when the two regions are presented with *different* patterns or colours.

Sensitivity to stimuli relative to those surrounding is known to result in a number of useful perceptual effects (such as size constancy) and illusions (Julesz [128]). Zeki [252] found that V4 colour selective neurons showed significant colour constancy—perceived object colour remains stable over wide variations in illumination chromaticity—a capacity which may be facilitated by such a centre-surround arrangement.

A further property shown by V4 neurons, which is increasingly important at higher levels of the vision system, is modulation by focal attention. Experiments with awake, behaving monkeys has shown that the responsiveness of V4 neurons is modulated by the extent of attention to a stimulus within the receptive field, but only when there is also a distractor stimulus present (Desimone *et al.* [67]). This form of attention is different from a physical shifting of the eyes and involves neural mechanisms. Attention processes in humans have been studied by Sagi and Julesz [200, 201], Treisman [226], Kröse and Julesz [138] and Julesz [127], and they find that the vision system is essentially a bottom-up process, with a top-down mechanism that allows for selection and discrimination of stimulus specifics when there are a lot of distractors present. Such attention mechanisms in vision are believed to involve the pulvinar nucleus of the thalamus, and a number of models have been advanced (LaBerge [141]; Crick and Koch [49]; Niebur *et al.* [181]).

V4 is known to project to area IT in the temporal cortex, where individual neurons have receptive fields that cover most of the visual field. Neurons have been found that show strong selectivities for *faces* and occasionally *hands* and these selectivities are even present in infant monkeys (Gross [90]; Rolls [197]). Neurons that respond to other recognisable forms have

[7] About half the length of one's thumb when the arm is outstretched.

not been discovered, but the crucial importance of faces and the difficulty in picking up the subtle differences between them is likely to result in a large number of neurons which are similarly selective. IT receptive fields are also found to respond to shapes with a large degree of size, position and contrast invariance.

The current view in visual recognition, memory and visual imagination, is that neurons in areas like IT use a population code to represent the variety of different forms that we experience in the world. Individual neurons do not represent images in memory completely, but rather this representation is spread out over a number of broadly selective cells (Tanaka [223]). With regard to short-term memory, Miller *et al.* [169] found that there was a sub-population of IT cells that behaved differently, depending on whether or not the monkey was involved in matching to a remembered stimulus. They proposed the idea of "adaptive memory filters"—a concept which involves a pair of populations in IT, one of which can hold a short-term memory trace, and the other which is able to respond continuously to sensory input. Comparisons between these two populations forms the basis of match decisions. Further developments in this area are certain to be interesting.

2.3 Previous Simulation Work

I now describe previous work in modelling and particularly in simulating neural behaviour in the primate vision system.

2.3.1 Grossberg and Mingolla

The neural dynamics of perceptual grouping are described by Grossberg and Mingolla [91], concentrating on a number of texture and boundary effects, such as illusory contours, brightness filling-in, texture segmentation, grouping in textures, neon spreading, the café wall illusion and various Gestalt rules. A computer simulation of illusory contour generation is described, to illustrate some of their points about grouping. This simulation uses an oriented grouping field to combine together co-linear responses from units which are selective for the ends of lines.

There is no explicit attempt to model known receptive field properties, however, the scheme does start with with a low-level representation in terms of oriented boundary segments and was probably inspired by mechanisms in V1 as well as by perceptual grouping effects.

2.3.2 Finkel and Edelman

The simulation of Finkel and Edelman [79] bears some initial similarity to that of Grossberg and Mingolla above, but is much more involved and is closely related to visual neurophysiology. Finkel and Edelman investigate the role of reentry within functionally segregated visual areas using a simulation centering around 222,000 cells drawing input from a 64×64 pixel image. The three areas that they simulate are correlates of V1, V2/3 and MT, specialised for orientation selectivity, occlusion determination, and motion extraction.

The reentry paths covered are: those connecting MT to V1, responsible for suppressing local directional responses in layer 4B of V1 that are at odds with the more global MT picture; those communicating information about occlusion back from V2/3 to V1 in order to resolve conflicts and stop illusory contours from being generated across boundaries; those communicating information form MT to V2/3 which allow the system to respond to boundaries produced by differential motion (two dimensional structure from motion).

This simulation demonstrates three important functions for reentrant connections: (1) Resolution of conflicts between representations of the same stimulus held in different cortical areas. (2) Cross-modal construction allows the outputs of one process to be fed back to a common area so that a differently specialised area can use the results for its own computations. (3) Recursive synthesis—constructs derived in a higher area can be recycled to a lower area for the recursive generation of new constructs. Functions (1) and (2) are incorporated in the demonstration of Figure 2-12, for which Finkel and Edelman's simulation can generate the outline of an illusory square (grey) by sensing the motion discontinuity and, via reentry, using this as an input to an illusory contour generation process in an area unconcerned with motion.

Finkel and Edelman make no mention of simple cells, complex cells or spatial frequency channels and contrast does not play a strong role in their simulation, which appears to act only on binary images, but their results are impressive. The work in this thesis differs substantially in direction, in that I concentrate on a smaller amount of the vision system (principally the retina, LGN and V1) and use a more faithful quantitative model of neurons in these areas.

2.3.3 Neural Contour Processing

Heitger *et al.* [101] present a simulation of neural contour mechanisms that is very similar in spirit to the work described in this thesis. The difference is that they have a much stronger

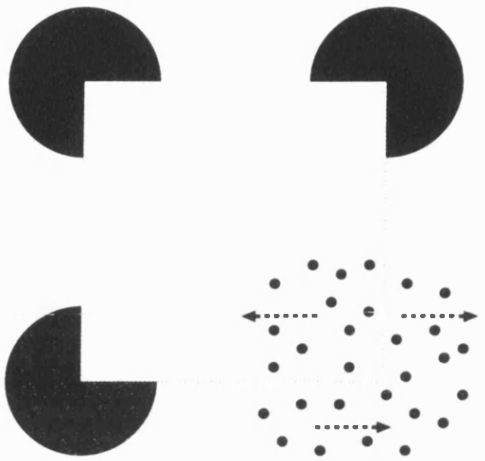


Figure 2-12: Illusory square defined by three inducing corners and two patches of dots which move in opposite directions. The motion boundary produced has the same effect as the other corners, implying a cross-modality input to contour generation.

emphasis on computer vision, do not model the retina or LGN, and are not interested in colour vision.

The first stage of their simulation is a Gabor filter simple cell model. They have modified the Gabor formulation to produce "stretched" versions of the kernel which, like simple cells, do not show DC responses at any bandwidth. A two dimensional operator is constructed in the Fourier domain. Complex cell responses are made by a local energy calculation based on the output of simple cell filters in quadrature.

End-stopped cells are simulated by a process that effectively takes first and second derivatives along the direction of orientation preference, but this is combined with a complicated suppression stage in order to avoid false responses resulting from stimulation by off-orientation boundaries. This end-stopped stage is crucial to their work because it communicates information about edge terminations and their orientation—information which is important for occlusion detection.

In a later paper, Heitger and von der Heydt [102] describe how responses from end-stopped cells can be grouped by neurally plausible processes to generate receptive fields that can respond to straight and curved illusory contours. This involves a very complex set of grouping rules which are embodied in the neural connectivity. Their method is successful in finding illusory contours in natural images, as well as in Kanisza-style test patterns. The method is also successful at determining which is the foreground and which is the background from occlusion cues alone.

I have evaluated the work of Heitger and van der Heydt by implementing all their methods (in parallel) and various references are made to this later in the text.[8]

2.3.4 Barrow and Bray

Barrow and Bray [12, 13] are interested in demonstrating how certain patterns of connections and cell specialisation in V1 could arise by activity induced self-organisation out of a sensitivity to the principal components in image data, or the responses of previous neural levels. They use a neural network model to demonstrate self-organisation into colour selective blob and orientation selective inter-blob regions. The most complex model contains 57,000 cells and uses a 40×40 pixel input image. It can account for the development of complex cells in response to moving stimuli, using an unsupervised time-derivative adaptation rule.

The main difference between the work of Barrow and Bray and that presented here, is that they are interested in explaining how various receptive field profiles could arise, whereas I am more concerned with what the static image processing consequences are.

2.3.5 Beaudot and Hérault

Beaudot and Hérault [16] describe a model of the vertebrate retina which is intended to be a realistic simulation tool for early visual processing. Their model (Beaudot *et al.* [17]) includes foveation and uses a resistive grid to account for lateral interactions in the outer plexiform layer. They concentrate on the spatiotemporal filtering aspects of retina-level vision (Beaudot and Hérault [15]) using a network that is loosely based on retinal architecture, but they do not attempt to explicitly model electrophysiological data. This is because the scheme is eventually intended for computer vision applications and practical realisation in silicon.

2.3.6 Brightness Coding Models

I include in this section two models of brightness coding due to Watt and Morgan [243] and Kingdom and Moulden [136], even though they are not in fact presented as two dimensional simulations. Both models attempt to explain why our perception of point-by-point brightness in various patterns, e.g. gratings, ramps, triangular waves, is not quite like the stimulus waveform and also departs from the result that is obtained by applying a single linear contrast sensitivity filter.

[8] No simulation results for illusory contours are presented in this thesis.

Both models use a bank of narrow-band channels to model spatial filtering in the retina or cortex, on and off-responses, and thresholding. Watt and Morgan use a set of interpretation rules to synthesise a tokenised representation in terms of edges and bars, but do not consider the filters to carry independent information. Kingdom and Moulden apply interpretation rules to the output of each spatial filter channel prior to lumping these descriptions together to produce the final percept.

There is some success in both models because they are able to predict Mach banding and various aspects of contrast threshold data. They are, however, very loosely based on neurophysiology and leave many two dimensional problems unsolved—in particular, Kingdom and Moulden do not explain how brightness "filling-in" might be extended to two dimensions.

These two models attempt to answer an important question: how do space-local concepts like "edge" or "bar" get produced from a one dimensional brightness profile, and how does our perception of brightness in a spatial pattern arise from banks of narrow-band spatial frequency tuned channels, such as those found in V1.

2.3.7 Du Buf

Du Buf [33, 32, 34] has extensively investigated the Gabor model of simple cell receptive fields, originally proposed by Marčelja [157]. In particular, he has concentrated on the Gabor response to real features, with and without blur for different positions within the receptive field, and when two features come close enough together to interfere with each other. He finds that for interfering stimuli, the phase response from pairs of quadrature Gabor filters is more accurate than the peak of the magnitude as a positional estimate.

This work is related to the brightness coding work described earlier because du Buf investigates how the outputs of narrow-band filters could be used to produce knowledge of composite feature types (edges/bars) and their positions.

Du Buf [34] makes a very worthwhile point that bears repeating:

> "The goal of the vision system is not to reconstruct the input image in some neural 'image plane' but to correlate incoming syntactical information with semantical information in order to confirm or update the internal world model. Yet we 'see' an almost perfect reconstruction."

In this paper he is able to demonstrate a simple Gabor image processing scheme that can be used to detect edge and bar features with much greater accuracy than peak-detecting Gabor magnitude information. This has relevance to the work presented in this thesis because I use

a modified Gabor model of simple cell behaviour. Du Buf [29] considers some methods of texture segmentation based on Gabor filtering.

2.3.8 Contrast Gain Control

When subjects are allowed to view high contrast sine wave grating stimuli for a long time, the slope of the contrast increment threshold curve (plotted on log-log axes) is found to reduce. Wilson and Humanski [248] propose a model of contrast gain control operating among cortical cells that can account for this data. Their model includes a set of synapses with weights that become modified with adaptation.

This model, along with the model due to Heeger [100] plays a large part in simulations presented later in this thesis. A detailed description is deferred until Chapter 7. Neither model is a simulation in the image processing sense, rather, they attempt to model some psychophysical or neurophysiological data. Here I use some aspects of the structure of both models because they are successful and plausible.

2.4 Conclusions

In this chapter I reviewed a variety of literature relating to biological vision and also prior simulation attempts. It can be seen that work on simulating the neural processing that underlies visual perception is still in its infancy. There is clearly room for a large-scale image processing simulation that faithfully attempts to draw together evidences from neurophysiology, in order to study interactions in the retina (receptive fields, contrast gain control), LGN (chrominance and luminance transfer), and primary visual cortex (spatial filtering, sampling theory, contrast gain control, end and side-stopping, surround effects). Such a directed and realistic simulation should help us to see how the ensemble of neurons in each visual area reacts to the stimuli that we experience from day to day.

Chapter 3

Simulation Strategy

3.1 Introduction

This chapter is split into two halves which detail the context and development of the simulation work described later. The first half describes the level at which simulation takes place, what is being modelled and assumptions that are made about the nature of neural calculations. The second half describes the hardware and software platform on which the simulation is built.

3.2 Context

In this section I consider: (1) a novel way of simulating a realistic viewing configuration, (2) assumptions about what neurons can do, and how they represent information, (3) ways of treating the time dimension, in a model which attempts to exclude it, (4) convolution as a substrate for simulating neuron behaviour, (5) subsampling and reconstruction.

3.2.1 Situational Model

How should real images be viewed by a simulated retina? Starting with a photograph of some scene, one uses a flat-bed colour scanner to convert this image into an array of pixel values stored in the computer's memory. Previous image processing models have operated on these pixel values directly, implementing spatial processing stages which are found in the retina or cortex (e.g. Heitger *et al.* [101]; Marr [154]) without attempting to include a receptor stage. This situation is not very satisfactory because it is at the receptor stage that some important

non-linearities are incorporated.

In the case of a colour image, the scanned representation contains red, green and blue components which are used to drive the electron beams within the display screen, thereby exciting phosphors which emit coloured light to form the image. The light given off enters the eye and excites cone receptors which then influence the rest of the neurons in the retina. This sequence defines a model for conversion between pixel values in the computer and responses in the retina.

This simulation therefore relies on what I call a "situational" model, in which a virtual subject (man or monkey) is sitting at a fixed distance in front of a colour screen that is displaying an image of a certain size. Starting with pixel values inside the computer, we can use this model to calculate responses at any point in the subject's visual pathway.

Image Parameters

To implement the transformation we need to specify a number of parameters that have a bearing on the size of the simulation. The image size, resolution and viewing distance must be chosen carefully because these define how much of the visual field is being incorporated.

A "standard" image size of 512×512 pixels was chosen. This colour image was intended to be displayed at such a size and viewed from such a distance so as to subtend 10° of visual angle at the eye. There are a number of reasons for these choices:

1. Preliminary calculations suggested 512×512 pixels as the size of image that could be processed in reasonable time and stored in reasonable amounts of memory, given assumptions about the number of stages and simultaneous representations which were to be used throughout the project.

2. This image size can be subsampled in powers of two. Subsampling was thought to be a fundamental part of the economic coding of visual data.

3. When representing a 10×10 degree region, this image size allows spatial frequencies in the range 0.1–25.6 cycles per degree (cyc/°) to be represented. This covers the full range of spatial frequencies that are visible to humans for natural stationary viewing.

4. A 512 pixel square image, when displayed on the Silicon Graphics Indigo machines, used for the development work, yields an image which subtends approximately 10° for normal viewing distances. Consequently, image features that are visible under these conditions should also be visible to the simulation, and conversely, features that are not

visible by eye should not be picked up by the simulation. This provides a subjective check of simulation fidelity.

5. Parafoveal cone spacing is reported to be about 0.6 minutes of arc whereas, under the conditions described above, pixels are spaced by 1.2 minutes of arc. However, it should be remembered that there are red, green and blue values for each pixel since a pixel can be any colour, therefore we can allow each one to represent a cluster of three cones giving a similar resolution (without saying anything explicit about relative densities of different cone types).

Retinal Mosaic

Image pixels fall on a rectangular grid, whereas the receptor mosaic is to a first approximation hexagonally packed. No specific measures were taken to simulate this arrangement since this would have been at great technical expense with no significant gain—the same sampling theory considerations apply to both situations.

In the same way, foveation has been left out of the model. It was thought that foveation would cloud the interpretation of effects due to neural interactions. It would have required differently sized convolution kernels at each spatial location—ruling out the use of Fourier transforms and making a parallel implementation difficult.

Although the neglection of foveation is a significant loss of realism, I consider that studying interaction effects at a single eccentricity is useful and, at least with regard to the relatively localised processing that takes place prior to V4, leads to generalisations that are applicable to a foveated system.

Another simplification regarding the retinal mosaic is that rod vision was excluded from the simulation. For the photopic light intensities involved in a computer display situation, cones are mediating vision. Since the aim was to explore some of the cortical image processing mechanisms, including a rod stage would have been a distraction.

An RGB pixel image implies that all three cone types have the same density, whereas this is not the case. A common estimate is that long, medium and short wave sensitive cones are parafoveally present in the ratio of 10:5:1 (De Valois and De Valois [58]). In the fovea, short wavelength cones are practically absent. For our purposes, however, relative cone densities are converted into weightings that are applied at each pixel location. This is further explained in Chapter 5, which deals with colour vision.

In conclusion, foveation presents problems, but I ignore these by concentrating on a

homogeneous region of the parafoveal retina at about 5° eccentricity, and by using receptive field size parameters that are consistent with this scenario.

3.2.2 Neurons

In this thesis, I use a linear connectionist approach to neural modelling which, although simple, is very flexible and allows for the simulation of many observed interaction types. This section outlines the level at which simulation of the vision system is carried out.

Background

In general, one neuron receives convergent input from a large number of other neurons. The output of this neuron diverges to synapse with a number of neurons downstream. At each stage, excitatory or inhibitory interactions are possible.

The computational behaviour of neurons can be analysed at a number of different levels. At the simplest level, when firing rates are taken to be the primary representational quantity, each neuron sums together individually weighted inputs and rectifies the result. This scheme captures much of the basic functionality but ignores the time dimension. A more complex model would treat each individual neuron or neural ensemble as a dynamic system, including temporal integration effects and adaptation, together with the possibility of non-linear input and output functions (Freeman [83]).

Linear processing can be carried out by neural elements of the first form, even though there is effective rectification of the signal. This is achieved in two ways. The neuron may be configured so as to have a maintained firing rate, in which case, excitatory inputs will increase the response and inhibitory inputs will decrease it. Alternatively, pairs of neurons may act in a differential manner to code for both positive and negative excursions of the signal. An example of the latter case is the simultaneous presence in the visual cortex of neurons that are excited by a dark bar and show no on-response to a light bar, and those that have the complementary response.

There has also been significant work on oscillatory EEG signals in the cortex arising from excitatory-inhibitory connections (Freeman [83]). The time-independent approach used here is not able to simulate this kind of behaviour because such signals arise out of the temporal dynamics of neural ensembles. High frequency oscillations (which are not visible in single neuron recordings) are known to synchronise activity across the cortical surface. Here, I concentrate on spatial processing and I therefore consider it valid to ignore the presence of

such overall modulatory signals. In the case of focal attention, oscillations may have a more direct impact on spatial processing, but the effects are still to be determined (Niebur *et al.* [181])

General Model

In this thesis, I do not employ any form of neural network training or allow for synaptic plasticity. Instead, neurons are simulated using image processing techniques that assume a fixed network of connectivity and inter-unit weights. A neuron's rate of action potential generation (measured in spikes per second) is, for many stages, held as a single precision floating point value. No attempt has been made to include the effects of noise, adaptation, or explicit temporal behaviour, e.g. the possibility that the precise timing of action potentials provides a temporal code that complements average firing rates. I would hope to include some of these effects in a future model.

The level of modelling used here concentrates on the mathematical behaviour of either single neurons or else small numbers of neurons which together perform a well defined function. I am concerned with functional properties that are evident from electrophysiological recordings, not with the more detailed mechanisms which give rise to these behaviours, e.g. types of neurotransmitters.

For each independent stage of response transformation, a model was chosen to encapsulate published response properties. I use the following formula to describe the behaviour of each neuron:

$$Y = Y_{max} f_{out}\left(\sum_i \frac{f_{in}(w_i X_i)}{1 + kD}\right), \tag{3.1}$$

where X_i is the set of inputs, D is an optional divisive influence and f_{in} and f_{out} are pre- and post-summation transfer functions. This function allows for common additive, subtractive and divisive interactions. Suitable choices for f_{in} and f_{out} can provide rectification, a compressive output function and also non-linear summation.

In particular, when $f_{in}(x) = x^n$, $f_{out}(x) = x^{1/n}$, $k = 0$ and $w_i X_i \geq 0$, a continuous range of non-linear summation types are possible (Quick [194]; du Buf [31]). When $n = 1$, this summation is linear, but when $n > 1$, the response produced by a number of inputs is less than their linear sum. For $n = \infty$, the summation is equivalent to a max function calculated across the inputs. For values of $n < 1$, summation of multiple inputs yields a response greater than their linear sum, and when n is small this results in a type of logical "and" function. This form of non-linear summation is therefore useful in describing a wide range of effects.

The functions f_{in} and f_{out} are chosen to best represent the spirit of neural behaviour at each stage of visual processing independently. For most neurons, $f_{in}(x) = x$, and $f_{out}(x)$ is a compressive function which is linear for small positive x. For neurons with a high spontaneous activity, such as those in the retina, f_{out} is chosen to be an odd compressive function, e.g. $f_{out}(x) = \tanh(x)$, because negative responses are usefully transmitted. For those in the cortex with low spontaneous activity, rectification is introduced, e.g. $f_{out}(x) = (x + |x|)/2$.

In conclusion, neural behaviour is modelled with a fair degree of flexibility so that the most appropriate functional level is targeted, and an optimal model of response properties can be adopted at each stage.

3.2.3 The Time Dimension

In producing a realistic spatial model of low-level vision one is constantly affected by temporal considerations. This is because vision is a spatiotemporal process and these dimensions are not in fact separable. Even though this is true, I attempt to concentrate on spatial processing and, to a large extent, leave time out of the picture. Burt [39] discusses the relationship between temporal and spatial vision.

Kelly [133] measured the contrast sensitivity of human vision for a number of temporal and spatial frequencies using stabilised vision. From this, a threshold surface was derived. Kelly points out that for normal unstabilised viewing (where eye movements are allowed), the contrast sensitivity curve is almost the same as that for a velocity of 0.15°/sec using stabilisation. This means that the most realistic approach to the simulation of normal viewing is for each spatial frequency component in the image to be given a temporal frequency consistent with a motion of 0.15°/sec.

If cells in the visual system had responses that were spatiotemporally separable, so that they were defined by $p(\omega)q(f)$, where $p(\omega)$ is the temporal response function (ω measured in Hertz) and $q(f)$ is the spatial response function (f measured in cycles per degree), then since $\omega = 0.15f$, the function $p(0.15f)q(f)$ would predict pure spatial sensitivity under natural viewing conditions. However, responses at the retinal and geniculate level are not completely separable in this way (Derrington and Lennie [64]). Also Foster *et al.* [82] find that cortical neurons show shifts in peak temporal frequency selectivity at non-optimal spatial frequencies, though such shifts are small.

Since details of the complete spatiotemporal response function for each type of neuron are not available, I make the assumption of separability and modify spatial frequency response

curves transforming them from constant temporal frequency curves (which are normally published) to constant velocity curves (which are not), for 0.15°/sec motion. For the results presented here, this provision makes very little difference to the responses of parvocellular and cortical cells, but magnocellular responses are found to be completely spatial frequency band-pass under constant velocity stimulation, even though they have significant responses to diffuse (low spatial frequency) stimulation when this is flashing at 5Hz.

In practice then, I ignore temporal effects for all cell types where they are not likely to affect the model used to describe the shape of the spatial frequency response function. Cortical cells and parvocellular LGN neurons fall into this category. Magnocellular neurons are simulated by giving them balanced receptive fields which do not show responses to diffuse light. At the level of cortical cells, constant velocity and constant temporal frequency contrast sensitivity curves have been observed to be very similar (Andrews and Pollon [5]).

The measures taken here provide a very rough acknowledgement of the effect of the temporal behaviour of the vision system, but it is difficult to go further on the limited information available (without full spatiotemporal characterisation).

3.2.4 Convolution Nodes

In order to find out how a set of neurons with identical receptive field properties responds to a test picture, two methods are theoretically possible: Either record from one neuron and drift the image past its receptive field, i.e. make a scan by moving the image; or else record simultaneously from many neurons, each with differently positioned receptive fields, while at the same time flashing a stationary image.

Considering these methods in turn: The first is more practical experimentally because only one cell needs to be isolated (although the result is dependent on the neuron's temporal response and stationary flashed stimuli are often used). This is the type of recording paradigm that has been used by investigators, who have presented a large variety of test patterns, including gratings, bars and edges. Li *et al.* [147] even used complete images.

For the simulation work presented in this thesis, the second, multiple recording paradigm, is more relevant—it may not be practical (yet), but its results are predictable by simulations that use single unit recording data. Here, I will consider the response of all neurons at each stage which simultaneously draw input from receptors, or previous stages, and these may have highly overlapping receptive fields. This is another way of looking at the system:

stationary images[1] and many identical receptive fields analysing the field of view in parallel and *en masse*.

Such a viewpoint represents a change from the single neuron centred world of the physiologist, whose response traces are for shifts in stimulus position, to an image processing paradigm, where response coordinates represent an indexing through neurons—each looking after a fixed location in the scene.

Convolution is the primary tool of image processing, and here two dimensional convolution is used to simulate the responses from many neurons. Starting with receptive field localised summation properties defined by a spatial set of weights from Equation 3.1 I create the convolution kernel by mirror-imaging this in both x and y to give $w(x,y)$. Next I define a general purpose non-linear summation operator for which discrete convolution is a special case:

$$(w \odot_n g)(x,y) = \left[\sum_i \sum_j [w(-i,-j)\, g(x+i, y+j)]^n \right]^{1/n} \tag{3.2}$$

When $n = 1$, linear convolution results and I define a special symbol for this: $\otimes = \odot_1$. The neural model from Equation 3.1 for a set of responses distributed over space, $Y(x,y)$, now becomes

$$Y(x,y) = Y_{max} \frac{f_{out}(\, w(x,y) \odot_n X(x,y)\,)}{1 + kD(x,y)}. \tag{3.3}$$

The function f_{out} is now an output transfer function over and above the power law implied by Equation 3.2.

When $n = 1$, either fast Fourier transforms or space-domain convolution can be used to implement Equation 3.3. For other values of n, a specialised space-domain algorithm is required.

For some of the stages involved in this thesis, Fourier transform convolution was avoided because subsampling was possible. Under these circumstances, space-domain processing can give a quicker result with less of a memory requirement because not many points need to be calculated.

3.2.5 Subsampling and Reconstruction

One belief which I attempt to express in this thesis, is the importance of subsampling for the economic coding and abstraction of visual information. Here, a sample is considered to be

[1] Of course in the real world, images will move across the receptor array, but this thesis concentrates primarily on spatial processing.

the response of a single neuron.

It is well known that receptive field sizes increase as one moves to higher visual areas. This increase reflects the fact that information about the scene is integrated over a growing region of visual space. A further aspect to this change is that neurons are selective for the same stimulus feature over a large area, and this means that few neurons with the same selectivity are required to cover the visual field—information transmitted varies slowly with receptive field position in the visual field. Extensive subsampling is possible because of this slow variation, and therefore reduces the number of neurons that need be allocated to any one stimulus selectivity at higher levels in the visual pathway. Sampling density is reduced as the wealth of selectivity increases.

In the work reported here, subsampling is used to reduce the computational burden where receptive fields are large. When a stage of processing introduces low-pass filtering, e.g. the centres of retinal ganglion cells, then it is always possible to invoke the Nyquist criterion to decide on the minimum rate of sampling (for the example case, ganglion cell density) that is required to represent without loss such a spatially band-limited signal. Essentially, sampling theory shows that a signal can be sampled without information loss providing that these samples are taken at a frequency that is greater than twice the maximum signal frequency. It is then possible to reconstruct the original signal at any time from such samples. Appendix A gives the standard mathematical proof behind this assertion.

In the vision system, I suggest that it is never necessary to reconstruct the original scene from the sampled neural representation—it is enough just to know that the samples are sufficient. This is because the process of vision is one of abstraction, and samples contain all the necessary information, including hyperacuities. The only case when some measure of interpolation between samples is necessary is when there are interactions between representations held at different sampling resolutions. In this case, there may never need to be an explicit reconstruction stage because divergent and diffuse neural connectivity provides the interpolative function.

Practical Considerations

All the band-pass and low-pass filters used here for modelling purposes have a smooth roll-off at high frequencies. This makes it difficult to set a sampling frequency—there is a tradeoff between minimising resolution and avoiding spatial aliasing. The 1% of maximum response point has been arbitrarily chosen as the minimum Nyquist frequency for sampling purposes.

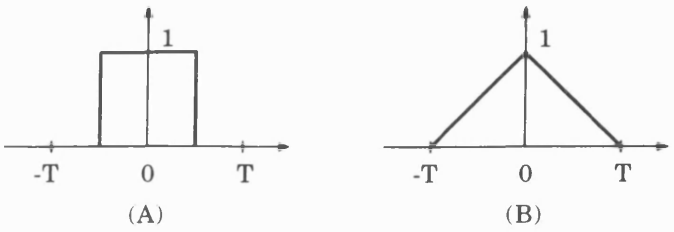


Figure 3-1: Spatial reconstruction kernels. (A) Rectangular function. (B) Linear interpolator.

The sampling frequency must then be at least twice this value.

When simulating gain control mechanisms in the cortex (see Chapter 7), interpolation is used to allow interactions between response maps held at different resolutions. This requires an interpolation filter. The perfect spatial interpolator is the $\sin(x)/x$ function as detailed in Appendix A, but this has such a large spatial support that it is unlikely that this function could be used in the brain. Appendix A demonstrates that it is often possible for a more relaxed spatial frequency roll-off criterion to be specified providing that the signal does not have much energy near the Nyquist frequency. This is often the case with the combination of sampling frequency and neural filter responses that are used here.

Figure 3-1 shows two functions which are often used for up-sampling. In the Fourier domain, the rectangular kernel results in a filter with response $2\sin(\omega T)/\omega$, where ω is the spatial frequency in radians per sample and T is the sampling interval. The linear interpolator has a response of $2[1-\cos(\omega T)]/\omega^2 T$. Figure 3-2 shows these two functions for $T = 1$.

The ideal reconstruction filter needs to suppress all signal energy above π radians per sample and it can be seen that both these functions are non-ideal. The linear interpolator is not too bad however because the lobes at higher frequencies are significantly attenuated and there is a faster roll-off above $\omega = \pi$. Other filters were evaluated, but the linear interpolator has been used because it is fast, has a localised support region, and gives good results.

3.3 Hardware and Software

In this section I introduce briefly the hardware configuration which is used to allow parallel computing, and then concentrate on the software development of a suitable environment for image processing. Selected implementation examples are also given.

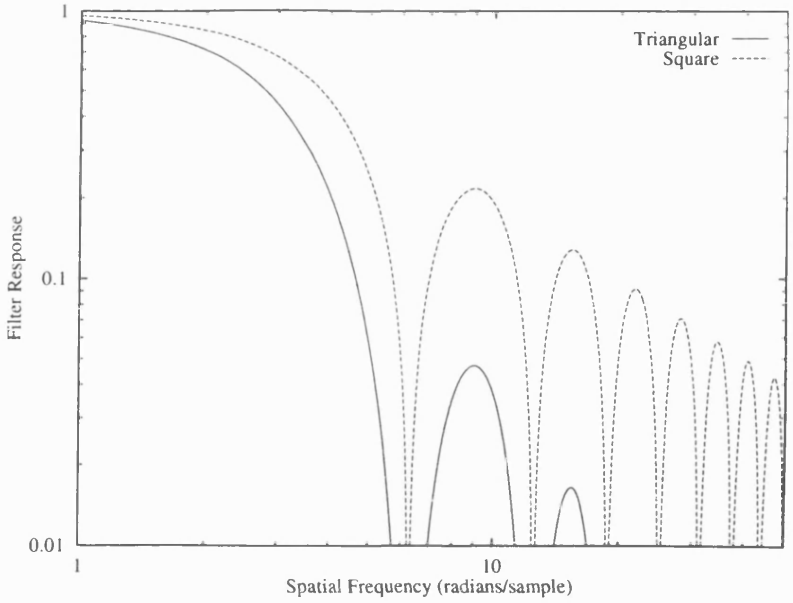


Figure 3-2: Filtering action of practical reconstruction kernels.

3.3.1 A Parallel Computer

All the large-scale simulation work presented in this thesis was carried out using a MASPAR MP-1 computer. This is a massively parallel machine with a SIMD architecture, and consists of the following major subsystems:

The Array Control Unit (ACU) This controls the processor array by broadcasting a single instruction stream and data to all the array elements. It is a self-contained processor which can carry out serial processing involving "singular" variables. It also handles communication with the front-end machine.

The Processor Element Array (PE Array) This executes the instruction stream broadcast by the ACU and handles all the data parallel processing. The PE array consists of 1024 processors arranged in a 32×32 array. Each one has 16K of local memory, a 192 byte register set and hardware floating point support. A variable held on the PE array is known as a "plural" variable and it can have different values on each PE. All PEs execute the same stream of instructions but these act on different data. Not all PEs need be active at the same time, and this allows scope for plural conditional execution.

Communication Mechanisms Communication between PEs can be achieved in two ways: Firstly, an eight way "Xnet" network allows communication between neighbouring processors along the eight compass directions. Secondly, a global router allows random

PE-to-PE communication via a hierarchical cross-bar. There is a speed advantage associated with using Xnet communication.

The I/O Subsystem This subsystem allows plural data to be DMA'd between the PE array and the host machine via a VME bus.

Front-End Host The host machine is a DECstation 5000 running UNIX with a local disk, and network support.

Programming Language

The MASPAR is programmed in MPL which is a language based on ANSI C. Various statements, keywords and library functions have been added to support data-parallel programming. The major addition is support for "`plural`" variable declarations. It is then possible to write "`a = b + c`", where the three variables hold different values on each processor. Other additions include constructs to allow communication between PEs and also plural conditional statements which change the active set of processors.

For further information consult the MASPAR MPL programming reference manual (MasPar [159]).

3.3.2 Creating an Image Processing Environment

Approximately 3500 lines of MPL code has been written in order to create a suitable environment for simulation work. From the beginning, a number of requirements were identified:

- The environment should efficiently exploit the data-parallel architecture of the MASPAR.
- It should allow efficient storage of image data because large memory use was predicted.
- It should provide fast image processing library routines.
- It should make possible the alteration of many parameters/constants used in the simulation without re-compilation.
- It should hide front-end and back-end communication from simulation programming.

To meet these goals, support software to handle image loading, saving, allocation, deallocation, and front-end/back-end image transfer was developed as an initial framework. This involved writing code to run on both the host (DECstation) in C, and back-end (MASPAR)

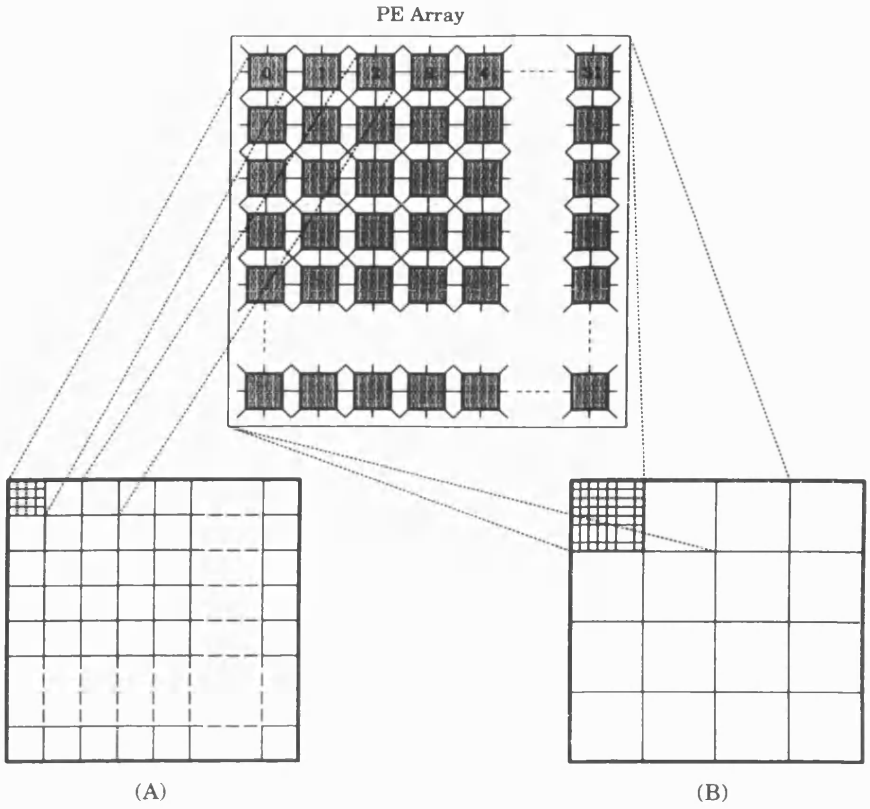


Figure 3-3: A 128×128 image can be mapped onto the PE array by using either the "patch" method (A) or else by cutting and stacking the image (B). Also shown are the Xnet inter-PE communication lines.

in MPL. Later, all image processing library and simulation software was written on top of this in MPL.

I now describe various aspects of the support software and library.

Image Format

On the front-end machine, response map images are allocated as a continuous block of memory associated with housekeeping and row pointer structures. On the PE array, each image is spread out so that part of the image is allocated in a small amount of memory on each processor. Images are generally larger than the processor array size (32×32), so many pixels elements are present on each PE. Image dimensions are always a multiple of 32 pixels. Information about each image on the PE array, together with associated row pointers which are used to speed up access, is maintained in ACU memory.

There are two useful ways that image data can be distributed across an array of processors so that each can work on part of the image separately. The first is the "patch" method of

Figure 3-3A, in which the image is divided up into 32×32 patches and each patch of pixels is allocated on the appropriate PE. For the "cut and stack" method (Figure 3-3B), each pixel is allocated on a different PE, and the image is cut up and stacked in 32×32 squares.

Here I have employed the "patch" method because nearby pixels are generally on the same processor, so inter-processor communication is reduced during convolution operations. The "cut and stack" method always requires extensive long distance communication.

The MASPAR MPL mathematics library provides a fast two dimensional Fourier transform routine and this is employed for convolution operations where the kernel would be too large. This routine uses a "cut and stack" representation, so conversion software has been developed to change between the formats. This relies on the router communication mechanism.

Memory Management

On the host machine, image memory allocation and de-allocation uses standard C library calls. On the PE array, allocation of image memory needs to be robust and separate from any other allocation scheme. This is because an image cacheing method is used which relies on knowledge of memory usage to clear space for priority image data. Memory management software was written to achieve this goal. Initially, a large block on each PE is requested from the run-time "malloc" routine. This memory is then managed as a heap so that images can be created and destroyed as requested.

Image Management

There are two different types of image: Temporary images are created and destroyed on demand on the PE array and are used to store partial results. Managed images store complete neural response maps and there may be up to 300 present during a large simulation of the primary visual cortex. These are maintained on the front-end, on the PE array, or in both places. They are referred to by name because pointers cannot reference objects across the IO connection.

In general, the front-end machine maintains a list of all managed images together with their status. These are transferred back and forth between the front-end and PE array as necessary. After images are transferred to the PE array, they are not normally written back unless they become swapped out because they have not been used for some time and there is a PE memory shortage. This swapping in and out is necessary because the PE array has limited memory, while the front-end machine has a large amount of disc-based swap space.

Code was written to make the swapping process as fast as possible and also transparent to the application software.

Parameter Database

Simulation work generally requires large numbers of parameters which must be specified and varied to show different effects, even though the underlying software does not need to change. In order to avoid excessive re-compilation, the support software was designed to read a database file containing name-value pairs. A typical example is shown in Appendix B. After loading, this database is maintained on the front-end machine and values are transferred to the ACU when requested by simulation software.

Library Routines

The image processing library contains around 30 routines which implement commonly used parallel operations. These include creation of convolution kernels, subsampling, spatial convolution, spatial non-linear summation, Fourier convolution, Gaussian smoothing, high-pass filtering, conversion between image types, fetching image data from offset locations, interpolation, image clearing, copying and conversion to and from "cut and stack" format.

3.3.3 Implementation Examples

In this section, I give implementation details for two library functions by way of example. These give a flavour for programming on the MASPAR.

Fetching and Sending

The simplest and fastest image transforms that can be handled by the MASPAR are those which do not require comparison of image data between processors. For example, square rooting all pixel values is a function that allows all processors to operate independently—each one sequences through the subset of image pixels for which it is responsible and applies the same function. This operation is then 1024 times faster than if only a single PE were employed.

More complex operations involve communication between adjacent processors because each one must gather or transmit pixel values some distance across the array.

Let us consider the problem of transmitting a plural floating point value from a set of processors to an general image coordinate offset (dx, dy) relative to a starting point (x, y)

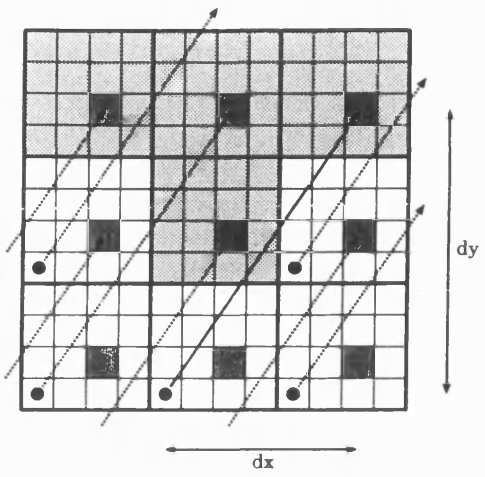


Figure 3-4: This diagram shows a nine processor section of the PE array. A value stored on the bottom centre PE is transmitted to a destination (dx, dy) away in image coordinates. Many other PEs also transfer data along the same vector (dotted arrows). Some PEs are inactive and do not send data (light shaded).

within the patch of image held on each PE. The destination image location may be on the same or a different processor, and we wish to carry out this operation simultaneously for all source PEs in an active set. Figure 3-4 illustrates the problem.

Prior to transmitting any data, we first use (x, y) and (dx, dy) to work out how many processors the data must be shifted in the x and y directions. We call these values `hopx` and `hopy`. When these are both zero, the destination location is on the same processor as the value to be stored. Next we must work out the location in the image patch held on the destination processor that is to receive the transmitted value. We put this coordinate in `newx` and `newy`.

When it comes to the shift, we can use the Xnet communication channels to transmit the data `hopx` processors east/west and `hopy` processors north/south, and store the result at (`newx,newy`). This is complicated, however, by the added requirement that not all the processors in the PE array should take part in this data transfer—only a subset of them. This is solved by careful handling of the active set, resulting in the following code segment:

```
/* Shift the value to the target processor (+ve y is down) */
ptr= &ip_float_value(image,newx,newy);
all active=0;
active=1;
all {
  if(hopy>0)
    {
```

```
        xnetS[hopy].val=val;
        xnetS[hopy].active=active;
      }
    else if(hopy<0)
      {
        xnetN[-hopy].val=val;
        xnetN[-hopy].active=active;
      }
    if(active)
      {
        if(hopx>0)
          xnetE[hopx].*ptr=val;
        else if(hopx<0)
          xnetW[-hopx].*ptr=val;
        else
          *ptr=val;
      }
}
```

The first line gets a singular (ACU) pointer to the destination location in PE memory for all processors. Next, all processors set their copy of the plural variable `active` to zero. After this, only those processors that were active prior to the code segment set `active` to one, and so the pattern of active processors at the start is saved in a variable.

The rest of the code proceeds with all PEs activated. Firstly, the value to be sent, `val`, is transmitted the correct distance in the north/south direction making use of the `xnet` construct. This occurs for all processors simultaneously. The value of the `active` flag is also sent because only the locations that receive data from source PEs that were initially active must actually store the value into the image.

After the vertical transfer, a plural "if" statement tests the `active` flag which is now on every processor that is the correct distance north/south of the source PE. The body of the "if" statement is only then executed by those processors that must store `val` into the correct destination PE image patch by transmission in an east/west direction.

This example illustrates some of the complexities of data-parallel programming.

Parallel Convolution

The parallel convolution algorithm implements simultaneous space domain convolution and subsampling. It is fast because Xnet communication is infrequent and only ever involves adjacent processor hops, and also because convolution results are only evaluated at the resolution of the output image.

The convolution algorithm uses three images: an input image, an output image and a convolution kernel. Input and output images must be a multiple of 32 pixels in size and the output image may be any integer factor smaller than the input. The convolution kernel may be any size, but commonly it will be square and have odd dimensions.

The input image and kernel are used only as source information. The output image starts off zeroed and is used to accumulate results. Data in the output image is shifted around between processors until all products have been accumulated and the output values are in correct spatial registration with the input. Fourier wrap-around is implemented in this algorithm by virtue of the fact that Xnet communication wraps around the PE array from east to west and from north to south. This means that there are no blank borders to the output image.

First consider the case when there is no subsampling and the input and output images have the same size. The algorithm then proceeds as follows: An outer loop chooses values from the kernel in an ordered fashion. One of these values is used to multiply the entire input image, and the resultant image is added to the output image. An adjacent kernel value is then chosen and the output image is shifted by one pixel. The new kernel value then multiplies the input image and the result is accumulated into the output image in a new position. This process repeats until all the kernel points are used. The trick is to employ a shifting strategy which coordinates choosing kernel values and output image data motions so that, firstly, convolution is achieved and, secondly, the output ends up in the correct registration with the input image.

In practice, I use a spiral motion for data motion and kernel value selection, with the output data spiralling in the anti-clockwise direction and the kernel values being chosen in a clockwise fashion, starting from the bottom right of the kernel and working in towards the middle. This removes the requirement for corrective output data shifts after the process is complete for kernels which have odd dimensions.

It should be noticed that in this scheme, only one Xnet communication to an adjacent processor is required per image patch row/column per kernel value, and this makes the process fast. When subsampling is included, the amount of communication is further reduced because output image shifting only occurs once for every *subsampling-factor* accumulates (approximately).

When subsampling takes place, only a fraction of the input pixels are added to the output image for each kernel value. The pattern of input image pixels that are used each time moves

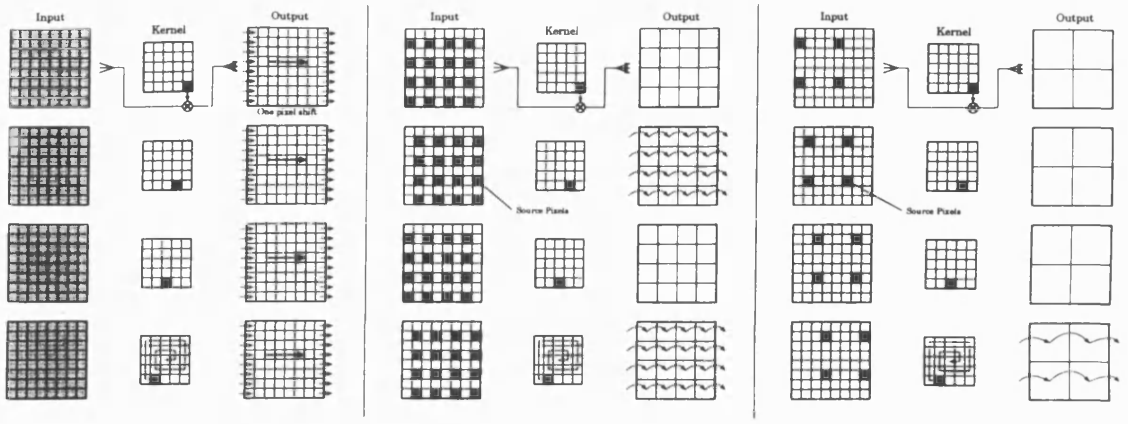


Figure 3-5: Subsampling convolution. The three panels show subsampling by 1, 2 and 4. The shift arrows indicate output data motions that happen after the accumulation of partial results. For subsampling by 2 or 4, fewer input image points are gathered for each kernel pixel, and fewer output image shifts occur.

in a spiral, as does the data in the output image, but the output data moves less frequently because of the subsampling factor. This is difficult to visualise, so Figure 3-5 attempts to illustrate the process.

Running on a 1024 processor MP-1, this algorithm can convolve a 512×512 image with a 27×27 kernel to give a 256×256 result in 2.5 seconds. The same operation takes 32 seconds on a Sun SPARCstation ELC.

3.4 Conclusions

In this chapter I have attempted to lay a background for the simulation work that is to be described. This background includes assumptions underlying the "situational" and neural models. I have also described the image processing and sampling theoretic approach, and the software development. We now have the tools in place for modelling vision.

Chapter 4

From the Retina

4.1 Introduction

In this chapter I am concerned with modelling the transformations that take place within the retina. Here, we find out how light entering the eye affects the photoreceptors and how signals from these are processed so as to produce the various types of response that have been recorded from the ganglion cell layer (Figure 2-4). The axons from these cells form the optic nerve which then transmits visual information to the rest of the brain.

Firstly, we take a brief look at the full simulation scheme that is used to model data transformations up to the output of retinal ganglion cells. Secondly, each transformation stage is introduced in turn. In Section 4.5, which deals with light adaptation, some of the implications of this non-linear stage for contrast and brightness perception are studied. Spatial opponency is described in Section 4.7 of this chapter but chromatic opponency is deferred until the next. Lastly, I concentrate on mechanisms of retinal contrast gain control which I believe are instrumental in producing useful signals from the M-cell pathway. Various simulation results are included throughout this chapter.

As we have already seen in Chapter 2, the retina contains a receptor layer made up of rods and cones, responsible for vision under *scotopic* (dark adapted) and *photopic* (daylight) conditions respectively. This simulation is restricted to normal daylight vision where photopic conditions apply, and therefore rod receptors are not included in the model. All three types of cone receptor are considered, however, since the aim of this thesis is to explore colour vision.

Only receptor and ganglion cell stages are explicitly included here. This is because the

simulation is targeted at transformations that take place within the retina as a whole, without being too concerned with the details of unit activity in the intermediate neural layers. A significant amount of simplification is present because temporal behaviour is bypassed in this work.

The lateral geniculate nucleus (LGN) is the target for optic nerve fibres and is the subject of the next chapter. There are not many differences between the receptive field properties of cells in the retina and geniculate nucleus, but some properties have been measured more thoroughly at the geniculate site. In the current chapter, I therefore concentrate separately on contrast, brightness, and spatial behaviour, leaving a full simulation of spatio-chromatic receptive field responses until Chapter 5.

4.2 A Model of the Retina

In Chapter 3 I introduced the "situational" model which is used to provide a context for simulation work, namely, that a subject is positioned in front of a TV screen which can display test images. Here, I attempt to calculate visually-evoked neural responses in the subject's retina.

Figure 4-1 shows various stages of the model retina. These are described below:

TV Model The television model converts from the pixel values of a colour image to a description of the light pattern that would be produced on the screen of a CRT display device when that image is displayed. This accounts for the screen gamma transform and background lighting. It includes a colour-space conversion from RGB values to CIE XYZ coordinates.

Cone Curves This stage transforms the light pattern incident at the retina into photopigment activation for the three cone types using a linear colour-space conversion to cone coordinates. Relative cone sensitivities are also taken into account.

Light Adaptation The effects of light adaptation are simulated using a non-linear model transforming light into response levels in the outer plexiform layer. This is the major forward-path non-linearity in the retina and it allows for significant contrast constancy.

Spatial Summation Activity from nearby cones is linearly summed together using a two-dimensional Gaussian spatial weighting function. This is used to create "centres" and "surrounds" which form the sub-units of ganglion cell receptive fields.

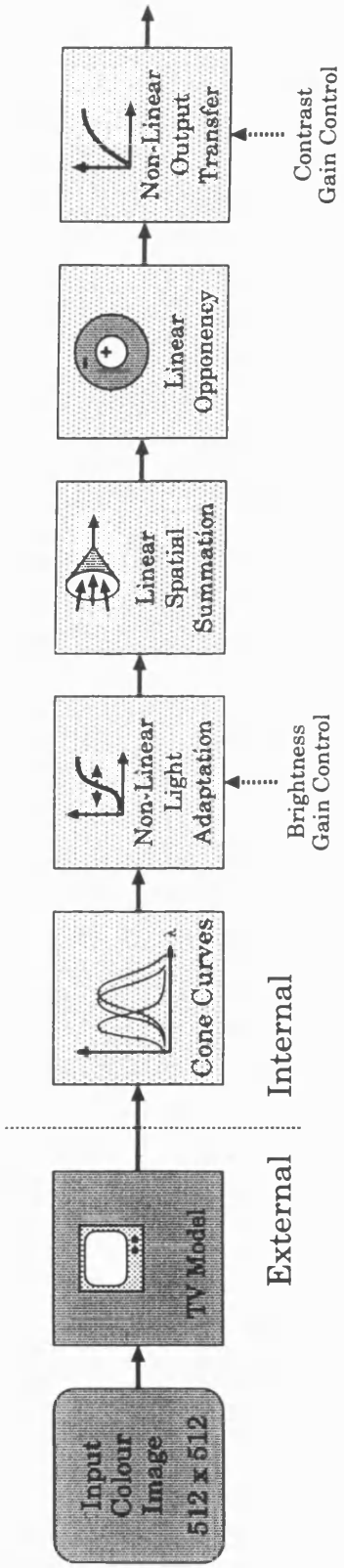


Figure 4-1: Stages in the Model Retina

Opponency This stage introduces spatial opponency by subtracting centre and surround responses. When centre and surround each pool responses from different cone types then chromatic opponency is also produced.

Output Transfer For ganglion cells with low contrast sensitivity, the output transfer function is modelled with a linear function. For those with high contrast sensitivity, local contrast measures are pooled and fed back to introduce divisive compression and contrast gain control.

4.3 TV Model

A *television model* is used to simulate the conversion of pixel data into light intensities that takes place when an image is being displayed on a computer screen. Figure 4-2 shows a block diagram of the process. The fundamentals of image display are described by Hall [94] chapter 5, and Foley *et al.* [81] chapter 13.

Conventionally, a computer display device does not provide a linear mapping between image data and light intensities. Instead a power law is applied with power index γ. This is often composed of two cascaded non-linear stages: Firstly there may be an adjustable power transformation at the level of the digital to analogue conversion within the display hardware. Secondly, there will be a fixed power law relating the voltage driving the electron beams to the output light intensity of the screen. The adjustable *gamma correction* (implemented either in hardware or software) can be used to tailor the overall power transformation. The two stages can therefore be treated as one composite stage.

Images taken from TV cameras or scanned-in are not usually intended for display on linear screens because the input device incorporates a gamma transformation that will be cancelled by a normal non-linear display unit. This means that to transform pixel values to light in the model, I incorporate an expansive power law to model the display hardware.

Vision scientists carefully ensure a gamma of one to produce a linear display for expermental purposes. When displaying images that have been scanned or captured from camera, however, the gamma should be set so as to undo the compressive input transform. The image displayed on the screen will then be faithful to the scanned original viewed under suitable lighting conditions.

A term to account for background illumination is also required. This is because images displayed by self-luminous display devices do not modulate the incident light as ordinary

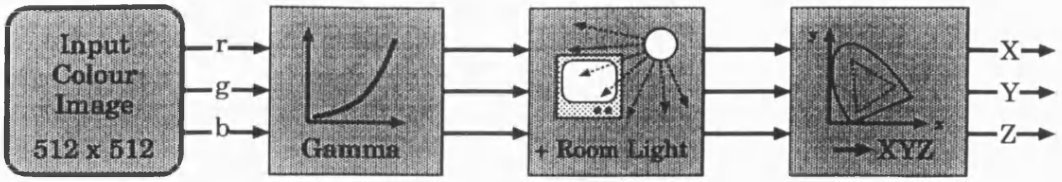


Figure 4-2: TV model: From pixel values to CIE coordinates

objects do—instead they generate their own light. Ambient room lighting therefore produces a fixed intensity pedestal onto which light from the display device is added. This background lighting tends to reduce the visibility of dark image features. We shall see the reasons for this later.

At this point the TV model can be described by the equation

$$I_r(x,y) = I_{back} + I_{max}(P_r(x,y)/255)^{\gamma}. \tag{4.1}$$

Here, I_{back} is the background intensity, I_{max} is the maximum phosphor intensity (measured in $\mathrm{cd}\cdot\mathrm{m}^{-2}$)[1] and $P_r(x,y)$ is the pixel value array (assuming a value range from 0–255). This equation describes the "red" channel. Three equations of this form define the emitted light in phosphor colour space. Knowledge of the emission spectra for each of the three phosphor types would allow us to calculate the amount of energy present at each wavelength in the resulting light mixture. In practice, this calculation is not necessary because of the trichromacy of colour vision.

In 1931 the *Commission Internationale de l'Éclairage* (CIE) defined a standard perceptual colour-space in which all lights[2] that can be distinguished are represented by positively weighted combinations of three virtual colour primaries: **XYZ**. The amount of contribution from each primary required to match a particular colour defines the XYZ coordinate for that light. Given the spectrum of a light source it is possible, by integration, to work out the coordinate in CIE colour-space where one finds all lights that appear to be the same colour as that source. This process is described more fully in Foley *et al.* [81].

[1] I_{back} and I_{max} together set the maximum drive to the CIE Y primary and are therefore luminance measures with units of $\mathrm{cd}\cdot\mathrm{m}^{-2}$.

[2] In this discussion the word "light" is used to refer to the physical superposition of electro-magnetic waves entering the eye, while "colour" refers to the perceptual experience which is produced.

The *chromaticity coordinate* (x, y) of a colour is defined to be

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z}. \tag{4.2}$$

This coordinate, together with the luminance Y, provide another way of specifying the colour of a light source. Chromaticity coordinates depend only on the dominant wavelength and saturation and are independent of the amount of luminous energy. The CIE chromaticity diagram is obtained by plotting x and y for all visible colours.

In order to transfer from the *rgb* colour-space of the display hardware to XYZ coordinates, which then provide a standard way of specifying colour, we need to know the chromaticity coordinates for each type of display phosphor and also the *white-point* of the monitor. This is the coordinate of the "white" that is generated when all three phosphors are driven equally (i.e. r, g and b are equal).

The standard NTSC phosphor coordinates are:

	Red	Green	Blue	White
x	0.67	0.21	0.14	0.31
y	0.33	0.71	0.08	0.316

These can then be used as parameters in the conversion to XYZ colour-space using the method described on pages 586–7 of Foley *et al.* [81]. This method gives the matrix equation

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} Y_r(\frac{x_r}{y_r}) & Y_g(\frac{x_g}{y_g}) & Y_b(\frac{x_b}{y_b}) \\ Y_r & Y_g & Y_b \\ Y_r(\frac{1-x_r-y_r}{y_r}) & Y_g(\frac{1-x_g-y_g}{y_g}) & Y_b(\frac{1-x_b-y_b}{y_b}) \end{bmatrix} \begin{bmatrix} I_r \\ I_g \\ I_b \end{bmatrix} \tag{4.3}$$

where $Y_r = 0.299$, $Y_g = 0.587$ and $Y_b = 0.114$ for the white-point given above.

The exact chromaticity coordinates for the phosphors and white-point are not critical because I am not attempting to simulate any particular piece of hardware. Setting up the parameters so that they mirror some actual display configuration is only used for convenience, to obtain a subjective assessment of the model's performance.

To summarise: Pixel values are converted to light intensities by using a power law. Background light is added and the resultant is converted to CIE XYZ coordinates. (Note that background light is specified in terms of image *rgb* coordinates. This means that "display-white" background illumination can be obtained by making I_{back} equal for each channel in Equation 4.1.)

4.4 Cone Model

I now present a model for the initial retinal stage which accounts for the spectral sensitivity of each receptor.

Three types of cone photopigment are present in the primate retina, having peak sensitivities in the long, medium and short wavelength regions of the visible spectrum. The corresponding receptors will be called red, green and blue cones throughout this discussion, although it should be understood that sensitivity curves are wide and so these labels are somewhat arbitrary. Figure 2-6 (in Chapter 2) shows the normalised Smith and Pokorny [218, 219] sensitivity curves. Absolute photopigment sensitivities are hard to obtain, but since the green is generally thought to be the most sensitive at its peak, we define κ_r and κ_b to scale the peak sensitivities of the red and blue pigments relative to that of green. Wald [238] has estimated that $\kappa_r \approx 0.9$ and $\kappa_b \approx 0.1$.

Since the cone curves can be well approximated by linear combinations of the CIE colour matching functions, it is possible to convert from light specified in XYZ coordinates to cone pigment activation by using a matrix multiplication:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 0.2435 & 0.8524 & -0.0516 \\ -0.3954 & 1.1642 & 0.0837 \\ 0.0 & 0.0 & 0.6225 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{4.4}$$

This transformation is from Smith and Pokorny [219] as used by Guth *et al.* [93] and gives the amount of light absorbed by each cone type.[3] The sensitivity factors κ_r and κ_b are then applied to attenuate the red and blue channels. In this way, we produce three transformed image arrays which represent the amount of pigment activation in the outer-segment of each cone type.

4.5 Adaptation Model

Following conversion of light to cone coordinates, the receptor responses, which, via the transforms above, are initially linearly related to light intensity, are subject to light adaptation mechanisms acting within the outer plexiform layer of the retina. I now make the simplifica-

[3]Smith and Pokorny's cone sensitivities are actually specified in terms of Judd's modification to the 1931 CIE colour matching functions. This difference affects the short wavelength end of the spectrum below about 450nm. I have compared the sensitivity curves and the difference is not significant for the type of simulation work presented here.

tion of assuming that light adaptation is a receptor-level phenomenon. The model which is used here results in a non-linear response relation to light intensity over the range 10^{-1}–10^{5} cd $\cdot$ m^{-2}. This single model is intended to account for all the brightness adaptation effects in the retina, and as such can only be a simplification. In practice, significant neurally-mediated adaptation occurs at the level of receptive field subunits (Enroth-Cugell and Shapley [73]).[4] However, there are a number of useful results that can be obtained using the model as it stands (e.g. cone difference signals are proportional to contrast) and these will be introduced in due course.

I employ an electrophysiological receptor model (EP model) that predicts the cone responses that result from light flashes, given a particular state of adaptation. This model was first used by Naka and Rushton [179] to describe S-potentials in the fish retina. Other researchers have found it useful in describing post-receptoral behaviour in the primate (Lee and Virsu [142]; Valberg *et al.* [231]; Cornsweet [48]) and for psychophysical modelling of brightness perception (Xie and Stockham [251]). It has the general form

$$R = \frac{R_{max} I^n}{I^n + \sigma^n} \tag{4.5}$$

where R is the response magnitude, I is the test flash intensity and σ is the half-saturation constant which is used as an adaptation parameter. Values for n in the range 0.7–2.0 have been used by various workers but here I use a value of 1.0, following Naka and Rushton [179] and also Xie and Stockham [251]. Guenther and Zrenner [92] were able to fit the suprathreshold responses to intensity shown by cat ganglion cells with this function when $n = 1.0$. An empirical relation between the adapting background intensity I_b and the adaptation parameter σ has been established as

$$\sigma = \beta I_b^{\alpha} + S_{min} \tag{4.6}$$

where the values for α and β are about 0.69 and 5.83 cd $\cdot$ m^{-2} respectively (see Xie and Stockham [251]). These figures assume that the receptors have a $V(\lambda)$ spectral sensitivity, i.e. they are luminance sensitive. Specific wavelength adaptation details will be considered

[4] Late in the writing of this chapter, I realised that evidence suggests that the adaptation and spatial summation stages shown on Figure 4-1 should really be reversed in order. Preliminary experimentation with this new ordering gave unsuccessful results with the current adaptation model because unrealistic non-linearites are introduced. The original order for these stages predicts that contrast sensitivity curves and response versus spatial frequency plots will have identical shapes. Campbell *et al.* [42] found that this appeared to be true in some cases. These observations show that the current model is acceptable, although further investigation is required.

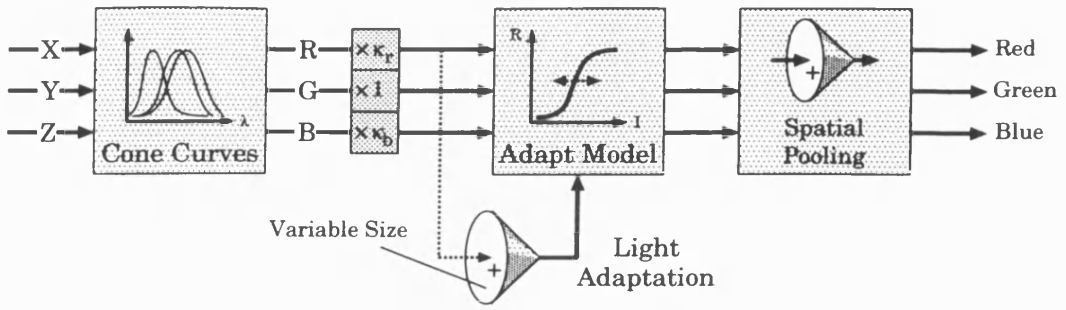


Figure 4-3: Adaptation model. Light adaptation occurs for each receptor channel separately. The cone-shaped symbol in the adaptation path represents summation of responses over a limited spatial region. This *virtual* summation region has an effective size which depends on eye-movements and other temporal factors.

later. In this equation, I have modified the relation used by Xie and Stockham [251] by adding S_{min} which limits the neural gain at low levels of adaptation luminance. This provision allows the model to fit the **P**-cell contrast *vs* retinal illumination plot of Purpura *et al.* [193].

Figure 4-4 shows how the cone response R in Equation 4.5 varies with light intensity at a selection of adaptation points (with $R_{max} = 1$). Each response is an approximately logarithmic function of test flash intensity, I, about the adaptation point, giving a range of about 1.5 log intensity units before saturation. Also shown is the response when the retina is allowed to adapt to the test luminance i.e. when $I_b = I$. This produces a response function that is logarithmically proportional to intensity over a wide range. (A response that is truly logarithmically proportional to intensity would be a straight line on this graph.)

Xie and Stockham [251] have attempted to unify a number of laws of brightness perception and they show that, when the retina is allowed to adapt to the stimulus, this model can predict the shape of just-noticeable-difference curves for brightness increments, Weber's law and, with some modification, the Stevens power law.

Where does the adaptation intensity come from? Primarily, adaptation occurs at the level of a localised post-receptoral site, so the adapting intensity is dependent on the light that is seen at each individual retinal location. The responsiveness of this site depends on its stimulus history, since adaptation to light levels involves temporal integration. This introduces the time dimension and means that under normal viewing conditions, eye movements become significant in determining the adaptation state. Depending on the extent of visual fixation, each receptor will sample intensities from a larger or smaller region of the scene. If the subject is fixating a target, small involuntary eye-movements will result in the summation

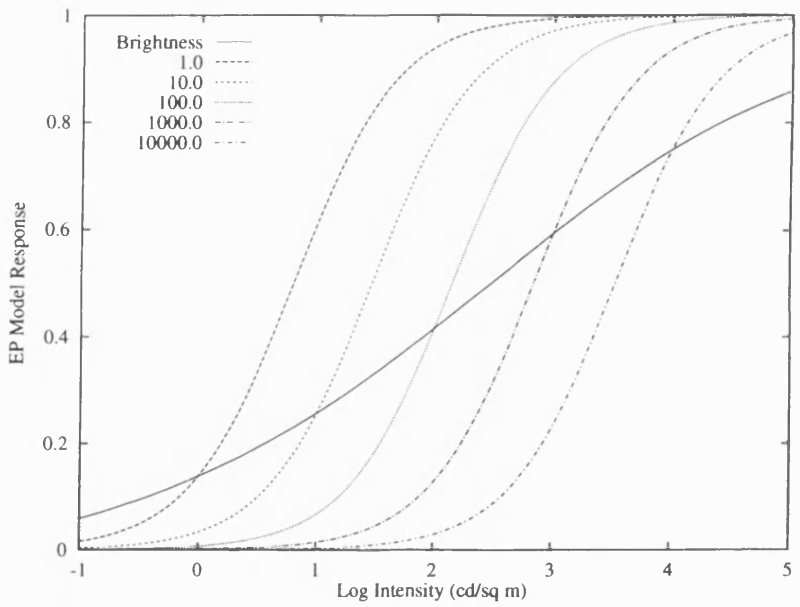


Figure 4-4: Response of the EP model of Equation 4.5 to test flash intensity at adaptation levels of 0–4 log cd $\cdot$ m^{-2}. Also shown is the "brightness" response that results when the system is allowed to adapt to the test.

of intensities from a local target-centred region and this will determine the level of adaptation. For more relaxed viewing conditions, the global average image intensity or background lighting will play a role, as will light scattered in the eye.

In order to cater for these different conditions, I include two possible types of adaptation: In the first, the local average photopigment activation over the entire scene is used to set the adaptation point for all receptors of one type. This is the simplest configuration. In the second, the receptor photopigment response map is first convolved with an isotropic low-pass filter with a Gaussian point-spread function. The resulting blurred image is used to set the adaptation point of each output. This option simulates more sustained fixation because the adaptation control is effectively gathered from a limited spatial area. The radius of the Gaussian filter can be set to vary the locality of adaptation. For the filter in question, the point-spread function is given by

$$G(x,y) = \frac{1}{\pi r_c^2} \exp\left(-\frac{x^2+y^2}{r_c^2}\right) \tag{4.7}$$

where r_c is the radius at $1/e$ and (x,y) are spatial coordinates measured in degrees. The choice of a Gaussian is not critical—Xie and Stockham [251] have used a different type of low-pass function—the main point being that the filter should approximately reflect the probability

density function of eye deviations from an intended fixation point.

The situation is complicated slightly by the fact that there are three types of cones. Since the three wavelength-sensitive mechanisms adapt more or less separately, we must treat them as three response channels, each with its own adaptation "map". This means that light intensity specified in CIE coordinates is first converted into cone coordinates (producing three cone pigment activation maps). Each of these channels is then convolved with a Gaussian and the resultant is used to set individual adaptation points in that colour channel. The EP model then receives one input from the adaptation map and one from the pigment activation map. In this way we accommodate colour and spatial adaptation effects.

If global rather than local adaptation is used, then this corresponds to the situation where a subject is viewing a blank screen which has the same luminance and colour as the average over the image to be displayed. When the eyes are fully adapted, the image is turned on and neural responses are sampled after some suitably short delay, during which time new adaptation effects are assumed to be insignificant.

The response for each cone type is therefore given by

$$r(x,y) = \frac{R_{max} I(x,y)}{I(x,y) + \beta I_b(x,y)^{\alpha} + S_{min}}. \tag{4.8}$$

I use $r(x,y)$ to represent the general receptor response or in conjunction with $g(x,y)$ and $b(x,y)$ to represent all three cone types. For global adaptation, for all (x,y), $I_b(x,y)$ is equal to the average of $I(x',y')$ over all (x',y'). For local adaptation,

$$I_b(x,y) = I(x,y) \otimes G(x,y), \tag{4.9}$$

the symbol $\otimes$ indicating convolution in two dimensions of space.

I will now illustrate the behaviour of the adaptation model by first considering the response to a gray-scale staircase, illustrating the results graphically. Next, I will apply it to a natural colour scene in order to obtain full two-dimensional response maps for the three receptor channels.

Figure 4-5 shows a one-dimensional greyscale staircase which can be considered as a cross-section through a greyscale test image, stored in the computer's memory. Figure 4-6 shows the result of applying Equation 4.1 to obtain the luminance staircase which is present at the observer's retina. The effects of the expansive power law and additive background illumination are evident. Also shown are graphs of local and global adaptation levels. The

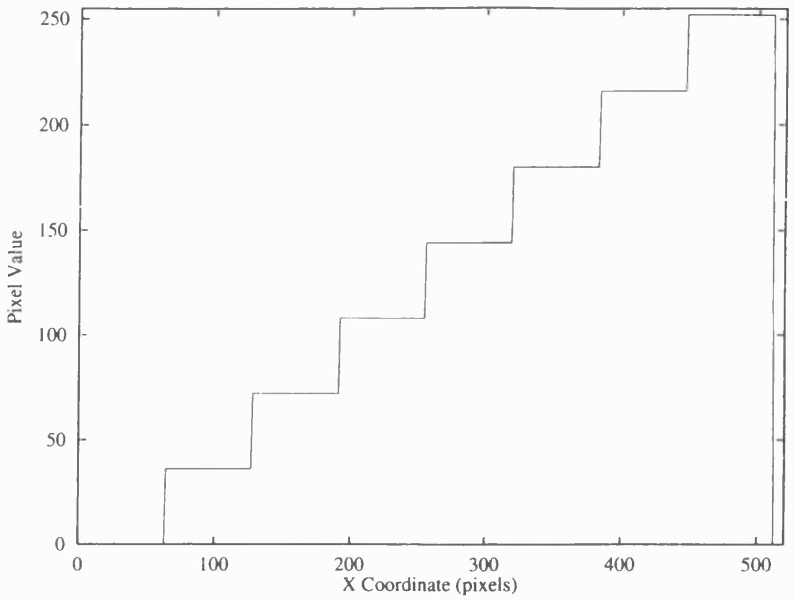


Figure 4-5: Greyscale staircase. Here the x-coordinate is given in terms of pixels, where 512 pixels are equivalent to 10° of visual angle under the viewing conditions used here.

global level is merely the average luminance of the staircase. The local adaptation graph results from convolving this staircase with a Gaussian, as defined by Equation 4.7, to give a smoothed staircase. (Note that in order to implement this convolution I have assumed that the image "wraps round" so that in effect it is as though the observer sees a line of greyscale staircases, or in the two-dimensional case, a tesselation of identical images. This provision simplifies the computation involved significantly.)

