# A BIASED COMPETITION COMPUTATIONAL MODEL OF SPATIAL AND OBJECT-BASED ATTENTION MEDIATING ACTIVE VISUAL SEARCH

by

LINDA JANE LANYON

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

**DOCTOR OF PHILOSOPHY**

School of Computing, Communications & Electronics,
Faculty of Technology
&
Centre for Theoretical & Computational Neuroscience,
Faculty of Science

**April 2005**

# Statement of Copyright

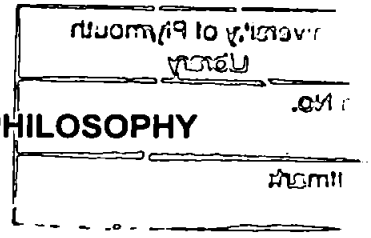
# A BIASED COMPETITION COMPUTATIONAL MODEL OF SPATIAL AND OBJECT-BASED ATTENTION MEDIATING ACTIVE VISUAL SEARCH

by

**LINDA JANE LANYON**

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

**DOCTOR OF PHILOSOPHY**



School of Computing, Communications & Electronics,
Faculty of Technology
&
Centre for Theoretical & Computational Neuroscience,
Faculty of Science

**April 2005**

# Abstract

**Linda Jane Lanyon**

## A BIASED COMPETITION COMPUTATIONAL MODEL OF SPATIAL AND OBJECT-BASED ATTENTION MEDIATING ACTIVE VISUAL SEARCH

A computational cognitive neuroscience approach was used to examine processes of visual attention in the human and monkey brain. The aim of the work was to produce a biologically plausible neurodynamical model of both spatial and object-based attention that accounted for observations in monkey visual areas V4, inferior temporal cortex (IT) and the lateral intraparietal area (LIP), and was able to produce search scan path behaviour similar to that observed in humans and monkeys.

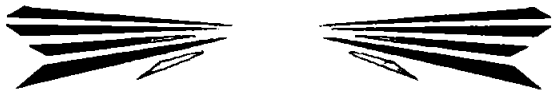
Of particular interest currently in the visual attention literature is the *biased competition hypothesis* (Desimone & Duncan, 1995). The model presented here is the first *active vision* implementation of biased competition, where attentional shifts are overt. Therefore, retinal inputs change during the scan path and this approach raised issues, such as memory for searched locations across saccades, not addressed by previous models with static retinas.

This is the first model to examine the different time courses associated with spatial and object-based effects at the cellular level. Single cell recordings in areas V4 (Luck et al., 1997; Chelazzi et al., 2001) and IT (Chelazzi et al., 1993, 1998) were replicated such that attentional effects occurred at the appropriate time after onset of the stimulus. Object-based effects at the cellular level of the model led to systems level behaviour that replicated that observed during active visual search for orientation and colour feature conjunction targets in psychophysical investigations. This provides a valuable insight into the link between cellular and system level behaviour in natural systems. At the systems level, the simulated search process showed selectivity in its scan path that was similar to that observed in humans (Scialfa & Joffe, 1998; Williams & Reingold, 2001) and monkeys (Motter & Belky, 1998b), being guided to target coloured locations in preference to locations containing the target orientation or blank areas. A connection between the ventral and dorsal visual processing streams (Ungerleider & Mishkin, 1982) is suggested to contribute to this selectivity and priority in the featural guidance of search. Such selectivity and avoidance of blank areas has potential application in computer vision applications.

Simulation of lesions within the model and comparison with patient data provided further verification of the model. Simulation of visual neglect due to parietal cortical lesion suggests that the model has the capability to provide insights into the neural correlates of the conscious perception of stimuli.

The biased competition approach described here provides an extendable framework within which further "bottom-up" stimulus and "top-down" mnemonic and cognitive biases can be added, in order to further examine exogenous versus endogenous factors in the capture of attention.

# List of Contents



# List of Figures & Tables

**Figures** **Page**

## Chapter 4



## Chapter 5

## Chapter 6

## Chapter 7

## Chapter 8



## Appendix A1



## Appendix A3

## Tables Page

### Chapter 4



### Chapter 6



### Appendix A3

# Acknowledgements

This thesis is dedicated to my mother, Val Lanyon, who sadly and unexpectedly passed away during its write-up. Mum, without your encouragement and support I would never have got to this point. Thank you so much for everything. I know you would have been very proud.

Also, to Amy and George Barcott ("Auntie and Georgie"), my great aunt and uncle in name but grandparents in reality, who also passed away during this work. I have so much to thank you for also, especially for our very special friendship throughout my life and for leaving me with the financial means to complete this PhD. Along with mum, you are always in my thoughts.

And, to my father, Barrie Lanyon, and Zippy, with much love and thanks.

Importantly, very many thanks go to my director of studies, Dr. Susan Denham, for her enthusiasm, advice and encouragement, which have supported me throughout the process. Also thanks to the other members of the Centre for Theoretical & Computational Neuroscience, which is my primary affiliation and where this work was based.

For organising generous studentship funding, I thank University of Plymouth's School of Computing, Communications and Electronics, in particular Prof. Patricia Pearce, and Centre for Theoretical & Computational Neuroscience, in particular Dr. Susan Denham and Prof. Michael Denham.

Thanks to the examiners of this thesis, Prof. Glyn Humphreys and Dr. Phil Culverhouse, for their support.

For administrative support, thanks to Carole Watson, Christine Brown and Veronica Montague.

Also, thanks to HM Land Registry, especially Ian Goodall & David Hemmings for allowing me a career break at an important stage of the national Electronic Conveyancing project, to complete this work.

Special thanks go to James Thompson for his caring support, encouragement and belief, which meant so much during the final write-up and beyond.

Very many special thanks also to Nada, Jane, Pauline, Huck, Loretta, Benoit, Karen, Tom, Sana, Michael, Stalin and Eva for their invaluable support and friendship; and to my other friends connected with the University who have been a wonderful source of fun and support, and will remain lifelong friends.

# Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Committee.

This study was financed with the aid of studentships from the School of Computing and Centre for Theoretical and Computational Neuroscience, University of Plymouth and partly self-funded by the author.

Relevant scientific seminars and conferences were regularly attended at which work was often presented; the author was invited to present a seminar talk about this work at another university and several papers were prepared for publication. Details are listed below:

## Publications:

- Lanyon, L. J., & Denham, S.L. (2004) A model of active visual search with object-based attention guiding scan paths. **Neural Networks (Special Issue: Vision & Brain)**, 17(5-6), 873-897
- Lanyon, L. J., & Denham, S.L. (2004) A biased competition computational model of spatial and object-based attention mediating active visual search. **Neurocomputing (Special Issue: Computational Neuroscience: Trends in Research 2004)**, 58-60C, 655-662
- Lanyon, L. J., & Denham, S.L. (2005) A model of object-based attention that guides active visual search to behaviourally relevant locations. **Lecture Notes in Computer Science (Special Issue: WAPCV 2004; L Paletta et al., Eds)**, Vol. 3368, 42-56
- Lanyon, L.J. & Denham, S.L. (2005) A model of spatial & object-based attention for active visual search. **Modelling Language, Cognition and Action: Proceedings of the 9th Neural Computation & Psychology Workshop. World Scientific, Progress in Neural Processing**, 16, 239-248
- Lanyon, L. J. & Denham, S.L. (2003) A biased competition computational model of spatial and object-based attention mediating active visual search. [Abstract] **Journal of Vision** (3rd Annual Meeting of the Vision Sciences Society), 3(9), 570a http://journalofvision.org/3/9/570/

- Lanyon, L. J. & Denham, S.L. A biased competition model of spatial and object-based cellular attentional effects that mediate visual search. *Submitted*

## Presentations, Conferences Attended & Papers Presented:

- **INVITED SEMINAR SPEAKER:** 'A Model of Spatial & Object-Based Attention for Active Visual Search'. Centre for Cognitive Neuroscience & Cognitive Systems, University of Kent, Canterbury, U.K., December 2004
- **CONFERENCE TALK:** A Model of Spatial & Object-Based Attention for Active Visual Search'. [Abstract] **Proceedings of the 9th Neural Computation & Psychology Workshop,** Plymouth, U.K., September 2004
- **CONFERENCE TALK:** 'A model of object-based attention that guides active visual search to behaviourally relevant locations'. **Proceedings of the 2nd International Workshop on Attention & Performance in Computer Vision,** Prague, Czech Republic, May 2004
- **CONFERENCE TALK:** 'A model of spatial and object-based attention for active visual search'. [Abstract] **Proceedings of the 11th Annual Workshop on Object Perception Attention & Memory,** Satellite Meeting to Psychonomics, Vancouver, Canada, November 2003
- **INVITED WORKSHOP TALK + CONFERENCE POSTER:** 'A biased competition computational model of spatial and object-based attention mediating active visual search'. [Abstract] **Proceedings of the Annual Computational Neuroscience Meeting 2003**, Alicante, Spain, July 2003
- **CONFERENCE POSTER:** 'A biased competition model of spatial and object-based attention mediating visual search'. [Abstract] **Proceedings of the 3rd Annual Meeting of the Vision Sciences Society**, Sarasota, Florida, U.S.A., May 2003
- **CONFERENCE ATTENDANCE: The Roots of Visual Awareness**, Oxford, U.K., September 2002
- **CONFERENCE ATTENDANCE: European Conference on Visual Perception**, Glasgow, U.K., August 2002

- **CONFERENCE POSTER:** 'Biased competition and object-based attention: Towards a unified model of visual attention'. [Abstract] **Proceedings of the 5th Annual International Conference on Cognitive & Neural Systems**, Boston, U.S.A., May 2001
- **SYMPOSIUM POSTER:** 'Towards a biased competition model of selective attention in area V4 during active visual search'. **Plymouth Institute of Neuroscience Symposium**, Plymouth, 2002
- **SEMINAR TALK:** 'A Model of Spatial & Object-Based Attention for Active Visual Search'. **Internal Seminar**, Centre for Theoretical & Computational Neuroscience, University of Plymouth, December 2003
- **SEMINAR TALK:** 'Biased competition and object-based attention: Towards a unified model of visual attention'. **Internal Seminar**, Centre for Neural & Adaptive Systems, School of Computing, University of Plymouth, April 2001
- **SEMINAR TUTORIALS:** 'Visual Attention I and Visual Attention II'. **Internal Seminars**, Plymouth Institute of Neuroscience, University of Plymouth, March 2003
- **LECTURE**: 'Visual Attention & Cognition'. Lecture as part of a new final year computing **undergraduate module in 'Brain Inspired Cognitive Systems'**, University of Plymouth, October 2004

Word count of main body of thesis:

**67,863**

Word count of textual part of appendices:

**4951**

Total word count:

**72814**

Signed .....

Date 22/9/05

# Chapter 1

# Introduction

This thesis uses a computational cognitive neuroscience approach to examine processes of visual attention in the human and monkey brain.

## 1.1 Theoretical and Computational Neuroscience in Context

Theoretical neuroscience, using computational techniques, is an increasingly important field that is needed to explain experimental results from a number of neuroscience-related disciplines, propose systems of processing in the brain and guide future experimental investigation. The processing of visual information is one of the most extensively studied areas of brain function and may provide an insight into other processes. The relative wealth of experimental data about the visual system leads it to be a suitable area to model computationally because biologically plausible computational models should be based on fairly extensive experimental data. The best models will provide an insight into processes within the brain and make predictions that can be tested experimentally. Therefore, such computational models should be able to be used to guide future experimentation in fields such as neuropsychology and neurophysiology.

The study of visual attention has enormous potential importance to technology in the fields of human-computer interaction (HCI, in order to inspire graphical-user interface design), information retrieval (in order to improve content-based information retrieval capabilities in images and video), robotics (robot guidance and object recognition) and other computer vision applications such as surveillance. Hence, models of visual attention, particularly robust computational models, are of interest in many fields beyond those traditionally connected with neuroscience.

## 1.2 Methodological Approach

Computational modelling of brain function involves forming a theory of the processes involved and testing this theory by implementing a biologically constrained computational model. Vision is a widely explored process in the brain and, therefore, an abundance of literature from various vision-related fields is available for the theoretical modelling process. Previously reported experimental evidence from various relevant fields, including neurophysiology (single cell recordings, neuro imaging, evoked potential recordings), psychology (principally psychophysics) was used to gain an insight into processes in the brain relating to visual attention. Such experiments provide information at various levels of abstraction:

- Examining behaviour in psychophysics treats the brain as a "black-box".

- Imaging, using techniques such as functional magnetic resonance imaging (fMRI), and evoked potentials, which are recorded on the surface of the scalp, investigate the function of particular brain *regions*.
- The behaviour of individual neurons can be examined by recording the activity of single cells in monkeys.

Methods such as single cell recordings and evoked potentials (normally event-related evoked potentials, or ERPs) provide information about neuronal activity with accurate temporal precision to the millisecond. However, the time course of the blood oxygen level dependant (BOLD) response used in fMRI is of the order of seconds rather than milliseconds and, thus, provides only a very coarse prediction of the time course of attentional effects at the cellular level. A combination of results from these various fields was used to develop a theory of visual attention processing in certain areas of the brain. Two research goals were formed and will be described in the next chapter. These goals were tested by the implementation of a computational model based on the proposed theory. This model forms the major contribution of this work and is described in the following three chapters, with results being presented in the subsequent three chapters.

The visual system comprises many brain regions and complex interactions, most of which are not yet fully understood. The level of modelling here, which includes systems level behaviour, requires some simplification and abstraction. Therefore, not all areas of the brain that could be involved in visual attention are included but reference to many of these is made throughout the text. There is an underlying

assumption about the legitimacy of animal to human mapping, i.e. that the animal model provides a reliable indication of human brain processing. The evidence from single cell recordings, on which much of this work is based, relates to experiments carried out on monkeys. However, many of the studies relating to visual search behaviour and brain imaging are reported from human as well as monkey studies. The human homologues for the monkey brain regions described here are commonly cited and fairly well understood. Therefore, it is not unreasonable to move between the two species for these particular brain regions and visual behaviours. The homologues will be described where relevant in the text.

## 1.3 Dorsal & Ventral Visual Pathways

A widely accepted neurobiological model suggests that cortical visual processing is performed in two main streams (Milner & Goodale, 1995; Ungerleider & Mishkin, 1982). Object recognition and the encoding of features belonging to objects is carried out in the "*what*" stream, which runs ventrally from the occipital lobe down to the inferior temporal lobe, including primary visual cortex (V1), visual areas V2, V4 and areas, such as TEO and TE, of inferior temporal cortex. This pathway is sometimes referred to as the ventral pathway. Spatial properties, on the other hand, are processed in the "*where*" (or "*how*") pathway, which runs dorsally from the occipital lobe up to the parietal lobe and includes visual areas V1, V2, V3, the middle temporal area (MT), the medial superior temporal area (MST), and areas

within posterior parietal cortex (within the inferior lobule of the parietal lobe) and superior temporal cortex. This pathway is sometimes referred to as the dorsal pathway. Parietal cortex represents space in many different coordinate systems in order to guide behaviour (see Colby & Goldberg, 1999, for review). For example, representations in the ventral intraparietal area (VIP) range from being retinocentric to being head-based. Neurons here are visually responsive or responsive to tactile stimuli or may be bi-modal. The medial intraparietal area (MIP) responds to stimuli within reach and is probably important in guiding arm movements. Response properties here range from being somatosensory to bi-modal to purely visual. The anterior intraparietal area (AIP) responds to stimuli that can be manipulated and the response relates to the desired shape of the hand. The lateral intraparietal area (LIP) represents the space explored by eye movements and, thus, has retinotopic receptive fields (a receptive field describes the spatio-temporal response properties of a neuron). This area is strongly linked with visual attention and will be one of the focuses in the work here. The work will also model feature and object representation in the "what" stream in order to produce object-based attentional effects.

## 1.4 Visual Attention

Attention operates as a "filter" in the brain so that the enormous amount of sensory information reaching cortex at any instant is subject to some restriction before

entering a level of processing where it can be reported or directly used to influence behaviour. A major area of research in psychology (specifically, psychophysics) is the nature of stimulus information and cognitive requirements that influence the capture of attention. Attentional capture appears to involve interplay between "top-down", voluntary (or *endogenous*) cognitive factors and "bottom-up" (*exogenous*) stimulus-related information, such that some types of stimuli can override top-down attentional control (see, for example, Remington, Johnston & Yantis, 1992) whereas others cannot (see, for example, Folk & Remington, 1999). The circumstance in which bottom-up capture overrides top-down control varies with the task, stimuli, and with the duration of presentation of the stimuli (Kim & Cave, 1999). Certain stimulus occurrences such as a sudden flash of light or movement in the periphery appear to be particularly *salient* in capturing attention in a bottom-up manner (see, for example, Remington et al., 1992).

The computational model that is the subject of the work reported here goes some way to resolving competition between top-down and bottom-up influences through competition in both the ventral and dorsal streams. However, it is primarily a model of top-down attentional influence.

The work is mainly concerned with *overt* control of attention, i.e. when an eye movement, or *saccade*, is made to the attended location. Some simulations of *covert* attention, where no eye movement is made, are provided later also.

### 1.4.1 Spatial verses Object-Based Attention

Traditionally, visual attention was thought to operate in a spatial manner (e.g. Posner 1980; Posner & Peterson, 1990) such that, metaphorically, a simple spatial "spotlight" (e.g. Crick, 1984; Helmholtz, 1867; Treisman, 1982) acted to enhance responses at a particular location(s). Spatially specific modulation of responses has been found in many experimental paradigms, including functional imaging (Brefcznski & De Yoe, 1999; Gandhi, Heeger & Boynton 1999; Martinez, Anllo-Vento, Sereno, Frank, Buxton, Dubowitz, Wong, Hinrichs, Heinze & Hillyard, 1999; Somers, Dale, Seiffert & Tootell, 1999), evoked potential recording (Anllo-Vento & Hillyard, 1996; Hillyard & Anllo-Vento, 1998; Luck, Fan & Hillyard, 1993; Luck & Hillyard, 1995), monkey single cell recording (Connor, Gallant, Preddie & van Essen, 1996; Luck, Chelazzi, Hillyard & Desimone, 1997; Moran & Desimone, 1985) and psychophysics (e.g. Posner, 1980; Treisman, 1998). However, many of these experiments also reveal modulation that is more complex than a simple spatial spotlight (e.g. Somers et al., 1999) and there is now a convincing body of evidence for attention operating in a more complex object- or feature-based manner. In psychophysics, performance gains are seen when subjects have to detect features relating to the same object rather than features belonging to spatially inseparable but different objects (e.g. Behrmann, Zemel & Mozer, 1998; Blaser, Pylyshyn & Holcombe, 2000; Duncan 1984; Mitchell, Stoner, Fallah & Reynolds, 2003; Reynolds, Alborzian & Stoner, 2003; Valdes-Sosa, Cobo & Pinilla, 2000). In one of the most convincing studies of object-based attention

(Blaser et al., 2000) "objects" were created by continually independently changing the colour, orientation and spatial frequency of two spatially inseparable Gabor patches, such that they occasionally passed each other in feature space. Observers reliably reported on the Gabor patch that was cued at the start, indicating that they had tracked this object, and were better at reporting two features from one of the Gabor patches rather than a feature from each. FMRI (O'Craven, Downing & Kanwisher, 1999) has revealed that functionally different brain regions related to properties of the same object are activated at the same time when attention is directed to one of two spatially overlapping objects. Event-related potential recordings (ERPs) show that attention to one of two overlapping surfaces (containing dots in motion) modulates even early ERP recordings (Pinilla, Cobo, Torres & Valdes-Sosa, 2001; Valdes-Sosa, Bobes, Rodriguez & Pinilla, 1998).

Many of the object-based attention psychophysical, fMRI and ERP results are confounded by issues relating to whether the selection is object-based, feature-based or surface-based. For example, the rapid serial object transformation (RSOT) paradigm introduced by Valdes-Sosa et al. (1998, 2000) and used by Mitchell et al. (2003), Pinilla et al. (2001) and Reynolds et al. (2003) creates two transparent surfaces containing dots of different colours and directions of motion, which have been shown to produce surface-based effects rather than feature-based effects (see Mitchell et al., 2003). Mounts & Melara (1999) found performance benefits when the feature (colour or orientation) to be discriminated was the same

one in which the target "pop-out"[1] occurred. A benefit was not found when the discrimination dimension did not correspond with the dimension in which the target popped-out. This suggests that not all features bound to objects were the subject of object-based attention and that a possibly earlier feature-based attention was operating. Human imaging shows that attending to a feature increases activity in areas that encode that feature type (Saenz, Buracas & Boynton, 2002), and that this increase may even occur in advance of the stimulus (Chawla, Rees & Friston, 1999). For the purposes of the work reported here, feature-based and object-based attention are considered to be effects originating from the same object-based process. Hence, the term object-based will cover non-spatial feature-based effects. At the cellular level, the work focuses on replication of monkey single cell recording data in extrastriate area V4 and inferior temporal (IT) cortex. Due to the nature of encoding in each of these regions (which will be described in later chapters), effects in IT will tend to be object-based (involving the entire object) and in V4 will tend to be feature-based (involving individual feature dimensions, before features are bound to objects).

In monkey single-cell investigations feature-based effects result in increased activity in cells in extrastriate area V4 when the cell's receptive field contain a stimulus feature that is currently attended (Haenny, Maunsell & Schiller, 1988; Haenny & Schiller, 1988; Motter 1994a,b). When cells in IT cortex (Chelazzi, Miller, Duncan & Desimone, 1993) or extrastriate area V4 (Chelazzi, Miller,

[1] Pop-out is the pre-attentive automatic detection of unique simple features that seem to "stand out" in the scene without cognitive effort.

Duncan & Desimone, 2001) are presented with two objects in a constant spatial arrangement within their receptive fields, responses are modulated according to which object is the current search target. Such an experiment is difficult in V1 due to the size of receptive fields. However, object-based modulation, possibly mediating early perceptual grouping, has been found in V1 (Roelfsema, Lamme & Spekreijse, 1998): Cells whose receptive fields contain different segments of a target line tended to be more active than cells containing segments of a distractor line.

Thus, the evidence from a range of experimental paradigms points to attention not simply acting as a spatial "spotlight" but being capable of operating in both a spatial and in a more complex object-based manner that may facilitate responses to behaviourally relevant features. However, a wealth of evidence for spatially specific attention also exists. McAdams and Maunsell (2000) have shown that both spatial attention and feature-based attention can concurrently affect the response of a V4 neuron and Treue and Martinez Trujillo (1999) found summable spatial and object-based effects in the attentional modulation of medial temporal (MT) neuron responses. Although within-subject comparisons are necessary to confirm the extent of anatomical overlap, comparison of fMRI activity across studies shows that similar regions are activated for both spatial and object-based attention (Corbetta, Kincade, Ollinger, McAvoy & Shulman, 2000; Hopfinger, Buonocore & Mangun, 2000; Serences, Schwarzbach, Courtney, Golay & Yantis, 2004). One of

the main aims of the current work was to explore possibilities for combining both spatial and object-based attention within the same model.

### 1.4.2 Feature Integration Theory and Biased Competition as Explanations of Visual Search Data

Attentional capture is often examined under conditions of visual search, i.e. when the subject is searching for a particular item in a display containing that target and several non-target stimuli known as *distractors*. Visual search for a target object containing a simple feature that is absent from distractors is normally effortless and appears to occur in parallel. Often the target is able to "pop-out" from the background and search time is normally invariant to the surrounding distractor set size. In contrast to this simple feature search, when the target is defined by a conjunction of features or by the absence of some feature the search takes longer and appears to require a serial process. This is explained by *Feature Integration Theory* (Treisman 1982, 1988, 1991, 1992, 1998, Treisman & Gelade 1980, Treisman & Gormican 1988) in which spatial selection is a prerequisite for the correct binding of visual features into objects. In this theory, the scene is organised in parallel into a set of "feature maps". Pre-attentive "pop-out" within a feature map allows fast search times for simple feature search. Feature conjunction search requires the integration of features at a specific spatial location through an attention mechanism. This is depicted in figure 1.1. This spatial attention mechanism is serial and thus requires more time. Serial attention may also be required for feature

search when targets and non-targets are closely similar (Treisman & Gormican 1988).

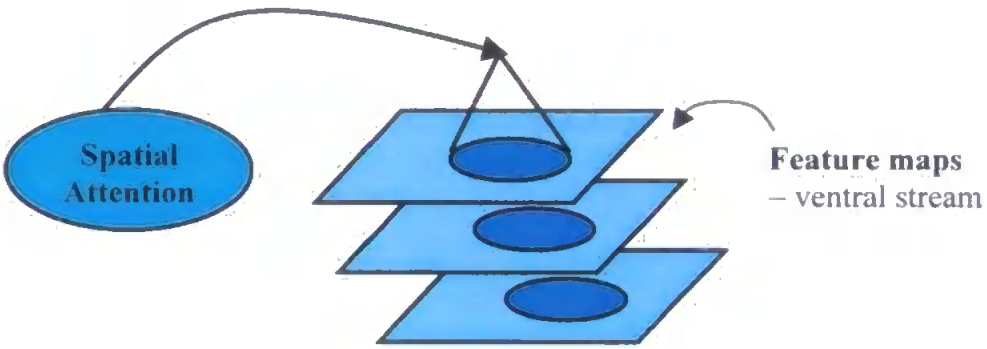



**Figure 1.1 The Spatial Spotlight Model of Attention**

The cartoon depicts the spatial spotlight model of attention, as proposed by Feature Integration Theory. Spatial attention selectively enhances activity in a specific spatial location across all feature maps and this serves to bind this information.

Feature Integration Theory proposes that feature processing is contingent upon the early spatial selection of stimulus attributes and that location plays a unique role in feature integration. However, an alternative view put forward by Duncan and Humphreys (1989) suggests that there is no early selection by location. Instead, all stimulus properties, including location, are processed in parallel across the visual field and compete for access to visual short-term memory according to how well

they match an *attentional template* that represents the behavioural requirement, i.e. the target representation (Duncan & Humphreys 1989, 1992, Duncan, Humphreys & Ward, 1997). This results in the suppression of non-target features. In this theory, increased search times are accounted for by the similarity of the target to distractors and decreased similarity amongst distractors.

When later elaborated in neurophysiological terms, this theory became known as the *biased competition hypothesis* (Desimone & Duncan, 1995) and suggested that multiple stimuli in the visual field activate in parallel populations of neurons that engage in competitive interactions. This competition is subject to a number of biases, such as "bottom-up" stimulus information and "top-down" cognitive requirements, for example a working memory template of the target object during visual search. Therefore, attending to a stimulus at a particular location or with a particular feature biases this competition in favour of neurons that respond to the feature or location of the attended stimulus. This external bias, therefore, produces the attention effect because the cells representing the attended stimuli are given an advantage such that they "win" the competition and cells representing distractor stimuli are suppressed. This idea is currently very influential in many fields, including modelling (Deco & Lee, 2002; de Kamps & van der Velde, 2001; Hamker, 1998; Reynolds, Chelazzi, & Desimone, 1999; Usher & Niebur, 1996) and is supported by a growing body of neurophysiological evidence, which is briefly reviewed next.

### 1.4.2.1 Neurophysiological Evidence for Biased Competition

A seminal study in monkey area V4 (Moran & Desimone, 1985) showed that attention acts to gate the information processed in extrastriate cortex. It appeared that the receptive field of neurons "contracted" around an attended stimulus so that, to a large extent, the presence of other stimuli in the receptive field did not influence the response. Further single cell studies (Chelazzi et al., 1993, 2001; Miller, Gochin, & Gross, 1993; Motter, 1993; Reynolds et al., 1999) showed that a high response to an effective stimulus (i.e. a stimulus that caused a strong response from the cell when presented alone) presented in the cell's receptive field was attenuated by the addition of a second ineffective stimulus (i.e. a stimulus that caused only a weak response when presented alone in the receptive field). Responses were eventually determined by which of the two stimuli was attended. This effect is shown in figure 1.2, which is reproduced from Chelazzi et al. (1993) and shows average population activity in IT over time. Further plots of cellular activity in IT and V4 will be shown in chapter 5, where biased competition effects are simulated.

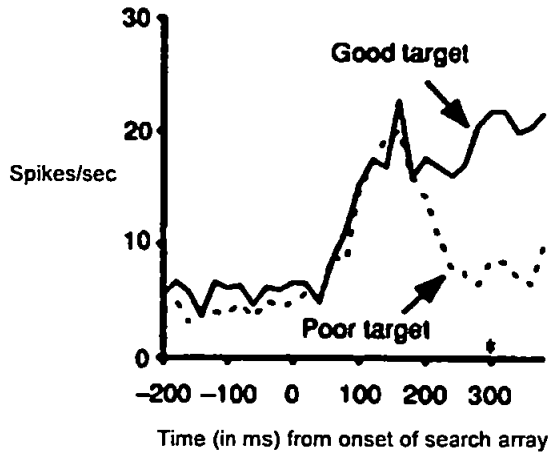



**Figure 1.2 Average population responses in monkey IT**

This plot is reproduced from Chelazzi et al. (1993: *figure 3a*) and shows responses in IT in two conditions where two stimuli are present in the cell's receptive field. One of the stimuli is effective in driving the cell (the "good" stimulus) and the other is ineffective (the "poor" stimulus). The solid line shows the response when the effective stimulus is the search target and the dotted line shows the response when the ineffective stimulus is the target. The initial response of the cell is the similar in both cases. However, later in the response, the firing rate depends on which object is the search target, i.e. there is an object-based attention effect developing late in the response. Figure used with the permission of Nature Publishing Group (http://www.nature.com/nature) and L. Chelazzi.

When an effective and an ineffective stimulus were present in the receptive field, figure 1.2 shows that the response was high when the effective stimulus was attended. This response was similar to the response when this stimulus was presented alone. However, if the ineffective stimulus was attended, responses were severely suppressed, despite the presence of the effective stimulus in the receptive field. This effect occurred both when the location of the stimulus to be attended was indicated by a spatial cue (Luck et al., 1997) and when the stimulus was

selected as a result of the search target object being held in memory (Chelazzi et al., 1993, 2001). These effects were strongest when both stimuli were present within the same receptive field and reduced when stimuli were moved apart such that one was no longer within the receptive field (Luck et al., 1997; Moran & Desimone, 1985). Therefore, it appears that stimuli engage in local competition that may be biased by spatial and target object information such that the response of the cell is determined by which of the stimuli are attended. Desimone (1998) outlines the tenets of biased competition in more detail. Further evidence for biased competition, as it relates to the model presented here, is considered in the next chapter.

#### 1.4.2.2 Integrating Approaches

The model presented in subsequent chapters integrates elements of Feature Integration Theory (a spatial "spotlight", which is due to an eye movement shifting attention to a new location, facilitates binding across features) within a biased competition framework. This is compatible with the biased competition approach since biased competition suggests that space is a feature capable of being preferentially biased. Here, spatial attention will be one of a number of different biases operating within the system and providing a "spotlight" bias to the dynamics of the system that enhances activity and could facilitate the integration of features within it.

### 1.4.3 Featural Guidance During Visual Search

When searching for an object, eye movements do not randomly scan the scene. Instead, during active visual search[2], saccades tend to be directed towards stimuli rather than blank areas (e.g. Gilchrist, Heywood & Findlay, 1999; Motter & Belky, 1998b). It appears that search can be *guided* by attributes of the target such that erroneous saccades (those not directed towards the target) are most likely to be directed towards a non-target (distractor) sharing a feature with the target (Findlay, 1997). Two early studies show that guidance by target colour is particularly strong. During search of simple geometric forms Williams (1967) found that specification of the target colour led to the search being much faster, with the majority of fixations falling on distractors sharing the target colour. Specification of the size of the target was much less effective and specification of shape provided almost no advantage. Luria and Strauss (1975) also found that subjects showed a marked tendency to direct saccades to distractors sharing the target colour. A similar but weaker effect was seen for target shape.

Most recent studies also support the idea of selectivity in search eye movements. Scialfa and Joffe (1998) used white and black bars tilted to the left or right and found that participants were three times more likely to fixate distractors that shared the target's contrast/colour than distractors that did not. Williams and Reingold (2001) found that saccades tended to be directed towards distractors sharing the

[2] Active visual search uses *overt* shifts of attention so that the eye position moves. This is distinguished from *covert* shifts of attention that occur in the absence of eye movements, i.e. where fixation is maintained but the spatial focus of attention moves.

target's colour in a single feature task (in which distractors shared only one feature with the target) or sharing the target colour and shape in a two-feature task (in which distractors shared two features with the target). This finding was robust across display sizes and a range of saccadic eccentricities. The strength of selectivity became stronger as the trial progressed, indicating an increasing bias towards the target's features. Also, subjects were able to use features to make the search more selective when the discriminability of the stimuli was increased, indicating that bottom-up stimulus factors are also important. Despite the majority of evidence supporting a selective search model, Zelinski (1996) has argued against search being selective and found only weak selectivity such that distractors sharing no target features captured attention almost as often as those sharing a target feature. However, this study did not separate the effects of colour guidance from that of orientation and, therefore, may underestimate the selectivity for colour.

The evidence for colour guiding search more strongly than other features is not conclusive, for example Bichot and Schall (1999) found that colour and shape information were used to similar degrees. However, a good deal of evidence indicates that colour is an important cue in driving search. This effect is found in monkeys as well as humans and the work here combines the replication of data from monkey single cell studies with replicating, at the systems level, a monkey psychophysical study (Motter & Belky, 1998b) that found strong guidance of search by colour. Motter and Belky (1998b) used a colour-orientation conjunction search task such that distractors shared either the target's colour or its orientation,

and there were an equal number of each distractor type. They found that the overwhelming majority (75%) of fixations landed near target coloured stimuli as opposed to non-target coloured stimuli having the target orientation (only 5% of fixations) or blank areas of the display (20% of fixations). This priority for colour guiding search was found even when the task was heavily biased towards orientation discrimination. Thus, it appears that colour is able to segment the scene (Motter, 1994a,b) and guide the search process.

Achieving a model of visual search guided by target colour was one of the main objectives of the work reported here. It is accomplished as a result of object-based attention effects in the model's ventral stream. Other factors that might influence search strategy are beyond the scope of the current work. For example, when distractor types are not in equal proportion in the display, saccadic selectivity may be biased towards the feature dimension that has the fewest distractors sharing this feature with the target (Bacon & Egeth, 1997; Shen, Reingold, & Pomplum, 2000, 2003). This is known as the *distractor-ratio effect*. Also, the discriminability of features distinguishing targets from distractors may provide a strong bottom-up effect such that, when a feature is less useful to distinguish the target, search proceeds through the group of stimuli defined by a more discriminable feature (Sobel & Cave, 2002). These bottom-up effects are not addressed here. However, an extension of the model to reproduce such bottom-up effects is discussed in the final chapter. The current work uses images similar to those used by Motter & Belky (1998a,b) in order to replicate the effects observed in that and similar

experiments. These images have equal numbers of each type of distractor and each feature dimension (colour and orientation) is highly discriminable.

## 1.5 Relation of the Work to Prior Models in the Field

As discussed above, attention can be captured by both top-down and bottom-up factors. Models of bottom-up attention capture typically use a "saliency map" (Koch & Ullman, 1985) or "master map" (Treisman & Gelade, 1980). Koch and Ullman's saliency map was the first neurally plausible computational architecture for controlling visual attention. Its saliency map was a two-dimensional topographic map that encoded stimulus saliency, i.e. how conspicuous the stimulus, at every location in the visual field. Many subsequent computational models of bottom-up attentional capture use this concept (see, for example, Itti & Koch, 2000; Niebur, Koch, Rosin, 1993). Typically, features are detected in various feature maps that feed into a single saliency map, which is similar in concept to Treisman's 'master map'. Desimone & Duncan's biased competition (1995) approach suggests that saliency-based attention can be achieved without the need for such a priority map. The current work is compatible with biased competition and results in a map of behavioural relevance (a behavioural "saliency" map) forming in a posterior parietal (LIP) module. This parietal area may form part of a network of cortical and sub-cortical structures that comprise the neural correlate of a saliency map, as will be discussed in the final chapter. The map described here is strongly

influenced by object-based attentional effects that progress in a top-down manner through the model's ventral stream.

Also, in contrast to the purely bottom-up approach, Wolfe's *Guided Search Model 2.0* (GS2: Wolfe, 1994, which replaced the original version of the model described by Cave & Wolfe, 1990, Wolfe & Cave, 1989, and Wolfe, Cave & Franzel, 1989.) suggested that the selection of relevant features during search could be performed top-down by means of spatially defined and feature-dependant weighting of various feature maps. In this model, saliency is determined by both the bottom-up occurrence of a feature and a top-down feature weighting. It builds upon Treisman's feature integration ideas but also has some similarity to the concepts outlined by biased competition. In Guided Search, maps for each feature type (e.g. colour, orientation) are subject to bottom-up stimulus information and top-down commands. Therefore, the activity in each feature map is determined both by bottom-up factors (based on local differences, i.e. how unusual an item is in its present context) and top-down task demands. The feature maps feed, in a summable manner, into an *activation map*, which determines where (to which location/object – these are considered to be the same) attention is guided in a serial fashion. Thus, the parallel feature-computation stage guides a serial attention stage. Importantly, in GS2, knowledge of task requirements is allowed to influence the relative contributions of different feature maps to the activation map. Attention is deployed to locations in order of decreasing activity in the activation map. The architecture of the model presented in later chapters has much in common with this

approach in that serial attention will be guided by features in the scene that are found to be behaviourally relevant following their parallel processing. However, the model will not have distinct pre-attentive and attentive stages. Instead, significant attentional effects gradually emerge from the dynamics of the system.

Deco (2001; Deco & Lee, 2002; Rolls & Deco, 2002) has produced an influential computational model that uses top-down biases to influence competition within the system. In this model the distinction between pre-attentive and attentive stages is blurred. Attentional effects develop within the system due to the ongoing dynamic interactions of the modules within it. This model is examined in more detail in the next chapter because the work reported here extends this seminal approach. However, in contract to Deco's model, the work here adopts an active vision approach, where eye movements produce overt shifts of attention, as opposed to covert attentional shifts.

Neurobiological computational models of visual attention have tended not to use active vision, i.e. overt attention. This approach is more common in the computer vision environment. However, some biologically-inspired models of attention specifically operate in an active vision manner. For example, Rybak, Gusakova, Golovan, Podladchikova, & Shevtsova (1998) designed a model with a moving attention window (AW) within which features were detected at specified locations in a grid arrangement. Locations competed on the basis of their luminance contrast and the winning location became the next fixation point. Thus, attention moved

according to bottom-up features without top-down control, except under *recognition mode* when the system carried out a memorised series of eye movements, under top-down control, once it had linked a bottom-up feature with one in a previously memorised scene. This system employs an efficient method of processing the scene and storing previously memorised images because it only processes the subset of the image that falls within the AW and stores just the features at the preset locations within the AW. The work described here similarly uses an active vision approach that is more efficient, and biologically plausible, than processing the entire image. However, the system has access to all features falling upon the retina at each fixation and analysis is not constrained to features at specific grid locations. An attention window forms within this retinal view and provides a spatial enhancement but does not restrict cortical processing of all retinal locations in parallel.

The work here is compared to other pertinent models in the field in the final chapter.

## 1.6 Motivation for the Work and Aim

The biased competition hypothesis (Desimone, 1998; Desimone & Duncan, 1995; Duncan, Humphreys & Ward, 1997) is currently of particular interest in the visual attention literature. Neurophysiological experiments have found evidence that

supports the hypothesis (e.g. Chelazzi et al., 1993, 2001; Miller et al., 1993; Moran & Desimone, 1985; Motter, 1993, 1994a,b; Reynolds et al., 1999) and it has been influential in other experimental paradigms, including brain imaging (e.g. Kastner, Pinsk, De Weerd, Desimone & Ungerleider, 1999) and evoked potentials (Anllo-Vento & Hillyard, 1996; Hillyard & Anllo-Vento, 1998; Luck et al., 1993; Luck & Hillyard, 1995). However, there is still much work to be done to explore this hypothesis computationally in biologically plausible systems. Early models tended to be small scale with only a few interacting units (e.g. Reynolds et al., 1999; Usher & Niebur, 1996) but recently systems level models have begun to emerge (Deco, 2001; Deco & Lee, 2002; Hamker, 1998). However, there has been no systems level computational modelling of the biased competition hypothesis for active visual search, where retinal inputs change as the focus of attention is shifted. The work presented here addresses this issue, using the biased competition framework and adopting an active vision approach such that, at any particular fixation, its cortical modules receive retinal inputs relating to a portion of the entire scene. In doing so, a serial spatial component is added to the parallel biased competition processing.

There is also an important requirement for models to explain the presence of both spatial (e.g. Bricolo, Gianesini, Fanini, Bundesen & Chelazzi, 2002; Connor et al., 1996) and object-based (e.g. Blaser et al., 2000; Chelazzi et al., 1993, 2001; Duncan, 1984; O'Craven et al., 1999; Roelfsema et al., 1998; Valdes-Sosa et al., 1998; Valdes-Sosa et al., 2000) attentional effects and the different latencies

associated with them at the single cell level. The model presented in subsequent chapters incorporates both spatial and object-based attention at the cellular level with the onset of these effects accurately replicating those seen in single cell recordings (Chelazzi et al., 1993, 2001; Luck et al., 1997; Motter 1994 a, b). It is novel in being the first to examine the timing of these effects in detail.

Avoidance of blank areas of a scene and guiding search to behaviourally relevant locations is important in computer vision as well as being something that biological systems appear to be designed to achieve. The scan path is the series of fixation points arrived at by saccades during the search process. Human and monkey scan paths are attracted to interesting locations in the scene and fixations tend not to land in blank areas. In particular, Motter and Belky (1998b) found only 20% of monkey's fixations during feature conjunction search were in blank areas of the display. A "cross-stream" interaction between the ventral and dorsal streams of the model presented later allows a parietal area (LIP) to act as a feature saliency map (Colby & Goldberg, 1999; Gottlieb, Kusunoki & Goldberg, 1998; Koch & Ullman, 1985; Kusunoki, Gottlieb & Goldberg, 2000) that is guided by object-based attention in the ventral stream so that the parietal representation becomes more behaviourally relevant, as has been found in single cell recordings (Colby, Duhamel & Goldberg, 1996). The nature of the saliency map in this model differs from that in some others (e.g. Niebur, Itti & Koch, 2001) normally formed on the basis of "bottom-up" stimulus information. The representation here develops from object-based attention information from the ventral stream, which is subject to a

"top-down" bias relating to the search target. Hence, this representation in LIP is used to guide search to behaviourally relevant locations and avoid blank areas of the scene.

A further aim of this work was to replicate, at the systems level, active vision scan path behaviour found in monkeys (Motter & Belky, 1998b) and humans (Lauria & Strauss, 1975; Scialfa & Joffe, 1998; Williams, 1967; Williams & Reingold, 2001), where attention is preferentially drawn to locations containing the target colour. During a colour-orientation feature conjunction search, Motter and Belky (1998b) found that colour (or luminance) appeared to segment the scene and fixations tended to land on distracters that were the same colour as the target (75% of fixations), rather than those that are the same orientation (only 5 % of fixations). Thus, surface features, such as colour, could have a preferential role in guiding search. Also, colour seems an important cue in attentional processing. Even abrupt onsets can be ineffective in overriding a "top-down" attentional set for colour and, thus, capturing attention in an exogenous manner (Folk & Remington, 1999). In some circumstances, colour segmentation of the scene appears to assist motion discrimination (Croner & Albright, 1997). Such a priority for one feature type (colour) over another (orientation) has not been specifically captured by leading computational models in this area (e.g. Deco & Lee, 2002; De Kamps & Van der Velde, 2001; Hamker, 1998; Niebur, Itti, Koch, 2001). However, it is useful to provide a model that allows such a priority not only to explain some psychophysical data but also because of the potential application in computer

vision. Therefore, the aim here was to allow a feature type (colour) to have priority in guiding the active vision scan path in addition to ensuring that the scan path targeted behaviourally relevant stimuli rather than blank areas.