Figure 4-7 shows the responses that result from the use of Equation 4.8. Since this transformation is compressive, the expanding luminance intervals are evened out. In the case of global adaptation, the staircase is steeper because we are following one of the dotted curves in Figure 4-4 about a fixed adaptation point. Local adaptation reduces the gradient to an extent which depends on the filter radius r_c. This also introduces a high-pass filtering effect in which the perceptual brightness steps are no longer flat but scalloped. At the higher response levels, the step differences reduce and the overall transform becomes compressive. This is primarily because the monitor non-linearity is a power function, rather than a logarithmic function.

The origin of the scalloped waveform is easily understood. Consider a point on the waveform of Figure 4-6 at the foot of a step. By virtue of temporal averaging, the system will be seeing an adaptation level from a region which includes the next step up. Since this is

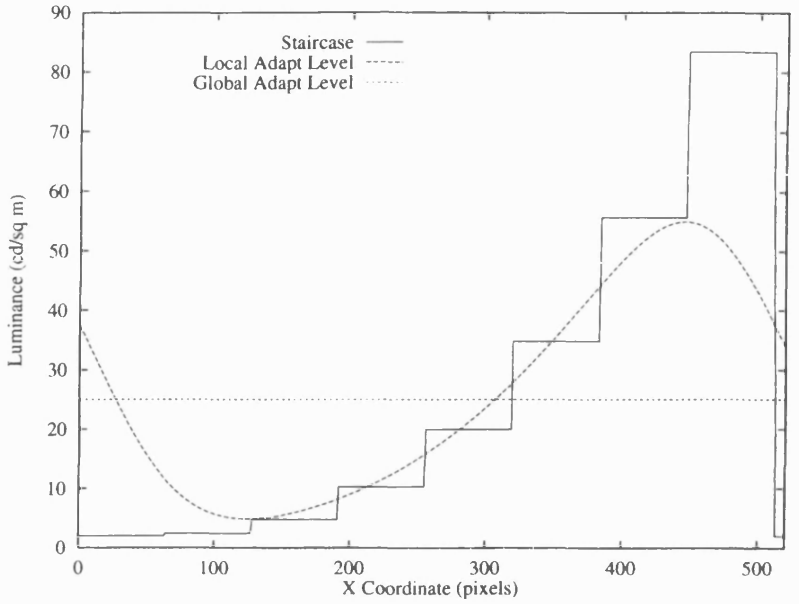


Figure 4-6: Luminance staircase and adaptation levels. The pixel value staircase has been subject to the power transform defined by Equation 4.1 ($\gamma = 2.7$, $I_{max} = 84\,\mathrm{cd}\cdot\mathrm{m}^{-2}$, $I_{back} = 2\,\mathrm{cd}\cdot\mathrm{m}^{-2}$). Dotted lines plot local and global adaptation luminances (for local adaptation, $r_c = 1.5°$ (77 pixels)).

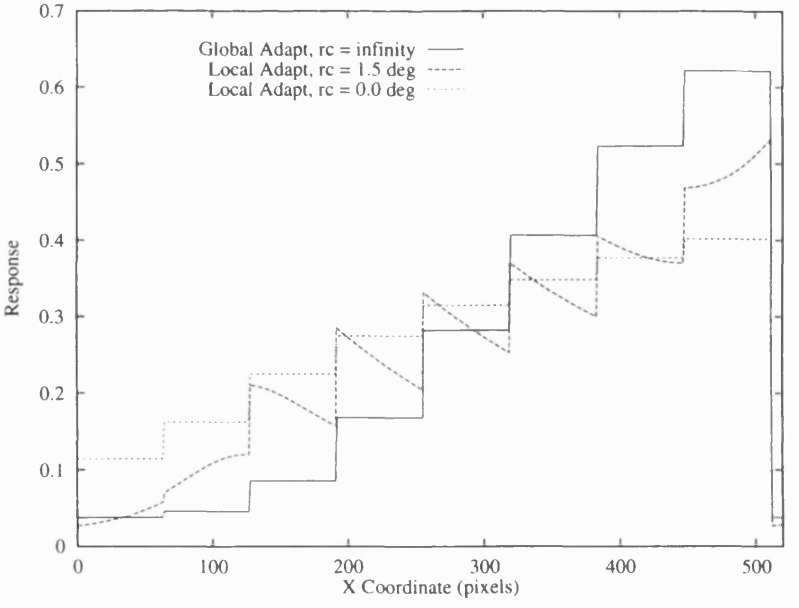


Figure 4-7: EP model response to greyscale staircase under conditions of local and global adaptation. Local adaptation may produce a induced brightness effect and it also reduces the apparent brightness range. ($\alpha = 0.69$, $\beta = 5.83$, $S_{min} = 0.25$)

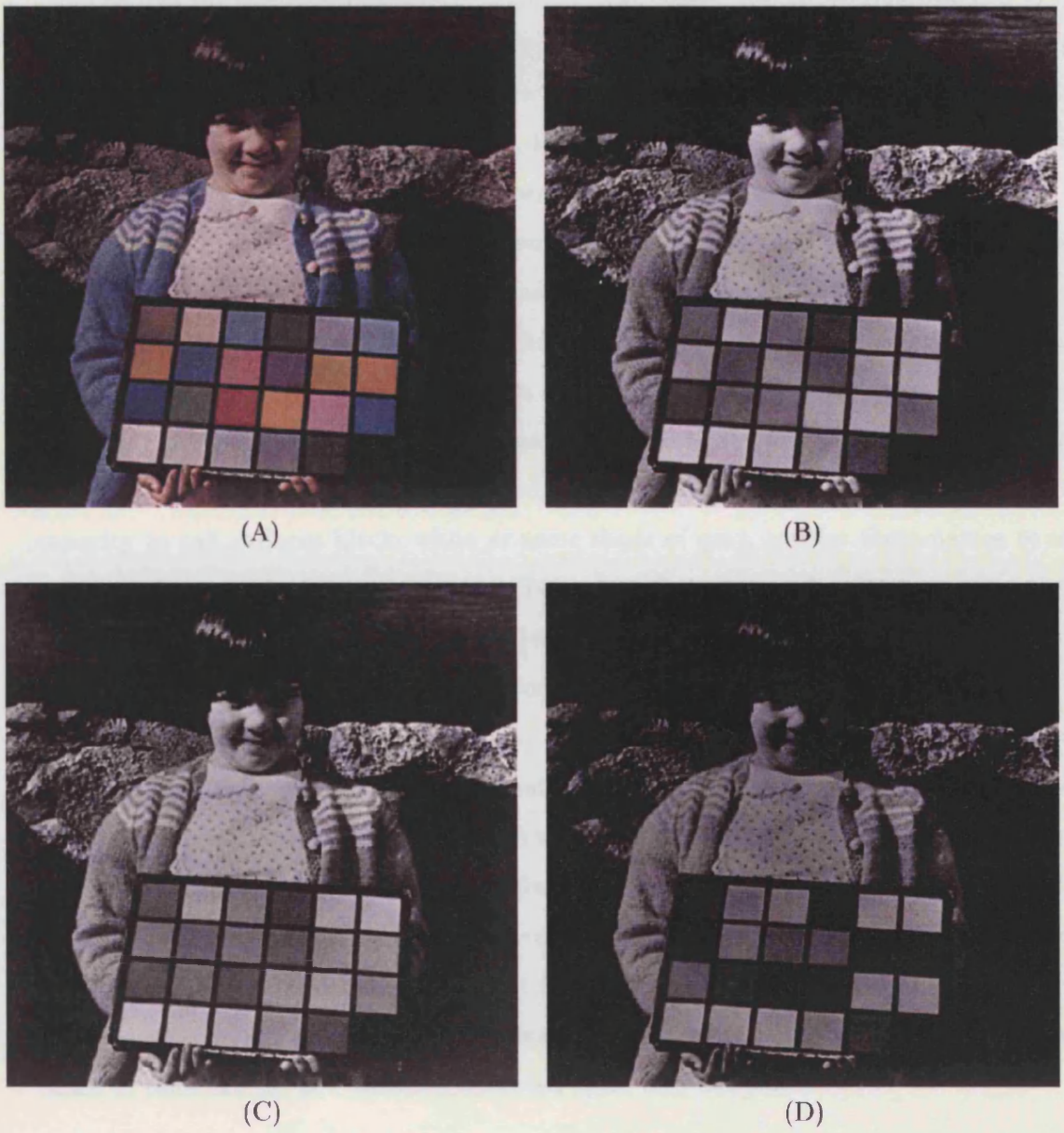
brighter, it acts to increase the divisor in Equation 4.5 and reduces the response to a constant luminance. In general, this effect causes a marked region of induced darkness to be generated around light areas and, by symmetry, a region of extra brightness on the light side of a step edge.

It should also be noticed that the model used here rolls two adaptation mechanisms into one, and these two mechanisms have different time courses. Adaptation due to pigment bleaching has a time course of the order of minutes whereas field adaptation (which is neurally mediated) takes 100–200ms (Bonds and Enroth-Cugell [25]; Shapley [211]). Pigment bleaching is primarily responsible for long lasting after-images. This mixing of two effects is a weakness of the current model.

When time is to be ignored, it is sometimes difficult to know what to simulate. In practice, a tonic ganglion cell (X-cell in the cat) fires a response burst after the stimulus onset and this settles down to a lower sustained firing level after about 100ms (Enroth-Cugell and Shapley [74]). Consider the case of viewing a grey screen on which an image is suddenly displayed. The initial burst is a reaction to the image content generated using the prior level of adaptation set by the grey field (the steeper greyscale in Figure 4-7), but it also incorporates a transient due to the temporal transfer function. The sustained response uses an entirely local adaptation intensity, providing that the eyes do not move (the shallower greyscale). If the eyes do move, the average response level will depend upon the extent of fixation and therefore the scalloped greyscale will result.

Figure 4-8 shows simulation results in the case of a natural colour image. The post-adaptation channel responses to this image are shown in Figures 4-8B–D for the red, green and blue receptor types respectively. It can be seen that there is not much difference between the red and green response maps. This is because the corresponding receptors have strongly overlapping spectral sensitivity curves. Subtle differences can be seen between the patches in the test card and in the child's face. The blue cone response is quite different, and it provides much of the contrast between the rightmost column of patches which are not differentiated by the red and green cones.

This section has described a non-linear model for determining receptor responses when given light levels and adaptation levels. There is evidence that this stage comprises the major non-linearity inherent in the first levels of sub-cortical processing and that the combined effect of later neural stages is to transform signals with a fair degree of linearity until the cortex is reached.



(A) (B)

(C) (D)

Figure 4-8: Receptor responses to a colour image. (A) Original 512 × 512 colour image. (B) Red channel response. (C) Green channel response. (D) Blue channel response. Local adaptation was used with $r_c = 1.5°$.

4.6 Contrast and Brightness

Here, we consider some of the implications of the EP model, with regard to brightness perception and the formation of contrast measures.

The real world is characterised by surfaces that reflect light in varying degrees. A typical scene will contain objects which diffusely reflect from 10% to 90% of incident light, giving a range of about one order of magnitude of light intensity when illuminated from a single source. Although this range is not large, levels of illumination can change by several orders of magnitude from scene to scene—indoors, outdoors, night and day.

It is generally recognised that the visual system is concerned with properties of objects and surfaces rather than with large inconsequential changes in illumination. Judgements of light and dark and contrast differences need therefore to be rooted in object reflectances, which remain constant, despite changes in overall light level. Contrast measures must encapsulate the difference between the reflectance of adjacent regions and remain invariant with illumination. Similarly, the vision system should show brightness constancy. This is the capacity to call surfaces black, white or some shade of grey, and for these names to remain constant across large changes in lighting, between different places in the same scene and with different viewing geometries (Cornsweet [48]).

Contrast and brightness constancy both hold to a large extent when the human vision system operates under natural conditions. Figure 4-9[5] shows three test spots, each presented on a surround together with associated intensity profiles. Figures 4-9A and 4-9B appear to have about the same contrast even though their luminances differ, but Figure 4-9C has a lower contrast despite the fact that the step luminance difference between test and background is the same as in Figure 4-9B. The luminance ratio (or log difference) between test and surround is the same for the first two figures and is smaller in Figure 4-9C, so this is one possible measure of contrast which rejects illumination levels—as the illumination changes, all the ratios of reflectances, and therefore reflected light, stay the same.

4.6.1 Contrast Measures

A commonly used contrast definition is Michelson contrast given by the formula

$$C = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = \frac{\Delta I}{2I_{ave}} \tag{4.10}$$

[5]Note that the accuracy of this figure is strongly dependent on the faithfulness of print reproduction.

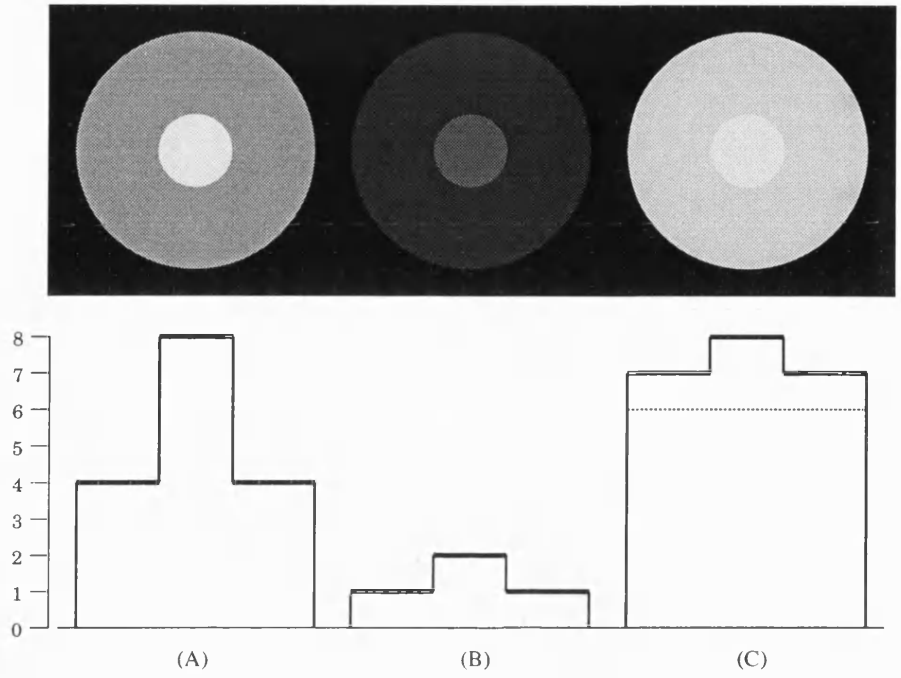


Figure 4-9: The ratio between the reflected luminances of the test spot and its surround should be the same in (A) and (B) and so they appear to have the same contrast even though the step luminance difference is larger in (A). In (C), the step difference is the same as in (B) but 6 units of light have been added to the figure. The apparent contrast is reduced.

yielding contrast values in the range 0–1. In the mammalian vision system, it is very likely that contrast is signalled by the arithmetic difference between neural responses at neighbouring retinal positions. We have already seen that the receptor response to light is approximately logarithmic about the adaptation point, so this difference represents a luminance ratio. I will now compare this measure with Michelson contrast.

If Equation 4.6 is substituted into Equation 4.5 and the adaptation luminance I_b is set to the average luminance I_{ave} at a contrasting boundary then the EP model response difference between spatial locations on either side of the boundary will be given by:

$$\Delta R = \frac{I_{max}}{I_{max} + \beta I_{ave}^{\alpha} + S_{min}} - \frac{I_{min}}{I_{min} + \beta I_{ave}^{\alpha} + S_{min}}. \tag{4.11}$$

Substituting $I_{max} = I_{ave}(1 + C)$ and $I_{min} = I_{ave}(1 - C)$ from Equation 4.10 we get

$$\Delta R = \frac{I_{ave}(1 + C)}{I_{ave}(1 + C) + \beta I_{ave}^{\alpha} + S_{min}} - \frac{I_{ave}(1 - C)}{I_{ave}(1 - C) + \beta I_{ave}^{\alpha} + S_{min}} \tag{4.12}$$

which gives response difference as a function of contrast, C. Plotting ΔR as a function of C

for a number of values of I_{ave} gives the graph of Figure 4-10. For values of C less than about 0.5, the difference response is almost linearly related to Michelson contrast.

The compressive power in Equation 4.6 means that the adaptation point does not linearly track the average scene intensity and this results in an increase in difference signal with overall illumination, up to about 1000 cd $\cdot$ m^{-2}. If this difference signal is taken to be a measure of human perceptual contrast then we find that the dependence of perceived contrast on average intensity should be small providing that field adaptation has occurred. This is shown by the fact that the line gradient only changes slowly with five orders of magnitude of average luminance. We will now see this demonstrated more explicitly.

The gradient of each contrast *vs* response line of Figure 4-10 at the origin can be obtained by differentiation of Equation 4.12 about $C = 0$. This gives a measure of contrast gain:

$$\Gamma(I_{ave}) = \frac{2I_{ave}(\beta I_{ave}^{\alpha} + S_{min})}{(I_{ave} + \beta I_{ave}^{\alpha} + S_{min})^2}. \tag{4.13}$$

When $\Gamma(I_{ave})$ is plotted against I_{ave}, the continuous line in Figure 4-11 is obtained. The main point to notice is that contrast gain changes slowly with adaptation luminance, and for luminances in the daylight range, the function is almost flat—contrast does not vary with luminance, and Weber's law holds. Purpura *et al.* [193] investigated the dependency of contrast gain on retinal illumination for ganglion cells in the macaque's retina. The data points in the figure show the actual change in gain for one **P**-type neuron. The shape and position of the theoretical curve is very close to this data set.[6]

4.6.2 Veiling Illumination

Most conventional image processing is carried out on raw pixel values but Kaushal [132] has suggested using a conversion to luminances followed by the Michelson contrast formula to yield a more perceptually realistic dependence on contrast for some applications. Kaushal also shows that when background light falls on a computer display, image contrast is reduced and he calls this *veiling illumination.* Its effect is to add to I_{ave} in Equation 4.10 and therefore to reduce the amount of contrast, especially when I_{ave} is already small. This is demonstrated by the difference between Figures 4-9B and 4-9C where a luminance pedestal has effectively been added. Kaushal specifically modifies the Michelson contrast formula to take this effect

[6] Purpura *et al.* [193] presented their results as contrast gain (in impulses per second per % contrast) plotted against retinal illumination measured in macaque photopic trolands with a 5mm pupil diameter. I have divided their scale by a factor of 1.4 to convert to human trolands (since the EP model parameters were obtained by psychophysical measurement) and a factor of 19.6 to convert to cd $\cdot$ m^{-2}.

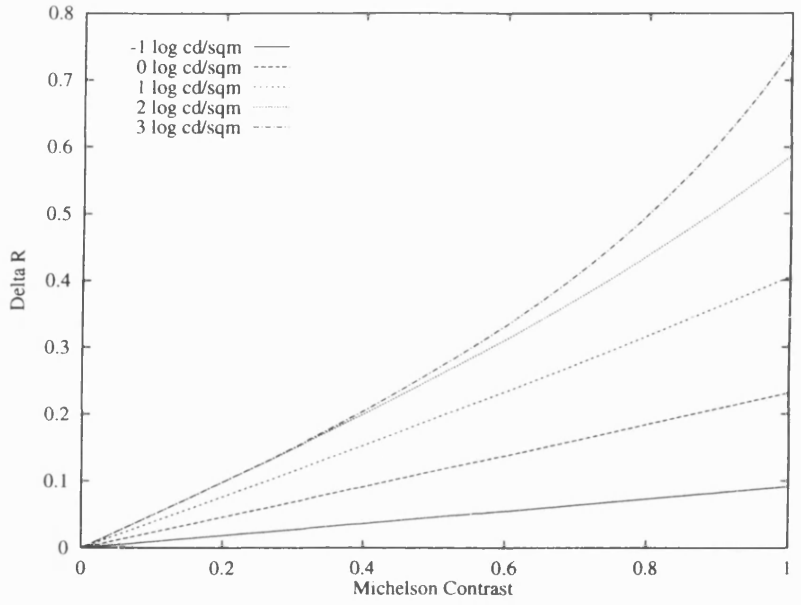


Figure 4-10: Contrast measure produced by differencing receptor signals (ΔR) plotted against Michelson contrast for a variety of adaptation points.

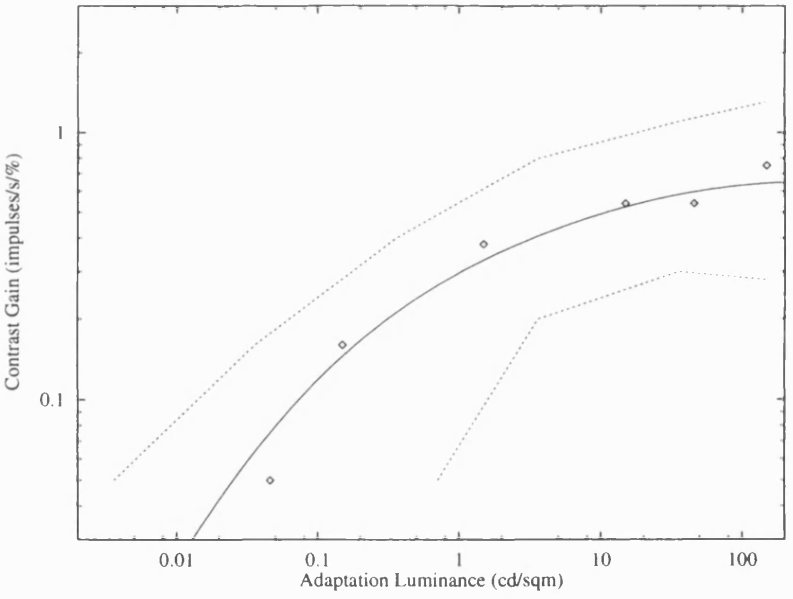


Figure 4-11: Plot of variation in contrast gain with adaptation luminance (continuous line, $R_{max} = 260$ impulses $\cdot s^{-1}$, S_{min}=0.8, $\alpha = 0.69$, $\beta = 5.83$). The contrast gain varies over less than one log unit with a 4 log unit variation in adaptation luminance. Dotted lines represent the limits of the distribution of contrast gain values for **P**-type macaque ganglion cells as determined by Purpura *et al.* [193]. Data points are for a representative neuron from the same source.

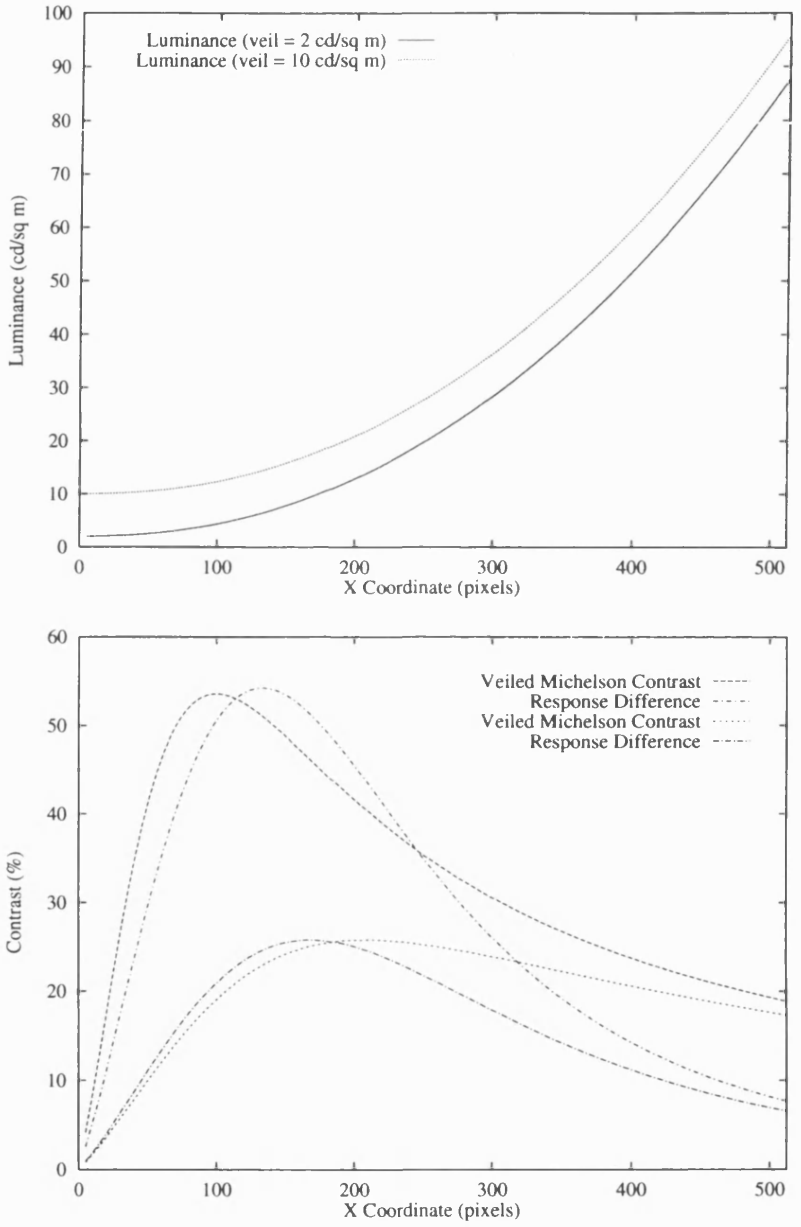


Figure 4-12: Incremental contrast along a greyscale ramp for two levels of background illumination (2 $cd \cdot m^{-2}$ and 10 $cd \cdot m^{-2}$). Figure A shows the luminance profiles for the two cases and B compares incremental contrasts predicted using Michelson and EP model difference formulas. Adaptation levels were set to the average luminance.

into account. Here the effect is allowed for by the inclusion of I_{back} in Equation 4.1.

Figure 4-12 demonstrates the effect of two levels of veiling illumination on Michelson contrast and on contrast computed from EP model differences. These graphs show the incremental contrast produced when moving up a greyscale ramp. Incremental contrast is the contrast between two adjacent points on the luminance curve. Both contrast measures show a similar response profile and reduction at low intensities (on the left of the graph).

4.7 Spatial Summation and Opponency

Two stages produce the well-known centre-surround receptive field layout: The first introduces spatial pooling of cone outputs over many small circular overlapping regions of the retina. In this way, a number of spatial units are formed which differ in the size of the area from which they pool receptor signals. I use the Gaussian receptive field model originally employed by Rodieck [196] to account for these summation properties. The Gaussian model describes the weight applied at each receptor-to-summing-neuron connection using the isotropic Gaussian function that I defined in Equation 4.7.

The second stage introduces spatial opponency. It has been known since Kuffler [140] and Hubel and Wiesel [112] that retinal ganglion cells in the cat and monkey subtract signals from differently-sized receptor pools to create a centre-surround receptive field arrangement (see Figure 2-5). For an on-centre cell, the neuron will fire a stream of action potentials if the receptors in the centre of its receptive field are stimulated by light, but if the surround is stimulated, what activity there is, will be reduced. Off-centre cells also exist for which the situation is reversed.

From Section 4.6, it is evident that this spatial arrangement could be used to calculate the contrast between centre and surround luminances. Although the optimum stimulus is a spot of light, it does not necessarily follow that this stage is a "spot detector" because the arrangement is what we might expect if the vision system needed to reduce its responsiveness to diffuse illumination. It is best to view this configuration as a spatial filter.

In the monkey retina, spatial pooling is often restricted to receptors of one type. For instance, all green cone activity within a region may be summed by the centre mechanism and all red cone activity by the surround (De Monasterio and Gouras [52]). Other mechanisms may sum together cone responses having different wavelength sensitivities to produce a signal that is sensitive to a wide spectral bandwidth. For the rest of this chapter, however, I ignore any dependency on stimulus wavelength and use a single luminance sensitive receptor type.

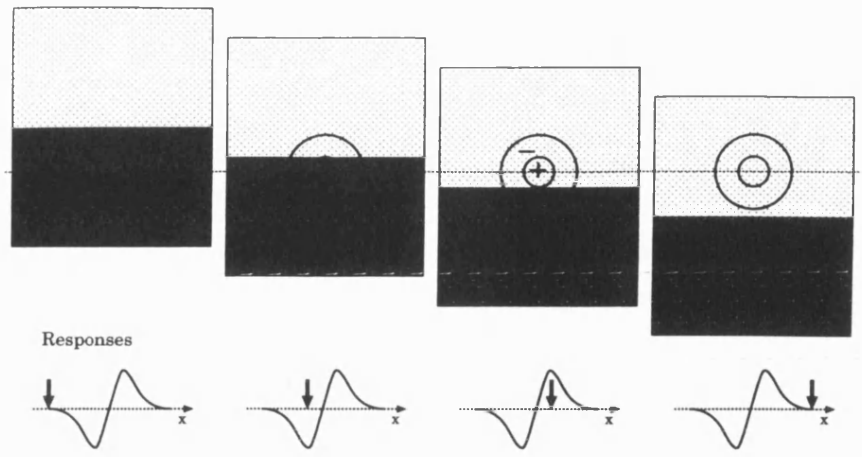


Figure 4-13: Testing a single on-centred neuron's response to a step edge. As the light half of the edge invades the receptive field it produces first suppression then excitation.

Figure 4-13 shows the kind of stimulation paradigm that is often used by experimenters when recording from single neurons. If the stimulus is a step edge and it is flashed at different positions across the receptive field then the response is first suppressed, then excited as the light-half invades the surround and then the centre. Finally, the light half will cover the whole field and produce a balanced response. Enroth-Cugell and Robson [72] carried out such an experiment using an edge stimulus.

Here I am interested in the responses from many neurons, each of which draws its input from a different spatial position. We therefore must change our mental reference point to accommodate a stationary stimulus pattern producing responses in a bank of adjacent neurons. (The receptive fields of these neurons overlap to a large extent because the spatial region that is analysed by each one is only slightly offset from that of its neighbour and this offset is small compared with the receptive field size.)

4.7.1 Simulation Results

Figure 4-14 shows the response of a collection of neurons to a light pattern consisting of a step edge and two lines (one light, one dark). Here there is one neuron for every receptor and receptive fields therefore overlap considerably. This graph is a cross-section through the neural response map produced using a two-dimensional simulation.

The peak-to-peak strength of the oscillation in the response trace is indicative of stimulus contrast. Also shown on this graph is a dotted line at -0.1. This is intended to represent a hypothetical response floor. There are often high rates of spontaneous activity present in the

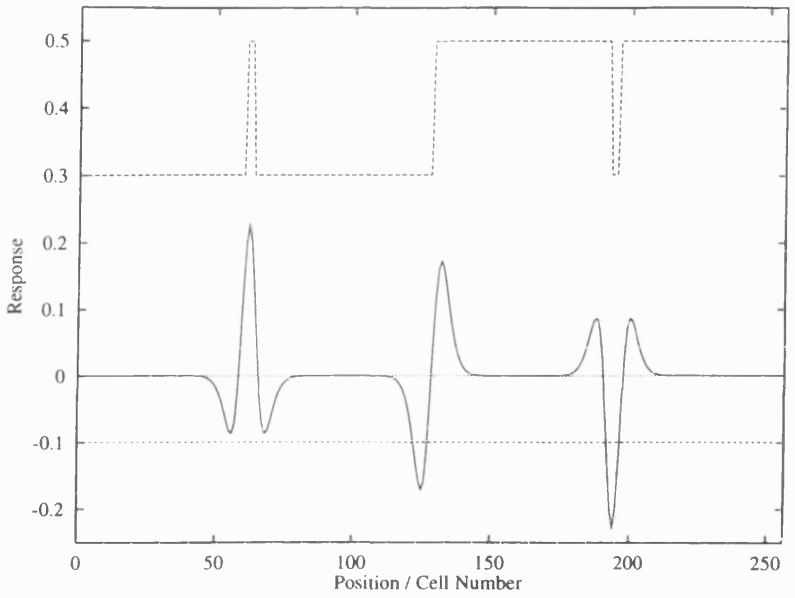


Figure 4-14: Responses (continuous line) from 256 cells having identical DC balanced on-centre receptive fields modelled using a difference-of-Gaussians, when stimulated with bar and edge patterns (dotted line). Each x-coordinate represents the response of one cell.

monkey optic nerve without stimulus (Hubel and Wiesel [112]) and this activity provides a DC level onto which excitation is added. The dotted line arbitrarily represents zero absolute activity and responses below this level cannot be signalled. The downward peaks to the right of Figure 4-14 must therefore be transmitted by off-centre cells for which this stimulus is excitatory.

The simulation shown here uses the difference-of-Gaussians (DOG) receptive field model which was originally employed by Enroth-Cugell and Robson [72] to fit the contrast sensitivity functions of retinal ganglion cells in the cat. It has proved to be a good model for spatial interactions in the retina and LGN of a variety of species. It has also been used to model psychophysical mechanisms underlying spatial vision in humans.

The receptive field centre and surround are both modelled using Gaussian spatial weighting functions and these act in opposition:

$$D(x,y) = k_c \exp\left(-\frac{x^2+y^2}{r_c^2}\right) - k_s \exp\left(-\frac{x^2+y^2}{r_s^2}\right). \tag{4.14}$$

Here, r_c and r_s are the $1/e$ radii of the centre and surround regions respectively, $k_c = S/(\pi r_c^2)$, $k_s = \mu S/(\pi r_s^2)$, S is the gain (sensitivity factor) and μ is the balance factor. When $\mu = 1$ then the receptive field is said to be balanced because it is not affected by diffuse illumination—

the centre and surround give equal and opposite signals as in the case of Figure 4-14. More realistically, retinal ganglion cells and cells in the lateral geniculate nucleus are found to have some response to diffuse light, giving us our ability to see subtle shading as well as sharp boundaries.

Equation 4.14 is used as a spatial weighting function to filter the output of the receptor model stage $r(x, y)$, Equation 4.8, giving the response

$$d(x, y) = D(x, y) \otimes r(x, y). \tag{4.15}$$

This results in neural response maps like the one shown in Figure 4-15 for the monkey's retina. Shown on the left is a test pattern and on the right is the neural activity map that is produced in response to it. In this response map, mid-grey represents a zero-level output (corresponding to maintained activity), bright responses are positions of excitation and dark regions correspond to neurons that are suppressed below the global level of maintained activity. Figure 4-16 shows two horizontal cross sections through this response map.

Note that here a ganglion cell response has be computed for each receptor position. In practice the receptive field centre size (r_c) may be large compared with the cone spacing, and therefore it may not be necessary to have a ganglion cell density that is this high. I will consider this issue shortly.

Figure 4-17 shows a response map for neurons in the cat's retina when the animal is presented with the same test image as Figure 4-15 under identical viewing conditions. The larger receptive field sizes possessed by the cat, when compared with the monkey, result in a much lower acuity and the out-of-focus look.

In considering these computed response maps, it is worth noting that such images are not for viewing! They indicate signals that are present *within* the visual pathway and are displayed as images only because this provides an effective way of visualising a large amount of response data. When looking at such response maps, we should keep in mind that they do not give rise to a perceptual response within the viewer that is related to the information that they represent to the simulated nervous system.

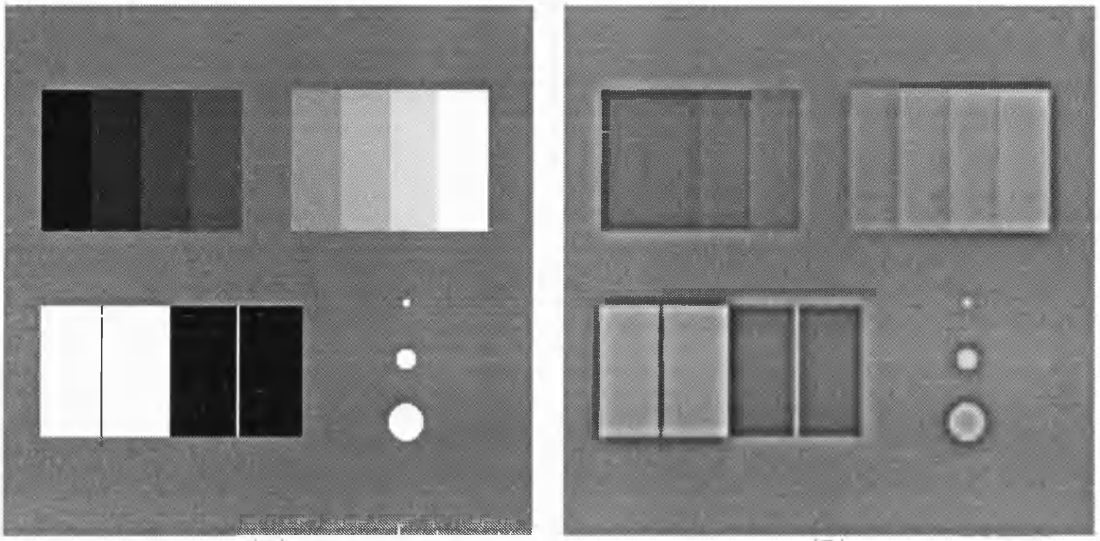


Figure 4-15: (A) Black and white test image with greyscale staircase. (B) Response map for neurons in the monkey's retina having an out-of-balance difference-of-Gaussians receptive field arrangement with 10% response to DC. This simulation is for an image subtending 5×5 degrees with $r_c = 0.025°$ and $r_s = 0.100°$.

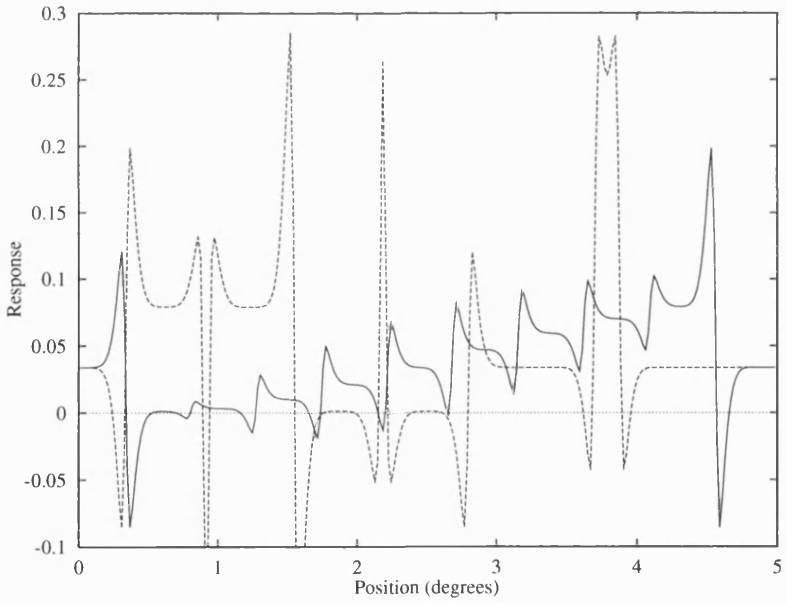


Figure 4-16: Horizontal cross-section through the the response map of Figure 4-15B at two different y-coordinates.

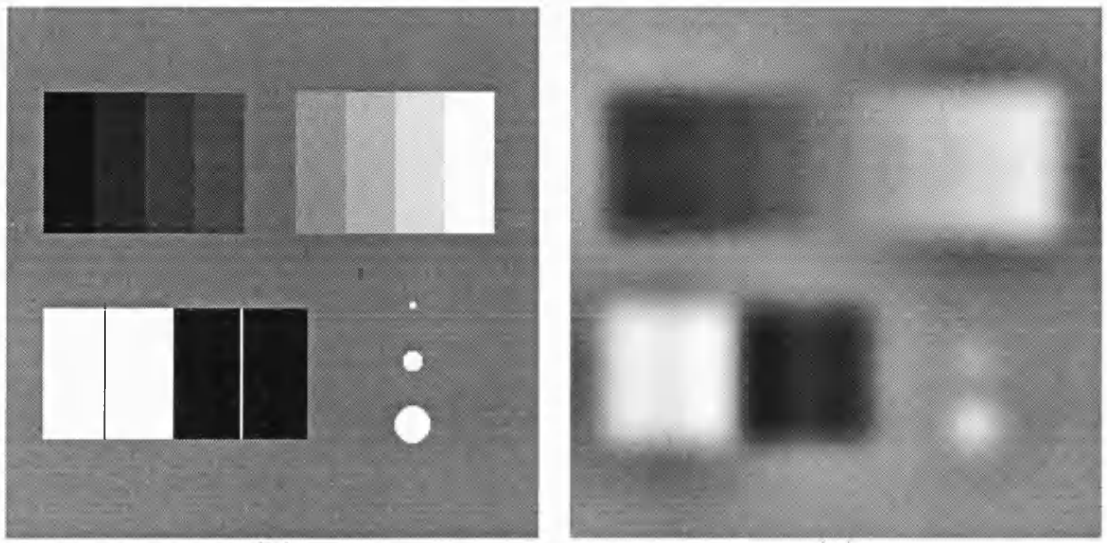


Figure 4-17: (A) Black and white test image. (B) Response map for neurons in the cat's retina having an out-of-balance difference-of-Gaussians receptive field arrangement. This simulation is for an image subtending 5×5 degrees with $r_c = 0.200°$ and $r_s = 0.800°$.

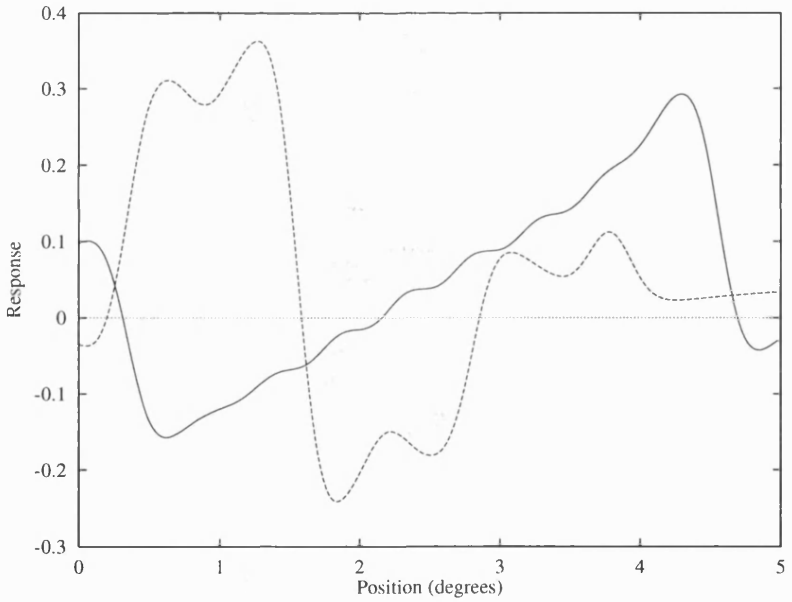


Figure 4-18: Horizontal cross-section through the the response map of Figure 4-17B at two different y-coordinates.

4.7.2 Frequency Domain Responses

A more useful form of Equation 4.14 is obtained by transforming it into the Fourier domain:

$$\mathbf{D}(\omega_x, \omega_y) = k_c \pi r_c^2 \exp\left(-\frac{r_c^2(\omega_x^2 + \omega_y^2)}{4}\right) - k_s \pi r_s^2 \exp\left(-\frac{r_s^2(\omega_x^2 + \omega_y^2)}{4}\right). \tag{4.16}$$

When x and y are measured in degrees then the spatial frequency variables, ω_x and ω_y, have units of radians per degree (rad/°).

Alternatively, it is possible to write $\omega_r^2 = (\omega_x^2 + \omega_y^2)$ and $f_r = 2\pi\omega_r$, to give a single radial spatial frequency variable expressed in cycles per degree (cyc/°). This can be done because the receptive field is circularly symmetric. This results in the more usual centre-surround equation (Enroth-Cugell and Robson [72]; Linsenmeier *et al.* [148]), which is

$$\mathbf{D}(f_r) = k_c \pi r_c^2 \exp(-r_c^2 \pi^2 f_r^2) - k_s \pi r_s^2 \exp(-r_s^2 \pi^2 f_r^2) \tag{4.17}$$

$$= S\left[\exp(-r_c^2 \pi^2 f_r^2) - \mu \exp(-r_s^2 \pi^2 f_r^2)\right]. \tag{4.18}$$

The response to diffuse illumination ($f_r = 0$ cyc/°) is zero when $\mu = 1$; the centre and surround are perfectly balanced. For a representative ganglion cell, μ is typically about 0.9 when sine wave gratings drifting at 1Hz are used as a stimulus. For 2Hz gratings, Linsenmeier *et al.* [148] found a value of around 0.8.

Figure 4-19 shows the normalised response predicted by Equation 4.18 when the stimulation consists of a sine wave grating that can range in spatial frequency from 0.1–26 cyc/°. The response is plotted on log-log axes. For small luminance differences and $\mu = 0.9$, this curve has the same shape as the contrast sensitivity function for a typical cell of this type.

The general effect of this filtering stage is to reduce the signal energy at low spatial frequencies. This means that more response dynamic range is given over to high frequency edge contrast information. Variations in r_c affect the horizontal position of the high frequency roll-off. A large value of r_s will make the response wide-band on the low frequency side.

4.7.3 Receptive Field Parameters, Channels and Eccentricity

We have already seen from Chapter 2 that receptor density reduces with eccentricity. Foveation is also a property of later stages of the neural hierarchy, in that all receptive field diameters increase with eccentricity. The overall size of a receptive field is not necessarily a measure of its capacity for resolving fine detail. The value of r_c affects the acuity of the ganglion cell

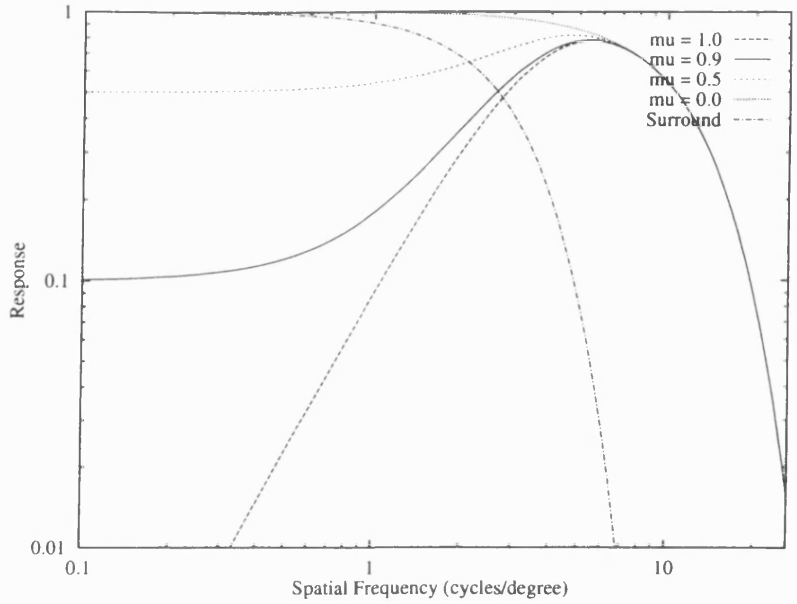


Figure 4-19: Spatial frequency response curve for a centre-surround receptive field with $r_c = 0.025°$ and $r_s = 0.100°$. The value of μ affects the DC response, changing the filter from low-pass to band-pass.

and small centre diameters result in high acuity. The surround size only selects the range of low frequencies that are attenuated. Early studies measured the entire receptive field size, but later experimenters concentrated on the way r_c changes with eccentricity.

Figure 4-20 shows that for both cat and monkey, the general trend is towards increase in r_c with eccentricity—and this is true for sustained and transient cell types. This figure shows regression lines that have been fitted through a set of distributions and it should be noted that there is always a random spread of centre sizes at each eccentricity. Since r_c sets the cell's acuity, spatial resolution reduces with eccentricity. This has been confirmed for the macaque by Crook *et al.* [50]. To set this point in context, it is important to realise that receptive field sizes increase far faster with eccentricity in higher visual areas (V2, V4) than they do in the retina. This suggests that there may be reductions in our ability to make spatial discriminations in the periphery that are not reflected in measurements of acuity made with sine wave gratings.

Linsenmeier *et al.* [148] and Cleland and Levick [46] report that the transient **Y**-cells of the cat generally have receptive field centres that are about 2.5 times as large as those of **X**-cells at the same eccentricity. In the central retina, **X**-cells have centre diameters of around 0.5° ($r_c \sim 0.160°$) or less and **Y**-cells about 1° ($r_c \sim 0.330°$). For each type the range of diameters is small (Lennie [143]).

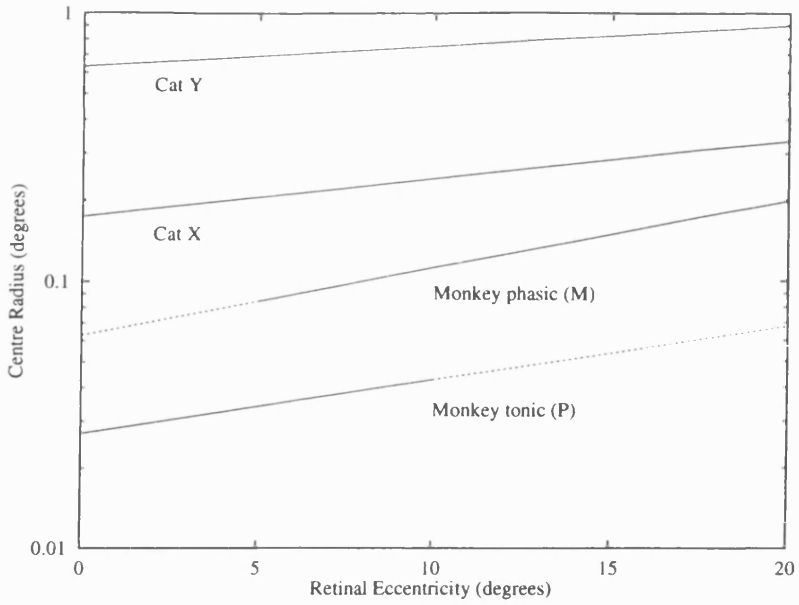


Figure 4-20: Variation of centre-size with eccentricity for ganglion cell receptive fields in cat and monkey. Regression lines for the cat are from Linsenmeier *et al.* [148]. For the monkey, lines were estimated by comparing the data of Hubel and Wiesel [112] (spider monkey retina), De Monasterio and Gouras [52] (rhesus retina), Crook *et al.* [50] (macaque retina) and Derrington and Lennie [64] (macaque LGN).

For the monkey, the situation is less clear and there may be species differences. For the rhesus monkey, De Monasterio and Gouras [52] report a clear distinction between centre sizes of phasic (**M**, transient, broad spectral bandwidth) and tonic (**P**, sustained, colour opponent) cells, with the transient cells having centres about 2.5 times larger, as is the case with the cat. Derrington and Lennie [64] report a similar separation when recording from the macaque LGN. Crook *et al.* [50] measured the visual resolution of macaque ganglion cells and found very little difference between the two resolution distributions. This is at least partly due to the large sensitivity difference between tonic and phasic cells. Since the phasic variety have a larger contrast sensitivity they can have larger centre sizes and still maintain the same threshold acuity.

A further point is implied by the dotted lines in Figure 4-20. Few **M**-cells are found near the centre of vision and the density of **P**-cells appears to reduce in the periphery (De Monasterio and Gouras [52]. **M**-cells have been identified with the motion pathway and the periphery is known to be more sensitive to motion that the fovea, so this is not entirely surprising. Similarly, **P**-cells are identified with sustained foveal colour vision.

Surround sizes have been measured for the cat. These vary slightly with distance from

the centre of vision, but show significant spread at any one eccentricity. Linsenmeier *et al.* [148] find that on average, the ratio of r_s to r_c is 4.0 for **X**-cells and 1.5 for **Y**-cells. They also find the centre-surround sensitivity balance μ to be about 0.8 as mentioned before, but this factor is generally smaller for **Y**-cells and reduces with eccentricity, increased temporal frequency and reduced light levels. Such changes are not taken into account here.

It is useful to compare the spatial vision of cat and monkey because complete data is rarely available for one species. As a rule of thumb, spatial parameters in cat vision are about eight times larger than those for the monkey and the contrast sensitivity curve is therefore shifted to frequencies that are a factor of eight lower.

The contrast sensitivity curves for individual ganglion cells, and cells in the LGN (which are similar) are much narrower than the human psychophysically-obtained contrast sensitivity function. Wilson and Bergen [247] accounted for the shape of the human contrast sensitivity function by using four DOG mechanisms or *channels*—the outputs of which were combined using probability summation. Wilson and Bergen's model includes two low spatial frequency channels with transient temporal responses and two higher-frequency channels with sustained responses. These may correspond to the phasic/tonic cell distinction, although one would expect to find less evidence for the "phasic" channels in central vision.

For the purposes of the simulation presented here, the eccentricity is assumed to be fixed at around 5°. The observations in this section suggest that it is valid to simulate mechanisms with values of r_c based around the regression lines at 5° in Figure 4-20. In practice, I use a set of up to four mechanisms with definable centre and surround sizes to carry **P**-cell information, but only one mechanism for **M**-cells (since they are only used in the retinal gain control simulation).

4.7.4 Spatial Sampling

Image subsampling is introduced when the output cell density is lower than the receptor density. The value of r_c sets the spatial frequency at which a neuron's contrast sensitivity function has its high frequency cut-off—and according to sampling theory, the spatial sampling rate need not be more than twice this frequency. This means that when receptive field centres are large, fewer ganglion cells are needed to cover the same retinal area. Since the Gaussian low-pass filter roll-off is not sharp, it is a matter of choice where to place the maximum frequency that the vision system must represent. If cells give no significant response to a 10 cyc/° grating at 10° eccentricity then 400 cells per square degree are required. Crook *et*

al. [50] mention that there are around 1000 cells per square degree at this eccentricity. This higher density figure may well reflect extra pathways, e.g. both on-centre and off centre cells are present. Wässle *et al.* [239] studied ganglion cell density and they found around 600 cells per square degree at the same eccentricity of 10°.

Since r_c increases with eccentricity, sampling rates, and therefore cell densities, are reduced away from central vision—there may even be some under-sampling in the periphery. Since rods and cones both drive the same output cells—and during dark-adaptation, rod receptive fields are found to be large—considerations of sampling theory alone are able to account for the overall 100:1 ratio of receptors to optic nerve fibres found in the primate eye.

In the simulation, subsampling of retinal ganglion cell responses is introduced where appropriate. This is described in Chapter 5.

4.7.5 Non-Linearities

So far I have identified neural contrast sensitivity functions with simulated spatial frequency response profiles assuming that the system is linear. It is important to notice that this need not be the case. To produce a contrast sensitivity function, researchers adjust the contrast of a stimulus so that the neuron being monitored gives a criterion signal—this signal is kept small so as to approximate threshold behaviour. Such a paradigm does not reveal non-linearities in the system and so the DOG model may not be valid for large signals. For instance, if the surround was divisive instead of subtractive then the shape of the contrast sensitivity function may be unchanged.

Various cancellation experiments have revealed that summation is present within the centre and within the surround of each ganglion cell receptive field. For high temporal frequencies, Kaplan and Shapley [131] and Derrington and Lennie [64] find a linear relationship between stimulus contrast and neural response for **P** ganglion cells in the retina and parvocellular neurons in the LGN. For transient cells and for low temporal frequencies the contrast response is more compressive. Although Enroth-Cugell and Robson [72] did not employ a receptor non-linearity, they were fairly successful at predicting the supra-threshold response to a light-dark boundary using an entirely linear DOG model (their figure 14). In addition, Campbell *et al.* [42] found that LGN fibres recorded in the cortex produce response and contrast sensitivity curves that are very similar to each other.

The model that I have adopted here includes the assumption that for retinal receptive fields, the system is linear beyond the photoreceptor stage. In particular, summation and

opponency can be implemented by using a linear convolution. Further discussion of this issue is presented in the conclusion to this chapter.

4.8 Contrast Gain Control

The diagram of Figure 4-1 includes a neural output transfer function. I now consider the output properties of ganglion cells and in doing so, I introduce a simulation of retinal contrast gain control acting among M-cells. This is done partly as an introduction to contrast gain control, since Chapter 6 details a rather complex form operating within the cortex, but it is in fact a significant result in itself.

Neurons in isolation have a limited response range—from about 0 to 150 impulses per second for retinal ganglion cells. As a stimulus drives the response rate to higher levels, the output begins to saturate and it becomes progressively harder to produce a response increment. For P-cells which have low contrast sensitivity, the effective output transfer function is almost linear because the contrast gain is low, and these neurons do not therefore produce an output level that significantly enters the saturating region of their transfer function.

For M-cells, and neurons in the magnocellular layers of the LGN which receive their afferents, contrast sensitivity is high and these neurons have a contrast *vs* response function that starts to saturate above about 20% contrast. If this output function was just a compressive non-linearity, acting directly on linearly-filtered responses, then the spatial waveform shape would be significantly distorted with large signals. In order to avoid this, without reducing the contrast sensitivity of the mechanism, a gain adjustment could be inserted before the output transfer stage. For low contrasts, the gain would be set to one, but for contrasts approaching the point at which the output would start to saturate, the gain should be reduced progressively in order to keep the response down in the linear region. The input-output relationship would then be kept linear, but some global contrast measure would have to be used to "throttle back" as saturation is approached.

In reality, when M-cells are stimulated with gratings they show responses to contrast that are far more compressive at low temporal frequencies than at high temporal frequencies. This behaviour implies that there is some signal path with a low-pass temporal response which acts to reduce M-cell gain as contrast increases at low temporal frequencies. Contrast gain control effects have been studied in the retina of the cat (Shapley and Victor [212], and monkey (Benardete *et al.* [18]), and also in the cat's lateral geniculate nucleus (Sclar [208]). Cortical contrast gain control effects have also been reported by Ohzawa *et al.* [184].

I now present results for a simulation of M-cell responses at low temporal frequencies (1–8Hz) when gain control is operating. Temporal factors implicated in this behaviour are not included in the model because I am interested in exploring the spatial effects of contrast gain control. Instead, image processing techniques are used to simulate spatial convergence and contrast compression at a fixed temporal frequency.

4.8.1 Gain Control Model

The ganglion cell output transfer function is modelled using a compressive transformation given by

$$R(x,y) = R_{max} \tanh[k\, d(x,y)]. \tag{4.19}$$

The excitation level introduced by $d(x,y)$, from Equation 4.15, is assumed to be measured in contrast units (%) by means of a suitable choice of k_c and k_s in Equation 4.14. To model a typical M-cell, I set $k = 0.05$ and $R_{max} = 120$ impulses per second. This results in an output function that has a linear region for $|d(x,y)|$ less than 10% and a maximum possible contrast gain of 6 impulses per second per % contrast (Purpura *et al.* [193]). Equation 4.19 is then modified so that the input signal is divided by a factor that adjusts the gain of the stage, yielding a final output response

$$R(x,y) = R_{max} \tanh\left[\frac{k\, d(x,y)}{\beta\, c(x,y)^{\gamma} + 1}\right] \tag{4.20}$$

where $\beta\, c(x,y)^{\gamma}$ is the gain control feedback signal and $c(x,y)$ is a contrast measure. When this signal is zero, the gain is at its maximum value of $R_{max}k$.

Where does the gain adjustment signal come from? I will investigate two possible alternatives: feed-forward and feedback control. For both of these arrangements, $c(x,y)$ is a measure of local contrast and it is derived from the responses of all neurons in a sizeable region centred on the receptive field that drives Equation 4.20. The difference between the alternatives is that for feed-forward control, this signal is made by combining activity from neurons upstream from the ganglion cell layer, whereas for feedback control, ganglion cell outputs are pooled and sent back to provide a recurrent inhibitory influence (Figure 4-21).

The contrast measure is obtained by adding together responses from on- and off-centre neurons. These responses are summed over an area slightly larger than the DOG receptive

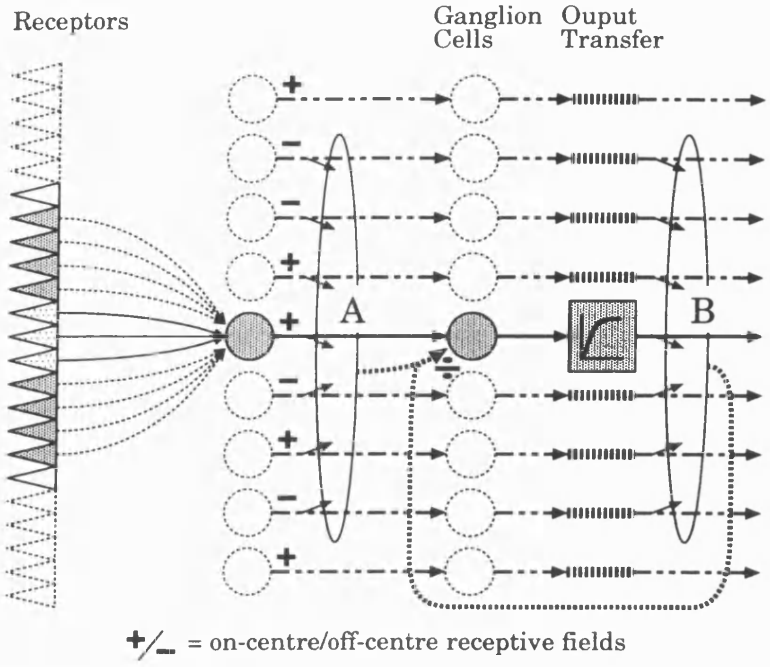


Figure 4-21: Retinal contrast gain control. Either: (A) response from intermediate layers in the retina are combined to provide contrast-dependent feed-forward inhibition, or (B) responses from on- and off-centred ganglion cells are pooled together and fed back to introduce divisive pre-synaptic inhibition.

field. For feed-forward control this gives

$$c(x,y) = |d(x,y)| \otimes \frac{1}{\pi r_g^2} \exp(-\frac{x^2+y^2}{r_g^2}) \tag{4.21}$$

and for the feedback case

$$c(x,y) = |R(x,y)| \otimes \frac{1}{\pi r_g^2} \exp(-\frac{x^2+y^2}{r_g^2}). \tag{4.22}$$

Note that taking the modulus introduces response rectification. Note also that I am assuming that the gain control summation region has a Gaussian weighting profile and employs linear additivity. Figure 4-22 summarises the theoretical processing arrangement for the two cases.

4.8.2 Simulation Results

The solid curve in Figure 4-23 shows the M-cell open-loop transfer relation that results when $\beta = 0$. This is just a rescaled tanh function and acts to compress the input, $d(x,y)$. Clearly, there is little departure from linearity for $|R(x,y)|$ less than about 90 spikes per second. For

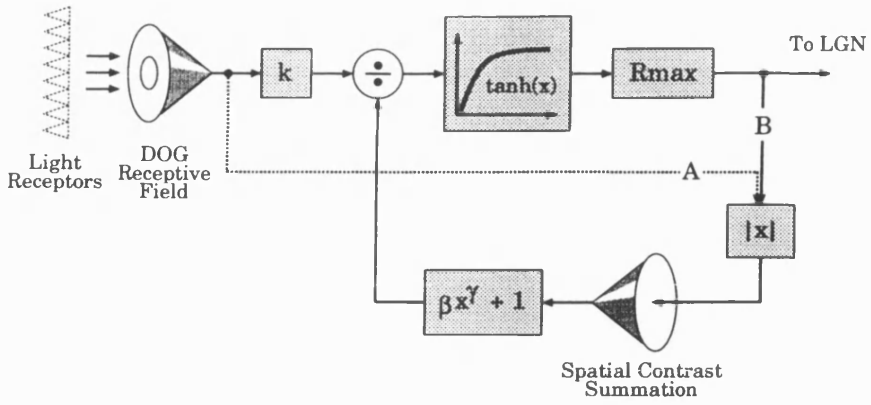


Figure 4-22: Theoretical model. Signals from light receptors are combined spatially to give a centre-surround receptive field with an output that is linearly proportional to boundary contrast. (A) and (B) indicate alternative feed-forward and feedback paths by which contrast information from a larger area is collected to adjust the gain of the ganglion cell stage.

P-cells the contrast sensitivity is ten times lower ($k = 0.005$), and so for these cells, the output can stay within the linear range for all contrasts.

When $\beta > 0$, **M**-cell gain is controlled by local contrast according to Equation 4.20. The important point to notice here is that the spatial summation in the gain control path ensures that the signal $c(x, y)$ changes only slowly with position. This means that we can treat $c(x, y)$ as being more or less constant over the extent of the $d(x, y)$ spatial response to any feature. The response of the **M**-cell is therefore linearly related to $d(x, y)$ if we avoid the tanh saturating region—no local distortion is introduced. From place to place, the signal $c(x, y)$ will vary, so the local gain will be turned up an down depending on how much contrast there is in each region.

Feed-Forward Case

For sine wave gratings that stimulate a large area, $c(x, y)$ is linearly related to grating contrast under feed-forward conditions. This means that as contrast is increased, $c(x, y)$ increases proportionally with the peak value of $d(x, y)$. The resulting peak value of $R(x, y)$ is therefore given by

$$R_{pk} = R_{max} \tanh[\frac{k\, d_{pk}}{\beta(A\, d_{pk})^{\gamma} + 1}] \tag{4.23}$$

where, for sine wave gratings, $A = 2/\pi$ because of the rectification and summation (Equation 4.21). Assuming that the output compressive region is not reached, we can substitute

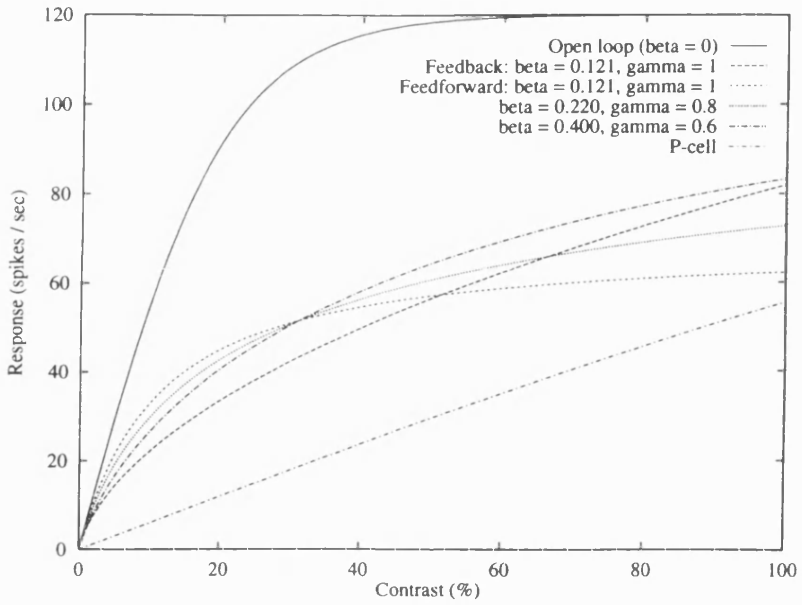


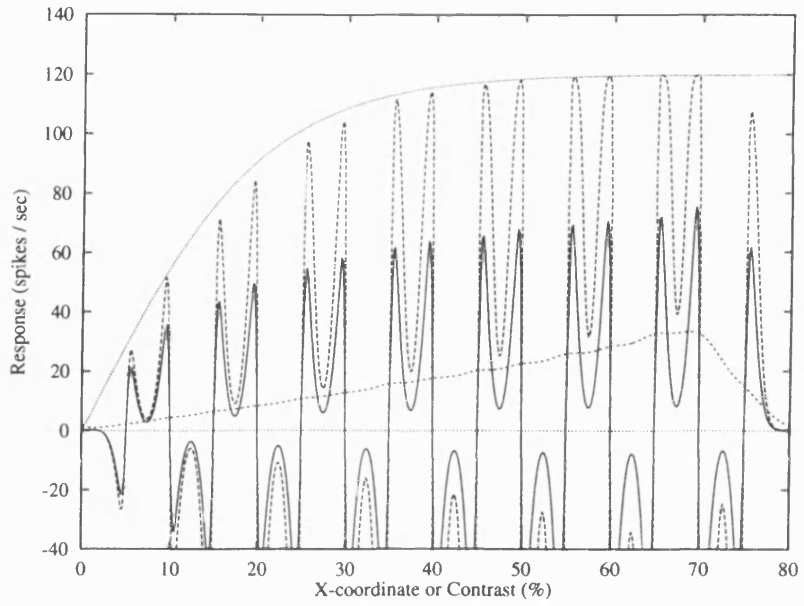
Figure 4-23: Excitation contrast *vs* response for an M-type retinal ganglion cell under open-loop conditions (solid line) and with contrast gain control (dotted lines). The solid line represents the transfer function of the cell, whereas the dotted lines show the level of activity at the response peaks given a sine wave grating input.

$\tanh(x) = x$ to give

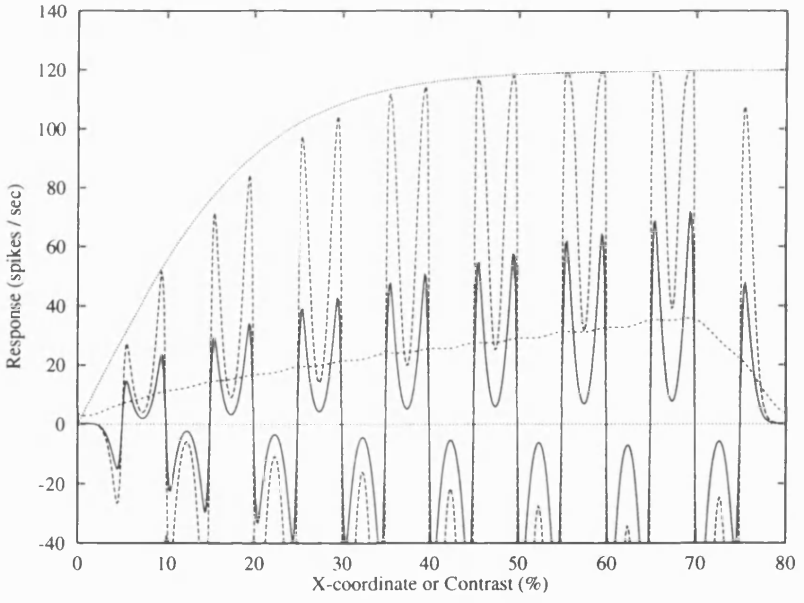
$$R_{pk} = R_{max} k \frac{d_{pk}}{\beta(A\, d_{pk})^{\gamma} + 1}. \tag{4.24}$$

For $\gamma = 1$ and $\beta = 0.121$ this equation can be rearranged to give $R_{pk} = 78\, d_{pk}/(d_{pk} + 13)$—an equation with the same form and half-saturation value as that used by Kaplan and Shapley [131] to fit a typical M-cell contrast curve. Figure 4-23 shows this curve, plus two others for different values of γ and β. The effect of decreasing γ is to stop the feed-forward control from balancing increases in stimulus contrast. The curve therefore develops a less saturating positive slope. Increasing β lowers the curve and reduces the half-saturation constant.

With $\beta = 0.121$, a grating stimulus cannot elicit a peak response that is greater than 78 impulses per second, and response modulations are therefore constrained to lie within the linear region of the neuron's output function. For non-grating stimuli, the value of A changes, and as a consequence so do the effective peak response and half-saturation constant—the neuron may still saturate if it sees a tiny white spot on a black background because A is very small for this stimulus. Large area grating stimuli are therefore not as effective as spots in producing a response, even after taking the DOG stage into account. As a result, the system displays a form of contrast surround suppression.



(A)



(B)

Figure 4-24: M-cell responses under conditions of feed-forward (A) and feedback (B) gain control (solid line). The stimulus is a square-wave grating increasing linearly in contrast from left to right. Also shown is the open-loop response (large, dashed) and the gain control signal, $c(x, y)$ (low ramp). The M-cell peak response saturates, but the signal is not distorted. ($r_c = 0.050°$, $r_s = 0.200°$, $\mu = 1$, $r_g = 0.400°$, $k = 0.05$, $\beta = 0.121$, $\gamma = 1$.)

Figure 4-24A shows the result of simulating M-cell behaviour when a square-wave grating stimulus is applied. This simulation was carried out in two dimensions of space, and a cross-section through the response map is shown. The grating steps linearly increase in contrast from left to right. Each x-coordinate on this figure represents the output from one neuron. Negative portions of the signal would be carried by off-centre neurons and positive portions by those with on-centres.

The outer plot which forms the envelope indicates the form of the tanh curve in Equation 4.19. Inside this is the open-loop signal which is distorted by the compressive transfer function. Note that at high contrasts the peaks begin to round, and the troughs do not stay in a fixed ratio with the peak heights. The low dotted curve indicates the rectified and smoothed contrast measure signal which increases from left to right. This control signal sets the gain for the forward transfer and produces the response shown with a solid line.

The rightmost grating step is at 40% contrast, and it is interesting to note that the height of this peak is less than the other 40% peak at $x = 40$. This is a surround-induced contrast effect which arises because the peak is next to a high contrast boundary (70%). The control signal is therefore larger than it would be if the right hand peak stood in isolation. This leads to a greater reduction in gain and a lower contrast sensitivity near high contrast boundaries.

Feedback Case

To simplify Equations 4.20 and 4.22 for the case of feedback, I assume that the system is effective in preventing saturation and therefore $\tanh(x) = x$ as before. Once again, the peak output values for a sine wave grating are considered:

$$R_{pk} = R_{max} \frac{k\, d_{pk}}{\beta (A\, R_{pk})^{\gamma} + 1}, \tag{4.25}$$

so therefore, when $\gamma = 1$ the solution becomes

$$R_{pk} = \frac{\sqrt{1 + 4\beta A R_{max} k\, d_{pk}} - 1}{2\beta A}. \tag{4.26}$$

The peak responses first rise linearly but soon compress with a power of 0.5 as contrast is increased (Figure 4-23).

Figure 4-24B shows the simulation results when feedback is in action. This simulation was achieved by first calculating $R(x, y)$ with $c(x, y) = 0$, then $c(x, y)$ was calculated followed by $R(x, y)$ and the process was repeated for 20 iterations, during which time the system

converged on a solution.

This feedback model behaves in a way that is not as simple as the feed-forward case because the forward saturation can have an effect on the signal $c(x, y)$. This gives scope for sub-additive non-linear summation in the feedback path when there are isolated high contrast features. For these features, A will change slightly with contrast because there is response compression prior to the rectify-and-sum stage of Equation 4.22.

Simulation Results

Figure 4-25 gives full two dimensional simulation results. Figure 4-25A shows a photograph of a bean flower and leaves which contains areas of high and low contrast. Figure 4-25B shows the neural response map for **P** retinal ganglion cells. These cells have a linear response to contrast, but much lower contrast sensitivity than **M**-cells. They also have a higher acuity, so r_c and r_s have been set to half the **M**-cell values, resulting in smaller receptive fields. The balance factor μ is made equal to 0.8, as in Section 4.7.2, in order to give these cells a significant sensitivity to DC brightness. White/black areas correspond to ± 120 impulses per second and mid-grey represents the zero level.

The **P**-cell map shows very little response to low contrast areas (e.g. in the dark regions) which were clearly visible on the original (this will depend upon the quality of reproduction). They also give a response which is subjectively similar to the original because of the linear contrast mapping.

Figure 4-25C shows the response that would result from M-cells without gain reduction ($\beta = 0$). There is a large amount of saturation present, but low contrast features are visible and are not affected by this. Simulated **M**-cell responses, with gain control in operation, are shown in Figure 4-25D. Saturation is no longer evident, but feature contrast is similar from place to place across the response map. Note also that the dark, low contrast regions do not change between Figures 4-25C and D.

4.8.3 Discussion

For this treatment of contrast gain control, I have not varied the size of the spatial summation region used to create the contrast measure; r_g is set so that this region is twice as large as each **M**-cell receptive field. If r_g is very large then isolated high-contrast features will easily saturate the output because they contribute little to the gain control signal. If r_g is too small, then the response no longer remains locally linear; compression is introduced and

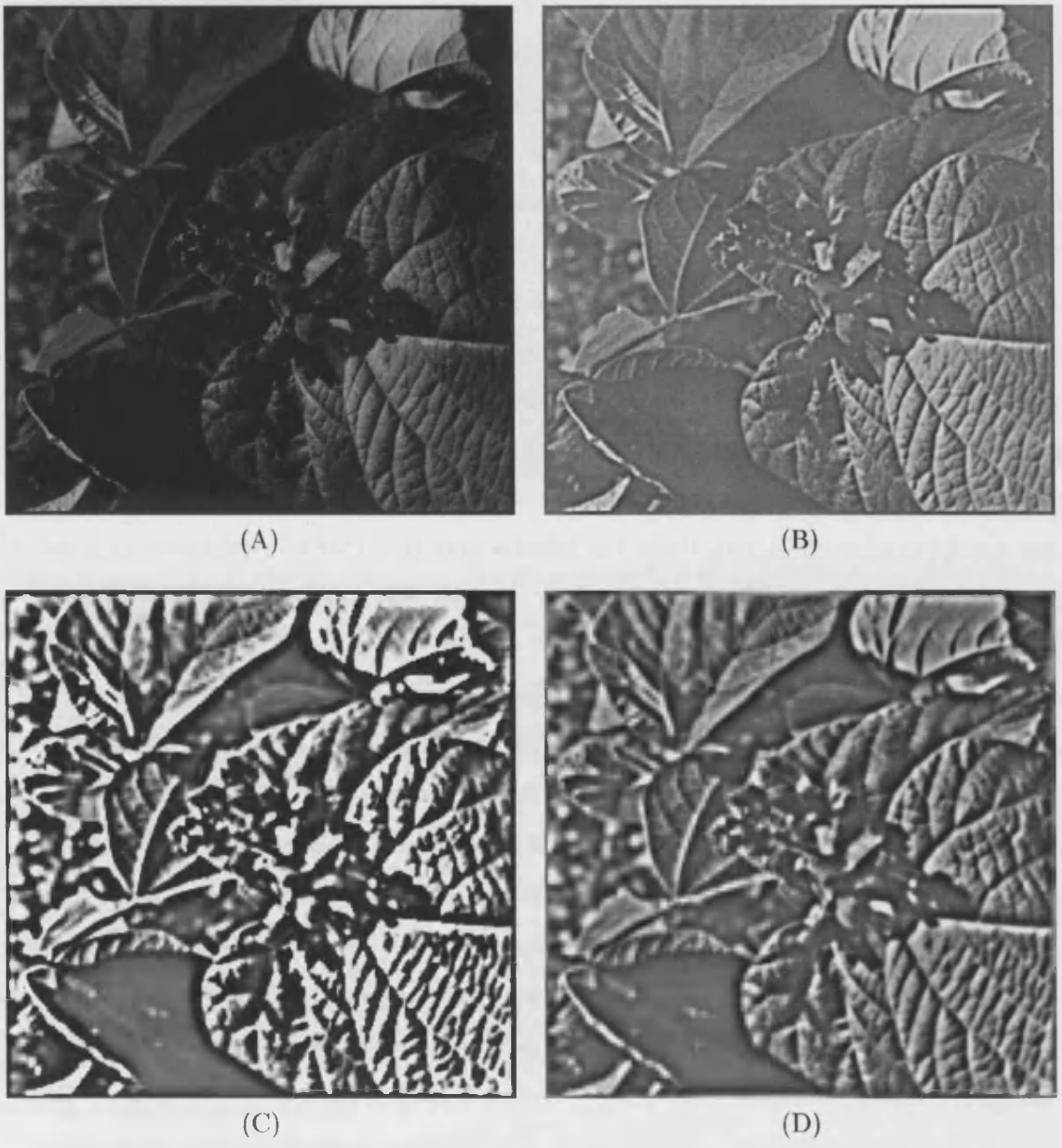


Figure 4-25: (A) Bean-flower image (512 × 512 pixels) which subtends a visual angle of 10° in the simulation. (B) P-cell response map. These cells have a higher acuity and show more sustained activity than M-cells, but have a much lower sensitivity to contrast. ($r_c = 0.025°$, $r_s = 0.100°$, $\mu = 0.8$, $k = 0.005$, $\beta = 0.$) Note that mid-grey represents the zero response level. (C) Saturated M-cell response when no gain reduction is applied. ($r_c = 0.050°$, $r_s = 0.200°$, $\mu = 1$, $k = 0.05$, $\beta = 0.$) (D) M-cell response when contrast gain control feedback limits the response to high border contrasts. The sensitivity to low contrast is preserved, without requiring a large response dynamic range. ($r_g = 0.400°$, $\beta = 0.121$, $\gamma = 1.$)

hence distortion in the response relationship between one neuron and its neighbour. I have therefore chosen r_g so that $c(x, y)$ is almost constant over the spatial extent of the DOG response to a boundary.

Receptive field centre-surround balance is an issue which has so far been avoided during this treatment. In Equation 4.15, the balance factor μ was set to unity for **M**-cells, and in Section 4.8.2 was set to 0.8 for **P**-cells. If **M**-cell receptive fields were not balanced, then the contrast measure formed after rectification would be contaminated by a response to the brightness of regions in the image.

It is known from Derrington and Lennie [64] that **M**-cells (in the LGN) have a non-zero response to low spatial frequencies or DC (diffuse light stimulation). However, this is for stimuli that are flashing or moving—and **M**-cell sensitivity to high temporal frequencies is very good. As described in Section 3.2.3, Kelly [133] found that, under normal viewing conditions, the human contrast sensitivity curve was similar to that obtained with stabilised vision at a velocity of 0.15°/s. At this velocity, a 5 cyc/° grating stimulates **M**-cell receptive fields at 0.75Hz, but a 0.2 cyc/° grating, at 0.03Hz—rather too low to elicit any response from these cells.

Therefore, for the purposes of this simulation, I have assumed that the combined effect of the spatial and temporal properties of **M**-cells is to produce a response as if their receptive fields were balanced. Under conditions of normal fixation, boundaries, not regions of light and dark, cause transient stimulation of **M**-cells, and therefore their response is primarily to contrast. Brightness perceptions are more likely to be transmitted to the brain by the sustained signals from out-of-balance **P**-cells.

Which of the feedback/feed-forward models is correct? Essentially, there is very little difference between the response maps that result from these models. The feed-forward map (not shown) is visually indistinguishable from Figure 4-25D. There is, however, slightly more evidence for the feed-forward model:

- Firstly, the contrast *vs* response relationship for the feed-forward case is very close to that observed by Purpura *et al.* [193], Kaplan and Shapley [131] and Derrington and Lennie [64]. The feedback model results in a relationship that is compressive at low contrasts and is not compressive enough at high contrasts.

- Secondly, one might question whether a feedback situation would be able to respond quickly enough in an application which is adapted to handling transients. It is true that the gain control signal appears to have a low-pass temporal response, but this is

not the same as a feedback delay. One might expect to see an initial response burst, followed by a tempering of activity as the feedback path cuts in. Such behaviour has not yet been reported.

- Lastly, Shapley and Victor [212] were able to hypothesise a gain control network that was feed-forward in conception, to account for their results for the cat.

Shapley and Victor's model considers the classical centre-surround receptive field to constitute a linear forward path filter, the gain and temporal phase shift being controlled by a contrast measure evaluated over the local region. This model is consistent with their experimental results, but is not very specific, and so I have simulated the spatial behaviour of such a network as if it operated by divisive inhibition, in the style of Wilson and Humanski's [248] model of gain control in the cortex. I suggest that evidence for divisive inhibition from a signal with a low-pass temporal response is provided by the results of Benardete *et al.* [18]: Contrast gain barely changes with contrast at high temporal frequencies, but (to a first approximation) reduces linearly with contrast at low temporal frequencies.

Where in the retina is the spatial contrast measure generated? Shapley and Victor [212] suggested that for the cat, **Y**-cell non-linear subunits are expressions of the same rectifying mechanism than mediates contrast gain control. Hochstein and Shapley [107] found that **Y**-cells were influenced by a pool of rectified responses over a region at least as large as the conventional receptive field. Summation over large regions involves lateral connections, and so for the primate, I wish to suggest that amacrine cells are good candidates for this role. They have transient responses and generally influence bipolar and ganglion cells over a large spatial region (Masland [158]). They are also in the right place: after spatial opponency and before the output layer. Amacrine cells were also discussed, among other possible candidates, by Hochstein and Shapley.

In summary, primate **M**-cells behave more linearly than one might expect from their contrast response function. The compression of this function appears to be strong when **M**-cells are tested with gratings because a grating stimulus excites the large gain control region which then reduces the local neural gain as contrast increases. Gain control keeps the input-output transfer function from saturation by setting the gain to an value appropriate to the local contrast conditions. High contrast sensitivity can therefore be preserved, even within a limited range of neural responses. **P**-cells, on the other hand, have a low contrast sensitivity and do not saturate under normal conditions. They do not need to be influenced by such a mechanism.

We therefore have two channels which are differently specialised: **P**-cells for acuity and sustained brightness vision, and **M**-cells for motion and low-contrast vision. These two streams are complementary, but not exclusive—**M**-cells are still useful at high contrasts. The results given in this section reinforce the idea that contrast gain control is a fundamental process by which **M**-cells maintain a high sensitivity and, simultaneously, avoid introducing local distortion of the space-domain waveform for high contrast stimuli.

4.9 Conclusions

In this chapter I have presented a six stage model which I have developed to transform image data into retinal response patterns. Within this framework, I have considered implications of the receptor adaptation model for brightness perception and forming contrast measures. In the next chapter the receptor model is related to chromatic receptive field properties. Here, results shown in Figure 4-11 demonstrate that such a model can usefully account for the change in neural contrast gain with retinal illumination.

In Section 4.7 I introduced the first level of spatial processing that we encounter in the vision system. There I presented simulations of spatial opponency for the cat and the monkey. I also considered frequency domain responses and contrast sensitivity curves for retinal neurons, and investigated receptive field sizes and spatial sampling so that realistic values could be assigned to parameters within the simulation.

Retinal contrast gain control was the subject of Section 4.8. I introduced a model of this process and provided simulation results that demonstrate that retinal **M**-cells carry a signal which is useful to general purpose spatial vision and does not become distorted and therefore unusable at high contrast levels.

4.9.1 Discussion

I now cover two issues that are relevant to the work at this stage. The first concerns the order of the light adaptation and spatial summation stages of Figure 4-1. Although adaptation was introduced as though it is a receptor-level phenomenon, there is some evidence that light adaptation may occur within the outer-plexiform layer as a complex, rather than just at isolated receptor sites. Problems arise because of this simplification:

Ricco's law (Cleland and Enroth-Cugell [45]; du Buf [30]) implies an equivalence between small test spots at high intensity and larger test spots at a lower intensity, whereas adaptation at the receptor level introduces saturation when followed by spatial summation.

Cleland and Enroth-Cugell [45] and Enroth-Cugell and Shapley [74, 73] found that for rod vision, field adaptation in the cat retina is a property of whole receptive field centres and not of individual receptors. The latter researchers proposed a model for rod vision in which light levels were summed linearly, and neurally mediated adaptive effects were introduced by feedback onto receptors from the summing junction (Enroth-Cugell and Shapley [74]). Later, rod cell bleaching adaptation was even found to show its effects at the scale of spatial summation units (Bonds and Enroth-Cugell [25]). Even though we are actually considering cone vision here, these discoveries still provide some justification for placing the "centres" spatial summation stage before any non-linear transform.

Unfortunately, experimentation with this arrangement produced rather unsuccessful results because of the non-linearities that are introduced by separating spatial summation and opponency with an log-like stage. Responses become highly compressive and rather unlike any published response traces. Since Enroth-Cugell and Robson [72] were fairly successful at predicting supra-threshold responses with an entirely linear DOG model, and since Campbell *et al.* [42] found an equivalence between response and contrast sensitivity curves, the current simulation scheme was retained. It is likely that the problem lies in thinking of the adaptation model as being present at a discrete stage, when it is more suited to describing the brightness adaptation properties of the retina as a whole.

The second issue relates to a neuron's responsiveness to tests on a large white adapting field. If adaptation mechanisms are entirely local, then cells should not show any difference in responsiveness when a large white surround is presented outside the receptive field. Valberg *et al.* [231] projected a coloured test stimulus and small white field in temporal alteration and recorded from the LGN. The field was intended to set the adaptation state of the retina and responses were found to relate to the field intensity as expected. When a large white surround was included around the test stimulus, this created a dramatic reduction in response gain. Glare or scattered light could be an experimental factor; alternatively, this kind of response could well be the result of fast acting field adaptation which is probably also responsible for simultaneous brightness effects (e.g. Figure 4-7), but the influence is very long-range and contradicts a local adaptation assumption. [7]

Li *et al.* [147] researched the role of the area outside the conventional receptive field on brightness transfer in cells at both retinal and LGN sites. Strangely, they found that the effect of increasing the size of a white stimulus far beyond that of the conventional receptive

[7] It is difficult to know to what extent such effects are introduced by interactions within the LGN.

field was to increase the responsiveness of these neurons as if their surrounds were being neutralised. This contradicts the report above since a large stimulus would be expected to introduce more light into the adaptation region and reduce response gain. The mechanism mediating these long-range adaptation effects remains unclear to me.

4.9.2 Function of the Retina

To conclude this chapter I give a summary of some properties of the retina that are of perceptual or engineering significance.

- The retina is foveated and provides good static resolving capacity in the centre of vision and good dynamic responsiveness towards the periphery.
- **M**and **P**-cell types represent an early stage of specialisation into structure and motion streams—concepts that have a clear high-level perceptual independence, even though both streams start off as properties of the same light pattern. In addition, **M**and **P**-cells serve different contrast ranges, but by virtue of the effect of retinal contrast gain control, **M**-cells are useful at both high and low contrasts.
- Rods and cones are specialised for dark and light conditions, respectively. The rod system is optimised for detection whereas the cone system is optimised for spatial resolution.
- On- and off-centre channels act as differential pairs coding complementary signals. This configuration can transmit visual signals along the optic nerve while rejecting common-mode noise and changes in resting response level.
- Chromatic opponency is introduced at the earliest possible stage in the vision system (see the next chapter). This is important because cone difference signals are small and might otherwise be swamped by noise. Chromatic opponency also introduces a coordinate system for colour perception.
- The centre-surround arrangement of receptive fields provides a spatial frequency emphasis that allocates response dynamic range over to medium resolution scene structure. This is achieved by band-pass spatial filtering and ensures that the strongest signal variations are associated with middle-range spatial frequencies. The next chapter explores this point in relation to spatio-chromatic interactions.

- Adaptation mechanisms reduce the effects on the vision system of large-scale fluctuations in absolute light levels. This results in a high degree of contrast and brightness constancy which is a first requirement for object-centred perceptual stability.

It is very difficult to place a specific representational significance on any response from the primate retina, as Barlow did with units in the frog's retina (Barlow [8, 9]). As the very front end of a general purpose vision system, this part of the brain is proving to be a rather complex and specialised structure.

Chapter 5

Through the Lateral Geniculate Nucleus

5.1 Introduction

In the last chapter, I was concerned with spatial response properties of retinal ganglion cells. This chapter covers the primate lateral geniculate nucleus (LGN), and in particular, I now concentrate on the responses of parvocellular neurons to chromatic stimulation. Neurons in the retina and in the LGN show similar chromatic behaviour, so the discussion here relates to both sites.

Initially, I describe the concept of colour opponency and I show how the suprathreshold response non-linearities which arise from the electrophysiological cone model that I use are consistent with previous findings. Next, I introduce the equations that are intended to model parvocellular and magnocellular receptive fields. Simulation results are included making use of these. Lastly, I present an analysis of the chrominance and luminance transfer properties of type I receptive fields in order to show how colour and brightness information can be decoded from their responses.

This chapter does not contain any material relating to integrative action in the LGN. Discussion of the role of cortical feedback into this area is deferred to Chapter 8. It is clear that there is also lateral inhibitory interaction between cells in the LGN (as noted in Chapter 2). There are a number of possible functions for this behaviour, which include: increasing of low spatial frequency attenuation (Hubel and Wiesel [113]), extension of large field brightness adaptation capabilities, and contrast gain control. Further research is needed

to clarify these modulatory effects.

5.2 Colour Opponency

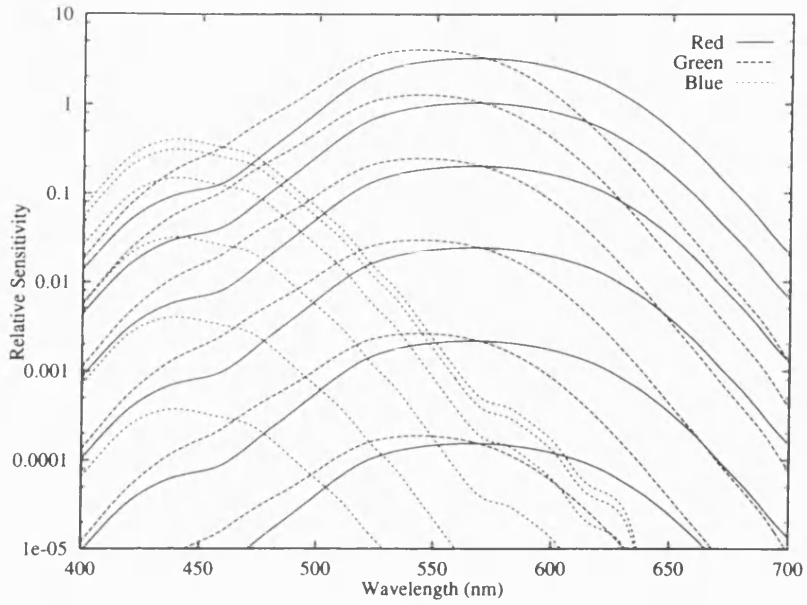
In this section I consider chromatic opponency in isolation from spatial receptive field properties. A very useful reference to psychophysical aspects of this topic can be found in Wyszecki and Stiles [250]. The aims here are to show how the adaptation model of Equation 4.5 affects the response *vs* intensity and response *vs* wavelength curves of retinal and geniculate **P**-cells and to introduce canonical chromatic mechanisms in order to prepare the ground for later results.

Many electrophysiological studies of the primate retina and LGN have shown that signals from cones having different spectral sensitivities are subtracted (De Valois *et al.* [56]; Wiesel and Hubel [246]; De Monasterio and Gouras [52]; Schiller and Colby [205]; Zrenner [257]; Zrenner and Gouras [258]; Derrington *et al.* [63]). This is believed to form the basis of colour vision because such differences are far more sensitive to chromatic than achromatic stimuli. In particular, two commonly observed interactions are between the red and green cones and between the blue cone and a combination of red and green types. We can view these interactions as the start of two colour pathways, one coding for the *redness* or *greenness* of a stimulus and the other, its *blueness* or *yellowness* in accordance with opponent-colours theory. These chromatic signals are both zero for "white" stimuli and therefore there exist neurons sensitive to light and dark in order to complete the perception of colour. In the retina, we see the first part of this process, but the full experience of colour cannot yet be explained in terms of neural processing, although threshold models have been advanced (De Valois and De Valois [58]).

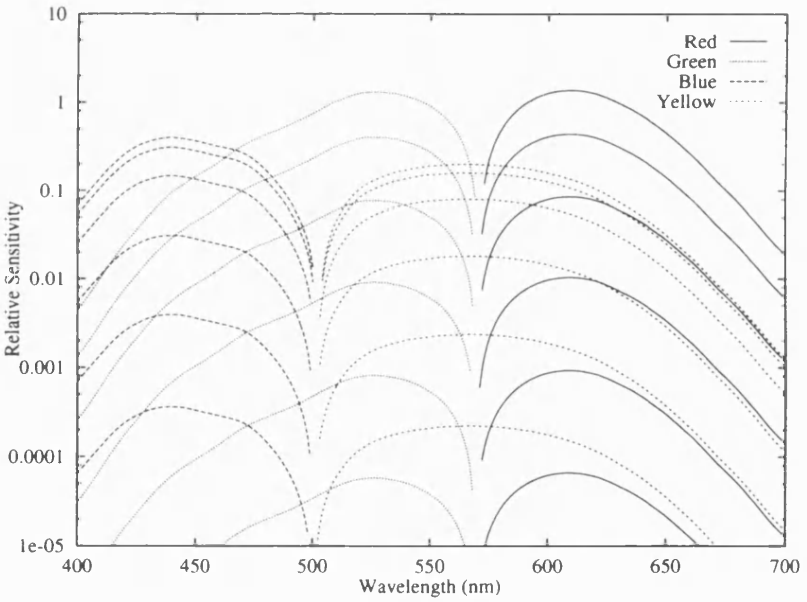
It has become common in the neurophysiological literature to present spectral sensitivities (e.g. Wiesel and Hubel [246]) rather than suprathreshold responses (e.g. De Valois *et al.* [56]). This is not always ideal because useful information concerning linearity is lost. Here I am interested in both threshold and suprathreshold behaviour. We start by considering a general three-cone interaction using the EP model:

$$R = A\frac{I\,\kappa_r r(\lambda)}{I\,\kappa_r r(\lambda)+\sigma_r} + B\frac{I\,\kappa_g g(\lambda)}{I\,\kappa_g g(\lambda)+\sigma_g} + C\frac{I\,\kappa_b b(\lambda)}{I\,\kappa_b b(\lambda)+\sigma_b}. \tag{5.1}$$

Here, A, B, C are the post-adaptation weights; $r(\lambda)$, $g(\lambda)$, $b(\lambda)$ are the normalised Smith and Pokorny [218, 219] cone spectral sensitivity functions (Figure 2-6); κ_r, κ_g, κ_b are the relative



(A)



(B)

Figure 5-1: Relative sensitivities for cone and opponent systems with different levels of white adaptation using the EP model. Uppermost curves are for the threshold condition, other curves (from top to bottom) are for log increases in background luminance from 0.1 to 1000 $cd \cdot m^{-2}$. (A) Red, green and blue cone systems ($\alpha = 0.69, \beta = 2.0, S_{min} = 0.25$, $\kappa_r = 0.8, \kappa_g = 1.0, \kappa_b = 0.1$). (B) Retina/LGN-level red-green and blue-yellow opponent systems ($\alpha = 0.69, \beta = 2.0, S_{min} = 0.25$; red-green: $\kappa_r = 0.8, \kappa_g = 1.0$; blue-yellow: $\kappa_b = 0.1, \kappa_r = 0.05$).

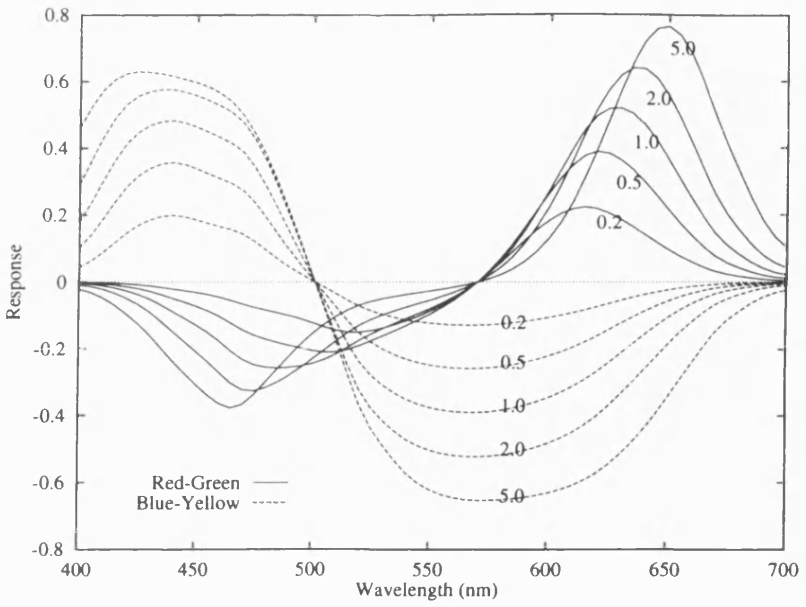


Figure 5-2: Theoretical suprathreshold red-green and blue-yellow opponent response curves for 200 cd $\cdot$ m^{-2} white adaptation with five fixed stimulus-to-background radiance ratios (red-green: $\kappa_r = 0.8, \kappa_g = 1.0$; blue-yellow: $\kappa_r = 0.05, \kappa_b = 0.1$).

pigment sensitivities, σ_r, σ_g, σ_b, are the cone adaptation points. I is the intensity (radiance) of the stimulus of wavelength λ. It is important to note that since the σ-values are properties of individual cone classes, these three adaptation points will be different under the conditions of white adaptation and, because of the nature of this equation, making a change to σ is equivalent to making a change to κ.

Figure 5-1A shows how the individual cone sensitivity curves change with the luminance of a white background. Note that the red and green cone sensitivity changes are almost in step. Note also that the blue cone increases in sensitivity relative to the other two cones as the background level is increased. Figure 5-1B shows the canonical red-green and blue-yellow opponent systems that I have used throughout my work.

5.2.1 Red-Green System

First let us consider a two-cone red-green opponent system which responds positively to red, negatively to green, and for white or yellow produces a zero result. In this case $C = 0$, and we also want $R = 0$ for all I at λ_{yellow} in Equation 5.1 to make a "unique hue" which is independent of stimulus intensity (providing that the adaptation points do not change relative to each other). At this unique hue, stimuli may appear blue, yellow or white, but will never appear to contain any red or green. To produce this null point, I find that the only

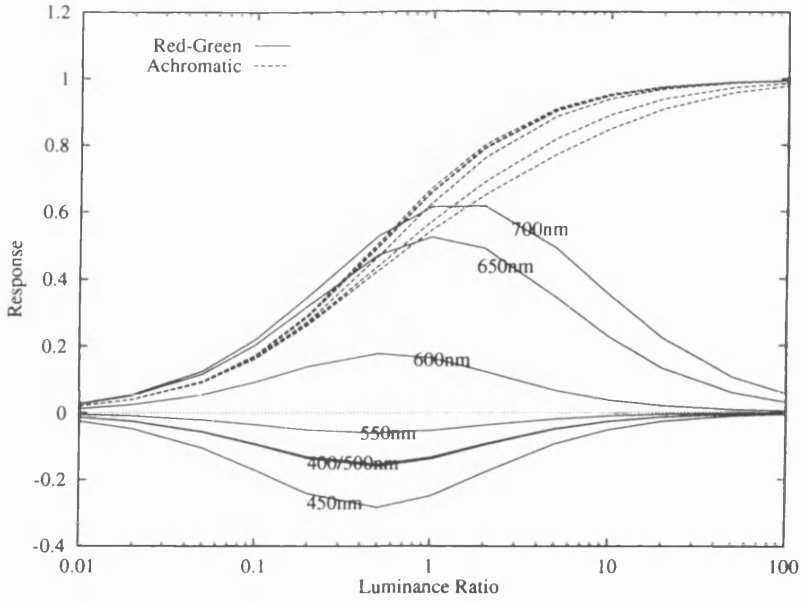


Figure 5-3: Red-green opponent and achromatic non-opponent suprathreshold responses to variation in stimulus luminance relative to a 200 $cd \cdot m^{-2}$ white background.

intensity independent solution occurs when $A = -B$. Figure 5-2 shows the suprathreshold red-green response *vs* wavelength plot for a few different radiances (expressed as ratios to that of a white adapting background).[1] These response curves are similar to those observed by De Valois *et al.* [56] in that they show a Bezold-Brücke-like hue shift (wavelengths to the right of the neutral point (yellow) of about 570nm become less red as stimulus intensity increases, and perceptually equal reds shift towards longer wavelengths).

Figure 5-3 shows a red-green response *vs* intensity graph for a number of different wavelengths. Intensity is expressed as luminance ratio between stimulus and white background.[2] A number of points arise from this plot:

1. These graphs compare favourably with the data obtained for the primate by Valberg *et al.* [231] and Lee and Virsu [142]. The peaks in red-green response coincide near a luminance ratio of one as Valberg *et al.* [231] found with a white surround that gives good field adaptation.

2. Providing that field adaptation has occurred, these plots predict that red-green opponent responses should be almost invariant with simultaneous changes in stimulus and

[1] Actually, these ratios are stimulus radiance divided by the radiance of a 555nm light giving the same luminance as the white background (200 $cd \cdot m^{-2}$).

[2] This treatment is very similar to that of Cornsweet [48] except that I am using a modified cone model which puts the result on a more quantitative basis and allows us to explore light adaptation.

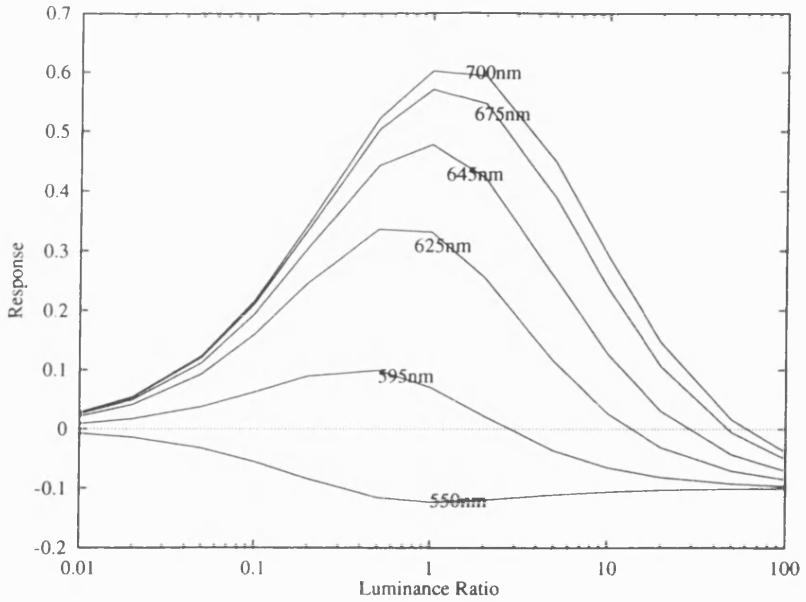


Figure 5-4: Mildly unbalanced red-green opponent responses to variation in stimulus luminance relative to a 120 cd $\cdot$ m^{-2} white background ($A = 1.0, B = 1.1$). This theoretical result is very similar to the recordings of Valberg *et al.* [231] (their Figure 3).

background luminance (the curves shift left and right along the intensity axis with adaptation, to maintain the peak at a luminance ratio of one).

3. For a fixed adaptation luminance, the red-green signal is almost constant with stimulus to background ratio over a one log unit range (because the curves have a maximum).

In conclusion, response level is independent of stimulus luminance[3] providing that it does not vary very far from that of the background. This means that, to a first approximation, chromatic saturation along the red-green direction is being coded for by this mechanism, and even though this is a very early stage in the vision system, we begin to see a split into separate streams that convey the perceptions of saturation and brightness. These curves predict that saturation should appear maximum for a chromatic stimulus with a luminance near that of its background.

Variations in Opponency

The majority of recordings from red-green opponent cells do not show balanced opponency, despite the fact that unique hues can be identified by psychophysical techniques. Here, I use

[3]We can talk of luminance here because red and green cones are adapted to an almost equal extent by white stimuli and have a similar spectral sensitivity curve to the luminance spectral sensitivity function, $V(\lambda)$.

the terms "out of balance" responses or "cone dominated" responses to refer to this case, for which it appears that $A \neq -B$ in Equation 5.1. The response signal from one cone is much stronger than the other, and such cells often show "hidden" opponency that can be uncovered by a strongly adapting chromatic background which weakens the dominant cone's influence (Zrenner and Gouras [258]). In these circumstances, the single-wavelength stimulus that yields zero response (if it exists) will vary with intensity and adaptation. Figure 5-4 shows theoretical responses, calculated using Equation 5.1, for such a cell which receives slightly excess input from green cones. It is clear that some wavelengths can be either excitatory or inhibitory depending on intensity (e.g. 595nm in the figure).

Since the majority of red-green **P**-cells are in this class, showing random cone balance, it would seem difficult to understand how useful chromatic information is carried. This question will be answered in later sections which deal with the way in which luminance and chrominance information are combined in responses of this type.

Adaptation

Figure 5-1A shows the cone spectral sensitivities under different levels of white-adaptation. For a balanced two-cone red-green system, the position of the crosspoint is invariant with stimulus intensity, but not invariant with adaptation intensity.[4] On Figure 5-1A, the point of intersection of the red and green cone sensitivity curves actually shifts horizontally very slightly with background level. This is because the cones have different initial sensitivities and therefore adapt at different rates down the figure. This shift is so small as to be insignificant. The position of the crosspoint is therefore fixed by the sensitivity ratio of the two cone photopigments and the colour of the adapting stimulus.

5.2.2 Blue-Yellow System

To model a two-cone blue-yellow system, I put $A = -1$, $B = 0$ and $C = 1$ in Equation 5.1. Figure 5-1B shows the threshold, and Figure 5-2, the suprathreshold response curves that result. The blue cone system is thought to be much less sensitive than the other two at threshold and this is probably due to the small number of blue cones and pre-retinal absorption of short wavelengths (Wyszecki and Stiles [250]). This presents a problem when blue responses are to be combined with those of red and green cones.

[4] Except, of course, for adaptation with a light equivalent to the crosspoint wavelength, which adapts both cones equally.

We have already seen that for a two-cone system we can always make a cross-point that is stable with intensity providing that white adaptation does not change the relative receptor sensitivities. Unfortunately, Figure 5-1A shows that the blue cone increases in sensitivity relative to the other types under these conditions. This is because white adaptation has a greater effect in depressing the responses of the more sensitive cones. Inspection of the points of intersection of the red cone and blue cone sensitivity curves show that these shift from 475nm to 505nm down the figure, and this would cause a corresponding change in neutral point if a blue-yellow opponent system were formed directly.

Zrenner and Gouras [258] found that ganglion cells with blue-yellow opponency were always well-balanced, with cross-points near 500nm. Others have found that the blue-yellow system is very insensitive to white light regardless of intensity (Wiesel and Hubel [246]; Derrington and Lennie [64]). This fact points to an attenuation of the long-wavelength cone input to this system at lower intensities, otherwise white stimuli would provoke a significant response. Experimentation with the value of A for this system yields no useful results because suprathreshold responses are badly affected. I have therefore assumed that somehow the pre-adaptation sensitivity constant κ_r is made small for input to the blue-yellow system, implying two separate post-receptoral adaptation mechanisms for long wave cones.

Figure 5-1B includes this modification. From the figure, it can be seen that the sensitivity of the blue-yellow system increases dramatically with adaptation level when compared to that of the red-green system. This is in line with the propositions of Guth *et al.* [93].

5.2.3 Three Cone Opponency

Some models of the early stages of opponency include three cone interactions (e.g. De Valois and De Valois [58]). Unfortunately, there is no set of values of A, B, and C in Equation 5.1 that allows a conventional opponent system drawing from three cones to have a cross-point that is invariant with I.[5] This means that, using the EP model, it is not possible to produce two stable spectral cross-points from one mechanism in a single stage of interaction. In particular, this precludes a retinal mechanism with both unique blue and unique yellow.

Experimentation with different values of A, B and C shows that departures from intensity invariance are tolerable for some three-receptor combinations. For instance, including the sum of green cone and red cone input into a blue-yellow system does not strongly affect the stability of the 500nm cross-point. For the purpose of the simulation presented here, however,

[5]There is one solution but it does not give a useful opponent system.

three cone interactions are avoided.

5.2.4 Luminance System

Various cells in the retina receive non-opponent input from one, two or three cone types. This results in a broad spectral sensitivity function which is likely to form the basis for brightness perception. These broad-band responses usually arise from **M**-cells which project to the magnocellular LGN layers. **P**-cells can also have broad spectral bandwidth, but most show some opponent influence.

In the past, luminance was a term used to describe the perceptual brightness of a stimulus. Since the CIE's 1924 definition of the luminance spectral sensitivity function, $V(\lambda)$, various sensitivity curves have been found as different methods of psychophysical measurement are employed. Lennie [146] gives a review of these methods. The conventional way of determining $V(\lambda)$ is to use heterochromatic flicker photometry—a high temporal frequency method which is likely to favour responses from **M**-cells.

The standard $V(\lambda)$ function is closely approximated by a combination of red and green cone sensitivities in an approximate ratio of 2:1 and there is thought to be little input from blue cones (Eisner and MacLeod [71]). This ratio reflects the relative densities of these receptor types in the retina (Sekiguchi *et al.* [210]. For this reason, I model a luminance-responsive **M**-cell using $A = 2$, $B = 1$ and $C = 0$ in Equation 5.1. For 200 $\mathrm{cd \cdot m^{-2}}$ white adaptation, this results in a red to green sensitivity ratio of 1.6378:1 with $\kappa_r = 0.8$ and $\kappa_g = 1.0$. This is very close to the ratio of 1.6330:1 used by Guth *et al.* [93].

Figure 5-3 shows how the luminance system response changes with luminance ratio. The wavelength splitting of this group of curves at high luminance ratios arises from the 2:1 weighting being applied at the post-adaptation site—the effect produces a subtle change the shape of the suprathreshold response curve. I would suggest that this is not very significant from a perceptual point of view.

Various researchers have pointed out that the **P**-cell red-green system may be capable of carrying luminance information by virtue of its receptive field geometry and out-of-balance responses (Ingling and Martinez-Uriegas [122]; Ingling and Martinez [120]; De Valois and De Valois [58]). This is undoubtably so, but I leave this topic to later sections.

5.2.5 Discussion

The modelling of cone opponency is complicated by two possible sites of sensitivity weighting. The pre-adaptation sensitivity factors, κ_r, κ_g and κ_b are probably set by absorbtion properties of the receptor pigmentation and ocular medium, whereas the post-adaptation sensitives, A, B and C, are most likely a function of neural connectivity. Difficulties arise because the EP cone model is intended to include neurally mediated adaptation, and so factors like receptor densities could actually affect the weighting at either of these two sites. This is particularly relevant to the blue-yellow opponent signal because unbalanced post-adaptoral weighting is a disaster for cross-point stability.

Here, I have assumed that neural mechanisms have access to the pre-adaptive sensitivities and can, at least, reduce them in the case of the blue-yellow system. Mechanisms receiving blue cone input have caused many problems already for experimenters—for example Stiles found three blue-mediated field sensitivities, including the elusive π_2 mechanism (Wyszecki and Stiles [250]; Thornton and Pugh [224]), Ikeda and Ayama [118] found non-linearities in the yellow chromatic valence, and unique red is known to be unstable with intensity—so there are clearly more complexities to be resolved.

5.3 Receptive Field Simulations

Now that spatial and chromatic opponency have been introduced, I shall examine the way that these properties are expressed in the receptive fields of LGN cells. This leads to a presentation of simulation results.

5.3.1 Types of Receptive Field

Figure 5-5 gives a summary of the most common LGN receptive field types. Here, the classification system introduced by Wiesel and Hubel [246] will be adopted. The following discussion relates to both retinal and LGN neurons.

Type I cells form the majority in central vision (De Monasterio and Gouras [52]). These have centres that are driven by one cone type and surrounds that are driven by a cones of a different type. In this way, they manifest both spatial and chromatic opponency. The majority of type I cells are red-green opponent. Type I cells with any contribution from blue cones are not very common. Both on-centre and off-centre varieties are present, as shown by the $+/-$ signs in the figure. (On-centres cause the cell to be excited by light, whereas

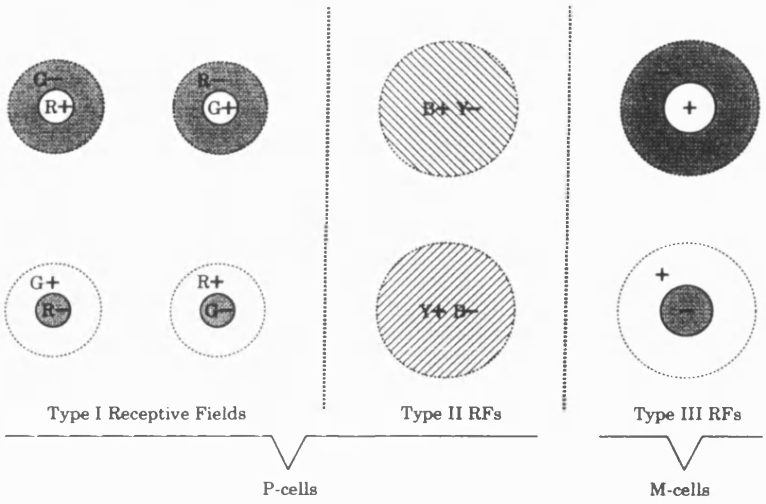


Figure 5-5: Spatial arrangement of common receptive field types in the primate retina and lateral geniculate nucleus. R, G, B: red, green, blue cone inputs, Y: "yellow" input from red and/or green cones. Signs indicate excitation or inhibition. Type III cells have a variety of cone inputs.

off-centres cause suppression when stimulated, and this suppression often results in rebound excitation, or *off-responses*, when the stimulus is removed (Hubel and Wiesel [112]).)

Type II cells have weak or non-existent surrounds. They do not show a spatially concentric arrangement, but have large centres which are driven by two or more cone types in opposition. This produces a receptive field with a single colour-opponent centre. The majority of this class of cells receive excitatory input from short-wave (blue) cones and have large receptive fields which are 3–5 times the diameter of those of type I neurons. Blue cones therefore contribute only to low frequency cortical spatial mechanisms (Humanski and Wilson [117]). Blue-yellow type II cells are much less common than red-green type I cells, and are also found towards the centre of vision.

Type III cells increase in number with eccentricity. These have concentrically organised receptive fields which generally show a wide-band spectral sensitivity, although random opponency is sometimes evident. The receptive field centres are about 2–3 times larger than those of type I cells.

Retinal cells in the first two categories (**P**-cells) project to the parvocellular layers of the lateral geniculate nucleus. They have sustained (tonic) temporal responses (similar to **X**-cells in the cat) and low contrast sensitivity. Their response to contrast is linear and does not saturate at high contrasts, as we have already seen in Chapter 4. Together, these cells form the major colour and form pathway to the cortex (Livingstone and Hubel [149]), and their

loss severely reduces contrast sensitivity for chromatic gratings as well as achromatic gratings at low spatial and temporal frequencies (Merigan [164]).

Retinal M-cells project to type III cells in the magnocellular LGN layers and have transient (phasic) temporal responses. They have high contrast sensitivity but their responses saturate at high contrasts (Kaplan and Shapley [131]; Purpura *et al.* [193]). Type III cells are generally thought to mediate colour-insensitive form and motion vision, especially in the periphery.

Type I neurons often show wide variations in opponent balance (Zrenner and Gouras [258]). Many are red-cone or green-cone dominated, so that their spectral neutral points can vary over a wide range either side of 560nm, reflecting an almost continuous gradation in cone weight. These neurons are most often dominated by the centre mechanism, meaning that the surround has a lower sensitivity, as we have already seen in Chapter 4. In the centre of vision, spectral-opponency is more likely to be balanced between red and green cone inputs, whereas further out there is an increase in cone dominance, especially for neurons having red centres. Off-centred neurons are less likely to show balanced opponency. In addition, retinal type I cells can often show intrusion from the blue cone system (De Monasterio and Gouras [52]) whereas this appears to be less evident at the LGN (Derrington *et al.* [63]).

Blue-yellow type II cells show little variation in cone balance when recording from sites at the retina (Zrenner and Gouras [258]) or the LGN (Derrington *et al.* [63]). Most have a neutral point at about 500nm and show very little response to white light (De Monasterio and Gouras [52]; Wiesel and Hubel [246]; De Valois *et al.* [56]). Neurons in this class that are excited by yellow rather than blue light are rare.[6]

De Monasterio and Gouras [52] found an extra class of cells in the retina which responded very transiently to light, had broad-band spectral sensitivities and very large non-concentric receptive field centres (up to 2° diameter). Since there have been no cells in this category observed at the LGN, I expect that these project to the superior colliculus, which is responsible for reflex responses outside of the main stream of visual awareness.

5.3.2 Simulation Equations

In order to carry out a practical simulation, some simplification is required so that essential mechanisms can be encapsulated in the model, without being overwhelmed by detail. The

[6]This most likely because yellow is signalled by red or green cone excitation, and since there are so many red/green cone driven neurons, it would make more sense to transmit information about blue cone excitation since response range is limited for inhibitory signals.

major simplification here is that I only model neural behaviour for one eccentricity. In addition, canonical forms of receptive fields are defined according to Figure 5-5 so as to limit the number of channels.

In Figure 5-5, the top row shows on-centre and the bottom row, off-centre receptive fields (for types I and III). Since retinal cells often have high levels of spontaneous activity, and, to a first approximation, on-centre and off-centre receptive fields carry complementary parts of the same signal, their combined responses are modelled using signed floating point arithmetic and linear two-dimensional space-domain convolution. Red- and green-centred concentric colour opponent type I, non-concentric blue-yellow opponent type II and concentric broad-spectral-band type III receptive fields are included since these form the majority types at 5° eccentricity, especially in the LGN.

Type I Neurons

Up to four channels of Type I receptive fields are provided for in the model. This allows the range of receptive field parameters that might realistically exist at one eccentricity to be approximated by four discrete mechanisms. The simulation software allows centre radius, surround radius, sensitivity and balance ratio to be independently set for each of these channels. This means that up to eight channels analyse the scene at once (including both centre cone types). The centre response of each mechanism is always weighted more strongly than the surround. This simulation strategy does not make explicit possible asymmetries in the numbers of cells dominated by each receptor type, or possible input from blue cones.

A type I red-centred receptive field is simulated by convolving the red cone map with a small diameter Gaussian and convolving the green cone map with a larger Gaussian kernel.[7] The surround contribution will cancel The resultants are subtracted giving the response:

$$d_{RG}(x,y) = S\,G_c(x,y) \otimes r(x,y) - \mu S\,G_s(x,y) \otimes g(x,y) \tag{5.2}$$

where $G_c(x,y)$ and $G_s(x,y)$ are the centre and surround Gaussian kernels in the form of Equation 4.7 with r_c and r_s as the centre/surround space constants, and $r(x,y)$ and $g(x,y)$ represent the red and green post-adaptation cone response maps. For a green centred receptive field (d_{GR}), the cone maps are swapped. Note that these cone responses are calculated

[7]Some experimentors have suggested that the surround may receive contributions from cones of all three types. De Valois and De Valois [58] include this possibility in their colour model. This change is not a problem for the cancellation scheme presented in this chapter. At high spatial frequencies the surrounds do not contribute, and at low frequencies they cancel. This leaves a pure $r - g$ signal in both cases.

with $\kappa_r = 0.8$ and $\kappa_g = 1.0$ so that if $\mu = 1$ then the balanced red-green spectral sensitivity shown in Figure 5-1B results, and the cross-point is near 570nm. Reducing the value of μ makes the system centre-cone dominated and the opponent cross-point shifts accordingly.

Type II Neurons

For type II receptive fields, the situation is simpler. Only one type II blue-yellow channel is included because a single centre produces a low-pass spatial frequency response and there is little advantage in introducing extra simulation channels at this level. Values from the red and blue cone maps are subtracted and the resultant is convolved with a Gaussian kernel to give the response:

$$d_{BY}(x,y) = S\,G(x,y) \otimes [b(x,y) - r(x,y)]. \tag{5.3}$$

In this case, the cone maps are generated with $\kappa_b = 0.1$ and $\kappa_r = 0.05$. Of course, for this linear model, it makes no difference whether the opponency stage is before, or after, the spatial summation.

Type III Neurons

Type III receptive fields are simulated by first summing $r(x,y)$ and $g(x,y)$ in a 2:1 ratio, and then convolving the result with the DOG filter of Equation 4.14 giving:

$$d_3(x,y) = D(x,y) \otimes [2r(x,y) + g(x,y)] \tag{5.4}$$

where $k_c = S/(\pi r_c^2)$ and $k_s = \mu S/(\pi r_s^2)$. The cone maps are calculated with $\kappa_r = 0.8$ and $\kappa_g = 1.0$. The balance factor μ is set to unity to give perfectly balanced fields. The reasons for this were discussed in Section 3.2.3. In addition, this response is subject to retinal contrast gain control as detailed in Section 4.8.

Receptive field size parameters for type I, II and III mechanisms are given in Appendix B.

5.3.3 Sub-Sampling

All the responses generated were subject to spatial subsampling. This provision compresses the size of the response map down to the minimum that is required to represent faithfully the information, according to sampling theory, as discussed in Chapter 3.

The sampling rate for each mechanism was chosen by inspecting the high frequency roll-off produced by the respective Gaussian filter (receptive field centre). The maximum resolvable

spatial frequency was chosen to be the frequency at which the centre response was 2 log units down from its peak at DC. The minimum sampling frequency is therefore twice this, according to the Nyquist criterion. The cone sampling frequency throughout this work was 51.2 cones per degree, as described in Chapter 3, and so the sub-sampling factor was calculated as the ratio between this and the ganglion cell sampling frequency measured in cells per degree. All sub-sampling factors were reduced to the nearest power of two for programming convenience.[8]

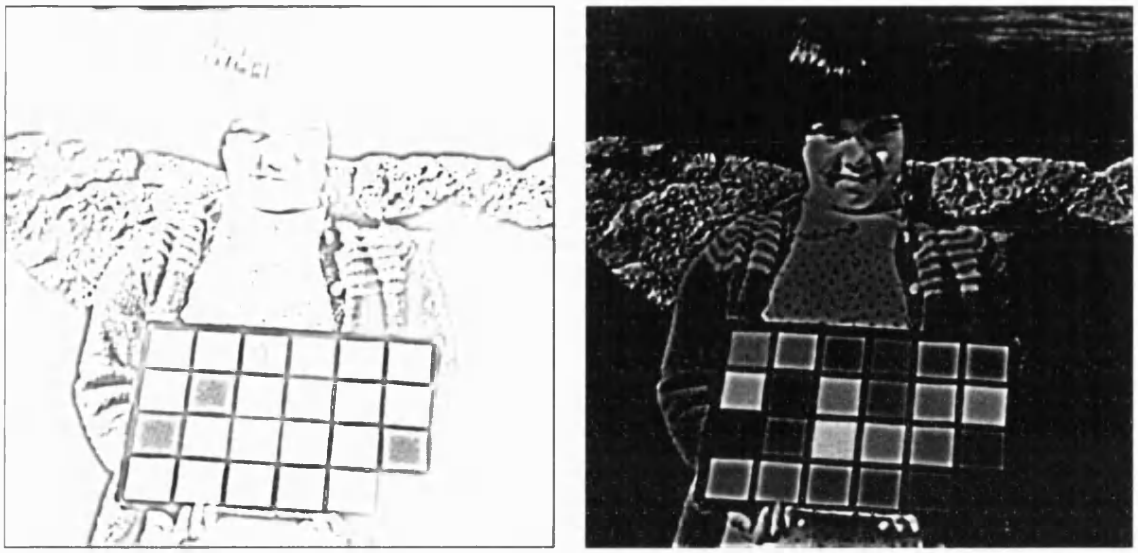
5.3.4 Simulation Results

Figures 5-6 and 5-7 show neural response maps that result when the simulated vision system is stimulated with the colour test image of Figure 4-8A. The first set of images shows responses of type I cells, and for this figure, on-centre and off-centre responses have been separated to show the complementary signals that are carried by these two neural sub-populations. Essentially, the positive parts of the signals d_{RG} and d_{GR} are carried by on-centre cells and the negative parts by off-centre cells. In Figure 5-7A and B, these responses have been combined to give a linear (non-rectified) response. This figure represents positive excursions of the signal by pixels lighter than mid grey, and negative portions by pixels which are darker than mid grey.

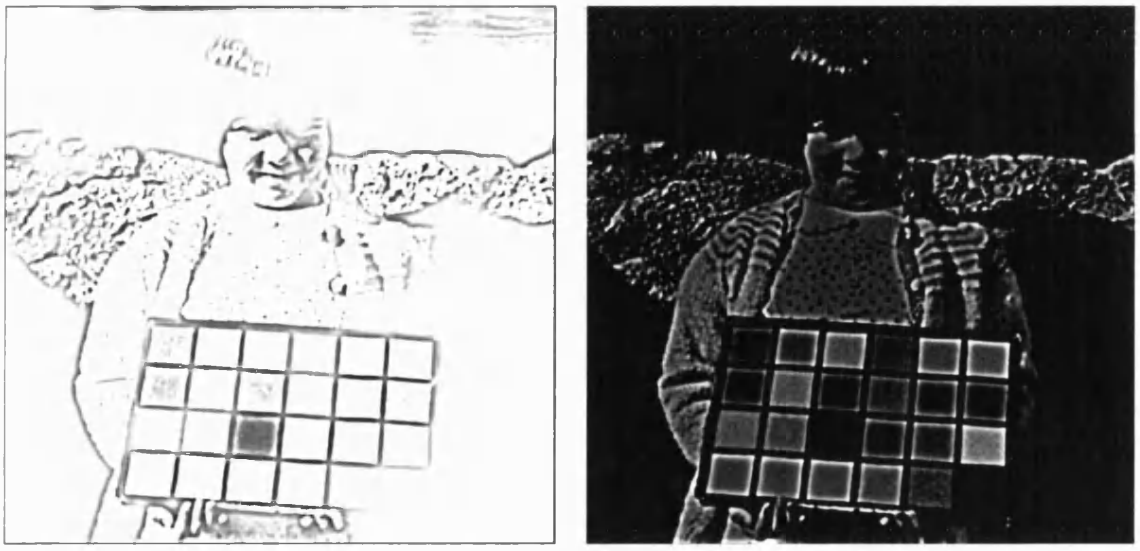
First of all, these two type I maps show contrast enhancement around edges, and an attenuated response to the brightness of regions. Secondly, both maps show much stronger differences in response to the test card colours than the cone maps of Figure 4-8B and C. This is because both d_{RG} and d_{GR} are colour opponent systems which accentuate differences in cone activity. System d_{RG} is excited by red and inhibited by green/blue, while the opposite is true for d_{GR}. Type I signals are confusing in that they mix together colour and luminance information. This is evident in their response to the greyscale sequence on the bottom of the test card. This response occurs because $\mu \neq 1$ in Equation 5.2. Both response maps also show activity at edges formed between achromatic stimuli, and this is due to the fact that the receptive field centre and surround are of different sizes. Colour difference signals dominate in the middle of large homogeneous regions. This behaviour is shown clearly by the graphs of Figure 5-8 which plot responses from a horizontal cross-section through Figures 5-7A and B.

Figure 5-7C shows responses from simulated type II blue-yellow opponent neurons. These

[8] In general, type I receptive fields were not sub-sampled, type II fields were sampled by a factor of 2 or 4 and type III, by a factor of 2. This reveals the difference in centre size between the three mechanisms.



(A) (B)



(C) (D)

Figure 5-6: Response maps for LGN neurons. "Off" centre cell excitation is shown as negative responses (dark) on a zero (light) background, and "On" centre excitation as positive (light) on a zero (dark) background. (A) Red off-centres (d^-_{RG}). (B) Red on-centres (d^+_{RG}). (C) Green off-centres (d^-_{GR}). (D) Green on-centres (d^+_{GR}).

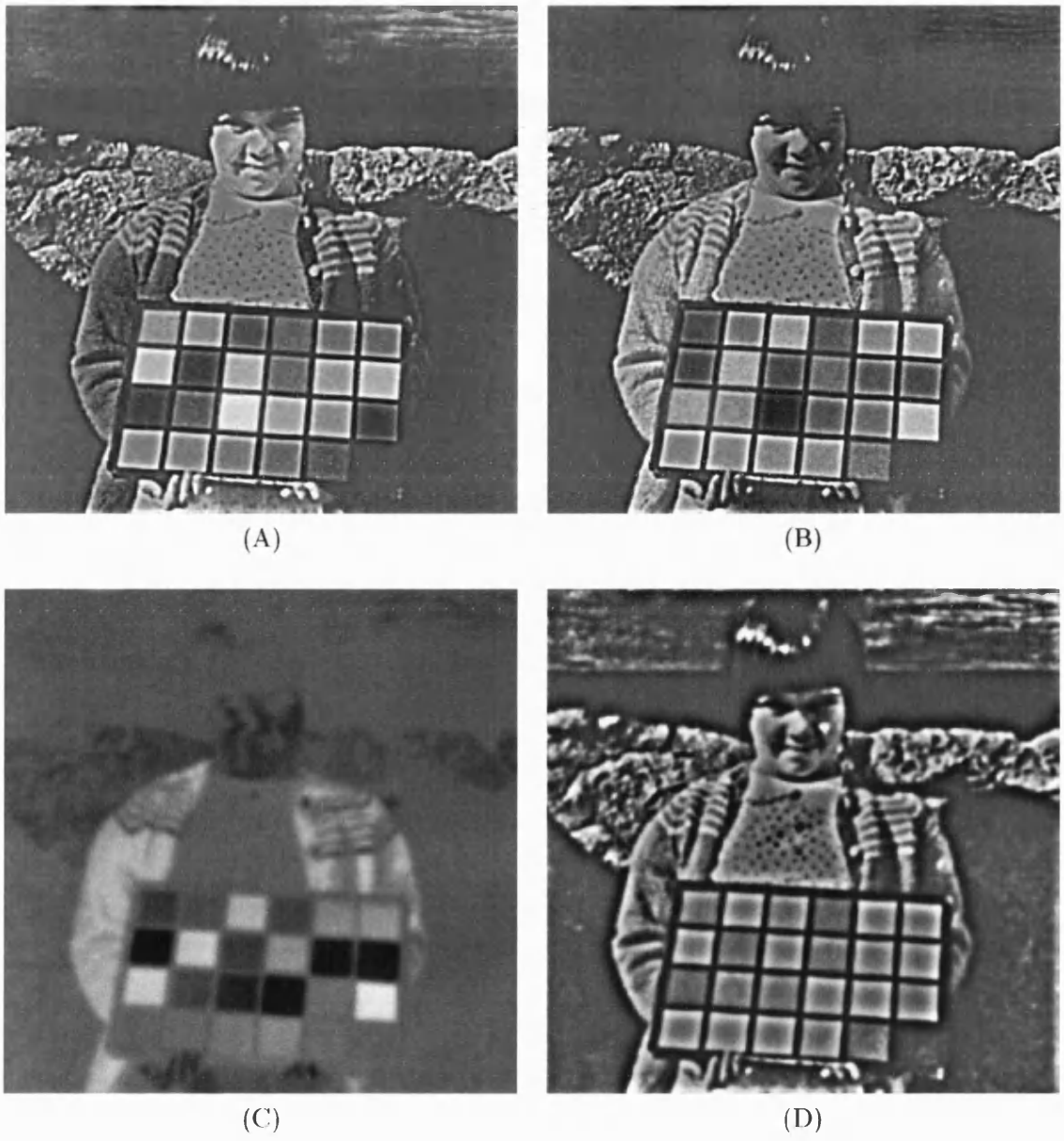


Figure 5-7: Composite response maps for LGN neurons. (A) Type I red-centred cells. (B) Type I green-centred cells ($r_c = 0.025°$, $r_s = 0.100°$, $\mu = 0.9$). (C) Type II blue-yellow opponent cells ($r_c = 0.070°$). (D) Type III broad-band M-cells ($r_c = 0.053°$, $r_s = 0.200°$, $\mu = 1.0$, $r_g = 0.400°$).

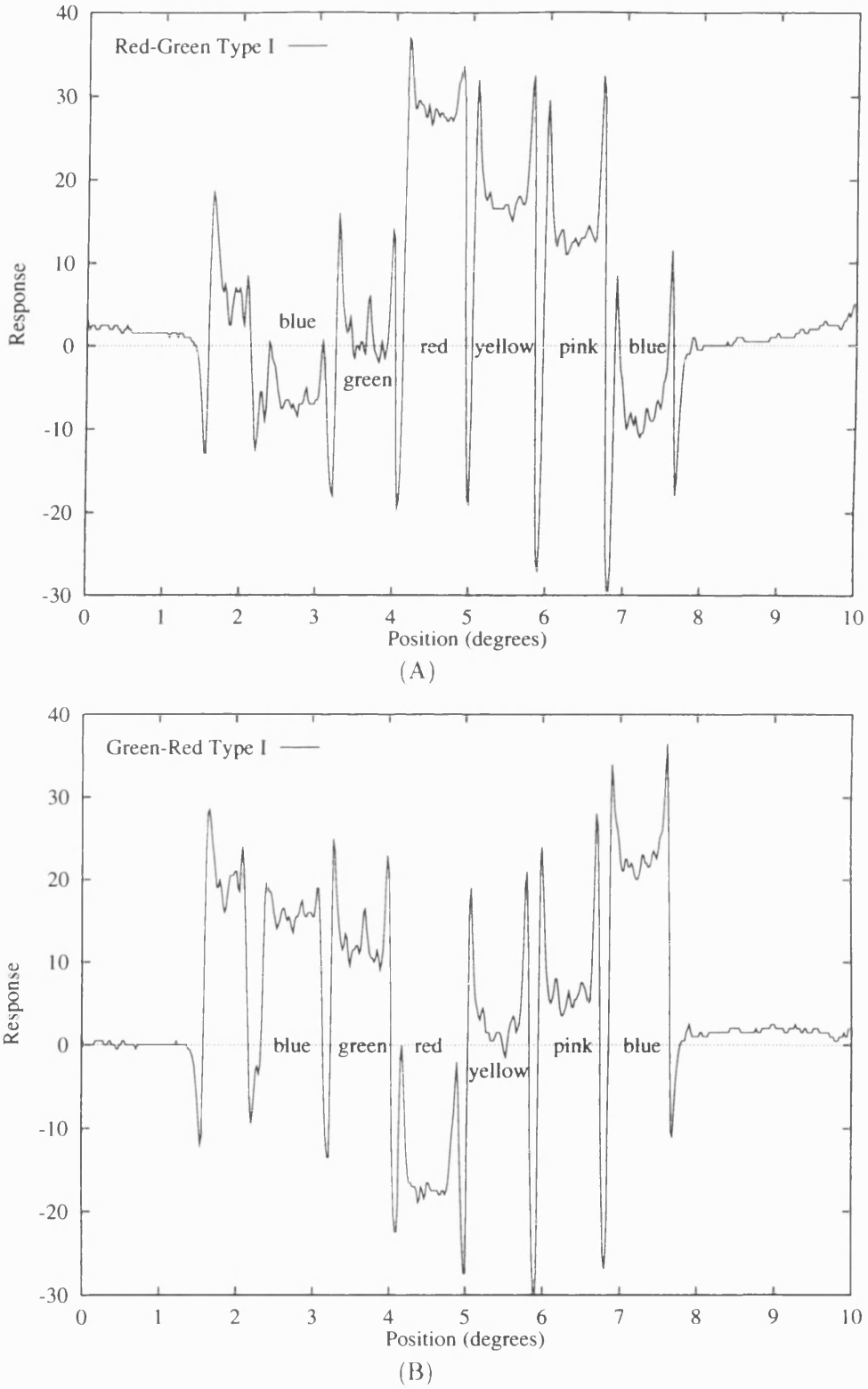


Figure 5-8: Type I LGN cell responses: Horizontal cross-section through the response maps for image line 412 which includes the third row in the test card. (A) Red-centred receptive fields. (B) Green-centred receptive fields.

responses are much simpler because they are excited by blue, suppressed by yellow/red, but are unresponsive to pure changes in luminance. Such cells have a low acuity, and this is seen in the blurred response to colour boundaries. Clearly, this channel codes chromatic information in a more direct way than do type I responses.

Figure 5-7D includes magnocellular type III responses for the same image. The response is strongly sensitive to boundaries but reacts poorly to large regions of brightness. The spectral sensitivity is equal to $V(\lambda)$ and so luminance variations are transmitted. The contrast sensitivity is higher than that of parvocellular neurons, and the contrast response is compressive as we have already seen in Chapter 4.

In conclusion, we can regard the four response maps of Figure 5-7 as representing the primary mechanisms by which all visual information about the scene is transmitted to the cortex.[9]

5.4 Chrominance and Luminance Transfer

In the rest of this chapter I concentrate on the luminance and chrominance transfer properties of the parvocellular type I receptive field arrangements. In particular, I show that this channel carries sufficient information to allow both achromatic and red-green chromatic vision, even though these dimensions are mixed in a rather complicated way for both red-centre/green-surround (RG) and green-centre/red-surround (GR) neurons.

It is very likely that chrominance and luminance information is separated from the responses of type I neurons in the cortex. Various researchers have suggested that this separation could be carried out by later spatial filtering of either of the RG or GR channels in isolation, but here I argue that it is not possible to use spatial filtering alone to achieve a useful red-green opponent perceptual system. This is because the position of unique yellow would then change dramatically with spatial frequency. Further, such a filtered RG resultant would be strongly influenced by spectral sensitivity changes that take place in the parvocellular responses during conditions of flicker and dark adaptation. Instead, cancellation between red-centred and green-centred responses is necessary.

Later, I demonstrate that subtractive cancellation, to generate a stable red-green opponent channel with wide spatial frequency bandwidth, requires extreme precision of neural convergence to give good colour acuity. Given the random nature of neural interconnections,

[9]There is, of course, a distribution of centre and surround sizes, so I have introduced three other pairs of type I maps to account for this in a flexible way.

mis-convergence may well upset a fragile measure such as this. Cortical spatial filtering is shown to improve the robustness of this cancellation stage, but at the expense of chromatic acuity—in line with observations.

Lastly, in this chapter I show that luminance, as defined by flicker photometry, cannot be extracted from parvocellular RG and GR responses either directly or by cancellation without introducing spatial frequency dependent spectral sensitivity changes. However, it is possible to have "luminance" responsive band-pass neurons in the cortex, providing that they draw input from RG and GR LGN neurons in the correct ratio and have peak spatial filter sensitivities that coincide with the spatial frequency at which opponent contributions are cancelled. Indeed it is possible to generate any red-green opponent or non-opponent response this way. These assertations are supported by mathematical arguments.

5.4.1 Background

Psychophysical research has lead to the adoption of zone models of colour perception, in which signals from the three types of cone receptor are transformed into an internal colour-opponent coordinate system (Judd [125]; Massof and Bird [160]; Guth *et al.* [93]) Since this coordinate system forms the basis of colour detection and discrimination, it is reasonable to suppose that cortical neurons exist that encode sensations in this way. Neurons have indeed been discovered that are good candidates for such a role (Michael [166, 167]; Gouras [86]; Zeki [252, 253]; Livingstone and Hubel [149]).

In general, electrophysiological results confirm psychophysical models of human colour perception in that they suggest that chromatic information is communicated to the brain via two channels similar to the "deuteranopic" (blue-yellow) and "tritanopic" (red-green) systems of opponent-colour theory. These systems, together with an achromatic system, account for many visual discrimination functions and they are thought to mediate colour and brightness perception (Guth *et al.* [93]).

The qualitative correlation between measured responses and psychophysical models appears quite satisfactory until the mechanism mediating brightness perception is considered. How is this information communicated to the brain? Cortical colour responses must be derived from those of the LGN as this is the primary gateway to the visual cortex. The parvocellular channels are strongly implicated—they alone can support the sustained scrutiny of high spatial frequencies required for primate foveal acuity. Magnocellular neurons, although possessing broad-band spectral sensitivities, respond transiently and serve a lower range of

spatial frequencies (Kaplan and Shapley [130]). In this chapter, however, we have already seen that parvocellular type I neurons respond in a complicated manner, confounding together the spatial and chromatic properties of stimuli (e.g. Figure 5-4). Other researchers have commented on this fact (Wiesel and Hubel [246]; Derrington *et al.* [63]; Billock [20]).

Is there another parvocellular mechanism dedicated to carrying spatial luminance modulation? Wiesel and Hubel's [246] original classification of response types included a group with broad-band non-opponent spectral sensitivities accounting for 16% of their sample. The largest group (76%) was made up of spectrally opponent RG neurons. Later, Livingstone and Hubel [149] claimed that 90% of parvocellular neurons were of this latter type. De Monasterio and Gouras [52] asserted that 63% of their sample of retinal ganglion cells had colour-opponent properties although it is not possible to decide from their results whether any of the remainder were tonic, as opposed to phasic units. De Valois, Abramov and Jacobs [56] found that one third of geniculate neurons were broad-band, however their sample may have also included the magnocellular layers. These neurons had sensitivities that were similar to the photopic luminance curve—a fact which lead them to suggest that there may be a separate channel communicating brightness information. More recently, Derrington *et al.* [63], in an experiment designed specifically to investigate chromatic mechanisms, found no evidence for a clear group of non-opponent parvocellular neurons. All neurons in their sample showed some degree of cone antagonism and were either blue-yellow or else red-green opponent. Most of this latter class showed some responsiveness to luminance modulation indicating out-of-balance opponency.

Zrenner and Gouras [258] reported that there was considerable variation in cone balance among retinal type I cells, with the majority being clearly dominated by one of the cone mechanisms (Section 5.3.1). Such evidence tends to suggest that there probably is not a discrete system of neurons coding for luminance alone—rather, there is a wide variation in cone balance so that some appear to be wide-band. In the central retina, type I neurons are often quite well balanced.

In conclusion, cells summing red and green cone inputs do not appear to be present in sufficient numbers to communicate spatial luminance modulation at foveal resolution. The BY system is not likely to play a role since Derrington and Lennie [64] found great difficulty obtaining a white-light contrast sensitivity curve from any blue driven geniculate neurons.