Thus, the aim of this work was to produce a computational model for active visual search that was able to replicate data from monkeys or humans at both the cellular and systems levels. The specific research question addressed was:

***Can a neurodynamical model of visual attention in the ventral and dorsal cortical processing streams account for spatial and object-based attention observations in monkey areas IT, V4 and LIP; and can such effects result in attentive scan paths similar to those seen in monkey feature conjunction active visual search?***

Therefore, object-based attentional effects at the cellular level of the model should replicate those found in monkey single cell recordings (e.g. Chelazzi et al., 1993, 1998, 2001; Luck et al., 1997) and lead to systems level behaviour of the model that replicates active visual search behaviour where attention is drawn to target coloured locations (in particular, Motter & Belky, 1998b; and also Lauria & Strauss, 1975; Scialfa & Joffe, 1998; Williams, 1967; Williams & Reingold, 2001 for human behaviour). This would provide "top-down" object-based guidance of an active vision scan path. In order to achieve this, a novel "cross-stream" interaction between the ventral "what" pathway and the dorsal "where" pathway is

proposed. This connection will allow the search scan path to be drawn to behaviourally relevant locations rather than less relevant stimulus locations or blank areas.

Thus, the model developed here uses object-based attention effects at the cellular level in the ventral stream to reproduce active visual search behaviour at the systems level such that colour is given priority in the featural guidance of the scan path to behaviourally relevant locations.

The model is implemented in Matlab (version 6.5), in order to allow the work to focus on specific research questions rather than the implementation issues involved with third generation languages (such as C++). The model is introduced in the next chapter and is described in detail in chapters 3 and 4. Cellular level results and predictions are presented in chapter 5, and systems level results and predictions are presented in chapter 6. The behaviour following lesion is described in chapter 7 where patient psychophysical data is replicated. Finally, the results and implications of the model are discussed in the final chapter. Supplementary information is provided in the appendices and will be referred to, when relevant, in the main body of the text. Copies of selected published papers are included after the list of references. A full list of publications resulting from this work and a list of conference presentations was included in the author's declaration in the introductory pages.

# Chapter 2

# Introduction to the Model

The following items are introduced in this chapter:

- The research goals forming the basis of the modelling work
- The computational model that forms the central component of the work

The model is introduced here at a high level, by describing its overall architecture and the active vision approach, and is then defined more comprehensively and formally in the following two chapters.

## 2.1. Research Goals

Underpinning this work are two research goals, which were investigated by implementing a biologically-constrained computational model. This model will be described in this and subsequent chapters and is the major contribution of this thesis.

Modelling focuses on three cortical regions. Ventral stream modelling is concerned with extrastriate area V4, which encodes form and colour features, and anterior inferior temporal cortex (IT), which encodes invariant object representations. In

the dorsal stream, posterior parietal area LIP is modelled. Attentional effects in these areas are relatively well studied, as will be described in this and subsequent chapters. The primary visual cortex, area V1, is also modelled. It provides feature detection in a purely feedforward manner to V4. However, the main focus of the work is on attentional effects in intermediate stages of visual processing. Other areas, such as prefrontal cortex are modelled as external biases to the system. The architecture of the model will be introduced in this chapter and described formally and in detail in the next two chapters.

The two fundamental research goals and associated working hypothesises, which the model was built to investigate, are as follows:

### 2.1.1. Object-based attention, operating concurrently with spatial attention, becomes significant later in the response in area V4 than a spatial effect associated with the allocation of attention to a location following an eye movement.

***Specific Research Goal:***

*To investigate computationally whether and how spatial and object-based attention can operate concurrently in the ventral stream and produce attentional effects over time courses similar to those found in neurophysiological studies of area V4.*

Event-related potential and single cell recordings suggest that spatial attention may be able to modulate responses earlier in the stimulus-related response than object-

based attention. For example, in monkey lateral intraparietal area (LIP), an early spatial enhancement of responses has been recorded from single cells in anticipation of the stimulus (Colby et al., 1996) and also has been seen in imaging of the possible human homologue of LIP (Corbetta et al., 2000; Hopfinger et al., 2000; Kastner et al., 1999). Imaging of extrastriate cortex (which includes the human homologue of monkey area V4) also shows enhanced activity in anticipation of a stimulus at a specific spatial location (Kastner et al., 1999). In monkey area V4, spatial attention can modulate baseline responses (by ~42%) in advance of a sensory response, thus enhancing incoming sensory data at an attended location such that the earliest stimulus-invoked response at 60ms post-stimulus can be modulated (Luck et al., 1997). This effect occurs even when the location marker is removed and the animal is simply trained to expect a stimulus to appear at a particular location. In this case, a baseline shift in responses occurs before the stimulus appears and continues into the sensory responses throughout the trial. The baseline shifts do not depend on whether the stimulus is an effective or ineffective stimulus for the cell and happen only when the animal is attending to a location inside the receptive field. Therefore, this reflects the allocation of spatial attention.

Object-based effects in V4 appear to take longer to be resolved and have not been recorded until at least 150ms post-stimulus (Chelazzi et al., 2001; Motter, 1994a,b). When an effective and an ineffective stimulus are presented simultaneously within the cell's receptive field, the initial response of the cell is the same regardless of

which object is attended (or is the search target). Then, from about 150ms after the onset of the array, the cell responds as if the effect of the non-attended stimulus is being filtered out. Therefore, when the ineffective stimulus is the target, the cell's response is suppressed, despite the presence of the effective stimulus in its receptive field. However, if the target is the effective stimulus, the cell continues to respond strongly. Such modulation, leading to the response of the cell being determined by which object within its receptive field is attended, is known as the "target effect" (Chelazzi et al., 2001). These object-based (or feature-based) effects lead to enhanced responses to target features. When searching for an object with a particular feature (colour or luminance), it appears that responses to that feature are enhanced in parallel across V4 (Motter 1994a,b) and in the human V4 homologue (Saenz et al., 2002).

In monkey IT, the average initial sensory response, like baseline responses during a delay period following a cue, is slightly modulated by the target object. However, this target effect in the early part of the sensory response in IT is not generally significant compared to the strong target effect developing, very similarly to that in V4, from 150-200ms after the onset of the search array (Chelazzi et al., 1998). No similar delay period baseline shift is found in V4. The shift in the early response in IT is less than the baseline shift in V4 relating to spatial attention (Chelazzi et al., 1998 compared with Luck et al., 1997). Thus, object-based attentional modulation of ventral stream signals appears to take at least 150ms into the sensory response to become substantial, despite a prior memory trace of the target object. In contrast,

anticipation of a stimulus at a specific spatial location significantly modulates baseline responses (by ~42%) in advance of the stimulus-related response and from its onset (Luck et al., 1997). The late onset of significant object-based effects has not been captured by previous models in the field (e.g. Deco, 2001; Deco & Lee, 2002; Hamker, 2004; de Kamps & Van der Velde, 2001; Rolls & Deco, 2002; Usher & Niebur, 1996; van der Velde & de Kamps, 2001). The current work investigates the mechanisms necessary to instantiate object-based attention in the ventral stream over the appropriate time course. Object-based effects may take longer to develop because they depend on feedback from higher levels of the ventral stream hierarchy and the resolution of competition between objects at those levels. So, these effects in V4 could depend on the resolution of competition between objects in IT. As will be seen later, the timing of prefrontal feedback to IT was found to be crucial to the timing of target object effects in IT. Although, the prefrontal response is typically modelled as simply a sustained mnemonic response (Deco, 2001; Deco & Lee, 2002; Rolls & Deco, 2002; Renart, Moreno, de la Rocha, Parga & Rolls, 2001; Usher & Niebur, 1996), it was found here to be necessary to consider the sensory response in prefrontal cortex along with its mnemonic response in order to accurately replicate the time course of object-based effects.

Support for spatial effects preceding object-based ones, such that there may be a prior selection of location, is found in event-related potentials (ERP) studies (Anllo-Vento & Hillyard, 1996; Anllo-Vento, Luck & Hillyard, 1998; Hillyard &

Anllo-Vento, 1998). These data indicate that, whilst spatial attention can affect extrastriate responses from ~80ms post-stimulus, paying attention to a non-spatial feature, such as colour, motion or shape, affects ERP patterns with a latency of at least 100-150ms. Spatial effects influence the P1 component (ventral-lateral extrastriate cortex) at 80ms post-stimulus, whereas feature-based effects result in a widespread "selection negativity" from 150-300ms post-stimulus and a "selection positivity" peaking at 200-240ms in frontal sites (Anllo-Vento & Hillyard, 1996). Furthermore, these data suggest that feature selection may be contingent upon prior spatial selection (Anllo-Vento & Hillyard, 1996; Hillyard & Anllo-Vento, 1998). The "selection negativity", related to attention to colour or motion, was found to be reduced (or even absent) for task-relevant stimuli positioned at unattended locations (Anllo-Vento & Hillyard, 1996). Also, responses elicited by stimuli lacking the attended feature but at an attended location were initially as high as stimuli with the attended feature at the attended location. Only later stages of processing became contingent upon the attended feature. Thus, the ERP data appears to support a model in which there is an early spatial selection bias and later non-spatial feature selective processing. Single cell results suggest that feature-based modulation may occur in combination with a spatial modulation (McAdams and Maunsell, 2000; Treue & Martinez Trujillo, 1999) and human PET measures suggest that similar areas are activated for both types of attention (Fink, Dolan, Halligan, Marshall, & Frith, 1997).

The *premotor theory of attention* (Rizzolatti, Riggio, Dascola & Umilta, 1987) suggests that movements of attention activate the same brain structures as do eye movements and that attention is allocated to a location when the oculomotor programme for moving the eyes to this point is ready to be executed. Therefore, in this theory, the allocation of spatial attention and an eye movement (whether actually carried out or not) are intimately linked. This tight coupling of the neural correlates of spatial attention and eye movement is an extreme view. However, it is supported by recent evidence showing that microstimulation of superior colliculus (an area directly involved in the generation of saccades), when the gaze is held in a fixed position, produces covert spatial attention (Muller, Philiastides. & Newsome, 2005). There is also psychophysical evidence to suggest a spatial attention effect exists at the saccade target location immediately before the saccade takes place and lasts some time after the saccade (Shepherd, Findlay & Hockey, 1986). Further, the eye movement appears to limit the allocation of attention elsewhere so that target detection at the location to which the saccade is directed is superior to that at other locations to which subjects have been told to attend (Hoffman & Subramiam, 1995).

In the work here, it is assumed that there is an early spatial attention effect associated with the eye movement. It is suggested that this effect is able to spatially enhance the development of object-based attention, which occurs in parallel across the scene. Over time, object-based attention develops and is facilitated, but not constrained, by the aperture of the initial spatial attention

window. The benefit for features within the initial spatial focus in V4 is that they are activated slightly more strongly and may receive preferential access to higher visual areas, such as IT and prefrontal cortex. Both types of attention cooperate concurrently in the model but result from different forms of bias to the competitive dynamics within the system such that the effects emerge over different time scales. Also, the spatial effect allows the focus of attention, or perceptual span, to be scaled according to the density of stimuli, as will be described later.

### 2.1.2. The enhancement of target features in parallel across V4 as a result of object-based attention is able to bias parietal cortex to represent behaviourally relevant locations and attract the scan path.

***Specific Research Goal:***

*To investigate whether object-based effects occurring in parallel across V4 (Motter 1994a,b) can be used to bias spatio-featural representations in LIP towards behaviourally relevant locations and, hence, guide the attentive scan path. In addition, to test the feasibility of using this route to provide a priority for certain features in attracting the scan path.*

LIP encodes behaviourally relevant salient stimuli (Gottlieb et al., 1998) and such a "salience map" has been postulated to guide visual search (Koch & Ullman 1985, Wolfe 1994). LIP neurons are able to encode both spatial and featural information when the information is behaviourally relevant (Toth & Assad 2002). Also,

parietal cortex in humans has been linked to both spatial and object-based attention (Fink et al., 1997).

Featural information relevant to the search target is enhanced in parallel throughout the visual field in area V4 (Motter 1994a,b; Saenz et al., 2002). Due to its retinotopic organisation, it is possible that V4 could have an important role during visual search in guiding the focus of attention towards relevant (e.g. target coloured) objects throughout the scene. Connections from V4 to LIP (Shipp & Zeki, 1995) could convey information about the location of behaviourally relevant stimuli so that LIP is able to represent these locations most strongly (Gottlieb et al., 1998) and become an integrator of spatio-featural information. Under certain conditions colour is able to guide search more strongly than orientation in a simple feature conjunction search (Motter & Belky, 1998b; Lauria & Strauss, 1975; Scialfa & Joffe, 1998; Williams, 1967; Williams & Reingold, 2001). The connection from V4 to LIP is one method by which such a priority could be implemented through the relative strength of synaptic connections, either being weighted innately or modified dynamically depending on task.

In the work here, these ideas are used to guide visual search by top-down target object knowledge. When a subject has exact knowledge of the target object search is faster than if there is no knowledge of the target or the information is inexact (Vickery, King, Jiang, 2005). In principle the guidance of the attentive scan path here does not rely on the spatial bias discussed above and would operate

successfully on the basis of the object-based bias alone. However, the spatial bias, as implemented here, is advantageous to some aspects of attentive scanning. This will be demonstrated later in relation to facilitating shorter amplitude saccades for detailed inspection in areas of dense stimuli.

The following sections progress these goals into a more concrete computational model. First, the foundation of the model, in terms of prior computational work in the field, is explained.

## 2.2. Previous Modelling Work Fundamental to the New Model

The work reported here extends the approach taken by the seminal biased competition model developed by Deco (2001; Deco & Lee, 2002) and described most comprehensively in Rolls and Deco (2002, chapter 9). In that model, attentional effects are developed within the dynamics of the system. Whilst not built directly upon Deco's model, the work here draws much inspiration from it and there are many similarities between the models. There are, however, significant differences from Deco's model and these will be outlined here and in relevant parts of later chapters. In particular, the model reported here is the first biased competition model to adopt an active vision paradigm where retinal inputs to the system change during the scan path (although a covert attention mode also exists). Significantly here, spatial and object-based biases operate concurrently to influence

the system's dynamics and allow the time course of spatial and object-based attentional effects to be investigated in detail. This section briefly outlines Deco's model and a smaller predecessor by Usher and Niebur (1996). A review and comparison with other influential models in this area (e.g. Heinke and Humphreys, 2003; Itti and Koch, 2000; Tsotsos, 1993, 1995) will be made in the final discussion chapter.

Biased competition was first implemented computationally by Usher and Niebur (1996). They presented a small-scale (in terms of number of interacting units) model in which object representations in IT competed with one another. These objects were activated by bottom-up inputs from an input layer representing V1. The competition was biased by a prefrontal signal relating to working memory of the target object. This bias had a direct excitatory influence on the target object representation such that, over time, it became more active than other objects in IT. The model was implemented at two levels: A *micro* level using an integrate-and-fire scheme and a *macro* level using mean-field population dynamics. The model showed that a small prefrontal bias was capable of affecting competition between objects in IT, so that the target object became more active than distracters in a similar manner to that measured in monkey neurons (Chelazzi et al., 1993).

Following this seminal but small-scale contribution, Deco (2001; Deco & Lee, 2002; Rolls & Deco, 2002) presented a larger-scale biased competition model using the mean-field approach. The model consisted of a ventral, object-recognition,

pathway consisting of IT and V1, which detected orientation features at different spatial frequencies using Gabor jets (Lee, 1996). A Gabor jet consists of a series of Gabor wavelets whose spatial frequencies increase in octaves. For certain simulations a V2-V4 module was also included. This pathway formed a reciprocally connected hierarchy. A posterior parietal module (PP) existed in the dorsal stream and was reciprocally connected to the V1 module. Competition operated between objects in IT, between spatial locations in PP and between different orientations within the same spatial frequency in V1. This was the first time spatial and object-based attention had been integrated within a single model. The model operated in either object recognition or visual search modes. In *object recognition mode*, a top-down spatial bias from prefrontal cortex was applied to the PP module, allowing this location to win the competition in PP. The connection from PP to V1 created a bias towards representation of the object at the selected location in the model's ventral stream (V1 and IT) so that the system recognised the object at the chosen location. In *visual search mode*, a top-down object-specific bias from prefrontal cortex was applied to the IT module. There was no spatial bias in this mode. The object bias drove the competition within IT such that the selected object won. The effect of feedback from IT to V1 was then to bias the competition in V1 towards features, relating to this object, which were detected from retinal inputs across the scene (the static retina encompassed the entire scene). After convergence, only the V1 pools associated with the chosen object remained active and the location in PP at which the object was found was most active. So, in this model, attention was not a specific input to the competition but "an emergent effect

that supports the dynamical evolution to a state where all constraints are satisfied" (Rolls & Deco, 2002, chapter 9, p. 335). Also, the distinction between pre-attentive and attentive (Neisser, 1967) processing is blurred, a concept that is now becoming more widely accepted (see Shipp, 2004) and is pursued in the work here. In common with Deco's model, the work described here seeks to unify different levels of neuroscience: Neurobiological, neurophysiological and psychophysical (see Deco, Pollatos & Zihl, 2002; Rolls & Deco, 2002, for discussion).

## 2.3. Overview of the New Model

This section is intended as an introduction to the model in order to set the context for the formal definition of the model given in the next two chapters.

### 2.3.1 Active Vision Approach

In contrast to Deco's model, where attention is covert (i.e. the retina is static), the model here operates in an active vision manner by moving its retina around the image so that its view of the world is constantly changing. Cortical areas receive different "bottom-up" input at each fixation. Here, this is referred to as the "retinal view" or "retinal image". This reflects the way one would carry out search in everyday life. This is the first time that the biased competition approach has been applied in an active vision paradigm. The active vision approach raises issues not

addressed by models with static retinas and covert attention (e.g. Deco, 2001; Deco & Lee, 2002; Niebur, Itti & Koch, 2001; Hamker, 1998, 2003, 2004), such as keeping track of locations in the scene already visited but no longer available within the current retinal coordinate system. This issue will be addressed in chapter 4. The majority of the work reported here uses an active vision approach where attention is overt. However, the system is also able to operate in a covert attention mode and some simulations in this mode will be presented in chapter 7.

### 2.3.2 Model Architecture

The focus of the model is on intermediate stages of visual processing, and attention arises as a result of the on-going dynamics of the system, which are implemented within each module in a similar way to the model described by Deco (2001; Deco & Lee, 2002; Rolls & Deco, 2002, chapter 9). However, in contrast to Deco's model, the model here does not operate in two separate attention modes (object-based for locating the target, and spatial for object recognition). Instead, it is assumed that both forms of top-down bias may be available to the system. Thus, spatial and object-based biases operate concurrently to influence the dynamics of the system. In contrast to Deco's purely parallel approach, there is a serial spatial component related to the eye movement here. A spatial bias relating to the re-positioning of attention following an eye movement influences the system's dynamics. Attentional effects emerge from the dynamics of the system, which are subject to both forms of top-down bias in addition to the "bottom-up" stimulus

information. This is the first time that the time courses of spatial and object-based attentional effects at the cellular level have been investigated in detail in a computational model.

The model consists of two portions: A "dynamic" portion within which the attentional effects develop and a prior "pre-processing" stage, which comprises modules that do not participate in the dynamic interactions but provide featural information in a feedforward manner. Within the dynamic portion of the system are three modules which interact over time to achieve the attentional effects and performance within the system. The modules in the dynamic portion of the system represent brain regions from both the ventral pathway leading to temporal cortex and the dorsal pathway leading to parietal cortex (Milner & Goodale, 1995; Ungerleider & Mishkin, 1982). "Cross-stream" interactions between areas in the two pathways are an important part of the model. In the ventral pathway are modules representing areas V4 and (anterior) IT and, in the dorsal pathway, is a module representing posterior parietal cortex area LIP. These modules interact as represented schematically in figure 2.1, where the dynamic portion of the model is shown in the colour blue. These interactions, bottom-up stimulus information from the retina and V1, and external biases related to the target object (from ventrolateral prefrontal cortex) and the currently attended location (a spatial bias through LIP to V4) influence the competitive dynamics within the system. This dynamic portion of the model is modelled using mean field population dynamics, where representation is at the level of populations of cells, known as *pools* or *assemblies*.

Each assembly is assumed to contain neurons with similar properties and is updated by means of a differential equation, as will be described in detail in chapter 4. Within each of the dynamic modules are sets of these cell assemblies, each representing a pool of excitatory pyramidal neurons. Each module also contains one or more inhibitory interneuron assemblies to mediate competition between the pyramidal assemblies. Figures 2.2, 2.3 and 2.4 show the inhibitory interneurons and external biases that influence competition in each of the dynamic modules. These dynamic processes will be described in more detail in chapter 4.

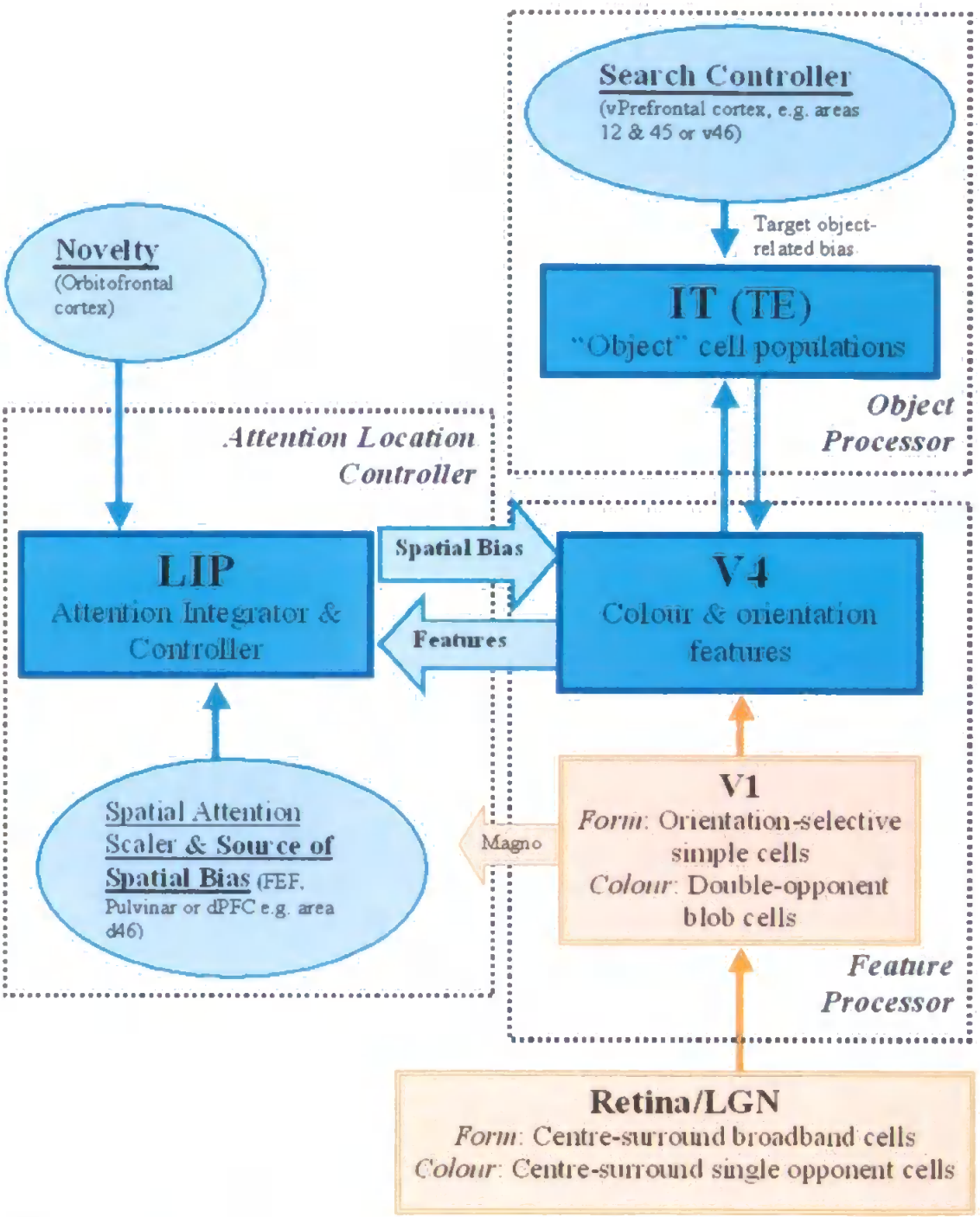



**Figure 2.1 Model architecture**

The architecture of the model shows the processing modules and connections. The dynamic portion of the model is shown in blue. V1 and retina/LGN, shown in orange, form a "pre-processing" stage and do not take part in the dynamic attentional interactions. External biases to the system are shown as ovals.

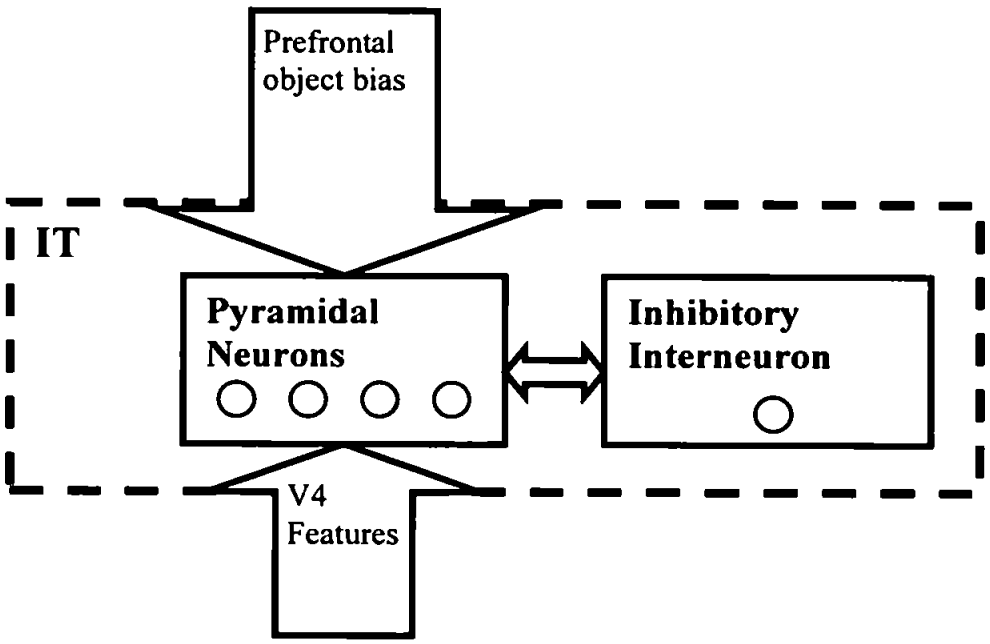



**Figure 2.2 Competition in IT**

The inhibition and biases influencing competition in the IT module are shown. This module encodes invariant object representations and the representations are activated by bottom-up inputs from V4. Competition between objects is mediated by a single inhibitory interneuron assembly and is biased by target object-related feedback from prefrontal cortex.

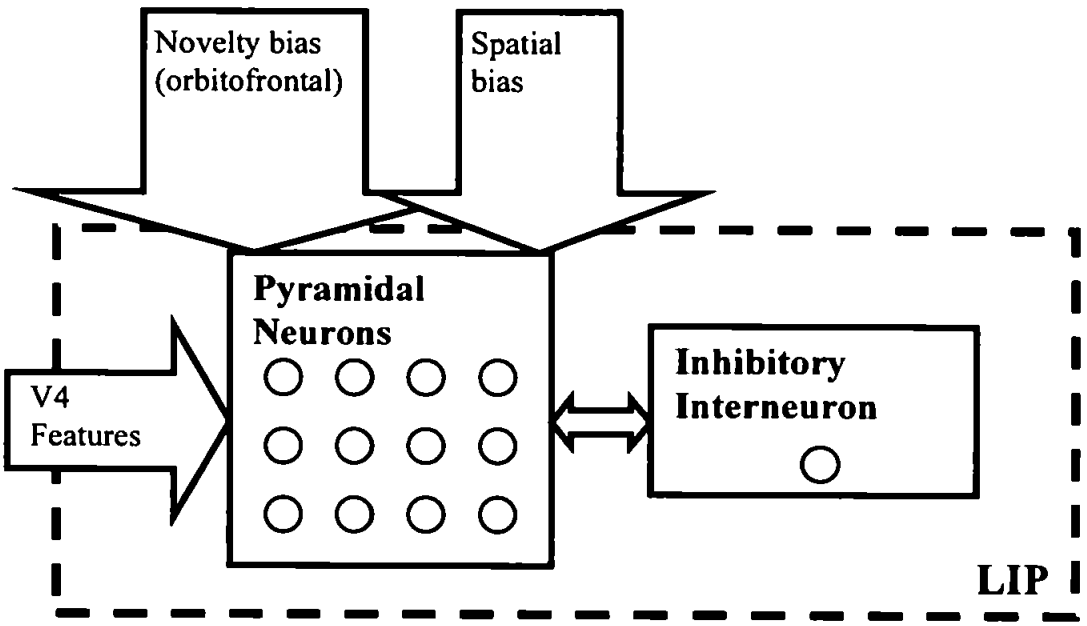



**Figure 2.3 Competition in LIP**

The inhibition and biases influencing competition in the LIP module are shown. LIP represents retinotopic locations competing with one another to attract attention. Locations in LIP receive a spatio-featural input from V4 and are subject to two spatial biases relating to the attention window and the novelty of the location in the scene. Competition between retinotopic locations is mediated by a single inhibitory interneuron assembly. The number of pyramidal assemblies in LIP varies according to the retinal image size, which can be adjusted. The normal size of the matrix of pyramidal assemblies in LIP is 20x20 = 400 assemblies.

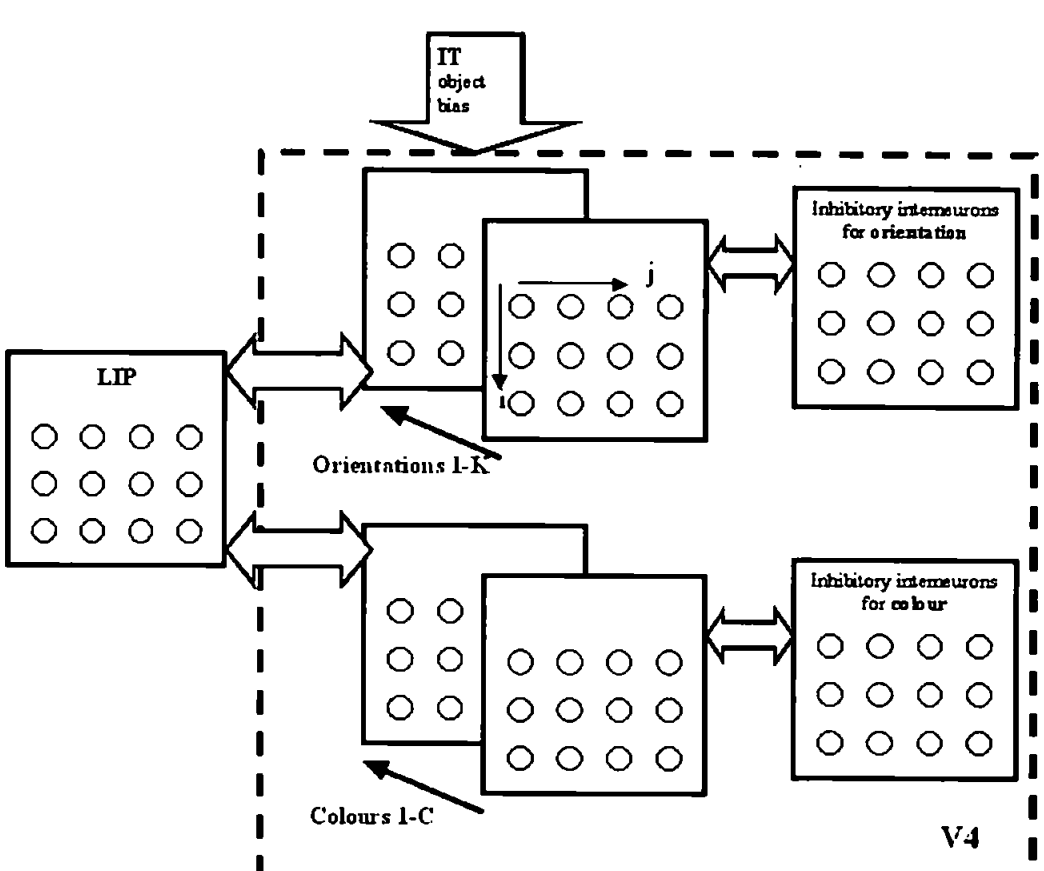



**Figure 2.4 Competition in V4**

The inhibition and biases that influence competition in the V4 module are shown. V4 encodes feature representations relating to colour and orientation (although, in principle, other features could be represented). These representations are activated by bottom-up inputs from V1 over the receptive field of each V4 assembly. V4 contains a retinotopically arranged layer (or map) of pyramidal assemblies for each feature and a retinotopically arranged set of inhibitory interneuron assemblies for each feature *type*. The size of each of these matrices is flexible in the same way as that in LIP, being normally set to 20x20 = 400 assemblies in each feature map and in each inhibitory interneuron layer. The inhibitory interneurons mediate local competition between different features of the same type at the same retinotopic location. This competition is biased by object-based feedback from IT and a spatial bias from LIP.

In order to provide "bottom-up" featural input to the dynamic portion of the system, retinal and V1 modules are simulated but do not take part in the dynamic interactions. These "pre-processing" stages are shown in orange in figure 2.1. Both orientation and colour features are processed within the retina, V1 and V4. These features are detected in the retina and V1 and are fed forward to the model's ventral stream hierarchy. To some extent, the V4 module could be viewed as incorporating early to intermediate ventral stream processing and representing areas V1, V2 and V4. V1 feeds forward to V4 via V2, but also has direct projections to V4 (Nakamura, Gattass, Desimone & Ungerleider, 1993).

The model could be extended to include V1 in the dynamic portion of the system in order to show possible attentional effects developing therein. Although such effects in V1 have been a matter of debate, it now seems clear that attention can affect V1 responses, at least in cluttered scenes (e.g. Brefczynski & DeYoe, 1999; Ghandi, Heeger, & Boynton, 1999; Motter, 1993; Roelfsema et al., 1998; Somers, Dale, Seiffert, & Tootell, 1999). ERP recordings suggest that attention does not modulate the earliest visual responses in V1 but can affect later responses, possibly as a result of feedback from other areas (Clark & Hillyard, 1996; Martinez et al., 1999). However, V1 is currently excluded from the dynamic portion of the system for reasons of computational simplicity[1], speed and to maintain the focus on intermediate stage of visual processing.

[1] There are three interacting dynamic modules (IT, V4 and LIP) and, within each of these, there are numerous cell assemblies. Therefore, the dynamics within the system are quite complex and it was important to make these stable before adding additional components. However, the model now has the potential to be extended.

A key suggestion in the model is that the development of object-based attention in the ventral stream (area V4), which highlights target features across the visual field (McAdams & Maunsell, 2000; Motter 1994 a, b; Saenz et al., 2002), guides the representation of behaviourally relevant locations in the dorsal stream (area LIP). LIP is then used to determine the location of the next fixation. The concept of a saliency map to accomplish pre-attentive selection of locations to be attended was introduced by Koch and Ullman (1985). Itti and Koch (2000) provided a very influential implementation of a pre-attentive saliency map, which embodied the principles suggested by Koch and Ullman. As discussed above, there is no clear-cut distinction between pre-attentive and attentive stages in the model here. The result of processing in the LIP module here produces a retinotopic map, which is similar in concept to a saliency map. However, the representation of salience here is related to behavioural relevance and it develops as attentional effects emerge rather than being simply a representation of the bottom-up stimulus salience. Thus, an early representation of all salient stimuli in the scene transforms into a representation linked to behavioural relevance as object-based attentional effects emerge from the system dynamics.

#### 2.3.2.1 The Ventral Stream

The ventral stream (V1, V4 and IT) operates as a feature/object processing hierarchy with receptive field sizes increasing towards IT and encoded features becoming more complex. Object-based competition operates within the ventral

stream. Features compete for similarity to the target object, information about which is assumed to be held in short-term memory in prefrontal cortex. The V1 and V4 modules encode colour and orientation features in a retinotopic manner but representations in IT are invariant to position. Neuronal properties *in vivo* along the ventral occipital-temporal pathway progress from sensitivity to local object features to invariant holistic representations (Lerner, Hendler, Ben-Bashat, Harel & Malach, 2001). Anterior regions in the temporal lobe (IT) are activated during the processing of intact, but not scrambled, faces (Ungerleider, 1995; Lerner et al., 2001). Thus, during the course of processing in the hierarchy from V1 to IT, cortical representation evolves from simple to complex features and then invariant object representation. It is necessary for cortex to transform the representation of the scene into such invariant object-based cell responses in order to limit the amount of information projected onwards into structures such as the amygdala and hippocampus. These memory systems can then remember or form associations about *objects*. Some neurons in the primate hippocampus, for example, respond to a particular objects wherever it is presented in space (Rolls, Xiang & Franco, 2005). In order to provide a full episodic memory system, it appears that the hippocampus provides separate representations of objects and space in addition to combined representations, which indicate where an object was seen (Rolls et al., 2005).

Features detected in the V1 module are fed forward to V4 in a convergent manner over the area of the V4 receptive field, which is larger than the receptive field size

in V1: The size of filters used in V1 determines the ratio of pixels to degrees of visual angle so that V1 receptive fields cover approximately 1° of visual angle (Wallis & Rolls, 1997). Throughout the model, biologically plausible receptive field sizes and characteristics are used. These will be described in more detail with justification during the formal definition of the model in the next two chapters.

Cells in monkey area V4 are sensitive to colour (Schein and Desimone, 1990; Zeki, 1993) and form, such as orientation, spatial frequency (Desimone & Schein 1987, Merigan 1996), disparity (Hinkle and Connor, 2001), and contour features (Pasupathy & Connor 1999). A few are selective for direction of motion (Desimone & Schein 1987). Interestingly, when V4 is lesioned, detection of low-salience stimuli is poor and the filtering of distracters is less efficient so that attention is captured by stimuli with strong bottom-up salience regardless of behavioural relevance (Schiller and Lee, 1991; De Weerd, Peralta, Desimone & Ungerleider, 1999). V4 neurons have been shown to be modulated by behavioural requirement (Haenny & Schiller, 1988) and by attention to a feature where the cue was not presented visually, suggesting a top-down source of the modulation (orientation: Haenny, Maunsell & Schiller, 1988; colour/luminance: Motter 1994b). When multiple objects are presented in a cell's receptive field, competition in V4 results in the response being determined by the attended object (Moran & Desimone 1985; Luck et al., 1997; Chelazzi et al., 2001).

Representation of selectivity in the model's V4 is simplified such that the V4 module encodes colour and orientation in separate feature layers, or maps. Competition, shown in figure 2.4 and described in detail in chapter 4, operates locally between different colours and different orientations at the same retinotopic location. This provides competition between features within the same receptive field, as has been found in single cell recordings (Chelazzi et al., 2001; Luck et al., 1997; Moran & Desimone, 1985). In addition to the bottom-up featural information from V1, competition in V4 is subject to two external biases: A spatial bias from LIP and an object-related bias fed back from IT. V4 features project forward to activate invariant object representations in the IT module.

The IT module represents encoding in anterior areas of IT, such as area TE, where receptive fields span much of the retinal image and populations encode objects in an invariant manner (Wallis & Rolls, 1997). Objects represented in the IT module compete with one another. In addition to the featural inputs from V4, IT is subject to a target object bias from prefrontal cortex. This acts to suppress non-target objects so that the target object is able, over time, to win the competition between objects in IT.

Stimuli, chosen to match those used by Motter and Belky (1998b), are vertical and horizontal, red and green bars of size 1° x 0.25° (a consistent correspondence of degrees of visual angle with pixels is used within the system and creation of these stimuli is described in appendix A1.1), with distractors differing from the target in

one feature dimension only. These simple stimuli allow the principles of the model to be demonstrated in a manner that is reasonably computationally straightforward. Many other models, such as the Selective Attention for Identification Model (SAIM: Heinke & Humphreys, 2003), also tend to use simple stimuli similar to those used in most psychophysical experiments. Chelazzi et al. (1993, 1998, 2001), whose single cell data are replicated later, used more complex stimuli. However, the principles of detection of features, feedforward connectivity through the ventral stream to invariant object representations in IT, and feedback to target features are the same for simple and complex objects.

#### 2.3.2.2 The Dorsal Stream

The model's dorsal stream contains only one module representing posterior parietal cortex, specifically area LIP. LIP encodes both spatial and colour featural information (Toth & Assad, 2002) when the information is behaviourally relevant (Gottlieb et al., 1998). It also has some shape sensitivity (Sereno & Maunsell, 1995). LIP has reciprocal connections with both V4 and other parietal areas, such as 7a, which are important in spatial analysis.

As discussed above, LIP is known to encode the locations of behaviourally relevant features (Colby et al., 1996) and may act as a saliency map (Colby & Goldberg, 1999; Gottlieb et al., 1998; Kusunoki et al., 2000). Responses in LIP reflect attentional priority rather than purely signalling a saccade target (Bisley &

Goldberg, 2003) and, therefore, are not dependant on motor response (Bushnell, Goldberg, & Robinson, 1981). However, LIP has projections to superior colliculus (Lynch, Graybiel & Lobeck, 1985) and FEF (Blatt, Andersen & Stoner, 1990), which are thought to be involved in generating saccades. Therefore, LIP is ideally suited to form part of the cortical network responsible for selecting possible targets for saccades, and direct electrical stimulation of the area has elicited saccades (Their & Andersen, 1998). Here, it is assumed that outputs from LIP would influence the selection of saccade trajectory in areas, such as superior colliculus and FEF, to which it projects. Assemblies within the LIP module compete and the location represented by the assembly of highest activity at the end of processing at each fixation is selected as the target for the next saccade.

Connections between the ventral and dorsal streams form an important component of the model. There is a reciprocal connection between V4 and LIP, which directly connects assemblies at the same retinotopic location. This contrasts with Deco's model, which uses a Gaussian kernel for connection strengths. The simpler form of connection is designed to test the principle without additional complication. The use of these connections to provide spatial and object-based biases within the system is now described. A spatial bias provides enhancement of activity at the location to which the eyes have just moved. An object-based bias, relating to the search target, is not spatially specific but operates in parallel across the retinal view. These two biases produce effects that are designed to investigate the first research goal above, i.e. that these two effects coexist such that an early spatial

effect can facilitate, but not constrain, the development of object-based effects that are occurring in parallel. Of specific importance in driving the scan path is the connection from the ventral to the dorsal stream (the second research goal), which is able to guide the search process towards behaviourally relevant locations.

### 2.3.3 Spatial Attention Window

Within each retinal image at every fixation an "Attention Window" (AW) is formed. This window is initially a spatial spotlight but, over time, becomes object-based. Thus, the initial spatial aperture of the AW primes development of object-based attention, which occurs in parallel, and object-based attention responses are strongest within the AW causing a combined spatial and object-based attentional modulation, as found by McAdams and Maunsell (2000) and Treue and Martinez Trujillo (1999).