More recently, attention has turned to the type I channels in order to see what they might be good for. A number of researchers have suggested that the type I responses code

for colour at low spatial frequencies and luminance at high frequencies (Billock [20]; Ingling and Martinez-Uriegas [122]). This is because the centre and surround regions are of different sizes: their sensitivities fall at different rates as the spatial frequency is increased, altering the balance of red and green contributions. For zero frequency diffuse light stimuli, both centre and surround are sensitive generating an $r - g$ colour difference signal, while at high frequencies when the surround contribution is negligible, only the centre wavelength response is present, resulting in a wide-band non-opponent spectral sensitivity. This means that type I spectral sensitivity is strongly dependent on stimulus size (Derrington *et al.* [63]; Finkelstein [80]).

This apparent multiplexing of chrominance and luminance information in the same cell response has led to a number of schemes for separating out this information. Neurons have been discovered in the cortex that have spectral responses that appear to be less sensitive to stimulus dimensions; in particular, the double opponent-colour cells in layer IV (Michael [166, 167]) and layers II and III (Livingstone and Hubel [150]) of the primary visual cortex, and the type II cortical blob-cells of Ts'o and Gilbert [227].

How are these responses constructed from the earlier geniculate stage? Some researchers have suggested low-pass and band-pass filtering as a way of recovering the two response components (Ingling and Martinez-Uriegas [121]; Ingling and Martinez [122]; Billock [20]). Others have suggested cancellation between RG (red-centred) and GR (green-centred) channels (Lennie [144]; Martinez-Uriegas [155]; Martinez-Uriegas and Kelly [156]). Both of these operations are considered shortly, following a closer examination of the RG receptive field properties.

5.4.2 Analysis of Parvocellular Responses

Ingling and Martinez-Uriegas [122] were the first to present a useful analysis of the spatial and chromatic responses of RG neurons, showing that their responses consisted of a low-pass hue signal and a band-pass luminance signal summed together in the same channel. They used one dimensional rectangular point spread functions in their model rather than the two dimensional Gaussian functions which I have already employed here. I extend the analysis of Ingling and Martinez using the Gaussian model to show how the spectral sensitivity of RG and GR geniculate neurons changes with spatial frequency.

For the purposes of this analysis and for the rest of this chapter, I define a balanced red-green (tritanopic) system $T(\lambda)$ with spectral sensitivity given by $T(\lambda) = t_r\, r(\lambda) - t_g\, g(\lambda)$

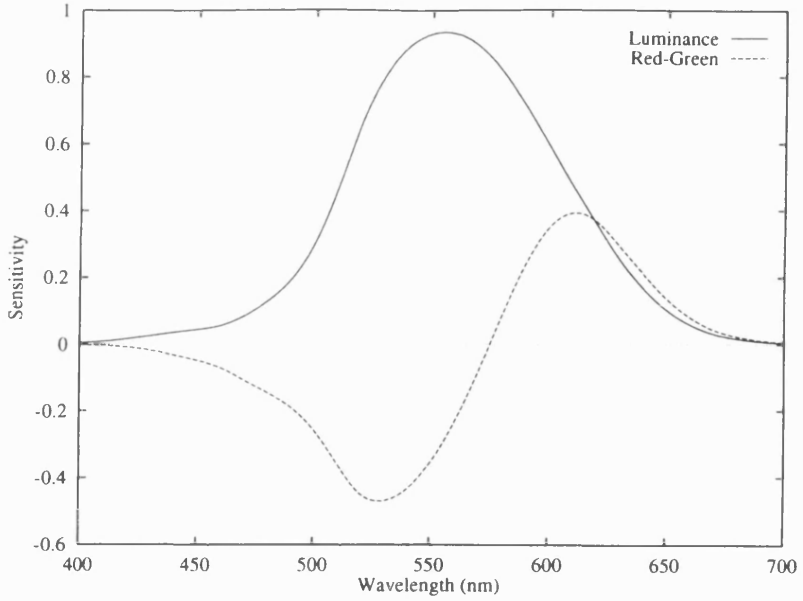


Figure 5-9: Spectral sensitivities of $T(\lambda)$, the cone balanced red-green opponent system, and $V(\lambda)$, the psychophysical luminance system (linear sensitivity scale).

and a luminance (achromatic) system given by $V(\lambda) = a_r\, r(\lambda) + a_g\, g(\lambda)$, where $t_r = 0.9553$, $t_g = 1.2836$, $a_r = 0.5967$ and $a_g = 0.3654$. These values are taken from the threshold model of Guth *et al.* [93].[10] Figure 5-9 plots the resulting sensitivity curves.

Substituting cone spectral sensitivities for receptor responses in Equation 5.2, and transforming this into the frequency domain in a similar manner to Equation 4.18, gives the spatial frequency dependent RG and GR channel sensitivities,

$$\mathbf{d_{RG}}(f,\lambda) = C(f)\, t_r\, r(\lambda) - \mu S(f)\, t_g\, g(\lambda) \tag{5.5}$$

$$\mathbf{d_{GR}}(f,\lambda) = C(f)\, t_g\, g(\lambda) - \mu S(f)\, t_r\, r(\lambda), \tag{5.6}$$

where $C(f) = \exp(-r_c^2\pi^2 f^2)$ and $S(f) = \exp(-r_s^2\pi^2 f^2)$. In this two-dimensional formulation, we have a single radial spatial frequency variable f, measured in cycles per degree. The functions $C(f)$ and $S(f)$ are the Fourier transforms of the centre and surround Gaussian spatial weighting functions, each spatially normalised to unity. The cone weights, t_r and t_g, are set so as to obtain the tritanopic spectral sensitivity at zero spatial frequency—but only if the centre and surround spatial functions happened to be balanced. The factor μ determines

[10] The resultant R:G ratios are 0.7442:1 and 1.6330:1 respectively. Sensitivity ratios derived from Equation 5.1 for 100 cd $\cdot$ m^{-2} white adaptation are 0.8184:1 and 1.6368:1, for $k_r = 0.8$, $k_g = 1.0$, $\alpha = 0.69$, $\beta = 2.0$ and $S_{min} = 0.25$.

the relative surround strength so that the spatial functions are balanced for $\mu = 1$. Here, I use $\mu = 0.9$ and put $r_c = 0.025°$ and $r_s = 0.100°$ as before.

Figure 5-10 shows the theoretical change in wavelength sensitivity to gratings as spatial frequency is varied in log intervals from 0.1–20 cyc/°.[11] At very low spatial frequencies, the curves are similar to the tritanopic system but are biased positively with slightly shifted cross points due to μ being less than unity. As spatial frequency increases the curves change shape, becoming approximately equal to the receptor curves above 5 cyc/°. The cross points are significantly shifted from the DC value for spatial frequencies around 0.5 cyc/°, with greatest change being from 1–5 cyc/°.

Clearly a scheme that tries to generate a spatial frequency invariant second-stage chromatic response by low-pass spatial filtering a single channel in isolation would have to have a very low spatial frequency cutoff in order to ensure reasonable invariance of the cross point with spatial frequency, given that unique yellow must stay at the same wavelength. Such spatial filtering would not be able to remove the achromatic contribution at DC that invariably exists with these neurons (because $\mu < 1$). For these reasons, mechanisms that code for red-green chromaticity cannot be limited to one type of receptive field centre.

The relative contribution of "chromatic" and "achromatic" information to the RG responses can be examined by using a modified version of the Ingling and Martinez-Uriegas [122] relation:

$$Cr - \mu Sg = 0.5(r - g)(C + \mu S) + 0.5(r + g)(C - \mu S). \tag{5.7}$$

Here, r and g are the weighted cone inputs, and C and S are the centre and surround spatial frequency functions as before. The $C+\mu S$ and $C-\mu S$ factors determine the contributions of $r-g$ and $r+g$ components at different frequencies. It can be seen from the way in which these factors change with spatial frequency (Figure 5-11), that the purely chrominance mediating $r-g$ term is low-pass filtered because the centre and surround functions add, while the $r+g$ brightness term is approximately band-pass. This makes the relative contribution of the $r-g$ component dominant for low frequencies. Figure 5-12 shows how the relative contributions change with spatial frequency. It is of particular importance that these figures indicate that it is theoretically possible to extract chrominance information up to a spatial resolution similar to that of brightness information, with the proviso that it will not be high frequency enhanced. Note however, that even though the centre and surround spatial frequency contributions add

[11] Receptor adaptation to the grating wavelength is ignored here but may have a significant effect (Hicks *et al.* [105]).

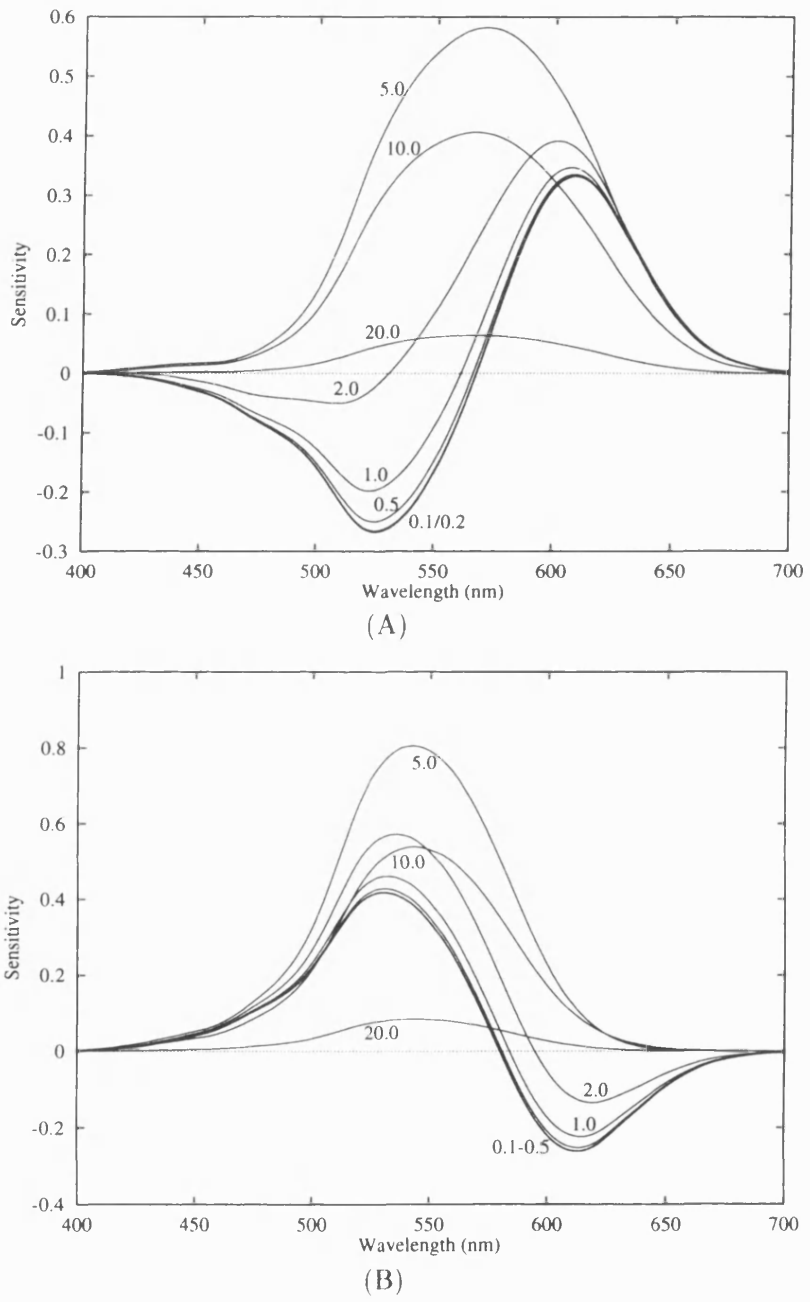


Figure 5-10: Predicted wavelength sensitivity curves for (A) red-centred and (B) green-centred type I parvocellular neurons, for log increments in spatial frequency between 0.1 and 20 cyc/°.

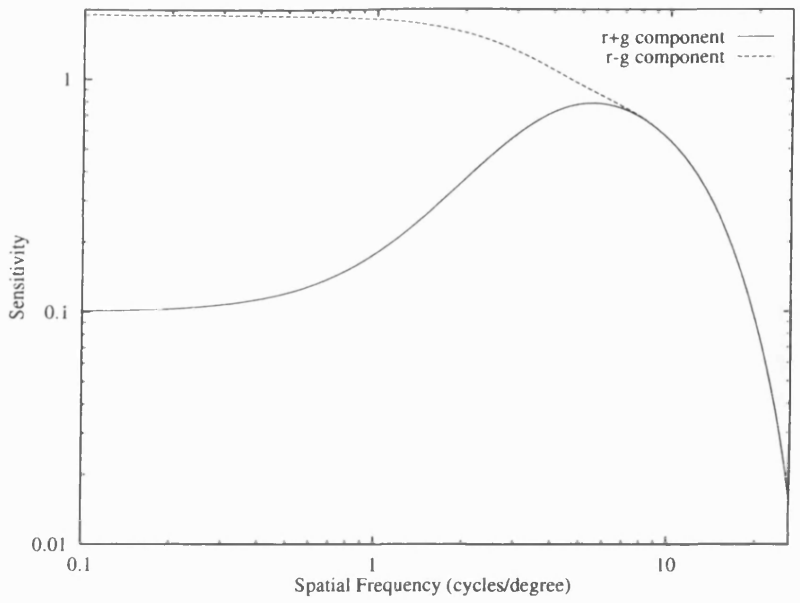


Figure 5-11: Chromatic $(r-g)$ and achromatic $(r+g)$ components comprising RG parvocellular responses.

together to boost the $r-g$ term, its peak sensitivity will generally be less than that of the $r+g$ component, because in the latter case *the cone responses are summing.*

Psychophysical experiments have shown that sensitivity to chrominance gratings has a low-pass form and sensitivity to luminance gratings has a band-pass form in line with the behaviour described here (Burbeck and Kelly [35]; Kelly [135]; van der Horst and Bouman [232]; Granger and Heurtley [87]).

The model presented here differs from that used by Ingling and Martinez-Uriegas [122] regarding the question of whether the surround is co-extensive with the centre. In other words, is the surround a ring or a disc? I have modelled the surround with a Gaussian and this fully overlaps with the centre, giving a very slight spectral opponency for centre-only stimulation. Since the peak sensitivity of the surround *spatial* function used here is about one twentieth of that of the centre this difference is not very significant. Cutting out a centre-sized hole in the surround results in its contribution becoming additive at high spatial frequencies, but this amounts to less than 5% of the centre contribution. It is difficult to establish which of the two options is correct from published responses, but in any case this model is fairly insensitive to the outcome.

The model used by Ingling and Martinez-Uriegas is rather more sensitive to this distinction because of their one-dimensional box-like treatment. Their surround adds a significant

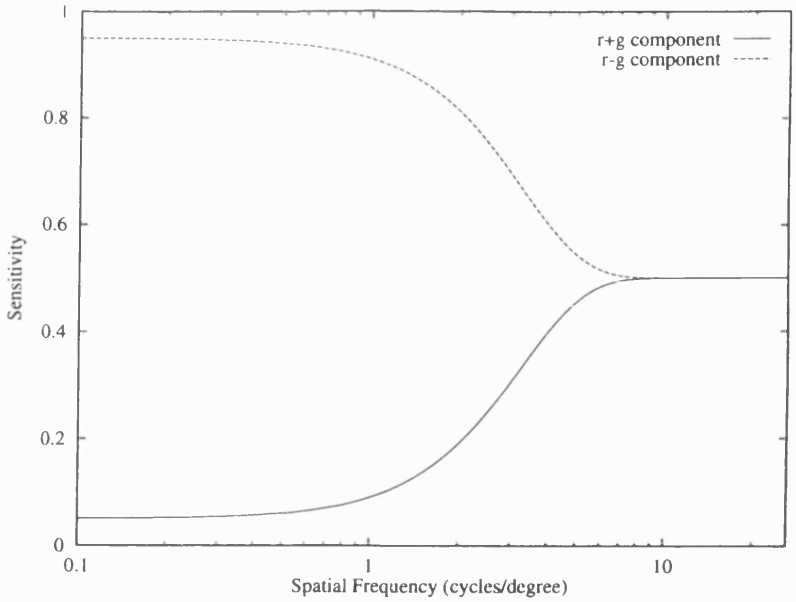


Figure 5-12: Relative amounts of chromatic $(r-g)$ and achromatic $(r+g)$ contribution to RG parvocellular responses.

excitatory component at high frequencies and this results in an overall high spatial frequency spectral sensitivity that is $r+g$ instead of centre-cone dominated. By contrast it can be seen from Figure 5-12, for the model described here, that above 10 cyc/° the $r+g$ and $r-g$ terms contribute equally, yielding a response that is just equal to r.

In summary, there are two groups of neurons in the parvocellular layers of the LGN. One group transmits a $b-y$ chrominance signal that is low-pass filtered and the other group transmits a mixture of $r+g$ brightness and $r-g$ chrominance information in proportions that vary with spatial frequency. In theory, this latter group does convey enough information about colour and brightness to account for psychophysical performance without the need for a separate achromatic parvocellular channel. This is true providing that the two mixed components can be separated.

5.4.3 Decoding Type I Responses

Having established the properties of these neurons, I now approach the problem of separating the two types of information. In general we wish to separate neatly the signals carried by the RG and GR type I parvocellular neurons in such a way that they have spectral properties that resemble the psychophysical achromatic and tritanopic systems over a wide spatial frequency bandwidth. Here, I demonstrate that it is possible to generate the red-green system $T(\lambda)$,

but not the true luminance signal $V(\lambda)$.

Ingling and Martinez [121] were of the opinion that in generating the RG responses, the vision system had found a way of communicating two types of information down a *single channel* by multiplexing in the spatial frequency domain. Decoding the two signals again in the cortex is difficult in this case, and must rely on some kind of spatial filtering. They concentrated on the problem of extracting an $r+g$ signal, not on ensuring the availability of a cross-point-stable colour-opponent signal.

Russell filtering (after Russell [199]) was therefore suggested as a way of separating out the luminance component. According to this scheme, the criterion for treating the signal as luminance-dominated is that the response should appear locally edge-enhanced—there should be a clear peak above the $r-g$ signal level on the lighter side of the edge and a clear trough below this signal level on the darker side. Unfortunately, if both luminance and hue signals are changing together at a boundary, depending on the relative signs of the two types of contrast, this can result either in enhancement of the hue edge responses or in a kind of de-emphasis. It is not clear how Russell filtering is expected to cope with this latter case.

Billock [20] describes a scheme for constructing cortical responses that also involves spatial filtering. Low-pass filtering extracts the chrominance signal and band-pass filtering is used to extract luminance. The difficulty is that neither of these signals are clearly segregated in the spatial frequency domain. We must ask what is to be done about the middle-range spatial frequencies when both components are present.

I now present a few other points that demonstrate why a single channel is not sufficient for communicating hue and luminance to the cortex:

Firstly, luminance contrast is visible down to 0.2 cyc/°, even for stationary stabilised gratings when the transient magnocellular neurons are unlikely to be contributing (Kelly [134]). Parvocellular neurons with a wide spectral bandwidth may be implicated, but it has already been argued that these do not form a clearly distinct subgroup and do not exist in sufficient numbers in the fovea to be devoted to this task. Clearly luminance contrast and an opponent colour measure cannot both be extracted from the same channel at the same spatial frequency. Neither will high-pass filtering to select for luminance boundaries followed by extrapolation, as some have suggested, achieve the desired result because in this case, there is no high frequency energy to select—an extra channel is required. In this situation, I suggest that, providing field centres are slightly more dominant than surrounds, both RG (red centred) and GR (green centred) channels must be summed together to give

luminance at these low spatial frequencies. This summation removes chromatic information by cancellation.

Secondly, at high spatial frequencies, a single RG channel does not communicate a $V(\lambda)$-like response. Instead, with rising frequency, the spectral sensitivity tends towards that of cones in the receptive field centre as the surround becomes ineffective. High frequency filtering will not therefore extract an $r + g$ signal, as Ingling and Martinez-Uriegas [122] state, but a predominantly red cone signal. Once again, some form of summation between RG and GR responses is needed to create the luminance signal.

Thirdly, the information carried by RG neurons is drastically changed during conditions of flicker or dark adaptation. Zrenner [257] points out that while targets flickering at 5Hz give rise to colour-opponent responses in the retina, flicker rates of 33Hz result in a $V(\lambda)$-like sensitivity in the same cell due to different time delay properties in the centre and surround. At this frequency, the centre and surround act in synergism rather than antagonism. This is also true for very brief stimuli—all trace of spectral opponency being lost. Alternatively, Wiesel and Hubel [246] have shown that RG opponent cells often receive input from rods, so that during conditions of dark adaptation these neurons switch from being narrow-band cone-opponent, to wide-band rod excited, giving excitatory responses for wavelengths that previously inhibited them. Clearly, there is no hope of separating hue and luminance measures when the spectral sensitivity can change to this extent. Any filtering scheme would have to take into account temporal and spatial variables as well as the state of dark adaptation.

All of these problems can be solved by combining RG and GR responses in order to obtain luminance and subtracting them to obtain a balanced hue signal. A switch to rod vision would make the RG and GR spectral sensitivities identical and so no differential activity would be present at any one retinal position to drive the hue channel. Colours would therefore not be seen at low light levels but luminance modulation would remain. During flicker, when receptive field surrounds lose their suppression and eventually become excitatory, providing that RG and GR neurons are affected in an equal and opposite manner, there would be only a slow desaturation of colour, rather than a gross shift in hue as the LGN neurons lose their opponency. Cancellation has already been suggested by Lennie [144] and by Martinez-Uriegas and Kelly [156]. Although Billock [20] rejected it on pharmacological grounds, the arguments just presented suggest that there is no viable alternative to the combination of responses from neurons with different centre types.

It can be seen Figure 5-10 that red-centred and green-centred neurons carry chrominance

information with opposite signs and so additive cancellation between these two channels can separate the signals. In this way we can generate $T'(f,\lambda)$, with the same spectral sensitivity as the tritanopic system and $A'(f,\lambda)$, an approximation to the luminance system:

$$\begin{aligned} T'(f,\lambda) &= \mathbf{d_{RG}}(f,\lambda) - \mathbf{d_{GR}}(f,\lambda) \\ &= [t_r r(\lambda) - t_g g(\lambda)]\,[C(f) + \mu S(f)] \quad \text{and} \qquad (5.8) \\ A'(f,\lambda) &= \mathbf{d_{RG}}(f,\lambda) + \mathbf{d_{GR}}(f,\lambda) \\ &= [t_r r(\lambda) + t_g g(\lambda)]\,[C(f) - \mu S(f)]. \qquad (5.9) \end{aligned}$$

Note that $A'(f,\lambda)$, when generated in this way, will have different sensitivity weights from $V(\lambda)$. This is because the weights are actually inherited from the red-green system, but with a sign change in the green contribution. Such a result will still be called an achromatic channel here, because it has a wide-band wavelength response produced by summing red and green cone excitations. This topic is discussed later.

It can be seen from these equations that cancellation allows factorisation into spatial and spectral terms, giving channel spectral sensitivities that are independent of spatial frequency. In particular the tritanopic cross-point is not shifted away from near 575nm.

A possible weakness of this scheme is that cancelation needs fairly precise matching in terms of receptive field spectral and spatial properties. With regard to spectral matching and producing the correct weights to cancel achromatic contributions in the cortex, I suggest that cortical synaptic learning and lateral inhibition may act to force the achromatic and reg-green axes to be orthogonal. I discuss spatial matching later in this chapter.

5.4.4 Simulation Results

Having analysed the responses of the parvocellular type I channel and also having considered cancellation as a method for recovering stable brightness and colour information, I now present simulation results for the process.

Figure 5-13A shows a close up of a red bean flower on a green leafy background. Figure 5-13B and C show the results of simulating the activity of neurons having red-centred (RG) and green-centred (GR) receptive fields, in the same fashion as Section 5.3.4, but for a 5° field of view. From these images, it may be seen that for large regions of colour, RG and GR responses are very different because the chrominance components are of opposite sign. At boundaries, the achromatic component dominates and has the same sign for both centre

types. The sum and difference images of Figures 5-13D and 5-14A show that these can be successfully separated. The purely low-pass quality of the recovered $r-g$ response is evident from the lack of edge-enhancement in Figure 5-14A, but it can also be seen that the spatial resolution is still fairly high. I refer to this reconstructed red-green signal as a type II response since it is low-pass and purely opponent like the LGN type II blue-yellow system.

Figure 5-15A shows the same recovery technique applied to the 10° field of view test image of Figure 4-8A. Graphs of red-green and blue-yellow type II responses taken from these two maps are also shown in Figure 5-16. The blue-yellow type II response has the lower spatial resolution, as can be seen from the flatter "pink" response.

5.4.5 Colour Acuity

A significant feature revealed by the graph of Figure 5-11 is that in theory, the acuity of the chrominance channel can be very good. For high spatial frequencies only the receptive field centres are contributing, but red centres and green centres are still subtracted to give a chrominance signal. The psychophysical upper spatial frequency limit for detection of pure chrominance gratings is found to be around 20 cyc/°—about half that for luminance (Kelly [134]). It may be supposed that part of the chromatic performance could be due to the lower sensitivity of this system or to chromatic aberration, but Sekiguchi *et al.* [210], using a procedure that isolated neural factors, found that the visual system has a neurally imposed spatial bandwidth 1.8 times lower for chrominance stimulation than for luminance stimulation and that this was the case for both red-green and blue-yellow colour dimensions. Since both $r+g$ and $r-g$ signals are filtered by the same centre mechanism we must ask why the psychophysical tritanopic system is observed to have a lower colour acuity than predicted.

One possible explanation is advanced here. It arises out of a consideration of how easy it could be for the cortex to implement the cancellation scheme described. The scheme relies on subtracting responses from LGN neurons that have precisely overlapping receptive fields with identical spatial properties. Billock [20] has already pointed out that schemes relying on exact receptive field registration may be prone to error. Although cortical architecture shows considerable organisation, and indeed self-organisation, there is still scope for randomness at the interconnection level. Indeed, because red and green cones cannot occupy the same retinal location, receptive field centres driven by different cones can never be truly in register. Receptive field positional, spatial and spectral properties are subject to random variation from one neuron to the next, and if a scheme depends heavily on the detailed repeatability of these

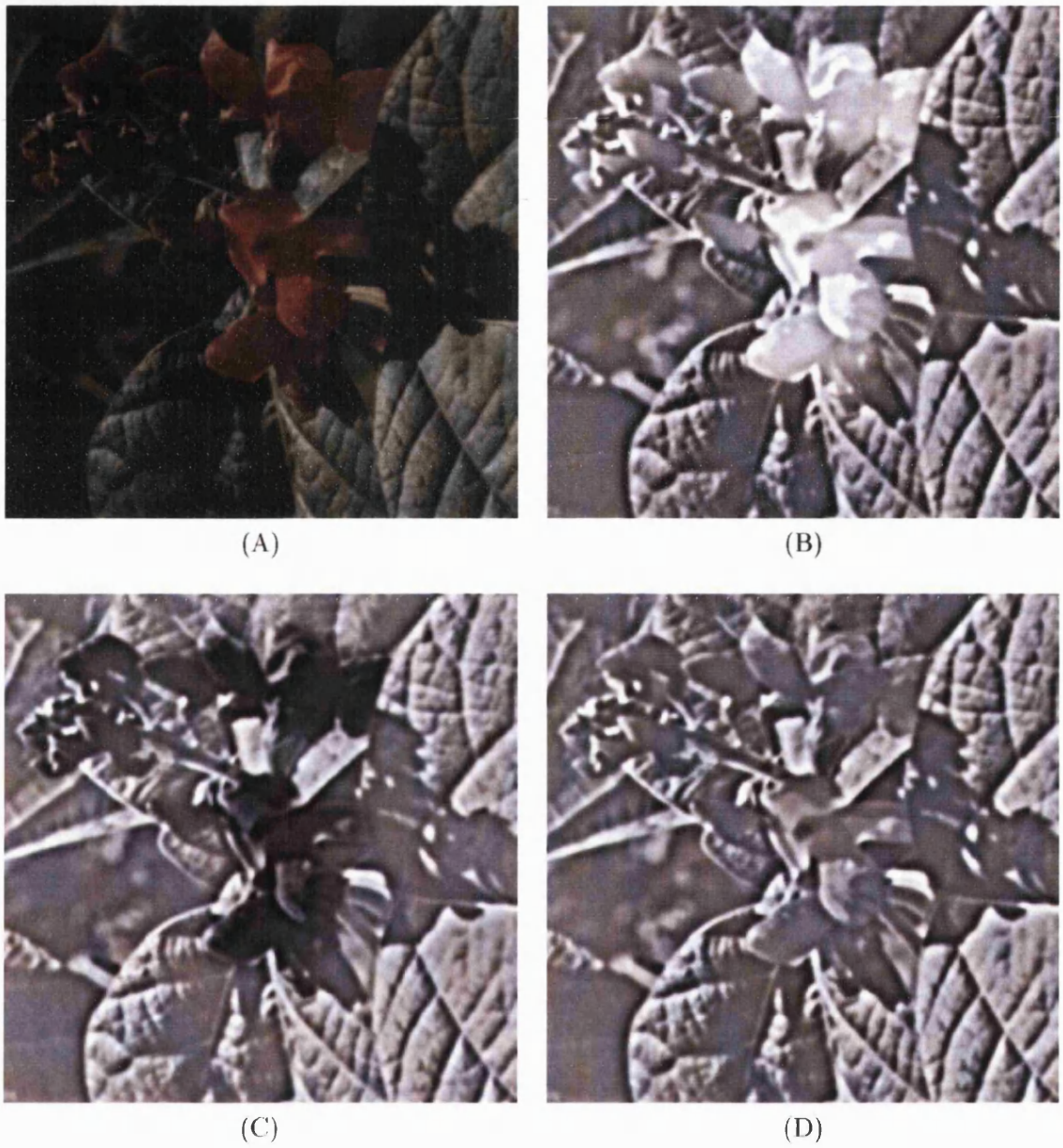


Figure 5-13: (A) Original 256×256 bean-flower image. (B) Simulated type I RG (red centre) responses. (C) Simulated type I GR (green centre) responses. (D) Reconstructed achromatic component with spectral sensitivity $A'(\lambda)$ produced by addition of B and C.

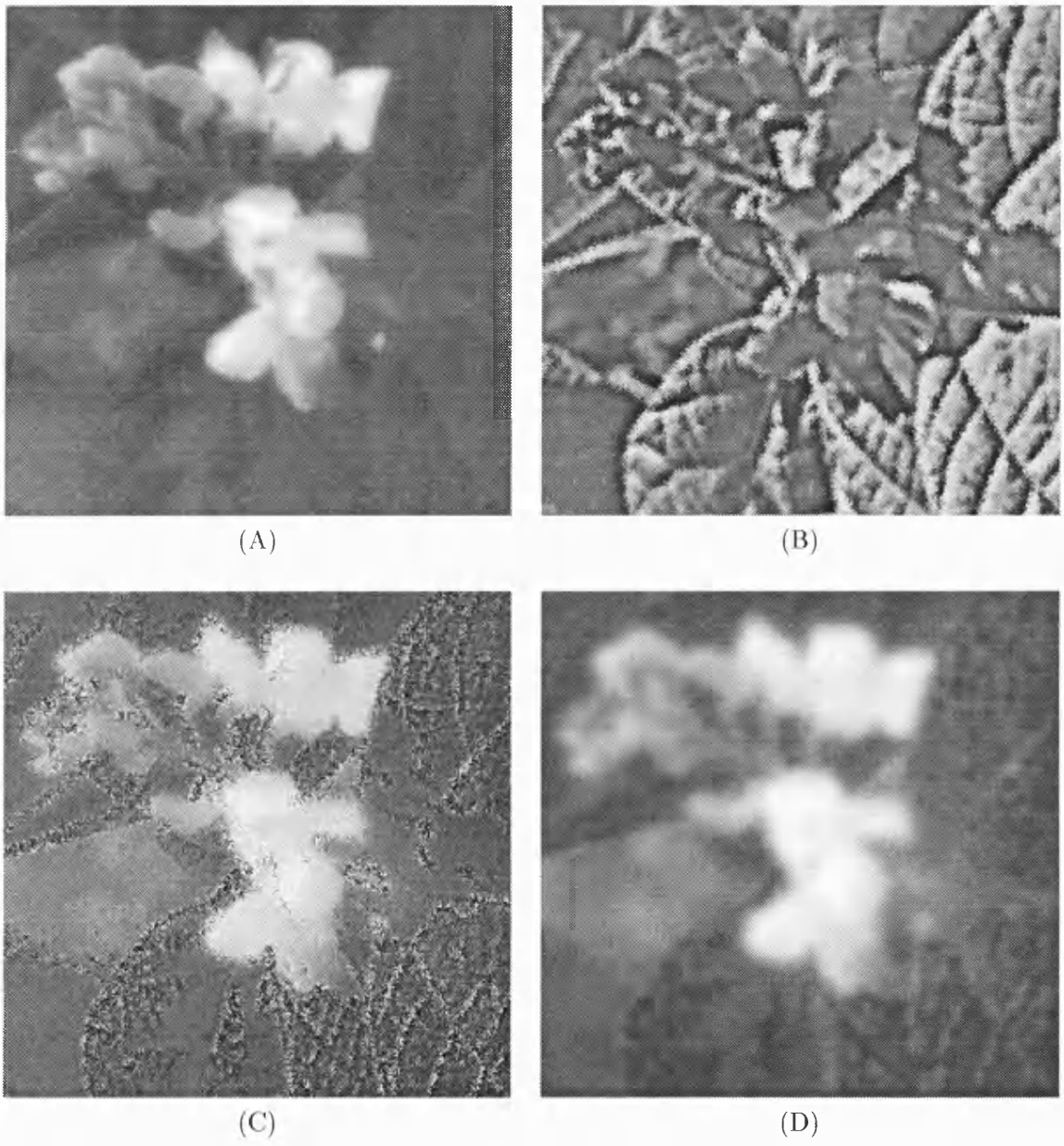


Figure 5-14: (A) Reconstructed tritanopic component (red-green type II response) with spectral sensitivity $T(\lambda)$ produced by subtracting Figures 5-13B and C. (B) Corruption of achromatic component by "noisy" cortical routeing. (C) Corruption of red-green type II response. (D) Removal of noise in C by spatial convergence, at the expense of chromatic acuity.

(A) (B)

Figure 5-15: Type II responses to the test image of Figure 4-8A. (A) Red-green type II response derived in the cortex from RG and GR type I LGN inputs. (B) Blue-yellow type II response reproduced from Figure 5-7C for comparison.

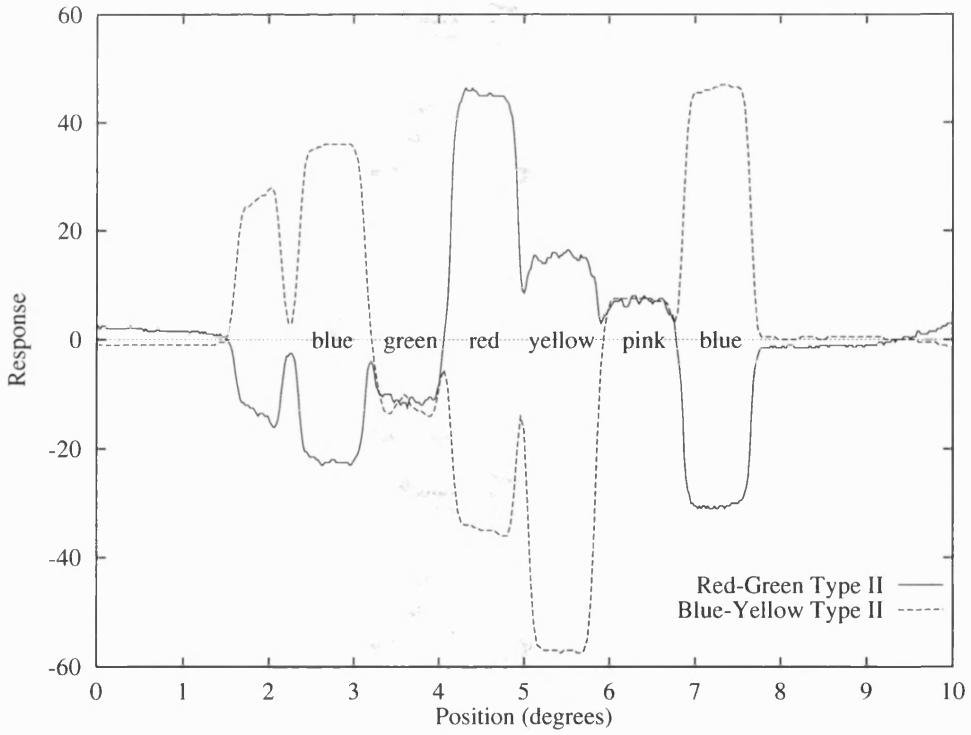


Figure 5-16: Type II blue-yellow LGN and cortical red-green cell responses: Horizontal cross-section through the response maps for image line 412 which includes the third row in the test card.

properties or else on interconnection accuracy, it may not be viable.

In view of this, I next study the stability of the cancellation approach to variations in receptive field overlap. As previously mentioned, cortical colour measures can be obtained from RG and GR neurons having coincident receptive fields. In practice, great precision is unlikely to be achieved, so the effect of having fields that are not in exact registration will now be simulated.

Figure 5-14C shows the result of subtracting green- and red-centred LGN responses with a random positional offset up to $\pm 0.04°$. This offset value corresponds to about half the surround radius. (This figure is probably larger than actually encountered, but it serves to illustrate the point.) Considerable noise is introduced as luminance information breaks through, especially at light-dark boundaries. The achromatic channel response of Figure 5-14B appears to be less affected by random offsets.

To explain this effect we first note that subtracting values at neighbouring points in an image is a high frequency boosting process similar to differentiation, while adding together such points results in smoothing which reduces high frequency content. In generating the cortical measures by cancellation, components that are being selected are added and therefore smoothed when a positional offset is introduced, whereas components that are being cancelled are subtracted and therefore differentiated. For the achromatic system, differentiation of the unwanted $r - g$ component introduces only a small amount of noise because most of its signal energy is present at low spatial frequencies and this is removed by the effective differentiation. For the tritanopic system however, differentiation of the unwanted $r + g$ component serves to boost the already strong presence of this signal at high spatial frequencies. This results in random breakthrough of luminance information which appears as irrelevant chromatic noise, particularly along edges since this is where the high frequency content is dominant.

Given that colour responses are badly affected by interconnection randomness, it follows that one reason that high resolution red-green colour measures are not derived could well be because they cannot easily be generated. The cortex can overcome this problem for a lower range of spatial frequencies by introducing a spatial smoothing stage. In this scheme, single cortical neurons with large receptive fields pool responses from a number of neighbouring LGN neurons. Incoming signals from RG neurons are excitatory and those from GR neurons are inhibitory. Such inputs are intended to be present in pairs with receptive fields that overlap precisely but in practice they may not do so. Lack of overlap will not be so devastating, because on the whole, cortical neurons will receive a covering of their own receptive fields .

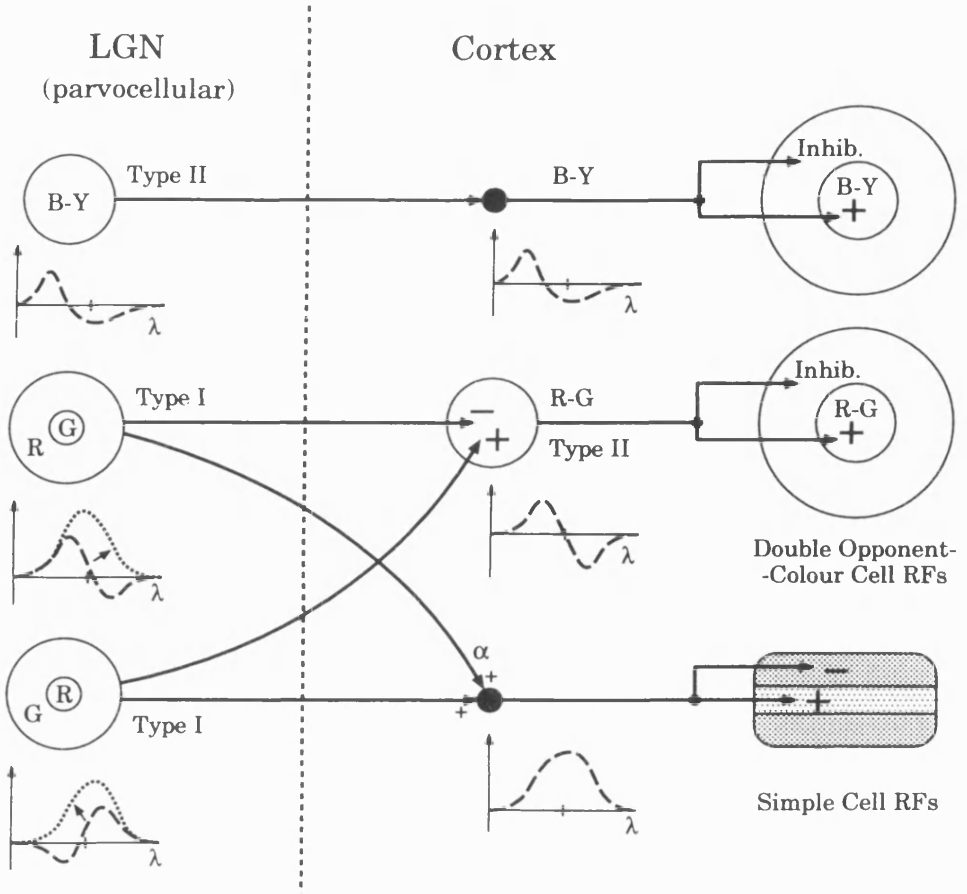


Figure 5-17: Interconnection scheme. Responses from opponent LGN neurons are subtracted over an area to yield a stable spectral sensitivity. Double opponent-colour responses can then be obtained by spatial filtering. Conceptually, cortical simple-cell neurons can sum together and band-pass filter α-weighted inputs from parvocellular RG neurons to obtain, $V(\lambda)$, or any other opponent or non-opponent spectral-sensitivity. This diagram represents a linear processing scheme and is not intended to imply any particular connectivity at the level of on- and off-centred neurons.

from both LGN receptive field types. Note that spatial filtering is being introduced here to clean up noise in the colour responses, not to ensure spatial frequency independence of their spectral-sensitivity—cancellation has already achieved this. Figure 5-17 illustrates the arrangement in diagrammatic form.

The configuration described above is equivalent to following the theoretical cancellation stage with a low-pass filtering stage in which any high frequency noise is smoothed out. It is important to note that such filtering, with its resulting reduction in chromatic acuity, is a *countermeasure against*, rather than a *consequence of* interconnection misalignment. Figure 5-14D gives the simulation result when a Gaussian smoothing stage is introduced to

clean up the corrupted tritanopic response. For this smoothing stage I have set $r_c = 0.090°$ giving a spatial low-pass filter with a response that falls to 10% at about 6 cyc/°. The simulation shows that this amount of smoothing does in fact remove the noise introduced by ±0.04° of random receptive field offset. This may be the function of the cortical 'type II' receptive fields described by Ts'o and Gilbert [227] among others. These neurons sum over a region large enough to encompass a number of parvocellular receptive fields.

5.5 A Generalised Decoding Strategy

In this section I introduce a method for deriving arbitrary spectral response functions (involving red and green cone input) from parvocellular type I responses. As an example, I consider the problem of producing a $V(\lambda)$ luminance sensitivity at the level of cortical simple cells. There is evidence, however, that even in the cortex, there is still considerable mixed opponency (Lennie *et al.* [145]).

5.5.1 Luminance

The classical photopic luminance function, $V(\lambda)$, plotted on Figure 5-9, is produced by adding together red and green cone sensitivities using the achromatic weights, a_r and a_g. Lennie *et al.* [146] review a number of psychophysical methods for measuring spectral sensitivity functions that produce a $V(\lambda)$-like curve. The CIE $V(\lambda)$ function itself is based primarily on measurements obtained with the use of heterochromatic flicker photometry. This technique is likely to reflect heavily the wavelength sensitivity of the magnocellular pathway since this is the channel most sensitive to flicker.

Using an acuity criterion that isolates the parvocellular channel, Ingling *et al.* [119] have demonstrated that at high spatial frequencies this channel also carries luminance with a $V(\lambda)$-like sensitivity. At first sight such a result is difficult to explain because the cancellation scheme described earlier can only achieve a spatial frequency stable sensitivity that sums cone inputs using weights inherited from the red-green opponent system. The weight ratio involved in this system is approximately $R/G = 2/3$ whereas for $V(\lambda)$ the ratio is about $R/G = 5/3$ favouring red cones (Ingling and Tsou [124]). The achromatic and tritanopic axes formed from $V(\lambda)$ and $T(\lambda)$ are almost orthogonal in cone space, whereas the pseudo-luminance axis $V'(\lambda)$ made using red-green opponent weights (Equation 5.9) is only 63° away from the tritanopic axis. This is less than ideal.

It is known that there are about twice as many red cones than green cones in the retina

and that this cone ratio may be the factor that results in magnocellular channels having a $V(\lambda)$ sensitivity. Ingling and Tsou [124] have pointed out that the retinal cone distribution could result in there being twice as many parvocellular neurons having red centres than have green centres. At high spatial frequencies only the centres will be responding and so cortical summation among LGN afferents could result in the correct $R/G = 5/3$ luminance sensitivity for acuity targets. I now extend this argument to show how a true luminance response, or indeed any desired RG spectral sensitivity can be derived at any spatial frequency in the cortex from type I parvocellular input.

5.5.2 Ratio Cancellation

In the scheme presented next, cancellation between red-centred and green-centred neurons is still used but the responses are now summed in some ratio α:

$$A''(f,\lambda) = \mathbf{d_{RG}}(f,\lambda) + \alpha\,\mathbf{d_{GR}}(f,\lambda) \tag{5.10}$$

$$= [C(f) - \alpha\mu S(f)]\,t_r\,r(\lambda) + [\alpha C(f) - \mu S(f)]\,t_g\,g(\lambda). \tag{5.11}$$

This only results in a broad-band spectral sensitivity that is spatial frequency invariant when $\alpha = 1$. For other values of α there will be some colour-opponent contribution that varies in magnitude with spatial frequency but the overall sensitivity will be predominantly broad-band. The important observation is that for any single spatial frequency of interest, it is possible to select a value of α that yields $V(\lambda)$ at that one frequency. If this value of α is then fixed, raising or lowering the stimulus spatial frequency will cause the sensitivity to shift slowly away from its luminance dominated form.

Equating the cone weight ratio predicted by Equation 5.11 to that of the achromatic system $(a_r : a_g)$ yields the value of α needed to obtain a luminance response at any chosen spatial frequency:

$$\frac{[C(f) - \alpha\mu S(f)]\,t_r}{[\alpha C(f) - \mu S(f)]\,t_g} = \frac{a_r}{a_g}. \tag{5.12}$$

Rearrangement gives

$$\alpha = \frac{a_g t_r\,C(f) + a_r t_g\,\mu S(f)}{a_r t_g\,C(f) + a_g t_r\,\mu S(f)}. \tag{5.13}$$

Figure 5-18 plots α against spatial frequency. It can be seen that choosing $\alpha = 0.7$ yields a balancing spatial frequency of about 3 cyc/°. If we fix this value of α we can express the resulting $A''(f,\lambda)$ in terms of the colour-opponent and luminance components

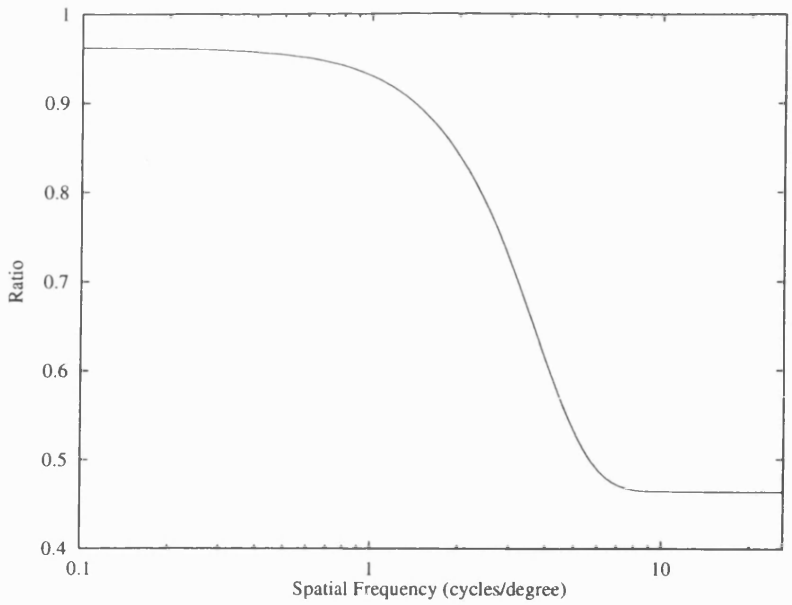


Figure 5-18: Plot of cancellation-ratio α required to achieve a luminance response for a given spatial frequency.

that make up its response. Figure 5-19 shows how the two contributions vary with spatial frequency and inspection reveals that at 3 cyc/° the chromatic contribution is completely nulled. By resolving the responses along the luminance and chrominance axes in this way, we can also obtain the relative contribution of luminance for a variety of values of α. Figure 5-20 shows that it is always possible to achieve a 100% luminance driven response at some spatial frequency; and for wide ranges of frequency, the signal can be more than 80% luminance dominated for a fixed value of α.

Also shown on Figure 5-20 is the relative sensitivity curve for a 1.2 octave Gabor filter. Such spatial filters have often been used to model the band-pass qualities of cortical neurons (Daugman [51]). This filter can be used to select the region of achromatic response where the $V(\lambda)$ component is most dominant.

This results in a method for recovering luminance information that involves summation and spatial filtering. A cortical neuron, e.g. a simple cell, receives inputs from both red-centred and green-centred LGN neurons. These inputs are summed in such a ratio as to obtain the desired spectral sensitivity at the spatial frequency to which the cell is most sensitive. It is therefore possible for the whole ensemble of cortical neurons, having different weights and spatial frequency peaks to decode collectively a luminance or colour-opponent signal for the whole spatial frequency range represented in the cortex. Any spectral sensitivity

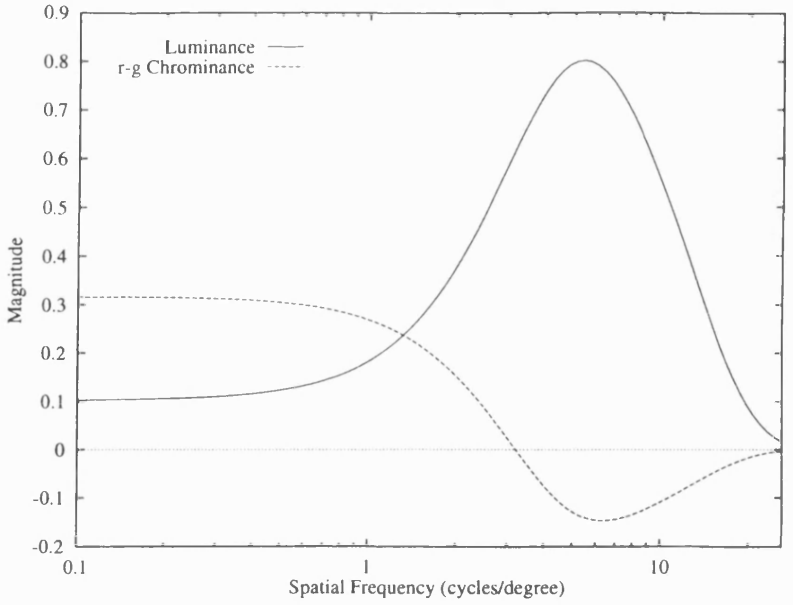


Figure 5-19: Luminance and red-green chrominance contributions to the response $A''(f, \lambda)$ for $\alpha = 0.7$.

arising from red and green cones can be stabilised by piecewise construction across a number of frequency channels. This recovery technique differs from previous filtering attempts which tried to demultiplex luminance carried at high spatial frequencies from the responses of single LGN receptive field types.

At the highest spatial frequencies the summation ratio required to achieve a luminance response (using centre-normalised geniculate fields) matches the cone distribution ratio. Wiesel and Hubel [246] found about twice as many red-centred as green-centred receptive fields in the parvocellular layers of the LGN, so it is very likely that cortical simple cells would receive cone inputs in this ratio. In view of this, at least for cortical neurons responsive to high spatial frequencies, a $V(\lambda)$ parvocellular mediated sensitivity is very likely. This is in line with observation (Ingling and Tsou [124]). For lower spatial frequencies, where the weights need to be different, opponent effects are known to contribute to brightness perception when measures are taken to discount magnocellular input (Lennie *et al.* [145]).

In summary, spatial filtering has been introduced to select the luminance dominated responses from a signal obtained by combining RG and GR parvocellular outputs in some suitable ratio. Such filtered responses will still be driven by colour-opponent signals despite having a wide-band sensitivity and ought therefore to betray their parvocellular origin under suitable test conditions: for example, Krüger and Gouras [139] found that slit-length some-

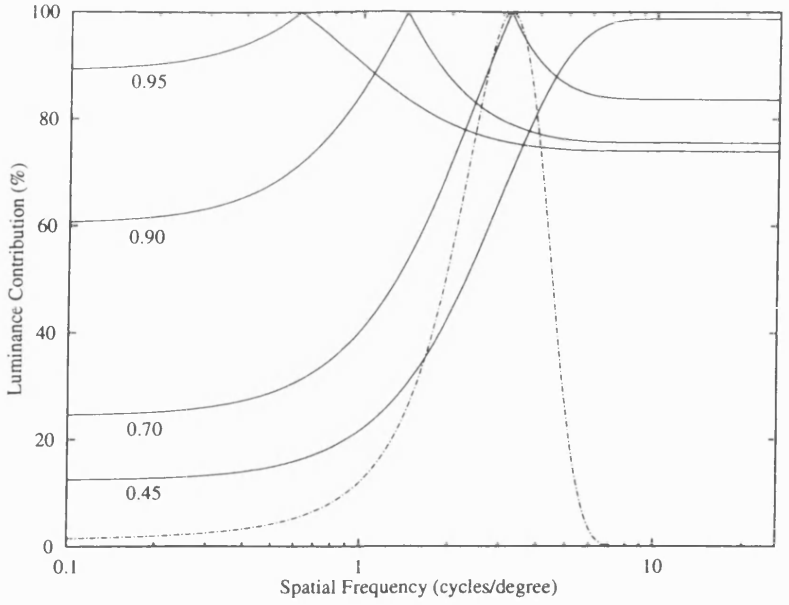


Figure 5-20: Solid curves: Relative contribution of luminance to $A''(f, \lambda)$ for four values of α. Dashed curve: Normalised sensitivity curve for a Gabor band-pass filter with 1.2 octave bandwidth centred near 3 cyc/°.

times altered the sensitivity of these neurons in a way that suggested summation between different LGN centres.

5.6 Discussion

I now summarise a few points regarding the spatio-chromatic behaviour of neurons in the LGN.

Red-green driven neurons represent the vision system's trade-off between two signals that originate from the same cone classes. The spatiochromatic organisation of their receptive fields mixes luminance and colour signals and, since the relative proportion of the two signals in the mixture changes with spatial frequency, the type I channel dynamic range is allocated differently at each frequency. Luminance is favoured at high spatial frequencies because it forms the principle component of natural scenes and is therefore the most informative quantity for determining scene structure—a task that requires good spatial resolution.

The spread of cone domination in large-field spectral responses is evidence for another tradeoff: In the fovea, which is specialised for colour discrimination, opponent responses are more closely balanced (Zrenner and Gouras [257]), giving dynamic range over to the $r - g$ signal at the expense of luminance. Both colour and luminance information can still be sepa-

rated by cancellation and summation of RG and GR responses in the cortex, but propagation through the retina and geniculate nucleus becomes the limiting factor for sensitivity. Away from the fovea, there is more spread in cone balance, but RG and GR signals can still be combined appropriately to form a chromatic signal at the cortical level, as long as for every opponent cross-point shifted to short wavelengths there is a neuron nearby with a corresponding shift towards long wavelengths. Out-of-balance responses represent a bias in favour of luminance since the opponent-colour signal is then less explicitly coded at the geniculate nucleus.

Blue-yellow LGN neurons show well balanced opponency and have receptive fields without a centre-surround organisation. This is explained by noting that this channel is specialised for transmitting a stable $b - y$ signal with the same low-pass spatial characteristics as the cortical $r - g$ signal.

Since the magnocellular pathway carries transient signals, it is unlikely to be part of the mechanism that processes the colour and brightness of large long duration test fields. The sustained activity of parvocellular neurons is more clearly implicated. For such low spatial frequency (i.e. large) test fields, the majority of parvocellular neurons, having a colour opponent organisation, will be carrying a predominantly chromatic signal. Luminance information for this test configuration will be carried by virtue of the out-of-balance nature of the opponency—either red-biased or green-biased—and also by the small number of neurons with wide-band $V(\lambda)$-like spectral responses. The chromatic signal will however dominate the detection process.

The geniculate nucleus carries signals that are ideal precursors to cortical visual processing: Motion updates are relayed to the cortex via the magnocellular pathway; luminance information is high-frequency boosted to give channel dynamic range over to components that are significant in the detailed structural analysis of scenes; perceptual colour information is encoded in RG and GR channels in such a way that it can be extracted in the cortex; a further colour dimension is available in the blue-yellow responses. Although some of these receptive field properties are initially obscure (e.g. Marr [154] page 262, thought that type I cells represented bad engineering), they form a suitable starting point for the process of vision.

5.7 Conclusions

In this Chapter, I have introduced chromatic opponency and shown how the non-linear cone model produces opponent-responses that are strongly influenced by stimulus saturation and weakly by stimulus radiance, given the correct adapting conditions. These responses show a good correlation with experimental results. I have developed a natural formulation of two cone opponency which results in cone sensitivity weights that are similar to those of previous second-stage zone models, and this opponency has been included in a model of three LGN receptive field types.

Extensive simulation results have been presented in this chapter. I considered the encoding of chrominance information in type I responses and demonstrated that this information could be successfully decoded—even when "sloppy" routeing was present. Lastly, I advanced a plausible scheme which involves ratio cancellation and spatial filtering in order to show that any spectral sensitivity, including $V(\lambda)$, can be possessed by individual cortical neurons and can be made very insensitive to spatial frequency when narrow-band spatial filters are employed.

Chapter 6

Into the Striate Cortex

6.1 Introduction

In this chapter I am concerned with neural responses in the primary visual cortex (V1). Cortical cells are different from their precursors in the retina and LGN in that most show orientation selectivity and have narrow spatial frequency tuning functions. They are unresponsive to diffuse illumination, and respond primarily to edges or other significant features in the scene. In V1, a vast number of cells analyse the scene in parallel, computing information about orientation, spatial frequency, spatial phase, length, colour, localised motion direction, and in part, binocular disparity. This area of the brain is concerned with creating a detailed spatial map, to which later vision centres refer when they construct more generalised perceptions.

Most of the neurons in V1 fall into the simple/complex cell groups originally described by Hubel and Wiesel [114, 115]. Here, I introduce models for these two types of behaviour, and these models are used in the remainder of the thesis. The simple cell model makes use of a modified Gabor function. Seven spatial frequency channels are defined in Section 6.2 and simulation results are presented. In Section 6.3 I introduce models for complex cell behaviour and present further results. In Section 6.4, which covers representational issues, I use these simulation results to demonstrate that a limited number of orientation channels can successfully code for edge orientation in a way that is consistent with human performance. Simple and complex cell responses to wide-band "features" (edges/bars) are also discussed, and various observations are made on the spatial localisation of such features. This chapter ends with the presentation of a novel spatial sampling theory for cortical cells. This theory

allows for the lossless representation of image information within a scheme that includes extensive under-sampling of simple cell outputs.

6.2 Simple Cells

Simple cells are found primarily in layer 4 of the striate cortex, but also in layer 6. Layer 4 is the layer in which geniculate fibres terminate most densely, suggesting that simple cells are driven by geniculate neurons directly. Parvocellular and magnocellular afferents arrive in different sublamina of layer 4 and there is segregation of the two streams at this level.[1]

From now on, I concentrate on the cortical cells that are driven from parvocellular inputs. In addition, I do not consider the mixed chromatic properties of simple and complex cells. These are made luminance sensitive by driving them from type I RG and GR receptive fields combined in the appropriate ratio.

Linear Filtering

There is still speculation on how simple and complex cells achieve orientation selectivity and narrow, DC-insensitive, spatial frequency tuning functions. The simplest interpretation is that simple cells apply a pattern of weights to geniculate input and sum linearly. This means that all their selectivity is derived from a bottom-up filtering action.

It is clear, however, that there are a number of other influences on simple cell behaviour: Inhibition between cells having different orientation selectivity acts to narrow orientation tuning curves (e.g. the gain control model described by Wilson and Humanski [248] which does not require good orientation selectivity to be achieved by input filtering), and inhibition between cells tuned to different spatial frequencies acts to narrow the bandwidth in the spatial frequency domain. De Valois and Thorell [59] found that neurons show "surround" inhibition in both spatial and spatial frequency domains, and Sillito [213] demonstrated that inhibition within the cortex is necessary to achieve good orientation selectivity. A combination of linear filtering and inhibition is therefore used in practice. I consider some of these orientation and spatial frequency domain interactions in the next chapter, but for now I adopt a linear filter model which can be made to account for as much of the final selectivity as necessary.

Due to their low spontaneous activity, simple cells have rectified responses. Ignoring this

[1] There is a significant fraction of cortical cells which mix parvocellular and magnocellular input. I find that the process of retinal contrast gain control which linearises M-cell responses, also makes the magnocellular and parvocellular systems compatible with regard to mixed input to cortical receptive fields.

obvious output non-linearity, the underlying filtering action is found to be linear.[2] This has been demonstrated by a number of studies: Andrews and Pollen [5] and Movshon *et al.* [176] both showed that simple cell spatial line-weighting functions could be predicted by an inverse Fourier transform of their spatial frequency tuning curves. Albrecht and De Valois [2], De Valois *et al.* [53] and Albrecht *et al.* [3] found that cells were selective for the Fourier components in wide-band stimuli, rather than for the wide-band stimuli themselves.

Previous Models

The earliest model of how simple cell receptive fields are constructed from unoriented geniculate inputs is the "aligned-LGN" model of Hubel and Wiesel [114]. These authors propose that a set of centre-surround geniculate neurons with receptive fields arranged along a straight line provide excitatory input to each simple cell. The spatial weighting function of LGN input to the simple cell is therefore a thin oriented line of positive values, but the resultant receptive field is made by a convolution between this and the point-spread function of the centre-surround input. Hubel and Wiesel's scheme accounts for orientation selectivity, but cannot account for a narrow-band frequency response and lack of excitation by diffuse illumination, without additional suppression. A more complicated weighting function is required.

The Gabor model (Marčelja [157]; Daugman [51]) was introduced as an efficient "sampler" of visual information and a good model for simple cell receptive fields. Support for this model comes from Webster and De Valois [244]. In general, however, the Gabor spatial frequency selectivity curve is found to fall off too slowly at low spatial frequencies (Hawken and Parker [98]). For wide-band Gabor functions, this results in a response at DC, so for their simulation work, Heitger *et al.* [101] engineered a "stretched" Gabor function which does not display this behaviour. I do not adopt their model here because it is not easily specified as a space domain kernel. This is because their function only becomes two dimensional when constructed from radially separable parts in the frequency domain. Radial separability is not a necessity, and many neurons do not show this property (Webster and De Valois [244]).

Various other models for simple cell selectivity are described and evaluated by Hawken and Parker [98]. They find that a "d-DOG-S" model is able to fit the greatest variety of cortical contrast sensitivity functions. This one dimensional model is comprised of three spatially separated DOG units (arising from geniculate receptive fields). The outer units

[2]It is linear in the width direction. Hammond and Mackay [95, 96] have cast doubt on the linearity of length summation. This topic is covered in Chapter 8.

are weighted with the opposite sign to the middle unit. In essence, the two dimensional interpretation is that of a three-row extension of Hubel and Wiesel's original aligned-LGN model. Hawken and Parker find that this model can actually be used to manufacture simple cell receptive fields having odd and even phases, while maintaining the same spatial frequency tuning curve.

In this chapter, I am less concerned with the exact mechanism by which simple cell fields are formed than with modelling their spatial frequency responses and resulting receptive field arrangements. For this reason, I adopt a modified form of the Gabor model (m-Gabor model) which allows for easy setting of receptive field phase, centre frequency, bandwidth and orientation bandwidth. This model will now be described.

6.2.1 The Modified Gabor Model

The Gabor model was originally formulated as a complete receptive field description. In this section, I use the filter as a spatial weighting function to derive simple cell responses from geniculate input. In this case, the resulting spatial frequency response curve is merely the product of the filter selectivity curve and that of the preceding subcortical stage(s).

For this simulation I pool together input from a number of parvocellular channels (each with different centre and surround sizes) by addition to create a wide-band frequency response which acts as a contrast sensitivity function. Simple cell filters each select a narrow spatial frequency range from this. In real life, LGN neurons with approximately the same peak frequency as the simple cell filter may provide the major input, although this is not entirely certain. The "combine and split" arrangement, adopted here, provides the same essential results: filters with narrow-band selectivity and an overall retinally-sourced contrast sensitivity function.

To construct a linear filter that is suitable for modelling odd- and even-phased simple cells, I start with a partial filter function which is specified in the frequency domain by

$$\mathbf{W}(f_x, f_y) = \exp\left(-\eta^2 \ln 2 \, \frac{(f_x - f_c)^2 + \lambda^2 f_y^2}{f_c^2}\right), \tag{6.1}$$

where f_c is the centre (peak response) frequency, λ is the receptive field aspect ratio (which indirectly sets the orientation bandwidth), and $\eta = (2^b + 1)/(2^b - 1)$, where b is the spatial frequency bandwidth measured in octaves at half height. This partial function is used in the

frequency domain definition of odd and even Gabor filters according to

$$\mathbf{Q}_e(f_x, f_y) = \mathbf{W}(f_x, f_y) + \mathbf{W}(-f_x, f_y) \tag{6.2}$$

$$\mathbf{Q}_o(f_x, f_y) = j\,\mathbf{W}(f_x, f_y) - j\,\mathbf{W}(-f_x, f_y). \tag{6.3}$$

These filters correspond to a two dimensional spatial receptive field that is maximally selective for vertically oriented Fourier components having a frequency of f_c. The space domain convolution kernels are given by

$$Q_e(x, y) = \frac{2\pi f_c^2}{\eta^2 \lambda \ln 2} \exp\left(-\frac{\pi^2 f_c^2}{\eta^2 \ln 2}\left[x^2 + \frac{y^2}{\lambda^2}\right]\right) \cos(2\pi f_c x) \tag{6.4}$$

$$Q_o(x, y) = \frac{2\pi f_c^2}{\eta^2 \lambda \ln 2} \exp\left(-\frac{\pi^2 f_c^2}{\eta^2 \ln 2}\left[x^2 + \frac{y^2}{\lambda^2}\right]\right) [-\sin(2\pi f_c x)]. \tag{6.5}$$

These functions are simply sine or cosine gratings that are multiplied by a localised two dimensional Gaussian envelope, which is circular when $\lambda = 1$.

In order to modify the frequency response of the Gabor function so that the low frequency limb falls off at a faster rate and so that there is never any response at DC, I introduce an oriented Gaussian high-pass filter, defined by

$$\mathbf{H}(f_x, f_y) = 1 - \exp\left(-\frac{\pi^2 f_x^2}{f_c^2}\right). \tag{6.6}$$

This filter shows attenuation below f_c and its space domain convolution kernel is given by

$$H(x, y) = \delta(y)\,[\,\delta(x) - \frac{f_c}{\sqrt{\pi}} \exp(-x^2 f_c^2)\,].^3 \tag{6.7}$$

The frequency response of the complete m-Gabor kernel (even phase) is then given by $\mathbf{Q}_e(f_x, f_y)\mathbf{H}(f_x, f_y)$, and shows a more symmetrical band-pass response (on log frequency axes) than a standard Gabor function—particularly for wide bandwidths. Clearly, the resultant frequency response will have a half-height bandwidth less than b, but this constant can be used to set the m-Gabor bandwidth to any value in the range used here (0.82–2.15 octaves).

Linear responses from odd and even simple cells are therefore produced by convolving

[3] ILthough this function was included to modify the standard Gabor function, it could also have a structural interpretation. The filter could be viewed as selecting the DC response of the standard Gabor form, and subtracting this signal: a form of cortical surround suppression.

parvocellular inputs by both high-pass and band-pass kernels:

$$q_e(x,y) = H(x',y') \otimes Q_e(x',y') \otimes \frac{d_{RG}(x,y) + \alpha\, d_{GR}(x,y)}{1+\alpha} \tag{6.8}$$

$$q_o(x,y) = H(x',y') \otimes Q_o(x',y') \otimes \frac{d_{RG}(x,y) + \alpha\, d_{GR}(x,y)}{1+\alpha}. \tag{6.9}$$

Here, the orientation selective kernels are rotated using the transforms $x' = x\sin\theta - y\cos\theta$ and $y' = y\sin\theta + x\cos\theta$. This allows a simple cell having any preferred axis of orientation to be simulated. Orientation θ is defined so that when $\theta = 0$, the filter selects horizontal features in the input.

Filter Properties

Figures 6-1 and 6-2 show the one and two dimensional impulse responses for the m-Gabor filters defined by $H(x',y') \otimes Q_e(x',y')$ and $H(x',y') \otimes Q_o(x',y')$ for $\theta = 90°$. Figure 6-3 shows the one dimensional responses to an impulse and two step edges. The even filter responds with a positive peak to a positive going impulse, while at edges, the odd filter produces either a positive or negative peak, depending on edge contrast sign. Real simple cells show output rectification and can only carry positive signals. This means that four types are necessary to code this information, and these are referred to as: even-positive, even-negative, odd-positive and odd-negative—responding with peaks at light bars, dark bars, dark-light edges and light-dark edges respectively.

Gabor filters show a number of interesting properties, as discussed by Daugman [51]. Increasing the spatial frequency bandwidth reduces the size of the space domain Gaussian envelope and hence the number of cycles in Figure 6-1. Conversely, decreasing the bandwidth b results in a wider impulse response containing more cycles of the frequency f_c. The overall width of a vertically oriented simple cell receptive field is set by the space constant of the Gaussian envelope in Equation 6.8. This envelope has an effective Gaussian r_c (Equation 4.7) which is defined by

$$r_c = \frac{\sqrt{\ln 2}}{\pi f_c}\left[\frac{2^b+1}{2^b-1}\right]. \tag{6.10}$$

The aspect ratio λ indirectly sets the orientation bandwidth, so that for large values of λ the neurons have long receptive fields along the orientation axis and therefore have sharp selectivity for orientation. The orientation half-bandwidth is given by the equation

$$\theta_b = 2\tan^{-1}\frac{2^b-1}{2\lambda\, 2^{b/2}}. \tag{6.11}$$

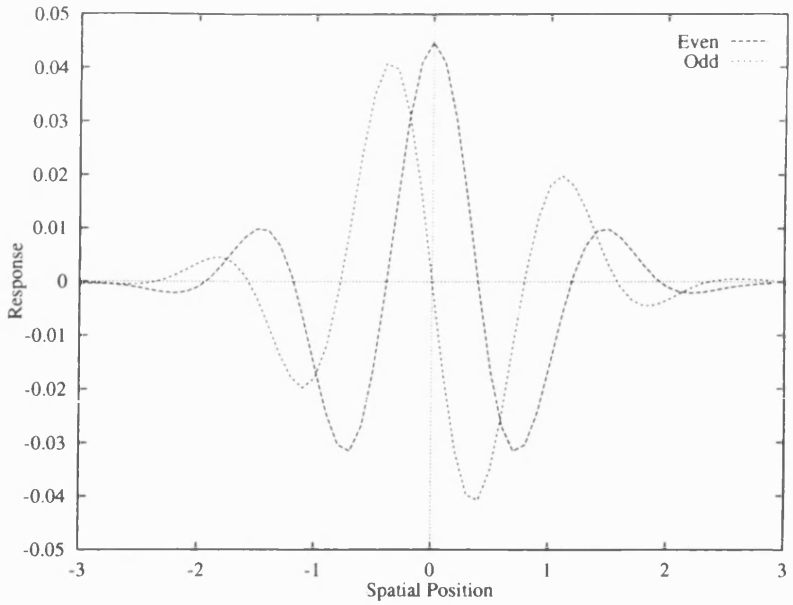


Figure 6-1: One dimensional impulse responses for even and odd m-Gabor filters (bandwidth = 1 octave, arbitrary x-scale).

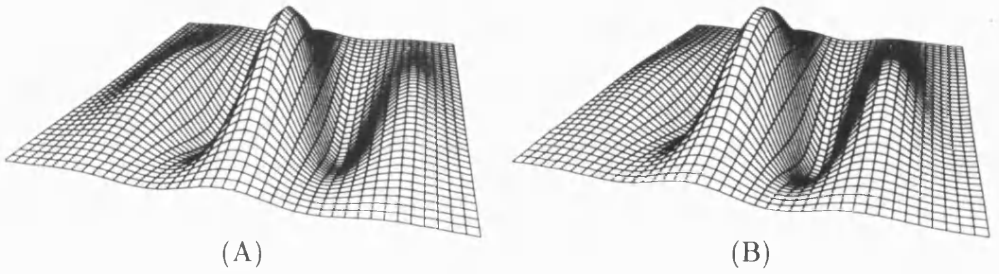


Figure 6-2: Surface plots of (A) even and (B) odd m-Gabor filter impulse responses (bandwidth = 1 octave).

The inclusion of a high-pass filter makes little difference to the values of r_c and θ_b defined above.

Channel Parameters

Psychophysical experimentation has determined that there are orientation selective channels in the vision system and that these channels have a bandwidth of about two octaves (narrower at high spatial frequencies). Wilson *et al.* [249] proposed a six channel model to account for the effects of spatial masking in the cortex. This multichannel approach is supported by neurophysiology: De Valois *et al.* [57], when recording from neurons in V1, found that at every retinal location there is a collection of cells with spatial frequency selectivities distributed over

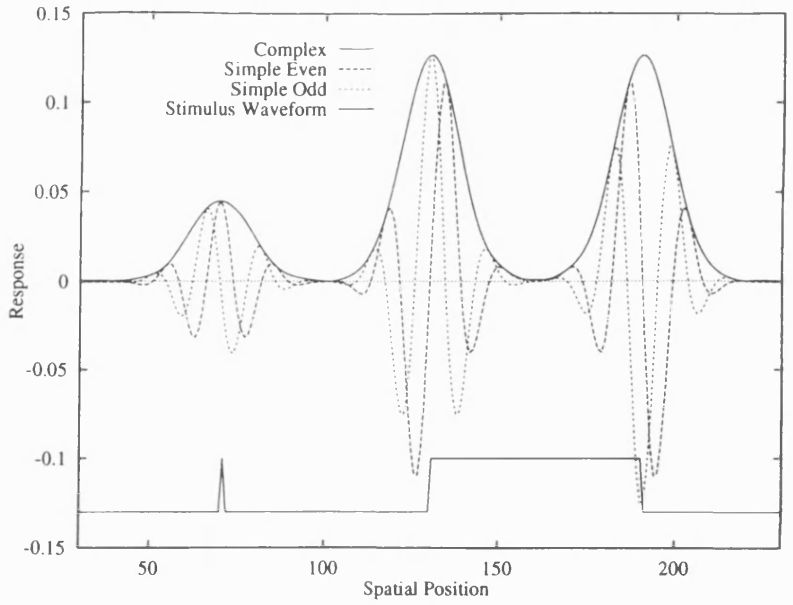


Figure 6-3: Responses from cortical simple and complex cells to an impulse and two step edges. Complex cells are introduced in Section 6.3.

the spectrum.

The simulation presented here includes seven spatial frequency channels with the normalised sensitivity curves of Figure 6-4. Foster *et al.* [82] find that, for the macaque, parafoveal neurons in V1 have peak sensitivities in the range 0.5–8.0 cyc/° and a mean bandwidth of 1.4 octaves. For V2, the spatial frequency range is 0.2–2.0 cyc/°. The channels used here cover the V1 spatial frequency representation and the lower two include V2. Even lower spatial frequency channels could not be usefully simulated because of the 10° window size. The simulation channel with highest frequency is similar to that of Wilson *et al.* [249].

For present purposes, $b = 2 - \log_{10}(f_c)$, so that bandwidth increases at the rate of 0.3 octaves per octave decrease in filter centre frequency (De Valois *et al.* [57]). I have also set $\lambda = 1$ for all channels, so that orientation selectivity half-height bandwidth reduces from 78° to 32° as centre frequency increases from 0.7–15.0 cyc/° (Webster and De Valois [244]). Parker and Hawken [187], when studying the spatial structure of receptive fields in primate V1, reported an orientation bandwidth range from 10° to 70° with a mode of 40°. In addition, they found that the modal ratio of receptive field length to sub-unit width was about 4 for both simple and complex cells. This corresponds approximately to a receptive field having an aspect ratio of one, assuming a spatial frequency bandwidth of 1.3 octaves.

Irrespective of the orientation bandwidth, eight orientation channels were used in the

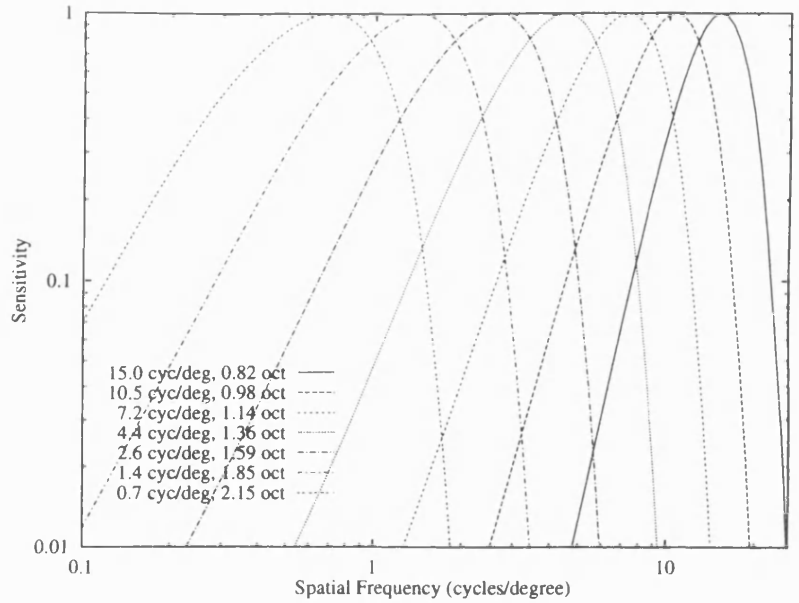


Figure 6-4: Cortical spatial frequency channels used in the simple cell simulation.

simulation at each spatial frequency. In all there are seven channels $\times$ eight orientations $\times$ two phases = 112 simulated response maps. Each map is representative of a pair of neurons: one carries positive excursions of the signal and the other carries negative excursions. Each spatial location is therefore analysed in parallel by 224 simulated neurons. Details of channel selectivities are found in Appendix B.

6.2.2 Simulation Results

Figure 6-5A shows a test image that was used in Chapter 4. When this is processed to obtain responses from simple cells (including also the retinal and geniculate stages) then Figure 6-5B is produced. The simple cells simulated here only respond to image features which are close to the horizontal. The channel has an orientation half-height bandwidth of 44° and has an even phase which results in a zero response from those cells which are positioned on an edge boundary. Figure 6-5C and D show responses for vertically selective neurons with even and odd phases. The odd-phased neurons respond with a positive peak at the dark-light staircase transitions. Figure 6-6 shows a pair of one dimensional sections through each of these two figures.

Figures 6-7 and 6-8 illustrate the behaviour of neurons having different spatial frequency tuning and different orientation tuning, and it can be seen that they each respond to a selected range of feature orientations and spatial scales.

In summary, simple cells respond in a phase dependent manner to narrow-band oriented Fourier components in an image. In Section 6.4 I examine more closely the information that they select, and what this means in terms of a representation of the visual world.

6.3 Complex Cells

In V1, complex cells are present outside of layer 4 (Gilbert [85]). They show similar spatial frequency and orientation selectivity to simple cells but do not respond in a linear way. In particular, they are not sensitive to stimulus phase. Figure 6-3 includes example responses obtained when complex cells are stimulated by edge and bar patterns. A single complex cell will respond to a sine wave grating if it has the correct orientation and spatial frequency, but as the grating is translated across the receptive field, the cell produces a constant firing level without modulation due to the grating luminance profile.

Hubel and Wiesel originally proposed that complex cells were driven by simple cells. Movshon *et al.* [174], using a two-line interaction technique, found that complex cells had sub-units that resembled simple cell receptive fields. It is also possible that some complex cells receive direct input from the LGN. In particular, a type of complex cell found sometimes in layer 3 but primarily in layer 5—the "special" complex cell—is very sensitive to small spots as well as being selective for the orientation of longer lines (Gilbert [85]; Livingstone and Hubel [150]; Hammond and Mackay [96]). Special complex cells are discussed extensively in Chapter 8.

6.3.1 Complex Cell Models

A selection of complex cell models is reviewed by Spitzer and Hochstein [221]. Several researchers have pointed out that complex cells calculate a local energy measure (Heitger *et al.* [101]; Du Buf [34]; Morrone and Burr [171]). Morrone and Burr also found that, for isolated edges or bars, the position of a feature is perceived to be at the maximum in local energy, rather than at some other candidate position; for example, at the zero-crossing in the second derivative. This suggests that complex cells have a role in the spatial localisation of features without regard to their spatial phase.

Here, I consider two models for simulating standard complex cells. The first model is the local energy model which Heitger *et al.* [101] used in their simulation. This has the form

$$u(x,y) = \sqrt{q_e(x,y)^2 + q_o(x,y)^2}. \tag{6.12}$$

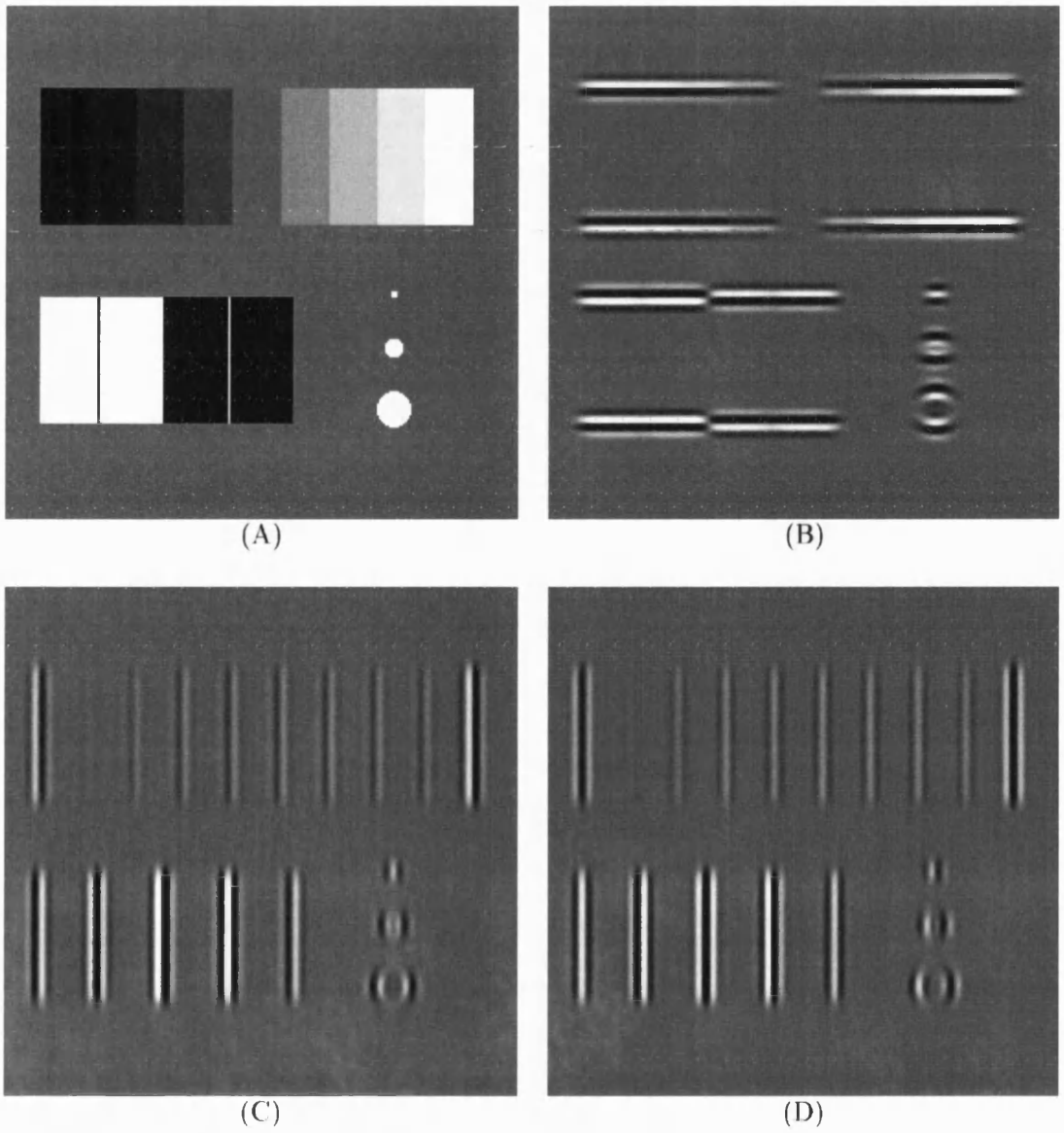


Figure 6-5: (A) Black and white test image. (B) Responses, $q_e(x, y)$, for simple cells having horizontally oriented, even phased (tripartite) receptive fields ($\theta = 0°$, $f_c = 7.2\,\mathrm{cyc}/°$, $b = 1.14$ octaves). (C) Responses for even phased receptive fields when $\theta = 90°$. (D) Responses for odd phased receptive fields. Mid-grey represents the zero response. Positive values are carried by "even-positive" cells and negative values by "even-negative" cells in the case of B or C, and by "odd-positive" and "odd-negative" cells respectively for D.

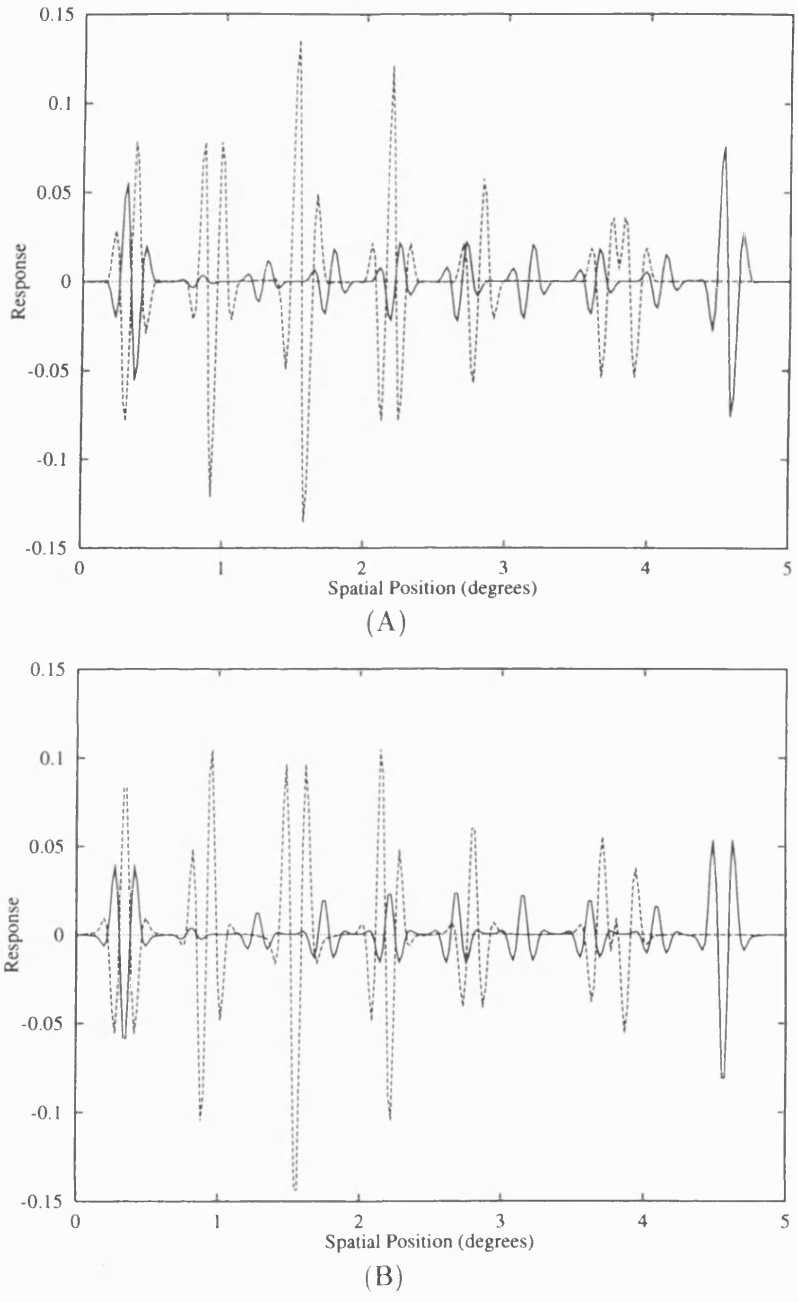


Figure 6-6: Horizontal cross sections through (A) even phased, and (B) odd phased, response maps for a simulation of simple cells selective to vertical boundaries ($f_c = 7.2$ cyc/°, $b = 1.14$ octaves). Cross sections were taken through the centre of the greyscale and black-white regions of Figure 6-5A as per Figure 4-16.

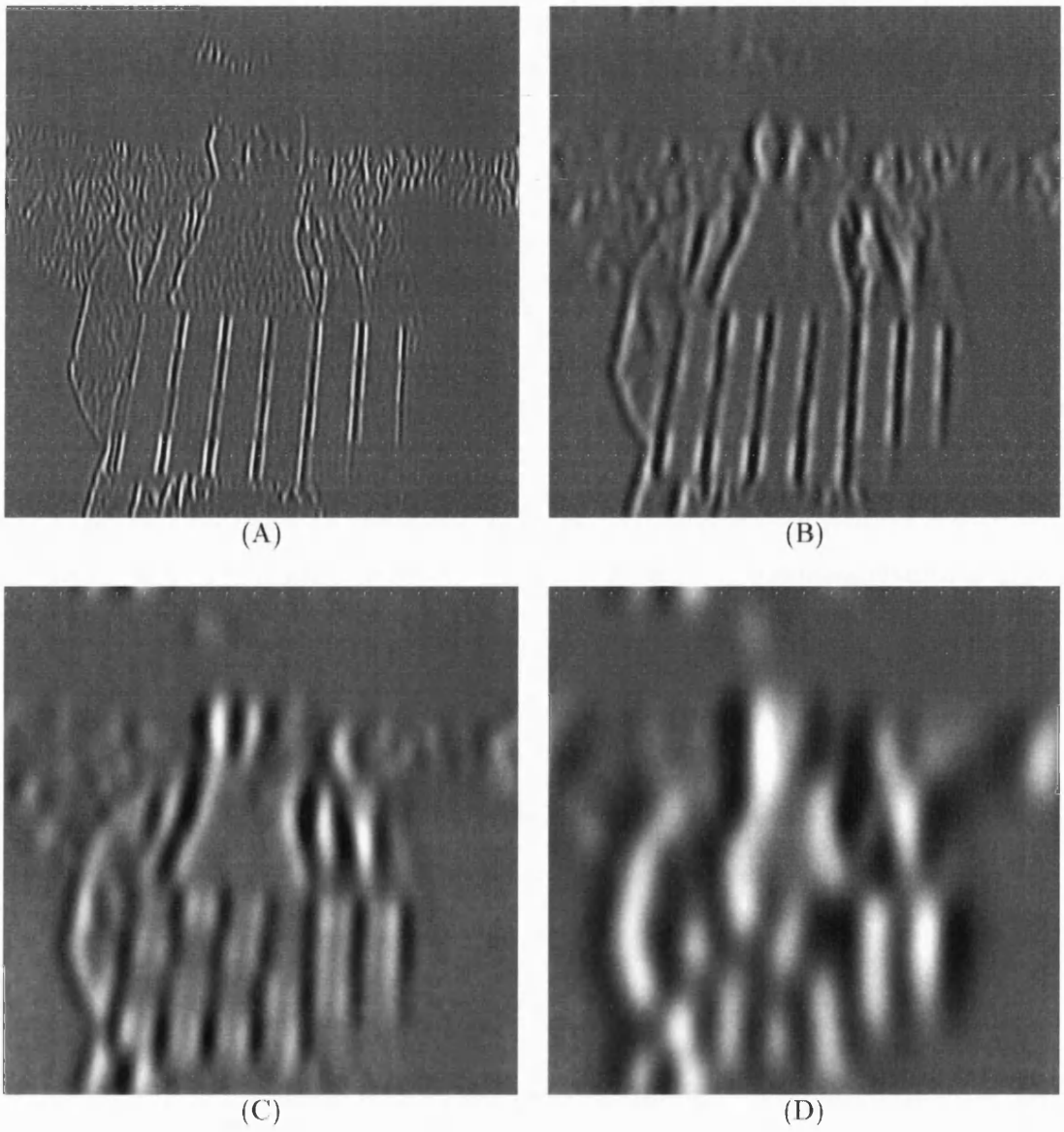


Figure 6-7: Spatial frequency channels: Simple cell responses are to the image of Figure 4-8A with $\theta = 90°$. (A) $f_c = 7.2$ cyc/°, $b = 1.14$ octaves. (B) $f_c = 2.6$ cyc/°, $b = 1.59$ octaves. (C) $f_c = 1.4$ cyc/°, $b = 1.85$ octaves. (D) $f_c = 0.7$ cyc/°, $b = 2.15$ octaves.

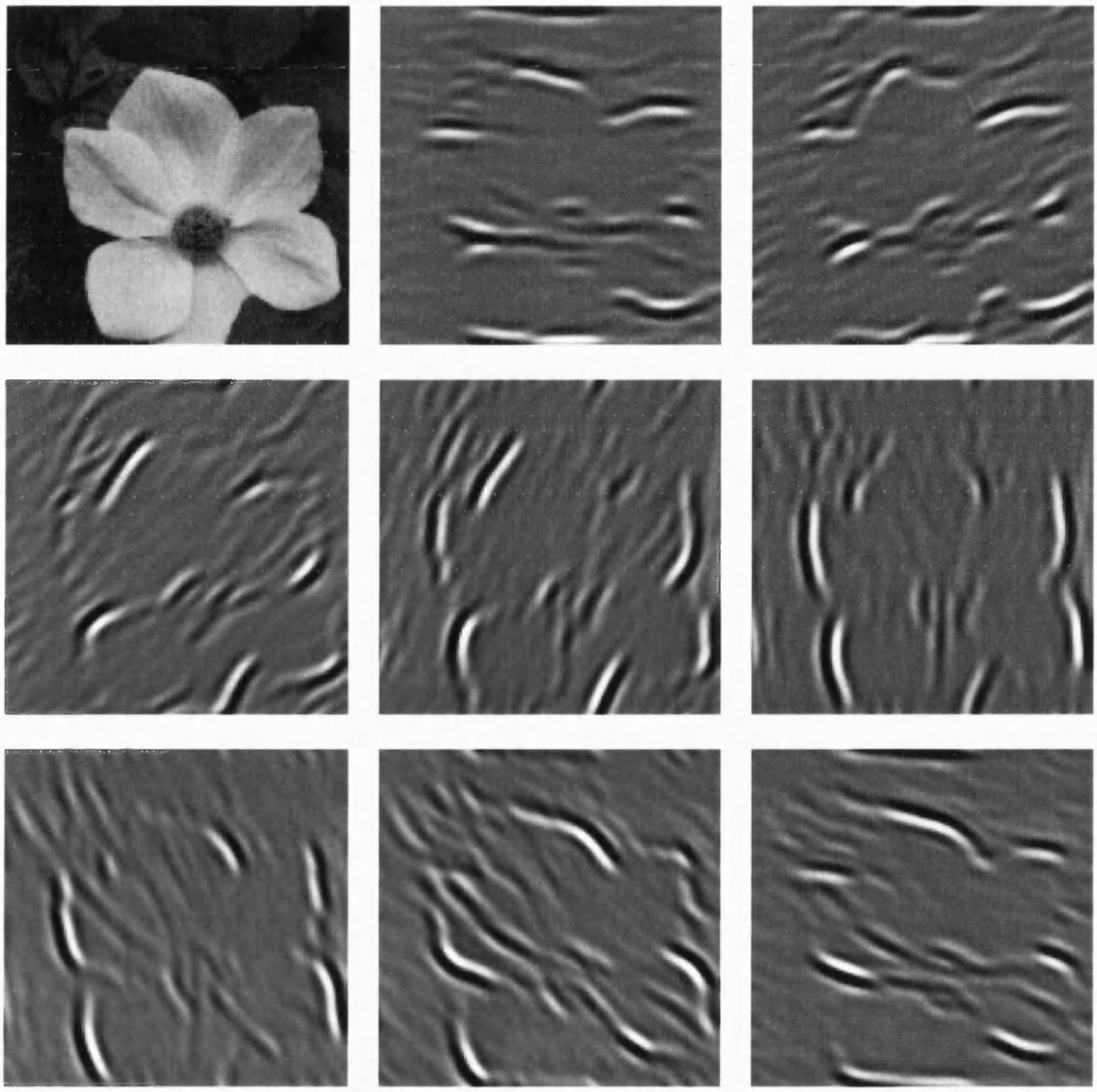


Figure 6-8: Orientation channels: Simple cell responses are to the flower picture at the top left. Orientations step round by 22.5° for each response map (f_c = 7.2 cyc/°, b = 1.14 octaves). The simulation was for a 2.5° square image.

Complex cell responses are therefore the magnitude of a vector defined by the responses of odd- and even-phased simple cells. Clearly, for sine wave stimuli, $u(x, y)$ has a constant value regardless of spatial phase. The impulse response produced by this model is a Gaussian with a width constant equal to the value of r_c defined in Equation 6.10. Convergent input from only two simple cells (or four with rectified responses) is needed to produce a single homogeneous complex receptive field.

An alternative model assumes that complex cells pool responses from simple cells over a spatially extensive region. Models using this assumption are described by Spitzer and Hochstein [221]. Since simple cell outputs actually appear rectified because the spontaneous activity in the cortex is low, response pooling of this type leads to a homogeneous excitatory region. Two equations are possible: When mixed input is obtained from simple cells with odd and even receptive fields then

$$u(x, y) = [\,|q_e(x, y)| + |q_o(x, y)|\,] \otimes G(x, y) \tag{6.13}$$

where $G(x, y)$ is the Gaussian kernel from Equation 4.7. On the other hand, when input comes from only even-phased simple cells then

$$u(x, y) = |q_e(x, y)| \otimes G(x, y). \tag{6.14}$$

At the visual field periphery, the incidence of odd-phased simple cells is thought to reduce (Hofmann and Hallett [108] comment further on this point). Complex cells at these retinal eccentricities probably take their input from even-phased simple cells according to Equation 6.14. In support of these two rectify-and-sum models is the fact that if the smoothing introduced by the Gaussian kernel is not very large, then ripple is present on the output because of the rectification implied by taking the modulus. This fits in well with the observations of Pollen and Ronner [192] who detected high frequency periodic excitation on complex cell response traces. Such ripple effects are exaggerated by aliasing if the simple cell input is not sampled at a high enough frequency. Equation 6.14 requires more spatial pooling (a larger Gaussian width constant) than Equation 6.13 because the ripple frequency is at $2f$ rather than at $4f$.

If the ripple amplitude is to be reduced to 1% by the Gaussian filtering, then the value

of Gaussian r_c needs to be set according to the formula

$$r_c = \frac{\sqrt{-\ln 0.01}}{\pi f_r}, \tag{6.15}$$

where f_r is the ripple frequency ($4f_c$ in the case of Equation 6.13).

Further support for the rectify-and-sum models comes from the fact that complex cells in V1 tend to have slightly larger receptive fields than simple cells, implying some convergence in space. Indeed, this "collection" of responses from distributed simple cell subunits results in an extra level of summation beyond that occurring within individual subunits. One can plot *length summation* curves which show how cell responses change as stimulus length increases. For complex cells, these curves will reflect the summation characteristics of both subunit and collection processes.

When stimulated by two parallel bar stimuli, the energy and rectification models are both found to produce two-bar interaction results that are similar to those described by Movshon *et al.* [174]. The rectification model, however, has a guaranteed high frequency cut-off because of the spatial pooling, so it is clear to what extent sub-sampling is possible. I shall consider sampling of these complex cell responses in Section 6.5.

6.3.2 Simulation Results

Figures 6-9 and 6-10 show the response maps for complex cells when these are derived from the simple cell outputs of Figures 6-7 and 6-8. These results were obtained by using the model defined by Equation 6.13 and differ only slightly from those obtained with the more frequently quoted local energy calculation of Equation 6.12. The main difference being that response troughs become slightly rounded by the spatial pooling stage.

It can be seen that complex cells are orientation selective and respond with a peak at isolated edges/bars (localised stimuli with a wide spatial frequency bandwidth). This behaviour may seem to suggest that complex cells perform edge-detection, but it should be remembered that they also produce a constant elevated response level to narrow-band stimuli, e.g. gratings with a frequency within the cell pass-band. When square wave grating stimuli are used, both types of response are possible: Complex cells tuned to high spatial frequencies respond only at the grating edges, while those with centre frequencies near to that of the grating fundamental produce a constant elevated response. Response maps from complex cells are therefore quite difficult to interpret. In the next section I consider representational issues raised by the simple and complex cell arrangement.

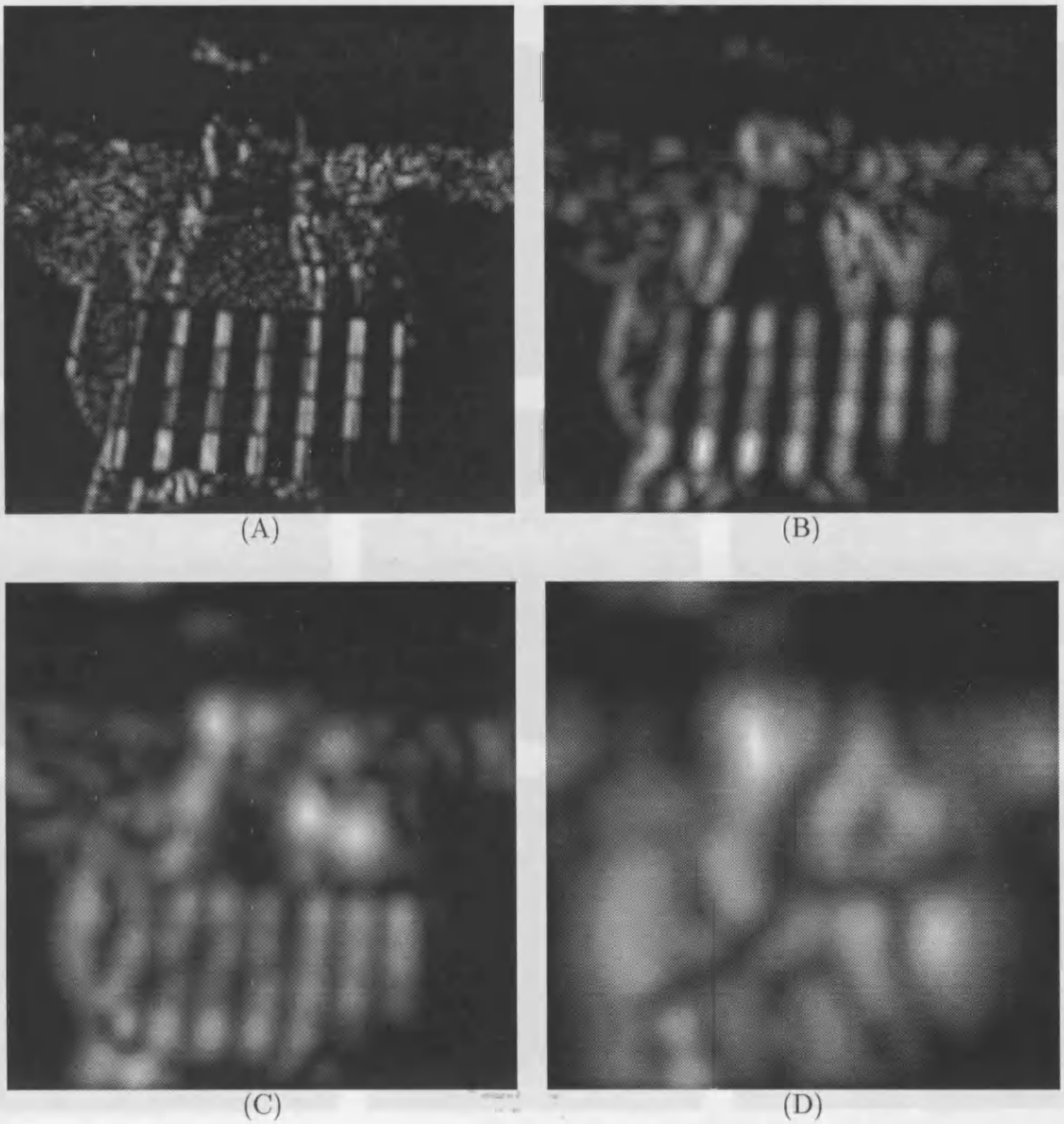


Figure 6-9: Spatial frequency channels: Complex cell responses to the image of Figure 4-8A with $\theta = 90°$. Light regions depict excitation. (A) $f_c = 7.2$ cyc/°, $b = 1.14$ octaves. (B) $f_c = 2.6$ cyc/°, $b = 1.59$ octaves. (C) $f_c = 1.4$ cyc/°, $b = 1.85$ octaves. (D) $f_c = 0.7$ cyc/°, $b = 2.15$ octaves.

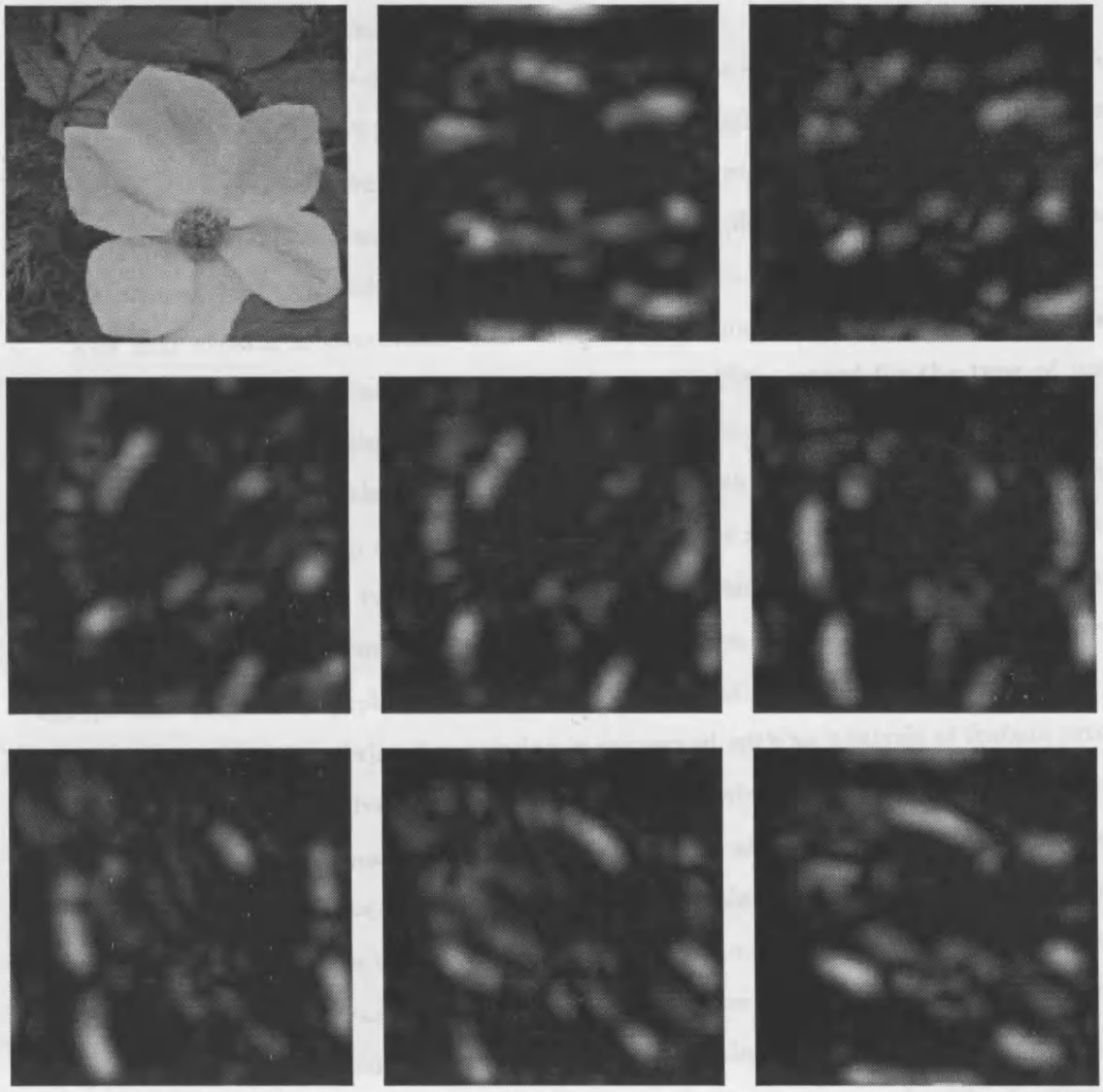


Figure 6-10: Orientation channels: Complex cell responses to the flower picture at the top left. Orientations step round by 22.5° for each response map (f_c = 7.2 cyc/°, b = 1.14 octaves). The simulation was for a 2.5° square image. Light regions depict excitation.

6.4 Cortical Representations

In this section, I aim to present an initial discussion on how the visual scene is represented in V1. Chapter 7 includes details of end-stopping and other attributes of simple and complex cells not yet covered, so this section deals only with an interpretation of the essential "linear filter" aspects of striate receptive fields.

I suggest that there are three interlinked streams of processing which arise in V1 and are subsequently elaborated in V2. These streams are concerned with (1) two dimensional contour formation, (2) one dimensional feature analysis, and (3) brightness and colour perception. In V1 they can be roughly identified with complex cells, simple cells, and unoriented cells in the cytochrome-oxidase rich blobs.

The first stream is concerned with contour formation, and eventually leads to shape recognition by extracting outlines and structures with little regard for the type of features used in their formation (edges, lines, illusory contours, motion or disparity discontinuities). In V1 and V2, I identify this process with complex cells that show little selectivity to feature type, but do respond to an oriented boundary. In V4 the same applies, except that cells can distinguish more complex two dimensional stimulus parameters (Desimone and Schein [65]). The initial stage in V1 is comparable to the edge detection processes used in computer vision, except that there is an explicit representation of orientation.

The second stream of visual processing is concerned with an analysis of feature profiles in one dimension. Cells involved in this stream need to be oriented because this analysis concerns relations between brightnesses orthogonal to the local feature orientation. In V1 the relevant neurons are phase sensitive, but in V2 they can be complex cells which show different levels of excitation to each feature type and are therefore able to distinguish between edges and bars together with their contrast sign (Peterhans and von der Heydt [190]). This stream requires a synthesis of band-limited responses across spatial scales to produce a syntactic description in terms of "edge", "bar", "grating" and other primitives concerned with one dimensional relative brightness percepts (du Buf [34]).

The third stream involves brightness and colour perceptions which need not be associated with contours and are not necessarily linked to stimulus form. Brightness and colour are zero dimensional attributes of a scene, but we are able to evaluate these over arbitrary sized regions and unconsciously experience colour constancy. These abilities undoubtably involve the unoriented receptive fields of cells which are mainly found in layers 2 and 3 of V1 (Ts'o and Gilbert [227]; Livingstone and Hubel [150]) and the complex unoriented colour-opponent

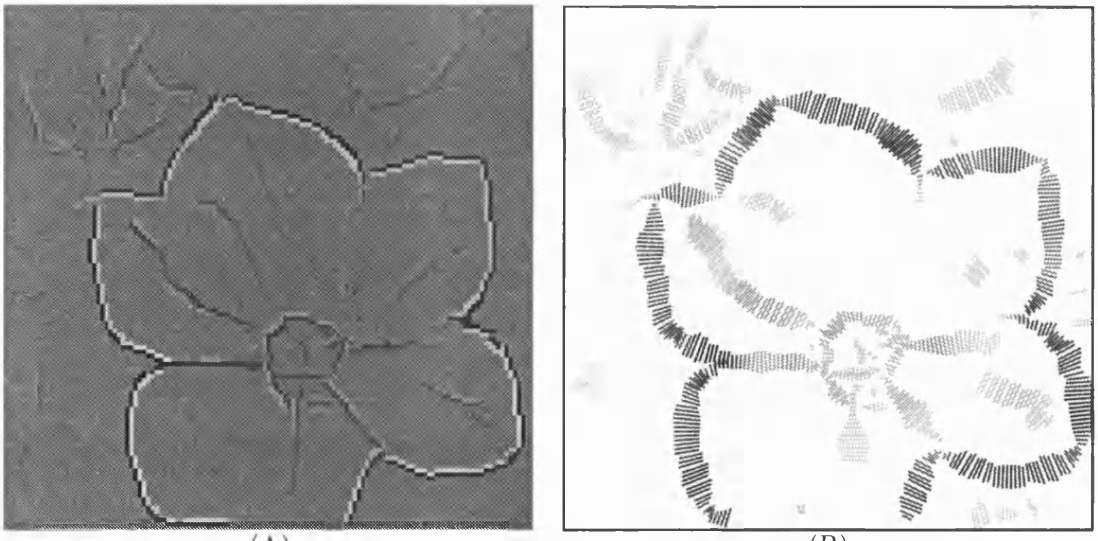


Figure 6-11: Interpretation of the complex cell responses of Figure 6-10. (A) Edges obtained by looking for response maxima in the dominant orientation channel. Underlying feature type is represented as light/dark lines (for bars) or adjacent light/dark pixels (for edges). (B) Orientation map. Lines are drawn orthogonal to local contour orientation. The length of each line represents a confidence measure and the shade represents the strength of the underlying complex cell response peak.

cells of V2 (Hubel and Livingstone [111]), and also certain cells in V4 (Zeki [252]).

I do not attempt to suggest that these three streams are isolated, because there is certainly some interaction—particularly between the feature stream and the other two. Rather, this outlook is intended to form a useful approach to the study of low level vision. The next two sections contain a brief study of the behaviour of simple and complex cells in V1 with regard to contour formation and feature analysis. Some observations regarding the colour/brightness stream are included in Chapter 8.

6.4.1 Contours and Orientation

Complex cell responses do not code for edge information. Instead, they have a maximum response to a grating of a particular spatial frequency and they cannot distinguish the "edges" in such a stimulus. For isolated step edges or impulses, which are both wide-band stimuli, complex cells do indeed produce a response maximum at the feature position, in line with Morrone and Burr [171]. Any contour grouping process which takes its input from complex cells must take into account the fact that sine wave grating stimuli can result in a homogeneous spatial region of complex cell activation across a limited set of orientation channels.

Despite this reservation, many localised features in a scene result in a response peak which can be artificially detected to provide a useful symbolic picture of the way that complex cells are responding. Figure 6-11 shows two results following interpretation of the eight complex cell maps of Figure 6-10. In Figure 6-11A, an edge map is formed by using a technique modified from Heitger *et al.* [101]. For each point in the image, this method involves (1) selecting the strongest responding orientation channel, and (2) non-maximal suppression by checking for a local peak along a line in space orthogonal to the channel orientation determined from (1). This method accurately localises features providing that they are not too close together (du Buf [33]), and has some similarity to computer vision edge detection techniques, especially that due to Canny [44] which also involves non-maximal suppression along a line orthogonal to the dominant orientation.

Here, I have extended this treatment to indicate the type of feature present at each detected location. This involves checking the even and odd simple cell responses to find out which of the even-positive, even-negative, odd-positive, odd-negative responses is greatest, and marking the map with the appropriate related feature (Venkatesh and Owens [233]; Morrone and Burr [171]).

Figure 6-11B illustrates the fact that a limited number of orientation channels can carry precise information about local orientation. The response maps from Figure 6-10 have been subject to an orientation evaluation method which I describe below. This is also used in the next chapter to demonstrate the effects of orientation-tuned surround suppression.

Multiple orientation channels imply that local orientation is coded by a population response at each point in visual space. The filter orientation bandwidth limits the number of discrete orientations that can be detected at a single location to about three—and these must be regularly spaced in the orientation domain.[4] I interpret the population code in the following way:

1. Find up to three orientation domain maxima.

2. Calculate a confidence measure by checking the adjacent two channels to find out how strong each peak is. This indicates how strongly oriented the stimulus is.

3. Use moments to interpolate for the actual orientation across the three adjacent channels for each peak present.

[4] Note that we are not normally restricted to analysing one location when perceiving the orientation of extended stimuli.

If y_i is the response of orientation channel i which has a peak, and if the channels are spaced at 22.5° intervals, then the angular offset $\delta\theta$ of the stimulus orientation from that of channel i is given by

$$\delta\theta = 22.5 \times \frac{y_{i+1} - y_{i-1}}{y_{i+1} + y_i + y_{i-1}}. \tag{6.16}$$

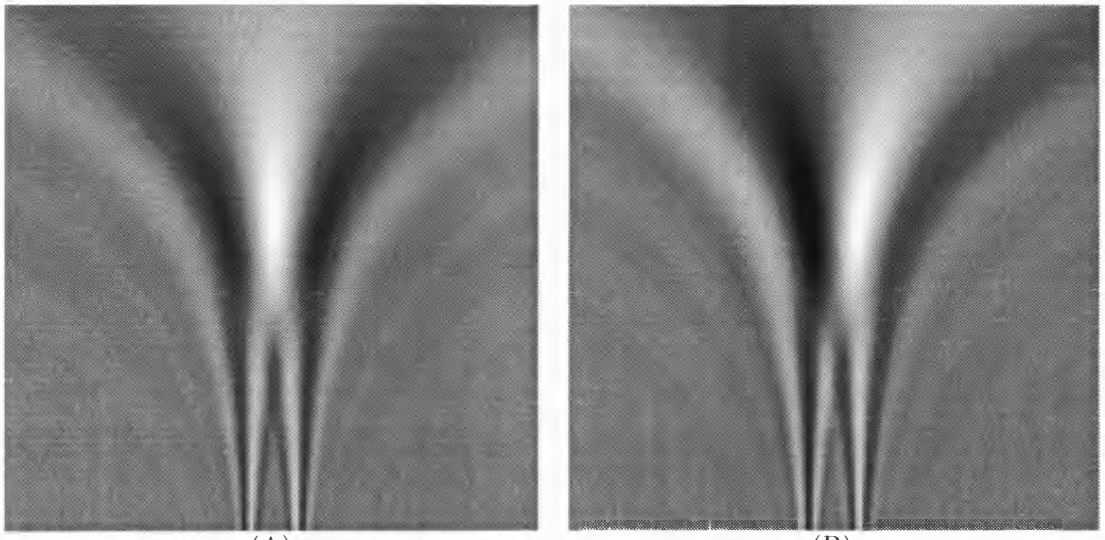
Here, moments are used because I consider the interpretation of a population response to be a "balancing act". Wilson and Humanski [248] use a quadratic interpolation to model shifts in orientation with adaptation.

In the simulation work presented here, I do not attempt to include noise or model threshold behaviour. It is interesting, however, to note in passing that psychophysical orientation discrimination thresholds are found to be in the range 0.4–0.7° and for spatial frequency discrimination, which also must rely on the interpretation of a population response, thresholds are around 0.1 octaves (Skottun *et al.* [216]). Individual striate cells have been reported that have thresholds of the same order of magnitude as behavioural studies for orientation discrimination, spatial frequency discrimination, spatial location and acuity tasks (Skottun *et al.* [216]; Bradley *et al.* [27]; Parker and Hawken [185]).

In conclusion, within the restrictions already discussed, complex cells code the position and orientation of wide-band stimuli, and the presence and orientation of narrow-band stimuli.

6.4.2 One Dimensional Features

Features are localised one dimensional relations between brightnesses, so it is possible to consider edges or bars to be high level concepts describing the behaviour of a luminance waveform without reference to the actual brightnesses involved or to the fact that the features form useful structural elements. The MIRAGE and MIDAAS models (Watt and Morgan [243]; Kingdom and Moulden [136]) attempt to describe how a particular luminance profile will appear to the human observer. This is achieved by first constructing a symbolic description in terms of edges and bars, and then using this to reconstruct the perceptual waveform. Both models work by interpreting the output of a bank of even-phased filters using an *ad hoc* collection of rules, without any kind of scale-space analysis. In this section, I review progress that has been made on understanding the scale-space behaviour of cell responses in V1 and suggest how later cortical processing may represent information about contours and features.



(A) (B)

Figure 6-12: Responses in frequency space to a one dimensional bright bar with a width of 0.625°. The x axis represents spatial position and the y axis represents spatial frequency ranging from 0.2 cyc/° at the top to 6.4 cyc/° at the bottom, in equal log intervals. (A) Responses from simple cells with even receptive fields and a bandwidth of 1 octave. (B) Responses from simple cells with odd receptive fields.

Interpretation of Simple Cell Responses

Odd and even simple cell responses can be usefully analysed using a polar representation, with magnitude given by $|\text{Even} + j\text{Odd}|$ and phase by $\arg(\text{Even} + j\text{Odd})$, where $j^2 = -1$ (du Buf [33]). For a grating stimulus, the magnitude indicates the grating contrast and the complex phase indicates the grating phase relative to the centre of the composite receptive field.

A very useful method of determining the behaviour of simple cells when presented with wide-band stimuli is to consider their responses in scale-space. By way of illustration I use one dimensional receptive fields and compute the responses to a 0.625° bar produced by cells tuned to many spatial frequencies in the range 0.2–6.4 cyc/°. Figure 6-12 shows the *frequency space* response maps that result for simple cells using the m-Gabor model. At the bottom of each map, simple cells tuned to high spatial frequencies respond at the bar edges. Just above the centre of each map, cells tuned to 0.8 cyc/° respond strongly because the bar is the optimum width to stimulate them. If the bar width was varied then the resonant peak would move up or down the diagram and the "feet" of the response would move further apart or closer together.

The resonant peak can be seen more clearly on the complex cell response map of Figure 6-

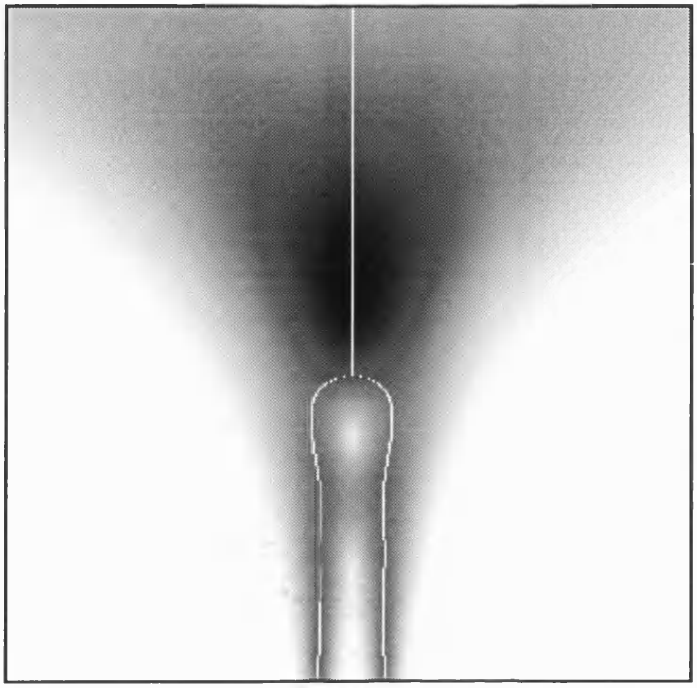
13A. A secondary peak can be seen at the centre of the lower part of the map. This is due to the third harmonic of the bar Fourier spectrum which is still distinguishable because of the narrow filter bandwidth. There are essentially two useful groups of complex cells responding to visually relevant aspects of the bar: those at high frequencies that respond well to the edges, and those at 0.8 cyc/° that indicate the size of the bar. Finding complex cell response maxima has been suggested as the way in which human vision locates features (Morrone and Burr [171]; Venkatesh and Owens [233]). On Figure 6-13A, a line is drawn to mark the feature position as determined by finding the complex cell maxima at each spatial frequency. This position is stable at high frequencies (where it marks the edges of the bar) and near 0.8 cyc/° (bar centre), but shifts around at intermediate frequencies.

Du Buf [33] pointed out that feature position can be located more accurately from the simple cell phase response if the feature type is known beforehand. Figure 6-13B shows the positions of edge or light bar phases determined from a general phase map. For high and low spatial frequencies, the feature positions determined from Figure 6-13A and B are the same, but the straight portions in Figure 6-13B, which correspond to the real features, remain stable over a wider range of frequencies. For pure sine wave stimuli, the complex cell response provides no information about feature position since there are no maxima, and so the simple cell phase is needed to mark all the feature positions. Note that in Figure 6-13B, no indication of which lines correspond to real features is present, other than that for real features the phase remains stable across a range of spatial frequencies.

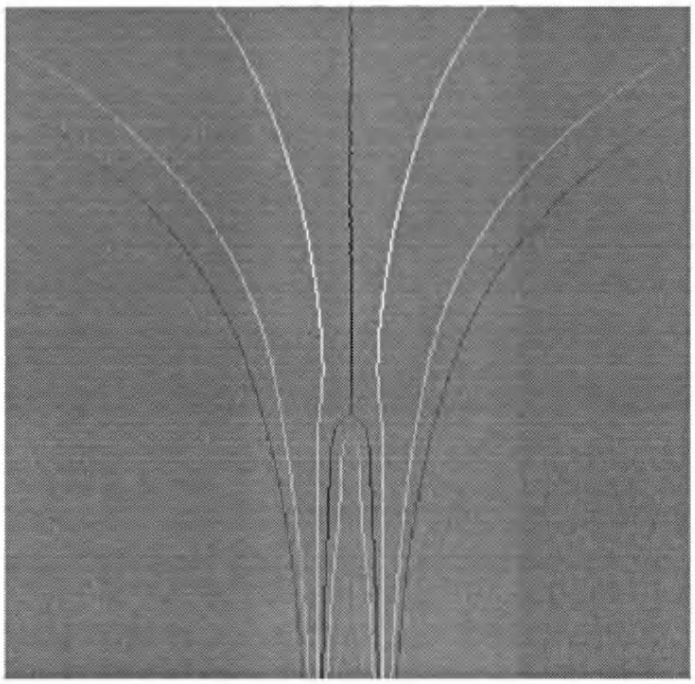
Du Buf [34] proposes a syntactical description of features in terms of Gaussian blurred line and step edge "events". In order for this to be derived from the image using an automatic data-driven process, event detection would require some sort of stability analysis to determine the reliable scales at which each feature could be extracted. The fact that neurons in V2 and V4 can distinguish between edges and bars (unlike the simple cells in Figure 6-12 which respond at both) indicates that something of this nature is actually taking place in the vision system.

From Figure 6-13A it may be seen that there is a requirement to ignore responses from cells tuned to mid-range spatial frequencies for which events are interfering. This, I suggest, may be a role for the extensive high-to-low spatial frequency inhibition that is observed in V1 and also implied by psychophysical masking experiments.

Both simple and complex cells in V1 are known to show response suppression by high spatial frequency gratings outside their excitatory bandwidth (De Valois and Tootell [55];



(A)



(B)

Figure 6-13: Marking feature positions: (A) Complex cell responses to the 0.625° light bar. Excitation is indicated by the dark region. The white line shows the position of the local response maxima at each spatial frequency. (B) Positions of light-bar phase (dark) and edge phase (light) computed from the simple cell responses in Figure 6-12. These have been scaled by the responses in A.

Albrecht and De Valois [2]) and this is a phase independent effect. Masking studies have shown that psychophysical sensitivity to test gratings with a spatial frequency one octave lower than that of a simultaneous mask grating is generally lower than when the test is one octave higher than the mask (Tolhurst and Barfield [225]; De Valois and Switkes [54]). I suggest that suppression of spatial frequency channels by those tuned one octave higher may be a necessary part of reducing the saliency of ambiguous, non feature-related responses. Such suppression would have to reduce the responses from cells which are not part of the resonant peak or high frequency response limb. This introduces a form of scale-space cuing in which complex cells interact to produce a frequency domain peak at scales for which feature primitives must be extracted. For instance, an isolated edge would produce a frequency domain peak rather than an almost flat profile, where the centre frequency of the peak is indicative of the extent of edge blur. At present, this proposal is speculative and further research is necessary.

Phase Discrimination and Appearance

Phase discrimination is the ability to distinguish between two stimuli with identical amplitude spectra on the basis of phase differences. Badcock [6] found that this ability may be based on local contrast discrimination rather than on measuring relative phase *per se*. However, relevant local features still need to be isolated so that their contrast can be compared between presentations. Burr *et al.* [36] found evidence for edge and bar detectors in human vision, in that discrimination thresholds were lower for arrival phases of 0° and 90° (edge/bar) than at 45°. Hofmann and Hallett [108] showed that a model using odd and even Gabor-like receptive fields can predict pre-attentive texture discrimination. Du Buf [29] described a similar result with application to machine vision. For brief presentations, for which focus of attention is unlikely to play a role, Caelli and Bevan [40] found that quantising the phases of texture pairs to 45° intervals did not improve discrimination performance. Precise local phase knowledge may require the presence of focus of attention. Morgan *et al.* [170] found that images were still recognisable after the phase information from local patches had been replaced by local phase information from another image. This implies that structural appearance is strongly dependent on local energy, whereas local phase is important for accurate determination of contour position and feature type.

6.4.3 Conclusions

To conclude this section on cortical representations, I would like to emphasize again the suggestion that there are three streams responsible for pre-attentive vision. These are concerned with contours, features and colour. The first is based on complex cell responses and allows for structural recognition. The second is concerned with establishing a feature description (perhaps in terms of edges and bars) drawing from simple cells tuned to a number of spatial frequencies. Neither mechanism is able to measure local phase (except where it alters the contrast of pre-attentive features) or provide spatial localisation below the size of a complex cell receptive field. For these tasks we require focus of attention and a prior knowledge of the feature type. Future research would need to concentrate on working out a reliable data-driven method of extracting stable edge and bar primitives from information coded in the cell responses of V1.

6.5 Spatial Sampling

In this section I introduce a simple cell spatial sampling criterion which shows that it is possible to have a large apparent under-sampling of visual space by these cells without information loss, providing that both odd and even receptive fields are present. This technique is novel within the field of neurophysiology but is extensively used in telecommunications engineering. Appendix A provides a mathematical proof of the theory.

6.5.1 Theory

To demonstrate this theory, I consider the one-dimensional case. This simplification is valid because simple cells show a band-pass response only for stimuli that are correctly oriented. Conventionally, if simple cell responses were to be sampled (orthogonally to the orientation axis of the receptive field), the highest frequency component present would dictate the sampling frequency and hence the number of cells per linear degree. Sampling below twice the Nyquist frequency would result in information loss because frequency components become mixed up and there is no way of separating them (see Appendix A).

In the case of simple cells, however, there are two things which change the situation: Firstly, these cells have even or odd receptive fields and are therefore organised in such a way as to compute two signals in phase quadrature. Secondly, the resulting signals are band-limited because of the spatial frequency band-pass filter action. This scenario allows for a

more relaxed sampling criterion.

At this point, I first give an intuitive presentation of why it is possible to subsample without information loss, and later I demonstrate this fact in a more rigourous way.

We have already seen from the previous section that together, even- and odd-phased simple cell responses can be represented using a polar coordinate system. Consider each sample to consist of the phase and magnitude responses derived from an odd/even pair. Suppose we assume that the output waveform from the simple cell filter is primarily described by a sine wave component with a frequency that is equal to the centre frequency of the filter. (This assumption is valid if the bandwidth of the filter is narrow.) If we have a single sample which tells us about the local phase and magnitude of this frequency component, then since we know the frequency, we can predict what the waveform will do at points near to the sample point.

If the filter has an appreciable bandwidth our frequency estimate may lose validity because other frequencies can be passed. In this case, for points further away in space from the sample point our estimate of the waveform will start to be in error due to uncertainty of the actual frequency. Under these circumstances, other samples will be needed to get a new fix on the local phase and magnitude of the waveform. The filter bandwidth constrains the rate at which the response waveform can change phase/frequency with position and hence sets the distance allowable between samples. If the filter were to have zero bandwidth then only one frequency would be passed, and only one sample of magnitude and phase would be necessary to define the entire waveform. (This is essentially what happens in a standard Fourier transform.) As the bandwidth increases, more samples are needed because the signal then has a greater freedom to deviate from a constant frequency. It is therefore the filter bandwidth that sets the sampling rate, not the highest frequency component present. This departure from standard sampling theory is only possible because we possess knowledge of both the magnitude and phase of the waveform at each sample position.

I now move on to a demonstration of the validity of these statements. Figure 6-14 shows the process of sampling and reconstruction of simple cell responses. Figure 6-14A shows a wide-band signal that has been band-pass filtered by even-phased simple cells. The left and right hand halves of the diagram indicate negative and positive frequencies and the slope of the top of each conceptualised response is intended to indicate the order of frequencies within the pass band—the slope rises from low to high frequencies. This has nothing to do with the actual band-pass shape of a real filter. Figure 6-14B shows the spectrum that results

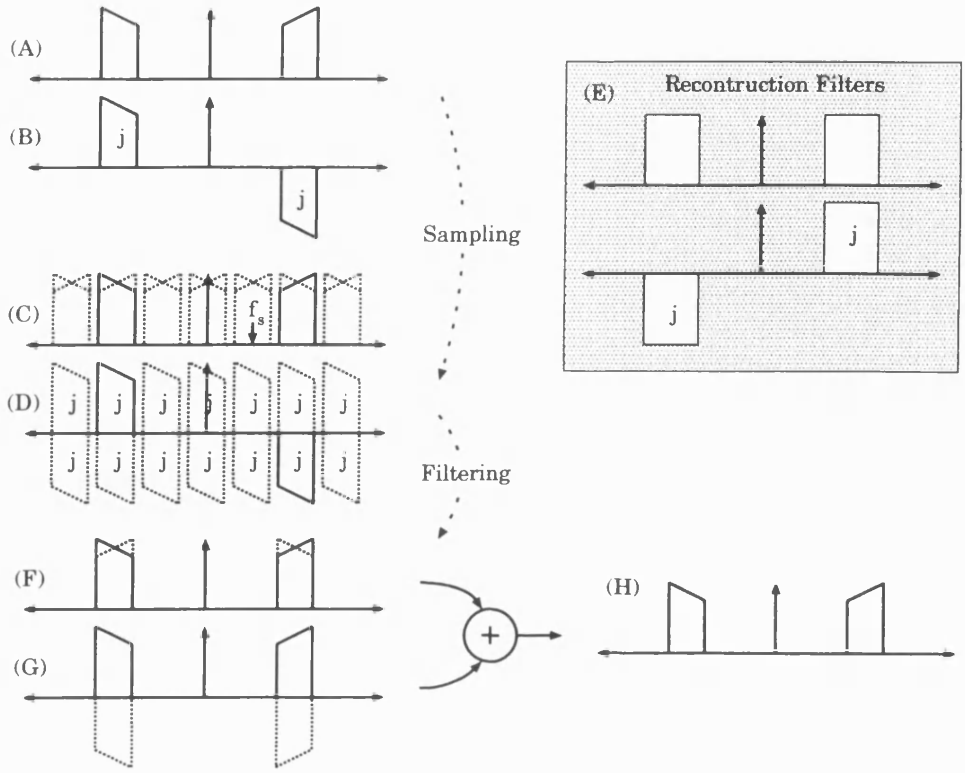


Figure 6-14: Sampling and reconstruction of simple cell responses. Even and odd responses shown in the frequency domain (A,B) are under-sampled to give (C,D). Reconstruction filters (E) then select part of the spectrum to give (F,G). When these results are added, the original band-pass filtered signal is produced (H).

from an odd-phased filter. The "j" symbols indicate the extra 90° of phase shift given to the components within the pass-band.

Next these responses are subsampled. Appendix A shows that subsampling is equivalent to convolving the frequency spectrum with a comb of impulses spaced at intervals equal to the sampling frequency. Figures 6-14C and D show the results of sampling in this way. The sampling frequency, f_s, is below the Nyquist limit, but it is chosen so that there is no overlap of the band-pass sections that arise from the same half of the frequency domain. Obviously, the responses corresponding to positive and negative frequencies interfere and there is no way of recovering the original signal by simply filtering either of Figure 6-14C or D.

In order to reconstruct the original band-pass signal (defined by Figure 6-14A), I use the reconstruction filters shown in Figure 6-14E. In the space domain, these filters are sine and cosine waves that have been modulated by a $\sin(x)/x$ function. An approximation can be produced using Gabor filters.

When these reconstruction filters are applied to Figure 6-14C and D, this results in the

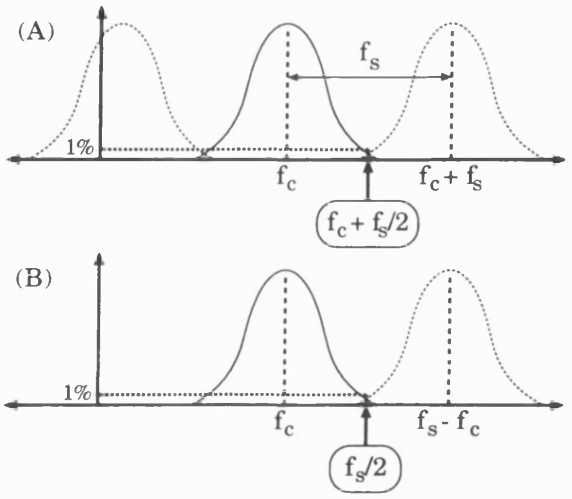


Figure 6-15: Sampling criterion. (A) Quadrature sampling. The minimum sampling frequency is equal to the minimum allowable distance between adjacent lobes in the sampled spectra. (B) Nyquist sampling. The minimum sampling frequency is twice the frequency at which the filter high frequency cutoff falls to a small value. If the filter bandwidth is small, and centre frequency high, then A will give a much lower value of f_s than B.

spectra of Figure 6-14F and G. The extra frequency components have been removed, but the negative and positive frequency lobes still interfere. When Figure 6-14F and G are added together, however, cancellation occurs and the original signal is produced (Figure 6-14H). I refer to this method of subsampling as "quadrature sampling" from now on.

6.5.2 Predictions for Cell Densities

To work out the minimum sampling frequency, I assume that the filter pass-bands are nearly symmetrical about their centre frequency f_c. In order to avoid overlap of the positive (or negative) frequency lobes, the filter response must be almost zero at $f_c + f_s/2$ (see Figure 6-15). If the filters are Gabor functions then this converts into the constraint that

$$f_s > 2f_c \frac{2^b - 1}{2^b + 1} \sqrt{\frac{\ln 100}{\ln 2}} \tag{6.17}$$

given that an attenuation to 1% at $f_s/2$ is sufficient. The same considerations applied to standard Nyquist sampling yield

$$f_s > 2f_c \left[1 + \frac{2^b - 1}{2^b + 1} \sqrt{\frac{\ln 100}{\ln 2}} \right]. \tag{6.18}$$

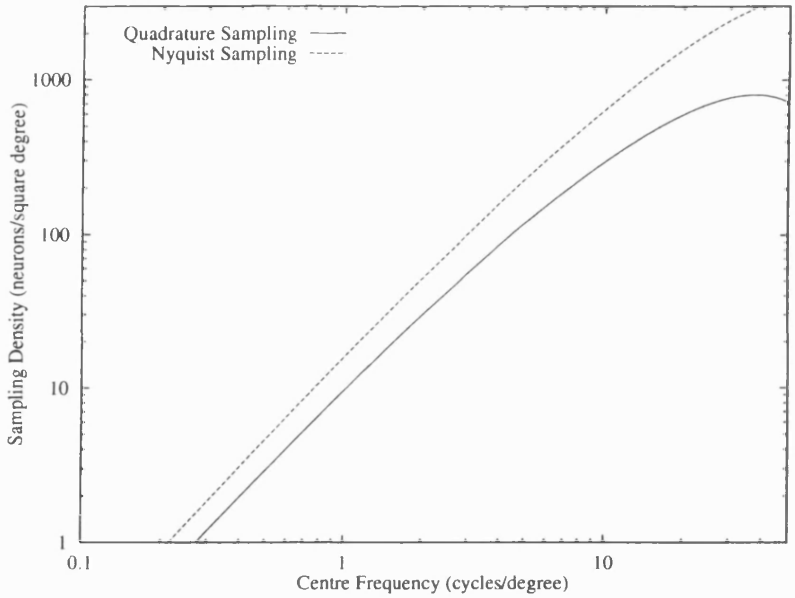


Figure 6-16: Graph showing predicted sampling densities for simple cells tuned to a single orientation under two sampling regimes.

The graph of Figure 6-16 plots the number of simple cell odd/even pairs needed to cover visual space for both types of sampling strategy. Extension to two dimensions has been achieved here by multiplying the sampling rates from the equations above by the rate required to sample parallel to the orientation axis, given the length summation properties of simple cells with an aspect ratio of one. The maximum on the lower curve is due to the effect of a slow reduction in bandwidth with spatial frequency, modelled by the equation $b = 2 - \log_{10}(f_c)$, as previously defined in Section 6.2.

6.5.3 Is Quadrature Sampling Used?

Does the vision system use the quadrature sampling method? It is difficult to give a clear answer to this. It is known that the incidence of odd-phased simple cells is low in the periphery—but the periphery is sensitive to a low range of spatial frequencies and the benefit gained from using the method is small, according to Figure 6-16. Towards the centre of vision, both odd and even-phased cells are apparent. Since high foveal acuity requires large numbers of cells, it would make engineering sense to take advantage of the more efficient coding offered by quadrature sampling as both cell phases are present there.

Both sampling methods predict that the number of cells required to represent visual space faithfully (including hyperacuities) must rise with spatial frequency, whereas in fact

cell densities found in the striate cortex of the macaque monkey peak at middle spatial frequencies (around 5-8 cyc/°). Foster *et al.* [82] provide a useful discussion on issues of sampling. They include the point that there may well be a significant over-representation of middle spatial frequencies (for which contrast sensitivity is highest in any case). It is worth noting that experimental cell populations are not normally obtained from exactly one eccentricity and so an average of parafoveal and foveal populations may well doubly include neurons covering middle frequencies but inadequately represent neurons that are tuned to high spatial frequencies.

With regards to the complexity of representation, there is no reason to rule out quadrature sampling as being too complicated. Normal sampling produces a representation that needs interpolation by an extensive $\sin(x)/x$ function if points between samples are needed. In the same way, quadrature sampling needs interpolation of the sampled even and odd responses followed by addition. In general, I believe that interpolation does not need to be performed by the vision system (in disagreement with the views of some researchers, e.g. Hirsh and Hylton [106]). Instead, abilities that seem to rely on interpolation, e.g. hyperacuities, are actually performed by means of the comparison of suitable population responses at a higher level in the vision system. Interpolation is not needed—it is enough to know that the information is faithfully coded in the representation. Parker and Hawken [185] concluded that there was enough information in individual simple cell responses to account for performance in visual resolution tasks without interpolation. In the case of quadrature sampling, I have demonstrated that the necessary information is present.

For the simulation work described here, quadrature sampling was used successfully to reduce the amount of memory required to store 224 simple cell response maps from the 20.7 megabytes required for Nyquist sampling down to 7.2 megabytes.

6.5.4 Complex Cell Response Sampling

In general, once simple cell density has been reduced by subsampling, complex cell responses should not need to be evaluated at Nyquist rates either. Since complex cells are modelled as responding to the magnitude of the simple cell quadrature pair response, then, providing that this magnitude is complemented with knowledge of spatial phase, no information is lost when subsampling.

In the rectify-and-smooth model of Equation 6.13, high spatial frequency energy is attenuated by the smoothing stage. Since complex cell receptive fields are often larger than those

of simple cells, the amount of post-rectification smoothing may well be great, and this leads to a reduction in the sampling rate required. Complex cells can always be sampled at a rate less than or equal to the rate used for quadrature sampling simple cell responses.

The only difficulty with such a low rate of sampling is in manufacturing complex responses from simple cell sub-units without an intermediate interpolation stage. For the model described by Equation 6.12, this is not a problem. For the rectification model, however, great difficulty is caused by the fact that harmonics at $4f_c$ and above, produced during rectification, wrap around to cause spatial aliasing, and this cannot be removed by later smoothing. It should be noted that this aliasing problem is present not just for quadrature sampling but for any sampling method that cannot represent such a high frequency.