The aperture of the initial window is scaled according to coarse resolution information, reflecting stimulus density, which is assumed to be conveyed rapidly by the magnocellular pathway to parietal cortex, including LIP, and other possibly involved areas, such as FEF. This information is shown as the magnocellular output from V1 in figure 2.1. The serial deployment of attention for conjunction search appears to require the magnocellular pathway because search reaction times are much slower under isoluminant conditions where stimuli are defined by colour difference only, i.e. the information is not available to the magnocellular pathway

(Cheng, Eysel & Vidyasagar, 2004). This magnocellular input could provide fast information, at a coarse spatial resolution, which could be used to scale the initial focus of attention. The spatial resolution of attention has been suggested to be much coarser than that available for visual perception (Intriligator & Cavanagh 2001). The algorithm used to scale the AW will be formally defined in chapter 4. All other information within the system is assumed to be derived from the parvocellular pathway. The koniocellular pathway is not specifically addressed, but may also contribute to some of the functions attributed to the parvocellular, and possibly magnocellular, pathways here.

During a scan path, the size of the AW is dynamic, being scaled according to local stimulus density found around any particular fixation point (Lanyon & Denham, 2004b, 2005a,b). In principle, the distance over which targets are most likely to be identified is similarly scaled. Motter & Belky (1998a) found that, during overt conjunction search (but not feature search), objects could be identified only within an "*area of conspicuity*" surrounding a monkey's fixation point. The size of this area was dependant on stimulus density so that it effectively zoomed in and out (see also the "zoom-lens" model of attention: Eriksen & St James 1986) and only stimuli within a distance of 2 average-nearest-neighbour-distances (ANNDs) were processed. Here, this "area of conspicuity" is interpreted as the maximum spatial aperture of the initial attention window, which follows the allocation of attention to a location with the eye movement to that location (Hoffman & Subramiam, 1995; Shepherd, Findlay & Hockey, 1986).

The spatial AW is implemented as a bias applied to LIP, creating an early anticipatory spatial enhancement, as seen when a stimulus is expected at a location in single cell studies (Colby et al., 1996) and imaging of the intraparietal sulcus (IPS), which is the possible human parietal homologue of LIP (Corbetta et al., 2000; Hopfinger et al., 2000; Kastner et al., 1999). This is also consistent with psychophysical findings (Posner, Walker, Friedrich, & Rafal, 1984; Shimozaki, Hayhoe, Zelinski, Weinstein, Merigan & Ballard, 2003), single cell recordings (Robinson, Bowman, & Kertzman, 1995), positron emission tomography (PET) results (Corbetta, Shulman, Miezin & Petersen, 1995), fMRI (Giesbrecht, Woldorff, Song & Mangun, 2003) and event-related potential (ERP) studies (Martinez et al., 1999) that suggest a parietal control signal for spatial attention. LIP provides a bias to V4 that results in a spatial attentional enhancement in V4, as seen in single cell recordings in this area (Luck et al., 1997) and imaging of human extrastriate cortex (Hopfinger et al., 2000; Kastner et al., 1999).

Although it is suggested that LIP receives this spatial signal very rapidly, possibly via a magnocellular route, the exact source of the spatial bias is uncertain. A plausible source of the bias to LIP is FEF because microstimulation of FEF produces a spatial enhancement of signals in V4 (Moore & Armstrong, 2003) and the suggested circuitry through LIP offers one possible route to produce this effect in V4 when an eye movement is made. It is important to note that FEF is also able to represent stimulus salience (Bichot & Schall, 1999; Schall, 2002) and, therefore,

may contribute to other functions represented by the LIP module here. Other cortical and sub-cortical areas offer possible sources of the spatial bias to LIP. Dorsolateral prefrontal cortex has extensive connections with parietal cortex (Blatt, Andersen, & Stoner, 1990) and is a possible source of a short-term mnemonic spatial bias. FMRI recordings following a spatial cue (Corbetta et al., 2000; Hopfinger et al., 2000) show activity in prefrontal cortex in addition to IPS, although the prefrontal activation appears to be more transient (Corbetta et al., 2000). This information also could be conveyed sub-cortically, and superior colliculus or the pulvinar nucleus of the thalamus could be the source of this spatial effect in LIP and V4. The pulvinar receives retinal input and is thought to be involved in activating neurons in both V4 and parietal areas (Mountcastle, Motter, Steinmetz & Sestokas, 1987; Adams, Hof, Gattass, Webster & Ungerleider, 2000). Alternatively, the calculation of the size of the spatial attentional focus could be performed within parietal cortex, on the basis of the low-resolution information about stimulus density received rapidly from the magnocellular pathway. This would then enable an area such as LIP to project a varying size "spotlight" onto ventral stream areas, such as V4.

The spatial bias from LIP to V4 also allows feature binding to take place at the spatial resolution of the V4 receptive field, in a manner similar to that proposed by Feature Integration Theory (Treisman, 1982; Treisman & Gelade, 1980), but within a parallel feature-processing environment that is subject to this spatial bias. As a V4 cell assembly encoding the target colour becomes active, this retinotopic

location becomes active in LIP, due to the excitatory bias from V4 to LIP. Reciprocal connections from LIP bias both the colour and orientation layers in V4 and this results in the activity of the V4 orientation selective cell assembly at the target coloured location tending to be enhanced in parallel with the colour selective assembly. In humans, PET measures show that parietal cortex is active during spatial attention shifts and for feature conjunction tasks (Corbetta et al., 1995). FMRI has shown that parietal cortex is involved in both spatial attention and feature binding tasks (Shafritz, Gore & Marois, 2002). Furthermore, parietal damage is linked to the inability to bind form with surface features, such as colour, but does not affect the binding of form into shape (Friedman-Hill, Robertson & Treisman, 1995; Humphreys, Cinel, Wolfe, Olson & Klempen, 2000). Thus, parietal cortex could be instrumental in binding only when it occurs across different feature dimensions in the ventral stream. Similar in principle to that demonstrated in imaging (Lerner et al., 2001) and modelled elsewhere (e.g. Riesenhuber & Poggio, 1999, 2000; Wallis and Rolls, 1997), binding of form into shape is achieved within the ventral stream hierarchy of the model here. Therefore, disruption to the parietal portion of the model would affect binding across feature dimensions but would not affect the binding of form into shape, similar to the effect observed in parietal patients (Friedman-Hill et al., 1995; Humphreys et al., 2000). The issue of feature binding is not investigated in any further detail until the discussion chapter but this possible role for a "cross-stream" signal from parietal cortex is simply highlighted.

### 2.3.4 Object-Based Attention

The source of object-based attention in the model is an object-related bias from prefrontal cortex applied to IT. This biases the competition between objects in IT in favour of the target object such that this object wins the competition over time. Attention appears to require working memory (de Frockert, Rees, Frith & Lavie, 2001) and, due to its sustained activity, prefrontal cortex has been suggested as the source of a working memory object-related bias to IT. Whereas IT responses are disrupted by incoming sensory information during visual search, prefrontal cortex is able to represent the target object in a more sustained manner (Miller, Erickson & Desimone, 1996). "What" and "where" information remains anatomically segregated in prefrontal cortex[2]. However, these neurons appear able to integrate spatial and object-based information (Rao, Rainer & Miller, 1997) and may adapt their responses depending on the behavioural requirement (Duncan, 2001: *Adaptive Coding Model*). Prior knowledge of a target object, stored in working memory, influences visual search by making it faster (Vickery et al., 2005) and improving accuracy compared to that when the target is specified by intrinsic display properties (Findlay, 1997). Single cell recordings using split-brain paradigms support the existence of a top-down bias from prefrontal cortex to IT (Tomita, Ohbayashi, Nakahara, Hasegawa & Miyashita, 1999) and other computational

[2] Posterior parietal areas of the dorsal stream project to the principle sulcus of the dorsolateral prefrontal cortex, to areas 9 and 46. However, the ventral stream projects to the inferior convexity in ventrolateral prefrontal cortex so that ventrolateral areas 12 and 45 receive input from IT areas TEO and TE and have similar response properties to those found in IT (Logothetis, 1998).

models have implemented such a bias (e.g. Deco, 2001; Deco & Lee, 2002; Usher & Niebur, 1996).

External mnemonic prefrontal biases are normally modelled as a static value over time (e.g. Deco, 2001; Deco & Lee, 2002; Usher & Niebur, 1996). However, many prefrontal neurons do not display static activity when maintaining memory during a delay period. Such delay-period activity is temporally irregular (Compte, Constantinidis, Tegner, Raghavachari, Chafee, Goldman-Rakic & Wang, 2003) and possibly dependant on previous experience of the object retained in working memory (Rainer & Miller, 2002). Many neurons, known as "late" neurons have increased activity during the late portion of a delay period (Rainer & Miller, 2002; Romo, Brody, Hernandez & Lemus, 1999), i.e. the activity builds during the delay period. Significantly, modelling the prefrontal bias as a static value tends to produce an early target object effect in IT, developing strongly from the onset of the sensory response (e.g. Deco, 2001; Deco & Lee, 2002). However, as previously discussed, significant target object effects are not normally found in single cell recordings until later in the IT response, for example from ~150-200ms after the onset of a search array (Chelazzi et al., 1993, 1998). In order to reproduce the temporal aspects of this effect computationally, it is necessary that the strength of prefrontal feedback to IT builds over time or is delayed in its onset. Experiments using a split-brain procedure[3] have revealed that the latency of response in IT to top-down information from prefrontal cortex (178ms) is longer than that resulting

[3] The posterior corpus callosum was dissected leaving the anterior corpus callosum interconnecting the prefrontal cortices. Bottom-up responses in IT to contralaterally presented stimuli were compared with responses, to ipsilaterally presented stimuli, that were due to top-down prefrontal feedback.

from bottom-up information (Tomita et al., 1999). Also, a new visual cue takes up to ~200ms to become fully effective (presumably by entering working memory) as a target for search when compared to search performance in a blocked design when the target is already memorised (Wolfe, Horowitz, Kenner, Hyle & Vasan, 2004). Once a search array is presented, prefrontal cortex begins to receive feedforward information about objects in the display from IT and discriminates targets and non-targets (Everling, Tinsley, Gaffan & Duncan, 2002; Miller et al., 1996; Rainer, Asaad & Miller, 1998). Therefore, during the presentation of the search array, activity of relevant prefrontal neurons increases, in response to this stimulus information, beyond the level required to hold an object in memory. The latency of this prefrontal response varies across tasks and conditions: The match response discriminating a remembered target versus non-targets begins 110-120ms after stimulus onset (Everling et al., 2002; Miller et al., 1996) but identification of a target not already held in working memory begins approximately 135ms after onset of the search array (Hasegawa, Matsumoto & Mikami, 2000). Prefrontal cells also determine, from ~130ms after presentation of an object, whether it matches the category of another object remembered during a delay (Freedman, Riesenhuber, Poggio & Miller, 2003).

In view of this latency in the prefrontal sensory response, it was decided to incorporate a similar latency in the modelled prefrontal bias. Therefore, the prefrontal signal is modelled as a sigmoid function over time (a function commonly used to model neural responses), so that the strength of the prefrontal bias to IT

builds over a biologically plausible time from fixation. Such a function represents prefrontal activity in response to the incoming stimulus information and its match with the target. The bias acts to inhibit non-target objects in IT, but is also effective if modelled as an excitatory bias to the target object. The sigmoidal bias enables the simulation of the target effect in the late (from ~150ms post-stimulus) portion of the sensory response in IT (Chelazzi et al., 1993, 1998) and V4 (Chelazzi et al., 2001). The effect of different forms of prefrontal feedback to IT will be discussed in chapter 5, where results from the simulation of single cell recordings in IT are presented.

Responses in area V4 are known to be modulated by both attention to space and attention to features (McAdams and Maunsell, 2000). In the model, area V4 is subject to both a spatial bias, from LIP, and an object-related bias, from IT. Both inputs are applied to V4 throughout the temporal processing at each fixation. However, the spatial bias results in an earlier spatial attention effect in V4 than the object effect because the latter is subject to the resolution of competition between objects in IT. The object-related bias from IT to V4 results in non-target features being suppressed allowing target features to win featural competition. This process occurs in parallel across V4. Such target object effects have been seen from ~150ms in monkey single cell studies of V4 (Chelazzi et al., 2001; Motter, 1994 a, b), and have also been observed in human imaging (Chawla et al., 1999). Although not modelled here, prefrontal cortex could provide another source of direct target

feature bias to V4, for example "priming" a particular colour so that it becomes enhanced in V4 even in the absence of visual stimulation (Chawla et al., 1999).

### 2.3.5 Guidance of the Scan Path

In order to carry out effective visual search, areas of the brain involved in the selection of possible target locations should be aware of object-based effects taking place in the retinotopic visual areas because these effects will be suppressing non-target features across the visual scene. An important suggestion here is that object-based effects occurring in area V4 are able to influence the spatial competition for next fixation location in LIP. It is this *cross-stream* interaction between the ventral and dorsal visual streams that allows visual search to select appropriate stimuli for examination. Thus, object-based attention, at least at the featural level, may be crucial for efficient search.

Once object-based attention effects become significant in the modelled V4, locations containing the target colour or orientation are "highlighted" across the scene, as has been found in single cell recordings in the area (McAdams & Maunsell, 2000; Motter 1994a,b; Saenz et al., 2002) and non-target representations are suppressed. V4 provides a spatio-featural bias to the spatial competition in LIP that enables it to represent the locations of the behaviourally relevant target features most strongly and drive the scan path towards them. Thus, the outputs of object-based attention in the ventral stream are used to "guide" the scan path to possible

target locations. This is similar in concept to the "Guided Search Model" (Wolfe, 1994; Wolfe, Cave & Franzel, 1989) but does not require the conjunction of features (although this happens over time). Connections from the V4 colour features to LIP are slightly stronger than those from the orientation features. This allows a priority to be given to a feature type (colour) in attracting the scan path. The difference in connection strength need only be very marginal in order to achieve this effect (Lanyon & Denham, 2004b; 2005b), as will be shown in chapter 6. The strength of these connections could be adapted subject to cognitive requirement or stimulus-related factors, such as distractor ratios (Bacon & Egeth, 1997; Shen et al., 2000, 2003). However, with the proportion of distractor types equal, Motter and Belky (1998b) found that orientation was unable to override colour even during an orientation discrimination task. This suggests that the bias towards a stronger colour connection could be learnt during development and be less malleable to task requirement. The V4 to LIP connection provides one (not exclusive) method by which a feature priority may be implemented. Such a cross-stream connection could be relevant in the case of a feature like colour that seems to have a strong influence on search under various conditions. Further bottom-up saliency effects that could influence feature priorities are not examined in the current model. However, the competition for attention in LIP is extendable to include other factors, for example the receipt of bottom-up features direct from V1. These issues will be returned to in the discussion in chapter 8.

Hence, the dorsal stream of the model (LIP) acts as a featural saliency map (Koch & Ullman, 1985) that is guided by object-based attention in the ventral stream. Evidence from PET measures of regional cerebral blood flow and fMRI indicate that parietal cortex is involved in object-based attention as well as spatial attention (Fink et al., 1997; Serences et al., 2004), and damage to parietal cortex affects the ability to use featural information to guide search (Shimozaki, Hayhoe, Zelinsky, Weinstein, Merigan. & Ballard, 2003). Further evidence for this area being involved in non-spatial attention comes from human imaging of the IPS area (a possible homologue of LIP), which has been shown to become active during shifts in an object feature, such as colour or shape (Le, Pardo, & Hu, 1998), when attending to the direction of motion before target onset (Shulman, Ollinger, Akbudak, Conturo, Snyder, Petersen & Corbetta, 1999) or between percepts during binocular rivalry (Lumer, Friston & Rees, 1998).

This "cross stream" interaction provides a very effective method of guiding search towards behaviourally relevant locations on the basis of bottom-up stimulus information and top-down attentional guidance. In order to prevent revisiting locations, an *inhibition of return* mechanism is required. This is implemented as a further bias to the spatial competition in LIP and raises interesting issues of non-retinotopic memory for locations previously fixated. This topic is addressed in chapter 4.

## 2.4. Chapter Summary

This is the first biased competition model to adopt an active vision approach where the retinal view changes with shifts in attention. It is also the first to address in detail the timing of spatial and object-based effects at the cellular level. These effects, in particular the object-based effects, are used to produce search scan path behaviour at the systems level of the model. At this level, behaviour results from interactions between the modules in the system. The outcome of object-based processing in V4 is used to bias spatial competition in LIP so that attention is attracted to behaviourally relevant locations within the current retinal view.

At the start of this chapter two research goals and supporting data from the literature were presented. As some of this data will be replicated in later chapters, the goals are re-stated and re-phrased in more concrete terms here, along with a summary of how key points from the literature will be replicated. The contribution that the current work makes, at a theoretical level (results will be discussed in later chapters), is also summarised.

| *Research Goal 1* | **Spatial Attention, biased from LIP, appears earlier in V4 than object-based attention, biased from IT** |
|---|---|
| *Experimental Evidence* | *Model Performance* |
| **Colby et al. (1996)** | The activity of cells assemblies in LIP whose receptive fields overlap the initial spatial focus of attention will be enhanced, due to the requirement to attend at fixation, preceding and during the stimulus-related response. |
| **Luck et al. (1997)** **(Hopfinger et al., 2000)** | As a result of spatial attention, the activity of V4 cell assemblies whose receptive fields overlap the initial spatial focus of attention will be selectively enhanced at an early stage, even preceding the onset of stimuli. |
| **Motter (1994a,b)** | Cell assemblies selective for target features will be enhanced, and those selective for non-target features suppressed, across V4 beginning ~150ms into the sensory response. |
| **Chelazzi et al. (1993, 1998, 2001)** | Object-based attention will lead to the responses of V4 and IT cell assemblies being determined by the target object. This should be significant only in the late time window beginning from ~150ms post-stimulus. |

| | |
|---|---|
| **Contribution of work:** | Apart from the seminal contribution from Deco (2001; Deco & Lee, 2002, Rolls & Deco, 2002), few computational models of both spatial and object-based attention currently exist. This model is the first in which spatial and object-based attention develop concurrently such that an early spatial attention window facilitates object-based attention. Prior to the work described here, there was no specific simulation of both spatial and object-based attention with different time courses for each. This is the first time that the time course of object-based affects at the cellular level in V4 and IT have been replicated in detail. |

| ***Research Goal 2*** | **The enhancement of target features across V4 as a result of object-based attention is able to bias LIP to represent behaviourally relevant locations and attract the scan path.** |
|---|---|
| *Experimental Evidence* | *Model Performance* |
| **Gottlieb et al. (1998), Colby et al. (1996)** | The activity of LIP cell assemblies whose receptive fields overlap with the locations of behaviourally relevant (target object) features will be increased. One possible route is suggested to enable LIP to be an integrator of spatial and |

| | |
|---|---|
| | featural information. |
| **Motter & Belky (1998b)** | Scan path simulations will qualitatively replicate those seen by Motter and Belky (1998b): Most fixations will land near target coloured objects, some non-target colour distractors may also be able to capture attention and occasionally fixations will fall in blank areas of the scene. A quantitative analysis will be performed to examine the model's behaviour in different circumstances |
| | |
| **Contribution of work:** | The proposal that locations containing the target colour are enhanced in parallel in area V4 and that this information is conveyed to LIP is a key proposal. Thus, cross-stream interaction from the ventral to dorsal pathway is an important part of the model to guide active visual search scan paths.<br><br>Many models, such as Hamker (1998), Niebur et al. (2001) and Deco & Lee (2001) attract attention at least partly in a "bottom-up" fashion on the basis of incoming featural information. However, colour, luminance and orientation features are normally treated equally when forming a saliency |

map or equivalent device and none give a special bias to colour in attracting attention. There is evidence (e.g. Motter & Belky 1998b; Lauria & Strauss, 1975; Scialfa & Joffe, 1998; Williams, 1967; Williams & Reingold, 2001) to suggest that surface features such as colour play a special role in attracting the attentive scan path during active visual search. In this model, featural information encoded in the ventral stream (in V4) is integrated with saliency map information in the dorsal stream (in LIP) such that colour is given precedence over orientation and has the strongest effect in driving the active vision scan path. This provides one method of implementing a feature priority in the guidance of the scan path. However, it is not suggested that this is the only method by which such effects could be achieved. Further bottom-up saliency factors may also influence search and such factors are discussed in relation to this model in the final discussion chapter.

The model is designed to simulate monkey and human data at two levels:

- **At the cellular level**, the activity of individual cell assemblies within the model will monitored over time during the course of one fixation in order to simulate recordings of individual neurons in awake behaving monkeys.
- **At the systems level**, the collective system will produce behaviour that will simulate active visual search behaviour observed in humans and monkeys. During these simulations, fixation will shift around the scene, i.e. the model's retina will move to simulate active eye movements. The system will autonomously determine when to perform a saccade and find the location that will be the target, i.e. end point, for the saccade.

A formal description of the retina and V1 pre-processing stages is presented in the next chapter. Then, the formal description of the dynamic modules and associated external biases will be given in chapter 4. Results of simulation at both the cellular and systems levels will be presented in chapters 5 and 6 respectively.

# Chapter 3

# Description of the Model
# A – Retina & V1 Pre-Processing

In this and the following chapter, the computational model that forms the central component of this work will be described. The model was introduced in chapter 2 but is described more formally in this and the next chapter. Here, the retinal and V1 stages that form the pre-processing for the dynamics of the system are described. The dynamic modules are described in the next chapter.

Following the random creation of an image, as described in appendix A1.1, retinal and V1 processing is applied across the entire image as "pre-processing" for the subsequent dynamic active vision processing. The retina and V1 do not form part of the dynamic portion of the system, for reasons of computational simplicity and speed, and because the aim of this model is to focus on attention operating in extrastriate areas and beyond. Retinal processing here subsumes thalamic processing in the lateral geniculate nucleus (LGN), although it is recognised that this is a simplification. It is possible to apply the retinal and V1 processing across the whole image in advance of the active vision portion of the system, which moves the virtual retina around the scene, because there is no feedback to V1 or attempt to replicate attention in this area. Therefore, V1 is not involved in the dynamics of the system and its contribution remains constant over time. Hence, it is computationally efficient to pre-process the entire scene within the retina and V1 and then pass the required V1 inputs, i.e. those relating to the current retinal image, to V4 during the

active vision processing. Separation of the retinal and V1 processing allows the dynamic portion of the system to run faster and it may be run multiple times against the same "pre-processed" V1 output, thus generating many scan paths over the same image.

Although the dynamic portion of the model, described in the next chapter, adopts a similar computational approach to that described by Deco (2001, Deco & Lee, 2002; Rolls & Deco, 2002), the retina and V1 stages described here are substantially different to Deco's model. Modelling of the retina and V1 is inspired by Grossberg and Raizada (2000). This means that the method of feature detection here differs from Deco's model and does not adopt the use of Gabor jets (Lee, 1996) to detect orientations at different spatial frequencies. The normalisation of the initial image performed by Deco is not carried out here. Instead, V1 inputs to the dynamic portion of the model, specifically V4, are subject to normalisation as will be described in the next chapter.

## 3.1 Retina

Retinal ganglion cells and LGN neurons can be classified into: broad-band cells, concentric single-opponent cells, and co-extensive single-opponent cells:

- **Broad-band cells** are not colour-specific but respond only to brightness. They receive input from red and green cones and have a centre-surround antagonism.

Blue cones do not connect with broad-band cells and, therefore, short wavelengths are not used for the processing of form.

- **Concentric single-opponent cells** display not only the centre-surround antagonism but also antagonism between red and green cone inputs. They process information about both colour and achromatic brightness contrast.

- **Coextensive single-opponent cells** transmit information from the blue cones. They have a uniform receptive field without centre-surround antagonism. However, the inputs from the blue cones antagonizes the combined inputs from the red and green cones.

The stimuli to be processed here consist of red and green bars. For the purposes of this model, blue cone (short wavelength, S, cones) inputs will be ignored. Thus, colour processing focuses on the red-green channel and not the blue-yellow channel. The outputs of the red (long wavelength, L, cones) and green (medium wavelength, M, cones) cones are represented by the feature maps $\Gamma^{red}$ and $\Gamma^{green}$ respectively. This is a simplification in that the output for the non-preferred wavelength is assumed to be zero, whereas cones actually respond to some extent to non-preferred wavelengths because their output is dependant on photon absorption: An intense stimulus of a non-preferred wavelength could result in as much photon absorption as a less intense stimulus of the preferred wavelength. References to red and green throughout this thesis refer to long and short

wavelengths. The greyscale image, $\Gamma^{grey}$, used for form processing, is a composite of the colour arrays and provides luminance information.

### 3.1.1 Form processing in the Retina

At each location in the greyscale image, retinal ganglion broad-band cells perform simple centre-surround processing, according to Grossberg and Raizada (2000) as follows:

**On-centre, off-surround broadband cells:**

$$u^{+} = \Gamma_{ij}^{grey} - \sum_{pq} G_{pq}(i,j,\sigma_1)\Gamma_{pq}^{grey} \qquad [3.1]$$

**Off-centre, on-surround broadband cells:**

$$u^{-} = -\Gamma_{ij}^{grey} + \sum_{pq} G_{pq}(i,j,\sigma_1)\Gamma_{pq}^{grey} \qquad [3.2]$$

where $G_{pq}(i,j,\sigma_1)$ is a 2-dimensional Gaussian kernel, given by:

$$G_{pq}(i,j,\sigma_1) = \frac{1}{2\pi\sigma_1^2}\exp\left(-\frac{1}{2\sigma_1^2}\left((p-i)^2 + (q-j)^2\right)\right) \qquad [3.3]$$

The Gaussian width parameter is set to: $\sigma_1 = 1$

For the image in figure 3.1, the outputs from the retinal monochromatic processing are given in figures 3.2 and 3.3:

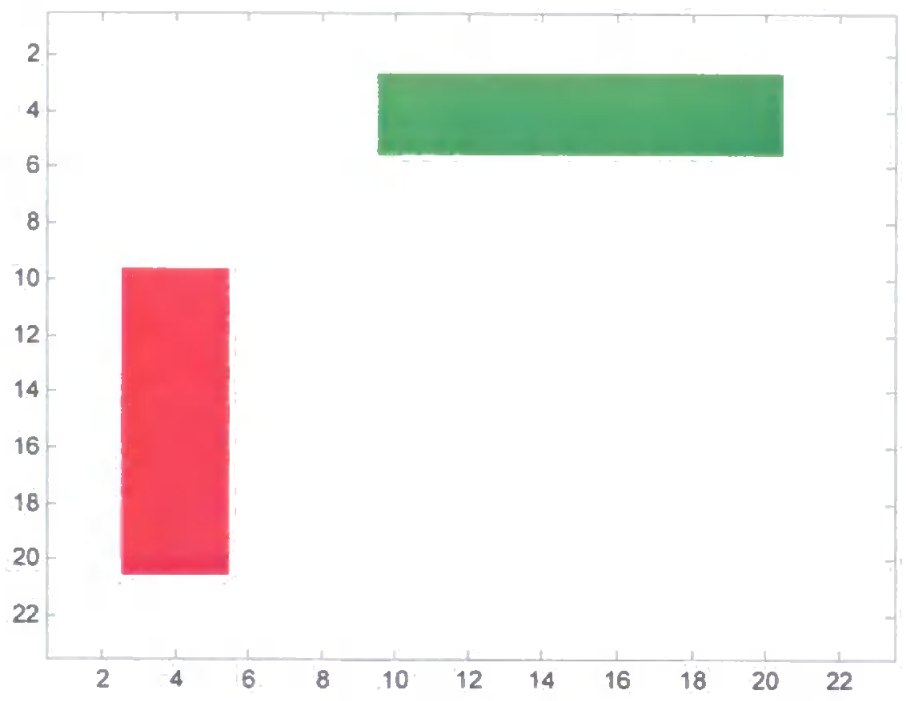



**Figure 3.1 Example Image with Red Vertical Bar and Green Horizontal Bar**

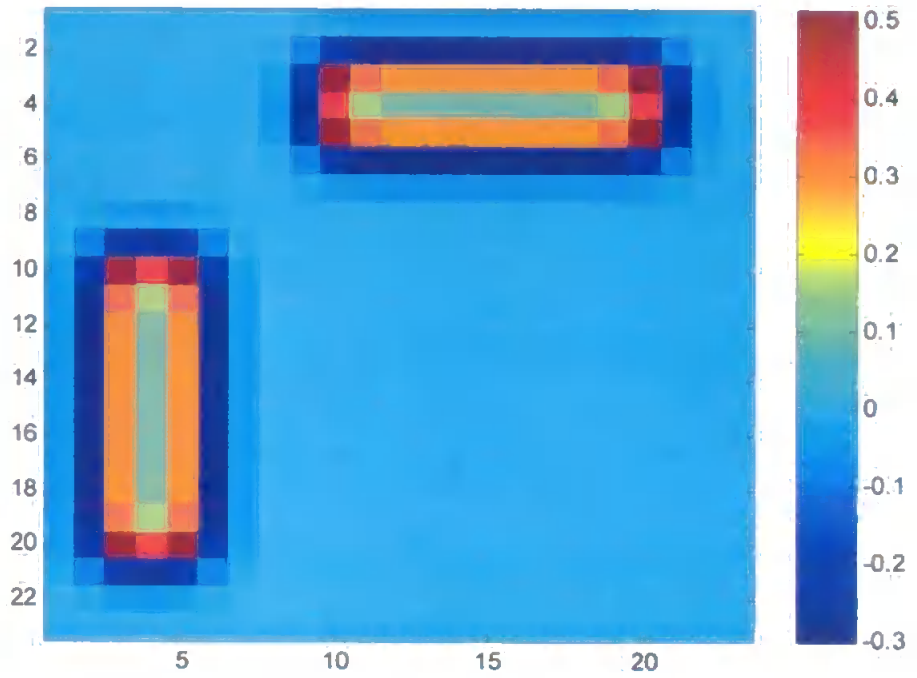



Figure 3.2 On-Centre Retinal Broad-Band Cells

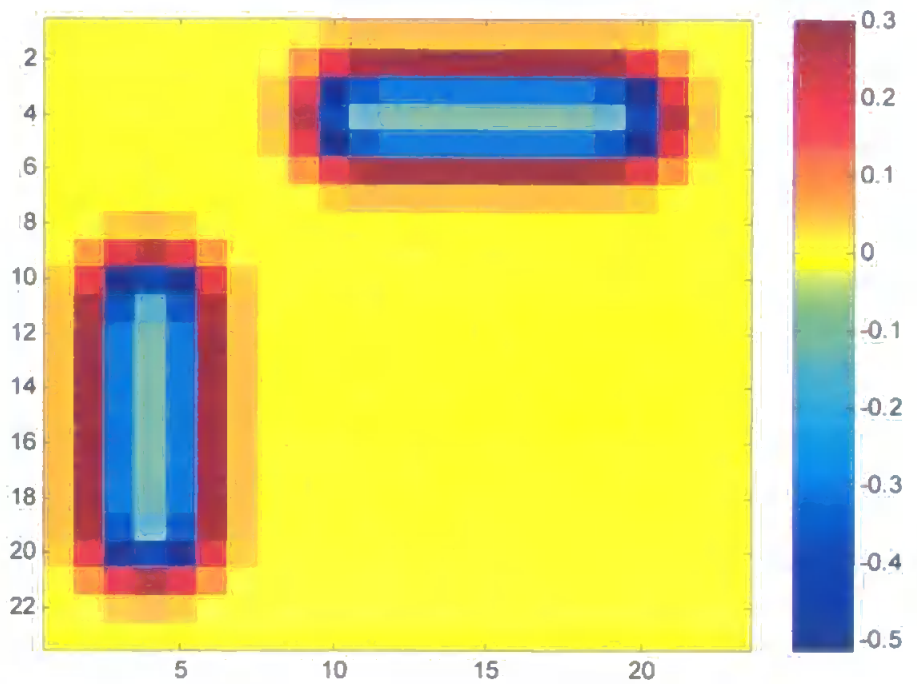



Figure 3.3 Off-Centre Retinal Broad-Band Cells

These broadband cells provide luminance inputs to V1 interblob simple cells that are orientation selective.

### 3.1.2 Colour processing in the Retina

Retinal concentric single opponent cells process colour information as described below. This model is based on the simplified description of colour processing presented by Rolls & Deco (2002) and is inspired by the model of orientation processing described above.

**Red on-centre, off-surround concentric single-opponent cells:**

$$v_{ij}^{redON} = \Gamma_{ij}^{red} - \sum_{pq} G_{pg}(i,j,\sigma_1)\Gamma_{pq}^{green} \qquad [3.4]$$

**Red off-centre, on-surround concentric single-opponent cells:**

$$v_{ij}^{redOFF} = -\Gamma_{ij}^{red} + \sum_{pq} G_{pg}(i,j,\sigma_1)\Gamma_{pq}^{green} \qquad [3.5]$$

**Green on-centre, off-surround concentric single-opponent cells:**

$$v_{ij}^{greenON} = \Gamma_{ij}^{green} - \sum_{pq} G_{pg}(i,j,\sigma_1)\Gamma_{pq}^{red} \qquad [3.6]$$

**Green off-centre, on-surround concentric single-opponent cells:**

$$v_{ij}^{greenOFF} = -\Gamma_{ij}^{green} + \sum_{pq} G_{pg}(i, j, \sigma_1)\Gamma_{pq}^{red} \qquad [3.7]$$

For the image shown in figure 3.1, the outputs from the retinal colour processing are given in figures 3.4 to 3.7:

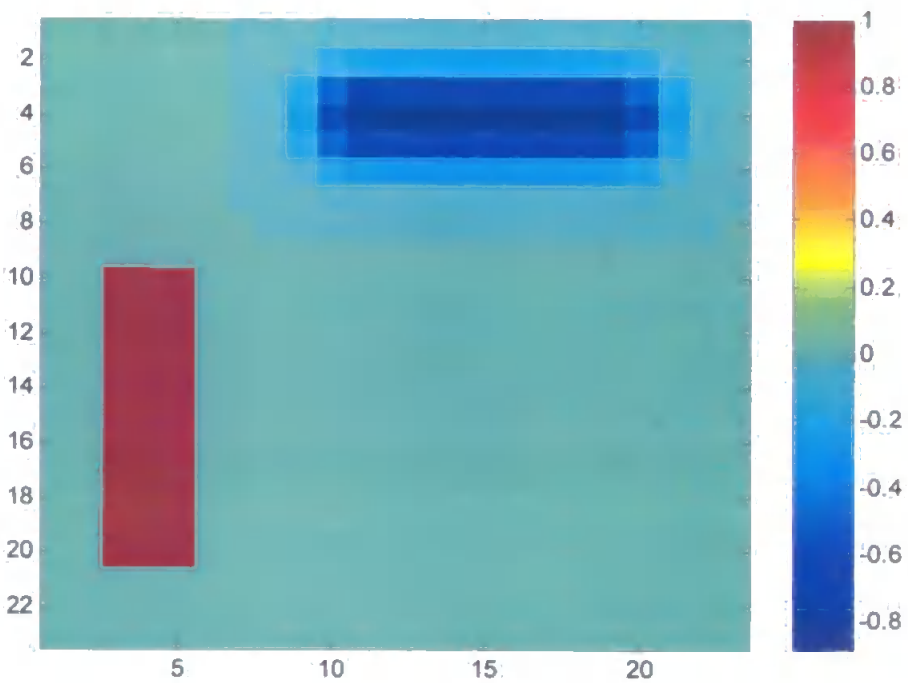



**Figure 3.4 Red On-Centre Retinal Single-Opponent Cells**

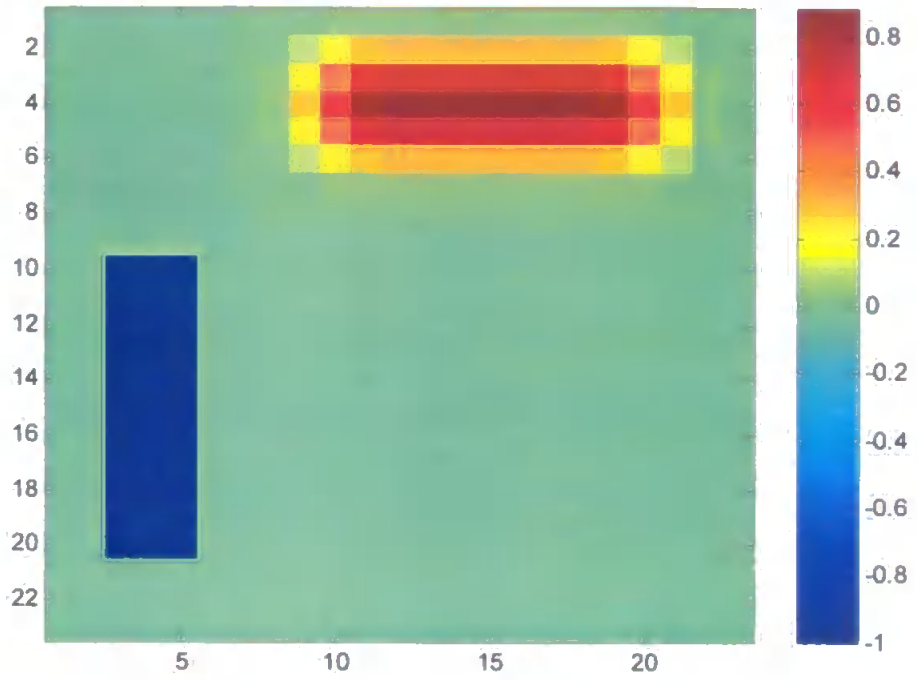



**Figure 3.5 Red Off-Centre Retinal Single-Opponent Cells**

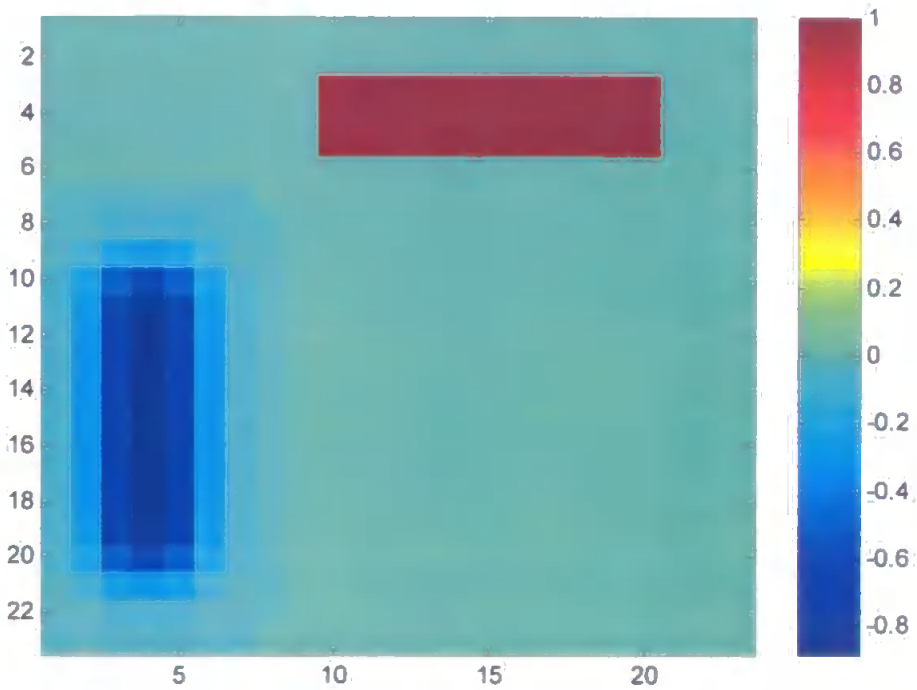



**Figure 3.6 Green On-Centre Retinal Single-Opponent Cells**

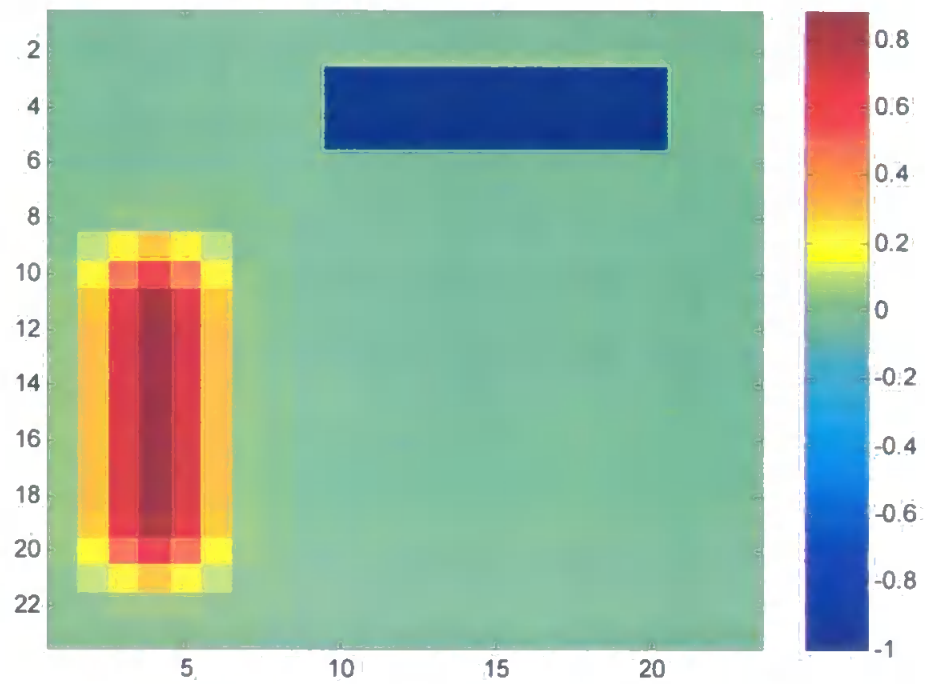



**Figure 3.7 Green Off-Centre Retinal Single-Opponent Cells**

These concentric single-opponent cells provide colour-specific inputs to V1 double-opponent blob neurons

## 3.2 V1

The purpose of V1 here is simply to provide featural input to V4. V1 interblob cells provide input to V4 orientation-selective cells and V1 blob cells provide input to V4 colour-selective cells.

The V1 module consists of K + C neurons at each location in the original image, so that neurons detect K orientations and C colours. K and C are normally set to 2 so that vertical and horizontal orientations, and red and green colours are processed.

### 3.2.1 Form Processing in V1

For orientation detection, V1 simple and complex cells are modelled. The receptive fields of simple cells can be modelled in a number of ways, for example difference-of-Gaussian (DOG, as used by Wallis & Rolls, 1997, in VisNet, for example), difference-of-offset-Gaussian (DOOG, as used by Grossberg & Raizada, 2000) or Gabor (as used by Deco & Lee, 2002, for example). Grossberg and Raizada's simple and complex cell model implementation of the DOOG was implemented and compared with an implementation of Gabor filters comprising a Gabor Jet described by Deco and Lee (2002). The comparison criteria included biological plausibility, processing speed and accuracy. Whilst, the Gabor filter fits neurophysiological data very well, particularly in relation to varying spatial frequencies (de Valois & de Valois, 1988), it is less accurate at orientation

detection than the DOOG, which is a precise edge detector. Gabor-filtered inputs to V4 led to the orthogonal orientation-selective assemblies becoming quite highly active. Due to the fact that the stimuli used here are simple, with orientation being one of their main characteristics, accurate orientation detection was deemed important and would have significant effect on the object-based feedback within the system. Thus, it was decided to develop the model using the DOOG filter, described by Grossberg and Raizada (2000). This model fits neurophysiological data about simple and complex cell response properties (see Grossberg & Raizada, 2000). In contrast to Grossberg and Raizada (2000), two spatial resolutions are calculated here. This forms a hierarchy of DOOG filters similar to the idea of the Gabor Jet (Lee, 1996). A Gabor Jet is a hierarchical series of Gabor filters at increasing octaves, i.e. increasing spread and decreasing spatial frequency, where the Jets for all orientations at a particular spatial location represent a hypercolumn in V1, if ocular dominance is ignored. The Gabor filters of different scales detect different spatial frequencies due to their excitatory and inhibitory lobes. However, the concept of varying spatial frequencies cannot be applied to the DOOG filters because they lack the lobes of the Gabor filter. The DOOG filters detect orientation inputs over varying spatial extents and, therefore, filter at different spatial resolutions. Therefore, the term spatial resolution is used here rather than spatial frequency.