6.6 Conclusions

In this chapter I introduced models for simple and complex cell receptive fields in V1. Simulation results have been presented using these models and the basic representation of the visual scene has been described. In particular, the following areas were covered:

- A modified Gabor model of simple cell receptive fields has been proposed which has all the benefits of the standard Gabor model but does not suffer from the DC response problem.
- The simulation of V1 has been based on eight orientations, seven spatial frequency channels and two phases of simple cell receptive fields At Nyquist sampling rates and for a 10° square image, this representation includes 21 million neurons.
- Local energy and rectification models for complex cell receptive fields have been described.
- I proposed that three streams of processing have their origin in V1. These are concerned with two dimensional contours, one dimensional features, and localised colour and brightness.
- With regard to contours: A method has been described for interpreting the complex cell population responses to orientation.
- With regard to features: Frequency domain responses of simple cells have been presented, and some of the problems involved in using these to create a high level feature description were discussed.

- The spatial sampling of simple and complex cells has been studied. In particular, I have presented a reduced sampling rate method for the lossless encoding of simple cell responses.

In the next chapter, I introduce complications to the linear filter picture of V1 by including the effects of contrast gain control and divisive spatial inhibition.

Chapter 7

Horizontal Interactions

7.1 Introduction

So far, I have considered only the excitatory influences that lead to the form selectivity of simple and complex cells in V1. In the preceding chapter, I introduced a linear filter model of simple cells that was based on a modified Gabor receptive field profile and this yielded basic orientation and spatial frequency selectivity. Simple cell responses were then combined using either a local energy calculation or else a rectify-and-smooth model, to simulate standard complex cell receptive fields, employing a hierarchical scheme in the spirit of Hubel and Wiesel [114].

In this chapter I attempt to provide a more complete account of primate striate neurophysiology by concentrating on the numerous non-linear influences which act to modulate a neuron's sensitivity. The topics covered here are contrast gain control and end, side and surround-stopping.

Throughout the chapter, results are obtained from an extensive computer simulation that allows for neural interactions in the domains of space, spatial frequency and orientation. The simulation predicts the ensemble behaviour at each stage of neural processing. This analysis technique provides insight into the problems that are faced by the vision system and the solutions that are employed within V1.

Although in previous chapters I have been concerned with primate vision, a large proportion of the data discussed in this chapter and the next originates from studies on the cat. This is unavoidable because few quantitative reports are available for the monkey. I therefore use receptive field parameters suitable for modelling primate vision and view results obtained

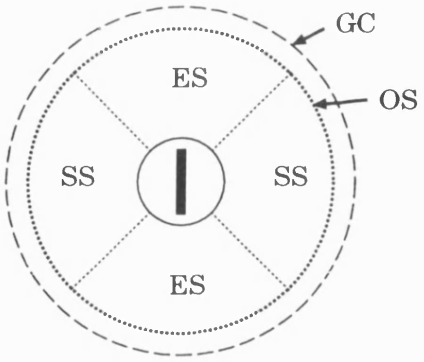


Figure 7-1: Inhibitory regions surrounding a receptive field centre which is selective for vertically oriented patterns. ES: end-stopping region, SS: side-stopping region, OS: surround-stopping region, which includes both ES and SS, GC: contrast gain control region. Region GC is the only one to overlap the receptive field centre.

from the cat as applicable to both animals where there is no clear evidence for a species difference.

7.1.1 Background

Many of the modulatory influences on a cell's receptive field are inhibitory. Studies have shown that this inhibition is generally divisive, with extra stimuli introducing a change in the contrast gain of the central response mechanism (Morrone *et al.* [172]; Heeger [100]; Bonds [24]; DeAngelis *et al.* [61]). Inhibition can arise from inside or outside the receptive field, and from stimuli at various spatial frequencies, orientations and spatial phases.

Figure 7-1 shows the regions of visual space from which inhibition can originate, relative to the excitatory receptive field. *End-stopping* suppression results from extending the length of the stimulus along the cell's orientation axis so that it invades the inhibitory end-zones marked "ES". *Side-stopping* results from increasing the width (number of cycles) of a grating stimulus until the regions marked "SS" are encountered, whereupon the cell's responsiveness is progressively reduced, or abolished. Some cells are *surround-stopped*, that is, they are inhibited by large grating patches that include the annular ring marked "OS", but are excited by stimuli confined to the central receptive field. These three types of inhibition appear to have minimum overlap with the receptive field centre and are broadly tuned for both spatial frequency and orientation, with maximum inhibitory effect coincident with the excitation tuning curves of the centre mechanism (DeAngelis *et al.* [61]).

Non stimulus specific inhibition is strongly associated with contrast gain control and arises

throughout the entire region shown in Figure 7-1, including the excitatory centre. The spatial extent and details of this type of inhibition are difficult to determine, but it appears that such suppression should be related to contrast only and should not show stimulus selectivity (Heeger [100]). In the orientation domain, Morrone *et al.* [172] reported that suppression is similar for *all* orientations of an inhibitory stimulus outside the excitatory orientation tuning bandwidth, and DeAngelis *et al.* [61] stated that this is also the case when such stimuli are confined within the otherwise excitatory receptive field centre.

Evidence for the interactions described above comes from a variety of places: Rose [198] studied length summation and found some cells that were suppressed and some that were excited by bar stimuli extending outside the receptive field (when the size of the field was defined by the completion of the first stages of length summation). Gilbert [85] showed that both simple and complex cells can be end-stopped and that end-stopping can vary in strength.

Foster *et al.* [82] found that neurons in macaque V1 are suppressed by grating stimuli confined to the regions outside the conventional receptive field, and this suppression is broadly tuned for spatial frequency. Albrecht and De Valois [2] and De Valois and Tootell [55] reported that a cell's response could be reduced by the simultaneous presence of spatial frequencies outside its excitatory tuning range. The strongest inhibition was found to arise from higher spatial frequencies. Bauman and Bonds [14] reported a similar type of spatial frequency domain suppression that is strongest for frequencies that are at the limit of a cell's excitatory tuning range.

Morrone *et al.* [172] demonstrated that cells are inhibited by stimuli over a broad range of spatial frequencies and the full range of orientations. They showed that cross-orientation inhibition is most effective when produced by broad-band stimuli, e.g. noise bars, and the phase of the inhibitory stimulus is not important.

Knierim and Van Essen [137] investigated the effect of static textured surrounds on the responsiveness of neurons which were otherwise receiving optimal stimulation. They found that surrounds consisting of oriented segments produced a suppressive effect that increased with texture density. This suppression was often maximal when the surround segments matched the centre element in orientation.

DeAngelis *et al.* [61] studied the influence of grating stimuli placed in the receptive field surround. They found four distinct classes of neurons: those that were not affected by surrounding stimuli, and those that showed side-stopping, surround-stopping or end-stopping, as described above. They reported that inhibitory influences, when present, were divisive,

involving a change in contrast semi-saturation constant; were independent of spatial phase; were broadly orientation selective, with peak suppression centred at the same orientation as the peak excitatory response of the centre mechanism; and were broadly spatial frequency selective. In addition, neurons generally showed similar length and width tuning curves for both eyes and suppression was found to be binocular.

7.1.2 Two Mechanisms

In this chapter, I separate the effects described above into two mechanisms:

The "stopping" mechanism is concerned with producing new stimulus selectivities in a subset of neurons, beyond those implied by the excitatory receptive field. This involves inhibition which is (1) confined to the surround quadrants, (2) broadly tuned for orientation, (3) broadly tuned for spatial frequency, (4) phase independent.

The "gain control" mechanism is concerned with adjusting the sensitivity of the excitatory field to obtain the best contrast dynamic range. This involves inhibition which (1) overlaps the centre (and can thus be isolated using suitable stimuli confined to the centre) but includes the surrounding regions to an unspecified distance, (2) is not selective for orientation, (3) is broadly tuned for spatial frequency, (4) is phase independent. In addition, for the correct contrast gain control effect, as we shall see later, this inhibitory signal should be sensitive to small spots as well as to extended stimuli.

Such a division into two mechanisms is primarily for conceptual convenience because it is possible that the same underlying interactions produce both types of suppression and the difference may just be a change in connectivity weights with position. In addition, the mechanism described as being responsible for contrast gain control may also be responsible for narrowing the excitatory spatial frequency tuning functions for some cells by frequency domain "side-suppression" as described by Bauman and Bonds [14].

7.2 Contrast Gain Control

Contrast becomes increasingly less important as visual processing proceeds to higher levels. We recognise a shape providing it is defined with sufficient contrast to be seen. In area MT, Sclar *et al.* [209] found that response *vs* contrast curves saturate more quickly and at lower contrasts than in V1, implying a progressive shedding of information about contrast. The most important point to notice is that if the response ratios between neurons having different selectivities remain the same, then the population response will keep the same

shape as contrast varies. It is not important that the population response as a whole scales compressively with contrast, and this may even be an advantage because contrast compression reduces the response dynamic range.

In V1 (and also in other areas), the relation connecting grating contrast to neural response can be well characterised by the Naka-Rushton function (see also Equation 4.5) given by

$$R = R_{max}\frac{C^n}{C^n + C_{50}^n}, \tag{7.1}$$

where C is contrast, R_{max} is the maximum discharge rate, C_{50} is the half-saturation constant and n is an exponent which determines the steepness of the curve on log-log axes. Sclar *et al.* [209] found that in V1, average values of C_{50} and n were 33% and 2.4 respectively. The average curve therefore starts off looking like $C^{2.4}$ at low contrasts, appears linear for contrasts up to about 35% but also appears logarithmic around the half-saturation value. This goes some way towards explaining why the relation has been variously described as linear with a threshold (Dean [60]) or logarithmic (Maffei and Fiorentini [153]; Skottun *et al.* [216]).

Many neurons in the striate cortex receive input from parvocellular geniculate neurons which have responses that do not saturate with contrast in the manner described, and this suggests that a simple linear filter is not sufficient to model the cortical stage. Such a filter can account for the spatial frequency and orientation selectivity possessed by simple cells, but this filter cannot then be followed by a saturating non-linearity without causing the spatial frequency and orientation tuning functions to vary in shape with contrast. Real simple and complex cells do not show this variation (Skottun *et al.* [216]). This leads to the conclusion that, in order to behave in the way they do, real neurons must be influenced by their neighbours along all selectivity dimensions so that the linearity of local response relations is preserved. In particular, this influence is thought to take the form of gain adjustments.

I have already described a model for contrast gain control operating among M-cells in the primate retina. This model included a divisive influence from neighbouring cells. Here, a similar principle is used, except that the number of selectivity dimensions is greatly increased.

7.2.1 Models of Contrast Gain Control

Ohzawa *et al.* [184] first applied the term "contrast gain control" to the cat's striate cortex, after observing that the response *vs* contrast curves for neurons in this area tended to shift laterally along the log-contrast axis in such a way as to centre on the adapting contrast.

Bonds [24] reported similar results when investigating the temporal dynamics of contrast gain adjustments. The semi-saturation constant of a cell's response *vs* contrast relation was found to increase when the stimulus history had included high contrasts, and decrease when the stimulus history included low contrasts, indicating a change in gain to reflect prevailing stimulus conditions. This effect was not seen in the LGN. In addition, gain changes were found to depend on local contrast and not on the response level of the neuron under observation. Simple and complex cells were similarly affected.

Heeger [100] put forward a model of contrast gain control which involved the normalisation of neuronal responses against contrast. In this model, responses from simple cells are combined by half-squaring to create complex cells responding to stimulus energy. This signal is then pooled across all orientations, space and across a three octave spatial frequency range. The pooled signal is then either fed forward or fed back to divide the simple cell responses and Heeger's theoretical treatment uses a feed-forward arrangement. Because the pooled signal is not stimulus specific, responses from differently selective neurons are divided by the same signal and therefore their response ratios remain constant even though the contrast response becomes compressive.

Wilson and Humanski [248] have developed a slightly different theoretical model of this process operating in the orientation domain: Neurons apply an oriented linear filter to input from the geniculate nucleus and the output of each filter passes through a compressive Naka-Rushton function. Responses pooled over a local subset of orientation channels are fed back via interneurons to apply divisive pre-synaptic inhibition. This divisive feedback has three principle effects: it keeps the response confined to the linear region of the output function; it produces a saturating response to contrast because division reduces the cell's gain as contrast increases; it narrows the orientation tuning of the cell. Wilson and Humanski's model was developed to explain the fact that the psychophysical contrast increment threshold curve reduces in slope (on log-log axes) following prolonged adaptation to grating stimuli. Their model separates changes that occur as a result of synaptic modification caused by adaptation, from fast-acting gain adjustments that involve dynamic properties of the immediate network.

Neither Heeger [100] nor Wilson and Humanski [248] extended their models to include the space domain. Here I employ both feed-forward and feedback gain control models in the spirit of Wilson and Humanski, extending their treatment to cover the domains of orientation, space, spatial frequency and edge/bar phase.

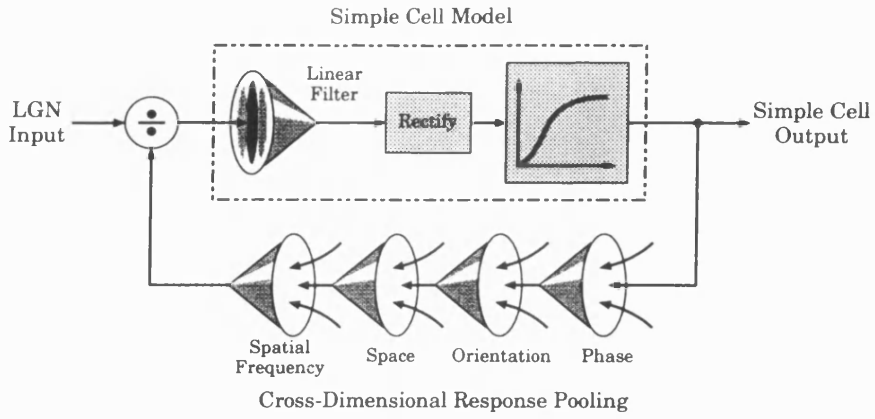


Figure 7-2: Contrast gain control model. Input from the LGN is linearly filtered by simple cells to produce stimulus selectivity. Responses are pooled between neighbouring cells with various selectivities and are fed back to divide the input. This results in contrast normalisation and maximises the response dynamic range of the cell assembly.

7.2.2 Theory

Figure 7-2 illustrates the basic features of the gain control network. In the model employed here, the non-specific control signal is formed for each spatial frequency channel by: combining outputs from simple cells having even and odd receptive fields using the complex cell equation (Equation 6.12); summing this signal for all eight simulated orientations; pooling over a local region of space, thereby combining outputs from different neurons having receptive fields that are nearby; and summing the result over five spatial frequency bands (about 3 octaves) centred on the channel in question. This scheme differs from that used by Wilson and Humanski [248] in that all orientation channels contribute equally to the pooled signal. This change is made because several researchers have reported that cross-orientation inhibition is not orientation tuned (Morrone *et al.* [172]; DeAngelis *et al.* [61]).

I start with response maps for simple cells having receptive fields with even or odd symmetry, as as defined by Equations 6.8 and 6.9. These are then subject to half wave rectification to give the generalised simple cell response map $q_{ijk}(x,y)$ where i, j and k index the four phases, the eight orientations and the seven frequency channels respectively. Division is then introduced, and the result is passed through a Naka-Rushton function to give the normalised outputs, $\overline{q}_{ijk}(x,y)$, according to the formula

$$\overline{q}_{ijk}(x,y) = \frac{R_{max}\left(\frac{q_{ijk}(x,y)}{1+\beta\, c_k(x,y)}\right)^n}{\left(\frac{q_{ijk}(x,y)}{1+\beta\, c_k(x,y)}\right)^n + \sigma^n}. \tag{7.2}$$

This can be rearranged to give

$$\overline{q}_{ijk}(x,y) = \frac{R_{max}\, q_{ijk}(x,y)^n}{q_{ijk}(x,y)^n + \sigma^n \left[1 + \beta\, c_k(x,y)\right]^n}, \tag{7.3}$$

where β controls the strength of the divisive feedback. It can be seen that the effect of dividing by the contrast measure $c_k(x,y)$ is to change the half-saturation value of the Naka-Rushton function (Wilson and Humanski [248]). Note that, σ sets the maximum gain and the exponent n is assigned a value of 2.4 in line with observations (Sclar *et al.* [209]).

In the cortex, $c_k(x,y)$ is a local contrast estimate and for any one neuron it is obtained by summing responses from many nearby neurons via horizontal interconnections and passing this sum through a low-pass temporal filter. Ignoring this filter for the moment, I define the measure used to control spatial frequency channel k for the feedback case to be

$$c_k(x,y) = \sum_{m=k-2}^{k+2} \left[G_{km}(x,y) \odot \psi_{km} \sum_{j=0}^{7} \left[\sum_{i=0}^{3} \overline{q}_{ijm}(x,y)^2 \right]^{1/2} \right]. \tag{7.4}$$

The function $G_{km}(x,y)$ is the Gaussian kernel from Equation 4.7, and has a space constant r_c that depends on k and m. The operator $\odot$ is the general purpose summation operator, as defined in Equation 3.2, but is initially considered to be $\odot_1$ which provides linear spatial summation, i.e. convolution. Constant ψ_{km} is used to weight each spatial frequency channel prior to summation in the spatial frequency domain. This is needed to adjust for the different channel orientation bandwidths and contrast sensitivities. For the feed-forward case, Equation 7.4 is driven by q (the inputs) instead of $\overline{q}$ (the outputs) and the low-pass temporal filter is not actually necessary.

7.2.3 Implementation

The contrast gain control stage was implemented by first of all forming complex cell response maps (thus removing sensitivity to spatial phase), then adding together maps for neurons having different orientation preferences. This produces one control map per spatial frequency channel. These are then processed to introduce pooling in the space domain, and the resultants are combined in the frequency domain to give seven final control maps which are used to adjust the gain of the simple cells according to Equation 7.3. Figure 7-3 illustrates this process.

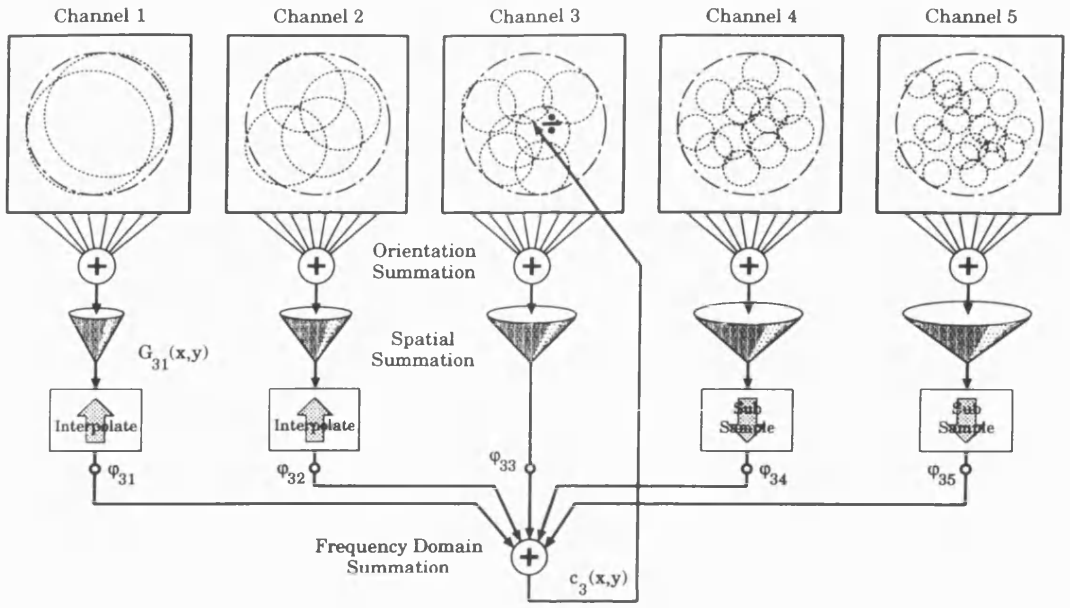


Figure 7-3: Generating the gain control signal. Complex cell response maps (top) are combined to give one control signal per channel ($c_3(x, y)$ is shown). Summation over space is designed so that the same region contributes to the control signal, regardless of the underlying receptive field size. A random selection of receptive field outlines is shown for each channel (dotted circles).

Orientation Summation

When adding together maps in the orientation domain a weighting factor is required to offset the effects of differences in orientation bandwidth between channels. Grating stimuli excite more orientation units in a spatial frequency channel with wide orientation bandwidth, so adding their responses produces a larger gain control measure than for a channel with narrow orientation tuning. Tuning curves are assumed to have a Gaussian shape, and an adjustment is incorporated into factor ψ_{km} to equalise the gain control responses to grating stimuli across all seven channels.

Spatial Summation

Following orientation domain pooling, spatial summation is introduced to make the gain control region larger than the receptive field size. For each spatial frequency channel, this summation region is made three times larger than the receptive field itself. This is achieved in the first instance by convolution with a Gaussian kernel. The size ratio of three was a somewhat arbitrary decision based on the limited information available (see Heeger [100]). Further discussion of this point is deferred until Section 7.2.5.

Two kinds of summation were tested: Linear, with $\odot = \odot_1$ (convolution), and non-linear, with $\odot = \odot_4$. This latter type of summation is sub-additive and this means that given multiple inputs, the output is much less that if the inputs were linearly added. Using summation of this form, it is possible to reduce the impact of grating patch size on the contrast measure.

Frequency Domain Interactions

Gathering a gain control measure in the spatial frequency domain presents a number of technical difficulties because response maps are not all held at the same sampling resolution. The gain control map for each channel is constructed by summation over that channel plus its two neighbours on each side—five channels in all, representing about 3 octaves. Interpolation and subsampling are used to match response map resolutions where necessary. In addition, to create the final gain control map for each channel, spatial summation of signals from adjacent channels is carefully controlled so that the gain signal is obtained from the same sized region in each one. To achieve this, the spatial extent of the summation kernel must be less for low frequency input than for high because receptive fields increase in size as spatial frequency reduces, and therefore less spatial pooling is needed to cover the same area using input from these low frequency channels.

An additional problem when interactions across frequency are considered is the fact that the channels have different peak sensitivities. Factor ψ_{km} is used here to reduce the strength of the inhibitory gain control influence from channels with high contrast sensitivity to those with a low sensitivity. If this is not done, then strongly responding mid-range channels drive down responses from those at higher frequencies where the contrast sensitivity is already low. The sensitivity ratio between channels k and m is incorporated into ψ_{km} to counteract this effect.[1]

Feedback

Both feedback and feed-forward versions of the model have been tested. Feedback is simulated by iteration: Gain control maps are calculated from the linear simple cell inputs, $q(x, y)$. These are then used to apply gain control, resulting in the modified responses $\overline{q}(x, y)$. New gain control maps are then made from these responses and the inputs are again subject to

[1] Having said this, all results in this chapter, except for Figure 7-6, were generated with the contrast sensitivity curve "turned off", i.e. all channels had the same peak sensitivity. This was done to clarify the essential gain control behaviour.

gain control. This process is then repeated. In order to converge on a solution, and to prevent oscillation, I found it necessary to make each new gain control map equal to the average of the old one and the freshly calculated map. This is analogous to the inclusion of a temporal low-pass filter in the feedback path, and agrees with the models presented by Heeger and by Wilson and Humanski. Here, a stable response was normally reached after five iterations.

7.2.4 Simulation Results

Throughout this section, I assume that the primary purpose of contrast gain control is to produce a compressive (log-like) contrast response in the cells of V1, in order to maximise output dynamic range, without contrast-related distortion of the tuning curves that define stimulus selectivity. If this process functions correctly, there will be no difference between contrast normalised response functions (e.g. orientation tuning or length response curves) and their contrast-linear precursors. In order for this to work perfectly, the contrast feedback signal must be in no way stimulus specific. I also assume here that neurons covering distant portions of the visual field are controlled by separately evaluated (local) contrast measures. The consequence of this assumption is that neurons with receptive fields in close spatial proximity have linearly related responses that stay in a fixed ratio and can be interpreted without contrast dependence by later visual processing, whereas those that are widely separated are independently normalised and have a log response relation—they cannot be thus compared. This is not a problem if neurons with larger receptive fields that derive new structure selectivity, in V2 for instance, draw from V1 responses within each linearly related area.

Initial Results

Figure 7-4 shows the response functions that result when a full-field grating stimulus (7.2 cyc/°) of variable contrast is used to excite a complex cell. Bar stimuli produce a very similar plot. Cell responses were obtained by monitoring an individual data point in the appropriate response map, while running the full simulation. The lower two curves show the results for the feed-forward and feedback models. The theoretical cell output function is also plotted (Equation 7.1). In essence, gain control tries to move this theoretical function left and right along the log contrast axis so that the stimulus contrast always produces a response near to 50% (Ohzawa *et al.* [184]). Because the normalisation does not employ perfect division, the output function fails to track the input completely, and the slower, log-like responses are

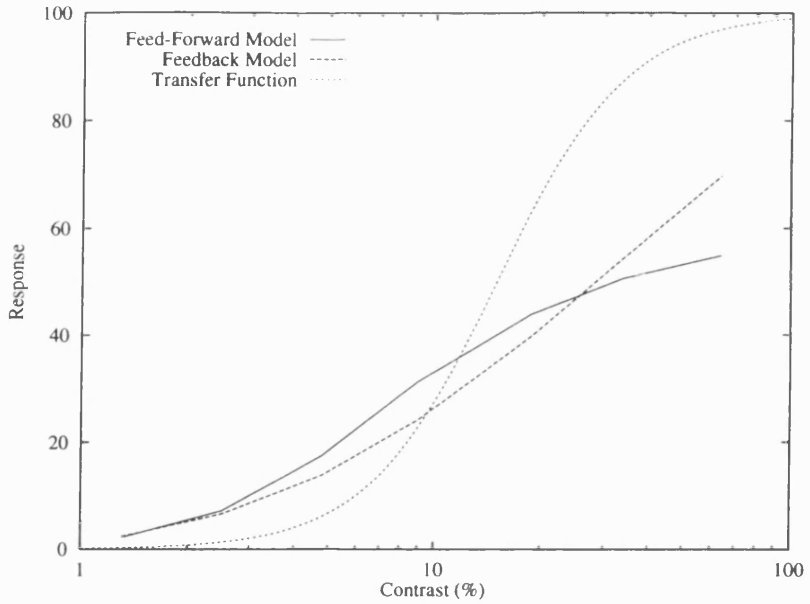


Figure 7-4: Response *vs* contrast relations under conditions of feed-forward and feedback contrast gain control ($R_{max} = 100$, $\sigma = 5$, $\beta = 0.16$). Also shown is a typical forward transfer function ($n = 2.4$, $C_{50} = 15\%$).

generated.

In Wilson and Humanski's approach, the essential role of the feedback path is to prevent the response signal from driving into the saturating region of the output function (i.e. above about 65%). This is achieved by turning the gain down to a progressively greater extent as local contrast increases. The compressive response to contrast that results is therefore not produced by the cell transfer function. The transfer function could be viewed as essentially linear with a threshold at low contrasts because the compressive region is not normally reached.

Heeger [100] provides an alternative model which gives an *overall* contrast transfer function having the form of a Naka-Rushton function. In order to do this, the model employs a feed-forward configuration which, in addition to having less neurophysiological plausibility, also relies on individual simple cells having a purely C^2 response to contrast. This relation seems undesirable because an expansive power law would appear detrimental to the local response dynamic range, and it would also prevent later summation stages from being linear, since they would then receive convergent C^2 inputs. At higher levels in the visual cortex where the contrast exponent increases, one would expect an even more extreme behaviour. I see these points as difficulties with Heeger's model and so I employ the model due to Wilson and Humanski because these authors were able to demonstrate its capability in matching

psychophysical results.

Figures 7-5 and 7-6 illustrate the effect that gain control has on simple cell responses. Responses shown are for simple cells selective for light bars (i.e. with response rectification) and it is important to note that these have been superimposed across the dimension of orientation for illustrative purposes only. Figure 7-5B shows the responses that occur using a straightforward linear model (Chapter 6). There is a wide response dynamic range and there are secondary responses to contours due to the narrow bandwidth of the filter. Figure 7-6A shows the effect of gain control operating over space, feature phase and orientation. Many of the contours now elicit similar levels of activity, but the integrity of local response relations is maintained. Here, the effect has been slightly exaggerated for illustration by using a small value of σ in Equation 7.3 and by using the feed-forward model. This results in a strong contrast compression. Some of the secondary responses are lost—an effect which is due to the "threshold"-like behaviour of the output function for low contrasts (Barlow *et al.* [10]). Figure 7-6A was generated without including response pooling in the spatial frequency domain. This results in a widening of the effective frequency selectivity bandwidth, and is visible as a response to the blurred background on the left of the original image. When spatial frequency interactions are also included then Figure 7-6B is produced. Some of the responses have now been suppressed because the spatial frequency tuning functions are restored to their correct width. In general, Figure 7-6B resembles Figure 7-5B except that there is now a compressive response to grating contrast.

Frequency Domain Analysis

Results have been gathered for a variety of model configurations: feedback or feed-forward gain control, with or without linear spatial summation, and with or without spatial frequency domain interactions. These tests provide insight into the behaviour of the model, and bring to light some problems with forming the contrast-related control signal.

Figure 7-7 summarises response levels that result in each spatial frequency channel before (row 1) and after (rows 7–11) the gain control stage for bar or grating stimuli. Also shown is the level of activity in the gain control path (rows 2–6). Summation was "turned off" for some runs by using the non-linear $\odot_4$ type which produces only a slightly larger response for stimuli covering the entire gain control region than for stimuli covering a small portion of it. This was used to probe the effects of additive spatial summation in forming the contrast measure, $c(x, y)$. Spatial frequency interactions were also turned off for some runs by ignoring

(A)

(B)

Figure 7-5: Results of simulating cortical contrast gain control mechanisms. Inputs: (A) Original bean flower image. (B) Simulated responses from simple cells sensitive to light bars using a linear filter model tuned to 7.2 cyc/°. Note that responses from eight orientations have been superimposed in this figure.

(A)

(B)

Figure 7-6: Results of simulating cortical contrast gain control. Outputs: (A) Simple cell responses following a gain control stage without spatial frequency interaction. (B) As for A, but including spatial frequency interactions over a ± 1 octave range. Note that responses from eight orientations have been superimposed in this figure.

adjacent channels when calculating the control signal (e.g. setting ψ_{31}, ψ_{32}, ψ_{34}, and ψ_{35} to zero for channel 3 in Figure 7-3).

Row 1 in both tables of Figure 7-7 gives the linear filter responses prior to gain control calibrated in contrast units. The grating stimulus (Table 1) excites channel 4 most strongly because it is the optimum stimulus for this channel. The isolated bar (Table 2) excites the same channel to a large extent, but since it is a wide-band stimulus, all channels respond to a significant extent. The linear channel responses to both stimuli are plotted in Figure 7-8.

Considering first the responses to a grating: Rows 2 and 3 show the gain control signal produced by the feed-forward model with and without additive spatial summation when spatial frequency interactions are turned off. Both rows are the same because gratings fill the gain control region, so the type of summation makes no difference.[2] When rows 2 and 3 are used to apply gain control to row 1, the results of rows 7 and 8 are obtained. These are based on a value of 100 for R_{max}. The independent normalisation in the spatial frequency domain results in a strong response from channels 3–6, rather than the narrow peak of row 1. Rows 4 and 5 show the values of each gain control signal when spatial frequency pooling adds together the five surrounding channels.[3] Since the feed-forward control signal now has a wide bandwidth, the output responses (rows 9 and 10) are now sharply tuned for spatial frequency because all responding channels are normalised by a similar value. If the output transfer function did not possess a "threshold-like" non-linearity then the output ratios would be approximately the same as the input ratios. The behaviour is similar when the feedback model is in force, except that the output responses are slightly larger. This difference can be seen in Figure 7-4 for gratings of 64% contrast.

Moving on to Figure 7-7, Table 2: This table shows activity levels obtained for neurons responding at the centre of a narrow bar. The main reason for experimenting with a bar stimulus is that it excites only part of the gain control region, so that the type of spatial summation in the control path now becomes important. Rows 2 and 3 show the control signals obtained without spatial frequency interactions. When additive summation is used (row 3), estimates of contrast are much lower than the input because the bar only activates about 40% of the gain control region in each channel. This results in outputs that are saturated (row 8). Without linear summation (row 2), control signal levels are higher, but

[2]The differences between between rows 1 and 2 are caused by the assumption of Gaussian orientation tuning when computing the orientation bandwidth correction factor.

[3]The result is also scaled by a constant factor to maintain an approximate 1:1 grating contrast transfer over all channels. This factor can be incorporated into the value of β.

Table 1 Grating (7.2 cyc/°, 64% contrast)

	Channel:	0	1	2	3	4	5	6	Sum	SFI
1.	Input	0.1	0.1	0.2	15.3	64.0	33.6	5.5	-	-
2.	Feed-	0.1	0.1	0.2	11.4	63.0	39.7	7.0	N	N
3.	Forward	0.1	0.1	0.2	11.4	63.0	39.7	7.0	Y	N
4.		0.2	6.5	41.5	63.5	67.4	67.3	60.9	N	Y
5.		0.2	6.5	41.5	63.5	67.4	67.3	60.9	Y	Y
6.	Feedback	0.0	1.8	34.9	50.0	50.3	50.4	48.5	Y	Y
7.	Output FF	0.0	0.0	0.0	54.8	58.6	44.6	17.2	N	N
8.		0.0	0.0	0.0	54.8	58.6	44.6	17.2	Y	N
9.		0.0	0.0	0.0	4.3	54.9	20.7	0.4	N	Y
10.		0.0	0.0	0.0	4.3	54.9	20.7	0.4	Y	Y
11.	Output FB	0.0	0.0	0.0	7.0	69.5	33.0	0.7	Y	Y

Table 2 Bar (0.06° wide × 5° long, 64% contrast)

	Channel:	0	1	2	3	4	5	6	Sum	SFI
1.	Input	14.6	26.4	43.0	60.2	70.7	61.9	28.2	-	-
2.	Feed-	10.5	19.7	33.1	47.8	57.8	53.1	28.4	N	N
3.	Forward	5.6	10.2	16.8	23.6	28.4	26.7	16.4	Y	N
4.		23.4	55.2	88.9	115.5	124.9	109.1	82.1	N	Y
5.		9.6	22.0	42.1	58.7	69.9	65.0	50.7	Y	Y
6.	Feedback	12.2	26.3	49.6	52.9	55.0	52.2	40.9	Y	Y
7.	Output FF	55.4	64.1	67.8	68.9	68.4	65.5	52.9	N	N
8.		74.0	84.1	88.4	90.2	90.5	88.6	75.7	Y	N
9.		16.8	18.4	20.2	24.0	29.9	27.7	10.9	N	Y
10.		58.4	59.2	56.3	58.8	58.9	55.0	25.9	Y	Y
11.	Output FB	49.3	50.7	47.7	64.1	70.7	66.3	35.6	Y	Y

Figure 7-7: Simulation results: Responses are shown for bar or grating stimuli, using five different gain control configurations. The model employs either feed-forward (FF) or feedback connections (FB), spatial summation (Sum) is either linear (Y) or non-linear (N), and spatial frequency interactions (SFI) are either present (Y) or turned off (N). Seven spatial frequency channels are shown, numbered 0–6. These have centre frequencies of 0.7, 1.4, 2.6, 4.4, 7.2, 10.5 and 15.0 cycles per degree, and half-height bandwidths of 2.15, 1.85, 1.59, 1.36, 1.14, 0.98 and 0.82 octaves respectively. Row 1 refers to contrast in each linear filter channel which provides input to the gain control mechanism. Rows 2–6 present levels produced in the gain control feed-forward/feedback path, while rows 7–11 indicate the (complex cell) output level ($R_{max} = 100$, $\sigma = 5.0$, $\beta = 0.16$). A note on the stimuli: The bar and grating both had the same peak and trough luminances. Within a neuron's limited RF, the bar looks like a section of square wave grating. Since the fundemental of this square wave grating has a higher contrast than the sine wave grating, it produces a greater channel response, even though, globally, the bar has less energy.

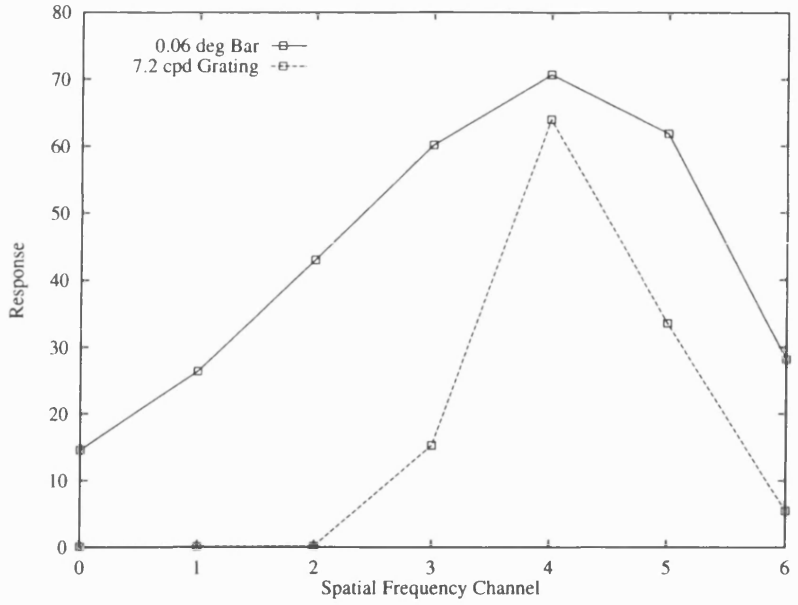


Figure 7-8: Linear spatial frequency channel responses given bar or grating stimuli.

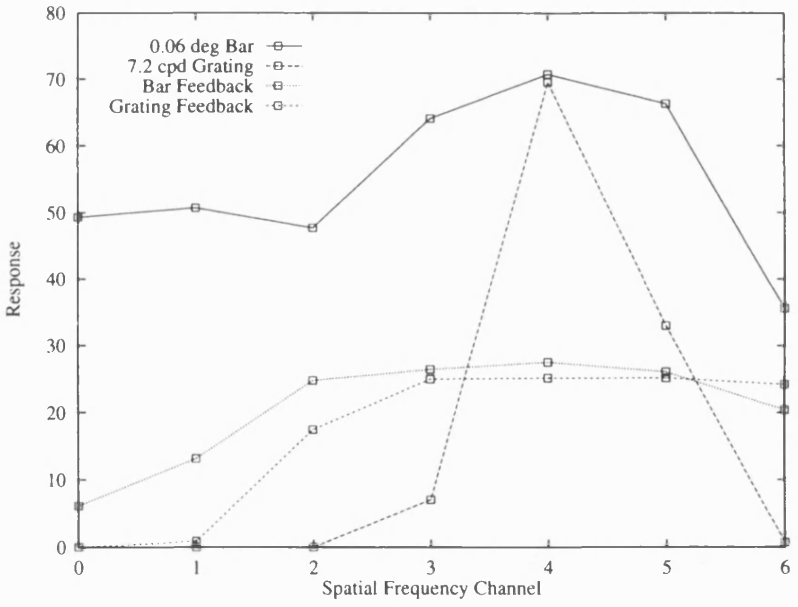


Figure 7-9: Spatial frequency channel responses given bar or grating stimuli following gain control (64% contrast, feedback model, linear spatial summation). Also shown are the pooled feedback signals, which have a wide frequency bandwidth (these have been scaled down by a factor of 2 to avoid cluttering the plot).

since we are using $\odot_4$ not $\odot_\infty$, there is still some reduction below the level of the input. When frequency domain interactions are included, row 4 has a very strong signal whereas the results of row 5 have risen to a level more appropriate for setting the operating point of the neural output when compared with row 3. In particular, because the grating and bar stimuli were chosen to have quite similar contrasts, the gain control signals produced for bars and for gratings should ideally be similar—the system should be gathering a contrast measure which is independent of stimulus form. What is happening, with regard to the row 5 signal, is that the reduction introduced by the effects of summation over space is to some extent being cancelled by summation in the frequency domain. Since stimuli range from being localised in space and wide-band in frequency (bar) to being localised in frequency and wide in space (grating), this reciprocal relationship can act to reduce the preference for gratings that would otherwise be introduced in the control path by having a large spatial summation region.

Figure 7-9 graphs the neural outputs for the feedback case and these can be compared with the ideal relations of Figure 7-8.[4] Figures 7-8 and 7-9 compare well, except for the channel 0 and 1 responses to the bar. These outputs are raised because the control signals are rather low (plotted in Figure 7-9). The control signals fall in channel 0 and 1 because there is a much larger ratio of receptive field size between channels tuned to low spatial frequencies than between those tuned to high frequencies. This means that to create the gain control input from channel 2 to channel 0 requires far more extensive spatial pooling than to create the input from channel 6 into channel 4. As a result, the bar poorly excites the channel 0 gain map via input from the other channels when linear spatial summation is in force.

In summary, we can see that the gain control model functions in the spatial frequency domain, but that there are also some problems which are mostly caused by linear spatial summation. The ideal solution may appear to be to reject linear summation in both space and frequency, in favour of a system which computes a max, or some other non-additive measure. There are, however, benefits to be obtained from linear summation in the frequency domain, including the rejection of wide-band noise, as described shortly. The reciprocal relationship between the size of a stimulus and its spatial frequency bandwidth does go some way towards nulling out some of the summation problems, but a way must be found to control the amount of additive pooling in the gain control region irrespective of this region's spatial extent.

[4] It should be noted that the output transfer function produces its own "distortion" by introducing an expansive and a compressive region. It is unclear how neurons downstream interpret their inputs—whether they can take this fixed output transfer function into account.

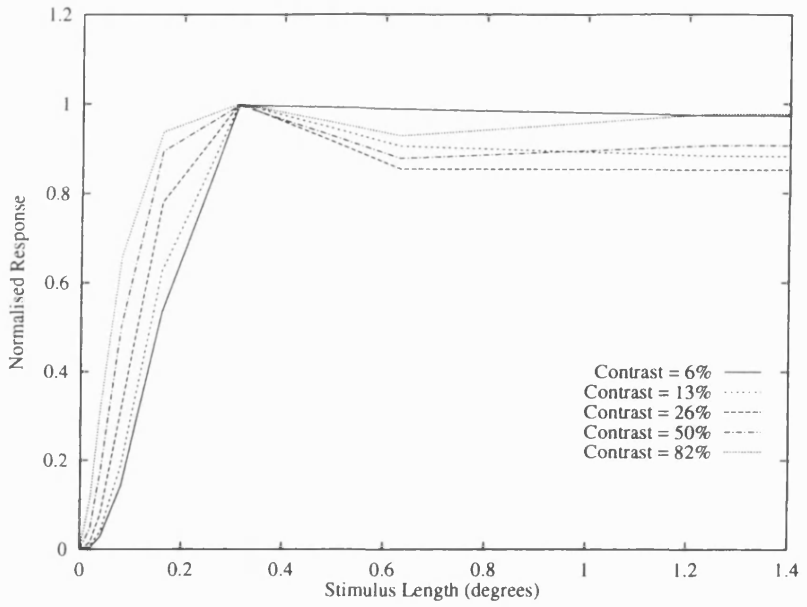


Figure 7-10: Length-response curves for complex cells under feedback gain control. As contrast increases, the model cells respond more to short stimuli. This is caused by a lack of sensitivity to such stimuli in the spatially-pooled feedback signal.

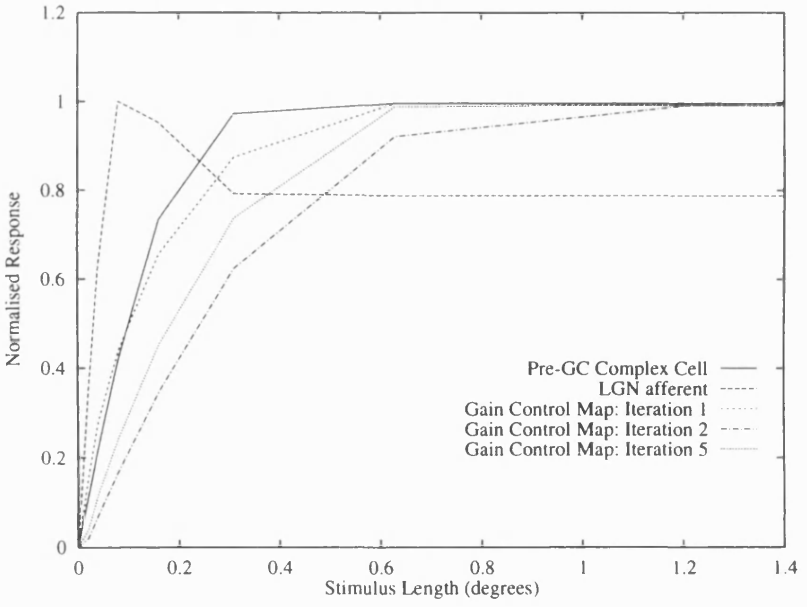


Figure 7-11: Problems with length summation: The gain control feedback signal is less sensitive to short stimuli than the cells that it controls. What is needed is a signal that sums as fast as a LGN afferent, but is sensitive over a large area.

Length Summation

In this section I introduce the length summation properties of cortical cells and show that summation in space to obtain a gain control measure introduces both end-stopping and also response saturation for short bar lengths. I consider that it is important to the judgement of size to maintain the integrity of length summation properties against contrast variations. Both these effects are therefore unwanted since they vary in magnitude with contrast. The solution is to employ a gain control measure that, although derived from a large region of space, is not selective for stimulus length.

Figure 7-10 shows response curves obtained by stimulating the channel tuned to 7.2 cyc/° with narrow bars of varying lengths and contrasts. Length summation curves are obtained from complex cells when subject to gain control and are shown normalised at their peak so that the shapes can be compared. For short bar lengths, the rate of length summation increases with contrast, and therefore the cells become more sensitive to spots. This happens because linear spatial summation in the gain control feedback path gives a contrast estimate that is too low for short bars, and so the cell output transfer function is centred at a lower contrast for short bars than it is for long bars. Contrast dependent end-stopping is also produced by a similar mechanism as demonstrated by the reduction in response at large bar lengths.

A length summation curve for a cell that is not influenced by gain control is plotted in Figure 7-11. This figure also demonstrates the essential problem by showing the length summation properties possessed by the contrast feedback signal. Since the pooling region is large, this signal integrates slowly with length—much more slowly than responses from the receptive field that it controls—and therefore the divisive contrast signal is not the same for all stimulus lengths. The best situation would be to have a gain control signal that depended on contrast, showed very fast summation with length, much like the LGN input (but without the associated end-stopping), and was also drawn from a large spatial region. In V1, there is a group of cells with large receptive fields, that show quick, flat-topped length summation. These are the "special complex" cells found by Gilbert [85] in layer 5. Considering the problems described here, these cells are ideal candidates for exerting a gain control influence over neurons with small receptive fields in the superficial layers. This point is taken up again later in this chapter.

Noise Immunity

Various researchers have commented that some neurons in the cortex show a particularly low sensitivity to noise textures (e.g. Hammond and MacKay [95]; Morrone *et al.* [172], but see Skottun *et al.* [217]). It was therefore considered worthwhile to test whether noise rejection may be one function of the gain control system. In particular, since noise is both wide-band and spatially extended, one would expect that the gain control feedback signal should be quite sensitive to noise if it is derived by summation in the frequency domain and over space. As gain control exerts a suppressive influence, this could act to reduce a neuron's sensitivity to noise.

These conjectures are borne out by simulation results. Figure 7-12A shows a bar pattern with added Gaussian noise. Figure 7-12B shows the response map to this stimulus using a linear simple cell model without gain control. These simple cells are selective for the vertical bars, but extensive oriented noise is also present. Figure 7-12C results when gain control is applied but interactions are not extended to cover the frequency domain. The noise is still evident, but has been partly "thresholded" by the neural output function. Figure 7-12D uses the full gain control scheme, including summation between spatial frequency channels. Noise is considerably suppressed in this response map.

Although the reduction of noise is a good thing, it may present problems for texture segmentation, since textures are often noise-like. One solution woud be to feed back an excitatory signal from high-level neurons selective for texture boundaries. This excitatory signal would lift some of the weaker simple cell responses to the texture above threshold and therefore improve the discrimination of texture difference at such boundaries.

It appears, therefore, that gain control does have a significant and useful role to play in the reduction of visual noise. This can only happen if the gain control feedback signal has a wide-band spatial frequency selectivity and shows summation within this bandwidth.

Contrast Dependent Surround Inhibition

When restricted grating stimuli are expanded in size beyond the bounds of their excitatory receptive fields, retinal M-cells subject to gain control show response reduction whereas neurons in the cortex do not, even though these cortical neurons exhibit gain control when tested in other ways (Ohzawa *et al.* [184]). This retinal "surround inhibition" effect has been described in Chapter 4. The fact that such an effect is not present in V1, even though the cortical gain control summation region is large, is probably a consequence of spatial non-

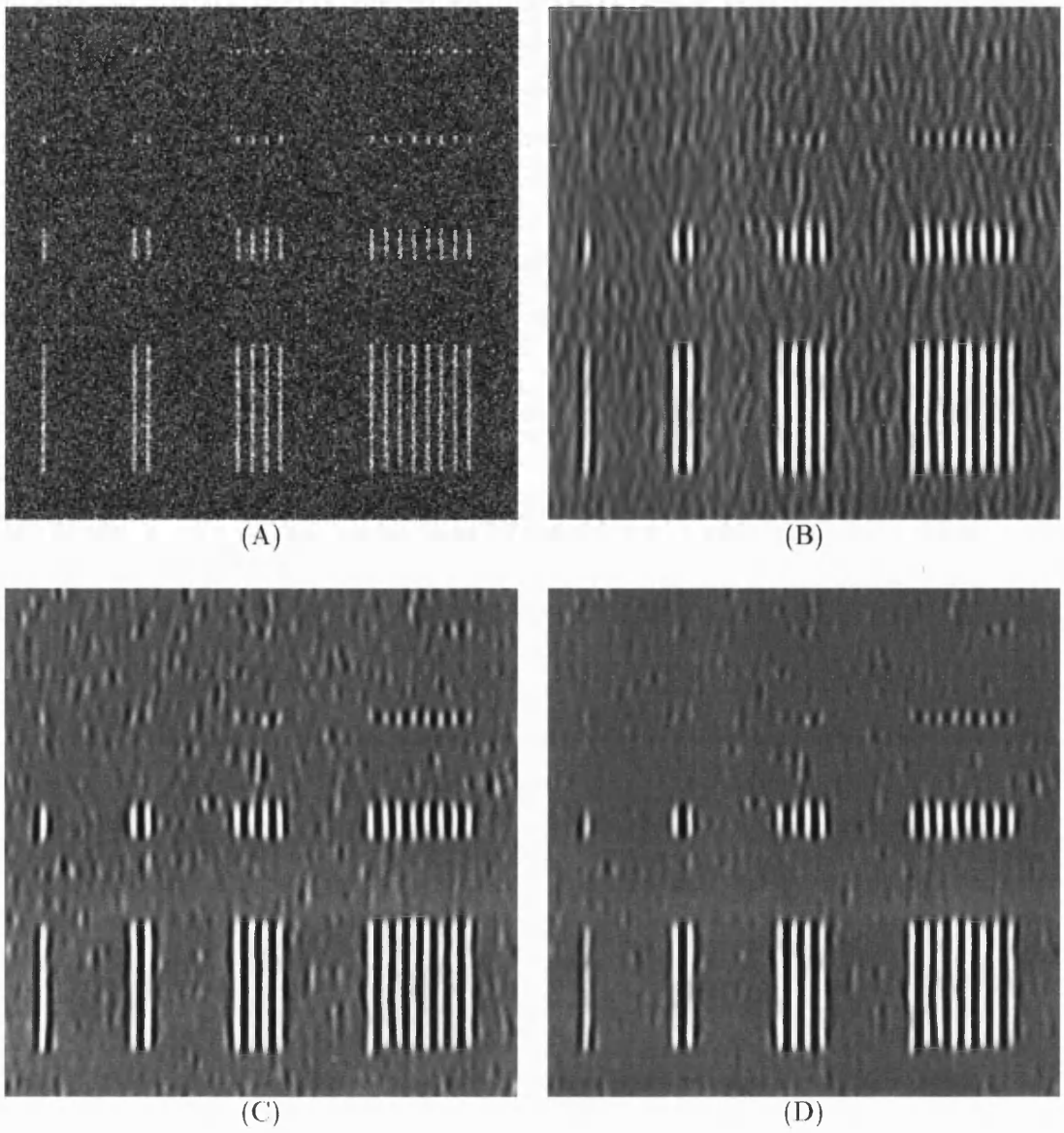


Figure 7-12: Demonstration of noise rejection by gain control interactions: (A) Noisy bars test stimulus ($5° \times 5°$). (B) Responses from simple cells tuned to vertical bars using the linear model of Chapter 6. (Responses from on-centre/off-centre units are displayed as white/black.) (C) Responses following gain control without frequency domain interactions. Reduction of some responses results from "thresholding" action of the output function. (D) Responses following gain control employing spatial frequency interactions. Responses to wide band noise stimuli are reduced in amplitude (feedback model, linear spatial summation).

additivity over this region leading to an inhibitory control signal that does not increase in strength with stimulus size.

Simulation results for gain control obtained using non-linear summation show that surround inhibition only occurs when a stimulus placed in the surround has a higher contrast than the stimulus that excites the centre. This is because the local contrast estimate is largely set by the stimulus with highest contrast within the gain control region and therefore the gain of the centre is reduced for this configuration. When the contrast of the surround stimulus is reduced below that of the centre stimulus then it ceases to have an inhibitory influence.

I find that a similar effect occurs in the spatial frequency domain and this can explain some of the non-linear behaviour reported by De Valois and Tootell [55], namely that a neuron can be both excited and inhibited by the same spatial frequency. The reasons for this behaviour are as follows: A neuron A selective for frequency f may be to a lesser extent excited by a frequency of $2f$ since this is within its excitatory bandwidth. If A is stimulated with f then the effect of $2f$ presented simultaneously will be inhibitory, particularly if f has a lower contrast. This is because $2f$ stimulates other suitably selective neurons that contribute to the suppressive gain control influence, reducing A's gain to f more than this frequency would excite via the conventional linear filter selectivity. Contrast dependent spatial frequency surround inhibition is the result of this interaction.

7.2.5 Discussion

The primary role for cortical contrast gain control appears to be the maintenance of high contrast sensitivity within a limited cell dynamic range without sacrificing local linearity. Although I have highlighted some problems with the current attempt at simulating the action of gain control, these simulations do in fact model the action quite well, and it is only upon quantitative examination that such problems come to light. All of the contrast-dependent changes in neural selectivity curves are the result of extensive linear spatial summation, since this produces an insensitivity to the contrast of small targets in the feedback path.

With regard to the difference between the feedback and feed-forward model: in general, the feedback configuration is found to produce a moderating influence on the problematic effects described, reducing their magnitude. Its behaviour is, however, rather complex and difficult to predict.

For one dimensional stimuli, additive spatial frequency domain and space domain summation tend to complement each other so that the inhibitory influence of gain control is

similar for isolated bars and extended gratings. Stimuli that are both wide-band and have a wide spatial extent produce stronger inhibition, and this has the beneficial effect of reducing noise sensitivity. The implication is that, when coupled with frequency domain summation, a controlled amount of spatial summation in the feedback path along an axis orthogonal to each cell's preferred orientation could well be a good thing.

Along the orientation axis itself, however, I find that additive summation in the gain control path creates a contrast dependent length summation characteristic which is unwanted because it interferes with the interpretation of response ratios.[5] The solution is to employ a feedback method that collects from a large area but does so without adding the signals.

One possibility is that the feedback signal could be created using a completely non-additive method over space and spatial frequency. This would indeed produce a stimulus independent contrast measure. Although $\odot_4$ was used to test this possibility, the ideal operator would be one that took the *maximum* response from a set of inputs. In space, this requires weighting the inputs with a Gaussian kernel and then finding the largest value. The gain control process would then find the maximum contrast in space and over a number of frequency channels and uses this measure to set the neural gain. I am not in favour of such an approach for a number of reasons:

- It seems unlikely that a neuron (e.g. an inter-neuron responsible for gain control) could carry out this operation at the same time as maintaining a linear contrast input-output relation.
- The neural map produced using this technique is full of value discontinuities—it is quite a harsh non-linearity, and so the resulting two dimensional signal must be maintained at the same sampling resolution as the input, which presents a problem during inter spatial frequency pooling.
- The noise rejection properties that seem to relate to additive pooling would be lost.
- There is a better solution to the problem which is supported by a significant amount of neurophysiological evidence. This solution is described later in Chapter 8.

In Section 7.2.3 I made the decision to extend gain control to an area three times larger than the size of the receptive field that is to be controlled. Heeger [100] points out that the gain

[5]This is true for the Wilson and Humanski type of divisive feedback, but other forms of divisive influence could be envisaged that may tame this effect.

control region must be larger than the receptive field because otherwise local spatial response relations would be disrupted. On the other hand, if the gain control region is very large then little advantage is obtained by normalisation because complete images have a reasonably wide contrast dynamic range. Following experimentation, Ohzawa *et al.* [184] reported that the gain control region was limited in size to the receptive field under test. They stated that there was no effect from the surround because extended gratings and gratings restricted to the receptive field produced the same contrast adaptation. I believe this statement is not a correct deduction from their test results, because if the gain signal is truly sensing contrast and not stimulus form, then this technique would be expected to set the gain equally for localised and extended gratings of equal contrast as explained earlier. A better test would be to set the contrast of an grating annulus outside the receptive field to a high value and see if this shifts the operating point for lower contrast stimuli at the receptive field centre.

As it stands, the model cannot in theory be used to provide sharpening of tuning curves in the spatial frequency and orientation domain without producing selectivities that change with contrast. In practice, such changes may be mild, symmetrical and non-critical. Wilson and Humanski [248] improved orientation selectivity using this method, and the recordings of Bauman and Bonds [14], which show that suppression appears to be largest at the low and high cut-off limits of spatial frequency tuning curves, suggest that feedback may be harnessed to improve selectivity in this dimension also. Similarly, Sillito [213] found that orientation curves were being sharpened by inhibitory mechanisms.

In summary, gain control provides a dynamic way of altering how each cell will respond to the contrast of a stimulus, and in doing so reduces the extent to which contrast modulates cell responses from place to place across the visual field. Forming the control measure to achieve this function is a difficult process that ideally involves gathering input from a wide number of cells to produce a signal that is not stimulus specific and, in particular, is not length tuned. This must involve direct input from the LGN because the LGN is the only source for signals that sum quickly enough with increase in length. I have already suggested that special complex cells have the correct receptive field sizes and summation rates to be "contrast" units. Chapter 8 describes how these receptive fields may be formed using additive spatial summation.

7.3 End, Side and Surround-Stopping

In Hubel and Wiesel's original account of receptive field properties in the cat's cortex, they described a set of cells, which they termed "hypercomplex". These showed response reduction as a bar stimulus was extended in length beyond the bounds of the excitatory receptive field. This property became known as "end-stopping", and various subsequent experimenters attempted to characterise the behaviour (Rose [198]; Sillito [214]; Sillito and Versiani [215]; Bolz and Gilbert [23]; DeAngelis *et al.* [61]) and produce theoretical models (Dobbins *et al.* [69]; Dobbins *et al.* [70]; Heitger *et al.* [101]).

Two other related effects are side-stopping and surround-stopping. Neurons that have side-stopped receptive fields show response reduction as the width of a grating is increased (more cycles are added) (De Valois *et al.* [59]). In the same way, surround-stopping is a reduction of response to gratings extended in any direction outside the excitatory receptive field (Foster *et al.* [82]).

Recent work (DeAngelis *et al.* [61]) has provided a clear characterisation of the behaviour of all three types of suppression. In particular, it is found to be divisive, phase insensitive and only broadly tuned for both orientation and spatial frequency. Previous theoretical models have assumed a subtractive influence drawing from highly stimulus specific input, so here I attempt to simulate divisive inhibitory mechanisms based on the new evidence.

7.3.1 Theory

As we have already seen, divisive inhibition presents a problem because it has the potential to introduce stimulus selectivities that change with contrast. In the case of end-stopping, for example, a neuron with an orientation selective receptive field might receive inhibitory connections from neurons with receptive fields that are offset along its orientation axis (Sillito [214]; Heitger *et al.* [101]). If the suppressive signal acts on the central neuron with a divisive effect of the form $e/(1+ks)$, where e is excitation and s is suppression, then we have end-stopping, but its strength increases with contrast. This effect would spoil the assumed contrast invariancy of population response ratios for the dimension of length tuning.

Careful analysis of the contrast-related behaviour of end-stopped cells described by DeAngelis *et al.* [61] suggests an alternative form of divisive influence. These authors reported the results of increasing surround contrast as centre contrast was held constant at 20%. A surround contrast of 5% was sufficient to halve the firing rate of one cell. I suggest that for this same cell, adding a surround contrast of 10% would halve the rate of firing if the centre

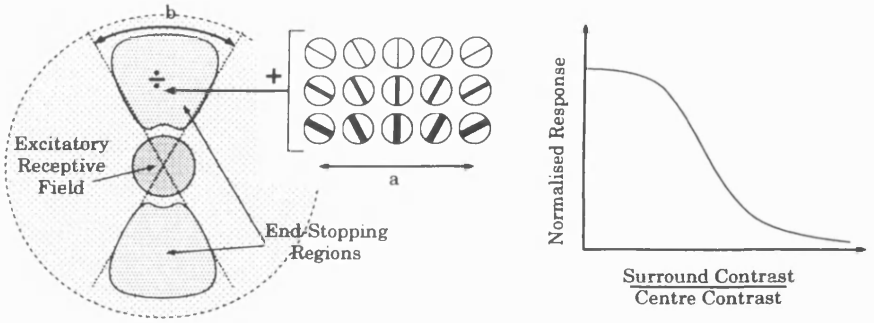


Figure 7-13: End-stopped neurons receive suppressive inputs from receptive fields spanning a range of spatial frequencies and orientations. Suppression is divisive and depends on the ratio of "centre" and "surround" contrasts. Constants a and b indicate the orientation bandwidths of the end-stopping pool and suppressive region, respectively.

was being driven by gratings with 40% contrast. In other words, the amount of suppression should be dependent on the ratio of centre and surround contrasts rather than on just the contrast of the surround. When tested, I find that this model results in contrast independent suppression, with no change in length tuning as contrast is varied. It is also able to predict the changes in semi-saturation constant and Naka-Rushton exponent reported by DeAngelis *et al.* when recording the excitatory contrast transfer function using various levels of surround contrast.

Here, I consider complex cells, with activity $u(x, y)$ that is linearly related to contrast (i.e. without including a Naka-Rushton output function or gain control), and allow these to be subject to an inhibitory signal $w(x, y)$. I then define the response from neurons with end, side or surround-stopping according to the formula

$$e(x, y) = u(x, y) \frac{u(x, y)^n}{u(x, y)^n + k^n w(x, y)^n}. \tag{7.5}$$

Here, k sets the ratio of surround to centre contrast that will reduce the output to half the value that it would have with no surround stimulation. In the example above $k = 0.25$.

The suppression signal $w(x, y)$ must be drawn from a range of spatial frequency channels centred on the channel in question (here I use the excitatory channel and its two neighbours—a bandwidth of about 2 octaves), and a range of orientation channels, to give broadly tuned inhibition. In the space domain, this inhibitory signal must arise from outside the receptive field and for end-stopping or side-stopping it should be drawn from the ends or sides of that receptive field. For spatial frequency channel k and orientation i (of eight), these requirements

result in the formula

$$w_{ik}(x,y) = M_{ik}(x,y) \otimes \sum_{j=i-3}^{i+4} \exp(-[j-i]^2/a^2) \sum_{m=k-1}^{k+1} [G_{km}(x,y) \otimes u_{jm}(x,y)], \qquad (7.6)$$

where a sets the orientation bandwidth of the suppression and $G_{km}(x,y)$ is a Gaussian used to match the receptive field sizes between channels prior to frequency domain pooling. The kernel $M_{ik}(x,y)$ determines the spatial arrangement of the suppression. It is best expressed in a polar coordinate system. For end-stopped cells this kernel is given by:

$$M_{ik}(r,\theta) = \begin{cases} G_k(r,\theta)\exp(-[\theta-\phi_i]^2/b^2) & \text{for } r > r_{min} \\ 0 & \text{otherwise.} \end{cases} \qquad (7.7)$$

In this equation, ϕ_i is the orientation angle of channel i and constant b sets the orientation bandwidth of the end suppression region. This is different from a, as illustrated in Figure 7-13. $G(r,\theta)$ is, again, a Gaussian which defines the spatial extent of the inhibitory region for spatial frequency channel k. Inhibitory input from the centre mechanism is prevented by the inclusion of a minimum radius for suppression. Note that the sharp cutoff below r_{min} implied by Equation 7.7 does not produce a sharp end-stopping effect because the underlying receptive fields already have a Gaussian profile. For side-stopped cells, kernel $M_{ik}(x,y)$ is rotated by 90° in space, whereas to produce surround-stopping, the dependence on θ is removed.

Bandwidths a and b are both set so that the overall orientation bandwidth for end-stopping suppression (judged using long lines) is well matched to the orientation bandwidth of the excitatory receptive field at the centre. This ensures that there are no "false responses" to off-orientation boundaries like those generated by the subtractive model due to Heitger *et al.* [101].

7.3.2 Simulation Results

Figure 7-14 shows response maps for end, side and surround-stopped complex cells when presented with the stimulus shown in Figure 7-14A. For the end-stopped cells, the responses are maximal for the third row of bars, and do not diminish significantly as the number of bars increases. Longer lines result in a reduced response, and there are local peaks near (but not actually at) the line ends. Side-stopped cells show the normal complex cell length tuning characteristics, but their responses are reduced by increasing the width (number) of the excitatory stimuli. In this example, two adjacent bars are about optimal to drive each

receptive field. Surround-stopping results in suppression when grating stimuli are extended in length or in number of cycles. In Figure 7-14D the maximum response is obtained from two short lines which are in close proximity (equivalent to a grating patch confined to the receptive field centre).

Although the role of end-stopped cells has been considered at length in the literature (e.g. Heitger *et al.* [101]; Heitger and von der Heydt [102]), side-stopping has seen very little exposure. I would suggest that side-stopped receptive fields are very important for structural vision because they are sensitive to contours that are isolated or present at texture boundaries. The side-stopped neurons of Figure 7-14C respond at the bars on the left, but they also respond at the edges of the "grating" regions. Although gratings are often thought of as the ideal stimuli for many V1 receptive fields, they are visually a form of texture, and the most useful contour information is really the outline boundary formed between a homogeneous grating patch and its background. Many structurally important visual features are isolated edges, and side-stopped receptive fields are arranged to show summation along the orientation of such contours, while rejecting line stimuli which are not isolated but are embedded in a similar background.

These assertations are illustrated to some extent by Figure 7-15 which shows a branch against a background of rubble. Figure 7-15B shows complex cell responses (without gain control) for which the eight orientation channels have been superimposed to give an overall picture. There is significant response to the background, which contains no strongly oriented elements or boundaries of large-scale structural significance. Figure 7-15C shows the side-stopped response map. These neurons are far more selective, and respond predominantly to the contours of the branch. Figure 7-15D shows the orientation domain responses, interpreted using the strategy described in Section 6.4.1. This figure demonstrates that useful local orientation information is present and is not disrupted by the background.

Reduction of the influence of noise is achieved here by means of a similar principle to that described for gain control (Section 7.2.4), except that here, responsiveness is only reduced along a line perpendicular to the orientation axis of the receptive field. This leaves room for length summation which is an important aspect of contour integration.

Length Summation Properties

End-stopped neurons have a clear selectivity for stimulus length. This is indicated in Figure 7-16 which shows length summation curves for neurons having different strengths of

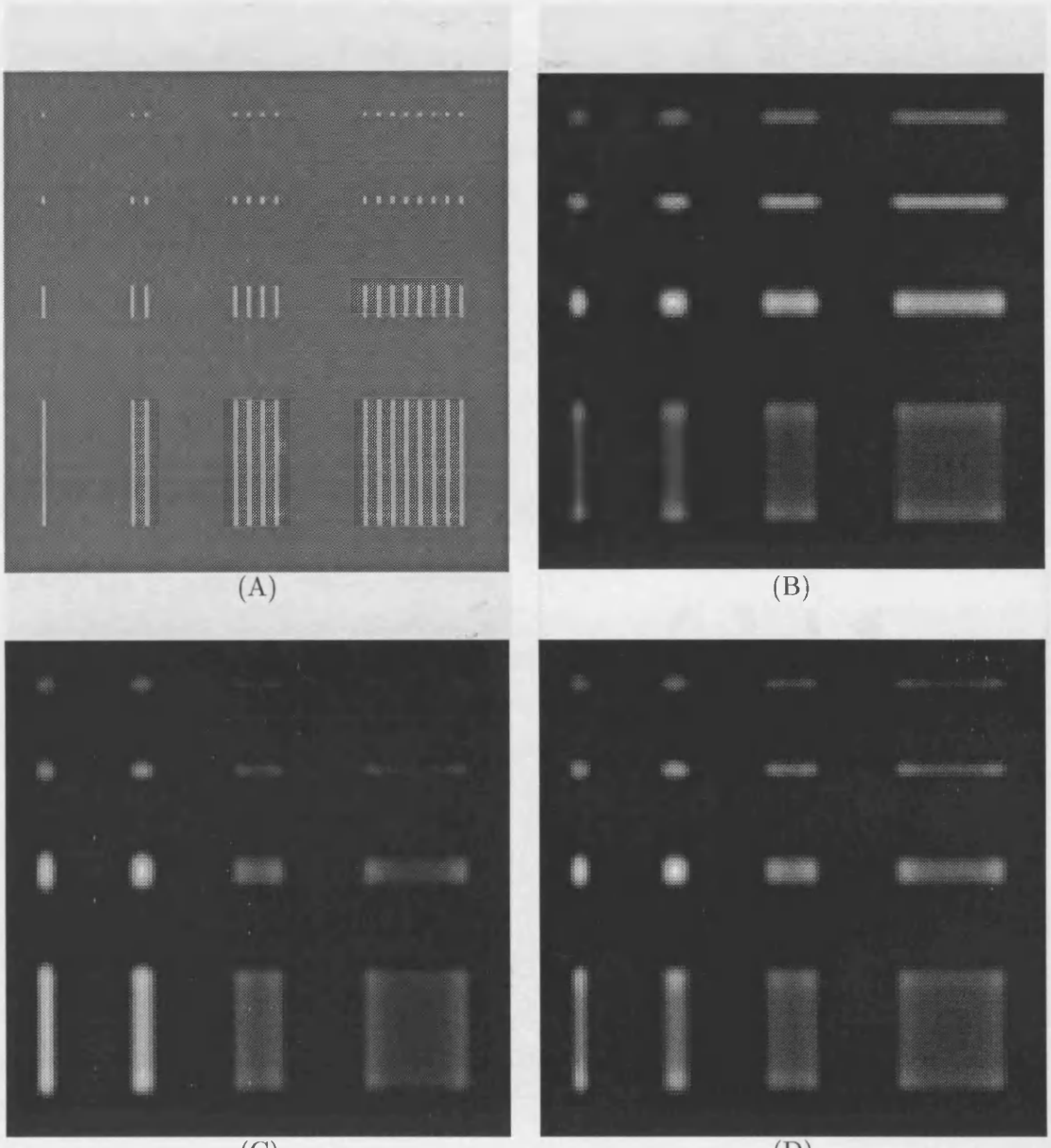


Figure 7-14: (A) Test pattern to demonstrate length and width tuning. Responses maps shown are for vertically-tuned complex cells with (B) end-stopping, (C) side-stopping and (D) surround-stopping. The test bars were spaced so as to provide a stimulus at the optimum spatial frequency of the channel (7.2 cyc/°).