Simple cells detect oriented edges using the difference-of-offset-Gaussian (DOOG) kernel (Grossberg and Raizada, 2000) as follows:

The right and left-hand kernels of the simple cells are given by:

$$R_{ijrk} = \Sigma_{pq}([u^{+}_{pq}]_{+} - [u^{-}_{pq}]_{+}) [D^{(lk)}_{pqij}]_{+} \quad [3.8]$$

$$L_{ijrk} = \Sigma_{pq}([u^{-}_{pq}]_{+} - [u^{+}_{pq}]_{+}) [-D^{(lk)}_{pqij}]_{+} \quad [3.9]$$

where:

$u^{+}$ and $u^{-}$ are the outputs of the retinal broadband cells above

$[x]_{+}$ signifies half-wave rectification, i.e.

$[x]_{+} =$ x if $x \geq 0$

0 otherwise

and the oriented DOOG filter $D^{(lk)}_{pqij}$ is given by:

$$D^{(lk)}_{pqij} = G_{pq}(i - \delta\cos\theta, j - \delta\sin\theta, \sigma_2) - G_{pq}(i + \delta\cos\theta, j + \delta\sin\theta, \sigma_2) \quad [3.10]$$

where:

$\delta = \sigma_2/2$ and $\theta = \pi(k-1)/K$, where k ranges from 1 to 2K, K being the total number of orientations (2 is used here).

$\sigma_2$ is the width parameter for the DOOG filter, set as below

r is the spatial resolution, such that

$r = 2$ and $\sigma_2 = 2.2$ for low resolution processing, used in the magnocellular (or sub-cortical) pathway for scaling the AW;

$r = 1$ and $\sigma_2 = 1.2$ for high resolution processing, used in the parvocellular pathway, which forms the remainder of the model. This results in the filter spanning 11 pixels. Throughout the model, this size is taken to represent 1° of visual angle because it is assumed that an average V1 cell's receptive field covers approximately 1° of visual angle (0.25-1° near the fovea, up to 1.3° in the periphery: Rolls and Deco, 2002). Therefore, the stimuli used in the original images were scaled accordingly, as described in appendix A1.1. (N.B. Most published models are do not overtly link pixels to

degrees of visual angle in order to create accurately sized stimuli. For example, implementation of Deco's (2001; Rolls & Deco, 2002) model results in a Gabor filter that extends 15 pixels for the first spatial frequency and 29 pixels for the second. If the filter size is taken to represent 1°, a typical V1 receptive field size, then the 2x4 pixel bar stimuli used to replicate the experimental data of Reynolds, Chelazzi and Desimone (1999) were too small (0.1x0.3°) compared to those used by Reynolds et al., which were 0.25x1°. Ideally, the simulated bar stimuli should have been 3.75x15 pixels in size.)

The DOOG filters for the lower spatial resolution are shown in figure 3.8. Note that, for each of the two orientations detected here, two direction-of-contrast sensitive filters are generated.

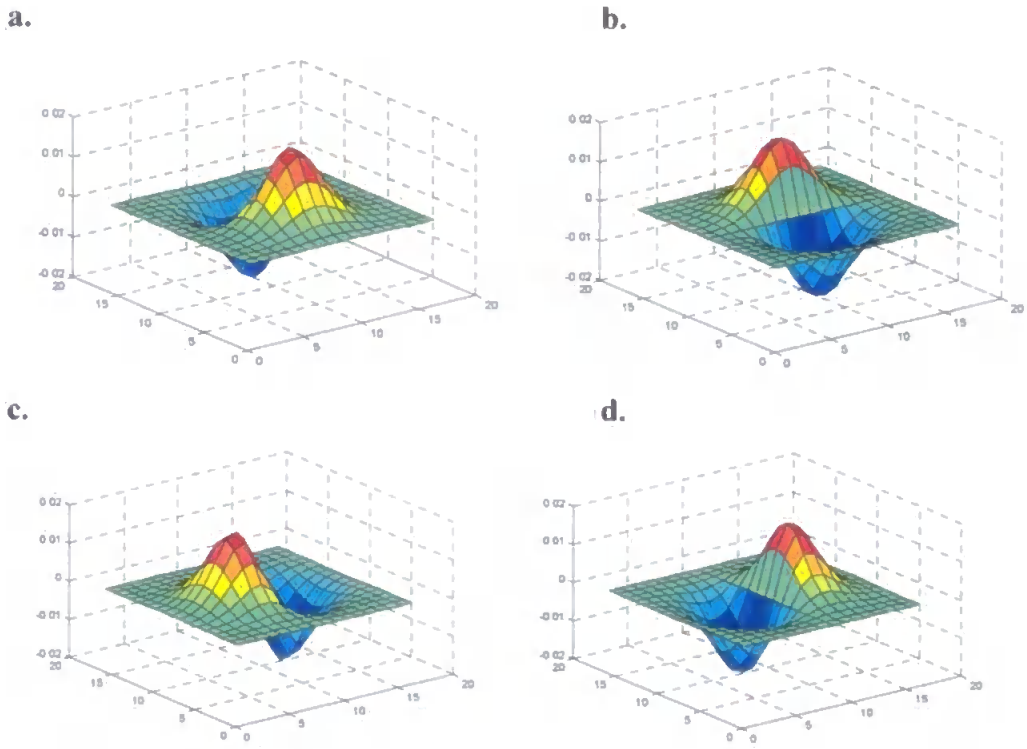



**Figure 3.8 Difference-of-Offset-Gaussian Filter**

The DOOG filter for the detection of two orientations: Vertical and horizontal.

**a.** Horizontal, direction of contrast 1

**b.** Horizontal, direction of contrast 2

**c.** Vertical, direction of contrast 1

**d.** Vertical, direction of contrast 2

Application of the above filters results in a direction-of-contrast sensitive response for each orientation. Therefore, where two orientations are detected, four right and four left hand responses are calculated (for each spatial resolution – only the finest resolution is shown in the plots here), as shown in figures 3.9 to 3.16, for the image in figure 3.1.

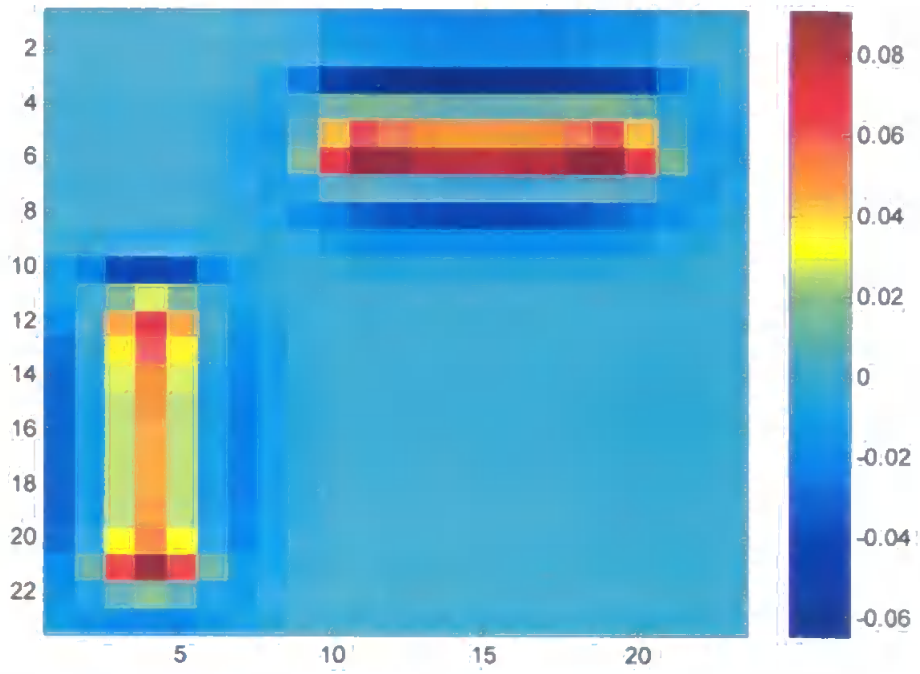



Figure 3.9 Horizontal Simple Cell Right-Hand Output – Direction of Contrast 1

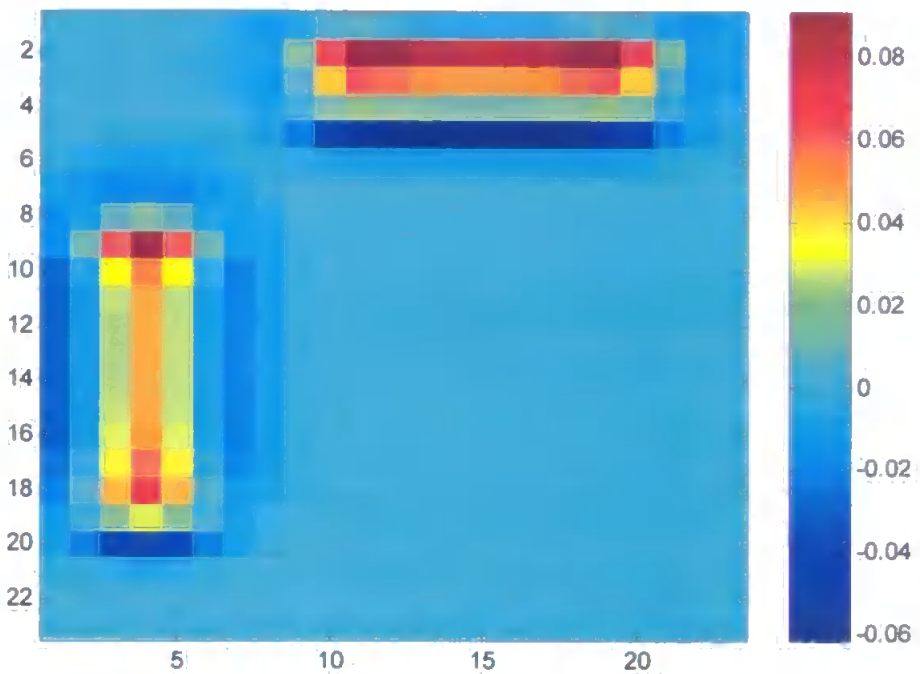



Figure 3.10 Horizontal Simple Cell Right-Hand Output - Direction of Contrast 2

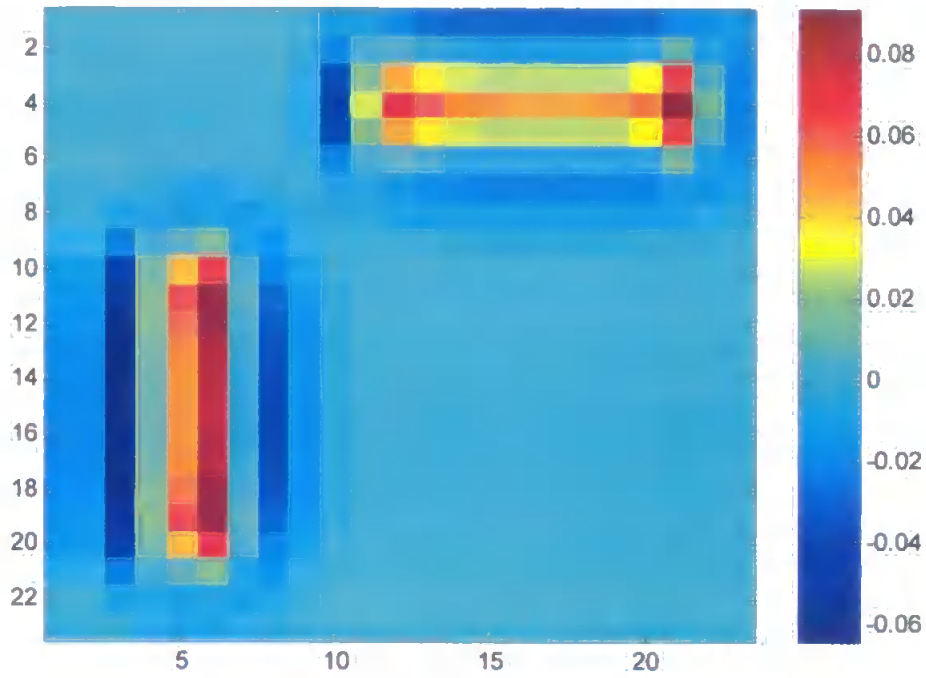



Figure 3.11 Vertical Simple Cell Right-Hand Output - Direction of Contrast 1

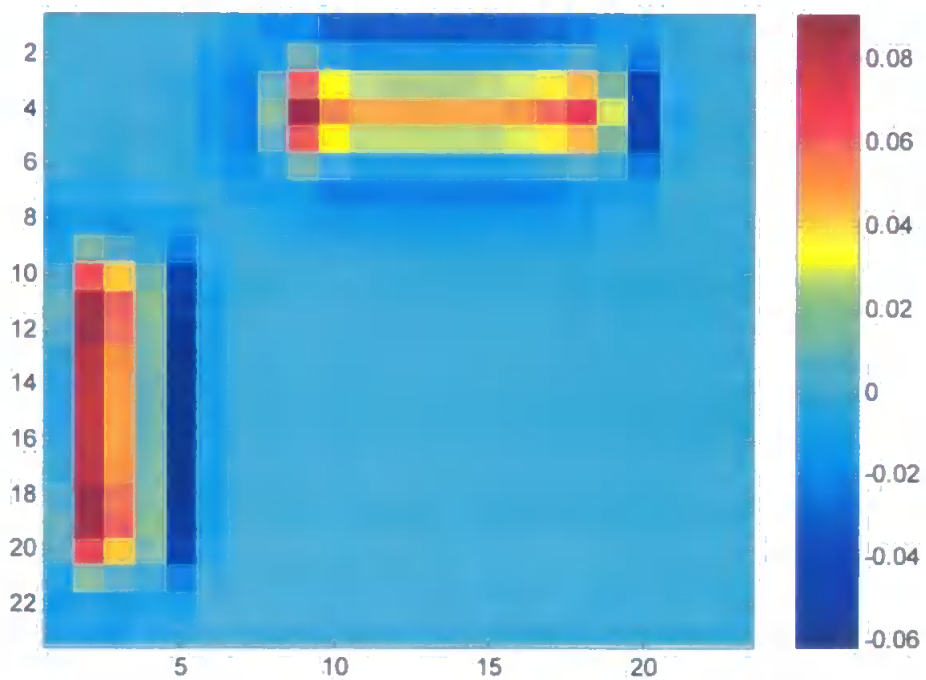



Figure 3.12 Vertical Simple Cell Right-Hand Output - Direction of Contrast 2

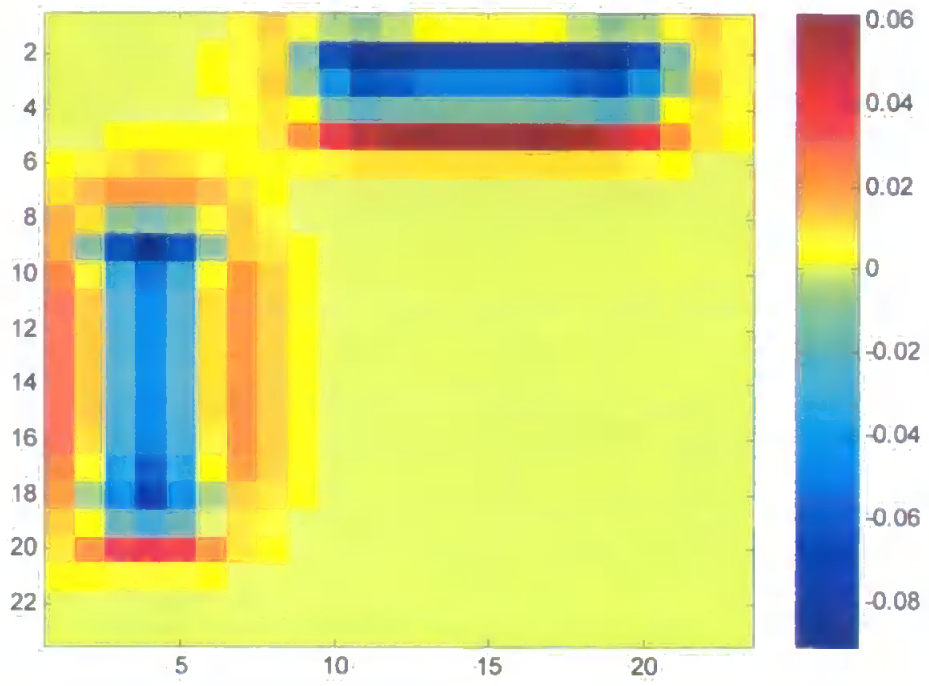



Figure 3.13 Horizontal Simple Cell Left-Hand Output - Direction of Contrast 1

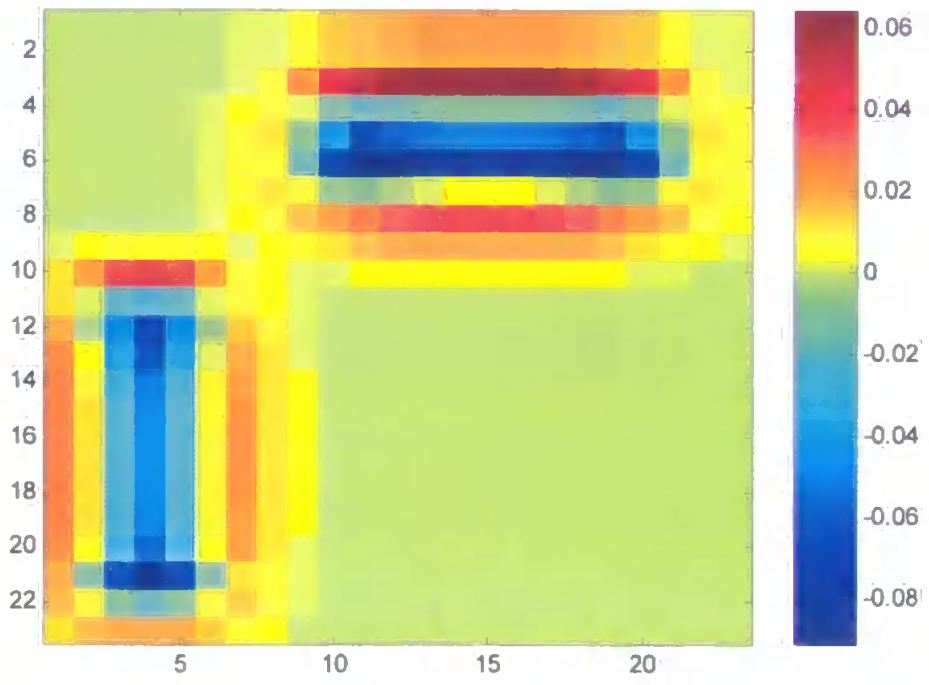



Figure 3.14 Horizontal Simple Cell Left-Hand Output - Direction of Contrast 2

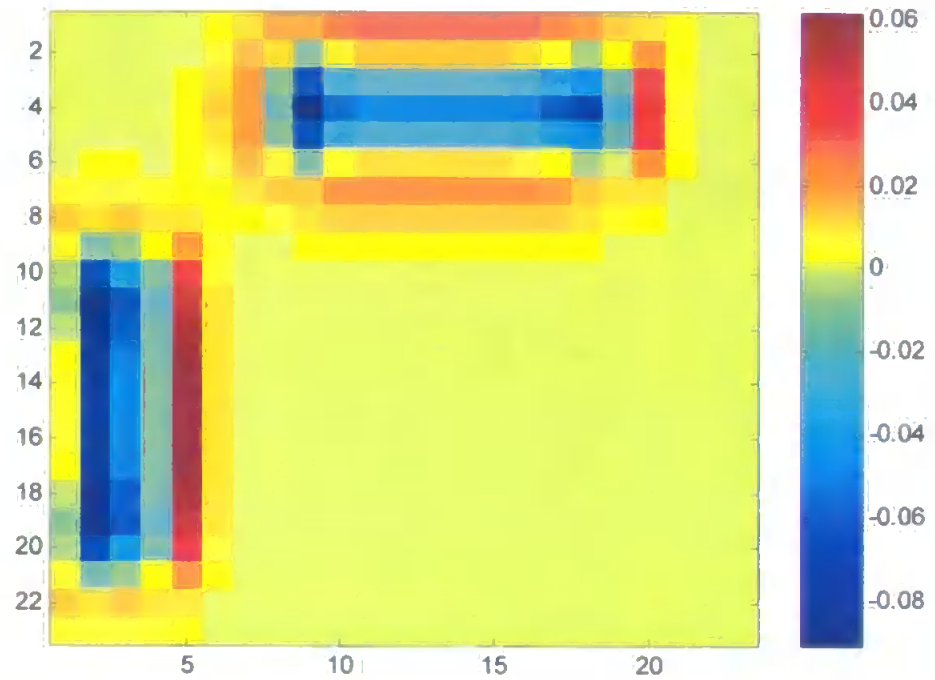



Figure 3.15 Vertical Simple Cell Left-Hand Output - Direction of Contrast 1

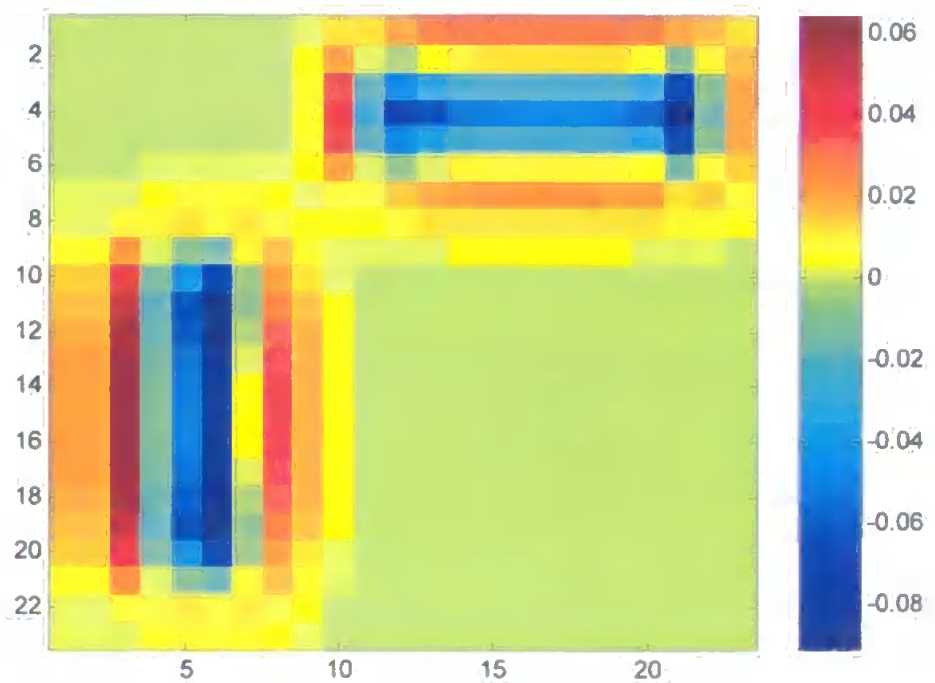



Figure 3.16 Vertical Simple Cell Left-Hand Output - Direction of Contrast 2

The plots in figures 3.9 to 3.16 show image edge effects that are normally prevented by the exclusion zone around the edge during image generation (see appendix A1.1). For the purposes of convolving image or retinal data with kernels, the edge of the data is handled by replicating the contents of the area near the edge. This tends to strengthen the response to a stimulus near the image edge.

The direction-of-contrast sensitive simple cell response is given by:

$$S_{ijrk} = \gamma[R_{ijrk} + L_{ijrk} - |R_{ijrk} - L_{ijrk}|]_+ \qquad [3.11]$$

$\gamma$ is set to10 (Grossberg and Raizada, 2000)

This leads to four (two for each orientation) simple cell responses at every location as shown in figures 3.17 to 3.20.

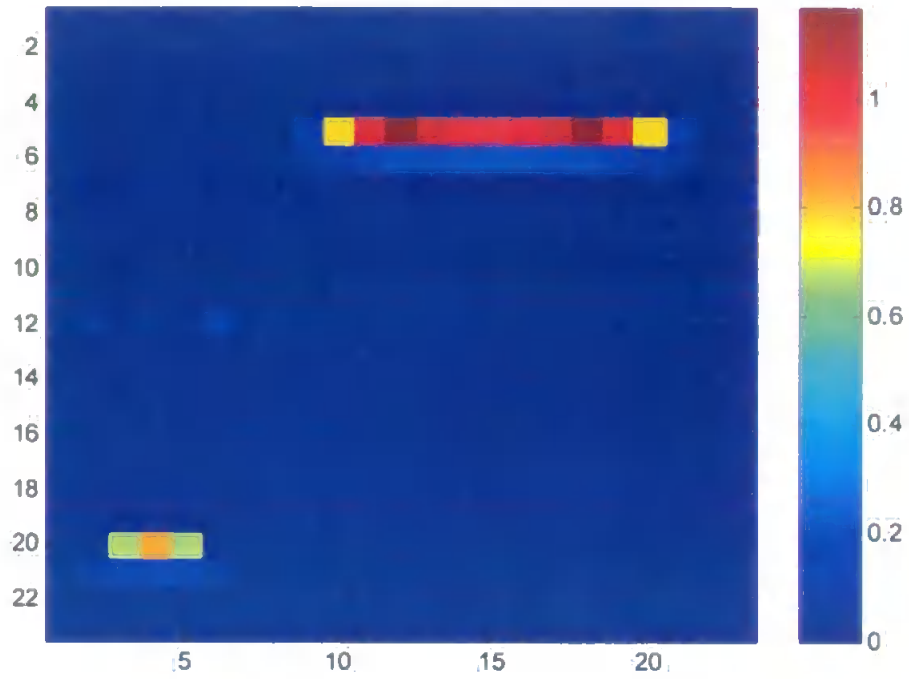



Figure 3.17 Horizontal Simple Cell Response - Direction of Contrast 1

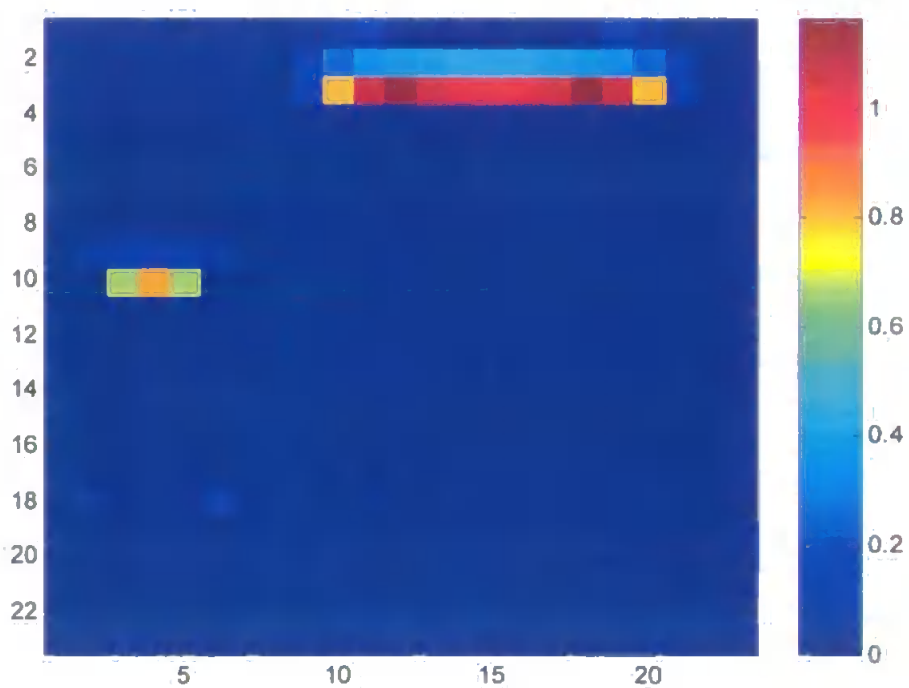



Figure 3.18 Horizontal Simple Cell Response - Direction of Contrast 2

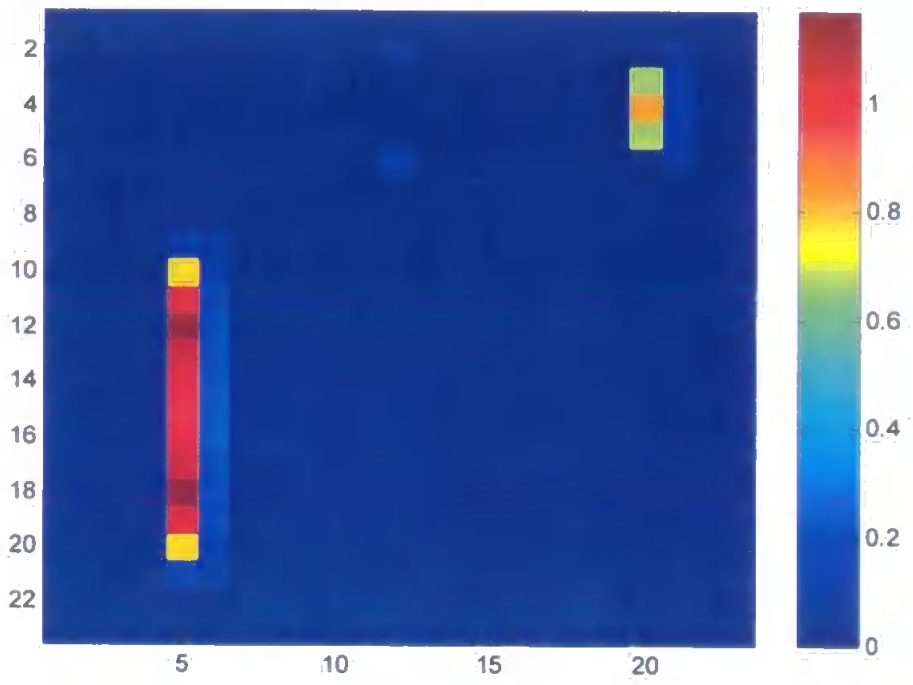



Figure 3.19 Vertical Simple Cell Response - Direction of Contrast 1

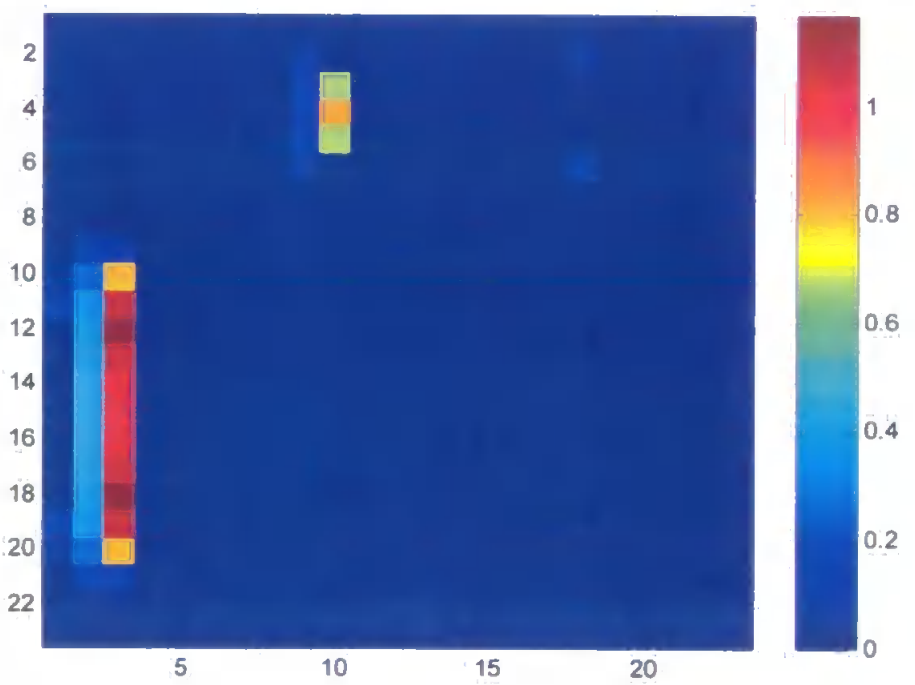



Figure 3.20 Vertical Simple Cell Response - Direction of Contrast 2

The complex cell response is invariant to direction of contrast and is given by:

$$I_{ijrk} = S_{ijrk} + S_{ijr(k+K)} \quad \text{where k ranges from 1 to K} \qquad [3.12]$$

Hence, the complex cell provides a single response for each orientation, as shown in figures 3.21 and 3.22.

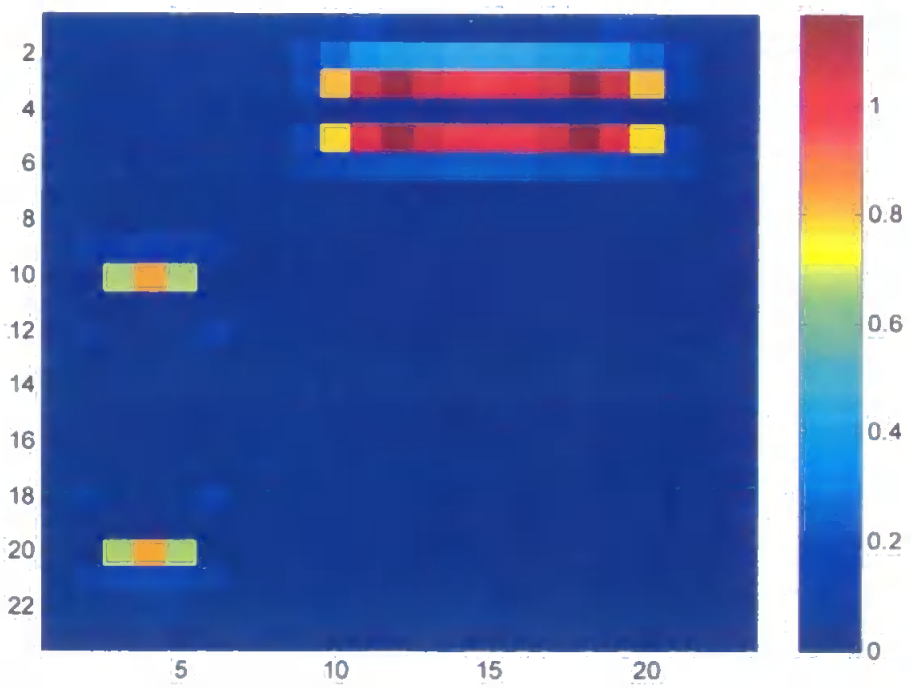



**Figure 3.21 Horizontal Complex Cell Response**

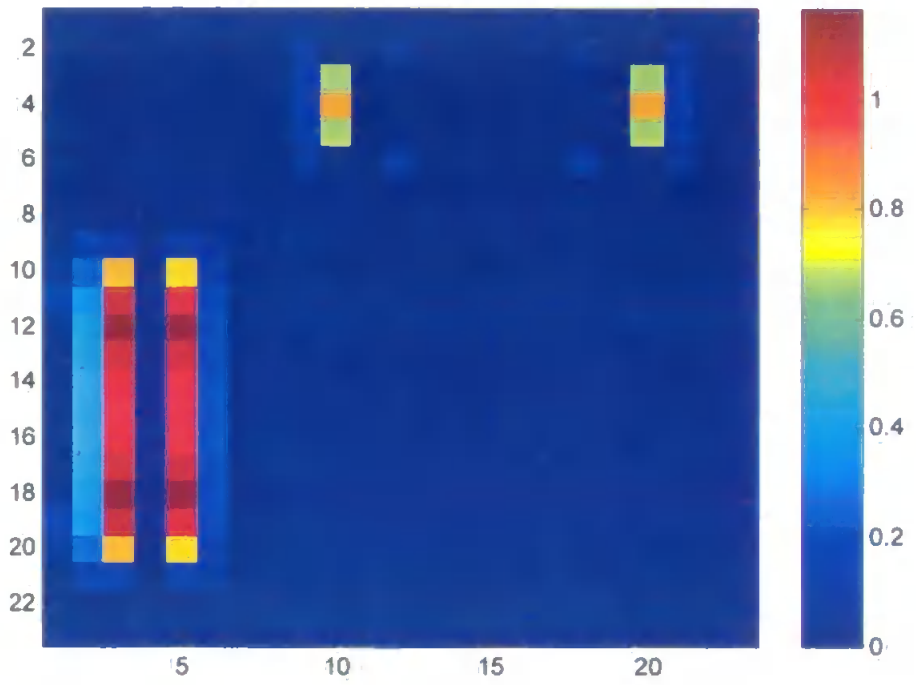



**Figure 3.22 Vertical Complex Cell Response**

The value of the complex cells, $I_{ijrk}$, over the area of the current retinal image, is input to V4.

### 3.2.2 Colour Processing in V1

The outputs of LGN concentric single-opponent cells (simplified to be the retinal cells here) are combined in cortex in the double-opponent cells concentrated in the blob zones of layers 2 and 3 of V1, which form part of the parvocellular system.

The outputs of blob cells are transmitted to the thin stripes of V2 and from there to colour-specific neurons in V4. For simplicity, V2 is not included in this model.

This section of the model is intentionally simple and is based on accepted descriptions of colour processing (in particular, Rolls & Deco, 2002; also Kandel, Schwartz, & Jessell, 1991). Double-opponent cells have a centre-surround antagonism and combine inputs from different single-opponent cells so that, for example, red cone inputs might excite the centre but inhibit the surround, whilst green cone inputs inhibit the centre and excite the surround. This is described below:

**Red on-centre portion:** [3.13]

$$\omega_{ij}^{redON} = \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{redON} + \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{greenOFF} - \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{redOFF} - \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{greenON}$$

**Red off-surround portion:** [3.14]

$$\omega_{ij}^{redOFF} = \sum_{pq} G_{pq}(i,j,\sigma_2)\nu_{ij}^{redOFF} + \sum_{pq} G_{pq}(i,j,\sigma_2)\nu_{ij}^{greenON} - \sum_{pq} G_{pq}(i,j,\sigma_2)\nu_{ij}^{redON} - \sum_{pq} G_{pq}(i,j,\sigma_2)\nu_{ij}^{greenOFF}$$

**Green on-centre portion:** [3.15]

$$\omega_{ij}^{greenON} = \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{greenON} + \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{redOFF} - \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{greenOFF} - \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{redON}$$

**Green off-surround portion:** [3.16]

$$\omega_{ij}^{greenOFF} = \sum_{pq} G_{pq}(i,j,\sigma_2) v_{ij}^{greenOFF} + \sum_{pq} G_{pq}(i,j,\sigma_2) v_{ij}^{redON} - \sum_{pq} G_{pq}(i,j,\sigma_2) v_{ij}^{greenON} - \sum_{pq} G_{pq}(i,j,\sigma_2) v_{ij}^{redOFF}$$

where:

$\sigma_1$=1.2, $\sigma_2$=1.5

These values are chosen to provide receptive fields of the same extent (11 pixels) as those provided by the high resolution DOOG filters for orientation detection above.

Note, the implementation of the model allows the addition of lower resolution colour processing similar to the two resolutions detected in the form stream. For this, $\sigma_1$= 2 and $\sigma_2$=2.7. The lower resolution colour information is not currently used and is not described further here. However, it could provide an alternative to the low-resolution monochromatic form information as input for scaling the AW (described in the next chapter). In this case, a subscript 'r' would be added to equations 3.13 to 3.18 to denote the different resolutions. This option will be discussed further in the final chapter.