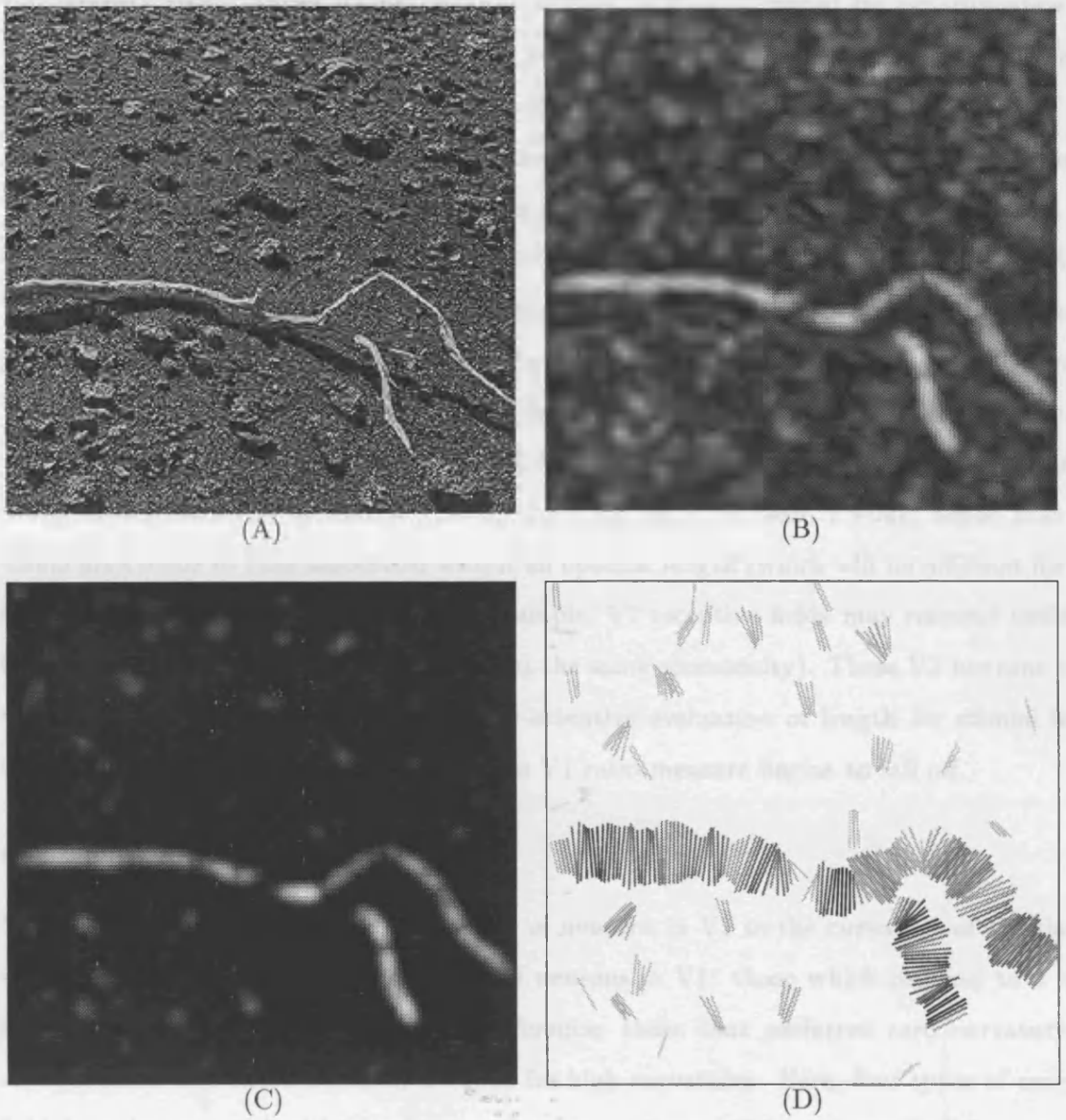


Figure 7-15: Neurons with side-stopping: (A) Original image of a branch on a stony background. (B) Responses from complex cells without side suppression (or gain control). Responses maps from eight orientation channels have been combined using a max function for illustrative purposes only. (C) Responses of complex cells with side-stopping. The best response is to a clearly oriented local feature, and not to the background texture. (D) Orientation domain responses for the neurons in C. Lines are drawn orthogonal to the local feature orientation. The spatial frequency channel shown here is tuned to 7.2 cyc/°. The shadow feature is primarily present on the 2.6 cyc/° channel.

end-stopping (changing values of k in Equation 7.5). Length summation curves are obtained by monitoring the response from a single neuron as bars of different lengths are centred over its receptive field. These curves are very similar to those recorded by experimenters (e.g. DeAngelis *et al.* [61]), but note that the receptive field size parameters used here for V1 neurons relate to the primate, rather than the cat.

From Figure 7-16, the length of the stimulus can be determined given non end-stopped and end-stopped responses because it is a monotonic function of the response ratio. The maximum sensitivity is then achieved for lines lengths between 0.3° and 0.7° because this is where the ratio changes fastest. It is not necessary to have a large spread in end-stopped cell optimal lengths. In fact, examination of the literature shows that in the superficial layers of V1 of the cat, most end-stopped receptive fields show a maximum response to bars between 1.5° and 2° long (Sillito [214]; Gilbert [85]; Sillito and Versiani [215]) and optimum lengths are generally clustered around 2° (DeAngelis *et al.* [61]). Instead, I would argue that each visual area tends to have associated with it an optimal length (which will be different for each spatial frequency channel), so that, for example, V2 receptive fields may respond optimally to bars around 0.6° long (for the primate, at the same eccentricity). These V2 neurons would then be responsible for taking up the pre-attentive evaluation of length for stimuli longer than 0.7° or so, where the sensitivity of the V1 ratio measure begins to fall off.

Selectivity for Curvature

Versavel *et al.* [234] studied the selectivity of neurons in V1 to the curvature of circular arc stimuli. They found three broad classes of neurons in V1: those which respond to a broad range of curvatures without showing preference; those that preferred zero curvature, i.e. straight lines; and those which were tuned for high curvatures. Here, four types of receptive field have been tested with circular arcs having various radii in order to build up a set of curvature response plots. Figure 7-17 shows the result of this exercise. The behaviour of standard complex, special complex and long complex receptive fields has been simulated. Special complex and long complex cells are discussed in Chapter 8, but essentially, these neurons show either an extremely fast, or a rather slow increase in response with stimulus length.

Figure 7-17 shows that neurons with standard complex receptive fields prefer low curvatures (straight lines) and their responses reduce as stimuli become tightly curved. Receptive fields with longer line integration lengths (long complex types) show a greater reduction at

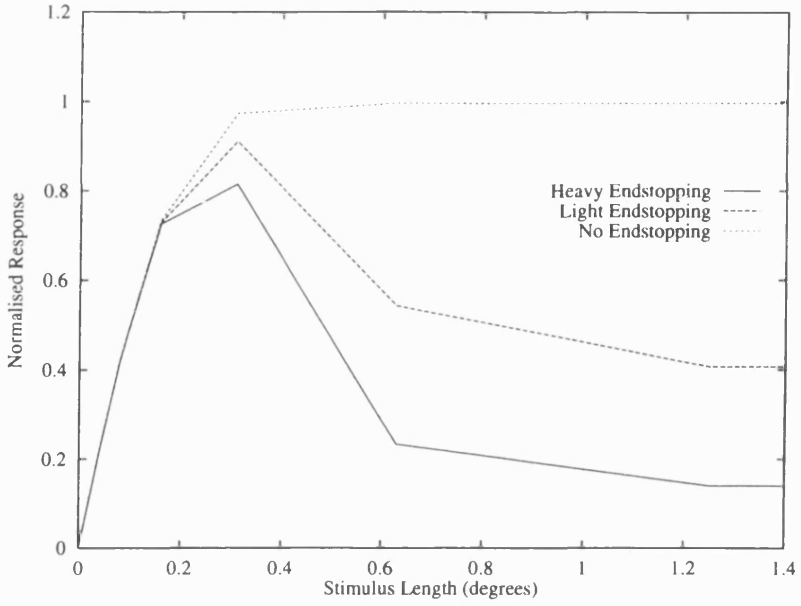


Figure 7-16: Length summation curves for neurons showing various degrees of end-stopping.

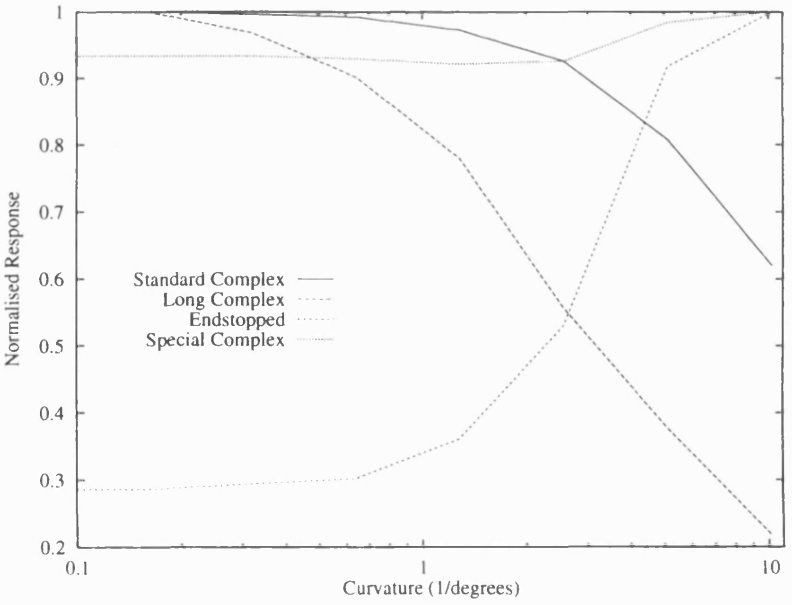


Figure 7-17: Selectivity for curvature: Various simulated cell types were tested for responsiveness to curves having a range of radii. Long complex and special complex types are described in Chapter 8.

high curvatures as might be expected. End-stopped cell responses increase with curvature, although they would be expected to fall again for spot-like stimuli which do not fill the entire excitatory receptive field (not tested here). Special complex cells, by virtue of their very short summation lengths, respond well over the entire curvature range. Thus, we have three groups of cells: special complex cells, which are curvature indifferent; long or standard complex cells, tuned for zero curvature; and end-stopped cells which prefer high curvatures. This behaviour corresponds nicely to the observations of Versavel *et al.* [234].

Dobbins *et al.* [69, 70] also attempted to model curvature selectivity in end-stopped cells, but their model is rather different to the one presented here because it is based on the assumption of a subtractive suppression mechanism. In addition, they supposed that end-stopping arose from neurons with large receptive fields that completely overlap the excitatory centre. This arrangement is problematic because it results in a large reduction in contrast sensitivity since the response is only a small difference signal. The selectivity and responsiveness of the end-stopping model I present here are both significantly compromised if the suppressive sites overlap the central mechanism to any great extent. It is, however, possible to allow this arrangement if the larger receptive field shows no response at all for short contours. Such behaviour would require a threshold, but unfortunately this then must introduce a contrast dependency.

In addition to an overlapping suppression source, Dobbins *et al.* [70] used phase sensitive inhibition. This causes their curvature tuning curves to differ greatly for stimuli positioned at various offsets across the receptive field width. Preliminary investigations have shown that the model presented here displays relatively little variation in curvature tuning under these conditions.

Effects of a Textured Surround

Knierim and Van Essen [137] investigated the behaviour of neurons in V1 when bar stimuli were embedded in a surround that was made up of similar elements. They found that responses were reduced by the simultaneous presence of such a texture, and suppression was often maximum when the surround elements matched the centre stimulus in orientation. Their results provide a possible physiological basis for perceptual "pop out" effects: A stimulus that differs from its surrounding elements in any domain is particularly salient and can be detected without a serial search (Bergen and Julesz [19]; Treisman [226]; Nothdurft [183]). Investigations were therefore conducted to determine if these effects were reproduced

by simulated neurons.

Since cells with various types of stopping receive inhibitory input from a receptive field pool which is broadly orientation tuned, there is more inhibition from a surround whose elements have, on average, the same orientation as the centre receptive field preference. Randomly oriented or orthogonally oriented surrounds produce less inhibition. The response to an isolated centre element is not inhibited at all. This mechanism could be seen as providing a context dependent reduction of the saliency of the centre stimulus.

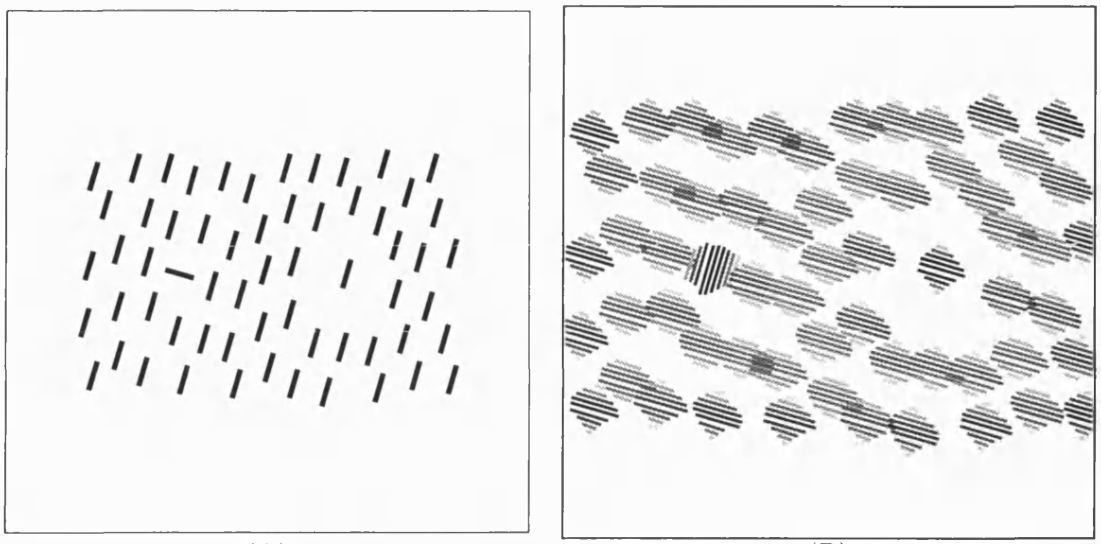
Figure 7-18 demonstrates this effect. In Figure 7-18A, elements that are isolated or differ from the background orientation are salient. Figure 7-18B shows the behaviour of simulated complex cells with receptive fields that are lightly surround-stopped. Both the salient elements produce stronger responses than their neighbours. This is also true for the elements that lie on the outside boundary of the textured region. The strongest responses are always to isolated stimuli.

I would suggest that the same approach might also work for the spatial frequency domain, since some neurons receive broad-band suppression from their surrounds. A neuron would therefore be less suppressed by a surround consisting of elements that are smaller or larger than the optimum excitatory element, than if the centre and surround elements had the same size. This has not yet been tested.

Figure 7-19 shows a side-effect of stimulus specific surround inhibition: the texture-induced orientation tilt illusion. In Figure 7-19A, the vertical lines are parallel but appear to be diverging up the page. Their perceived orientation appears to be "repelled" by the orientation of the surrounding texture. Figure 7-19B demonstrates that the computer simulation also suffers this illusion. The orthogonal elements which indicate local orientation for the parallel bars are each tilted by about 2°. This happens because the suppression signal from the surround, being orientation biased, acts asymmetrically on the population responses to the bars. The extent of biasing of local orientation is much greater with divisive suppression than when a subtractive influence is assumed. Preliminary results have shown a similar effect operates in the frequency domain. This can be interpreted as a reduction of the apparent thickness of bar elements when the surround consists of wide lines, and *vice versa.*

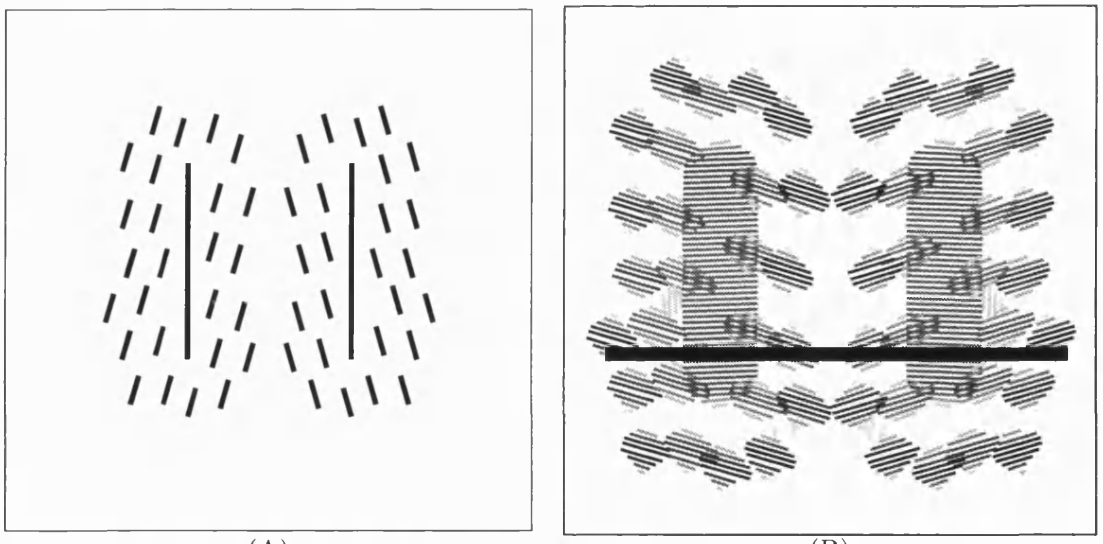
7.3.3 Discussion

The source of the end-stopping effect has been debated. I would suggest that there are four clear possibilities:



(A) (B)

Figure 7-18: (A) Differently oriented lines "pop out" of a textured background, as do elements which are significantly isolated. (B) The computer simulation predicts this effect. The elements in question both produce a stronger response (darker lines) than their neighbours.



(A) (B)

Figure 7-19: (A) Surround-induced orientation tilt illusion. The long lines are vertical and parallel. (B) Responses from mildly surround-stopped neurons, interpreted by the method of Section 6.4.1. Lines on the figure are drawn orthogonal to the local feature orientation. The black bar provides a straight horizontal reference edge. The computer simulation sees the vertical lines in A as being tilted by about 2°.

1. End-stopping (as well as surround and side suppression) could be generated by stimulus specific horizontal connections within layer 4 and perhaps also within layers 2 and 3. The inhibitory pool is not expected to be as form-selective as the recipient neuron, yielding broader orientation and spatial frequency tuning in the surround, as well as phase insensitivity.

2. End-stopping may result in part from inhibitory layer 6 to layer 4 connections, as suggested by Bolz and Gilbert [23]. Layer 6 cells have often been described as having a preference for long boundaries, and their activation could feed suppressively to layer 4 to reduce the response to these stimuli (Dobbins *et al.* [69]). This possibility has been called into question more recently by Grieve and Sillito [89] who find little change in the responsiveness to long bars in layer 4 during GABA blockade of layer 6. Instead, they find that the response to a short bar in layer 4 is reduced, reflecting a loss of excitatory drive rather than a loss of inhibition. This scenario is more consistent with the proposals that I advance in Chapter 8.

3. End-stopping in layers 2, 3 and 4 may result from end-stopping that is present already in the LGN by virtue of corticogeniculate feedback arising from layer 6 (Murphy and Sillito [178]).

4. End, side and surround-stopping could result in part from an inhibitory feedback from neurons having large receptive fields in V2. This has been suggested as a possible source for texture saliency effects in V1 (Knierim and Van Essen [137]). An argument against this proposal is that simple cells in layer 4 also show similar stopping effects to the superficial layers, but feedback from V2 avoids this layer (Zeki and Shipp [255]).

The model presented here has been developed as if proposal number 1 is the primary mechanism generating the specific inhibitory effects described so far. Further discussion on the creation of end-stopping is left until Chapter 8.

Researchers have described a number of behaviours that are not shown by the model as it stands. Sillito [214] and Sillito and Versiani [215] report that some end-stopped cells prefer stimuli that are shorter than the length of their excitatory receptive field. They also find that some cells are excited by a long bar with a gap in it when the gap is placed over the receptive field centre. Such phenomena could be explained by the model employed here if there was also some post-inhibitory spatial convergence along the length axis. Alternatively, these behaviours could arise from spatial pooling of an end-stopped LGN input as described

in the next chapter.

The model due to Heitger *et al.* [101] relies heavily on output from "single end-stopped" neurons to detect the ends of lines. These neurons have receptive fields that show suppression from only one end of their receptive field. Although Heitger *et al.* [101] provide an example recording, the incidence of this type is probably low because they are infrequently described in the literature. Rose [198] stated that experiments did not reveal any cells with an inhibitory flank on one side only; however, some of the recordings by Sillito and Versiani [215] appear to show an asymmetry of suppression. Although Heitger *et al.* [101] and Heitger and von der Heydt [102] included an input from doubly end-stopped receptive fields when detecting line discontinuities, this input does not contribute usefully to their final measure. This is because the doubly end-stopped signal has response peaks that occur away from the ends of lines, and the surround suppression that the authors include to suppress false responses also suppresses this signal. The resulting discontinuity measure becomes highly dependent on singly end-stopped cells.

Line ends are clearly important discontinuities and it seems that there ought to be a mechanism which yields a response maximum at the centre of the step change, rather than displaced some distance away, as is the case with double end-stopping. Preliminary studies have suggested that it may be possible to use three doubly end-stopped receptive fields to detect line ends. Two are excitatory and are selective for short lines and the third is inhibitory and tuned to lines that are twice this length. Results from these studies are promising, but are not included here.

Why do neurons receive broad-band suppression? One answer has been provided already for general surround-stopping, in that the surround tuning is important for producing saliency effects. Another explanation, which is of significance to end-stopped cells, is that a broad spatial frequency bandwidth for inhibition allows neurons to respond correctly to isolated spatial features. Real contours may change in "blur" along their length, especially when perspective effects and depth of field are taken into account. Since the image is broken up into spatial frequency bands for analysis, contour length judged from a single channel may be false since the feature could be shifting to another band. Having a sharply tuned centre and broadly tuned surround reflects the possibility that the peak spatial frequency may change along a contour and provides a limit on how fast this change can happen before the contour appears to be of limited length. A similar argument may apply to the orientation domain.

In summary, I have presented a model and simulation results for cells which are subject to long-range inhibitory influences from various locations outside of the excitatory receptive field. The model is successful in that it produces length tuning and curvature tuning results that are consistent with recordings from V1. In addition, the predicted effects of a textured surround and susceptibility to the tilt illusion provide further validation and give an insight into the mechanisms that underly the determination of visual saliency.

7.4 Conclusions

In this chapter I have concentrated on "horizontal interactions"—investigating contrast gain control and producing a model for the inhibitory processes that endow receptive fields with "stopping" properties.

Contrast gain control was found to produce response compression while (ideally) preserving response ratios. It was also found to produce contrast dependent surround inhibition in space and also in the spatial frequency domain. In addition, I demonstrated that contrast gain control reduces simple cell sensitivity to visual noise.

I uncovered a number of problems with forming a measure suitable for the cortical control of contrast gain. The ideal contrast signalling "receptive field" needs to have spatial properties that are similar to those of layer 5 special complex cells—in particular, a large spatial extent and a short summation length. I suggest a model for the formation of special complex cell receptive fields in the next chapter, which concentrates on a larger picture of V1 and its interaction with other areas.

In this chapter, I also introduced a novel divisive model for neurons that show end side or surround-stopping. Simulation results were included as well as length summation and curvature selectivity plots. Most significantly, I find that the computer simulation of divisive surround-stopping responds appropriately to stimulus saliency and is affected by the orientation tilt illusion.

Chapter 8

Summation Inside and Outside V1

8.1 Introduction

In this chapter I am concerned with summation and the way that it can be controlled by divisive inhibition. In expressing these ideas, I also find it necessary to consider the role of some of the forward and backward connections between V1 and the LGN, and between V1 and V2. This chapter contains original ideas which are more speculative than those of Chapter 7. Exploratory simulations have, however, given weight to many of these interpretations, and so they are presented here in the context of a broader look at the functionality of V1.

Section 8.2 describes a set of idealised length summation curves which provide a useful way of thinking about receptive field classes in V1. In Section 8.3 I propose a theory of how linear summation may be controlled in order to provide generalisation over space without necessarily introducing form selectivity. I also consider implications of this with respect to convergence into V2 from neurons with end, side and surround-stopped receptive fields. In Section 8.4 I look at end-stopping in the geniculate nucleus, and suggest how this affects the neural properties of V1, particularly with respect to the formation of special complex receptive fields and the derivation of a contrast measure suitable for gain control. The chapter concludes with a proposed model of V1 which, I believe, accounts for a good proportion of the functionality of this brain area—at least with regard to spatial processing.

8.2 Length Response Properties

Receptive fields in V1 are larger in area than their geniculate precursors. In the cat's visual cortex, receptive fields in layers 2 and 3 average about 2° in diameter (although this depends

on spatial frequency tuning and eccentricity in the visual field) whereas receptive field centres in dorsal LGN are only about 0.3° across (Gilbert [85]; DeAngelis *et al.*[61]; Lennie [143]; Linsenmeier *et al.* [148]). If V1 receptive fields were constructed by linear spatial summation of LGN input then one would expect that they would produce an increasing response when tested with progressively longer bars, up to a length equal to the receptive field size. In practice, this is true only for one class of cells, the "standard" simple or complex cell. Other classes show a more puzzling behaviour: Cells with special complex receptive fields produce a similar response to long or very short stimuli even though their receptive fields are large; and some end-stopped cells prefer a bar that is shorter than their excitatory receptive field (Gilbert [85]; Sillito [214]). In addition, there are often long-range facilitatory or inhibitory influences from outside the receptive fields of standard cells, as we have already seen in Chapter 7.

In this section, I propose a set of idealised behaviours that are helpful in that they allow us to tease apart a number of types of interaction that could be responsible for the summation properties of V1 neurons. Here, I am concerned with selectivity for stimulus length—an attribute that can be examined by constructing length tuning curves of the kind presented by Rose [198]. Three classifications are given here: standard V1 tuning curves, which show a medium summation length in a medium-sized receptive field; LGN tuning curves, which show a short summation length and a short receptive field; and special V1 tuning curves which have a short summation length but a medium to large receptive field.

8.2.1 Length Tuning for "Standard" Units in V1

Figure 8-1A shows an idealised length summation curve for cells in V1 which, superficially, show linear summation along the length of their receptive field. The overall summation limit is indicated by the vertical broken line marked "V1". For standard cells in layers 5 and 6, this is somewhat greater than the limit for cells in layers 2–4 because they have larger receptive fields (Gilbert [85]). When bars extend outside the excitatory receptive field, three things can happen: the cell response may reduce, reflecting remote inhibitory influences (end-stopping); it may increase, due to a facilitatory influence; or alternatively, it may stay the same (Rose [198]).

In the case of remote facilitation, it is important to realise that the receptive field itself need not be long, and although obviously preferring long stimuli, it may not be possible to stimulate the cell with long flanking bars placed outside the receptive field. This would be

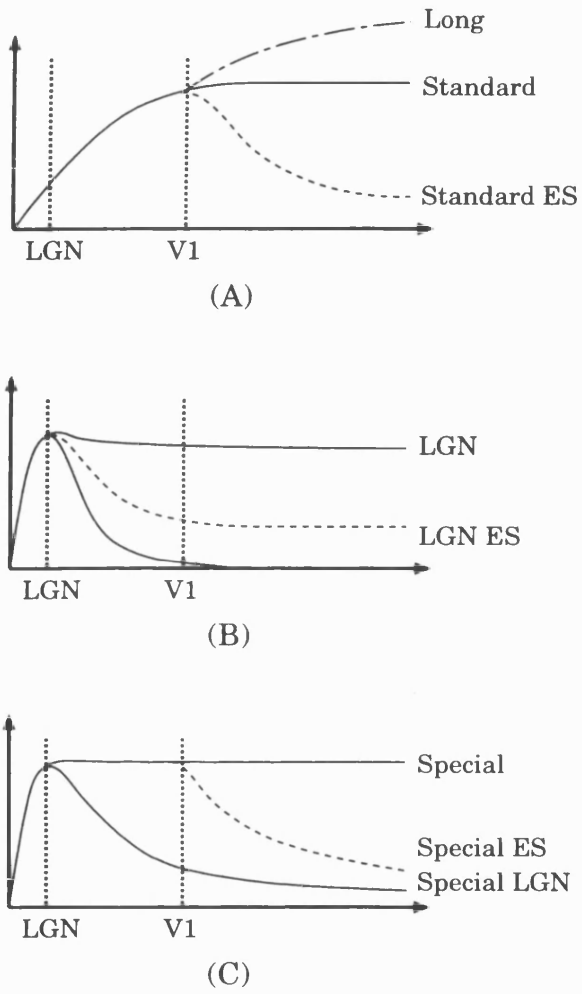


Figure 8-1: Idealised length-summation curves. (A) Standard V1 shape which can show facilitation (long) or inhibition (ES) from outside the receptive field but generally shows positive length summation within a 2° diameter field (cat). (B) Lateral geniculate nucleus. Summation lengths are short (about 0.3° for the cat). Length tuning curves would be fairly flat, but corticofugal inhibition from layer 6 renders them end-stopped for elongated stimuli. End-stopping can have different strengths. (C) Special complex cells. These show LGN-like summation lengths and can be end-stopped from outside the receptive field. Alternatively, they can be end-stopped within the large receptive field.

especially true for a multiplicative type of facilitation. Cells in layer 6 with long receptive fields appear to show this property (Gilbert [85]), and Rose [198] includes length tuning curves that display a clear change of slope between internal and external summation sections.

For neurons showing a standard length curve like Figure 8-1A, end-stopping or end-facilitation could arise from within V1 or they could be produced by back-projection from V2. Since layer 6 is a target for V2 feedback, the latter configuration seems more appropriate for generating "long" receptive fields in this layer. In either case, response ratios between neurons having curves like the three shown in Figure 8-1A could encode a pre-attentive measure of length for bars that extend outside the receptive field. For shorter bars, the length of the stimulus and its contrast cannot be separated because the curves are coincident.

8.2.2 Length Tuning for Units in the LGN

Figure 8-1B shows typical length tuning curves for neurons in the geniculate nucleus. They are characterised by a small receptive field and a short summation length. Such cells are not orientation selective, so responses to longer lines are just a reflection of local contrast information. If LGN neurons were not influenced by feedback from V1, they would only be mildly end-stopped, because bar stimuli only invade part of the receptive field surround and do not stimulate it optimally. In reality, there is an inhibitory drive to the LGN, arising from layer 6 of V1, that increases in strength with bar length (Murphy and Sillito [178]). This inhibition does not appear to be orientation specific, reflecting a convergence from layer 6 cells tuned to the full range of orientations. The net effect on neurons in the LGN is to improve their selectivity for spots and discontinuities, at the expense of edge information. Two other curves are shown in Figure 8-1B to illustrate this effect. A significant number of cells show complete suppression for 2° bars ("V1" on the figure) and, on average, the response to a long bar is about 30% of the response to a bar of optimal length.

The important point here is that LGN end-stopping must have a significant impact on the response properties of V1 neurons, and in particular, on the apparent linearity of their spatial summation. For example, an end-stopped excitatory drive is probably the reason why simple and complex cells are so strongly affected by small contrast-reversed segments inserted in medium-length bar stimuli (Hammond and Mackay [95, 96]). In later sections I develop the idea that having an end-stopped LGN signal is necessary for the cortical assessment of contrast.

8.2.3 Length Tuning for "Special" Units in V1

Figure 8-1C shows length response curves for special complex cells in V1. These cells have large receptive fields and they are present in layer 5 and at the layer 3/4 border (Gilbert [85]; Livingstone and Hubel [150]). The characteristic feature of special complex cells is a very short summation length and this strongly implies the presence of a direct LGN-like input. Ferster and Lindström [77] found that, for the cat, some neurons in superficial layer 5 were monosynaptically activated from the LGN. Other cells in the lower regions of layer 5 receive input from layer 4 via layer 3. Since Gilbert [85] found special complex cells at the layer 3/4 border, these could be receiving input from neurons in layer 4 that show a very fast summation with increase in stimulus length.

In addition to having a good response to short stimuli, special complex cells can exhibit end-stopping from outside the receptive field (Figure 8-1C, and see Gilbert [85] figure 2I, for a good example). This inhibition probably arises from specific long-range connections in the same way as end-stopping in standard complex cells. Special complex cells can also show end-stopping within the receptive field. I suggest that this property is the result of receiving a strongly end-stopped LGN drive.

I consider the detailed behaviour of special complex cells later in this chapter, but at this point it is worth noting that cells of this type (without end-stopping) are ideal for communicating information about the contrast of a feature irrespective of its position within a large region. This is because there is little response dependence on the size of a stimulus. In layer 5 there are also neurons which have responses that increase with stimulus length, but their responses are also contrast dependent. The ratio between responses from standard and special complex cells with coincident receptive fields would therefore convey information about contour length independently of contrast. This is true only for a range of lengths between points "LGN" and "V1" on the graphs of Figure 8-1. For very short lengths, special complex cells lose the length independence of their responses.

8.2.4 Discussion

In the preceding sections I introduced three classes of length response curve—two for cortical neurons and one for neurons in the geniculate nucleus. I have suggested that end-stopping can arise from outside the receptive field, or from within the receptive field by virtue of end-stopped LGN input.

Hammond and Mackay [96] report that they experienced difficulty in classifying cells

cleanly into the "special" and "standard" groups of Gilbert [85]. This suggests that, for a fixed size of receptive field, summation lengths may actually be distributed over a range from short (special), through a length matching the receptive field size (standard), to beyond (for units with apparently long receptive fields).

Given that V1 sums signals that are already end-stopped, two points are worthy of note. Firstly, if the LGN input is very strongly end-stopped and shows no response to long contours, then no amount of summation in V1 will be able to recover a response and so the convergence site is bound to appear end-stopped. Secondly, if the LGN signal is only 50% end-stopped, for example, the fact that a response still remains to a long contour allows suitable linear summation in V1 to completely cancel the overt effect of the end-stopping.

It should be remembered that length summation in V1 is necessary not only to provide stronger responses to long stimuli, but also to introduce orientation selectivity. Special complex cells are enigmatic in that they show progressively increasing orientation selectivity for longer stimuli without an associated length-response increment. They also have a short summation length and a large receptive field. In the next section, I begin to describe a general theory for behaviour of this type.

8.3 The Control of Summation

The ideas expressed in this section arose out of dissatisfaction with a truly non-linear convergence method (e.g. $\odot_4$) for producing a non-summing contrast measure suitable for contrast gain control. The main argument presented here is that the introduction of divisive end, side or surround-stopping can control the amount of summation introduced by neurons downstream that employ a purely additive form of spatial convergence.

Neurons in V2 or V4 have previously been described as responding to an optimal stimulus wherever it is placed in a large receptive field (Desimone *et al.* [66]; Desimone and Schien [65]; von der Heydt and Peterhans [235]; Peterhans and von der Heydt [190]). These neurons are able to *generalise over space*—that is, they lose information about absolute position but retain form selectivity. Here I suggest how this can be achieved using additive convergence, and I propose that the large suppressive surrounds which are possessed by such neurons are essential for this behaviour.

8.3.1 Generalisation Over Space

Consider a neuron somewhere in the visual system that receives its input from complex cells in V1. If this neuron receives convergent input then it will have a large receptive field. If the convergence process employs linear addition, then the optimal stimulus will be a large grating patch since this activates all the constituent complex cells and produces a strong summed response. This type of convergence, therefore, introduces a change in form selectivity: The underlying complex cells show short-range length and width summation, while their projection target shows long-range length and width summation.

Suppose that instead of the above, the neuron was expected to carry out generalisation over space. This would require a retention of the selectivity of the underlying complex cells, but the neuron must still have a large receptive field. The neuron would need to fire strongly if a grating stimulus filled the receptive field of at least one of the complex cells, but its firing rate should not increase for any further expansion of the grating patch. In addition, there should not be any non-linear response compression at the output which could distort the shape of the transmitted length or width tuning functions.

Superficially, it appears that the summing junction should just implement a logical *or* function. The continuous version of this is the *max* function. To employ this transform, a neuron would firstly need to apply a Gaussian weighting function[1] to its inputs (to define the receptive field size), and then it would have to chose the maximum of the weighted inputs, which would then become the response level. I have already indicated that I find this scenario an unlikely solution. The main reasons are, firstly, that this non-linearity is harsh and results in a spatially discontinuous output that cannot be legitimately subsampled, and secondly, it seems unlikely that the input-output transfer function for an isolated stimulus could remain linear for a single cell implementation. The subsampling issue is of particular importance since the role of generalisation over space is surely to reduce the number of individual receptive fields that are needed to cover the visual field.

I suggest that instead of using an *or* function, the vision system can use linear addition providing that the inputs to the convergence site are suitably pre-conditioned. The essential proposal is to pass the output of each complex cell through a stage that reduces its gain as more of its neighbours are stimulated. As the grating patch increases in area and excites more complex cells then their gains will reduce in proportion to the number excited. This provision keeps the signal at the higher level summing junction constant. This extra stage amounts

[1] Or some other similar pattern of weights.

to the introduction of divisive suppression of the input response map, using an inhibitory signal drawn from a region which has a size comparable to that of the output receptive field. Complex cells which are subject to such an influence will appear to have large suppressive surrounds, and the surround suppression will be divisive.

The solution therefore is to start with the signal you wish to generalise over space, introduce divisive surround suppression, and then carry out convergence using a linear summing junction. Because the last stage acts as a low-pass spatial filter, it is quite legitimate to subsample the final signal. Few receptive fields of this sort are needed to cover the visual field.

8.3.2 Simulation Results

I now demonstrate the process described above. Figure 8-2A shows a response map for complex cells in primate V1 which have receptive fields which have a diameter of 0.3°. The visual stimulus is the test card of Figure 7-14A which has already been used to assess length and width tuning. From Figure 8-2A it can be seen that these complex cells show increasing responses to the length and width of a grating-like stimulus over the first two rows and first column of the figure. This increased response can be seen as an increase in the brightness on the response map at the centre of each feature. A grating that extends outside the receptive field produces no additional response (bottom right 3×2 set of stimuli: all the neurons positioned over the centre of these 6 stimuli are responding maximally (white)).

When linear spatial convergence is used directly to create a receptive field that is four times larger, then the response map of Figure 8-2C is produced. The additive summation changes the length and width tuning from that of the input so that the optimal stimulus is now the large grating patch at the bottom right. (This can be seen as optimal because it produces the strongest response in neurons at its centre, when compared with the other 15 different stimuli shown) This is clearly not generalisation over space. In particular, the length summation curve will now tend to a maximum for lengths equal to the large receptive field diameter.

Figure 8-2B shows the response map of Figure 8-2A after each cell has been subject to divisive surround-inhibition using exactly the same method as described in Chapter 7. The effect of employing linear spatial summation to create a large receptive field from the responses shown in Figure 8-2B is very different from the previous case. Figure 8-2D shows the results. The strength of the centre neuron's response to each of the 16 stimuli is almost

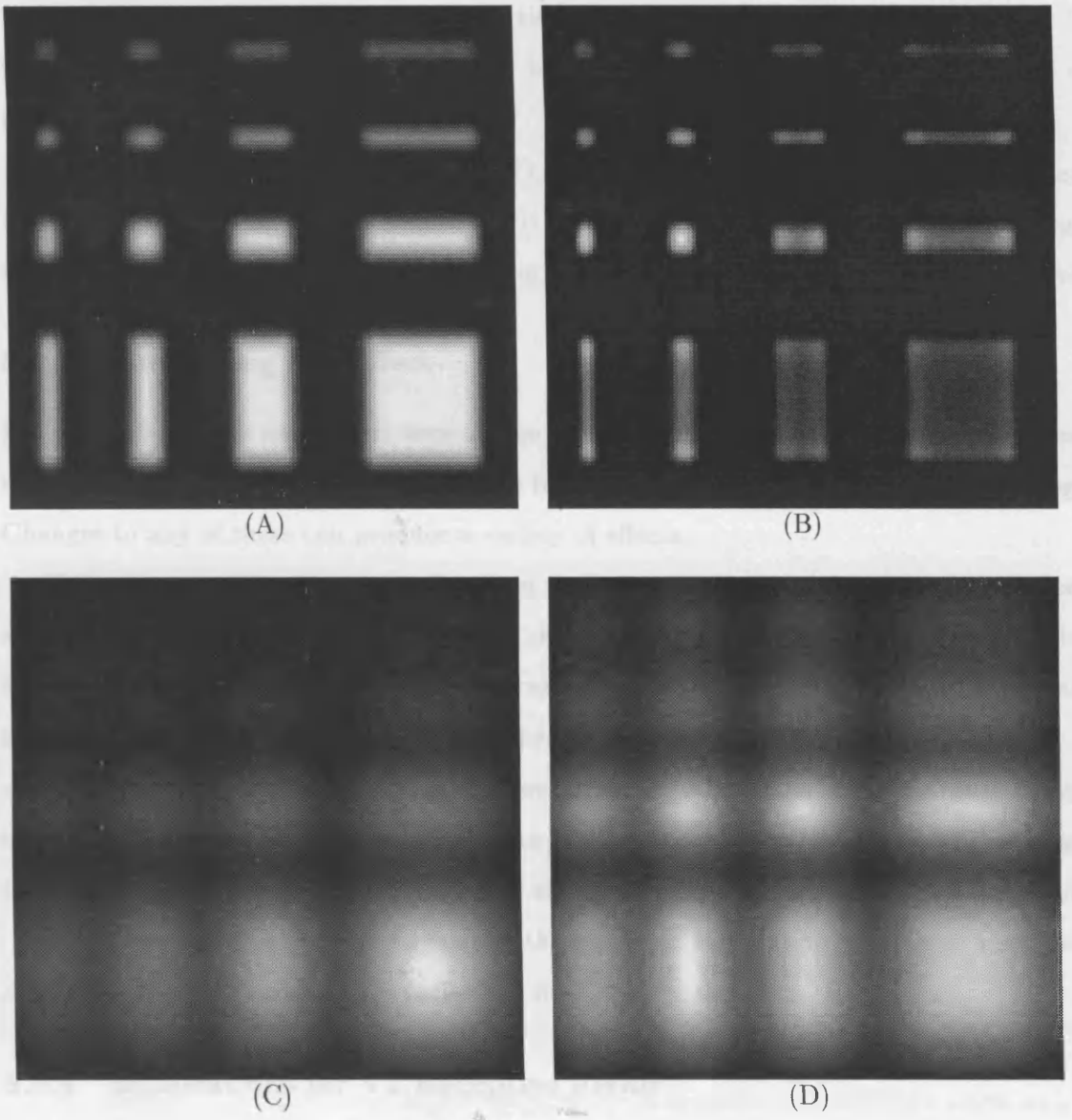


Figure 8-2: Demonstration of summation control. The stimulus was the length/width tuning test image of Figure 7-14. (A) Responses of complex cells having a 0.3° diameter receptive field to the test image. (B) Responses of surround-stopped complex cells to the same test image. Surround-stopping produces a reduction in response as the length or width of the "grating" patch is increased. (C) Responses of hypothetical neurons that receive convergent input from the neurons in A, to create a large 1.2° diameter receptive field. (D) Responses of hypothetical neurons that are driven by surround-stopped input to produce the 1.2° receptive field. These neurons retain the length and width summation properties of the original complex cells in A but have a large receptive field (generalisation in space), whereas the neurons in C prefer the large grating patch.

the same as in Figure 8-2A. The amount of "blur" apparent in Figure 8-2C and D is the same because both types of neurons receive convergence from equal areas. Unlike Figure 8-2C, however, Figure 8-2D does not show distortion of the original length and width tuning. This is because responses in Figure 8-2B have been reduced appropriately to account for each increase in stimulus area.[2]

The large amount of blur in Figure 8-2D, coupled with the fact that the map was created using a linear summing process means that it can be held at a much lower sampling resolution than shown here without loss of information. Generalisation over space has been achieved.

8.3.3 Controlling the Effect

For Figure 8-2, three parameters were chosen carefully. These were the size of the suppressive surround region, the strength of the divisive inhibition, and the size of the convergence region. Changes to any of these can produce a variety of effects.

If the convergence region is smaller, then some surround suppression remains uncancelled and the resulting neuron shows "stopping" within the bounds of its receptive field. This is also the case if the inhibition strength is increased. If the inhibition strength is very great, or if there is a response threshold prior to convergence, then the resultant large receptive field will fail to give any response to extended stimuli. On the other hand, if the convergence region is increased in size so that it is greater than the spatial extent of the surround suppression, then some new summation takes place and effects like those in Figure 8-2C are obtained.

In conclusion, it is possible to control the extent of summation to achieve a number of desirable effects by varying the spatial and inhibitory parameters.

8.3.4 Implications for V2 Receptive Fields

Apart from surround-stopping, DeAngelis *et al.* [61] have described two other inhibitory effects: end-stopping and side-stopping. It is of interest to consider what effects all three of these mechanisms would have if neurons possessing such properties converged on hypothetical neurons in V2 which did nothing other than sum over space, to create receptive fields that were 2–3 times larger than their precursors in V1. I consider the three classes mentioned above, and additionally, neurons which lack inhibitory influences.

[2] I have developed a more rigourous theoretical treatment of this behaviour, but there is not enough space here to include it.

Surround-stopping We have already examined convergence from receptive fields that receive suppression from their surrounds. This form of suppression can, under the correct conditions, prevent further length and width summation.

Side-stopping When the input to a large receptive field consists of drive from side-stopped cells then new width summation is prevented because of the side-stopping. Along the length axis, however, there is no inhibition, so the large receptive field has a preference for long stimuli. The implied orientation tuning is derived entirely from the location of the inhibitory side regions. Such a neuron will have a preference for longer lines without showing a response differential between bars and gratings. This has relevance to the "contour association field" ideas of Field *et al.* [78] and Polat and Sagi [191].

End-stopping When convergent input to V2 arises from end-stopped units in V1 then new summation is produced only along a direction orthogonal to the orientation axis. In particular, if the input is strongly end-stopped then the output cell is only activated by short lines or line ends, but summation still takes place orthogonally to the local orientation. This produces a receptive field which groups line discontinuities along an axis perpendicular to their local orientation. Obviously this has some relevance to the formation of subjective contours and the grouping of line terminations that takes place in V2 as discovered by Peterhans and von der Heydt [190]. The effect described here is not quite the same, however, because the current model does not really group the *centres* of line termination. This is worthy of further investigation.

No stopping effect I have already demonstrated the effect of leaving out the pre-convergent surround-stopping: This is found to produce a selectivity for larger grating patches as shown in Figure 8-2C.

In conclusion, it can be seen that restricting the suppressive influence to a single axis prior to convergence creates some interesting effects. Polat and Sagi [191] found an improvement in contrast sensitivity when oriented Gabor elements were aligned either at 0° or at 90° to their orientation axis, but not when the angle was 45°. This suggests that two contour grouping processes are present: one which sums along the direction of local orientation, and one which sums orthogonally to this. Such a finding is consistent with the ideas that are presented here.

8.3.5 Further Implications

A mechanism that allows a linear summing junction to implement generalisation over space has wide application throughout the vision system. I have already mentioned the possibility

that the suppressive surrounds observed in V4 may be a manifestation of this process. Receptive fields in IT are known to cover almost the entire visual field (Desimone *et al.* [66]). They receive convergent input from V4 receptive fields which are smaller and have silent suppressive surrounds with the same stimulus selectivity as the centres. It is not too difficult to see the proposed mechanism being used for this V4 to IT spatial convergence.

More significantly to interactions between V1 and V2, colour selective neurons in primate V1 often have large surrounds. These have been described as double opponent-colour cells by Michael [166] and Livingstone and Hubel [150]. Initially researchers reported that such neurons had surrounds from which excitatory responses could be obtained. Later, Ts'o and Gilbert [227] reported that very few were truly double opponent and most were variously described as having "three-quarter opponent" or "modified type II" receptive fields, reflecting the fact that the surround was only ever suppressive. Lennie *et al.* [145] could not find clear evidence of double-opponent receptive fields. In addition to this, Hubel and Livingstone [111] discovered unoriented complex cells in primate V2 which prefer coloured stimuli that are smaller than the size of their receptive field.

I would suggest, in view of these observations, that the double opponent cell silent surround is actually divisive rather than subtractive, and the same mechanism for generalisation over space is being implemented by the unoriented colour-opponent channels in V1 and V2. Preliminary simulation results suggest that this configuration can allow for a controlled amount of spatial summation in the colour-opponent channel which is then followed by spatial generalisation. If different centre sizes are used then we have neurons that can supply colour information averaged over various region sizes, and like other stimulus selectivities in the "recognition" pathway, these become decoupled from a dependence on absolute spatial location by the generalisation process. This has significant relevance to perception, but a further discussion is beyond the scope of this thesis.

8.3.6 Conclusions

In conclusion, the control of spatial summation by divisive surround suppression has significant application throughout the vision system. I suggest now that this is also the method by which special complex cells keep a short summation length even though they have a large receptive field. When combined with end-stopping and side-stopping, the theory described above may also be responsible for controlling form selectivity in V2. The same mechanism is most probably used to generalise colour responses over space and to produce larger receptive

fields at numerous stages within the visual pathway.

8.4 Layer 5, Layer 6 and Corticogeniculate Feedback

In this chapter, I have frequently referred to special complex cells, and this section presents a simulation of their behaviour. This simulation is included in order to illustrate the effect of corticogeniculate feedback and to show how the summation ideas I have expressed above may be usefully employed, even at a relatively low level, within V1. In the following sections, I summarise the properties of neurons in layers 5 and 6 before proposing a special complex cell receptive field model.

I have tended to refer to a mixture of studies relating to the cat and the monkey. This mixing of evidences is not really desirable, but is somewhat unavoidable since I am interested in primate vision but the cat has been studied rather more intensely. The primary species differences recognised here are as follows: In the primate, the termination site for geniculate fibres is less diffuse and is generally restricted to layer 4C; directionally selective cells are found in upper layer 4 and layer 6, rather than being present in all layers; significant numbers of colour selective cells are present; neurons cover a higher range of spatial frequencies; V2 receives its drive mainly from layers 2/3 and 4B of V1, rather than additionally from the LGN (Lennie [143]; Livingstone and Hubel [149]; Zeki and Shipp [255]; Hawken *et al.* [99]; Lennie *et al.* [145]). In general, this represents an increase in the segregation of functional streams and a sharpening of receptive field categorisation when compared with the feline visual system.

8.4.1 Layer 6

Gilbert [85] reported that layer 6 contained both simple and complex cells which had larger receptive fields than any present in the superficial layers. In particular many (63%) were found to be narrow and very long, responding with directional preference to the longest test stimuli available. Livingstone and Hubel [150] reported a similar thing for layer 6 in primate V1. Later, Grieve and Sillito [88] and DeAngelis *et al.* [61] found that the incidence of neurons with long receptive fields was lower than previously reported (about 25%) and DeAngelis *et al.* [61] demonstrated that neurons with long receptive fields in this layer tended to be part of a population that showed strong side-stopping.

Neurons in layer 6 are known to receive excitatory input directly from the LGN as well as via layers 4 and 5 (Lennie [143]; Ferster and Lindström [77]). Their responses feed back to the

LGN and appear to produce inhibition since neurons in this subcortical site are known to be end-stopped, and experimental inactivation of layer 6 reduces this end-stopping (Murphy and Sillito [178]). Local circuit neurons in layer 6 also project to simple cells in layer 4 producing an effect which is predominantly excitatory (Lund *et al.* [152]; Grieve and Sillito [89]; Ferster and Lindström [77]). Layer 6 sends its axons to MT which is an area heavily concerned with the analysis of motion (Zeki and Shipp [255]; Hawken *et al.* [99]).

Along with Grieve and Sillito [89] I would suggest that layer 6 both introduces end-stopping via its feedback to the geniculate nucleus and reduces its effect for selected cells in layer 4 via facilitatory local feedback to simple cells in this layer. Inactivation of layer 6 would then be expected to remove LGN end-stopping, but may not produce a large change in layer 4 and the superficial layers because the balancing correction is also removed. I would expect, however, that inactivation of layer 6 *should* produce an effect on special complex cells in layer 5, since I argue later that these require LGN end-stopping to prevent their responses to bars exceeding their responses to spots. Inactivation of layer 6 might therefore make special complex cells appear "standard". This is a testable hypothesis.

In addition to describing layer 6 as containing neurons with long receptive fields, Gilbert [85] also reported that, on average, cells in this layer showed the lowest orientation selectivity in V1. Cells in layers 2 and 3 are more sharply tuned even though they have smaller receptive fields and therefore shorter summation lengths. There is almost a contradiction here, because a long layer 6 receptive field that was created by linear filtering of LGN input would be highly orientation selective. This argues for a shorter receptive field that receives end-facilitation from within layer 6 or perhaps from cells with longer receptive fields in V2. This is consistent with the length summation curves of Rose [198] as I pointed out in Section 8.2.1.

The role of end-facilitation is uncertain—however, in accord with a suggestion by Ferster and Lindström [77] about the reason for a layer 6 to layer 4 projection, it seems likely that localised responses to weak colinear contour elements could be enhanced relative to surrounding texture by end-facilitation. This would improve the contrast sensitivity for oriented elements that were organised in a line—a suggestion that is consistent with the psychophysical findings of Polat and Sagi [191].

8.4.2 Layer 5

Layer 5 consists of large complex receptive fields that are a mixture of standard and special types (Gilbert [85]; Livingstone and Hubel [150]). This layer receives excitatory input from

layer 3 and 4B and makes return connections to layers 2 and 3. In addition, the superficial part of layer 5 receives input from layer 4C, and, at least in the case of the cat, directly from the LGN. Neurons in this part of layer 5 feed back to the layer 3/4 border (Lund *et al.* [152]; Lund [151]; Ferster and Lindström [77]). It is possible that neurons in superficial layer 5 are special complex cells since they have reciprocal connections with units at the layer 3/4 border where Gilbert [85] also found special complex cells.

Apart from sending a return projection to layers 2 and 3, neurons in layer 5 also drive layer 6 (cat) and project extensively to subcortical sites associated with the control of attention: the superior colliculus, pulvinar nucleus and other parts of the brain stem.

Although there is clearly some functional segregation within layer 5, I suggest two possible roles for this layer. Firstly, with reference to special complex cells, it provides a general purpose contrast reference which could form the basis for contrast gain control via the known projections to layers 2 and 3. Secondly, layer 5 could work as a collector for saliency information arising within layers 2 and 3, as described in Chapter 7. If this were the case, then one might expect to find some poorly oriented neurons in layer 5 that are nevertheless sensitive to centre-surround orientation differences in texture patterns. Some of the neurons tested by Knierim and Van Essen [137] appear to show this property, although these authors did not report in which layer these cells were found.

8.4.3 Special Complex Cells

Special complex cells were investigated by Gilbert [85] and Hammond and Mackay [96] in the cat, and reported by Livingstone and Hubel [150] in the monkey. The essential features of these cells are as follows: They exist in layer 5 of V1, alongside standard complex cells; they have large receptive fields (4° in diameter for the cat); they respond strongly to spots positioned anywhere within the receptive field; they show very little summation along the axis of orientation preference, so that their length tuning functions reach a maximum very quickly; as stimulus length increases, they become more selective to orientation, so for long stimuli they achieve similar tuning curves to normal complex cells; they respond well to random dot textures; they have similar spatial frequency tuning functions to other complex cells; they show high levels of spontaneous activity; they generally have a higher contrast gain than other complex cells (Dean [60]). Here I suggest two possible models for these cells, both of which employ the summation ideas that were introduced in Section 8.3.

In the first model, they receive convergent input from strongly end-stopped geniculate

cells. This end-stopping is assumed to be divisive, but since it is also very strong, stimuli of any significant length reduce the special complex cell input drive to zero. As it stands, this scheme gives a large receptive field that is excited by spots but barely at all by any oriented stimuli because of the end-stopping. In order to produce a flat length summation curve, such a cell would also need to receive secondary excitatory input from standard complex cells so as to balance the fall in response as stimulus length is increased. This model predicts an increase in orientation selectivity with length because there is a progressive changeover from LGN-like properties to standard complex cell properties.

The second model requires convergent input from geniculate neurons that show medium amounts of end-stopping. The amount of end-stopping needs to be just sufficient to balance the summation introduced by convergence and so prevent any increase in response with length after the initial rise. As described, this model would produce the same response to a contour regardless of its length or orientation. In order to produce an increase in orientation selectivity with length, the cell would also have to receive pooled *suppression* from spatially coincident standard complex cells tuned to orientations other than the one desired. In this case, as stimulus length increases then the special complex cell output will only reduce if the orientation is inappropriate—this reduction being mediated by an increasing suppression from other cells.

The second proposal is probably more in line with the neurophysiological data. Sillito [213] reported that some complex cells lost their orientation tuning when treated with the GABA antagonist bicuculline which blocks local inhibitory effects. Although Sillito did not assign a layer to these cells, the description of complex cells as having high spontaneous activity, large receptive fields and a good response to small stimuli fits the special complex group described here.

As they stand, both models introduced above have a problem: they are responsive to diffuse illumination. This arises because geniculate neurons are not entirely suppressed by diffuse stimuli and additive convergence into V1 exacerbates the problem. On the whole, cortical neurons do not respond to diffuse illumination. This problem can be solved by assuming that geniculate input receives some form of unoriented band-pass filtering prior to the convergence stage. As well as removing responses to diffuse illumination, this also has the advantage of matching the spatial frequency bandwidths of special complex cells to that of the standard cells in the same layer. This filtering may well take place in layer 4. A similar effect could also be achieved by spatial frequency domain side-suppression within layer 5.

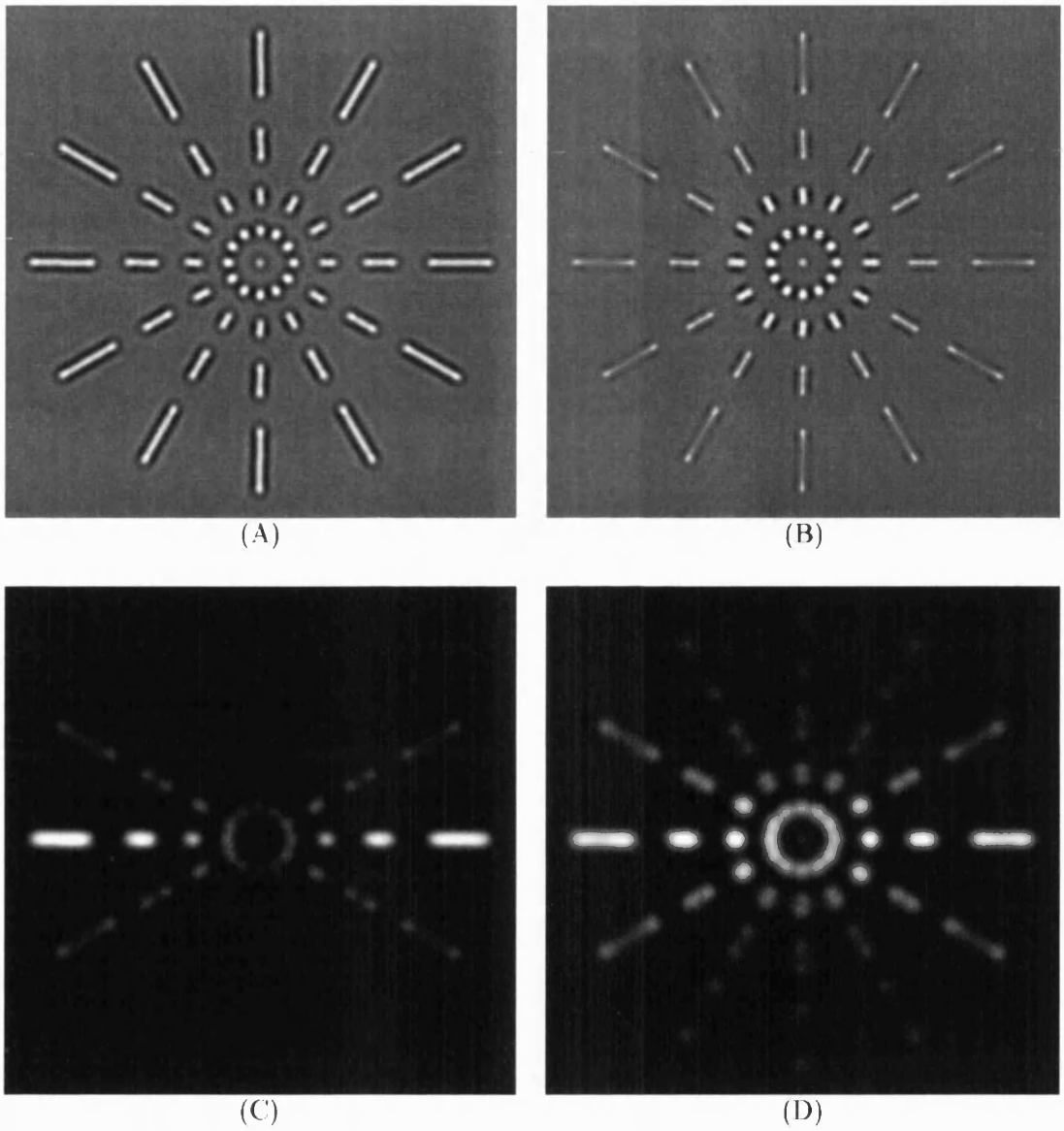


Figure 8-3: Special complex cells and LGN end-stopping. (A) LGN response map. On/off-centre responses are shown as lighter/darker than the grey background. (B) End-stopped LGN signal after suppression approximating the corticogeniculate feedback from layer 6 cells sensitive to long lines. (C) Standard complex cells tuned to horizontal features. Responses increase with bar length. (D) Special complex cell responses made by spatial pooling of a rectified version of the signal show in B, but also including additional excitatory input from C. Responses are similar for all lengths of bar, but the apparent orientation bandwidth narrows with increasing bar length.

8.4.4 Simulation Results

Figure 8-3 shows the results of a simulation of special complex cells using the first model described above. The simulation is simplistic in that it employs no actual feedback connections, but it does however serve to illustrate the arguments presented above.

Figure 8-3A shows a response map for neurons in the LGN when the vision system is presented with a pattern of white radial lines (the original is not shown because it is similar to this response map). Figure 8-3B shows the response map after it has been subject to end-stopping from complex cell receptive fields covering the complete range of orientations. The divisive inhibition used here was similar to that used for end-stopping in Chapter 7. It can be seen that the response to the longer lines has been reduced by this interaction. Figure 8-3C shows a response map for standard complex cells computed using the model described in Chapter 6, while Figure 8-3D shows the response map for special complex cells. This last map was made by applying a spatial low-pass filter to a rectified version of Figure 8-3B (simulating convergence). To this result was then added the complex cell map of Figure 8-3C. Whereas the complex cells are orientation selective and show an increasing response with line length, the special complex cells in Figure 8-3D become progressively more oriented as line length increases, but they produce about the same response to a horizontal line regardless of its length. This is in accord with the characteristic behaviour of special complex cells.

8.4.5 A Cortical Circuit for Gain Control

I stated in Chapter 7 that the ideal contrast measure for gain control should show no variation with stimulus length. In the width dimension, however, some summation can be tolerated, and may even be desirable since it helps to remove noise (Section 7.2.4).

Figure 8-4 details a neural circuit that would yield a special complex cell receptive field that showed no summation with increase in the length of a grating region, but some summation with an increase in its width. The key to this behaviour is the presence of a side-stopped layer 6 projection to the LGN. In addition, this projection must terminate in a diffuse manner. A diffuse termination means that many layer 6 neurons converge on a single LGN cell (probably an inhibitory inter-neuron), and this provides spatial summation.

In the feedback path, we then have side-stopping followed by spatial summation. The spatial summation balances the effect of layer 6 side-stopping (using the principles described in Section 8.3). When this resulting signal inhibits neurons in the LGN, it produces end-stopping, but it cannot produce much side-stopping because the inhibitory signal does not

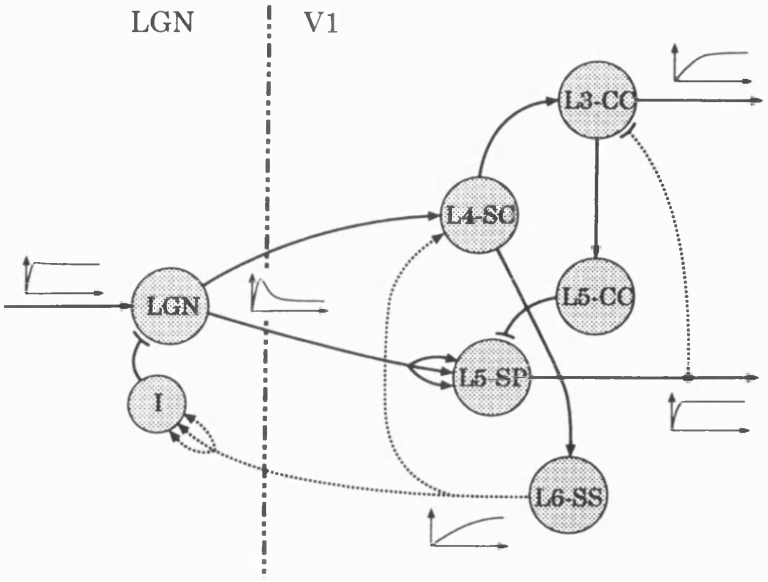


Figure 8-4: Special complex cells and corticogeniculate feedback connections. L6-SS side-stopped cells in layer 6, L4-SC simple cells, L3-CC complex cells, L5-CC standard complex cells in layer 5, L5-SP special complex cells in layer 5, I inhibitory inter-neuron.

increase with an increase in grating width. Convergence is necessary because otherwise, the feedback signal would reduce with grating width due to the layer 6 side-stopping, rather than stay the same. A reducing signal would cause LGN neurons to show less end-stopping with gratings than with isolated bars. In summary, the feedback path should cause geniculate neurons to be end-stopped but not side-stopped. It would be interesting to see these predictions tested experimentally.

The effect of a convergent drive that is end-stopped but not side-stopped from the LGN to special complex cells would be to prevent these cells from showing length summation, but not to prevent them from showing width summation (providing these neurons sum homogeneously over their large receptive field). This is the behaviour that I initially suggested was ideal for gain control: no length summation, but a controlled amount of width summation. Orientation selectivity may then be introduced where necessary by inhibitory input from neighbouring standard complex cells in layer 5.

Gain control could then be introduced to layers 2–4 by inhibitory connections from layer 5. At the same time, facilitatory influences from layer 6 cells to layer 4 could help overcome the subcortical end-stopping in those cells that need to show length summation (Figure 8-4).

This proposal, although complicated, is attractive in that it makes use of established cortical connections and the known properties of receptive fields in each layer. I have also

supported the proposal with some evidence from preliminary simulation work.

8.4.6 Conclusions

In this section, I have described one possible functional role for neurons in layer 5, layer 6 and for the corticogeniculate projection. There are surely other roles for this neural circuitry. I have already suggested that layer 5 might gather saliency information generated in the superficial layers and send it to appropriate subcortical sites. Ferster and Lindström [77] have suggested a role for layer 6 based on the facilitation of responses to colinear arrangements of line elements. However, it should be remembered that neurons in layer 6 are also selective for motion direction, and so the picture is doubtlessly more complicated than this. It will be interesting to see if the suggested roles can be demonstrated by electrophysiological recording.

8.5 Summary

The primary pathway mediating structure vision in primate V1 is from layer 4 to layers 3 and 2 and then on to V2. Receptive fields in these layers are small and provide a detailed analysis of the visual scene in terms of orientation, phase and spatial frequency. Neurons in layer 4 are simple cells which classify the type of feature present at each location. Complex cells in layers 2 and 3 generalise this information to provide oriented contour detection. All these layers contain neurons that are end, side or surround-stopped, producing selected sensitivity to terminations, features with high curvature, or contours with specific lengths. In addition, these mechanisms allow neurons downstream to generalise the same selectivities over space or to create new selectivities. They also can provide saliency-related responses.

Layer 6 produces end-stopping in the LGN via feedback connections. Some neurons in layer 6 are excited by extended edges or bars and could provide a mechanism for contour grouping. They may communicate this property to neurons in layer 4 by means of facilitatory connections.

Layer 5 contains neurons that seem ideally suited for controlling the contrast gain of the superficial layers and for providing a contrast reference for line length judgements. In addition, they may collect and pass on saliency information from layers 2 and 3 to subcortical sites where it plays a role in attention shifting,

In the primate, unoriented neurons in layers 2 and 3 which are selective for particular parts of the visible spectrum sum colour information over variously sized regions of space to create a range of local colour measures. Their suppressive surrounds may then allow colour

complex cells in V2 to collect this information over space without necessarily producing a change in the size of the underlying summation region.

This description is no doubt a large simplification of the spatial processing that takes place in V1. There are probably many subtleties left to discover. Here, I have not considered the processing of stereopsis or motion information and these are liable to add their own complications to the picture.

8.6 Conclusions

In this chapter, I have described various types of length summation properties that are encountered in V1. I have proposed a novel way in which the visual cortex might control its spatial summation so as to obtain a non-additive pooling of neural responses over space using an additive summing junction. I have also suggested some applications for this mechanism. I have concentrated on one which relates to the short summation length shown by some layer 5 neurons, and the fact that this may result from summation control applied within the corticogeniculate feedback circuit. Simulation has been used as a tool to provide support for these proposals.

Chapter 9

Conclusions

9.1 Introduction

In this thesis I have attempted to combine a variety of models of the functionality of parts of the vision system into one simulation. I have been able to demonstrate various interesting behaviours that occur when all the stages are in place. In addition, I have uncovered some problems with our current understanding and I have suggested solutions, that are consistent with our physiological knowledge, in the light of simulation results. Some of these implicate neural pathways for which a role has not previously been identified.

This chapter completes the thesis by reviewing the reasons for building a simulation of this nature and summarises the results that have been obtained: Section 9.2 includes a discussion on what computer simulation has to offer. In Section 9.3 I review the simulation environment used here and examine the role that parallel processing has in stimulating work of this kind. Section 9.4 lists the main results and achievements and also includes suggestions for future research. Section 9.5 contains my final remarks.

9.2 The Role of Simulation

In Chapter 1 I introduced some of the benefits to be obtained from carrying out computer simulations of the vision system. Here, I expand and summarise these points.

The major benefit of a large computational simulation of the operation of the vision system is that it helps to tie together work that is, for the most part, fragmented. Researchers working in experimental psychophysics and neurophysiology tend to create their own models and see the vision system from somewhat different viewpoints. It is advantageous to draw

together these models in order to see if they are mutually self-consistent by means of a unified simulation. Such a simulation then has significant predictive power.

Further advantages of work of this kind are:

Visualisation of Responses. Since neurophysiological recording is, as yet, only able to provide us with access to the world of the single neuron, it is not easy to visualise how the responses from a large number of neurons will relate to each other. Knowing how they relate is important for working out how new selectivities are constructed at higher levels of the visual hierarchy. Knowing how they relate is also important when considering how information is represented in the visual pathway, for instance: How is contrast related to stimulus selectivity in V1? This is a question which I have looked at in some detail when considering gain control in Chapter 7.

In general, image processing simulation of the type described is able to widen our view of neural responses so that we can see how an ensemble of neurons reacts to both artificial and natural stimuli. In addition, we can see whether there are any forms of emergent behaviour that would have been difficult to predict from existing theoretical models. An example is the responsiveness to visual saliency and the orientation tilt illusion described in Chapter 7. These results are entirely a consequence of constructing and simulating a model that is well grounded in the available neurophysiological data.

Many of the results shown in this thesis have been displayed in response-map form. This is not always desirable because there are a great many maps produced in one simulation run and patterns of activity do not always show clearly what measure each local population is encoding. Future research should develop new graphical ways to visualise the data produced by such a simulation.

Constraining Models. In order to carry out a simulation we must specify every parameter in every model that we are using. This forces a quantitative approach and requires us to make our assumptions explicit. It is generally very easy to use terms like "neurons in layer 6 have an inhibitory influence on neurons in the LGN" without specifying exactly what form this influence takes. To build a simulation, we need to include a quantitative model of such behaviour and this is often difficult. A side-effect is that, when quantitative studies are absent from the literature, new lines of experimentation are automatically suggested in order to obtain the required information.

Refining Models. Once we have a simulation in place and results have been obtained,

the fidelity of each constituent model can be evaluated. In particular, it is sometimes found that there are problems with the simulation, e.g. side-effects or inappropriate responses may be evident that do not seem to relate to the way that real neurons might reasonably be expected to behave. These effects highlight inadequacies in modelling the experimentally observed behaviour and lead to a refinement of the model responsible. An example of this is the problem experienced with spatial summation in the contrast gain control model introduced in Chapter 7. This difficulty led to the discovery of a new method of implementing spatial convergence as described in Chapter 8. It is very hard to identify such weaknesses without the aid of simulation.

As well as revealing the failings of our assumptions when creating models in the first place, the experience of continuing simulation cycles is able to teach us about some of the pitfalls that frequently occur when thinking about the functionality of the vision system.

Robustness is also an issue. The vision system is known to be very robust and able to show graceful degradation. Simulation results can tell us to what extent our models capture this ability. This is especially important when noise is introduced, or natural images are used as a stimulus.

Suggesting Future Work. Simulations allow us to visualise responses and clarify and refine our models. They also leads to a greater understanding of what the vision system is computing, which stimulus properties are explicitly represented, and which problems the system needs to solve at later stages of processing. All these insights provide their own prompts for further experimental and theoretical work.

In addition to the items described above, it is important that simulation attempts have a strong base in results from neurophysiology and psychophysics otherwise there is little reason for such attempts to be taken seriously by experimentalists. This points to a close collaboration between researchers who are gathering data and those who are concerned with computational modelling.