The colour filters above result in the responses shown in figures 3.23 to 3.26, for the image given in figure 3.1:

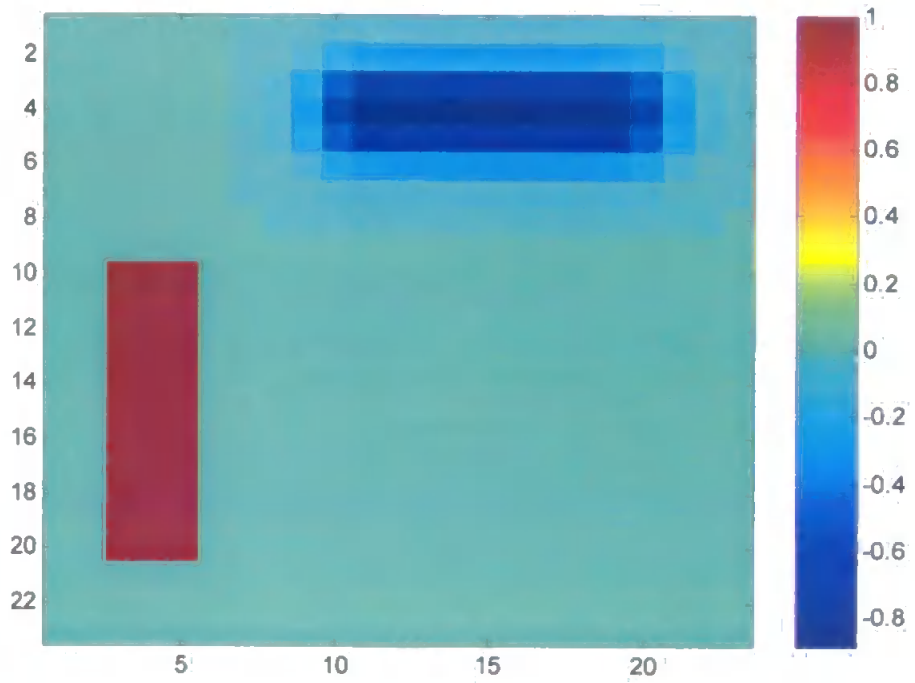



**Figure 3.23 Red On-Centre Blob Cells**

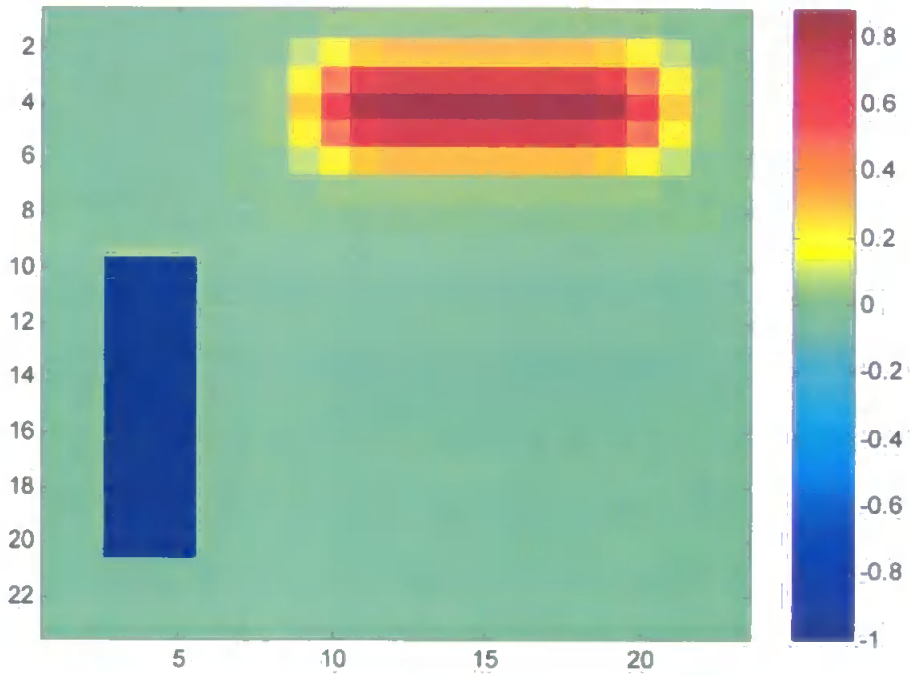



**Figure 3.24 Red Off-Centre Blob Cells**

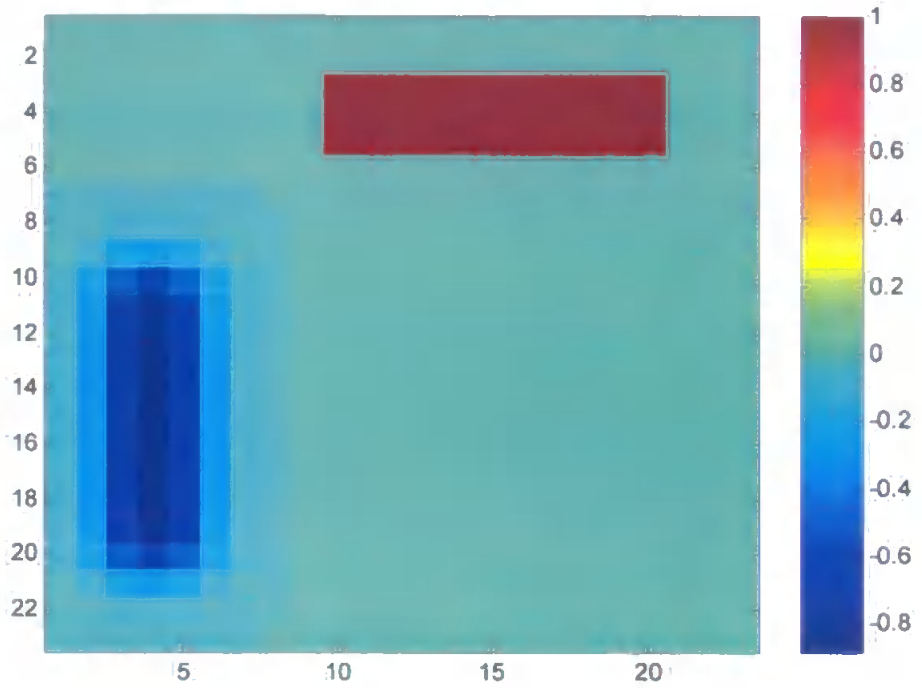



Figure 3.25 Green On-Centre Blob Cells

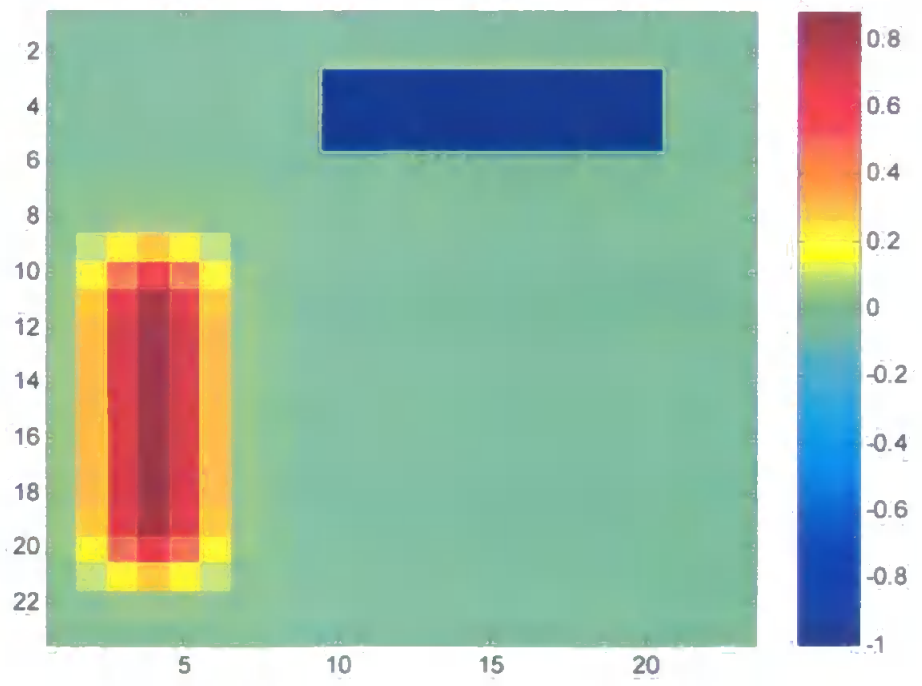



Figure 3.26 Green Off-Centre Blob Cells

The complete red-selective blob cell is given by:

$$I_{ijc^1} = \gamma\left[\omega_{ij}^{redON} - \omega_{ij}^{redOFF}\right]_+ \qquad [3.17]$$

The complete green-selective blob cell is given by:

$$I_{ijc^2} = \gamma\left[\omega_{ij}^{greenON} - \omega_{ij}^{greenOFF}\right]_+ \qquad [3.18]$$

where:

$\gamma = 0.2$ This scales the output of V1 blob cells to be consistent with that of the orientation-selective cells

$c^1 = K + 1$ This represents the position of the first colour input to V4 (i.e. red)

$c^2 = K + 2$ This represents the position of the second colour input to V4 (i.e. green)

This results in the responses shown in figures 3.27 and 3.28.

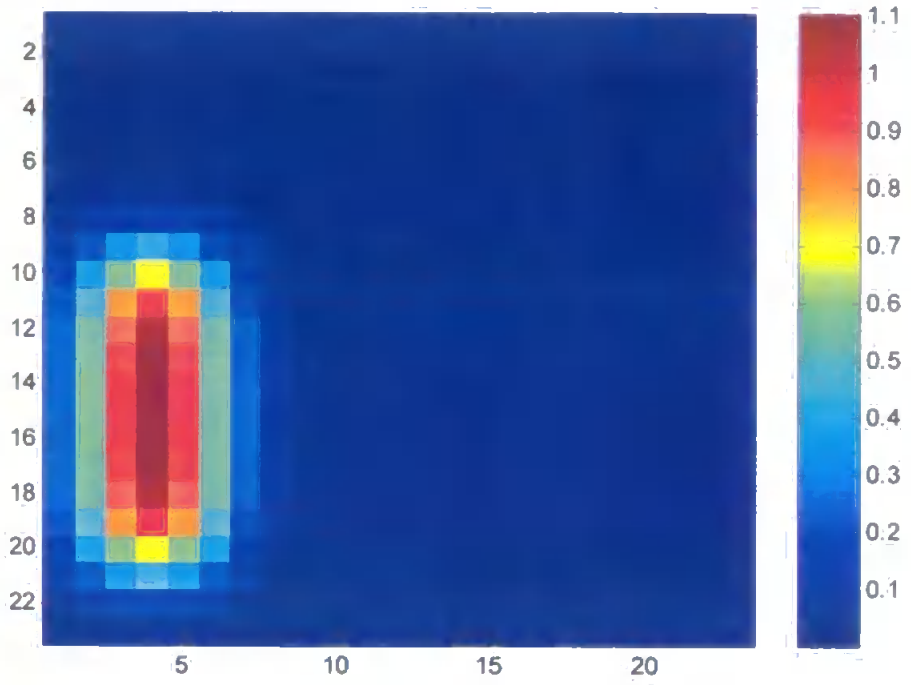



Figure 3.27 Red Complete Blob Cells

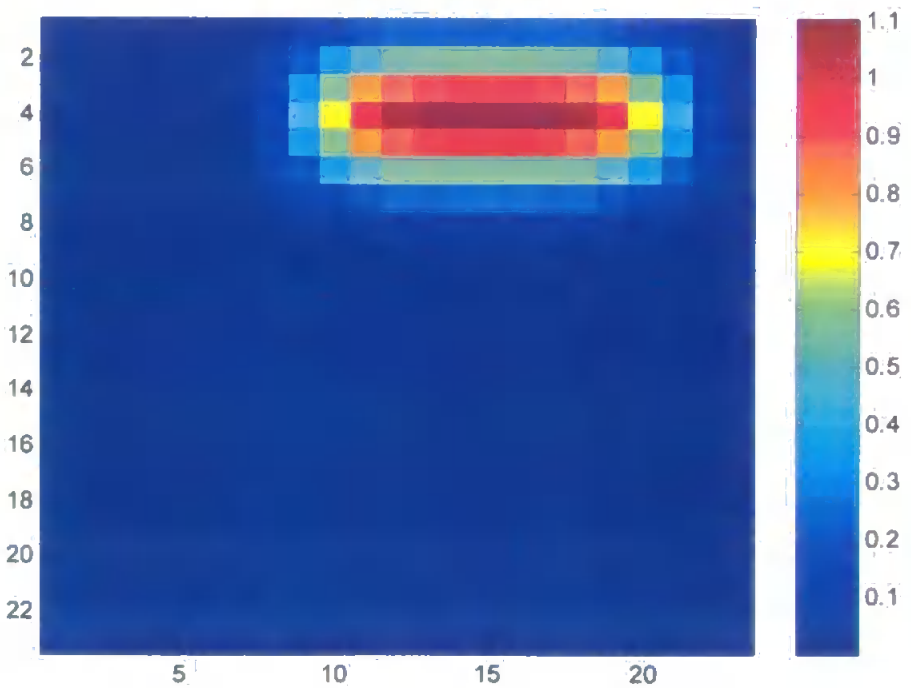



Figure 3.28 Green Complete Blob Cells

The blob cell outputs over the area of the current retinal image are input to V4.

## 3.3 Chapter Summary

Presented in this chapter was the portion of the computational model that simulates processing in the retina and V1 for the detection of orientation and colour features. As in all such computational models, the algorithms presented here are a simplification and generalisation of actual cellular processing, but are based on biological evidence. For reasons of accuracy in orientation detection, DOOG filters (Grossberg and Raizada, 2000) were used to detect form, with the modification that the filters were adapted to also detect features at a coarser spatial resolution, for the purposes of scaling the AW described later. For colour processing, a simple model was devised based on widely accepted views of colour processing in the retina and V1 (Kandel et al., 1991; Rolls & Deco, 2002), being also inspired by Grossberg and Raizada's retinal cell model.

Form and colour features detected here are input to V4 and enter the dynamic part of the system, described in the next chapter. Coarse resolution form information is used to inform the AW scaling process, which is also described in the next chapter.

# Chapter 4

# Description of the Model
# B – Dynamic Modules

In this chapter, the dynamic modules are described. These modules are the main feature of the model, which was introduced in chapter 2. The dynamic portion of the system runs after the retinal and V1 "pre-processing", described in chapter 3. First we examine the formation of the retinal image and how the AW is scaled within this. Then, the cortical processing of information within the retinal image is formally described.

## 4.1 The Retinal Image & Modes of Operation of the System

The retinal image refers to the "view" of the scene available to the dynamic cortical modules. "View" means the subset of V1 output that is passed to V4. Thus, during active search, a retinal image of a pre-determined size is moved around the scene and the cortical areas receive inputs relating to the current retinal image only. This is intended to reflect the real life situation of the subset of the 360° world that subtends the retina at any particular fixation.

In addition to being able to operate in overt mode for active search, the system can be operated in a covert mode where attentional shifts are made covertly and the retinal image remains the same throughout the simulation. The user is asked at the

start of a simulation to select the mode in which the system will operate. The covert mode is reported here for the lesion simulations in chapter 7 only. Similarly the user is requested to provide other information that determines the type of simulation, for example, whether it is monochromatic or colour. The user also determines whether the activity within each of the dynamic modules should be plotted at each time step or whether plots should be suppressed in order to speed operation of the simulation. Examples of these plots will be given as figures later in this chapter. The system can also run several scan path simulations in serial, i.e. multiple loops to produce several different scan paths on the same image, in order to collect statistics about the scan path behaviour under certain circumstances, as will be discussed in chapter 6. In this statistics mode, all parameters about the type of run are preset so that no human intervention is required during the simulation.

### 4.1.1 Size of the Retinal Image

The size of the retinal image can be varied for any particular simulation, but is normally set to 441 pixels (radius of 220 pixels), which equates to approximately 40° of visual angle (where the degree to pixel ratio is determined by the size of the V1 filters so that 1° equates to 11 pixels). This is smaller than our natural vision but this restriction is necessary in order to make the system computationally practical[1]. The size of the cortical areas V4 and LIP is scaled based on the retinal image size (as described in appendix A1.2.4). Therefore, a smaller retinal image

[1] To create scenes, images of maximum size 1000x1000 pixels were used (this was the maximum size allowed by Matlab to create the necessary arrays). Retinal images should be small enough to be able to move around such scenes under the active vision paradigm modelled here.

leads to fewer assemblies in these areas and, thus, faster active vision processing. The flexibility in retinal image size allows comparison of scan path behaviour when using a larger biologically plausible retina or optimisation of processing speed with a smaller retina. Stable performance across a range of retinal image sizes is possible due to the normalisation of inputs to IT, described later. The only restriction is that a very small retina tends to lead to a higher proportion of fixations landing in blank areas of very sparse scenes due to lack of stimuli within the limited retina. This will be discussed in chapter 6.

### 4.1.2 Acuity

The retinal image is simplified in that it assumes acuity is constant across the retina, rather than decreasing towards the periphery. A simple decrease in resolution could be modelled in the manner described by Rybak et al. (1998), based on Burt (1998). However, such an approach would mean that the raw image data would have to be sampled, and retinal and V1 processing performed, at every fixation point. This means that the retina and V1 stages could not be performed as pre-processing and would contribute to the processing time at each fixation. In addition, implementation of the Rybak/Burt algorithm proved to be computationally expensive due to the number of convolutions (multiplying by a Gaussian kernel in a recursive manner) required to gradually decrease resolution in peripheral vision and would dramatically increase processing time during the active vision portion of the system. Therefore, this system, in common with most similar

models (e.g. Deco, 2001; Deco & Lee, 2002; van der Velde & de Kamps, 2001), uses a retina of constant acuity. This is made more plausible by the limited size of the retinal image.

### 4.1.3 Formation of a Retinal Image at the Scene Edge

When a fixation is placed at a distance less than the retinal image radius from the original image edge, the retinal image extends beyond the scene (obviously, this is an artefact of the model and would not occur in real life due to the continuous nature of the visual world). When this happens, the dynamic cortical areas receive no "bottom-up" stimulus information for the area beyond the original image but other cortical processing remains unaltered. This is described in appendix A1.2.3. Importantly, this means that the next fixation point is never chosen at a location beyond the extent of the scene because LIP only represents the locations of behaviourally relevant stimuli, and a location not in receipt of stimulus related information never becomes strongly active.

Alternative solutions included padding the V1 inputs to the full size of the retinal image. However, the cortical areas are driven by the V1 inputs received within the retinal image and the spurious padded area could effect cortical processing, in particular making it possible to select a location in this area as the next fixation point. Another possibility is to limit the fixation position so that an area extending the radius of the retinal image around the periphery is excluded. However, this

could severely limit the positions available for fixation. Alternatively, the selected fixation position could be moved towards the centre of the image when it was found to be too close to the edge of the image but this would artificially alter the scan path towards features and objects that had not been selected on the basis of activity in LIP. Another alternative is to constrain the size of the retinal image wherever an edge of the original image is found. With this solution, the retinal image would not necessarily be square with the fixation point at its centre. This would further complicate the coordinate systems described in appendix A1.2.

## 4.2 Scaling the Attention Window (AW)

Within each retinal image, an initially spatial AW is formed, to reflect the extent of the *perceptual span* (Bertera & Rayner, 2000) or *area of conspicuity* (Motter & Belky, 1998a) around fixation. The aperture of this window is scaled according to coarse resolution information reflecting local stimulus density, which is assumed to be conveyed rapidly by the magnocellular pathway to parietal cortex, although the source of this spatial bias is uncertain, as discussed in chapter 2. Hence, during a scan path, the size of the AW is dynamic, being scaled according to stimulus density found within the current retinal image. The calculation of the radius of the AW is inspired by the psychophysical findings of Motter and Belky (1998a), who found that targets could be discriminated within 2 ANNDs (Average Nearest Neighbour Distance) of the fixation point. They derived their ANND empirically

but suggested that the square root of the image size divided by the number of stimuli provides an estimate of stimulus density.

Therefore, the radius of the AW here is given by the following:

$$AWrad = \min\left[ Round\left[ \sqrt{\frac{m * n}{d}} * 2 \right], \min(m, n) \right] \qquad [4.1]$$

where:

d is the number of non-zeros in f below

m,n are the dimensions of the retinal image.

$$f = \psi\left(\sum_{k=1}^{K} I_{ijrk}\right) \qquad [4.2]$$

where:

$I_{ijrk}$ is the output of the orientation-selective V1 simple cell process (selective for K orientations) over the area of the current retinal image. When fixation is close to the edge of the original image and the retina "overflows" the original image, the output of the retinal and V1 stages in the "overflow" area are considered to be zero so that no "bottom-up" inputs are provided to V4 from this area (as mentioned above and described in appendix A1.2.3).

However, for the purposes of scaling the AW, this region is ignored and only that portion of the retina and V1 that does not "overflow" is processed here.

r is the spatial resolution of the orientation information; set to 2, the lowest resolution detected

ψ is a function that removes the lowest 95% of activity and reduces to zero the activity at points where a neighbour has been found within a Euclidean distance equal to the length of the bar stimuli (11 pixels)

Motter & Holsapple (2000) later showed that the area of conspicuity for target detection was based on the density of stimulus as represented in cortex (i.e. accounting for the decrease in peripheral resolution, which is not modelled here). The process here is compatible with this, since it is based on the coarse resolution magnocellular output of V1.

A demonstration of the scaling of the AW according to local stimulus density is given in figure 4.1, which shows the AWs within equal sized retinal images associated with two fixation points in a scene containing a patch of dense stimuli in the top left corner. The AW is smaller when the retinal image covers the dense patch of stimuli. Thus, the attentional focus zooms in and out, according to the density of local stimuli, as the scan path moves around the scene (Lanyon & Denham, 2004a).

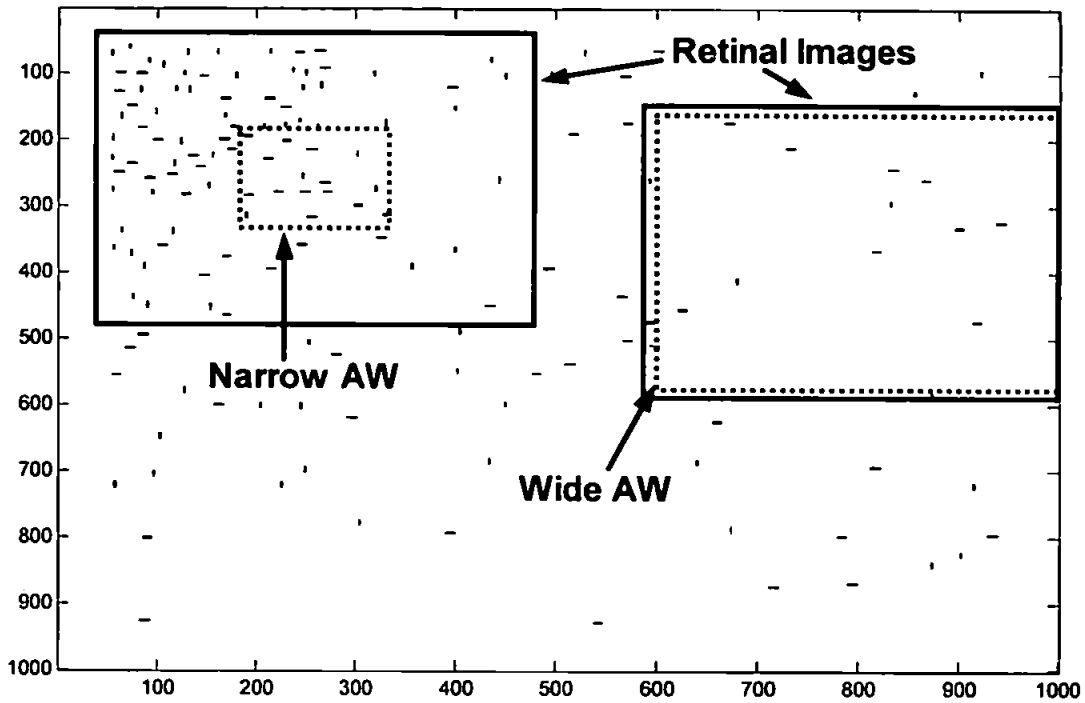



**Figure 4.1 Attentional Zoom**

Attentional zoom in areas of high stimulus density is demonstrated by the difference in aperture of the AW. For two fixation points, chosen to be located in dense and sparse regions of a scene of varying stimulus density, the retinal image (outer plotted boxes) and the associated spatial AW (inner boxes) are shown. The retinal image on the right extends slightly past the edge of the original image.

Further examples of AW scaling will appear in figures later in this chapter and an examination of scan path behaviour in scenes of mixed stimulus density will be outlined in chapter 6. Average AW radii obtained from multiple scan paths over each image are presented in appendix A3 (along with other statistics to be described in chapter 6) and summarised in table 4.1. Note that stimulus density is not uniform across the scene because stimuli are randomly positioned.

| Image (size 1000x1000 pixels) | Average Range of AW Radius (in Pixels) | Average Range of AW Radius (in Degrees) |
|---|---|---|
| **1000rv11 – contains 23 bars** | 303-345 | 28-31 |
| **1000gh50 – contains 101 bars** | 163-182 | 15-17 |
| **1000rh150 – contains 301 bars** | 91-95 | 8-9 |

**Table 4.1 Average AW Radii**

The average AW radii given in appendix A3 are summarised and this shows that the radius decreases as stimulus density increases. For reference throughout this thesis, image names comprise three components: A numeric identifier that signifies the size of the original image (in pixels; given by M in appendix A1.1), an alphabetic identifier of the target object, a numeric that signifies the number of distractors of each feature type. So, 1000rv11 is an image of size 1000x1000 pixels, which contains a red vertical target, 11 green vertical distractors and 11 red horizontal distractors.

## 4.3 Dynamic Modelling Approach

The dynamic cortical modules follow a similar approach to that described by Deco (2001) and Deco and Lee (2002), being modelled using mean field population dynamics (Gerstner, 2000; Wilson & Cowan, 1972). Such an approach was previously used in a smaller-scale model (Usher & Niebur, 1996), which proved the biased competition concept that a prefrontal bias could influence competition

between objects in IT. Under this approach, average ensemble activity is used to represent populations, or assemblies, of neurons with similar encoding properties. It is assumed that neurons within the assembly receive similar external inputs and are mutually coupled. Modelling at the neuronal assembly level is inspired by observations that several brain regions contain populations of neurons that receive similar inputs and have similar properties, and it is a suitable level at which to produce a model that has interacting modules representing brain regions (see Rolls & Deco, 2002, and Gerstner, 2000, for further details of this modelling approach). Population averaging does not require temporal averaging of the discharge of individual cells and, hence, the response of the population may be examined over time, subject to the size of the time step used in the differential equations within the model. The response function, which transforms activity within the assembly into a discharge rate, is given by the following sigmoidal function (Gerstner 2000). This has a logarithmic singularity, which reflects the threshold of the cell (given by $\tau$ below):

$$F(x) = \frac{1}{T_r - \tau \log(1 - \frac{1}{\tau x})} \qquad [4.3]$$

where:

$T_r$, the absolute refractory time, is set to 1ms

$\tau$, is the membrane time constant (where $\frac{1}{\tau}$ determines the firing threshold).

With τ set to 20 (as used by Usher & Niebur, 1996) this function takes the form shown in figure 4.2.

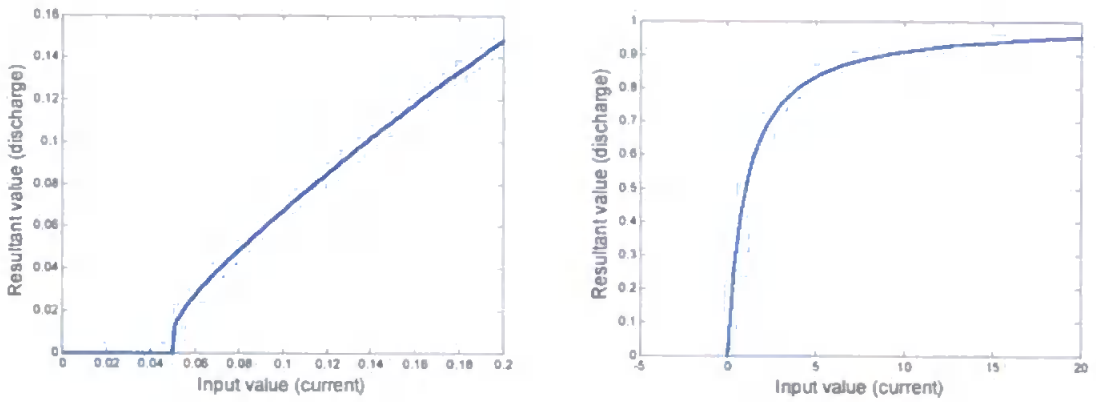



**Figure 4.2 Response Function with Low Threshold**

Response function with τ set to 20, i.e. threshold (1/τ) = 0.05. The plot on the left shows behaviour at low input values, as shown by Usher and Niebur (1996) and that on the right shows higher input values, which are more relevant to the model here.

Figure 4.2 demonstrates that the function produces output in the range (0,1). Output values above 1 can be obtained by decreasing the refractory time, $T_r$. However, such a modification is not biologically plausible because the absolute refractory time (during which the cell is unable to spike) lasts a few milliseconds and the relative refractory time (during which the cell is less likely to spike) lasts tens of milliseconds.

The initial slope of the function is steep but higher values of input current produce very little variation in output. Hence, once activity within the pool reaches more than a value of 6, the transfer function does not differentiate input activities well. The use of this function with the above parameters produced problems in LIP when it received the outputs from the V4 feature maps. The V4 output transformed by this function did not clearly differentiate the most active assemblies in V4 and, thus, did not necessarily attract the scan path to the correct locations. Altering the membrane time constant, $\tau$, shifts the position of the threshold and, hence, the steepest part of the curve. For example, smaller $\tau$ values raise the threshold. Therefore, here, the threshold value is set dynamically, being based on the maximum activity within a group of assemblies. The assemblies that form this group differ, to some extent, depending on cortical area, and will be described in the description of each area later. For example, all assemblies in IT form the group but only the assemblies in a particular feature layer in V4 form a group. The threshold (i.e. $1/\tau$) is set to half the current maximum activity in the group. For example, if the maximum group activity is 5, the threshold becomes 2.5 and produces the transfer function shown in figure 4.3.

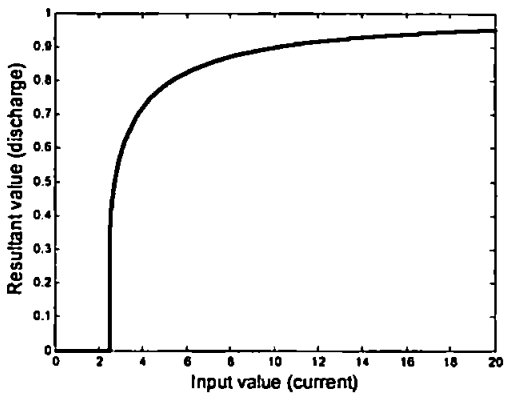



**Figure 4.3 Response Function with Higher Threshold**

Response function with $\tau$ set to 0.4, i.e. threshold $(1/\tau) = 2.5$.

Therefore, the function shifts leftward or rightward dependant on the maximum current activity within the group of assemblies. The output of each assembly is, therefore, normalised by the overall level of activity in the group. This modification provides a useful normalisation of activity within the model and prevents activity in any one area growing without bound. Stable performance within the model was found to depend on such local normalisation of activity. Normalisation of cortical activity dependant on overall activity in a local region of cortex is biologically justified. It is well known that the visual system adapts to the wide range of possible light levels and discounts the illuminant (Helmholtz, 1867; Craik, 1940). In particular, activity due to stimuli within the local region of the visual field appears to normalise neuronal response properties. For example, the perceived speed of a drifting grating is modulated by the frequency of nearby gratings (Smith & Derrington, 1996). When more than one stimulus is present within the receptive field of a neuron in area MT, its response is typically much less

than expected by summing the response of the component stimuli and depends on stimulus contrast (Heuer & Britten, 2002). In general, such gain control appears to be a feature of the visual system in many species (Brady & Field, 2000; Ohzawa, Sclar & Freeman, 1985).

## 4.4 The Dynamic Modules

The dynamic portion of the module comprises modules representing V4, IT and LIP as well as external biases to these modules. LIP contains the same number of neurons as V4 and has reciprocal connections with both the orientation and colour layers in V4. The size of V4 and LIP is determined by the size of the retinal image, which is flexible, as described above. For single cell simulations, V4 consists of one assembly in each feature layer and for scan path simulations, with a typical retinal image size of 441x441 pixels, V4 consists of 20x20 assemblies in each feature layer. For single cell simulations monochromatic stimuli are used and, therefore, V4 contains only orientation selective assemblies in this case. Otherwise, V4 contains both colour and orientation selective assemblies.

Each neuronal assembly in V4, LIP and IT is modelled by a differential equation, which has to be implemented as a difference equation with discrete time steps. System performance in terms of simulation run time speed is examined in appendix A2.1.

### 4.4.1 V4

Like V1, V4 is retinotopic and processes both colour and orientation. This area is known to be involved in the representation of colour as well as form (e.g. Zeki, 1993). V4 receptive fields are larger than those in V1 and each V4 neuron receives inputs from the relevant V1 feature detector neurons in its receptive field over an area of 23x23 V1 retinotopic positions. For each feature, a V1 neuron exists at every pixel position and its 1° receptive field extends 11 pixels. Thus, V4 receptive fields here extend over an area of 33x33 pixels, approximately 3°, which is typical of monkey V4 receptive fields that tend to be 2-4° wide in the central visual field (Moran & Desimone, 1985). V4 assemblies are arranged such that their receptive fields overlap by one V1 neuron, i.e. about 1°. However, the system is implemented in a flexible manner such that the sizes of receptive fields in all modules and the overlap of receptive fields here can be adjusted for any particular simulation. If the receptive field size is made much larger, there is a loss of spatial precision in the information passed from V4 to LIP and, hence, in the ability of the scan path to be able to fixate target coloured stimuli. This is because the stimulus could be present anywhere within the V4 receptive field and the centre of the LIP receptive field at the corresponding retinotopic position (i.e. the centre of the V4 receptive field) is selected as the saccade end-point. Thus, large receptive fields in V4 and LIP may reduce the spatial accuracy of saccade targeting. Introducing a V1 module to the dynamic portion of the model could provide a means of increasing

spatial accuracy. Intermediate ventral stream areas such as V4 could pass crude location information about potential target stimuli to LIP with this information being supplemented by more accurate spatial information from lower levels of the ventral pathway (areas V1 and V2). This would be possible if the V1 module (and possibly a V2 module), added to the dynamic portion of the model, received feedback from V4 that biased representations towards target features. However, the current implementation of V4 and LIP receptive fields covering 3°, allows the scan path to accurately locate stimuli without the addition of this V1 processing, as will be shown in detail in chapter 6.

V4 receives convergent input from V1 over the area of its receptive field with a latency of 60ms to reflect normal response latencies (Luck et al., 1997). In order to simulate the normalisation of inputs occurring during retinal, LGN and V1 processing, the V1 inputs to V4 are normalised by passing the convergent inputs (i.e. those within the assembly's receptive field) to each V4 assembly through the response function at equation 4.3 with its threshold set to a value equivalent to an input activity for approximately half a stimulus within its receptive field. This threshold was found empirically and set to 12 for orientation selective assemblies and 23 for colour selective assemblies. The difference is due to the different types of filters used in V1 (described in the previous chapter) and this action normalises these slight differences. Normalisation of featural inputs in order to eliminate feature-dependent amplitude differences due to different feature extraction methods is normal practise in modelling where different feature types are detected (see, for

example, Itti & Koch, 2000). If the V1 inputs to V4 were not normalised, a V4 assembly that had more than one stimulus *of the same type* (hence, V4 received strong inputs relating to the same features from V1) within its receptive field would become very active due to the strength its bottom-up inputs. In this situation, this bottom-up effect was found to be stronger than any other bias in V4 and tended to lead scan paths to locations that contained two stimuli of the same type within the same V4 receptive field. Normalisation of the feedforward inputs to the dynamic portion of the system removed the need for image normalisation pre-processing such as that applied by Deco (2001; Deco & Lee, 2002). The type of normalisation performed by Deco's model across the entire scene means that bottom-up inputs to each cortical assembly is dependant on stimulus density across the scene, since the activity for each stimulus is divided by total activity.

The V4 module is arranged as a set of feature "layers" or "maps" encoding each feature in a retinotopic manner. Each feature layer belongs to a feature *type*, i.e. colour or orientation. Although this is a simplification of the complex response properties of V4 neurons, they are known to be functionally segregated to some extent (Ghose & Ts'O, 1997), and the separation of feature maps is a common theoretical and computational modelling approach (e.g. Deco, 2001; Deco and Lee, 2002; Hamker, 1998; Itti & Koch, 2000; Koch & Ullman, 1985; Niebur, Itti, Koch, 2001; Schneider, 1999; Treisman, 1982; Treisman & Gelade, 1980). This representation is consistent with the feature layers/maps used in Treisman's Feature Integration Theory (Treisman, 1988, 1991, 1998; Treisman & Gelade, 1980) and

Wolfe's Guided Search Model 2.0 (Wolfe 1994). Furthermore, the seminal contribution of Koch and Ullman (1985) suggested that such feature maps exist in early representations, in particular, suggesting colour & orientation are represented in separate topographical maps. This theory has been implemented in a number of computational models, for example, Itti and Koch (2000). Hamker (2004) has recently implemented separate colour and orientation feature maps in V4, in a similar manner to those described here. Therefore, such separation is fundamental in a number of cognitive theories and is a standard way of implementing such feature encoding in computational models.

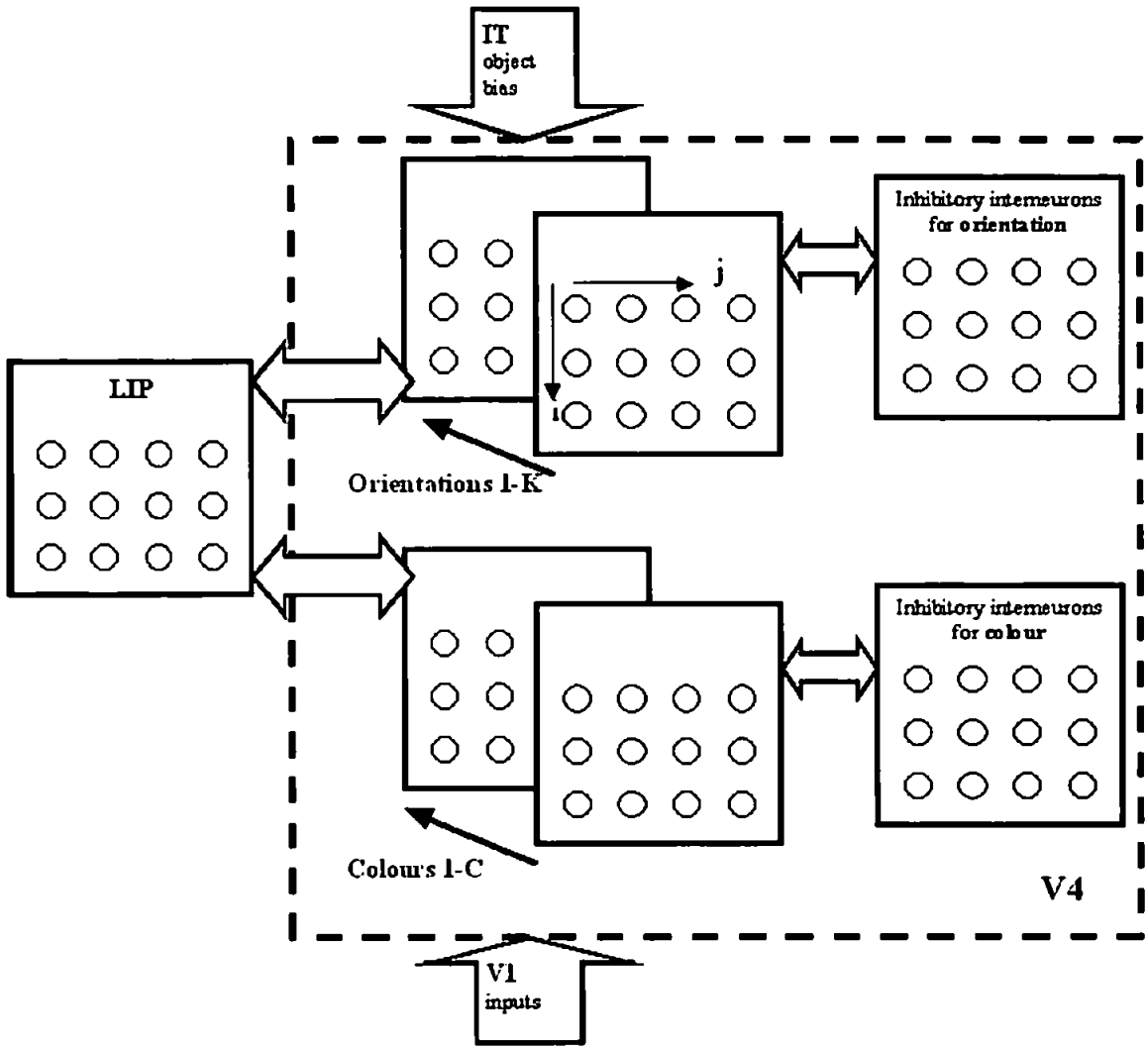



**Figure 4.4 Competition in V4**

V4 consists of i x j nodes represent retinotopic positions, in layers for C colours and K orientations. Two inhibitory interneuron pools exist to mediate competition: One for colours and one for orientations. Competition in V4 is influenced by external biases from V1, LIP and IT. Reciprocal connections exist with LIP for all retinotopic locations. This figure is the same as figure 2.4, presented in chapter 2.

V4 consists of a three dimensional matrix of pyramidal cell assemblies. The first two dimensions represent the retinotopic arrangement and the other represents the individual feature types: Orientations and colours. In the latter dimension, there are K + C cell assemblies, as shown in figure 4.4. The first K assemblies are each tuned to one of K orientations. The next C assemblies are each tuned to a particular colour. Two orientations (vertical and horizontal) and two colours (red and green) are normally used here. Competition operates within feature type so two sets of inhibitory interneuron pools exist: One set mediates competition between orientations and the other mediates competition between colours. The inhibitory interneurons are retinotopically organised in the same way as the pyramidal cells and are connected to those of the corresponding feature type at the same spatial location. Therefore, competition operates between different features of the same type at the same spatial location. This provides competition between features within the same V4 receptive field, as has been observed in single cells (Chelazzi et al., 2001; Moran & Desimone, 1985). With a single inhibitory interneuron for each feature type, as described by Deco (2001; Deco & Lee, 2002, Rolls & Deco, 2002), the competition was found to be too strongly winner-take-all such that scan paths were attracted to the same location many times. The nature of inhibition in V4 affects the representation of possible target locations in LIP. If V4 is implemented with a common inhibitory pool for all features or per feature *type* (i.e. one for colours and one for orientations), this results in a normalising effect similar to a winner-take-all process, whereby high activity in any particular V4 assembly can

strongly suppress other assemblies. V4 assemblies receive excitatory input from LIP, as a result of the reciprocal connection from V4 to LIP. When LIP receives input from all features in V4 as a result of two different types of stimulus (e.g. a red vertical bar and a horizontal green bar) being within a particular V4 receptive field, these feature locations in V4 will become most highly active and may suppress other locations too strongly. Thus, the common inhibitory pool tends to favour locations that contain two different stimuli within the same receptive field. Previous models that have used a common inhibitory pool (Deco, 2001; Deco & Lee, 2002; Rolls & Deco, 2002) may not have encountered this problem because only one feature type was encoded. Therefore, the scan paths here are most effective when features *within a particular feature type* compete locally. For example, an assembly selective for red stimuli competes with an assembly at the same retinotopic location but selective for green stimuli. Hence, inhibitory interneuron assemblies in V4 exist for every retinotopic location in V4 and for each feature type.

#### 4.4.1.1 Form Processing in V4

The output from the V1 simple cell process, $I_{ijk}$, for each position (i,j) at orientation k, provides the bottom-up input to orientation selective pyramidal assemblies in V4 that evolve according to the following dynamics:

$$\tau_1 \frac{\delta}{\delta t} W_{ijk}(t) = -W_{ijk}(t) + \alpha F\left(W_{ijk}(t)\right) - \beta F\left(W_{ij}^{IK}(t)\right) + \chi \sum_{pq} I_{pqk}(t) +$$
$$+ \gamma Y_{ij}(t) + \eta \sum_{m} B_{W_{ijk} X_m} . F(X_m(t)) + I_0 + \nu$$

[4.4]

where:

$\tau_1$ is set to 20ms

$\alpha$ is the parameter for excitatory input from other cells in the pool, set to 0.95

$\beta$ is the parameter for inhibitory interneurons input, set to 10

$I_{pqk}$ is the input from the V1 simple cell edge detection process at all positions within the V4 receptive field area (p,q), and of preferred orientation k

$\chi$ is the parameter for V1 inputs, set to 4

$Y_{ij}$ is the input from the posterior parietal LIP module, reciprocally connected to V4

$\gamma$ is the parameter for LIP inputs, set to 3

$X_m$ is the feedback from IT cell populations via weight $B_{W_{ijk} X_m}$, described later

$\eta$ is the parameter representing the strength of object-related feedback from IT; set to 5 for scan path simulations and simulations of single cell recordings in V4, but set to 2.5 for simulation of single cell recordings in IT (Chelazzi et al., 1993),

$I_0$ is a background current injected in the pool, set to 0.25

$\nu$ is additive noise, which is randomly selected from a uniform distribution on the interval (0,0.1)

The dynamic behaviour of the associated inhibitory pool for orientation-selective cell assemblies in V4 is given by:

$$\tau_1 \frac{\delta}{\delta t} W_{ij}^{IK}(t) = -W_{ijk}^{IK}(t) + \lambda \sum_k F\left(W_{ijk}(t)\right) - \mu F\left(W_{ij}^{IK}(t)\right) \qquad [4.5]$$

where:

$\lambda$ is the parameter for pyramidal cell assembly input, set to 1

$\mu$ is the parameter for inhibitory interneuron input, set to 1

Over time, this results in local competition between different orientation selective cell assemblies.

#### 4.4.1.2 Colour Processing in V4

The output from the V1 simple cell process, $I_{ijc}$, for each position (i,j) and colour c, provides the bottom-up input to colour selective pyramidal assemblies in V4 that evolve according to the following dynamics:

$$\tau_1 \frac{\delta}{\delta t} W_{ijc}(t) = -W_{ijc} + \alpha F\left(W_{ijc}(t)\right) - \beta F\left(W_{ij}^{IC}(t)\right) + \chi \sum_{pq} I_{pqc}(t) + \gamma Y_{ij}(t)$$
$$+ \eta \sum_m B_{W_{ijc} X_m} . F(X_m(t)) + I_0 + \nu$$

[4.6]

where:

$I_{pqc}$ is the input from the V1 blob cells at all positions within the V4 receptive field area (p,q), and of preferred colour c

$X_m$ is the feedback from IT cell populations via weight $B_{W_{ijc}X_m}$, described later

The remaining terms are the same as those in equation 4.4.

The dynamic behaviour of the associated inhibitory pool for colour-selective cell assemblies in V4 is given by:

$$\tau_1 \frac{\delta}{\delta t} W_{ij}^{IC}(t) = -W_{ijc}^{IC}(t) + \lambda \sum_c F\left(W_{ijc}(t)\right) - \mu F\left(W_{ij}^{IC}(t)\right) \qquad [4.7]$$

Parameters take the same values as those in equation 4.5.

Over time, this results in local competition between different colour selective cell assemblies.

### 4.4.2 IT

Anterior areas in IT, such as area TE, are not retinotopic but encode objects in an invariant manner (Wallis & Rolls, 1997) and the IT module here represents such encoding. All objects represented within IT compete with one another via a single inhibitory interneuron pool, as shown in figure 4.5.

The model IT encodes all possible objects, i.e. feature combinations, and receives feedforward feature inputs from V4 with a total response latency of 80ms to reflect normal response latencies (Wallis & Rolls, 1997). V4 inputs to IT are normalised by dividing the total input to each IT assembly by the total number of active (i.e. non-zero) inputs.

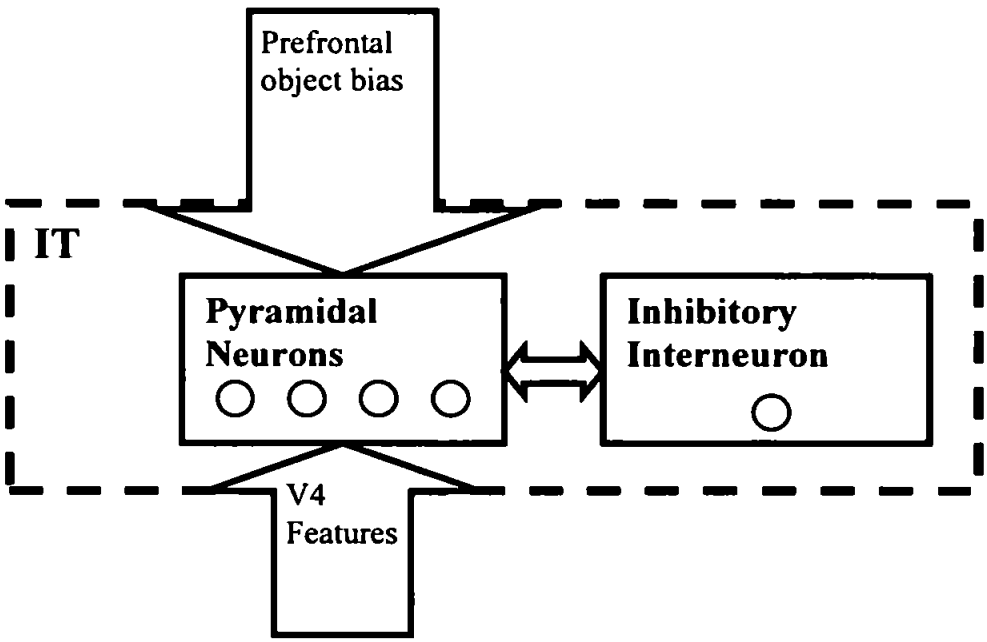



**Figure 4.5 Competition in IT**

IT consists of a pyramidal assembly for each object represented. Each object is represented invariantly. A single inhibitory interneuron mediates competition between objects. This figure also shows the biases that influence competition in the IT module. This figure is a reproduction of figure 2.2, presented in chapter 2, for ease of reference here.

#### 4.4.2.1 IT Feedback to V4

IT also feeds back an object bias to V4. This feedback is diffuse and inhibitory to features not belonging to the object represented in IT. Initially, inhibitory feedback was used in order to avoid illusory activity in V4 due to excitatory feedback from IT in the absence of bottom-up stimulus information. In this case, those assemblies also in receipt of bottom-up information were clearly more active than those only in receipt of top-down input, i.e. illusions were not present in the activity. However, in its initial form, the response function given at equation 4.3 above was not normalised based on activity in the V4 feature layer. Instead, it had a standard threshold that resulted in the problem described in section 4.3 above, where its output for high input activities was indistinguishable from that for low input activity. Therefore, the output from V4 assemblies that were activated by bottom-up stimulus information in addition to a top-down bias was similar to that from assemblies only receiving top-down excitatory input. This resulted in the stimulus locations not necessarily being the most active in LIP. Setting the threshold of the transfer function according to the activity in the layer, as described above, prevented this problem. However, the inhibitory feedback implemented for this reason was then found to produce an inhibitory effect (a suppressive effect after the initial sensory burst and before the onset of object-based effects) in the V4 single cell simulations similar to that observed in monkey cells, which will be described in chapter 5. Therefore, it was decided to retain inhibitory feedback but, in principle,

the model would be equally effective with excitatory feedback to features relating to the object represented in IT.

An experiment by McAdams and Maunsell (2000), who recorded V4 cells in behaving monkeys, suggests that feature-based attention in V4 might produce an inhibitory effect compared to spatial attention alone. During a space-and-feature task, the animal attended a feature at a location outside the cell's receptive field and the feature was different from that in the receptive field. This task produced lower responses than those in a spatial attention task in which the monkey attended the same location outside the receptive field but it contained features identical to those in the receptive field. This indicates that the featural component of the attentional effect involved inhibition of responses. In the context of the model here, this could be interpreted as inhibitory feedback from IT to V4 suppressing features not belonging to the currently attended object. Hence, this supports the current implementation of inhibitory feedback from IT to V4.

An alternative to the diffuse feedback from IT to V4 is for the feedback to be limited in spatial extent in V4, perhaps just within the AW. Lateral connections in V4 could excite neighbouring neurons with similar properties (in a similar manner to that modelled in V1 and V2 by Grossberg & Raizada, 2000) and thus, transmit the effect of the localised feedback across V4. This would result in a similar effect to that achieved by the current model, which does not contain local lateral

connections in V4. The model could be extended to include such connectivity and this will be discussed further in the discussion.

The strength of the IT to V4 feedback connections is given by weights described below, which are set by hand (to −1 or 0, as appropriate, for inhibitory feedback; although the model may also be implemented with excitatory feedback: 0, +1) to represent prior object learning. These simple matrices reflect the type of weights that would be achieved through Hebbian learning without the need for a lengthy learning procedure (such as Deco, 2001, where only one feature type was modelled and 1,000,000 presentations of an object were required to train IT), which is not the aim of this work. Such Hebbian learning is typically performed by presenting the object at certain locations in the scene, but it is doubtful that this provides translation invariant learning (e.g. van der Velde & de Kamps, 2001). This then restricts the object to being presented in the learnt positions of the scene, unless such learning is interpolated, for example by means of a *trace learning rule* (Wallis & Rolls, 1997). Other models of object recognition have shown how complex feature representations can be formed from simpler representations at lower stages of a hierarchy (Riesenhuber & Poggio, 1999, 2000). This is not the focus of the current work.

- The weight of the feedforward connections from V4 assemblies to IT assemblies is given by $A_{X_m W_{iz}}$, where z indicates orientation, k, or colour, c. This weight is used in equation 4.8.

- Similarly, the weight of the feedback connections from IT assemblies to V4 assemblies is given by $B_{W_{iz} X_m}$ and is used in equations 4.4 and 4.6.

The implementation of learning procedures for object knowledge has been avoided here also because psychophysical evidence using the paradigm of preferential learning (Fantz, 1964) in young infants suggests that they might not develop the capacity to learn about objects by viewing them statically. Although some knowledge may be innate (Spelke, Breinlinger, Macomber & Jacobson, 1992), in general young infants are unable to segregate objects on the basis of static Gestalt principles (such as good continuation) that are the norm in adults. Young infants seem to use the principle of common motion to individuate an object and, having seen an object move separately from surrounding objects, are able to segment it from the scene (Arterberry & Yonas, 2000; Bower, 1974; Craton, 1996; Hicks & Richards, 1998; Kellman & Spelke, 1983; Kellman, Gleitman, & Spelke, 1987; Mak & Vera, 1999; Spelke, Breinlinger, Jacobson & Phillips, 1993; Spelke, Kestenbaum, Simons & Wein, 1995). Therefore, the motion signal may be very important in infant object learning and sensitivity to rapid motion is available soon after birth (Kauffman, 1995). In contrast, static Gestalt relations have a protracted development course (Spelke et al., 1993). Featural representation appears to

develop over several post-natal months (Wilcox, 1999), with individuation of objects on the basis of object property and category information possibly being linked with language acquisition towards the end of the first year (Xu, 1999). This may indicate that the motion pathway through middle temporal (MT) and middle superior temporal (MST) areas influences learning of objects, possibly by mediating the development of feedback and lateral connections within the ventral processing stream. Lateral connections in V1 may facilitate responses to collinear movement in MT such that these responses are effectively phase shifted in MT (Chavane, Monier, Bringuier, Baudot, Borg-Graham, Lorenceau & Fregnac, 2000). Hence, MT could perform perceptual grouping through motion and feed back information about grouped objects to the ventral stream. MT feedback to parvocellular V1 arrives before LGN input to these areas, projects onto the same layers in V1 (4B & 6) that provide the feedforward pathway (Lamme, Super & Spekreijse, 1998) and affects the earliest response of V1 (Hupe, James, Girard, Lomber, Payne & Bullier, 2001). Thus, this feedback could potentially influence parvocellular processing and formation of connections. These developmental issues of object learning were considered to be a separate area of work that could be used to enhance the current model in future.

Therefore, it is assumed that object representations within the ventral stream have been previously learned and the matrices above reflect this learning. This "preset" object learning is used to demonstrate the principle of object-based attention in the model. Other models in the field, for example SAIM (Heinke & Humphreys,

2003), use similarly predefined object-related feedback connections. Here, these simple connections provide feedback that is diffuse to all retinotopic locations so that the connections that are active for excitatory feedback (or inactive for inhibitory feedback) are those features relating to the object wherever it appears in the scene. This allows target features to be enhanced and non-target features to be suppressed across V4 (Motter, 1994a,b).

#### 4.4.2.2 Prefrontal Feedback to IT

IT receives feedback from prefrontal cortex that relates to the target object. The purpose of this feedback is to bias the competition between objects in IT in favour of the target object so that the assembly encoding this object wins the competition and other assemblies are suppressed. Such a prefrontal bias to competition in IT was originally implemented in a computational model by Usher and Niebur (1996) and has since been used in other computational models, such as Deco's model (2001; Deco & Lee, 2002; Rolls & Deco, 2002) operating in visual search mode. In contrast to these models, the prefrontal bias here is related to the sensory response in prefrontal cortex, so that it increases in magnitude over time, and operates in an inhibitory manner to non-target assemblies in IT (although similar results may be achieved with excitatory feedback to the target object). This bias was discussed in chapter 2 and its implementation, as a sigmoid function over time, is presented here. The effects of object-based attention at the cellular level in IT are examined in chapter 5. Also, in chapter 5, the prefrontal bias is extended to

include an additional constant excitatory mnemonic component, reflecting memory of the target. This allows the less significant target effects in the early portion of the sensory response in IT to be replicated. The focus of the current chapter is on the sensory-related component of prefrontal feedback, which produces the most significant object-based effect in the late portion of the response in IT.

The pyramidal cell assemblies in IT evolve according to the following dynamics:

$$\tau_1 \frac{\delta}{\delta t} X_m(t) = -X_m(t) + \alpha F(X_m(t)) - \beta F(X'(t)) + \chi \sum_{ijk} A_{X_m W_{ijk}} . F(W_{ijk}(t)) + \chi \sum_{ijc} A_{X_m W_{ijc}} . F(W_{ijc}(t)) + \gamma P_M^v(t) + I_0 + \nu$$

[4.8]

where:

β is the parameter for inhibitory interneuron input, set to 0.01

$W_{ijk}$ is the feedforward input from V4 relating to orientation information, via weight $A_{X_m W_{ijk}}$, described above

$W_{ijc}$ is the feedforward input from V4 relating to colour information, via weight $A_{X_m W_{ijc}}$, described above

χ is the parameter for V4 inputs, set to 2.5

γ is the parameter for the object-related bias, set to 1.2

$P_M^v$ is the object-related feedback current from ventrolateral prefrontal cortex, injected directly into this pool

This feedback is sigmoidal over time as follows:

For the target object:

$P_M^v = 0$

Other objects receive inhibitory feedback as follows:

$P_M^v = -1/(1+\exp(\tau_{sig}-t))$ [4.9]

where t = time (in milliseconds) and

$\tau_{sig}$ is the point in time where the sigmoid reaches half its peak value: Set to 150ms for replication of most single cell simulations and all scan path simulations. However, it is set to 200ms in order to accurately reproduce the onset times of significant object effects and saccades recorded in IT by Chelazzi et al. (1993). This will be described in chapter 5.

The remaining terms and parameters are evident from previous equations.

A single inhibitory interneuron assembly mediates competition between objects in IT. Pyramidal neuronal assemblies representing each object within IT contribute to the activity of the inhibitory interneuron assembly. The dynamic behaviour of the inhibitory assembly in IT is given by:

$$\tau_1 \frac{\delta}{\delta t} X'(t) = -X'(t) + \lambda \sum_m F(X_m(t)) - \mu F(X'(t)) \quad [4.10]$$

where:

$\lambda$ is the parameter for pyramidal cell assembly input, set to 3

$\mu$ is the parameter for inhibitory interneuron input, set to 1

### 4.4.3 LIP

LIP is a retinotopic representation of locations within the retinal image and its purpose is to highlight those locations that might contain a behaviourally relevant stimulus. Competition in LIP, shown in figure 4.6, is spatial and the location that is most active when a saccade is made determines the end point of the saccade, i.e. the next fixation position. LIP receives featural input from V4 that, after the development of object-based attention, serves to increase activity at locations in LIP that contain target features. The nature of this connection is such that V4 colour features are marginally more strongly connected to LIP than V4 orientation features. This means that target coloured locations are most strongly represented in LIP and, thus, the scan path is preferentially attracted to these locations, in order to replicate the effect observed by Motter and Belky (1998b).

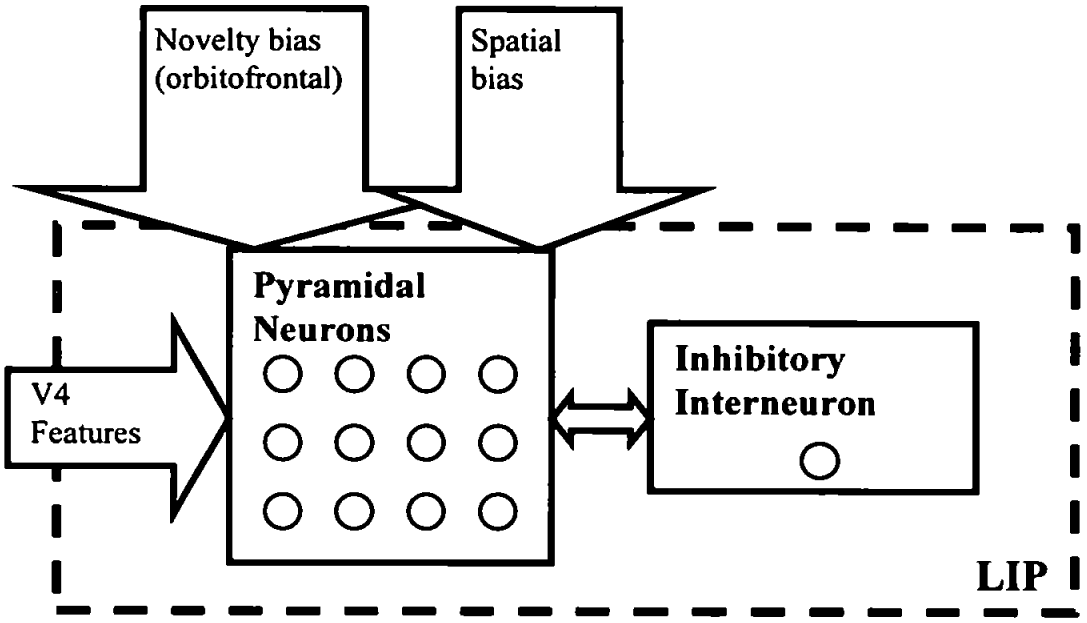



**Figure 4.6 Competition in LIP**

LIP is a retinotopic representation of space. Locations represented within LIP compete with one another. This competition is mediated by a single inhibitory interneuron and is biased by external signals relating to features, novelty and the AW, as shown. This figure is a reproduction of figure 2.3, presented in chapter 2, for ease of reference here.

It was mentioned above that Wolfe's (1994) feature maps might correspond to the representations in V4 here. Under this comparison, Wolfe's *activation map* may be likened to the LIP representation here. Wolfe's activation map was able to reflect both bottom-up and top-down activation for feature conjunctions and the LIP representation here represents bottom-up stimulus locations on the basis of top-down search requirements. In Guided Search 2.0 (Wolfe, 1994), knowledge of task

requirements is allowed to influence the relative contributions of different feature maps to the activation map. This could be interpreted here as allowing top-down task requirements to influence the weight of connections from the various V4 feature types to LIP.

The spatial bias to LIP relating to the AW results in a spatial enhancement in LIP at the currently attended location (at the centre of LIP for overt attention but at any possible location for covert attention). This tends to increase the activity of LIP assemblies whose receptive fields overlap the spatial AW. Input from the novelty map over the area of the receptive field acts to decrease activity for assemblies whose receptive field covers recently fixated locations and, conversely, increase activity for assemblies whose receptive fields are at novel locations. Thus, at the end of the fixation, target coloured locations within the AW not in previously fixated regions will be most active in LIP and, therefore, most likely to be chosen as the next fixation point.

Representation of receptive fields in LIP is simplified here since receptive field profiles are varied by the animal's state of attentional engagement and depend on eccentricity from the central fixation point. The receptive fields could have been modelled according to eccentricity. However, this variance is still a matter of debate with Blatt, Andersen and Stoner (1990) finding a bias to peripheral representation in anaesthetised monkeys but Ben Hamed, Duhamel, Bremmer and Graf (2001) finding an overrepresentation of central locations in awake fixating

monkeys, with a 4.5% decrease in neural representation per degree eccentricity from 0 to 15° eccentricity and a 1.8% decrease from 15° eccentricity. Ben Hamed, Duhamel, Bremmer & Graf, (2002) subsequently found that receptive field characteristics were altered from this central representation during fixation to a more peripheral one during free gaze. The retinal image size in the model here is limited for computational reasons and, therefore, complications resulting from peripheral representations are somewhat alleviated. Also, it is the dorsal part of LIP that is modelled here, because it connects to ventral visual stream areas and areas in the frontal lobe, and dorsal LIP tends to have more central receptive fields (Ben Hamed et al., 2001). Cells with more peripheral receptive fields are found in ventral LIP, which connects to MT and the frontal lobes (Ben Hamed et al., 2001).

Receptive field sizes in LIP are quite diverse, averaging 12.8° wide during fixation, with 67% being less than 15° wide (Ben Hamed et al., 2001). The modelled receptive fields in LIP are the same size as those in V4, to allow simple retinotopic mapping between the regions. An extension to the model could allow V4 inputs over a wider Gaussian area to be applied to LIP. In addition, V1 inputs could be directly projected to LIP for the purposes of fast responses to V1 magnocellular information. This topic is returned to in the final discussion chapter.

The pyramidal cell assemblies in LIP evolve according to the following dynamics:

$$\tau_1 \frac{\delta}{\delta t} Y_{ij}(t) = -Y_{ij}(t) + \alpha F\left(Y_{ij}(t)\right) - \beta F\left(Y'(t)\right) + \chi \sum_k F\left(W_{ijk}(t)\right) + \varepsilon \sum_c F\left(W_{ijc}(t)\right) + \gamma P_{ij}^d(t) + \eta \sum_{pq} Z_{pq} + I_0 + \nu$$

[4.11]

where:

$\beta$ is the parameter for inhibitory input, set to 1

$W_{ijk}$ is the orientation input from V4 for orientation k, at location (i,j)

$W_{ijc}$ is the colour input from V4 for colour c, at location (i,j).

In order to avoid the scan path being preferentially attracted to locations that contain more than one stimulus of different types within the same V4 receptive field (and activate many V4 features at that location), the V4 inputs to LIP are normalised, within each feature type, after being transformed by the response function *F*. This normalisation simply divides the input by the number of V4 inputs that were non-zero. Thus, the average non-zero input is attained. This has little effect after the effects of object-based attention have suppressed the activity of non-target features in V4.

$P_{ij}^d$ is the AW bias, i.e. spatial feedback current from FEF, dorsolateral prefrontal cortex or pulvinar, injected directly into this pool when there is a requirement to attend to this spatial location: When fixation is established, this spatial bias is applied and the spatial AW is formed

$\gamma$ is the parameter for the spatial AW bias, set to 2.5

$Z_{pq}$ is the bias from the area, pq, of the novelty map (which is the size of the original image, N). Area pq represents the size of the LIP receptive field.

$\eta$ is the parameter for the novelty bias, normally set to 0.0009

In order to attract the scan path to target coloured locations:

$\varepsilon > \chi$ so that colour-related input from V4 is stronger than orientation-related input. Normally set to $\chi = 0.8$, $\varepsilon = 4$

The remaining terms are evident from previous equations.

A single inhibitory interneuron assembly mediates competition within LIP. Pyramidal neuronal assemblies at all locations within LIP contribute to the activity of the inhibitory interneuron assembly. This provides competition between locations represented in LIP. The dynamic behaviour of the inhibitory assembly in LIP is given by:

$$\tau_1 \frac{\delta}{\delta t} Y'(t) = -Y'(t) + \lambda \sum_{ij} F\left(Y_{ij}(t)\right) - \mu F\left(Y'(t)\right) \qquad [4.12]$$

where:

$\lambda$ is the parameter for pyramidal cell assembly input, set to 1

$\mu$ is the parameter for inhibitory interneuron input, set to 1

## 4.5 Saccades

Saccades are especially prominent in humans and nonhuman primates, who must ensure that the target falls on the very small fovea in order to see a high-acuity image. Most saccades last less than 100ms (Scudder, Kaneko & Fuchs, 2002). During a saccade there is large-field motion on the retina and the relationship between object position in external space and the image position on the retina changes. How the visual system copes with this and allows us to perceive the world as stable is still largely unclear. The brain needs to be able to ignore the motion on the retina as a result of a saccade and also compensate for the change in retinal image. The first problem appears to be solved by *suppression* of visual sensitivity, particularly of the magnocellular pathway, which dampens the sensation of motion. The second involves the complex adjustments of coordinates, accompanied by a perceptual distortion of visual space during the eye movement, in order to provide the perception of *spatial constancy*. These two related issues are examined in appendix A2.2.

In the model, saccades are instantaneous and this avoids the issue of needing to eliminate the motion sensed during the eye movement. The dynamic cortical areas (IT, LIP and V4) are reset to baseline random values at the start of the subsequent fixation to reflect saccadic suppression of magnocellular, and possibly parvocellular inputs, in addition to the cortical dynamics during the saccade. These cortical areas reflect current retinal inputs in retinotopic coordinates; and issues of

spatial re-mapping and perceptual stability are left to specialised models, as described in appendix A2.2. These issues raise questions about the nature of spatial memory across saccades, which is pertinent to the next section about *inhibition of return*. The automatic triggering of saccades within the model is now described. This allows the model's saccade onset times to mirror those observed in single cell recordings.

### 4.5.1 Timing of Saccade Onset

The single cell recordings of Chelazzi et al. (1993, 1998 & 2001) showed that there was a link between the development of object-based attention and the subsequent onset of a saccade. On average, saccades took place ~70-80ms after a significant target object effect was established in IT and V4. The development of significant object-based effects in IT is used here to initiate saccade preparation. Object effects are deemed to be significant when the most active object assembly in IT is twice as active as its highest competitor. Such a quantitative difference is reasonable when compared to the figures in Chelazzi et al. (1993). Once a significant difference is established, a saccade is initiated 70ms later. This latency is attributed to motor preparation and is consistent with the Chelazzi et al. observations. Saccade onset times reported by Chelazzi et al. (1993, 1998 & 2001) can be replicated using this model, as will be shown in chapter 5.

## 4.6 Inhibition of Return

The *inhibition of return* (IOR) effect refers to a slowing of response time for a target that appears at a previously attended location. An inhibitory after-effect once attention was withdrawn from an area was first reported in 1984 (Posner, Walker, Friedrich & Rafal, 1984). This appears to enable efficient search in complex environments, i.e. facilitates foraging by creating a bias against returning to locations that have been already inspected (Danziger, Kingstone & Snyder, 1998; Klein & MacInnes, 1999). IOR also has a cross-modal component, being capable of being instigated by an auditory rather than visual cue (Schmidt, 1996). IOR develops in early infancy between 3 and 6 months of age (Clohessy, Posner, Rothbart & Vecera, 1991). The more specific term *inhibition of saccade return* (Hooge & Frens, 2000) is sometimes used to describe the tendency for saccades not to return to the same location during active search. However, most authors refer simply to IOR. Hence, IOR is the process by which visual search is biased towards novel rather than previously inspected items. Such an effect may be attributed to oculomotor processes and be linked with the superior colliculus (Sapir, Rafal & Henik, 2002; Sapir, Soroker, Berger & Henik, 1999). Further, it appears to be linked with activity in parietal cortex because parietal lesion patients exhibit an inability to reorient attention (Posner et al., 1984).

Here, it was necessary to implement an IOR mechanism to prevent revisiting locations in the scene during the scan path simulations. Parietal module LIP is

central to this feature of the model. Through its role as a possible saliency map and as an integrator of featural and spatial information (Colby et al., 1996; Gottlieb et al., 1998; Toth & Assad, 2002), LIP could be involved in spatial (Posner et al., 1984), featural (Law, Pratt, & Abrams, 1995) and object-based (Tipper, Driver & Weaver, 1991) IOR. Currently, the model is designed to provide only spatial IOR.

In models of covert attention, where the retina is static, IOR can be achieved by suppression of a previously highly active location, for example by specific IOR input in retinotopic coordinates (e.g. Itti & Koch, 2000; Niebur et al., 2001). Here, if the retina were to be held static, such a process within LIP would lead to colour-based IOR (Law et al., 1995) due to the suppression of the most active locations, which are the target-coloured ones in LIP. However, within an active vision model, such suppression is inadequate and higher-level cognitive and mnemonic processes are required to remember previously visited locations across eye movements when such locations are no longer represented in retinotopic coordinates in cortex.

The issue of whether visual search uses memory is debated, with there being some evidence to suggest that humans use very little memory during search (Gilchrist & Harvey, 2000; Horowitz & Wolfe, 1998, Woodman, Vogel & Luck, 2001) but other authors concluding that memory is used (Dodd, Castel & Pratt, 2003; Zelinsky, Dickinson & Chen, 2003). However, even authors advocating "amnesic search" (Horowitz & Wolfe, 1998) do not preclude the use of higher-level mnemonic strategies. Findlay, Brown and Gilchrist (2001) suggest that information

at every fixation is analysed with no information carried over from the previous fixation (except when fixation duration is so short that two saccades must have been pre-programmed together as a unit) but with the region closest to the fixation point being emphasised, as it is here via the AW. However, they recognise, but specifically exclude, issues of "keeping track" of which locations have been examined. It is this "keeping track" function that is required here in order to prevent re-fixation of locations. Specifically, the competition for attention in LIP must take account of this memory trace. There is evidence that details of the spatial structure of the environment is retained across eye movements and is used to guide eye movements (Aivar, Hayhoe, Chizk & Mruczek, 2005).

Parietal damage is linked to the inability to retain a spatial working memory of searched locations across saccades so that locations are repeatedly re-fixated (Husain, Mannan, Hodgson, Wojciulik, Driver & Kennard, 2001; Shimozaki, Hayhoe, Zelinsky, Weinstein, Merigan, Ballard, 2003). Parietal patients tend to have more fixations per trial and are also not able to make use of preview displays, indicating that they suffer a loss of spatial memory for potential target locations (Shimozaki et al., 2003). Computational modelling has suggested that units with properties similar to those found in parietal areas such as LIP can contribute to visuospatial memory across saccades (Mitchell & Zipser, 2001). Hence, it is plausible that a parietal area such as LIP is involved in inhibition of return across a scene inspected by a series of eye movement, i.e. under the conditions simulated by the scan paths here.

Also, IOR may be influenced by recent event/reward associations linked to orbitofrontal cortex (Hodgson, Mort, Chamberlain, Hutton, O'Neill & Kennard, 2002). Here, the potential reward of a location is considered to be linked to its novelty, i.e. whether a location has previously been visited in the scan path and, if so, how recently. Hence, IOR is implemented within LIP such that competition therein is biased by the "novelty" of each location. This produces a modulation of activity on the basis of the potential reward of the location, as has been found in single cell recordings (Platt & Glimcher, 1999). The source of such a bias could be a frontal area such as orbitofrontal cortex, due to its effect on IOR (Hodgson et al., 2002). Also, Damasio's 'Somatic Marker Hypothesis' (Damasio, 1996) suggests that ventromedial prefrontal cortex may be involved in attributing symbolic rewards to stimuli and actions in order to bias behaviour.

In order to implement IOR here, a mnemonic map of novelty values is constructed in a scene-based (world-based or head-based) co-ordinate frame of reference and converted into retinotopic co-ordinates when used in LIP (coordinate transforms are described in appendix A1.2). This reflects a simplification of co-ordinate transformations that may occur within parietal cortex in areas such as the ventral intraparietal area (VIP), where receptive fields range from retinotopic to head-centred (Bremmer et al., 2001; Duhamel et al., 1997) or area 7a, where neurons respond to both the retinal location of the stimulus and eye position in the orbit (see Colby & Goldberg, 1999, for a review of space representation in parietal cortex).

The re-mapping of the visual world across saccades was discussed in the previous section and appendix A2.2.

Initially, every location in the "*novelty map*" representation of the scene has a high novelty, but, when fixation (and, thus, attention) is removed from an area, all locations that fall within the spatial AW have their novelty values reduced. Hence the region that is inhibited for return is linked to the size of the AW, which is scaled according to stimulus density, as described at paragraph 4.2 above. This allows a smaller region to be inhibited in dense areas of stimuli so that the scan path is not inhibited from investigating these areas in detail, as will be shown for scenes of mixed stimulus density when discussing the scan path results in chapter 6. In the immediate vicinity of the fixation point (the area of highest acuity and, therefore, discrimination ability, in a biological system) novelty is set to the lowest values that gradually increase, in a Gaussian fashion, with distance from the fixation point so that novelty in the periphery of the AW is neutral. Klein and MacInnes (1999) suggested that there is "a gradient of inhibition around a previously attended or fixated region" because they found a monotonic function relating saccadic reaction time to the angular distance from the previous fixation point to the new one.

Novelty is allowed to recover linearly with time, i.e. over subsequent fixations. As attention is withdrawn from a location at the end of a fixation, every value in the novelty map is increased by 0.1 before the novelty update in vicinity of the fixation takes place. This allows IOR to be present at multiple locations, as has been found

empirically (Danziger, Kingstone & Snyder, 1998; Snyder & Kingston, 2001; Tipper, Weaver & Watson, 1996) such that the magnitude of the effect appears to decrease approximately linearly from its largest value at the most recently searched location and with at least five previous locations being affected (Dodd et al., 2003; Irwin & Zelinsky, 2002; Snyder & Kingston, 2000).

The novelty map bias provides effective IOR in the scan path simulations, which will be examined in more depth in chapter 6. Figure 4.7 shows a typical scan path simulation where the target colour is green. The effect of the novelty bias is to allow the scan path to widely explore the scene and prevent repeated revisiting of locations. Findlay et al. (2001) contrasted large *scanning* saccades with small *corrective* saccades, which occurred frequently when fixation landed near targets rather than distractors. It is interesting to note that simulated scan paths contain some (seven are shown in this scan path) short corrective saccades near target coloured stimuli, but these tend not to occur near distractors that do not share the target colour. A plot of the values in the novelty map after this scan path is shown in figure 4.8, which demonstrates the Gaussian increase in novelty value with distance from the fixation point and recovery of novelty during the course of subsequent fixations. A comparison of the novelty map after scan paths through sparse and dense scenes will be made in chapter 6.

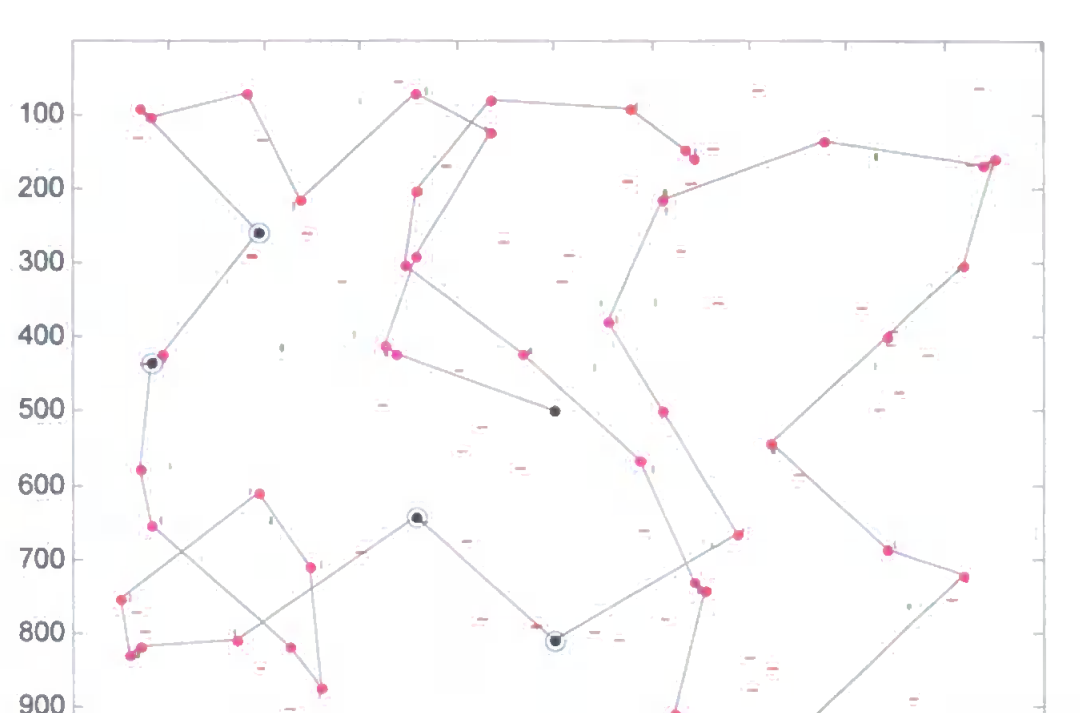



**Figure 4.7 Typical Scan Path**

A typical scan path produced by a model simulation is shown. The majority of fixations land near target coloured (green) stimuli. 92% of fixations here are within 1° of a target coloured stimulus and are shown as magenta dots. 8% of fixations here are within 1° of a non-target colour stimulus and are shown with a blue circle around the fixation point. Average saccade amplitude here is 12.1°.

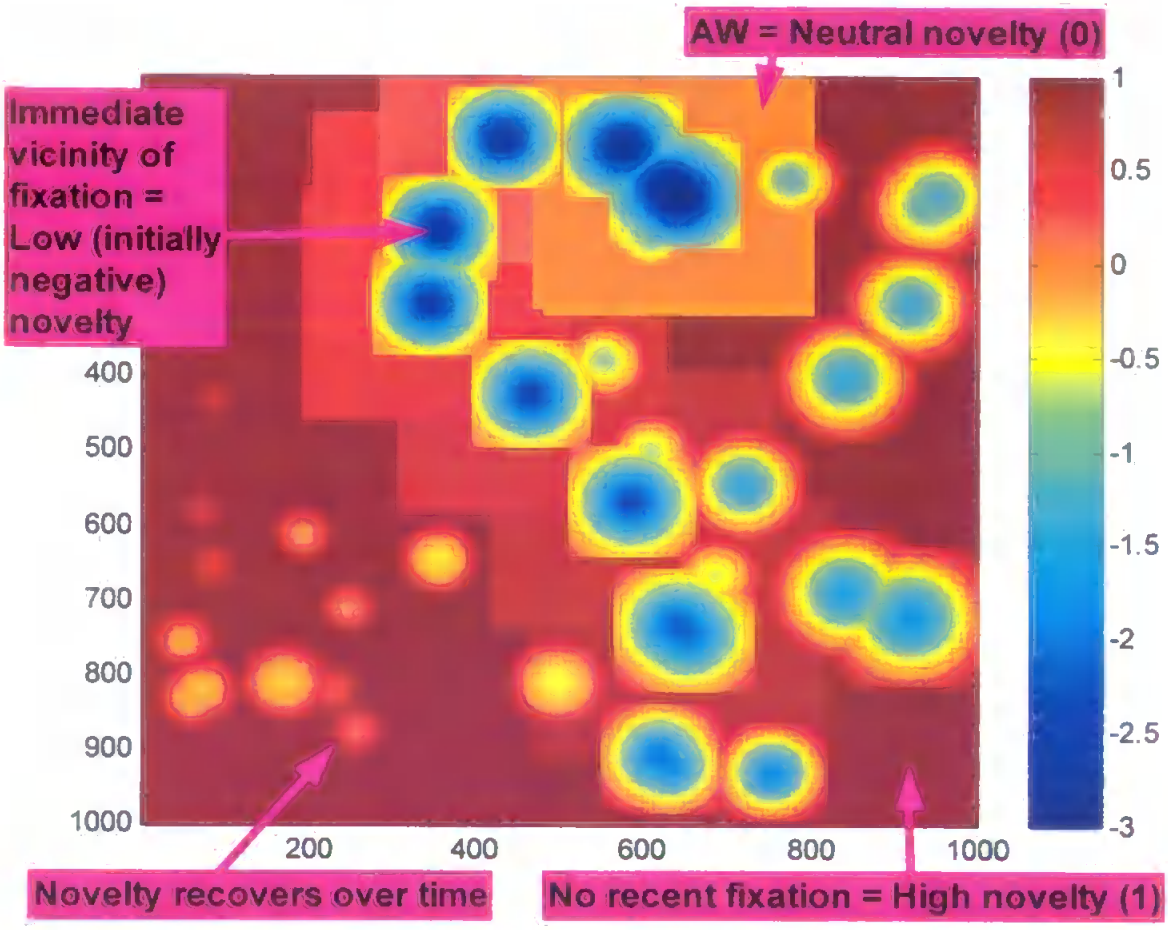



**Figure 4.8 Typical Novelty Map**

Values in the novelty map after the scan path shown in figure 4.7.

## 4.7 Overview of the Output of the System

This section shows output from the system at various time steps during a fixation, in order to provide an overall framework of how the system operates so that more detailed results can be presented in context during the following two chapters.

Figure 4.9 shows the extent of the retinal image and AW for the initial fixation at the centre of the image. The target object is a red vertical bar and there are equal numbers of red horizontal and green vertical distracters. The AW is scaled according to the density of stimuli within the retinal image, as described earlier. During this fixation, the activity in V4, IT and LIP at different times are shown in figures 4.10 to 4.12. These figures give an example of the visual output of the system at each time step (assuming such plots are not suppressed by the user, as mentioned at the start of the chapter). As IT is not retinotopic, its activity is given numerically at the top of the plot. The current position of the retina and AW in the original image is shown in the plot at the top left and this enables the user to follow the scan path around the image as the system moves its focus of attention. In overt mode, both the retina and AW move but in covert mode the retinal image is the same size as the original image and only the AW moves.

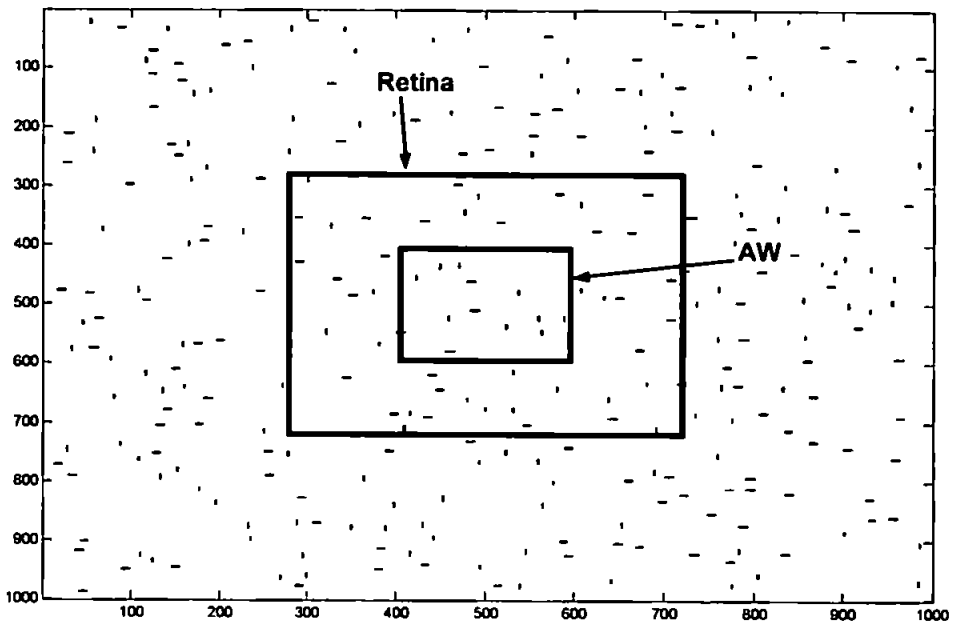



**Figure 4.9 Initial Fixation Retina and AW**

The fixation at the start of a scan path over a scene is shown. The initial fixation point is always at the centre of the image. The extent of the retinal image is shown by the outer of the two plotted boxes. Within the retinal image the initial spatial AW is shown as the inner box.

Figure 4.10 shows that, initially, spatial attention within the AW modulates representations in LIP and V4 and this is apparent even before the onset of the stimulus-related response. Early in the response (figures 4.10 and 4.11) object assemblies in IT are all approximately equally active. Figure 4.12 shows that, later in the response (from ~150ms post-fixation), object-based attention develops and the target object becomes most active in IT, whereas distractor objects are suppressed. At that time, features belonging to the red vertical target are enhanced

at the expense of the non-target features across V4 (Motter 1994a,b). Responses in V4 are modulated by both spatial attention and object/feature attention (Annlo-Vento & Hillyard, 1996; McAdams and Maunsell, 2000). This occurs because V4 is subject to both a spatial bias, from LIP, and an object-related bias, from IT. Both inputs are applied to V4 throughout the temporal processing at each fixation. However, in comparison to the object-based effect, the spatial bias results in an earlier spatial attention effect in V4. This is due to the fact that the object effect is subject to the development of significant object-based attention in IT resulting from the resolution of competition therein.

Once these object-based effects are present in V4, LIP is able to represent the locations that are behaviourally relevant, i.e. the red coloured locations, as possible saccade targets. Prior to the onset of object-based attention all stimuli are approximately equally represented in LIP (see figure 4.11) and a saccade at this time would select target and non-target coloured stimuli with equal probability. Saccade onset is determined by the development of object-based effects in IT. By the time this has occurred later in the response, the target coloured locations in LIP have become more active than non-target coloured locations, as shown in figure 4.12. Therefore, the saccade tends to select a target coloured location. Increased fixation duration has been linked with more selective search (Hooge & Erkelens, 1999) and this would be explained, in this model, by the time taken to develop object-based effects in the ventral stream and convey this information to LIP.