9.3 Simulation and Parallel Processing

In this thesis, I have obtained most of the results by using a computer which is capable of parallel processing. This section includes a number of reasons why such an approach is appropriate.

Using a parallel machine required the development of a specialised software environment suitable for image processing. Simulation programs were then written that made use of this framework. In general, implementing models in parallel is difficult, but once the environment is in place, testing out a theory or implementing a new model generally takes only a short amount of time.

A parallel architecture, particularly the single-instruction multiple-data type used here, is well suited to a simulation of the vision system because each processing operation is repeated across every spatial location and neurons analyse the image in parallel. When writing parallel software to simulate this configuration it is possible to concentrate on the operation to be performed while ignoring the fact that it is replicated over space. This permits better concentration on the modelling task.

The major benefit to be obtained from using a parallel approach is one of speed. If the simulation is very large then this is the only way in which results can be calculated in a reasonable time. A fast run is important when results that require a comparison across multiple test images are to be gathered.

Although parallel computation promises a speed increase over sequential processing, the future of visual simulation will require rather better performance than that which is currently achieved. The MASPAR machine used here ran the entire simulation in around 10 minutes. This is at least a factor of eight faster than any other machine available within the Computing Group research laboratory at the moment—but this is already too slow if the simulation were expanded to include motion processing since many frames of an input sequence would have to be analysed. Image processing hardware could be used to improve performance for low-level operations, but unfortunately, this loses some of the flexibility that is required for modifying the models that is available with a software simulation.

9.4 Summary of Results and Suggestions for Further Work

In this section I outline the major contributions made by the work presented in this thesis and I attempt to show which areas are deficient and what future research directions are possible.

Retina Model. In Chapter 4 I described a multistage model which accounted for the transformation from image values stored in a computer's memory to neural activation in the optic nerve fibres of a primate viewing the displayed image. This model included the effects of receptor adaptation and also spatial opponency. The inclusion of a real-

istic receptor model had the important consequence of allowing contrast and relative brightness signals to be obtained, while achieving a significant contrast and brightness constancy. This model could be improved by updating the cone model and spatial summation stages so as to separate the different sites of retinal adaptation.

Retinal Contrast Gain Control. Chapter 4 also included a novel simulation of the spatial effects of contrast gain control in the retina. The results of this simulation suggest that the compressive contrast response seen in the **M**-cell pathway is the result of contrast gain control, and that the apparent saturation need not be associated with local signal distortion. Further analysis of this mechanism would allow a full spatiotemporal comparison of signals carried by the **P** and **M** pathways.

Chromatic Mechanisms in the LGN. These mechanisms were described in Chapter 5. There, I was able to show that the type I **P**-cell receptive fields carry mixed chrominance and luminance signals that are separable, and form the basis for red-green colour and black-white brightness perceptions at low temporal frequencies. Simulation results for all LGN receptive field types were included. Preliminary simulations have shown how these signals may be treated in the cortex, and it would be interesting to include the colour dimension in a simulation of receptive fields in V1.

Simple and Complex Cells. In Chapter 6 I introduced a simple cell model which used a modified Gabor function in order to produce spatial frequency, phase and orientation selectivity. A number of complex cell models were also described and simulation results presented. In this chapter I included an examination of spatial representation in V1 and I suggested a novel sampling strategy. At present, the simulations of V1 do not include an input from the **M** pathway. The correspondence with neurophysiology could also be improved by using a simple cell model which made proper use of the spatial selectivity that is already present in the input received from the LGN. The LGN stage itself could be improved by including the effects of integrative action as these become clarified by experimental work.

Cortical Contrast Gain Control. I described contrast gain control in the first half of Chapter 7. I was able to show that this mechanism produces an overall log-like response to gratings in the cortex, but without necessarily introducing surround effects like those produced by contrast gain control in the retina. I also demonstrated that because contrast gain control includes summation over space and spatial frequency it reduces

the sensitivity of the target cells to visual noise—a highly desirable behaviour. The difficulty I found with current models of contrast gain control relates to the effect of stimulus selectivity in the control path. In Chapter 8 I proposed a mechanism to solve this problem, but this has not yet been included in the full simulation.

Stopping Mechanisms in V1. The second half of Chapter 7 included a simulation of end-, side-, and surround-stopping in V1. This was made possible by a new model for the divisive inhibition that is implicated in forming these three receptive field types. As well as demonstrating the selectivity for stimulus length and curvature that is shown by neurons which are end-stopped, I was able to present results which implicate surround-stopped neurons in the generation of saliency effects. Response to the orientation tilt illusion was found to be a side-effect of this behaviour. Further work is necessary to uncover how end-stopped neurons can locate edge discontinuities in a way that is useful for the extraction of occlusion cues.

Summation Control. In Chapter 8 I proposed a novel mechanism for spatial generalisation at many levels of the vision system. This mechanism involved surround suppression followed by linear spatial summation. This can then be followed by subsampling. I suggested that summation of this form may have a role in creating large receptive fields in IT from those in V4 with suppressive surrounds. I demonstrated that neurons in V1 which are subject to various types of spatial inhibition can make use of homogeneous summation to introduce a variety of new form selectivities in V2. There are certain to be wide applications for this mechanism.

Special Complex Cells. I described special complex cells in Chapter 8 as well as suggesting two possible models for their behaviour. This was done in part to demonstrate that summation control may well be taking place within the geniculo-cortical loop and in part because this receptive field type seems to be a good candidate for imposing a gain control influence on neurons in the superficial layers. Future simulation work should include the effect of corticogeniculate feedback on the formation of simple and complex cell selectivities.

In summary, there is plenty of scope for future simulation work, but a number of important results have already been obtained which taken together form a good base.

9.5 Conclusions

V1 contains a representation of the visual scene in terms of the colour, size, position, contrast, orientation, motion, and possibly, saliency of its elements. It is a representation that, although filtered through higher visual centres, is the most detailed of those available to consciousness, and therefore defines the limits for our ability to make localised spatial discriminations.

In this thesis, I have attempted to capture a part of this representation by means of a computer simulation. In doing so, I have had to identify and model the mechanisms that underly its formation. There is still much work to do, but I have been successful in generating the main spatial selectivities used in the neural representation and in explaining some of its idiosyncrasies, e.g. the orientation tilt illusion.

Some of the simulation results obtained have been constrained by the limited space that is available here: in a simulation run so many response maps are produced that it is often difficult to find a way of representing the important response properties with only a few diagrams. Finding a solution to this, is a research task in itself.

As simulations attempt to delve further into the vision system, it will become necessary to examine the method by which a neuron's stimulus selectivity begins to be dependent on visual experience. If a generic learning algorithm is applicable to the behaviour of neurons in areas V4, TE and IT then it may soon be possible to simulate neural activity in these areas too, and in doing so, we may be very close to an understanding of the mind's eye.

Appendix A

Sampling Theory

Introduction

This appendix contains the basic proofs of sampling theory: derivation of the Nyquist criterion, sampling and reconstruction. One dimensional treatments are given, but these are also applicable to two dimensions.

Impulse Combs

This section proves that a space-domain comb of impulses $\delta_T(x)$ remains as a comb of impulses when transformed to the Fourier domain. The function $\delta_T(x)$ has period T and frequency ω_0, where $\omega_0 T = 2\pi$. The delta function is $\delta(x)$ and the complex operator is j.

We can express any periodic function $f(x)$ as a complex Fourier series

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \exp(jn\omega_0 x) \tag{A.1}$$

where

$$c_n = \frac{1}{T}\int_{\theta}^{\theta+T} f(x)\exp(-jn\omega_0 x)\,dx. \tag{A.2}$$

Fourier transforming $f(x)$ we get

$$\mathbf{F}(\omega) = \sum_{-\infty}^{\infty} c_n\, 2\pi\delta(\omega - n\omega_0) \tag{A.3}$$

since $\mathcal{F}\exp(jn\omega_0 x) = 2\pi\delta(\omega - n\omega_0)$. Now if

$$f(x) = \delta_T(x) = \sum_{N=-\infty}^{\infty} \delta(x - NT) \tag{A.4}$$

we can express this as a Fourier series and calculate c_n:

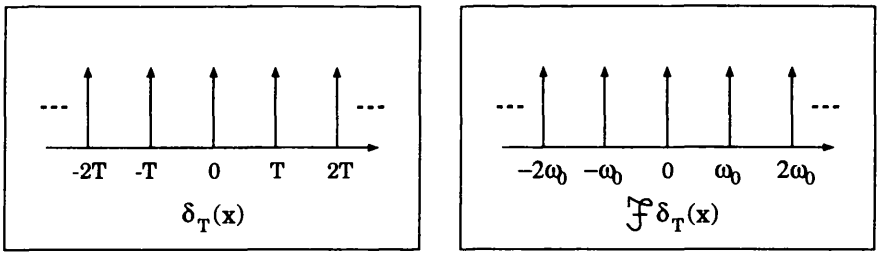
$$\begin{aligned} c_n &= \frac{1}{T}\int_{-T/2}^{T/2} \sum_{N=-\infty}^{\infty} \delta(x - NT)\exp(-jn\omega_0 x)\,dx && \text{(A.5)} \\ &= \frac{1}{T}\int_{-\infty}^{\infty} \delta(x)\exp(-jn\omega_0 x)\,dx && \text{(A.6)} \\ &= \frac{1}{T}\mathcal{F}\,\delta(x) = \frac{1}{T}. && \text{(A.7)} \end{aligned}$$

Therefore $\delta_T(x)$ expressed as a complex Fourier series is:

$$\delta_T(x) = \sum_{-\infty}^{\infty} \frac{1}{T}\exp(jn\omega_0 x). \tag{A.8}$$

From Equation A.3 above,

$$\begin{aligned} \mathcal{F}\,\delta_T(x) &= \sum_{-\infty}^{\infty} \frac{1}{T} 2\pi\delta(\omega - n\omega_0) && \text{(A.9)} \\ &= \sum_{-\infty}^{\infty} \omega_0\,\delta(\omega - n\omega_0). && \text{(A.10)} \end{aligned}$$



Sampling Theory

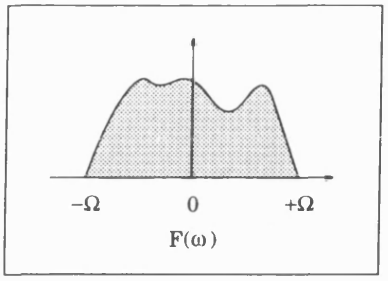
This section derives the frequency domain response when a band-limited signal is sampled by a train of impulses $\delta_T(x)$ with spatial interval T, giving a sampling frequency of ω_0.

A function $f(x)$ is sampled by multiplying by $\delta_T(x)$:

$$f(x)\delta_T(x) = f(x)\sum_{-\infty}^{\infty} \delta(x - NT) \tag{A.11}$$

$$= \sum_{-\infty}^{\infty} f(NT)\,\delta(x - NT). \tag{A.12}$$

Now suppose that $\mathbf{F}(\omega)$ $(= \mathfrak{F} f(x))$ is band-limited so that $\mathbf{F}(\omega) = 0$ for $|\,\omega\,| \geq \Omega$.



Then since $\mathfrak{F}[f(x)\,g(x)] = \frac{1}{2\pi}[\mathbf{F}(\omega) \otimes \mathbf{G}(\omega)]$, where $\otimes$ denotes convolution in space,

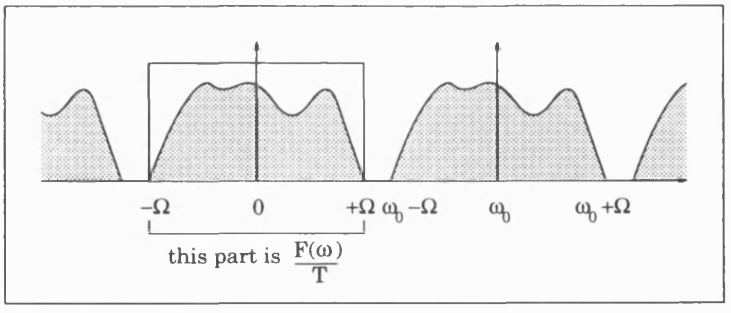
$$\mathfrak{F}[\,f(x)\,\delta_T(x)\,] = \frac{1}{2\pi}\mathbf{F}(\omega) \otimes \mathfrak{F}\delta_T(x) \tag{A.13}$$

$$= \frac{1}{2\pi}\mathbf{F}(\omega) \otimes \sum_{-\infty}^{\infty} \frac{2\pi}{T}\,\delta(\omega - n\omega_0) \tag{A.14}$$

$$= \frac{1}{T}\int_{-\infty}^{\infty} \mathbf{F}(\omega - u) \sum_{-\infty}^{\infty} \delta(u - n\omega_0)\,du \tag{A.15}$$

$$= \frac{1}{T}\sum_{-\infty}^{\infty} \mathbf{F}(\omega - n\omega_0). \tag{A.16}$$

If the shifted $\mathbf{F}(\omega)$ parts are separated then we can obtain $\mathbf{F}(\omega)$ by multiplying by $T[\,H(\omega + \Omega) - H(\omega - \Omega)\,]$, where $H(\cdot)$ is the Heaviside step function. Separation is possible if there is no overlap i.e. $\omega_o \geq 2\Omega$. It therefore follows that a band-limited spatial waveform may be reconstructed from its sample values providing that the sampling frequency is greater than twice the maximum signal frequency.



Reconstruction

Sampling a spatial signal has resulted in a train of samples with a spectrum containing multiple copies of the original spectrum. We can now reconstruct the oringinal signal from

these samples, as proven below.

For $\omega_0 = 2\Omega$:

$$\mathbf{F}(\omega) = T[\,H(\omega+\Omega) - H(\omega-\Omega)\,] \cdot \frac{1}{T}\sum_{-\infty}^{\infty}\mathbf{F}(\omega - n\omega_0) \tag{A.17}$$

therefore

$$f(x) = \mathcal{F}^{-1}\{\,[\,H(\omega+\Omega) - H(\omega-\Omega)\,] \cdot \sum_{-\infty}^{\infty}\mathbf{F}(\omega - n\omega_0)\,\} \tag{A.18}$$

$$= \mathcal{F}^{-1}[\,H(\omega+\Omega) - H(\omega-\Omega)\,] \otimes \mathcal{F}^{-1}\sum_{-\infty}^{\infty}\mathbf{F}(\omega - n\omega_0) \tag{A.19}$$

$$= \frac{\sin\Omega x}{\pi x} \otimes T\sum_{-\infty}^{\infty} f(NT)\,\delta(x - NT) \tag{A.20}$$

$$= \int_{-\infty}^{\infty} \frac{\sin\Omega(x-\lambda)}{\pi(x-\lambda)}\, T\sum_{-\infty}^{\infty} f(NT)\,\delta(\lambda - NT)\,d\lambda \tag{A.21}$$

$$= \frac{\Omega T}{\pi}\int_{-\infty}^{\infty}\sum_{-\infty}^{\infty}\frac{\sin\Omega(x-\lambda)}{\Omega(x-\lambda)}\, f(NT)\,\delta(\lambda - NT)\,d\lambda$$

$$= \frac{\Omega T}{\pi}\sum_{-\infty}^{\infty}\frac{\sin\Omega(x-NT)}{\Omega(x-NT)}\, f(NT)$$

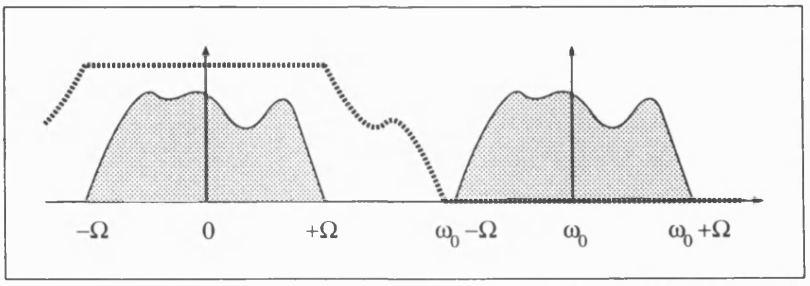
hence

$$f(x) = \sum_{-\infty}^{\infty}\frac{\sin[\,(\pi/T)(x-NT)\,]}{(\pi/T)(x-NT)}\, f(NT). \tag{A.22}$$

This means that reconstruction can be carried out by multiplying each sample by a $\sin(x)/x$ curve and summing these together.

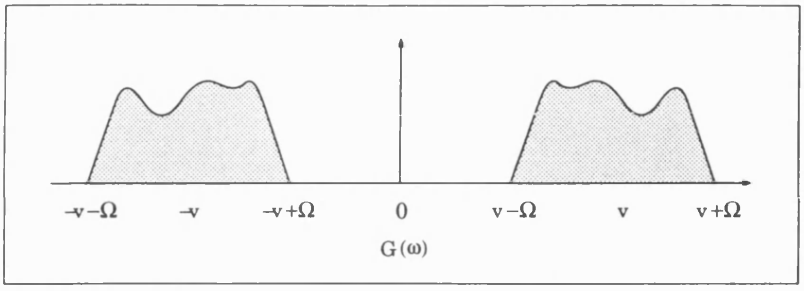
Note

Note that if $\omega_0 > 2\Omega$ then room exists between the instances of $\mathbf{F}(\omega)$ so a more relaxed specification can be used for the reconstruction filter. What is required is that the filter's frequency response equals T for $|\omega| < \Omega$ and equals zero for $|\omega| > \omega_0 - \Omega$.



Quadrature Sampling

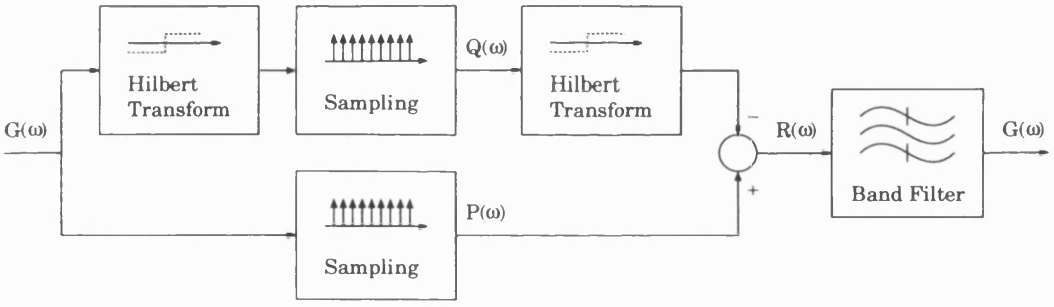
This section shows that it is possible to sample and reconstruct a band-limited signal using a sampling rate that is dependent only on the signal bandwidth.



Start with a band-limited real-valued signal $g(x)$ (centre frequency v, bandwidth 2Ω) with Fourier transform $G(\omega)$ given by

$$G(\omega) = F(-\omega - v) + F(\omega - v) \tag{A.23}$$

Here, $\Omega > 0$ and $v > \Omega$; also $F(\omega)$ is zero for $|\omega| > \Omega$.



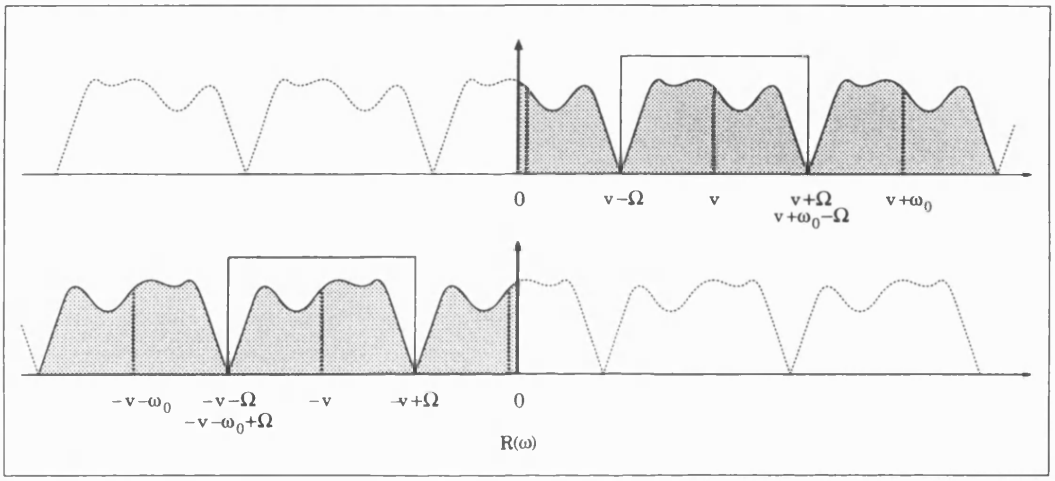
$P(\omega)$ and $Q(\omega)$ are then the phase and quadrature sampled versions of $G(\omega)$, where $Q(\omega)$ is produced using the Hilbert transform (equivalent to having both odd and even receptive fields).

$$\begin{aligned} P(\omega) &= \frac{1}{2\pi} G(\omega) \otimes \mathcal{F}\delta_T(x) && \text{(A.24)} \\ &= \frac{1}{T}\sum_{-\infty}^{\infty} [F(-(\omega - n\omega_0) - v) + F((\omega - n\omega_0) - v)] && \text{(A.25)} \end{aligned}$$

$$\begin{aligned} Q(\omega) &= \frac{1}{2\pi} G(\omega)\,[-j\,\mathrm{sgn}(\omega)] \otimes \mathcal{F}\delta_T(x) && \text{(A.26)} \\ &= -j\frac{1}{T}\sum_{-\infty}^{\infty} [-F(-(\omega - n\omega_0) - v) + F((\omega - n\omega_0) - v)] && \text{(A.27)} \end{aligned}$$

To reconstruct $g(x)$, $Q(\omega)$ is first modified by applying a Hilbert transform, then the difference $P(\omega) - Q(\omega)$ is formed to give $R(\omega)$

$$\begin{aligned} R(\omega) &= P(\omega) - Q(\omega)\,[-j\,\mathrm{sgn}(\omega)] && \text{(A.28)} \\ &= \frac{1}{T}\Big[\sum_{-\infty}^{\infty}[F(-(\omega - n\omega_0) - v) + F((\omega - n\omega_0) - v)] \\ &\quad + \mathrm{sgn}(\omega)\sum_{-\infty}^{\infty}[-F(-(\omega - n\omega_0) - v) + F((\omega - n\omega_0) - v)]\Big] && \text{(A.29)} \\ &= \begin{cases} \frac{2}{T}\sum_{-\infty}^{\infty} F((\omega - n\omega_0) - v) & \text{for } \omega > 0 \\ \frac{2}{T}\sum_{-\infty}^{\infty} F(-(\omega - n\omega_0) - v) & \text{for } \omega < 0 \end{cases} && \text{(A.30)} \end{aligned}$$



The diagram above shows that, rather than requiring $\omega_0 > 2(v+\Omega)$ as in standard Nyquist sampling, providing that $\omega_0 > 2\Omega$, the original band-limited signal $g(x)$ can be reconstructed using a suitable box filter since the $F(\omega)$ components do not overlap in the frequency domain.

$$\begin{aligned} G(\omega) &= R(\omega)\frac{T}{2}\,[\,H(\omega - (-v - \Omega)) - H(\omega - (-v + \Omega)) \\ &\quad + H(\omega - (v - \Omega)) - H(\omega - (v + \Omega))\,] && \text{(A.31)} \\ g(x) &= \mathcal{F}^{-1}R(\omega) \otimes [\,T\cos vx\,\frac{\sin \Omega x}{\pi x}\,]. && \text{(A.32)} \end{aligned}$$

Appendix B

Simulation Parameters

```
# ========================================================================
#  Maspar vision resource file
# ========================================================================

nyquist_limit           = 25.6          # cycles per degree
input_size              = 512           # pixels square

# ========================================================================
#  TV Model
# ========================================================================

phosphor.x[0]           = 0.67
phosphor.y[0]           = 0.33
phosphor.x[1]           = 0.21
phosphor.y[1]           = 0.71
phosphor.x[2]           = 0.14
phosphor.y[2]           = 0.08
phosphor.Wx             = 0.31
phosphor.Wy             = 0.316
intensity.max           = 100.0         # cd/m^2
intensity.background    = 0.5           # cd/m^2
intensity.power         = 2.2

# ========================================================================
# Cone Model
# ========================================================================

cone.model.flag         = 0             # 0 = EP model, 1 = linear
cone.adapt.flag         = 1             # 0 = manual, 1 = auto, 2 = local
cone.adapt.level        = 100.0         # (manual) cd/m^2
cone.adapt.gain         = 5.83
cone.adapt.power        = 0.69
cone.adapt.min          = 0.25          # cd/m^2
cone.adapt.rc           = 1.5           # deg
cone.sense[0]           = 0.80          # relative pigment
cone.sense[1]           = 1.00          #  sensitivities
```

```
cone.sense[2]           = 0.10
cone.sense[3]           = 0.05          # extra map for Red input to BY

# ============================================================================
# Opponent Colour Site
# ============================================================================

opponent.RG.ratio       = 1.0           # R/G ratio
opponent.BY.ratio       = 1.0           # R/B ratio
achromatic.RG.ratio     = 2.0           # R/G ratio

# ============================================================================
# Lateral Geniculate Nucleus
# ============================================================================

# parvocellular red-green type I system
parvo.RG.chan_count     = 4

#parvo.RG.rc[0]         = 0.015         # flat filter to about 10cpd
#parvo.RG.rs[0]         = 0.050
#parvo.RG.mu[0]         = 0.099
#parvo.RG.gain[0]       = 1.11

#parvo.RG.rc[0]         = 0.025         # DOG peaked at 5cpd
#parvo.RG.rs[0]         = 0.100
#parvo.RG.mu[0]         = 0.9
#parvo.RG.gain[0]       = 1.27

parvo.RG.rc[0]          = 0.015         # Four mechanism CS curve
parvo.RG.rs[0]          = 0.072
parvo.RG.mu[0]          = 0.89
parvo.RG.gain[0]        = 0.3952

parvo.RG.rc[1]          = 0.029
parvo.RG.rs[1]          = 0.069
parvo.RG.mu[1]          = 0.84
parvo.RG.gain[1]        = 0.2501

parvo.RG.rc[2]          = 0.029
parvo.RG.rs[2]          = 0.159
parvo.RG.mu[2]          = 0.83
parvo.RG.gain[2]        = 0.4437

parvo.RG.rc[3]          = 0.030
parvo.RG.rs[3]          = 0.202
parvo.RG.mu[3]          = 0.67
parvo.RG.gain[3]        = 0.2825

# parvocellular blue-yellow type II system
parvo.BY.rc             = 0.070
parvo.BY.gain           = 1.0

# magnocellular achromatic type III system
magno.enable            = 1
magno.rc                = 0.053
magno.rs                = 0.200
```

```
magno.mu                = 1.0
magno.gain              = 200.0
magno.gain.k            = 0.05
magno.gain.beta         = 0.121
magno.gain.gamma        = 1.0
magno.gain.rg           = 0.400
magno.gain.rmax         = 120.0

# ============================================================================
# Cortex Area V1
# ============================================================================

# Colour Properties
V1.sc.cone_ratio        = 1.6330        # R/G ratio for V(lambda) or other.

# Red-green type II neurons
V1.type2.rc             = 0.07          # (also m1 rc, equal to lgn BY)
V1.type2.gain           = 1.0

# Colour/Brightness channels (double opponent)
V1.cb.enable            = 0             # 0 = disable colour processing
V1.cb.rcrs[0]           = 0.14          # m1 rs, m2 rc
V1.cb.rcrs[1]           = 0.35          # m2 rs, m3 rc
V1.cb.gain[0]           = 1.5           # m2 gain
V1.cb.gain[1]           = 0.4           # m3 gain

# Simple cells
V1.orients              = 8
V1.sc.load              = 0             # 0 = make, 1 = load from disc
V1.sc.scale             = 515.7         # input conversion -> contrast (0-100)

# Complex Cells
V1.cc.load              = 0             # 0 = make, 1 = load from disc
V1.cc.smooth.flag       = 0             # 0 = off, 1 = slight smoothing
V1.cc.smooth.ratio      = 0.6           # rc compared to effective cx rc

# Gain Control
V1.gc.load              = 0
V1.gc.intersf           = 1             # 1 = include inter-channel effects
V1.gc.sum.orients       = 1
V1.gc.sum.space         = 1
V1.gc.iterations        = 5
V1.gc.intersf.adj       = 1.8
V1.gc.rcratio           = 2.8           # gain control radius ratio
V1.gc.sc.rmax           = 100           # max response
V1.gc.sc.gain           = 0.8           # gain control feedback constant
V1.gc.sc.sat            = 5.0           # gain control half saturation value
V1.gc.sc.power          = 2.4           # gain control power
V1.gc.cc.rmax           = 100
V1.gc.cc.gain           = 0.8           # high contrast gratings drive to
V1.gc.cc.sat            = 5.0           #  about 65% for feed-forward
V1.gc.cc.power          = 2.4
V1.gc.save              = 0

# End/Side/Surround Stopping
V1.stop.chan            = 2
```

```
V1.stop.obwratio.in     = 2.5          # relative to chan orientation bw

# Surround stopped Cells
V1.os.rc                = 1.9          # ratio to chan effective rc
V1.os.rs                = 3.0          # ratio to chan effective rc
V1.os.strength          = 8.0          # divisive surround strength
V1.os.thresh            = 0.0

# End stopped Cells
V1.es.obw.out           = 80           # (degs)
V1.es.rc                = 1.9
V1.es.rs                = 3.0
V1.es.strength          = 8.0
V1.es.thresh            = 0.0

# Side stopped Cells
V1.ss.obw.out           = 80
V1.ss.rc                = 1.9
V1.ss.rs                = 3.0
V1.ss.strength          = 8.0
V1.ss.thresh            = 0.0

# Special complex cells
V1.sp.offset            = 1.85
V1.sp.strength          = 9.0
V1.sp.gain_to_cc        = 1.5
V1.sp.save              = 0

# ============================
#  Spatial Frequency Channels
# ============================
V1.sc.enable[0]         = 1
V1.sc.freq[0]           = 0.7          # frequency (cpd)
V1.sc.bw[0]             = 2.15         # bandwidth (octaves)
V1.sc.orientbw[0]       = 78           # orientation bw (degs)
V1.sc.sample[0]         = 16           # sampling factor
V1.sc.save[0]           = 0
V1.cc.save[0]           = 1

V1.sc.enable[1]         = 1
V1.sc.freq[1]           = 1.4
V1.sc.bw[1]             = 1.85
V1.sc.orientbw[1]       = 68
V1.sc.sample[1]         = 8
V1.sc.save[1]           = 0
V1.cc.save[1]           = 1

V1.sc.enable[2]         = 1
V1.sc.freq[2]           = 2.6
V1.sc.bw[2]             = 1.59
V1.sc.orientbw[2]       = 60
V1.sc.sample[2]         = 4
V1.sc.save[2]           = 0
V1.cc.save[2]           = 1

V1.sc.enable[3]         = 1
```

```
V1.sc.freq[3]           = 4.4
V1.sc.bw[3]             = 1.36
V1.sc.orientbw[3]       = 52
V1.sc.sample[3]         = 4
V1.sc.save[3]           = 0
V1.cc.save[3]           = 1

V1.sc.enable[4]         = 1
V1.sc.freq[4]           = 7.2
V1.sc.bw[4]             = 1.14
V1.sc.orientbw[4]       = 44
V1.sc.sample[4]         = 2
V1.sc.save[4]           = 0
V1.cc.save[4]           = 1

V1.sc.enable[5]         = 1
V1.sc.freq[5]           = 10.5
V1.sc.bw[5]             = 0.98
V1.sc.orientbw[5]       = 38
V1.sc.sample[5]         = 2
V1.sc.save[5]           = 0
V1.cc.save[5]           = 1

V1.sc.enable[6]         = 1
V1.sc.freq[6]           = 15.0
V1.sc.bw[6]             = 0.82
V1.sc.orientbw[6]       = 32
V1.sc.sample[6]         = 2
V1.sc.save[6]           = 0
V1.cc.save[6]           = 1

# ============================================================================
# Misc Image Saving Flags
# ============================================================================

cone.save               = 0
parvo.RG.save           = 0
parvo.BY.save           = 0
magno.save              = 0
```

Bibliography

[1] E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2:284–299, 1985.

[2] D. G. Albrecht and R. L. De Valois. Striate cortex responses to periodic patterns with and without the fundamental harmonics. *Journal of Physiology*, 319:497–514, 1981.

[3] D. G. Albrecht, R. L. De Valois, and L. G. Thorell. Visual cortical neurons: Are bars or gratings the optimal stimuli? *Science*, 207:88–90, 1980.

[4] T. D. Albright. Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, 52:1106–1130, 1984.

[5] B. W. Andrews and D. A. Pollen. Relationship between spatial frequency selectivity and receptive field profile of simple cells. *Journal of Physiology*, 287:163–176, 1979.

[6] D. R. Badcock. How do we discriminate relative spatial phase? *Vision Research*, 24:1847–1857, 1984.

[7] J. S. Baizer. Receptive field properties of V3 neurons in monkey. *Investigative Opthalmology and Visual Science*, 23:87–95, 1982.

[8] H. B. Barlow. Summation and inhibition in the frog's retina. *Journal of Physiology*, 119:69–88, 1953.

[9] H. B. Barlow. Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1:371–394, 1972.

[10] H. B. Barlow, T. P. Kaushal, M. Hawken, and A. J. Parker. Human contrast discrimination and the threshold of cortical neurons. *Journal of the Optical Society of America A*, 4:2366–2371, 1987.

[11] H. B. Barlow and W. R. Levick. The mechanism of directional selective units in rabbit's retina. *Journal of Physiology*, 178:477–504, 1965.

[12] H. G. Barrow and A. J. Bray. Activity-induced "colour blob" formation. In I. Aleksander and J. Taylor, editors, *Artificial Neural Networks 2.* Elsevier Science Publishers B. V., 1992.

[13] H. G. Barrow and A. J. Bray. A model of adaptive development of complex cortical cells. In I. Aleksander and J. Taylor, editors, *Artificial Neural Networks 2.* Elsevier Science Publishers B. V., 1992.

[14] L. A. Bauman and A. B. Bonds. Inhibitory refinement of spatial frequency selectivity in single cells of the cat striate cortex. *Vision Research*, 31:933–944, 1991.

[15] W. H. A. Beaudot and J. Hérault. Motion processing in the retina: About a velocity matched filter. Technical report, Tirf Lab - Neuronics Report, Laboratoire de Traitement d'Images et Reconnaissance de Formes, Institut National Polytechnique de Grenoble, 46 avenue Félix Viallet, 38031 Grenoble Cedex, France, February 1993.

[16] W. H. A. Beaudot and J. Hérault. A neurobiological and psychophysical consistent model of the vertebrate retina. In *17th European Conference on Visual Perception*, Eindhoven, The Netherlands, September 4–8 1994.

[17] W. H. A. Beaudot, P. Palagi, and J. Hérault. Realistic simulation tool for early visual processing including space, time and colour data. In J. Mira, J. Cabestany, and A. Prieto, editors, *Lecture Notes in Computer Science 686: "New Trends in Neural Computation"*, pages 370–375. Springer Verlag, 1993.

[18] E. A. Benardete, E. Kaplan, and B. W. Knight. Contrast gain control in the primate retina: P-cells are not X-like, some M-cells are. *Visual Neuroscience*, 8:483–486, 1992.

[19] J. R. Bergen and B. Julesz. Rapid discrimination of visual patterns. *IEEE Transactions on Systems, Man, and Cybernetics*, 13:857–863, 1983.

[20] V. A. Billock. The relationship between simple and double opponent cells. *Vision Research*, 31:33–42, 1991.

[21] P. O. Bishop, J. S. Coombs, and G. H. Henry. Interaction effects of visual contours on the discharge frequency of simple striate neurones. *Journal of Physiology*, 219:659–687, 1971.

[22] P. O. Bishop, J. S. Coombs, and G. H. Henry. Responses to visual contours: Spatio-temporal aspects of excitation in the receptive fields of simple striate neurones. *Journal of Physiology*, 219:625–657, 1971.

[23] J. Bolz and C. D. Gilbert. Generation of end-inhibition in the visual cortex via interlaminar connections. *Nature*, 320:362–365, 1986.

[24] A. B. Bonds. Temporal dynamics of contrast gain in single cells of the cat striate cortex. *Visual Neuroscience*, 6:239–255, 1991.

[25] A. B. Bonds and C. Enroth-Cugell. Spatial consequences of bleaching adaptation in cat retinal ganglion cells. *Journal of Physiology*, 318:339–354, 1981.

[26] O. J. Braddick. Low- and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London, Series B*, 290:137–151, 1979.

[27] A. Bradley, B. C. Skottun, I. Ohzawa, G. Sclar, and R. D. Freeman. Visual orientation and spatial frequency discrimination: A comparison of single neurons and behaviour. *Journal of Neurophysiology*, 57:755–772, 1987.

[28] J. M. H. du Buf. Detection symmetry and asymmetry. *Spatial Vision*, 5:189–203, 1991.

[29] J. M. H. du Buf. Abstract processes in texture discrimination. *Spatial Vision*, 6:221–242, 1992.

[30] J. M. H. du Buf. Brightness versus apparent contrast 1: Incremental and decremental disks with varying diameter. *Spatial Vision*, 6:159–182, 1992.

[31] J. M. H. du Buf. Modelling spatial vision at the threshold level. *Spatial Vision*, 6:25–60, 1992.

[32] J. M. H. du Buf. Ambiguities in Gabor space. In A. Gale, editor, *Visual Search III*. Taylor and Francis, London, 1993.

[33] J. M. H. du Buf. Responses of simple cells: Events, interferences, and ambiguities. *Biological Cybernetics*, 68:321–333, 1993.

[34] J. M. H. du Buf. Ramp edges, Mach bands, and the functional significance of the simple cell assembly. *Biological Cybernetics*, 70:449–461, 1994.

[35] C. A. Burbeck and D. H. Kelly. Spatiotemporal characteristics of visual mechanisms: Excitatory-inhibitory model. *Journal of the Optical Society of America*, 70:1121–1126, 1980.

[36] D. C. Burr, M. C. Morrone, and D. Spinelli. Evidence for edge and bar detectors in human vision. *Vision Research*, 29:419–431, 1989.

[37] D. C. Burr and J. Ross. Visual analysis during motion. In M. A. Arabib and A. R. Hanson, editors, *Vision, Brain and Cooperative Computation*, chapter 5, pages 187–207. MIT Press, 1985.

[38] D. C. Burr, J. Ross, and M. C. Morrone. Seeing objects in motion. *Proceedings of the Royal Society of London, Series B*, 227:249–265, 1986.

[39] P. J. Burt. The interdependence of temporal and spatial information in early vision. In M. A. Arabib and A. R. Hanson, editors, *Vision, Brain and Cooperative Computation*, chapter 8, pages 263–277. MIT Press, 1985.

[40] T. Caelli and P. Bevan. Visual sensitivity to two-dimensional spatial phase. *Journal of the Optical Society of America*, 72:1375–1381, 1972.

[41] T. Caelli and B. Julesz. Psychophysical evidence for global feature processing in visual texture discrimination. *Journal of the Optical Society of America*, 69:675–679, 1979.

[42] F. W. Campbell, G. F. Cooper, and C. Enroth-Cugell. The spatial selectivity of the visual cells of the cat. *Journal of Physiology*, 203:223–235, 1969.

[43] F. W. Campbell and J. G. Robson. Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197:551–566, 1968.

[44] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–680, 1986.

[45] B. G. Cleland and C. Enroth-Cugell. Quantitative aspects of gain and latency in the cat retina. *Journal of Physiology*, 206:73–91, 1970.

[46] B. G. Cleland and W. R. Levick. Brisk and sluggish concentrically organised ganglion cells in the cat's retina. *Journal of Physiology*, 240:421–456, 1974.

[47] B. G. Cleland and W. R. Levick. Properties of rarely encountered types of ganglion cells in the cat's retina and an overall classification. *Journal of Physiology*, 240:457–492, 1974.

[48] T. N. Cornsweet. *Visual Perception.* Academic Press Inc., Harcourt Brace Jovanovich, 1970.

[49] F. Crick and C. Koch. Some reflections on visual awareness. *Cold Spring Harbor Symposia on Quantitative Biology*, 55:953–962, 1990.

[50] J. M. Crook, B. Lange-Malecki, B. B. Lee, and A. Valberg. Visual resolution of macaque retinal ganglion cells. *Journal of Physiology*, 396:205–224, 1988.

[51] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2:1160–1169, 1985.

[52] F. M. De Monasterio and P. Gouras. Functional properties of ganglion cells of the rhesus monkey retina. *Journal of Physiology*, 251:167–195, 1975.

[53] K. K. De Valois, R. L. De Valois, and E. W. Yund. Responses of striate cortex cells to grating and checkerboard patterns. *Journal of Physiology*, 291:483–505, 1979.

[54] K. K. De Valois and E. Switkes. Simultaneous masking interactions between chromatic and luminance gratings. *Journal of the Optical Society of America*, 73:11–18, 1983.

[55] K. K. De Valois and R. B. H. Tootell. Spatial-frequency-specific inhibition in cat striate cortex cells. *Journal of Physiology*, 336:359–376, 1983.

[56] R. L. De Valois, I. Abramov, and G. H. Jacobs. Analysis of response patterns of LGN cells. *Journal of the Optical Society of America*, 56:966–977, 1966.

[57] R. L. De Valois, D. G. Albrecht, and L. G. Thorell. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22:545–559, 1982.

[58] R. L. De Valois and K. K. De Valois. A multi-stage color model. *Vision Research*, 33:1053–1065, 1993.

[59] R. L. De Valois, L. G. Thorell, and D. G. Albrecht. Periodicity of striate-cortex-cell receptive fields. *Journal of the Optical Society of America A*, 2:1115–1123, 1985.

[60] A. F. Dean. The relationship between response amplitude and contrast for cat striate cortical neurones. *Journal of Physiology*, 318:413–427, 1981.

[61] G. C. DeAngelis, R. D. Freeman, and I. Ohzawa. Length and width tuning of neurons in the cat's primary visual cortex. *Journal of Neurophysiology*, 71:347–374, 1994.

[62] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman. Depth is encoded in the visual cortex by a speciallized receptive field structure. *Nature*, 352:156–159, 1991.

[63] A. M. Derrington, J. Krauskopf, and P. Lennie. Chromatic mechanisms in lateral geniculate nucleus of macaque. *Journal of Physiology*, 357:241–265, 1984.

[64] A. M. Derrington and P. Lennie. Spatial and temporal contrast sensitivity of neurones in lateral geniculate nucleus of macaque. *Journal of Physiology*, 357:219–240, 1984.

[65] R. Desimone and S. J. Schein. Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form. *Journal of Neurophysiology*, 57:835–868, 1987.

[66] R. Desimone, S. J. Schein, J. Moran, and L. G. Ungerleider. Contour, color and shape analysis beyond the striate cortex. *Vision Research*, 25:441–452, 1985.

[67] R. Desimone, M. Wessinger, L. Thomas, and W. Schneider. Attentional control of visual perception: Cortical and subcortical mechanisms. *Cold Spring Harbor Symposia on Quantitative Biology*, 55:963–971, 1990.

[68] M. C. Diamond, A. B. Scheibel, and L. M. Elson. *The Human Brain Coloring Book.* Barnes and Noble Books, Harper and Row, 1985.

[69] A. Dobbins, S. W. Zucker, and M. S. Cynader. Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, 329:438–441, 1987.

[70] A. Dobbins, S. W. Zucker, and M. S. Cynader. Endstopping and curvature. *Vision Research*, 29:1371–1387, 1989.

[71] A. Eisner and D. I. A. MacLeod. Blue-sensitive cones do not contribute to luminance. *Journal of the Optical Society of America*, 70:121–123, 1980.

[72] C. Enroth-Cugell and J. G. Robson. The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, 187:517–552, 1966.

[73] C. Enroth-Cugell and R. M. Shapley. Adaptation and dynamics of cat retinal ganglion cells. *Journal of Physiology*, 233:271–309, 1973.

[74] C. Enroth-Cugell and R. M. Shapley. Flux, not retinal illumination, is what cat retinal ganglion cells really care about. *Journal of Physiology*, 233:311–326, 1973.

[75] S. D. Erulkar and M. Fillenz. Single-unit activity in the lateral geniculate body of the cat. *Journal of Physiology*, 154:206–218, 1960.

[76] D. J. Felleman and D. C. Van Essen. Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. *Journal of Neurophysiology*, 57:889–920, 1987.

[77] D. Ferster and S. Lindström. Synaptic excitation of neurones in area 17 of the cat by intracortical axon collaterals of cortico-geniculate cells. *Journal of Physiology*, 367:233–252, 1985.

[78] D. J. Field, A. Hayes, and R. F. Hess. Contour integration by the human visual system: Evidence for a local "association field". *Vision Research*, 33:173–193, 1993.

[79] L. H. Finkel and G. M. Edelman. Integration of distributed cortical systems by reentry: A computer simulation of interactive functionally segregated visual areas. *Journal of Neuroscience*, 9:3188–3208, 1989.

[80] M. A. Finkelstein. Spectral tuning of opponent channels is spatially dependent. *Color Research and Application*, 13:107–112, 1988.

[81] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics, Principles and Practice.* Addison-Wesley, 2nd edition, 1990.

[82] K. H. Foster, J. P. Gaska, M. Nagler, and D. A. Pollen. Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *Journal of Physiology*, 365:331–363, 1985.

[83] W. J. Freeman. Tutorial on neurobiology: From single neurons to brain chaos. *International Journal of Bifurcation and Chaos*, 2:451–482, 1992.

[84] W. F. Ganong. *Review of Medical Physiology.* Appleton and Lange, 15th edition, 1991.

[85] C. D. Gilbert. Laminar differences in receptive field properties of cells in cat primary visual cortex. *Journal of Physiology*, 268:391–421, 1977.

[86] P. Gouras. Opponent-colour cells in different layers of foveal striate cortex. *Journal of Physiology*, 238:583–602, 1974.

[87] E. M. Granger and J. C. Heurtley. Visual chromaticity-modulation transfer function. *Journal of the Optical Society of America*, 63:1173–1174, 1973.

[88] K. L. Grieve and A. M. Sillito. Length summation properties of layer VI cells in the visual cortex and hypercomplex cell end zone inhibition. *Experimental Brain Research*, 84:319–325, 1990.

[89] K. L. Grieve and A. M. Sillito. A re-appraisal of the role of layer VI of the visual cortex in the generation of cortical end inhibition. *Experimental Brain Research*, 87:521–529, 1991.

[90] C. G. Gross. Representation of visual stimuli in inferior temporal cortex. *Philosophical Transactions of the Royal Society of London, Series B*, 335:3–10, 1992.

[91] S. Grossberg and E. Mingolla. Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception and Psychophysics*, 38:141–171, 1985.

[92] E. Guenther and E. Zrenner. The spectral sensitivity of dark- and light-adapted cat retinal ganglion cells. *Journal of Neuroscience*, 13:1543–1550, 1993.

[93] S. L. Guth, R. W. Massof, and T. Benzschawel. Vector model for normal and dichromatic vision. *Journal of the Optical Society of America*, 70:197–211, 1980.

[94] R. Hall. *Illumination and Color in Computer Generated Imagery.* Springer-Verlag, 1989.

[95] P. Hammond and D. M. Mackay. Influence of luminance gradient reversal on simple cells in feline striate cortex. *Journal of Physiology*, 337:69–87, 1983.

[96] P. Hammond and D. M. Mackay. Influence of luminance gradient reversal on complex cells in feline striate cortex. *Journal of Physiology*, 359:315–329, 1985.

[97] H. K. Hartline. The receptive fields of optic nerve fibres. *American Journal of Physiology*, 130:690–699, 1940.

[98] M. J. Hawken and A. J. Parker. Spatial properties of neurons in the monkey striate cortex. *Proceedings of the Royal Society of London, Series B*, 231:251–288, 1987.

[99] M. J. Hawken, A. J. Parker, and J. S. Lund. Laminar organisation and contrast sensitivity of direction-selective cells in the striate cortex of the old world monkey. *Journal of Neuroscience*, 8:3541–3548, 1988.

[100] D. J. Heeger. Normalisation of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–197, 1992.

[101] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kübler. Simulation of neural contour mechanisms: From simple to end-stopped cells. *Vision Research*, 32:963–981, 1992.

[102] F. Heitger and R. von der Heydt. A computational model of neural contour processing: Figure-ground segregation and illusory contours. In *Proceedings of the 4th International Conference on Computer Vision (ICCV4)*, pages 32–40. IEEE Computer Society Press, 1993.

[103] E. Hering. Beitrag zur Lehre vom Simultankontrast. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 1:18–28, 1890.

[104] C. A. Heywood, A. Gadotti, and A. Cowey. Cortical area V4 and its role in the perception of color. *Journal of Neuroscience*, 12:4056–4065, 1992.

[105] T. P. Hicks, B. B. Lee, and T. R. Vidyasagar. The responses of cells in macaque lateral geniculate nucleus to sinusoidal gratings. *Journal of Physiology*, 337:183–200, 1983.

[106] J. Hirsch and R. Hylton. Limits of spatial frequency discrimination as evidence of neural interpolation. *Journal of the Optical Society of America*, 72:1367–1274, 1982.

[107] S. Hochstein and R. M. Shapley. Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. *Journal of Physiology*, 262:265–284, 1976.

[108] M. I. Hofmann and P. E. Hallett. Preattentive discrimination of relative phase modelled by interacting Gabor or by difference-of-Gaussian filters. *Vision Research*, 33:2569–2587, 1993.

[109] D. C. Hood and M. A. Finkelstein. A case for the revision of textbook models of color vision: The detection and appearance of small brief lights. In J. D. Mollon and L. T. Sharpe, editors, *Colour Vision*, pages 385–398. Academic Press, 1983.

[110] D. H. Hubel. Single unit activity in lateral geniculate body and optic tract of unrestrained cats. *Journal of Physiology*, 150:91–104, 1960.

[111] D. H. Hubel and M. S. Livingstone. Complex-unoriented cells in a subregion of primate area 18. *Nature*, 315:325–327, 1985.

[112] D. H. Hubel and T. N. Wiesel. Receptive fields of optic nerve fibres in the spider monkey. *Journal of Physiology*, 154:572–580, 1960.

[113] D. H. Hubel and T. N. Wiesel. Integrative action in the cat's lateral geniculate body. *Journal of Physiology*, 155:385–398, 1961.

[114] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154, 1962.

[115] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195:215–243, 1968.

[116] D. H. Hubel and T. N. Wiesel. Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London, Series B*, 198:1–59, 1977.

[117] R. A. Humanski and H. R. Wilson. Spatial frequency mechanisms with short-wavelength-sensitive cone inputs. *Vision Research*, 32:549–560, 1992.

[118] M. Ikeda and M. Ayama. Nonlinear nature of the yellow chromatic valence. In J. D. Mollon and L. T. Sharpe, editors, *Colour Vision*, pages 345–351. Academic Press, 1983.

[119] C. R. Ingling, S. S. Grigsby, and R. C. Long. Comparison of spectral sensitivity using heterochromatic flicker photometry and an acuity criterion. *Color Research and Applications*, 17:187–196, 1992.

[120] C. R. Ingling Jr. and E. Martinez. The spectral sensitivity of the $r - g$ achromatic channel. *Investigative Opthalmology and Visual Science*, Supplement 22:17, 1982.

[121] C. R. Ingling Jr. and E. Martinez. The spatiochromatic signal of the $r - g$ channel. In J. D. Mollon and L. T. Sharpe, editors, *Colour Vision*, pages 433–444. Academic Press, 1983.

[122] C. R. Ingling Jr. and E. Martinez-Uriegas. The relationship between spectral sensitivity and spatial sensitivity for the primate $r - g$ X-channel. *Vision Research*, 23:1495–1500, 1983.

[123] C. R. Ingling Jr. and E. Martinez-Uriegas. The spatiotemporal properties of the $r - g$ X-cell channel. *Vision Research*, 25:33–38, 1985.

[124] C. R. Ingling Jr. and B. H. P. Tsou. Spectral sensitivity for flicker and acuity criteria. *Journal of the Optical Society of America A*, 5:1374–1378, 1988.

[125] D. B. Judd. Response functions for types of vision according to the Müller theory. *Journal of Research of the National Bureau of Standards*, 42:1–15, 1949.

[126] B. Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, 290, 1981.

[127] B. Julesz. Early vision is bottom-up, except for focal attention. *Cold Spring Harbor Symposia on Quantitative Biology*, 55:973–978, 1990.

[128] B. Julesz. Early vision and focal attention. *Reviews of Modern Physics*, 63:735–771, 1991.

[129] B. Julesz and B. Kröse. Features and spatial filters. *Nature*, 333:302–303, 1988.

[130] E. Kaplan and R. M. Shapley. X and Y cells in the lateral geniculate nucleus of macaque monkeys. *Journal of Physiology*, 330:125–143, 1982.

[131] E. Kaplan and R. M. Shapley. The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proceedings of the National Academy of Sciences USA*, 83:2755–2757, 1986.

[132] T. P. Kaushal. Towards visually convincing image segmentation. *Image and Vision Computing*, 10:617–624, 1992.

[133] D. H. Kelly. Motion and vision. II. Stabilised spatio-temporal threshold surface. *Journal of the Optical Society of America*, 69:1340–1349, 1979.

[134] D. H. Kelly. Spatiotemporal variation of chromatic and achromatic contrast thresholds. *Journal of the Optical Society of America*, 73:742–750, 1983.

[135] D. H. Kelly. Opponent-color receptive-field profiles determined from large-area psychophysical measurements. *Journal of the Optical Society of America A*, 6:1784–1793, 1989.

[136] F. Kingdom and B. Moulden. A multi-channel approach to brightness coding. *Vision Research*, 32:1565–1582, 1992.

[137] J. J. Knierim and D. C. Van Essen. Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, 67:961–980, 1992.

[138] B. J. A. Kröse and B. Julesz. The control and speed of shifts of attention. *Vision Research*, 29:1607–1619, 89.

[139] J. Krüger and P. Gouras. Spectral selectivity of cells and its dependence on slit length in monkey visual cortex. *Journal of Neurophysiology*, 43:1055–1069, 1980.

[140] S. W. Kuffler. Discharge patterns and functional organisation of mammalian retina. *Journal of Neurophysiology*, 16:37–68, 1953.

[141] D. LaBerge, M. Carter, and V. Brown. A network simulation of thalmic circuit operations in selective attention. *Neural Computation*, 4:318–331, 1992.

[142] B. B. Lee and V. Virsu. The interaction of opponent cone mechanisms in cells of the macaque lateral geniculate nucleus. In J. D. Mollon and L. T. Sharpe, editors, *Colour Vision*, pages 226–233. Academic Press, 1983.

[143] P. Lennie. Parallel visual pathways: A review. *Vision Research*, 20:561–594, 1980.

[144] P. Lennie. Recent developments in the physiology of color vision. *Trends in Neuroscience*, 7:243–248, 1984.

[145] P. Lennie, J. Krauskopf, and G. Sclar. Chromatic mechanisms in striate cortex of macaque. *Journal of Neuroscience*, 10:649–669, 1990.

[146] P. Lennie, J. Pokorny, and V. C. Smith. Luminance. *Journal of the Optical Society of America A*, 10:1283–1293, 1993.

[147] C. Y. Li, X. Pei, Y. Zhow, and H. C. von Mitzlaff. Role of the extensive area outside the X-cell receptive field in brightness information transmission. *Vision Research*, 31:1529–1540, 1991.

[148] R. A. Linsenmeier, L. J. Frishman, H. G. Jakiela, and C. Enroth-Cugell. Receptive field properties of X and Y cells in the cat retina derived from contrast sensitivity measurements. *Vision Research*, 22:1173–1183, 1982.

[149] M. Livingstone and D. Hubel. Segregation of form, color, movement and depth: Anatomy, physiology and perception. *Science*, 240:740–749, 1988.

[150] M. S. Livingstone and D. H. Hubel. Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience*, 4:309–356, 1984.

[151] J. S. Lund. Local circuit neurons of macaque monkey striate cortex. I. Neurons of laminae 4C and 5A. *Journal of Comparative Neurology*, 257:60–92, 1987.

[152] J. S. Lund, M. J. Hawken, and A. J. Parker. Local circuit neurons of macaque monkey striate cortex: II. Neurons of laminae 5B and 6. *Journal of Comparative Neurology*, 276:1–29, 1988.

[153] L. Maffei and A. Fiorentini. The visual cortex as a spatial frequency analyser. *Vision Research*, 13:1255–1267, 1973.

[154] D. Marr. *Vision.* W. H. Freeman and Company, 1982.

[155] E. Martinez-Uriegas. A solution to color-luminance ambiguity in the spatiotemporal signal of primate X cells. *Investigative Opthalmology and Visual Science*, Supplement 26:183, 1985.

[156] E. Martinez-Uriegas and D. H. Kelly. Chromatic and achromatic parvo channels. *Investigative Opthalmology and Visual Science*, Supplement 30:128, 1989.

[157] S. Marčelja. Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*, 70:1279–1300, 1980.

[158] R. H. Masland. The functional architecture of the retina. *Scientific America*, 255(6):90–99, 1986.

[159] MasPar Computer Corporation. *MasPar MP-1 Standard Programming Manuals.* MasPar Computer Corporation, Part Number 9300-9001-00, October 1990.

[160] R. W. Massof and J. F. Bird. A general zone theory of color and brightness vision. I. Basic formulation. *Journal of the Optical Society of America*, 68:1465–1471, 1978.

[161] H. R. Maturana and S. Frenk. Directional movement and horizontal edge detectors in pigeon retina. *Science*, 142:977–979, 1963.

[162] J. H. R. Maunsell and D. C. Van Essen. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, 49:1127–1147, 1983.

[163] J. H. R. Maunsell and D. C. Van Essen. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology*, 49:1148–1167, 1983.

[164] W. H. Merigan. Chromatic and achromatic vision of macaques: Role of the P pathway. *Journal of Neuroscience*, 9:776–783, 1989.

[165] C. R. Michael. Color-sensitive complex cells in monkey striate cortex. *Journal of Neurophysiology*, 41:1250–, 1978.

[166] C. R. Michael. Color vision mechanisms in monkey striate cortex: Dual-opponent cells with concentric receptive fields. *Journal of Neurophysiology*, 41:572–588, 1978.

[167] C. R. Michael. Color vision mechanisms in monkey striate cortex: Simple cells with dual opponent-color receptive fields. *Journal of Neurophysiology*, 41:1233–1249, 1978.

[168] C. R. Michael. Color-sensitive hypercomplex cells in monkey striate cortex. *Journal of Neurophysiology*, 42:726–, 1979.

[169] E. K. Miller, L. Li, and R. Desimone. Activity of neurons in anterior inferior temporal cortex during a short term memory task. *Journal of Neuroscience*, 13:1460–1478, 1993.

[170] M. J. Morgan, J. Ross, and A. Hayes. The relative importance of local phase and local amplitude in patchwise image reconstruction. *Biological Cybernetics*, 65:113–119, 1991.

[171] M. C. Morrone and D. C. Burr. Feature detection in human vision: A phase-dependent energy model. *Proceedings of the Royal Society of London, Series B*, 235:221–245, 1988.

[172] M. C. Morrone, D. C. Burr, and L. Maffei. Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proceedings of the Royal Society of London, Series B*, 216:335–354, 1982.

[173] A. Movshon. Visual processing of moving images. In H. Barlow, C. Blakemore, and M. Weston-Smith, editors, *Images and Understanding*, chapter 8, pages 122–137. Cambridge University Press, 1990.

[174] J. A. Movshon, I. D. Thompson, and D. J. Tolhurst. Receptive field organisation of complex cells in the cat's striate cortex. *Journal of Physiology*, 283:79–99, 1978.

[175] J. A. Movshon, I. D. Thompson, and D. J. Tolhurst. Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *Journal of Physiology*, 283:101–120, 1978.

[176] J. A. Movshon, I. D. Thompson, and D. J. Tolhurst. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology*, 283:53–77, 1978.

[177] K. T. Mullen. Sensitivity of the human red-green chromatic system to spatial frequency. *Journal of Physiology*, 332:14P, 1982.

[178] P. C. Murphy and A. M. Sillito. Corticofugal feedback influences the generation of length tuning in the visual pathway. *Nature*, 329:727–729, 1987.

[179] K. I. Naka and W. A. H. Rushton. S-potentials from luminosity units in the retina of fish. *Journal of Physiology*, 185:587–599, 1966.

[180] K. Nakayama. Biological image motion processing: A review. *Vision Research*, 25:625–660, 1985.

[181] E. Niebur, C. Koch, and C. Rosin. An oscillation-based model for the neuronal basis of attention. *Vision Research*, 33:2789–2802, 1993.

[182] C. Noorlander, M. J. G. Heuts, and J. J. Koenderink. Sensitivity to spatiotemporal combined luminance and chromaticity contrast. *Journal of the Optical Society of America*, 71:453–459, 1981.

[183] H. C. Nothdurft. Saliency effects across dimensions in visual search. *Vision Research*, 33:839–844, 1993.

[184] I. Ohzawa, G. Sclar, and R. D. Freeman. Contrast gain control in the cat's visual system. *Journal of Neurophysiology*, 54:651–667, 1985.

[185] A. Parker and M. Hawken. Capabilities of monkey cortical cells in spatial-resolution tasks. *Journal of the Optical Society of America A*, 2:1101–1114, 1985.

[186] A. J. Parker. Visual Neuroscience—Misaligned viewpoints. *Nature*, 352:109, 1991.

[187] A. J. Parker and M. J. Hawken. Two-dimensional spatial structure of receptive fields in monkey striate cortex. *Journal of the Optical Society of America A*, 5:598–605, 1988.

[188] E. Peli. Contrast in complex images. *Journal of the Optical Society of America A*, 7:2032–2040, 1990.

[189] J. A. Perrone. Model for the computation of self-motion in biological systems. *Journal of the Optical Society of America A*, 9:177–194, 1992.

[190] E. Peterhans and R. von der Heydt. Mechanisms of contour perception in monkey visual cortex. II. contours bridging gaps. *Journal of Neuroscience*, 9:1749–1763, 1989.

[191] U. Polat and D. Sagi. The architecture of perceptual spatial interactions. *Vision Research*, 34:73–78, 1994.

[192] D. A. Pollen and S. F. Ronner. Periodic excitability changes across the receptive fields of complex cells in the striate and parastriate cortex of the cat. *Journal of Physiology*, 245:667–697, 1975.

[193] K. Purpura, E. Kaplan, and R. M. Shapley. Background light and the contrast gain of primate P and M retinal ganglion cells. *Proceedings of the National Academy of Sciences USA*, 85:4534–4537, 1988.

[194] R. F. Quick. A vector-magnitude model of contrast detection. *Kybernetik*, 16:65–67, 1974.

[195] F. Ratliff and H. K. Hartline. The responses of *limulus* optic nerve fibres to patterns of illumination on the retinal mosaic. *Journal of General Physiology*, 42:1241–1255, 1959.

[196] R. W. Rodieck. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5:583–601, 1965.

[197] E. T. Rolls. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philosophical Transactions of the Royal Society of London, Series B*, 335:11–21, 1992.

[198] D. Rose. Responses of single units in cat visual cortex to moving bars of light as a function of bar length. *Journal of Physiology*, 271:1–23, 1977.

[199] P. W. Russell. Chromatic input to stereopsis. *Vision Research*, 19:831–834, 1979.

[200] D. Sagi and B. Julesz. "Where" and "What" in vision. *Science*, 228:1217–1219, 1985.

[201] D. Sagi and B. Julesz. Enhanced detection in the aperture of focal attention during simple discrimination tasks. *Nature*, 321:693–695, 1986.

[202] H. Saito, M. Yukie, K. Tanaka, K. Hikosaka, Y. Fukada, and E. Iwai. Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *Journal of Neuroscience*, 6:145–157, 1986.

[203] H. Sakata, H. Shibutani, K. Kawano, and T. L. Harrington. Neural mechanisms of space vision in the parietal association cortex of the monkey. *Vision Research*, 25:453–463, 1985.

[204] O. H. Schade. Optical and photoelectric analog of the eye. *Journal of the Optical Society of America*, 46:583–594, 1956.

[205] P. H. Schiller and C. L. Colby. The responses of single cells in the lateral geniculate nucleus of the rhesus monkey to color and luminance contrast. *Vision Research*, 23:1631–1641, 1983.

[206] P. H. Schiller and K. Lee. The role of primate extrastriate area V4 in vision. *Science*, 251:1251–1253, 1991.

[207] J. L. Schnapf and D. A. Baylor. How photoreceptor cells respond to light. *Scientific America*, 256(4):32–39, 1987.

[208] G. Sclar. Expression of "retinal" contrast gain control by neurons of the cat's lateral geniculate nucleus. *Experimental Brain Research*, 66:589–596, 1987.

[209] G. Sclar, J. H. R. Maunsell, and P. Lennie. Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research*, 30:1–10, 1990.

[210] N. Sekiguchi, D. R. Williams, and D. H. Brainard. Efficiency in detection of isoluminant and isochromatic interference fringes. *Journal of the Optical Society of America A*, 10:2118–2133, 1993.

[211] R. Shapley. The importance of contrast for the activity of single neurons, the VEP and perception. *Vision Research*, 26:45–61, 1986.

[212] R. M. Shapley and J. D. Victor. The effect of contrast on the transfer properties of cat retinal ganglion cells. *Journal of Physiology*, 285:275–298, 1978.

[213] A. M. Sillito. The contribution of inhibitory mechanisms to the receptive field properties of neurones in the striate cortex of the cat. *Journal of Physiology*, 250:305–329, 1975.

[214] A. M. Sillito. The spatial extent of excitatory and inhibitory zones in the receptive field of superficial layer hypercomplex cells. *Journal of Physiology*, 273:791–803, 1977.

[215] A. M. Sillito and V. Versiani. The contribution of excitatory and inhibitory inputs to the length preference of hypercomplex cells in layers II and III of the cat's striate cortex. *Journal of Physiology*, 273:775–790, 1977.

[216] B. C. Skottun, A. Bradley, G. Sclar, I. Ohzawa, and R. D. Freeman. The effects of contrast on visual orientation and spatial frequency discrimination: A comparison of single cells and behaviour. *Journal of Neurophysiology*, 57:773–786, 1987.

[217] B. C. Skottun, D. H. Grosof, and R. L. De Valois. On the responses of simple and complex cells to random dot patterns. *Vision Research*, 31:43–46, 1991.

[218] V. C. Smith and J. Pokorny. Spectral sensitivity of color-blind observers and the cone photopigments. *Vision Research*, 12:2059–2071, 1972.

[219] V.C. Smith and J. Pokorny. Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision Research*, 15:161–171, 1975.

[220] R. J. Snowden and S. T. Hammett. Subtractive and divisive adaptation in the human visual system. *Nature*, 355:248–250, 1992.

[221] H. Spitzer and S. Hochstein. Complex-cell receptive field models. *Progress in Neurobiology*, 31:285–309, 1988.

[222] M. Steriade, D. A. McCormick, and T. J. Sejnowski. Thalmocortical oscillations in the sleeping and aroused brain. *Science*, 262:679–685, 1993.

[223] K. Tanaka. Neuronal mechanisms of object recognition. *Science*, 262:685–688, 1993.

[224] J. E. Thornton and E. N. Pugh Jr. Relationship of opponent-colours cancellation measures to cone-antagonistic signals deduced from increment threshold data. In J. D. Mollon and L. T. Sharpe, editors, *Colour Vision*, pages 361–373. Academic Press, 1983.

[225] D. J. Tolhurst and L. P. Barfield. Interactions between spatial frequency channels. *Vision Research*, 18:951–958, 1978.

[226] A. Treisman. Features and objects in visual processing. *Scientific America*, 255:106–115, 1986.

[227] D. Y. Ts'o and C. D. Gilbert. The organisation of chromatic and spatial interactions in the primate striate cortex. *Journal of Neuroscience*, 8:1712–1727, 1988.

[228] S. Ullman. The interpretation of structure from motion. *Proceedings of the Royal Society of London, Series B*, 203:405–426, 1979.

[229] S. Ullman. *The Interpretation of Visual Motion.* MIT Press, Cambridge, Mass., 1979.

[230] L. G. Ungerleider and M. Mishkin. Two cortical visual systems. In D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield, editors, *Analysis of Visual Behavior*, pages 549–586. MIT Press, Cambridge, 1982.

[231] A. Valberg, B. B. Lee, O. D. Creutzfeldt, and D. A. Tigwell. Luminance ratio and the spectral responsiveness of cells in the macaque lateral geniculate nucleus. In J. D. Mollon and L. T. Sharpe, editors, *Colour Vision*, pages 235–243. Academic Press, 1983.

[232] G. J. C. van der Horst and M. A. Bouman. Spatiotemporal chromaticity discrimination. *Journal of the Optical Society of America*, 59:1482–1488, 1969.

[233] S. Venkatesh and R. Owens. On the classification of image features. *Pattern Recognition Letters*, 11:339–349, 1990.

[234] M. Versavel, G. A. Orban, and L. Lagae. Responses of visual cortical neurons to curved stimuli and chevrons. *Vision Research*, 30:235–248, 1990.

[235] R. von der Heydt and E. Peterhans. Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *Journal of Neuroscience*, 9:1731–1748, 1989.

[236] R. von der Heydt, E. Peterhans, and G. Baumgartner. Illusory contours and cortical neuron responses. *Science*, 224:1260–1262, 1984.

[237] H. von Helmholtz. *Handbuch der Physiologischen Optik.* Voss, Hamburg, 1st edition, 1867.

[238] G. Wald. The receptors for human colour vision. *Science*, 145:1007–1017, 1964.

[239] H. Wässle, U. Grünert, J. Röhrenbeck, and B. B. Boycott. Retinal ganglion cell density and cortical magnification factor in the primate. *Vision Research*, 30:1897–1911, 1990.

[240] A. B. Watson and A. J. Ahumada. Model of human visual-motion sensing. *Journal of the Optical Society of America A*, 2:322–341, 1985.

[241] R. J. Watt. Further evidence concerning the analysis of curvature in human foveal vision. *Vision Research*, 24:251–253, 1984.

[242] R. J. Watt and D. P. Andrews. Contour curvature analysis: Hyperacuities in the discrimination of detailed shape. *Vision Research*, 22:449–460, 1982.

[243] R. J. Watt and M. J. Morgan. A theory of the primitive spatial code in human vision. *Vision Research*, 25:1661–1674, 1985.

[244] M. A. Webster and R. L. De Valois. Relationship between spatial-frequency and orientation tuning of striate-cortex cells. *Journal of the Optical Society of America A*, 2:1124–1132, 1985.

[245] T. N. Wiesel. Receptive fields of ganglion cells in the cat's retina. *Journal of Physiology*, 153:583–594, 1960.

[246] T. N. Wiesel and D. H. Hubel. Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *Journal of Neurophysiology*, 29:1115–1156, 1966.

[247] H. R. Wilson and J. R. Bergen. A four mechanism model for threshold spatial vision. *Vision Research*, 19:19–32, 1979.

[248] H. R. Wilson and R. Humanski. Spatial frequency adaptation and contrast gain control. *Vision Research*, 33:1133–1149, 1993.

[249] H. R. Wilson, D. K. McFarlane, and G. C. Phillips. Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, 23:873–882, 1983.

[250] G. Wyszecki and W. S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley and Sons, 2nd edition, 1982.

[251] Z. Xie and T. G. Stockham Jr. Towards the unification of three visual laws and two visual models in brightness perception. *IEEE Transactions on Systems, Man, and Cybernetics*, 19:379–387, 1989.

[252] S. Zeki. The representation of colours in the cerebral cortex. *Nature*, 284:412–418, 1980.

[253] S. Zeki. The distribution of wavelength and orientation selective cells in different areas of the monkey visual cortex. *Proceedings of the Royal Society of London, Series B*, 217:449–470, 1983.

[254] S. Zeki. The visual image in mind and brain. *Scientific America*, 267(3):43–50, 1992.

[255] S. Zeki and S. Shipp. The functional logic of cortical connections. *Nature*, 335:311–317, 1988.

[256] S. M. Zeki. Functional specialisation in the visual cortex of the rhesus monkey. *Nature*, 274:423–428, 1978.

[257] E. Zrenner. Neurophysiological aspects of colour vision mechanisms in the primate retina. In J. D. Mollon and L. T. Sharpe, editors, *Colour Vision*, pages 195–210. Academic Press, 1983.

[258] E. Zrenner and P. Gouras. Cone opponency in tonic ganglion cells and its variation with eccentricity in rhesus monkey retina. In J. D. Mollon and L. T. Sharpe, editors, *Colour Vision*, pages 211–223. Academic Press, 1983.

[259] S. W. Zucker, A. Dobbins, and L. Iverson. On the computational neurobiology of curve detection. In *British Machine Vision Conference Proceedings*, pages 17–23, 1991.