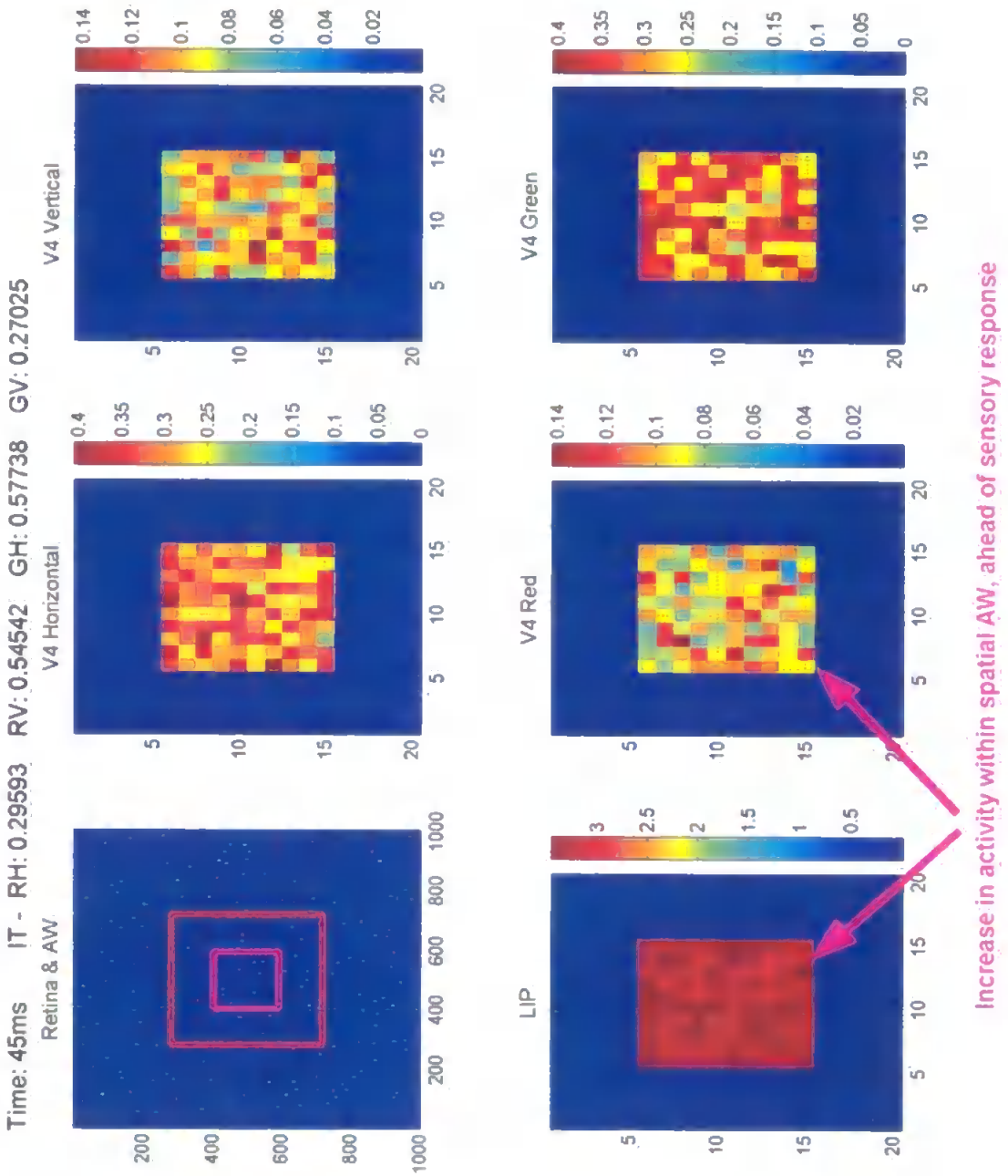



**Figure 4.10 Activity within the cortical areas 45ms after the start of the fixation**
This is prior to the sensory response in V4 (at 60ms). However, there is an anticipatory elevation in activity level within the spatial AW in LIP and V4, as seen in single cell studies of LIP (Colby et al., 1996) and V4 (Luck et al., 1997).

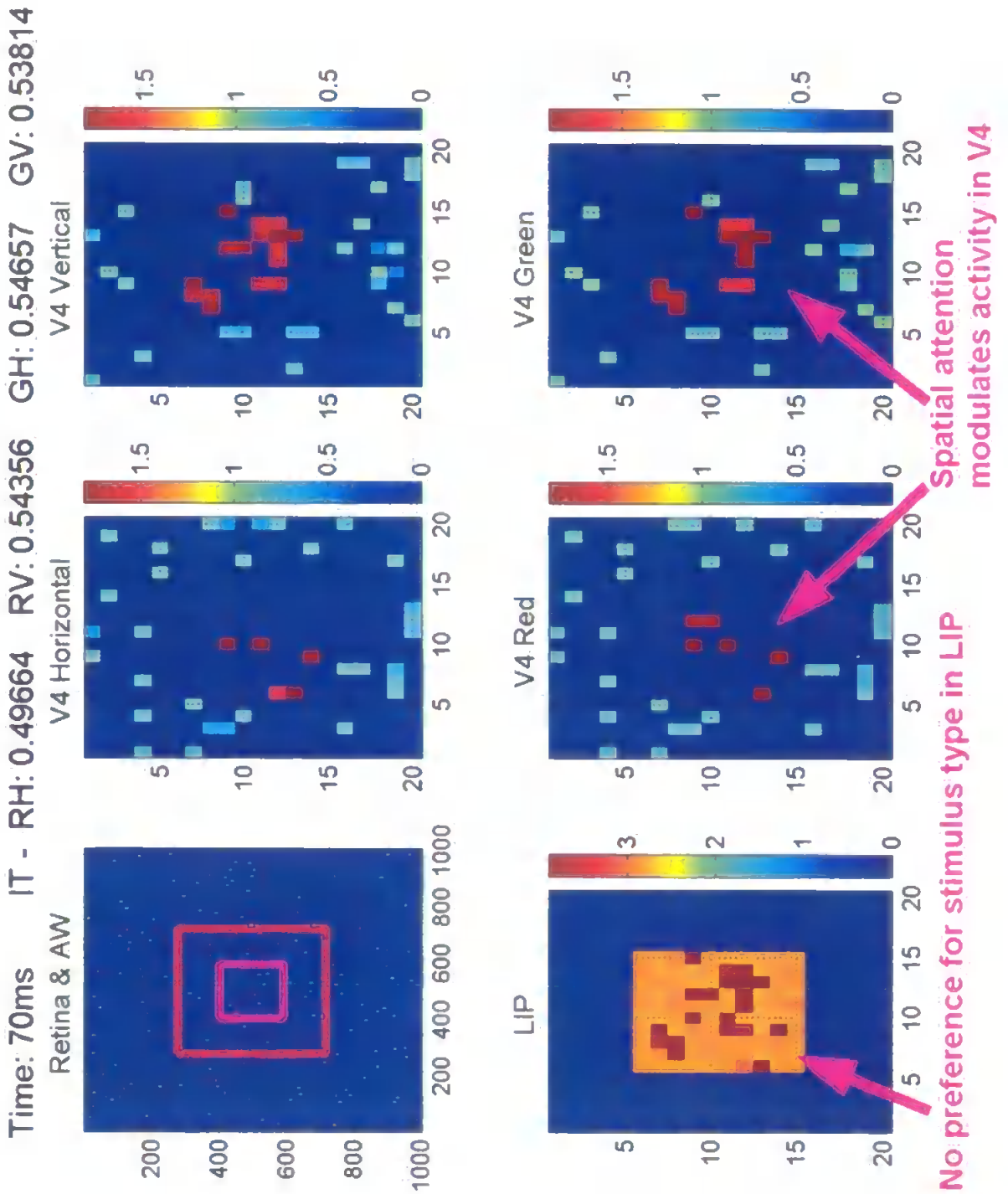



**Figure 4.11 Activity within the cortical areas 70ms after the start of the fixation**
At this time, spatial attention modulates responses in V4 and LIP. LIP represents both red and green stimulus locations.

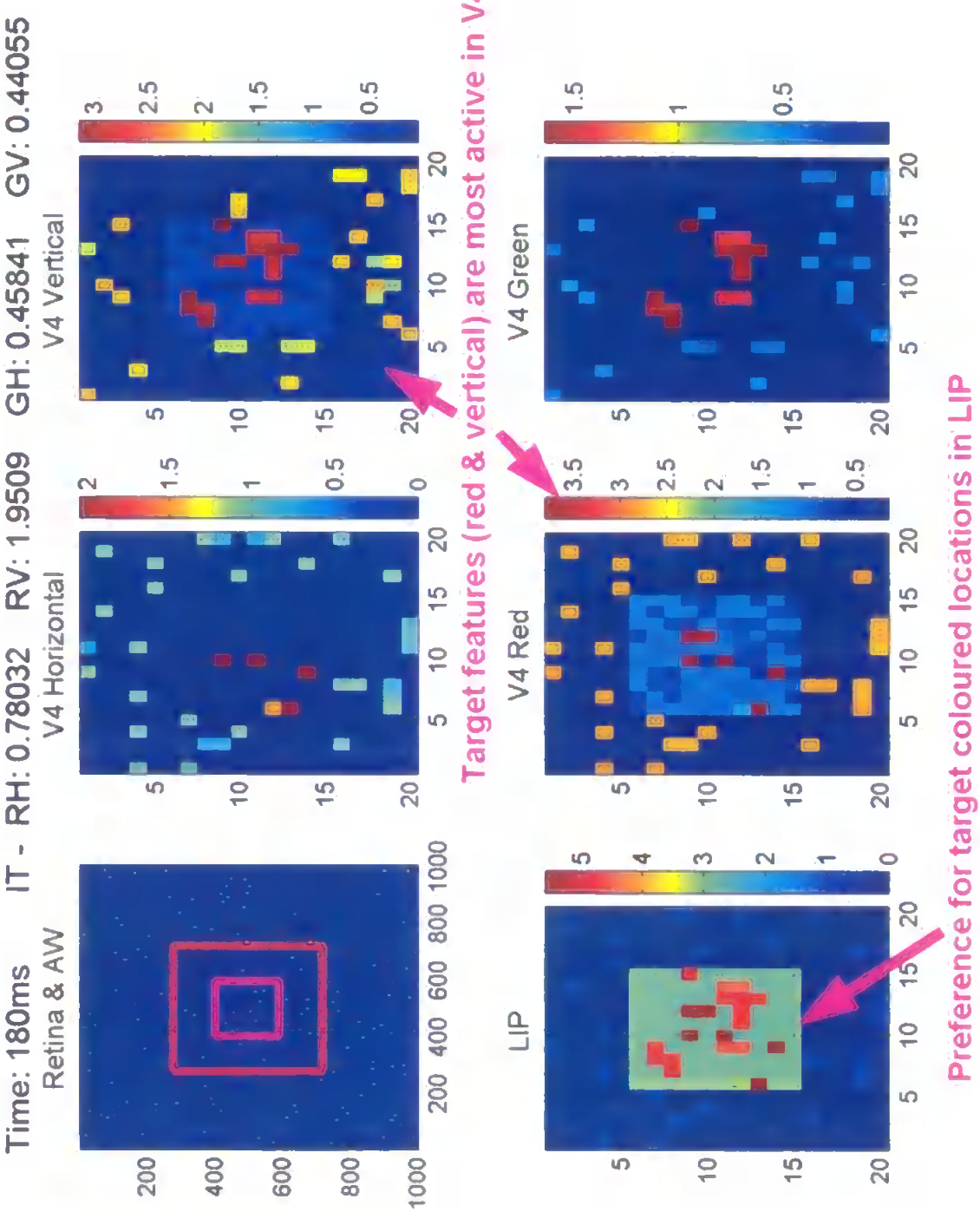



**Figure 4.12 Activity within the cortical areas 180ms after the start of the fixation**

Object-based attention has significantly modulated responses in IT and V4. LIP represents the location of target coloured (red) stimuli more strongly than non-target coloured (green) stimuli. Object-based effects are subject to a spatial enhancement in V4 within the AW.

The importance of local inhibition in V4 was explained earlier in this chapter. This is now demonstrated in figures 4.13 to 4.16, which show activity in the dynamic modules at different times during a fixation where two stimuli fall within the same V4 receptive field and the target is a red vertical bar. Figure 4.13 shows the retinal image and AW formed at this fixation point, which is near the edge of the original image. As the retinal image would extend beyond the edge of the image (a computational artefact not possible in the real world, which is continuous), the cortical areas receive no stimulus information for this "overflow" region, as described in appendix A.1.2.3. This can be seen in the activity in LIP and V4 shown in figures 4.14 and 4.15. Of particular note is that locations in LIP that relate to this "overflow" region are not highly active and, therefore, are unlikely to be selected as the next fixation point.

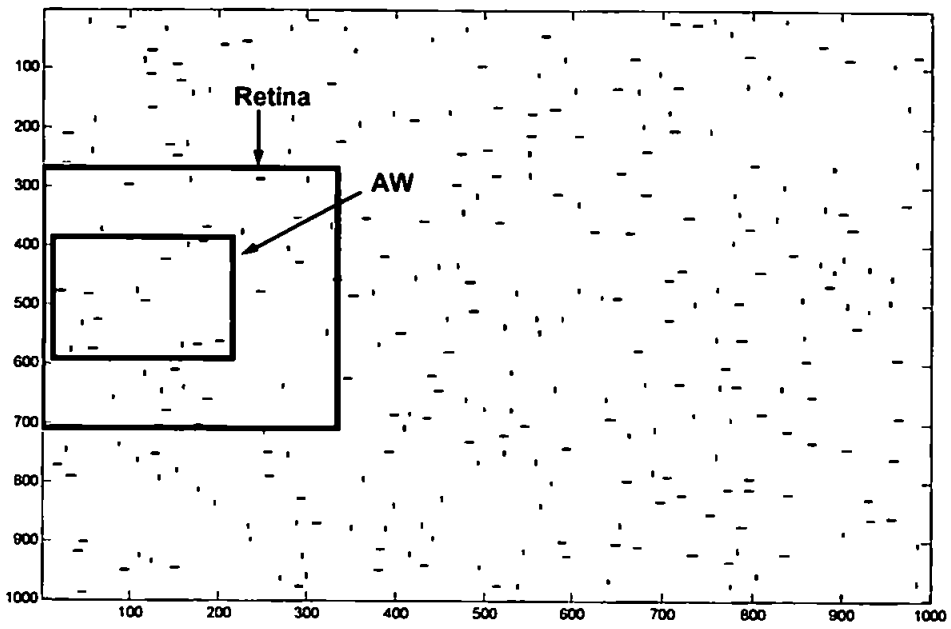



**Figure 4.13 Fixation Near Image Edge**

The retinal image (outer plotted box) and AW (inner plotted box) are shown for a fixation point near the edge of the original image. The retinal image extends beyond the edge of the image so the cortical areas receive no "bottom-up" stimulus information for the area beyond the original image but other cortical processing remains unaltered.

Within the receptive field of the V4 assemblies at (matrix coordinate) location [6,13] both a red horizontal bar and a green vertical bar are present in the retinal image and are within the spatial AW. During the initial stimulus-related response from 60ms post-fixation, the red and the green selective assemblies at this location were approximately equally active to each other and other assemblies in receipt of stimulus information. These are amongst the strongest V4 assemblies at this point because they are in receipt of stimulus information and are within the AW

positively biased by LIP. LIP represents the locations of all strong featural inputs from V4 and has no preference for stimulus colour at this point. However, by 70ms post-stimulus (shown in figure 4.14), strong competitive effects are taking place between both colours and between both orientations at this location because there are two stimuli with different features within the same receptive field. This results in overall responses in each V4 feature assembly at this location being lower than at locations that contain only one stimulus. This is due to there being less competition in the single stimulus case, i.e. the addition of a second stimulus lowers responses of the cell to the preferred stimulus, as found in single cell studies, such as Chelazzi et al. (2001).

By 200ms (shown in figure 4.15), object-based attention has become effective in V4 and the green assembly in V4 has become significantly suppressed compared to the red assembly at the same location. The representation in LIP is now biased towards the representation of the locations of target coloured stimuli and it is one of these locations that will be chosen as the next fixation point (Lanyon & Denham, 2004b).

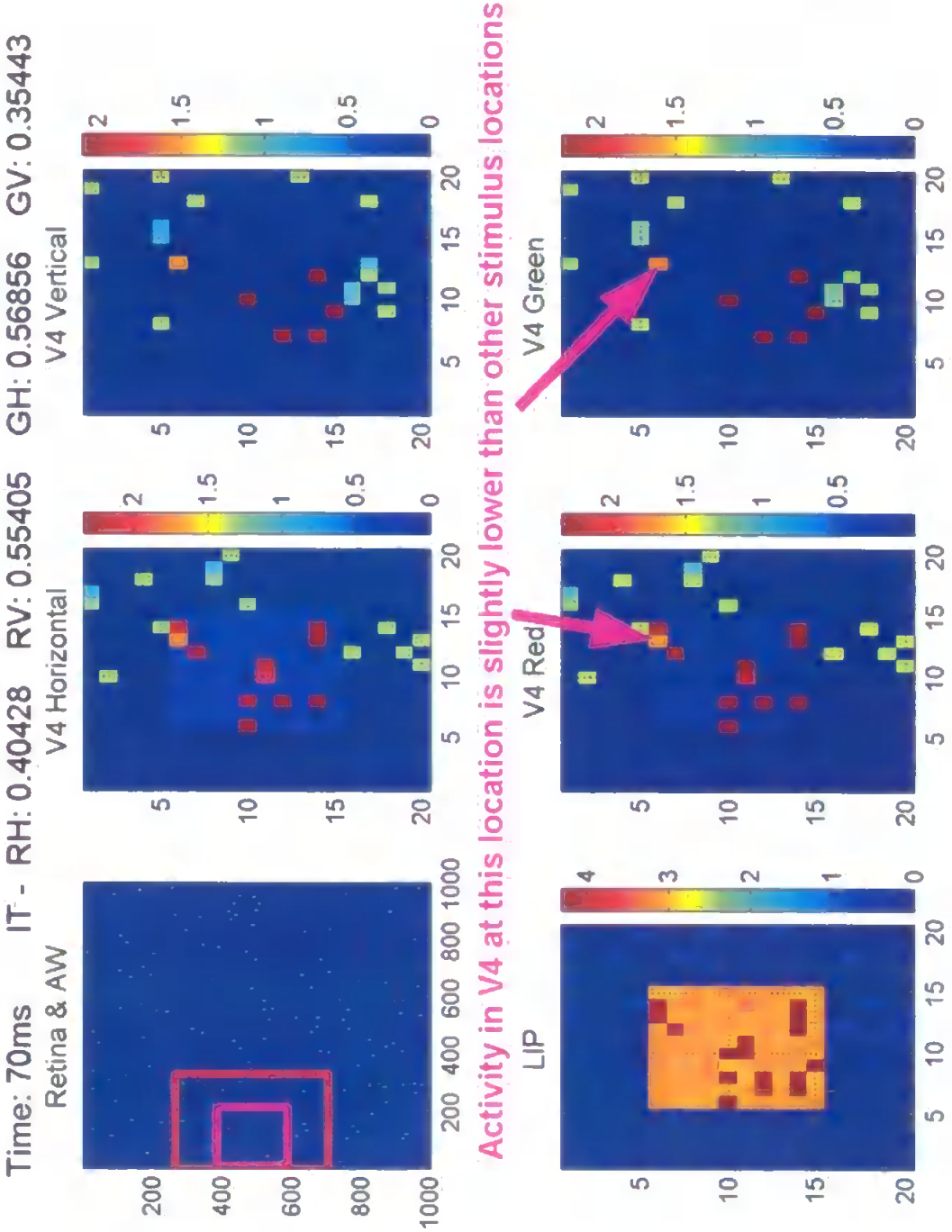



**Figure 4.14 Cortical Activity at 70ms with Two Stimuli in a V4 Receptive Field**

Activity 70ms after the start of the fixation at the position shown in figure 4.13. Within the receptive field of V4 assemblies at matrix coordinate position [6,13] there is both a red horizontal and a green vertical bar. Local competition has caused the activity here to be slightly lower than that of other assemblies that encode the features of a single stimulus within their receptive field. At this time all V4 assemblies are approximately equally active and there is no preference for target features.

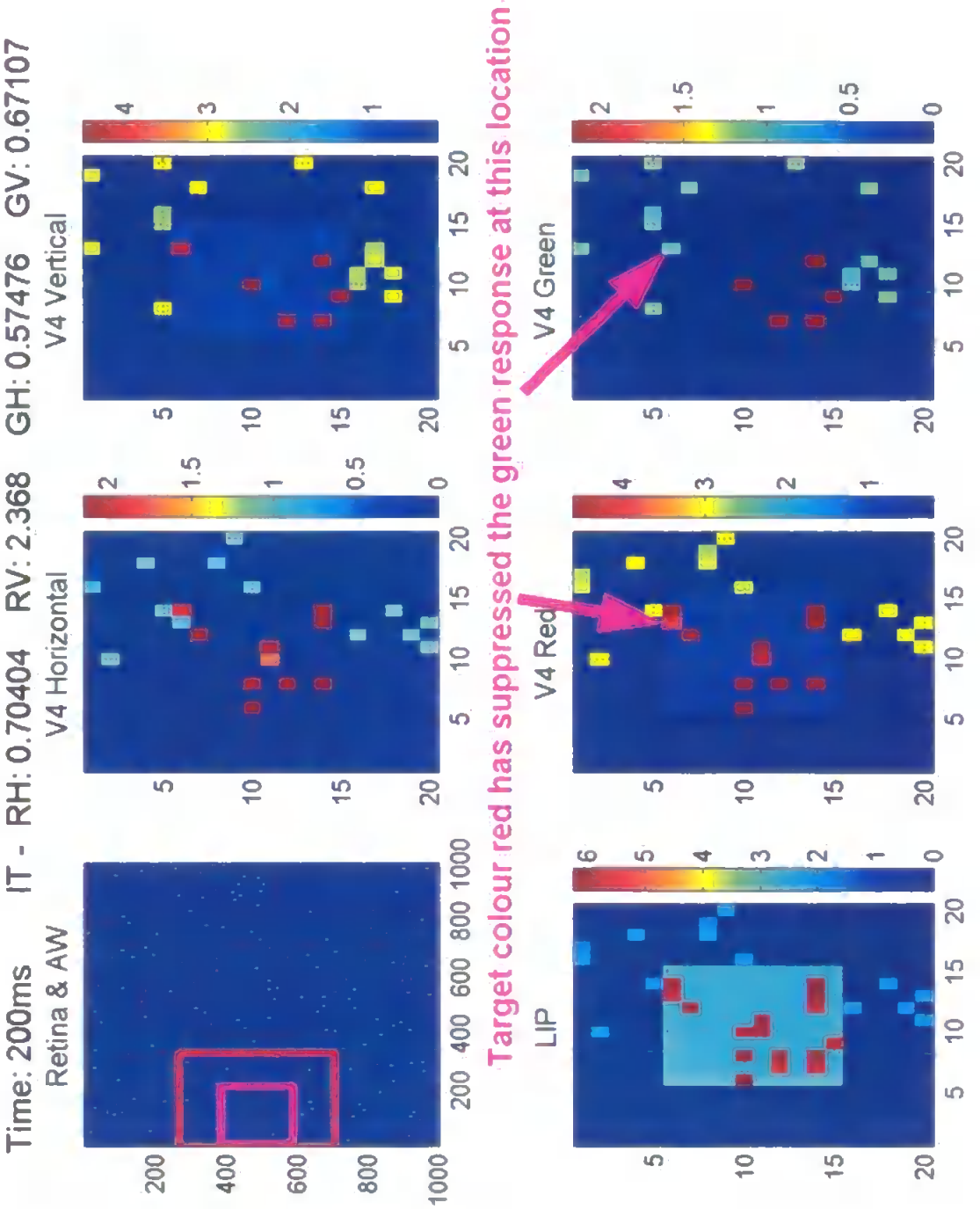



**Figure 4.15 Cortical Activity at 200ms with Two Stimuli in a V4 Receptive Field**

Activity 200ms after the start of the fixation shown in figure 4.13. Object-based attention has become significant and the target feature (red, vertical) assemblies at matrix coordinate position [6,13] have become more active than those for non-target features, which are suppressed, despite each receiving "bottom-up" stimulus information. LIP represents target colour locations most strongly.

Figure 4.16 shows the development of responses in the V4 assemblies at position [6,13]. From 150ms after the fixation begins, the response of assemblies selective for target features (red, vertical) are enhanced and the response of assemblies selective for non-target features (green, horizontal) are suppressed despite the presence of these features within the receptive field (Lanyon & Denham, 2004b).

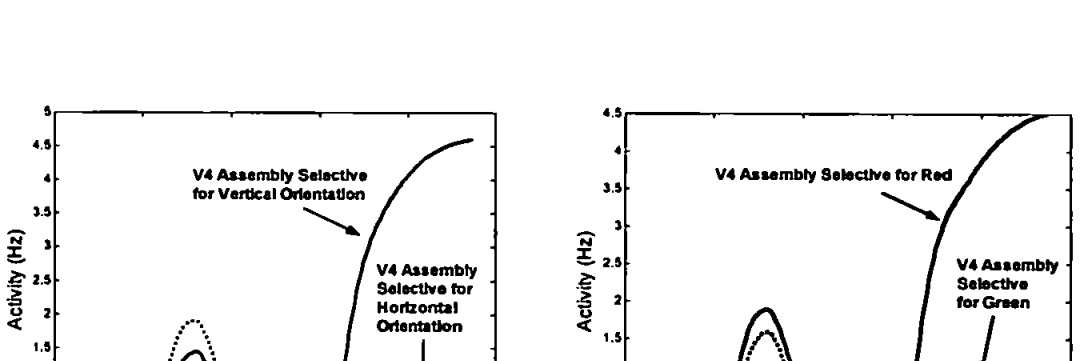



**Figure 4.16 V4 Individual Cells With Two Stimuli in Their Receptive Field**

Shows a plot of the activity of the V4 assemblies over time at position [6,13]. From ~150ms object-based effects suppress the non-target features.

## 4.8 Chapter Summary

The dynamic portion of the model has been described in detail in this chapter. As far as possible, biological justification has been given for all aspects of the model's implementation. Certain aspects of the model's design highlight areas where more information is needed and some predictions were made. For example, the need for a novelty bias to influence the retinotopic spatial competition within LIP suggests that active visual search requires a memory trace of locations visited in non-retinotopic coordinates. At the computational level, normalisation of inputs to each of the dynamic modules was found to be important in sustaining stable performance under different conditions.

In the final section, the behaviour of the system at both the cellular and systems levels was illustrated. Results are examined in more detail in the following chapters. At the cellular level, simulations of single cell recordings in monkey V4 and IT are discussed in chapter 5. Then, the systems level results and an examination of scan path behaviour, and the factors affecting it, are presented in chapter 6.

# Chapter 5

# Cellular Level Behaviour : Simulation of Single Cell Recordings

The cellular level behaviour of the model and a comparison with that recorded from individual monkey neurons are described in this chapter. Attentional effects in V4, IT and LIP are simulated. In particular, object-based attention in IT, and both spatial and object-based attention in area V4 are simulated over the correct time scales. First, the results are presented and, then, there is a discussion of the results in context in the final section.

For simplicity and to enable accurate replication of the monkey recordings within the complex dynamics of the model, the simulations in this chapter are run in monochrome mode. In this mode, colour is not represented and only 2 objects exist in IT: vertical bar and horizontal bar. Similarly, only one feature type (orientation) exists in V4, with two feature settings (vertical and horizontal). The weight of V4 inputs to IT is doubled to compensate for the halve in featural input to IT. Plots of activity in V4 in chromatic mode were shown in figure 4.16 of chapter 4 and this is discussed with the monochromatic simulations in section 5.3.2 below.

## 5.1 Attention in LIP

Attentional effects in LIP can be seen from the plots of activity in the system in figures 4.10 to 4.12 in the previous chapter. Colby et al. (1996) recorded from individual cells in LIP and found that baseline activity was enhanced before the stimulus appeared in cases where a stimulus relevant to behaviour was expected within the cell's receptive field, i.e. when attention was allocated to the spatial locus of the cell's receptive field. Figure 4.10 showed that activity in LIP was increased within the AW, even before the sensory response, due to the requirement to attend at the fixated location during the scan path. Colby et al. also found that responses to a stimulus in the cell's receptive field were higher if the stimulus was relevant to behaviour (in cases where the monkey had to press a bar when the stimulus dimmed as opposed to when no such action was necessary). Here, one of the key aspects of the model is that LIP is able to represent the locations of behaviourally relevant stimuli most strongly, as was described in previous chapters. LIP cell assemblies whose receptive fields contain potential target stimuli are enhanced as a result of object-based attention in the model's ventral stream. Figure 4.12 in the previous chapter showed this effect.

## 5.2 Spatial Attention in V4

Luck et al. (1997) found that spatial attention was able to modulate the earliest stimulus-related response of individual cells in monkey V4, at 60ms after onset of the stimulus, and increased baseline firing rates before this. Using fMRI, Hopfinger et al. (2000) found that activity increased in human extrastriate cortex areas representing the spatial locations of expected target stimuli before the stimuli appeared. Here, these early spatial attention effects are replicated in the model's V4 due to the effect of the AW, as shown in figures 4.10 to 4.12 of chapter 4. The spatial bias applied to V4 from LIP has the effect of increasing responses for V4 cells whose receptive fields overlap the spatial AW from the onset of the stimulus-evoked response and increases baseline responses before this. A plot of activity from a V4 assembly shows these effects in figure 5.1, which simulates the modulatory effects of spatial attention found by Luck et al. (1997). The figure compares the assembly's response when modulated by spatial attention to that when not subject to this effect.

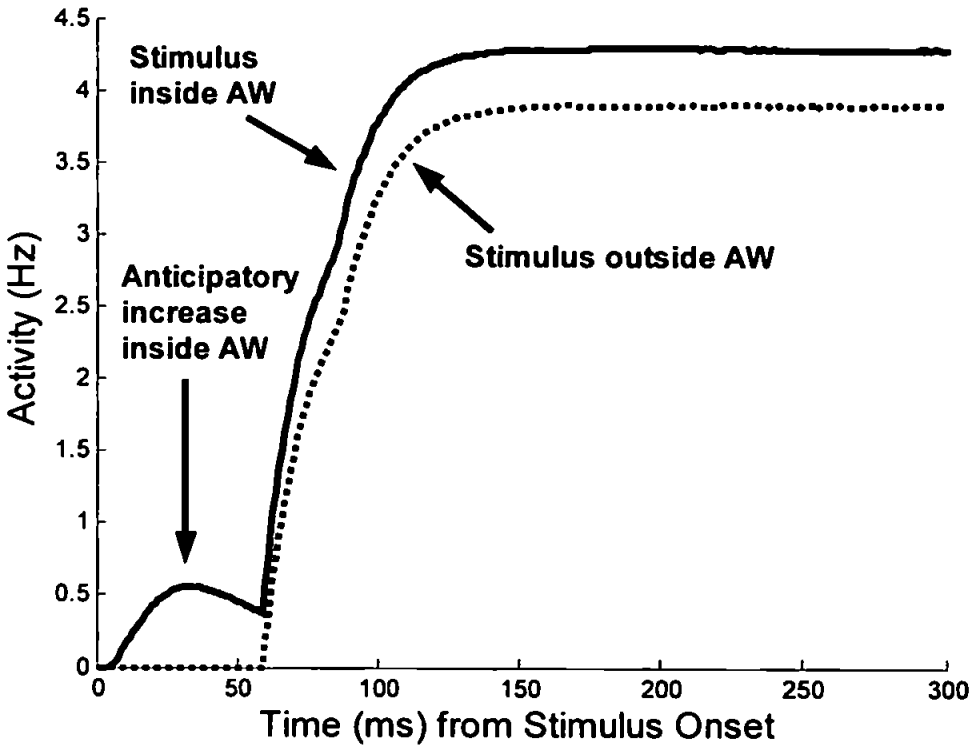



**Figure 5.1 Spatial Attention in V4**

Activity of a V4 assembly selective for the target stimulus presented within its receptive field. The plot shows that, under the same search and stimulus conditions, a cell assembly whose receptive field falls within the initial spatial AW will tend to be more highly active than one whose receptive field is outside the AW. This is due to the facilitating effect of the spatial bias to V4 from LIP. This spatial modulation affects the earliest stimulus-related response at 60ms post-stimulus and there may also be an increase in baseline firing before this, as found by Luck et al. (1997) following a spatial cue.

## 5.3 Object-Based Attention in V4 & IT

Object-based effects take longer than spatial effects to develop and the evidence for this was reviewed in chapter 2. Similar to monkey recordings (Chelazzi et al., 1993, 1998, 2001; Motter, 1994 a, b), object-based attention in the model develops in IT and V4 from ~150ms. The object-based effects in V4 are important in guiding the scan path, as will be seen in the next chapter. These effects are dependent on the resolution of competition between objects in IT. Therefore, object-based effects in IT are examined first below.

A typical stimulus configuration used for the cellular level simulations is shown in figure 5.2. The model is designed to work with stimuli of this type because the scan path simulations replicate psychophysical experiments where such stimuli are searched. The stimuli are less complex than those used by Chelazzi et al. (1993, 1998, 2001) but the principle of these stimuli activating feature and object representations in V4 and IT holds. In the simulations described here, the target is the red vertical bar stimulus.

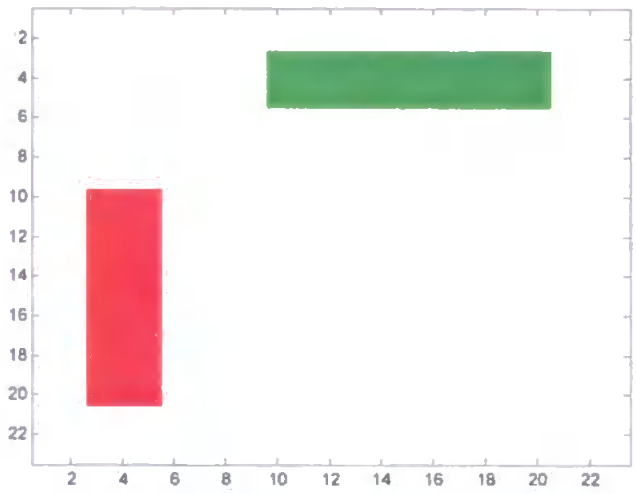



**Figure 5.2 Typical stimulus configuration used for cellular level simulations**

Both stimuli are within the IT and V4 receptive fields.

### 5.3.1 Replication of Object-Based Modulation of Responses in IT

Object-based effects in IT develop late in the cell's sensory response. Responses in the model IT replicate those seen in monkey single cell recordings by Chelazzi et al. (1993, 1998). The average population response recorded in monkey IT during a visual search task is shown in figure 5.3a. This was replicated as shown in figure 5.3b (Lanyon & Denham, 2005a), which plots activity in the model's IT when two stimuli (an effective and an ineffective stimulus for the cell) are present within the cell's receptive field. The initial sensory response, from 80ms post-stimulus until ~150ms, is unaffected by which stimulus is the search target. After about 150-200ms post-stimulus, the cell's response is suppressed if the ineffective stimulus is the search target, even though the effective stimulus is still present in the receptive

field. In contrast, if the effective stimulus is attended, responses remain high. This "target effect" was introduced in chapter 2.

This late object-based effect can be simulated because the prefrontal feedback signal increases over time according to a sigmoid function, as discussed in chapter 2. A similar effect is achieved with a prefrontal bias that increases exponentially, logarithmically, or with a step function such that feedback begins between approximately 130ms to 200ms post-stimulus (to reflect prefrontal response latencies). If the signal is constant over time, target effects begin from the onset of the sensory response in IT and V4. Although this is commonly the way such effects are modelled (e.g. Deco & Lee, 2002; Usher & Niebur, 1996), it does not reflect the temporal aspects of the object-based response observed in monkey cells (Chelazzi et al., 1993, 1998, 2001; Motter, 1994a,b).

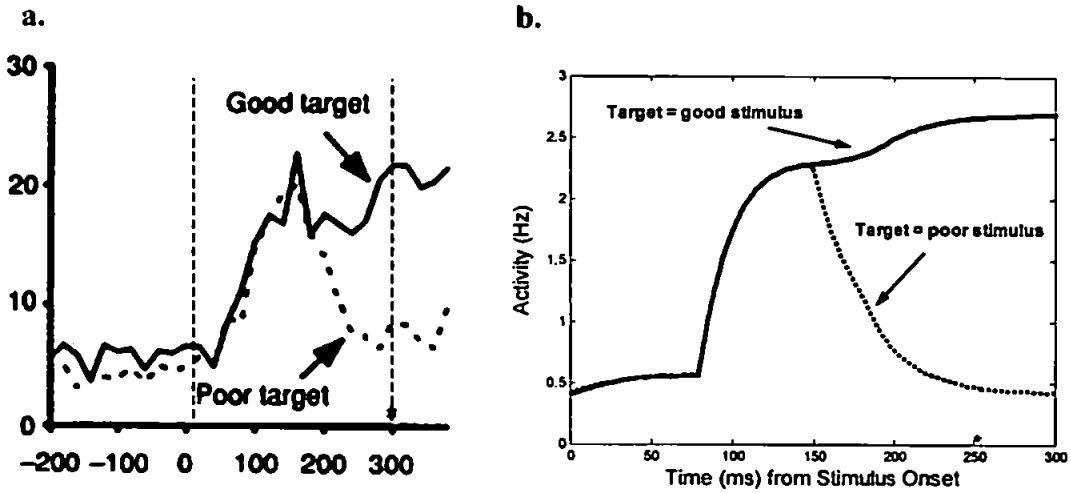



**Figure 5.3 Object-based Attention in IT – Chelazzi et al. (1993)**

**a.** Chelazzi et al. (1993) *figure 3a*: average population response in monkey IT. Shows the object-based modulation of the response that develops late in the response. Two stimuli are present in the IT cell assembly's receptive field: an effective (good) and an ineffective (poor) stimulus for the cell. The solid line shows the response when the search target is the effective stimulus and the dotted line shows the response when the target is the ineffective stimulus. The initial sensory response is independent of the search target but later in the response object-based attentional effects begin. If the target is the ineffective stimulus, the response is significantly suppressed to near baseline levels, whereas, if the target is the effective stimulus, responses remain high. Simulations relate to the period from the array onset until the time of the saccade (indicated by an asterisk) only. The extent of this period is indicated by vertical dashed lines. Figure used with the permission of Nature Publishing Group (http://www.nature.com/nature) and L. Chelazzi.

**b.** Activity in the model IT. Similar to that shown in **(a)**, object-based effects develop late in the response. The target effect becomes significant from 150ms, the time reported by Chelazzi et al. (2001), which will be discussed later. Parameter $\tau_{sig}$, which determines the time course of the prefrontal response in equation 4.9 of chapter 4, is set to 150. Parameter $\eta$, the weight of IT feedback to V4, in equations 4.4 and 4.6 is set to 2.5. For this simulation, saccades have been prevented but would have taken place ~250ms post-stimulus (a typical saccade onset time of 252ms is indicated with an asterisk). Replication of saccade onset times from IT cell recordings is discussed later.

In addition to the large significant target object effect late in the sensory response in IT, the average response of cells shows a smaller target effect in the early part of the response (Chelazzi et al., 1998), as shown in figure 5.4a. The authors suggested that this might be due to a continuation of elevated firing during the delay period. This can be simulated here by adding a constant excitatory prefrontal bias to IT throughout its response, as follows:

For the target object:

$$P_M^{vBackground} = 0.1 \qquad [5.1]$$

Otherwise:

$$P_M^{vBackground} = 0$$

This is added to the equation for the dynamics of IT pyramidal cell assemblies, given in equation 4.8 of chapter 4, so that it becomes:

$$\tau_1 \frac{\delta}{\delta t} X_m(t) = -X_m(t) + \alpha F(X_m(t)) - \beta F(X'(t)) + \chi \sum_{ijk} A_{X\_JW_{ijk}} . F(W_{ijk}(t)) + \chi \sum_{ijc} A_{X\_JW_{ijc}} . F(W_{ijc}(t)) + \gamma P_M^v(t) + \gamma P_M^{vBackground}(t) + I_0 + \nu \qquad [5.2]$$

All other aspects of the equation remain unchanged. For all other simulations, except where stated, this additional bias is omitted.

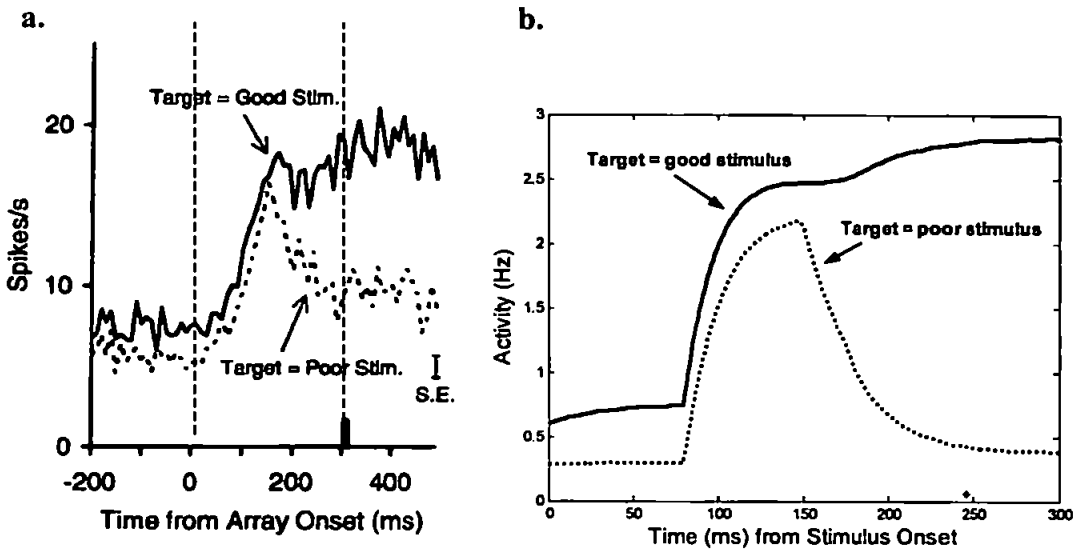



**Figure 5.4 Object-based Attention in IT – Chelazzi et al. (1998)**

**a.** Typical population response in monkey IT reported by Chelazzi et al. (1998; *figure 7a*) using the same experimental conditions as Chelazzi et al. (1993), shown in figure 5.3 above. Only a slight target effect is present in the early part of the response and this effect appears to be a continuation of elevated firing during delay intervals. The more significant target effect begins ~150-200ms after array onset. Simulations relate to the period from the array onset until the time of the saccade only. The extent of this period is indicated by vertical dashed lines. Figure used with the permission of the Journal of Neurophysiology (The American Physiological Society) and L. Chelazzi.

**b.** The same simulation as figure 5.3b but with an additional sustained excitatory component in the prefrontal bias to the model IT. This provides better replication of **(a)** than the simulation shown in figure 5.3b. In addition to the significant object-based effect late in the IT response (from ~150ms), there is a small effect in the early part of the response. Saccades have been suppressed in this simulation but would have occurred just before 250ms post-stimulus (as indicated by the asterisk at a typical latency of 246ms).

The additional prefrontal bias represents the sustained mnemonic part of the prefrontal response and results in an increase in baseline activity for the target object in IT. Figure 5.4b shows a simulation of the same situation as figure 5.3b but with the inclusion of this sustained bias (Lanyon & Denham, 2005a). The early part of the response shows a slight enhancement when the target object is the effective stimulus for the cell but the most significant target effect still appears later in the response, as found by Chelazzi et al. (1998). This suggests that the early and late effects observed in these monkeys may be due to a bias signal with more than one component. The early effect, a possible continuation of elevated firing seen during delay periods (Chelazzi et al., 1998), is likely to result from a small but sustained mnemonic bias from prefrontal cortex. The more significant effect later in the response requires a bias signal whose source has target-related activity that builds during the early sensory response following the onset of the search array. Such activity has been observed in prefrontal cortex (Everling et al., 2002; Rainer et al., 1998).

A minority of IT cells, for example the cell shown in figure 5.5a, do show a significant target effect in the early part of their response (Chelazzi et al., 1998). This can be simulated here by allowing prefrontal feedback to IT to become strong earlier, as simulated in figure 5.5b, where the sigmoidal response function is "shifted" to reflect an earlier sensory response in prefrontal cortex. This indicates that these IT cells may preferentially receive earlier prefrontal feedback, perhaps being connected to prefrontal cells with shorter sensory response latencies.

Prefrontal cortex receives inputs from various cortical and thalamic routes (Miller & Asaad, 2002) that have differing latencies.

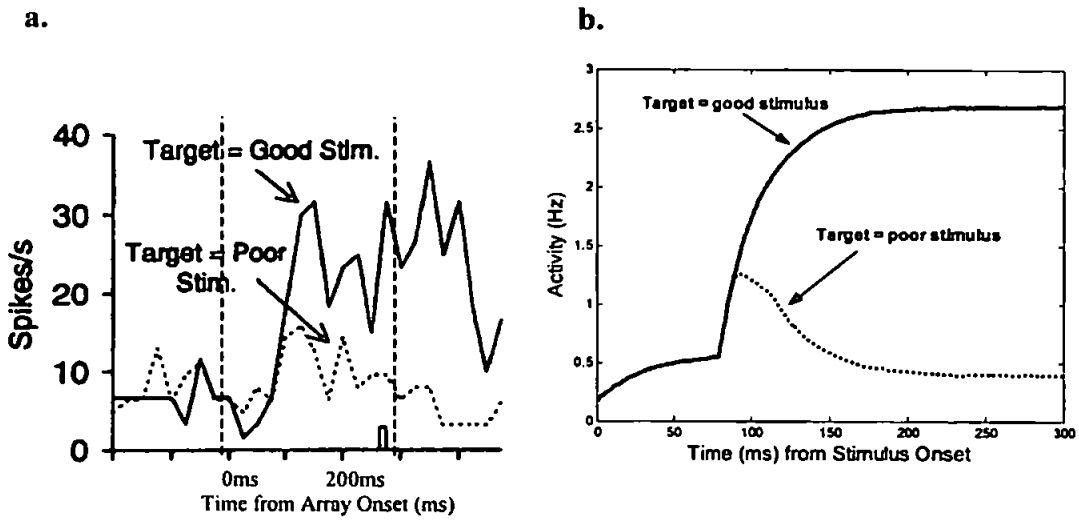



**Figure 5.5 Object-based Attention in IT – Earlier Effect**

**a.** Response of an individual cell, reproduced from Chelazzi et al. (1998; *figure 8b*). The target object modulated this cell's early response, from ~100ms after the onset of the array. An early significant target object effect was observed in only a minority of monkey IT cells. Simulations relate to the period from the array onset until the time of the saccade only. The extent of this period is indicated by vertical dashed lines. Figure used with the permission of the Journal of Neurophysiology (The American Physiological Society) and L. Chelazzi.

**b.** The same simulation as figure 5.3b but with prefrontal feedback to IT available earlier (parameter $\tau_{sig}$ in equation 4.9 of chapter 4 set to 90). This produces a more significant target effect in the early part of the response.

Chelazzi et al. (1998) *figure12a* is reproduced in figure 5.6.a in order to compare the response in IT when two stimuli are placed within the receptive field with that produced when the individual stimuli are alone in the receptive field. The initial sensory response in the two-stimulus case is slightly less than that when the good stimulus is presented alone in the receptive field. This is due to the competing influence of the poor stimulus also present in the receptive field. Once object-based attention effects become significant in the late portion of the response, the cell's response in the two-stimulus case when the good stimulus is the target approaches that in the single good stimulus case. However, if the poor stimulus is the target, the cell's response is suppressed almost to the level of its response to the poor stimulus alone, despite the presence of the good stimulus in the receptive field. This is the same target effect already presented in the figures above, where the effect of object-based attention in the two-stimulus case is to filter out the effect of the unattended stimulus in the cell's receptive field. Here, the resulting responses can be compared to those when the individual stimuli are presented.

This behaviour was replicated, as shown in figures 5.6b and 5.6c (Lanyon & Denham, 2005b). Two simulations are shown in order to illustrate the effects with and without the sustained mnemonic prefrontal bias used for figure 5.4. Figure 5.6b is an extension of figure 5.3, where the prefrontal bias is that described in chapter 4 and does not include the additional sustained excitatory component. Figure 5.6c is an extension of figure 5.4, where the prefrontal bias includes the additional sustained excitatory component. Both simulations capture the two

important observations of Chelazzi et al. (1998). The first is that the two-stimulus cases result in initial sensory responses that are not quite as high as the good stimulus presented alone case, due to the competing influence of the second stimulus. The second is that, following the development of object-based attention, the responses approach those that would be elicited by the target stimulus presented alone. Like figure 5.4, figure 5.6c also captures the slight early target effect in the two-stimulus case. The inhibitory prefrontal feedback to IT (resulting from the sensory sigmoidal bias) is possibly a little too strongly tuned for the single stimulus case (where IT receives less bottom-up stimulation) because the IT response in the case of the poor stimulus presented alone is too suppressed after the onset of the target effect at ~150ms. This could be due to the lack of a full sensory prefrontal module and is not a serious problem with the model.

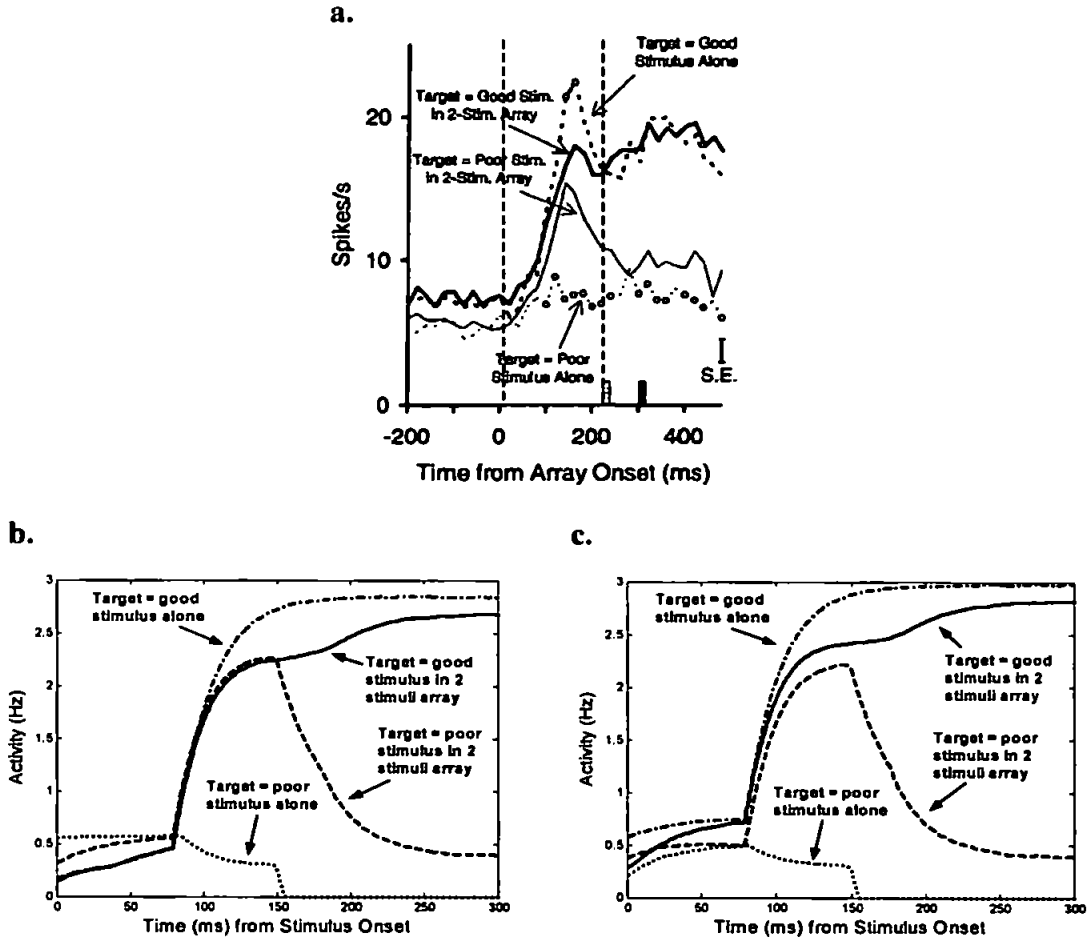



**Figure 5.6 Object-based Attention in IT – Comparison Plot**

**a.** Population response in IT comparing the response to a two-stimulus array with the response to the component stimuli presented alone. Reproduced from Chelazzi et al. (1998; *figure 12a*). Simulations relate to the period from the array onset until the time of the saccade only. The extent of this period is broadly indicated by vertical dashed lines. (Average saccade onset times are indicated by the stippled vertical bar in the single stimulus case, and by the solid vertical bar in the two-stimulus case.) Figure used with the permission of the Journal of Neurophysiology (The American Physiological Society) and L. Chelazzi.

**b.** Activity in a model IT assembly in the same cases as in **(a)**. When an effective stimulus for the cell (i.e. one that causes a high response) is presented alone in the receptive field, a high response is attained: Top (dotted dashed) line. When an ineffective stimulus for the cell is presented alone the cell's response is suppressed: Bottom (dotted) line. When both stimuli are simultaneously presented within the receptive field, the high response normally invoked by the effective stimulus is reduced due to the competing presence of the ineffective stimulus. However, once object-based effects begin from ~150ms the response is determined by which of the two stimuli is the target. The activity then tends towards that of the single stimulus case for the target object. The two-stimulus cases correspond to figure 5.3. Parameters are the same as those used for figure 5.3.

**c.** The same simulation as **(b)** but with the additional sustained prefrontal bias used for figure 5.4. There is an additional small early target effect. The two-stimulus cases corresponds figure 5.4. Parameters are the same as those used for figure 5.4.

The simulations above show that the model is very successful at capturing the biased competition attentional effects at the cellular level in IT. The two-stimulus simulations in this section replicated data from Chelazzi et al. (1993, 1998) when two stimuli were placed together in the contralateral field. Chelazzi et al. (1998) found that there was very little effect of attention on the population responses when one stimulus was placed in the contralateral field and the other in the ipsilateral field. In this case, the response was close to that for the contralateral stimulus presented alone, regardless of which stimulus was attended. This suggests that the contralateral stimulus has a strong competitive advantage that is not easily overcome by an attentional bias. The authors suggest that selection of a stimulus in one hemifield as opposed to the other may take place in higher areas of cortex, for example prefrontal cortex, or in the frontal eye fields, if it is related to the preparation for movement. Hemispheric differences are not modelled in IT here and the simulations presented in this section replicated the most commonly cited attentional effects from these papers, i.e. those where stimuli were placed in the same hemifield.

These effects in IT have an important effect on responses in V4, due to the object-related feedback from IT to V4. Target object effects in V4 are now examined. These effects at the cellular level in V4 are important in guiding the search scan path at the systems level.

### 5.3.2 Replication of Object-Based Modulation of Responses in V4

Object-based attention in V4 produces similar effects at the cellular level to those in IT. Thus, figures presented here will look similar to those just presented for IT. However, the nature of the representation in V4 is different from that in IT because V4 encodes features in a retinotopic fashion, rather than object representations that are invariant to spatial translation. Feedback from IT results in target features being enhanced, and non-target features being suppressed, in parallel across V4. This is crucial for driving the search scan path towards these behaviourally relevant locations, because V4 provides LIP with a spatio-featural bias.

Chelazzi et al. (2001) found that object-based attention developed from 150ms after the onset of the search array in V4, as shown in figure 5.7a. Figure 5.7b shows output from the model's V4 simulating this effect. Similar to that IT, the target object effect becomes significant from around 150ms after onset of the search array. However, in contrast to that in IT, there is no target effect found in the early part of the V4 response. When two stimuli are presented together in the V4 cell's receptive field, initial responses are the same regardless of which is the search target. However, from ~150ms, the response depends on which of the two objects is the target. If the effective stimulus is the target, the response remains high but, if the ineffective stimulus is the target, the response is strongly suppressed. Simulated saccade onset times, of around 240ms after onset of the array, are very similar to the average saccade onset time of 237ms found by Chelazzi et al. (2001).

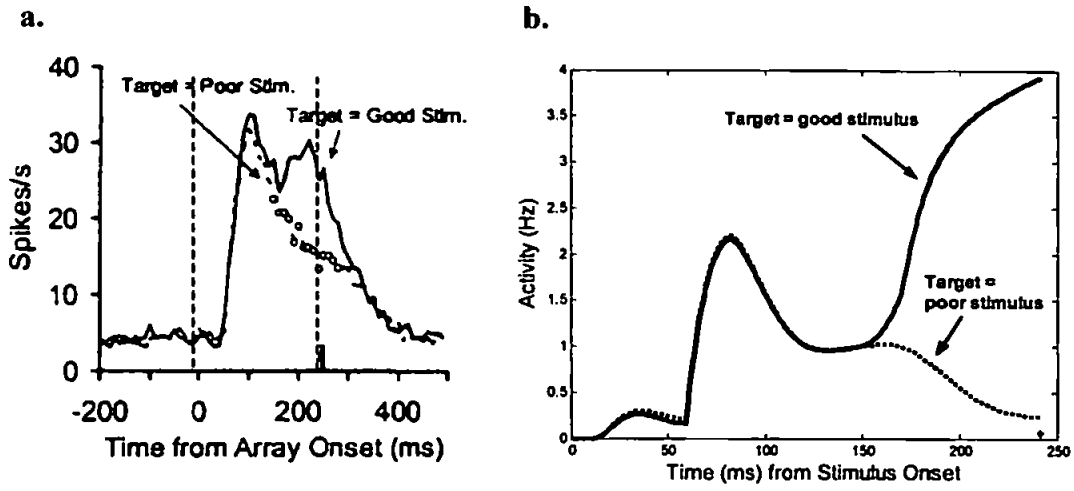



**Figure 5.7 Object-based Attention in V4**

**a.** Population response in monkey V4 from Chelazzi et al. (2001; *figure 3a*). Shows the object-based modulation of the response that develops late in the response. Average saccade latency is 237ms and is indicated by a short solid vertical bar. Simulations relate to the period from the array onset until the time of the saccade only. The extent of this period is indicated by vertical dashed lines. Figure from Chelazzi et al. 'Responses of neurons in macaque area V4 during memory-guided visual search' Cerebral Cortex, 11, 761-772, by permission of Oxford University Press and L. Chelazzi.

**b.** Response of a V4 assembly in the model when two stimuli (a good and a poor stimulus for the cell) are presented in the receptive field and the target is either the good or the poor stimulus, similar to **(a)**. Object-based effects develop from ~150ms and the saccade takes place at ~240ms (occurring here at 241ms, as indicated by an asterisk and cessation of the plotted lines). The parameter $\tau_{sig}$, which determines the time course of the prefrontal response in equation 4.9 of chapter 4, is set to 150, and the parameter $\eta$ in equations 4.4 and 4.6 is set to 5 for strong feedback from IT to V4. A lower value of the parameter $\eta$ still produces the target effect but causes less divergence in the plotted lines, i.e. a less significant object effect in V4.

Figure 5.8a shows a similar plot from Chelazzi et al. (2001, *figure 5*) but, this time, the activity when two stimuli are present in the receptive field is compared to that when these individual stimuli are presented alone in the receptive field. This is replicated in figure 5.8a (Lanyon & Denham, 2005b). In the two-stimulus cases, responses are suppressed compared to the effective stimulus being presented alone. This is the same outcome as was found in IT, shown in figure 5.6 above, and is again due to the competing influence of the second stimulus. Once object-based attention develops, responses in the two-stimulus cases tend towards those when the attended stimulus is presented alone.

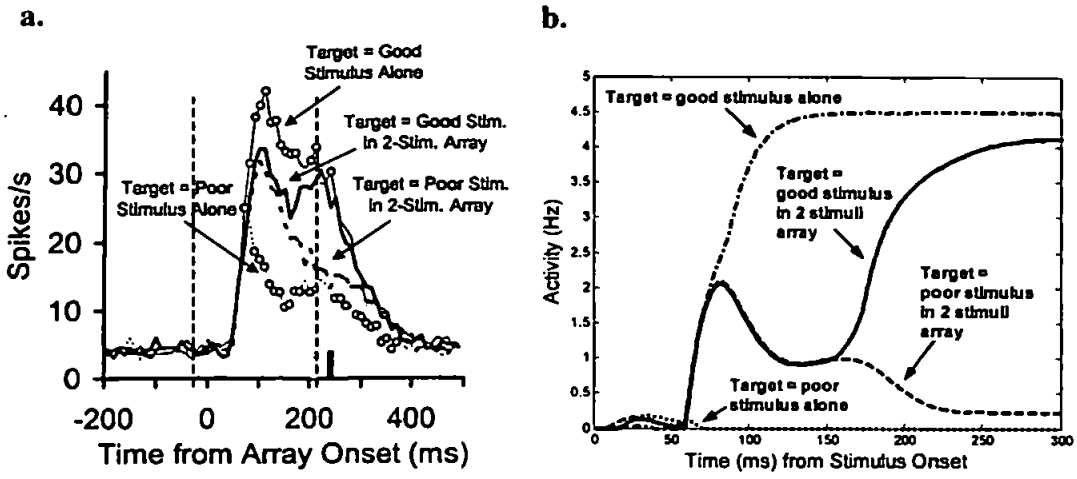



**Figure 5.8 Object-based Attention in V4 – Comparison Plot**

**a.** Population response in V4 comparing the response to a two-stimulus array with the response to the component stimuli presented alone, from Chelazzi et al. (2001; *figure 5*). Average saccade latency in the single stimulus and two stimuli conditions is shown by a vertical stippled (at 210ms) and solid bar (at 237ms) respectively. Figure from Chelazzi et al. 'Responses of neurons in macaque area V4 during memory-guided visual search' Cerebral Cortex, 11, 761-772, by permission of Oxford University Press and L. Chelazzi.

**b.** Shows activity in a model V4 assembly in four cases, to replicate Chelazzi et al. (2001) *figure 5* in the interval from array onset until the time of the saccade only. When a good stimulus for the cell is presented alone in the receptive field, a high response is attained: Top (dotted dashed) line. When a poor stimulus for the cell is presented alone the cell's response is suppressed: Bottom (dotted) line. When both stimuli are simultaneously presented within the receptive field, the high response normally invoked by the good stimulus is reduced due to the competing presence of the poor stimulus. However, once object-based effects begin from ~150ms the response is determined by which of the two stimuli is the target. The activity then tends towards that of the single stimulus case for the target object. The two-stimulus cases correspond to figure 5.7. Here saccades have been suppressed but, for the two-stimulus cases, would have occurred at ~240ms post-stimulus, as shown in figure 5.7. Parameters are the same as those used for figure 5.7.

The simulation of V4 data for two stimuli within the receptive field (figures 5.7b and 5.8b) shows that, prior to the onset of the object effect at ~150ms, there is a period of suppression of the response both when the target stimulus is an effective one for the cell assembly and when it is ineffective for the assembly. Such suppression mirrors that in the plots from Chelazzi et al. (2001). In the model here, the suppression is due to the feedback from IT being inhibitory. During this period all objects in IT that are in receipt of stimulus-related information from V4 are approximately equally active. Thus, all features in V4 are suppressed by the object-related feedback from IT. Once the target object begins to win the competition in IT, V4 cell assemblies representing its constituent features are released from inhibition whilst the assemblies representing non-target features continue to be suppressed. If feedback from IT to V4 is excitatory, rather than inhibitory, it is still possible to replicate the object-based effects seen in IT and V4, but this temporary suppression of all features does not occur. Therefore, the most accurate replication of the V4 recording is achieved here with inhibitory feedback from IT to V4. However, the effect seen in the cellular recordings may not necessarily be due to direct inhibitory feedback but could be an after-effect of the initial sensory burst or be due to a local inhibitory effect or to excitatory feedback from IT to inhibitory cells in V4. In principle, the target effect and its time course are unaffected by whether the feedback from prefrontal cortex to IT and from IT to V4 is excitatory to target features/objects or inhibitory to non-target features/objects. However, single cell recordings in V4 (McAdams & Maunsell, 2000) show that responses are suppressed when attending to a feature other than the

one in the receptive field (space and feature unattended condition), relative to when spatial attention is directed elsewhere (space only unattended condition) and this suggests that object-based modulation involves an inhibitory component for unattended features.

Figure 4.16 in chapter 4 showed plots of activity in V4 similar to figure 5.7 but for a simulation in chromatic mode. In this mode, two feature types (colour and orientation) are represented in V4 with two feature settings each. In figure 4.16, unlike figure 5.7, there was a slight increase in activity for non-target features between 150-200ms post-stimulus before this was suppressed again. This temporary effect is due to these features being released from suppression by a distractor object sharing one feature type, i.e. the opposite feature setting, same feature type object in IT. For example, for a RV target, V4 green assemblies become released from suppression by the RH object as the RV object wins the competition in IT. This is an issue of feature binding, which is discussed as a weakness in the current model in chapter 8. It does not cause a problem to the operation of the model during guidance of search but produces a slight deviation from the trends seen in the monkey single cell recordings, which is overcome by the use of monochromatic mode for the simulations in this chapter.

The object-based effects described in this section result in non-target features being suppressed and target features being enhanced in parallel across V4. This produces an effect like that observed by McAdams & Maunsell (2000) and Motter (1994a,b),

where the response of individual cells recorded across a local region of V4 were enhanced in parallel if a target coloured (or target luminance) stimulus was present in their receptive field. The effect across the scene in V4 can be seen in figure 4.16 in chapter 4. In addition to this parallel feature effect, there is a slight spatial enhancement over the initial area of the AW, which is supported by ERP data that suggests there is a process of an early spatial selection within which attention to features develops (Anllo-Vento & Hillyard, 1996).

### 5.3.3 Comparison of Simulations and the Timing of Saccade Onset

In comparison to the simulations of the recordings from IT shown in figures 5.3 to 5.6, the simulation of the V4 recording, shown in figures 5.7 and 5.8, used stronger object-related feedback from IT to V4: Parameter $\eta$ in equations 4.4 & 4.6 of chapter 4 was set to 5 instead of 2.5. Similar patterns of activity are seen with both settings of this parameter (see figure 5.9 in comparison to figure 5.3), but the higher value produces a stronger target object effect. It was necessary to tune the parameter to different settings to produce the optimal fit with the monkey data. The animals used by Chelazzi et al. (2001) for the V4 recordings were the same as those used by Chelazzi et al. (1993) for the IT recordings. However, the animals had received significantly more training during the intervening years. Thus, this suggests that the strength of object-related feedback from IT to V4 may be tuned by learning or affected by age.

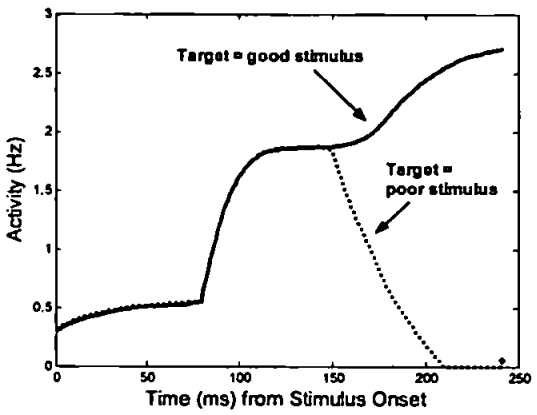



**Figure 5.9 Object-based Attention in IT – Strong IT Feedback**

Activity in IT for the V4 simulation given in figure 5.7. This plot shows the same situations as figure 5.3 and can be compared to that plot. However, here, the strength of IT feedback is set to the higher value used for the replication of V4 recordings shown in figures 5.7 and 5.8 (i.e. parameter $\eta$ in equations 4.4 & 4.6 is set to 5). The effect of stronger IT to V4 feedback is to suppress early responses slightly and, later, suppress the distractor object more significantly in IT, as well as V4. The saccade took place after 241ms, as indicated by the asterisk, which is similar to the times observed by Chelazzi et al. (2001).

The V4 recordings from the more highly trained monkeys (Chelazzi et al., 2001) also showed a slightly earlier object-based effect (from 150-160ms after array onset) compared to that in the IT recordings (Chelazzi et al., 1993), which began 170-180ms after array onset. The effects in both cases began ~70-80ms before the saccade onset. The slightly later onset of object-based effects in IT can be modelled by shifting the prefrontal sigmoid function such that effective feedback is delayed. This is achieved by altering the parameter $\tau_{sig}$ in equation 4.9 of chapter 4,

and results in the onset of object-based attention in IT and V4 being delayed until ~200ms, as simulated in figure 5.10. This leads to saccade onset occurring ~300ms (302ms in the simulation shown in figure 5.10). This is similar to the average onset times of 297ms found by Chelazzi et al. (1993; see figure 5.3a above); and 307ms found by Chelazzi et al. (1998; see figure 5.4a above) in this condition.

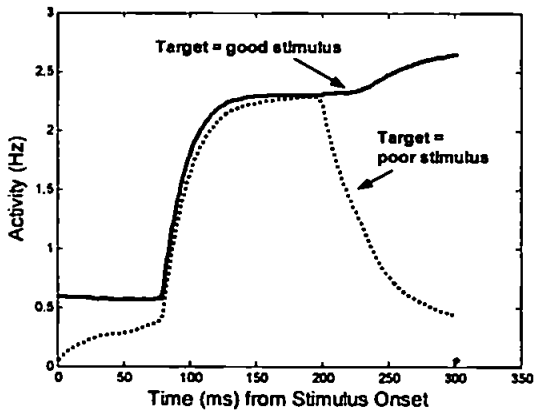



**Figure 5.10 Object-based Attention in IT – Later Effect**

Activity in IT during a simulation to replicate the time of onset of object-based attention and saccades found by Chelazzi et al. (1993, *figure 3a*; 1998, *figure7a*). This is similar to figure 5.3 but with the onset of object-based attention not occurring until ~200ms post-stimulus, which matches Chelazzi et al. (1993) most accurately (see figure 5.3a above) and is later than that found by Chelazzi et al. (2001) when recording in V4. Saccade onset is indicated by an asterisk at 302ms. Here, the parameter $\tau_{sig}$ in equation 4.9, relating to the timing of significant prefrontal feedback to IT, is set to 200. The weight of IT feedback to V4 (parameter $\eta$ in equations 4.4 & 4.6) is set to 2.5, the same value as in figure 5.3.

Therefore, the model's latency to saccade, being linked to the development of a significant target object effect in IT, produces saccade onset times that mirror those from the IT and V4 recordings by Chelazzi et al. (1993, 1998, 2001). The timing of prefrontal feedback to IT (i.e. the latency of the prefrontal sensory response) had a significant effect on the timecourse of development of object-based attention and, hence, the saccade onset time. Compared to those in less trained monkeys (Chelazzi et al., 1998), significant object-based effects and saccade onset occurred earlier in the more highly trained monkeys (Chelazzi et al., 2001). This was optimally replicated when prefrontal feedback was available to the system earlier (parameter $\tau_{sig}$ in equation 4.9 of chapter 4 set to 150). When object-based effects began in IT and V4 in the model simulations at ~150ms, saccades took place at ~240ms. This mirrors the times reported by Chelazzi et al. (2001). However, when prefrontal feedback was delayed such that object-based effects began in IT and V4 ~200ms, saccade onset occurred at ~300ms, which mirrored the timing of these events reported by Chelazzi et al. (1993) using the less highly trained animals. This suggests that the timing of prefrontal feedback to IT is also tuned by learning. As the sigmoidal bias represents the prefrontal sensory response, a plausible explanation is that the delay is due to bottom-up sensory information taking longer to arrive in prefrontal cortex for new objects or in an unfamiliar task.

The latency to saccade was also slightly longer when IT to V4 feedback was weaker, (parameter $\eta$ of equations 4.4 and 4.6 in chapter 4 is set to 2.5 instead of 5) as can be seen by comparison of figures 5.7 and 5.9 (strong feedback and saccade

onset at 241ms) with figure 5.3 (weak feedback and saccade onset at 252ms). However, this effect is less significant than the timing of prefrontal feedback to IT. In addition to the difference in training, the stimuli configurations used by Chelazzi et al. (1993, 2001) were slightly different in the two experiments (being closer together in the V4 recordings) and so these predictions are tentative.

Overall, the timing of object-based effects in IT and V4 and the associated saccade onset times were accurately replicated. Producing such accurate replication of cellular recordings whilst linking saccade onset time with the development of a significant object-based effect in the model's IT, suggests that such a link may exist in vivo. Therefore, the development of object-based effects in the ventral stream may have influence on the cortical and sub-cortical structures responsible for decisions about saccades (e.g. parietal cortex, FEF, and superior colliculus).

## 5.4 Effect of Increasing Distractor Numbers

The above simulations had one target and one distractor object in the receptive field. The image shown in figure 5.11 has four different objects present in the same receptive field. Compared to the earlier simulations there are more distractors present in this display (the target is still the red vertical) and four objects now compete in IT instead of the two used previously in the monochromatic simulations above. Therefore, there is increased competition in IT due to there being more

active object representations. The effect of this increased competition is to drive down the initial sensory response, as shown in figure 5.12, which can be compared to figure 5.3 above. However, the later response, which is subject to the effects of object-based attention, is similar in magnitude to the two-stimulus case.

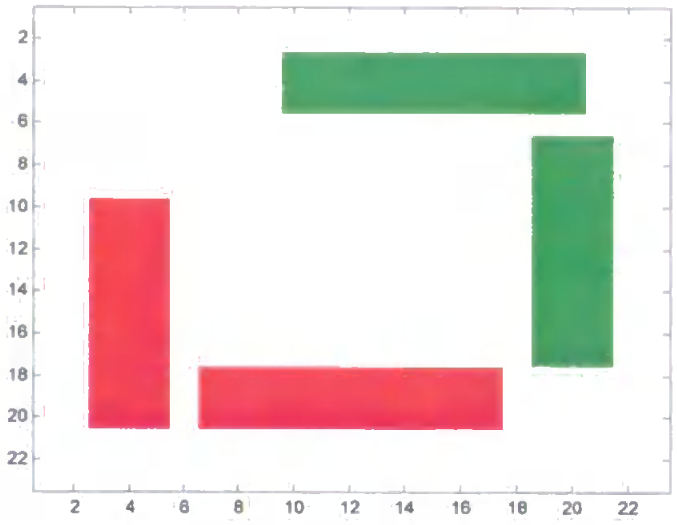



**Figure 5.11 Scene with Four Objects**

Original image with one of each of four object types in the receptive field. The IT receptive field spans the entire image.

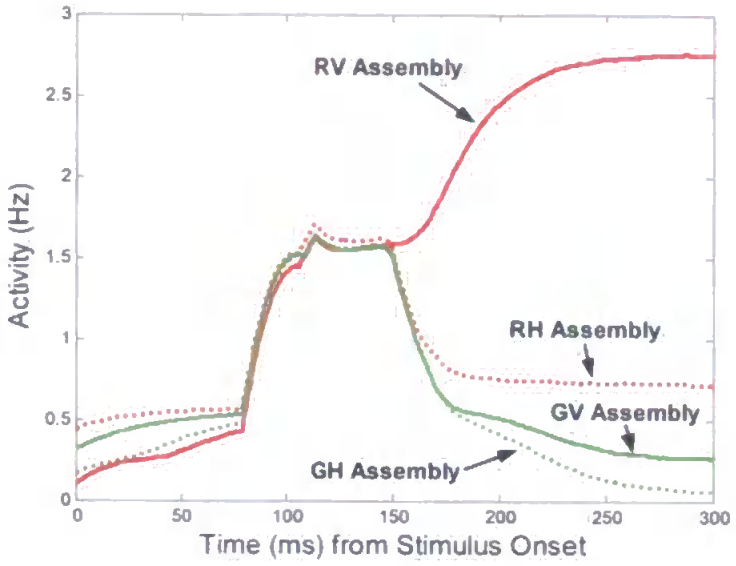



**Figure 5.12 Activity in IT for the Four Object Scene**

Plot of activity in IT cell assemblies when the objects shown in figure 5.11 are placed within their receptive field.

Activity in IT during the course of a scan path simulation appears to be slightly lower compared to the single cell simulations presented in this chapter. During a fixation on the image shown in figure 5.13, with the retina as shown, an example of activity in IT is given in figure 5.14. For this simulation, IT contains all four objects and V4 has 20x20 retinotopic positions in each feature map feeding forward to IT, with this input activity being normalised as normal. There are many more distractors in the scene feeding forward information to IT. This marginally reduces the initial sensory response, compared to figures 5.3 and 5.12. However, the most significant effect, compared to figures 5.3 and 5.12, is the reduction in magnitude of the response for the target object assembly in the later portion of the response.

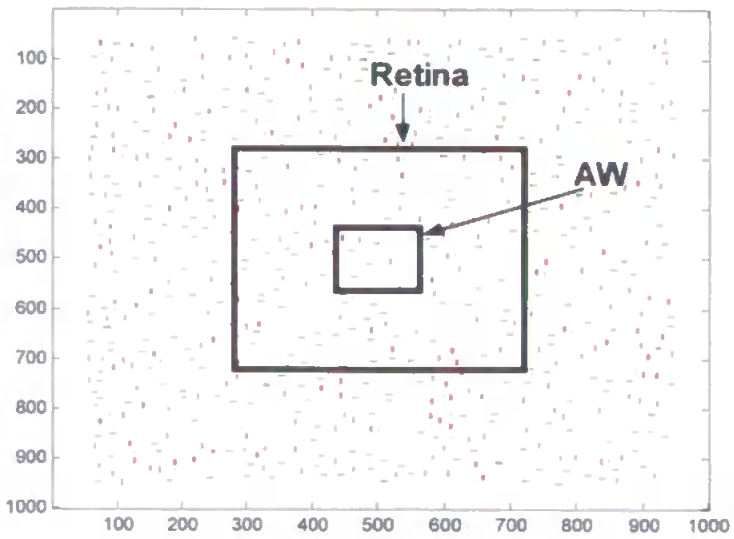



**Figure 5.13 Dense Scene with Numerous Distractors**

A dense scene (1000rh300, with 300 of each distractor type) with many distractors. The extent of the retina and AW are shown. The IT receptive field spans the entire retina. The target is a red horizontal bar.

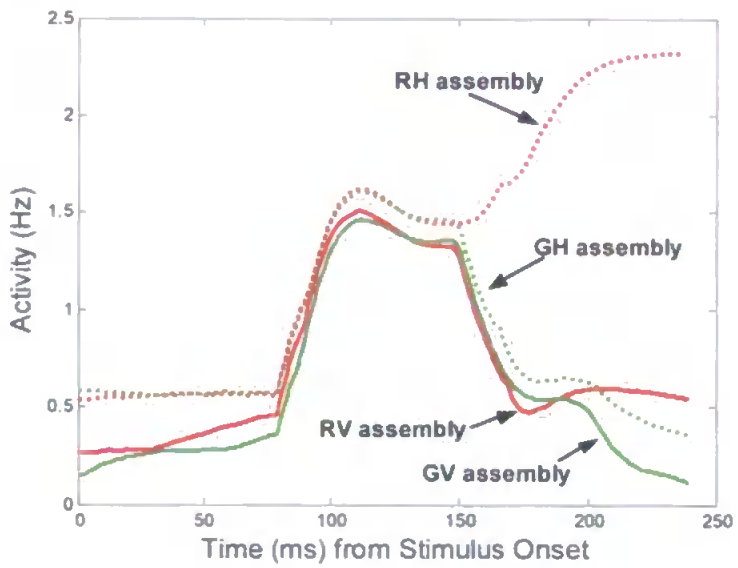



**Figure 5.14 Activity in IT for a Dense Scene with Numerous Distractors**

Plot of activity in IT cell assemblies during the fixation in a scene containing many distractors, shown in figure 5.13.

This effect of distractor density is demonstrated further by comparing the activity within IT during the fixation in the dense scene above with that during fixation in the sparser scene, as shown in figures 5.15 and 5.16. Activity is slightly lower in the more crowded scene.

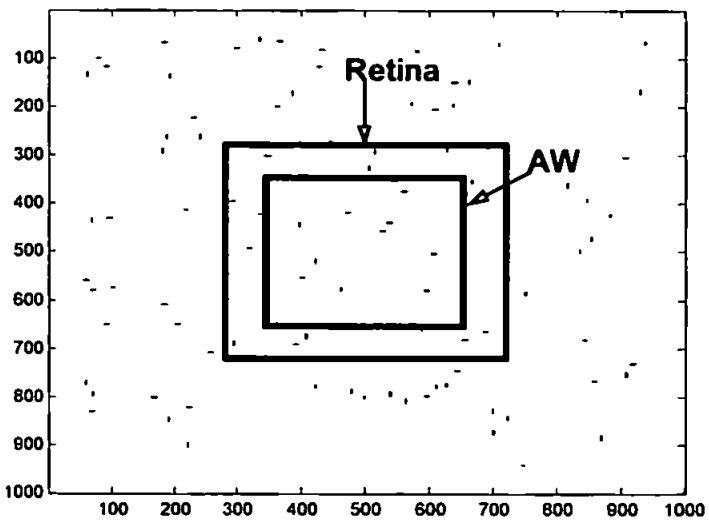



**Figure 5.15 Sparse Scene with Fewer Distractors**

A sparse scene (1000gv48, with 48 of each distractor type) with the retina and AW plotted. The retina is the same size as that used for figures 5.13 and 5.14 and, again, the IT receptive field covers the retina. The target is a green vertical bar.

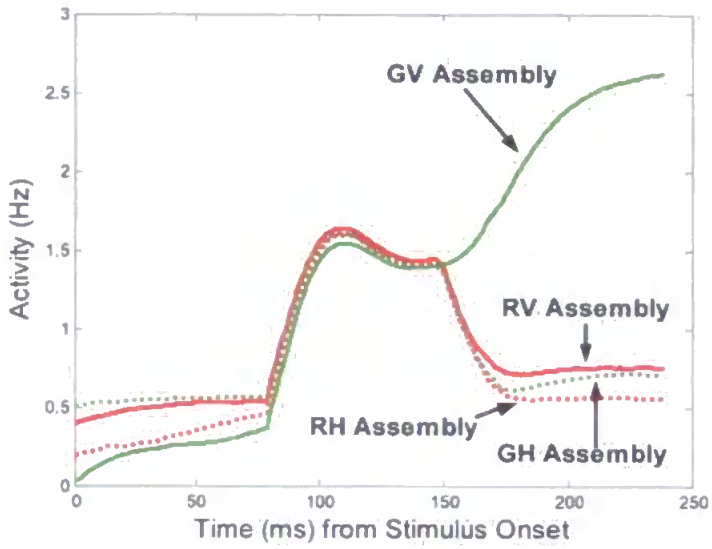



**Figure 5.16 Activity in IT for a Sparse Scene with Fewer Distractors**

Activity in the IT cell assemblies during the fixation shown in figure 5.15 in the sparser scene. Activity for the target object reaches a higher value than that shown in figure 5.14 for the denser scene.

If, however, the radius of the retina is extended in the sparse scene (from the normal radius of 220 pixels above, to 440 pixels), as shown in figures 5.17 and 5.18, more stimuli subtend the retina, and are processed by the dynamic cortical modules. This tests the theory that increasing the number of stimuli processed within the ventral stream affects the level of activity in IT. Figure 5.18 shows that this increased competition within the ventral stream does result in activity in IT being reduced. This effect is most prominent in the later part of the response, after the onset of object-based attention. Notice also the fact that the AW is scaled according to stimulus density and that this is consistent across figures 5.15 and 5.17

regardless of retina size. Therefore, any effect is not due to a variation in the number of stimuli within the initial spatial attentional "spotlight".

Thus, the effect of more distractors being viewed within the retina appears to be a marginal decrease in responses after the onset of object-based attention. In particular, the level of activity for the target object is lower. This could perhaps make the target object more difficult to identify and, therefore, increase reaction times. There is some evidence that a greater number of distractors in a display increases the time to saccade (McSorley & Findlay, 2003).

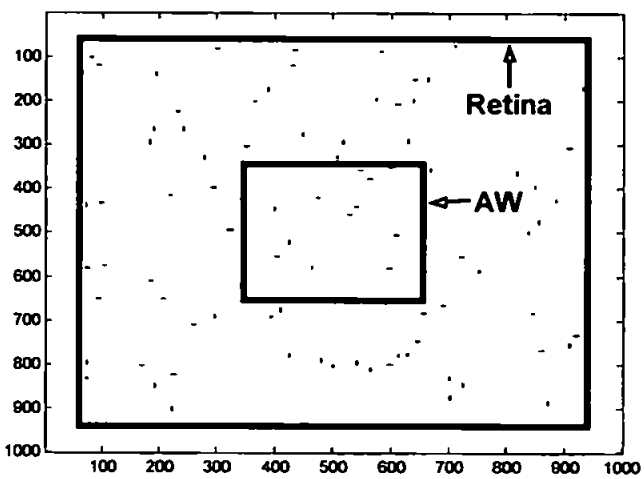



**Figure 5.17 Sparse Scene with Larger Retina**

The same sparse scene as that used in figure 5.15, but here the retina (and, hence, IT receptive field) is larger so that it encompasses more stimuli.

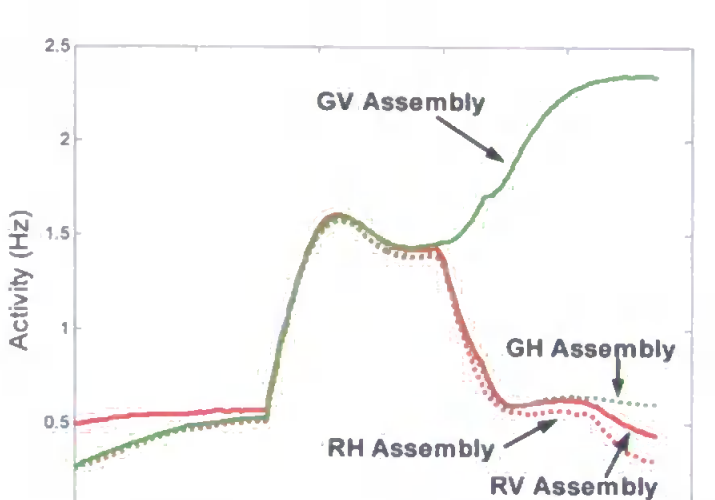



**Figure 5.18 Activity in IT for a Sparse Scene with Larger Retina**

Activity in the IT cell assemblies during the fixation shown in figure 5.17 in the sparser scene with the large retina. Activity for the target object is not able to reach such a high as that shown in figure 5.16 when the retina was smaller and competition was, therefore, weaker.

## 5.5 Chapter Summary

The cellular level behaviour of the model was described in this chapter. At this level, the times of onset of spatial and object-based attention were examined in detail so that these effects were replicated over time courses that reflect the monkey neurophysiological data discussed above (Chelazzi et al., 1993, 1998, 2001; Colby et al., 1996; Luck et al., 1997; Motter, 1994 a, b) as well human ERP data (Anllo-

Vento & Hillyard, 1996; Anllo-Vento et al., 1998; Hillyard & Anllo-Vento, 1998) reviewed in chapter 2. This is the first time that the time courses of these effects have been replicated.

Following fixation, the spatial Attention Window (AW) provided an early spatial effect in V4 from the start of the sensory response, as found by Luck et al. (1997). This was due to a spatial bias to V4 from LIP, which was subject to a bias related to the eye movement and new fixation position, possibly from FEF (Moore & Armstrong, 2003). Although the spatial bias was able to cause an early attentional effect (Anllo-Vento & Hillyard, 1996; Anllo-Vento et al., 1998; Hillyard & Anllo-Vento, 1998; Luck et al., 1997; Moore & Armstrong, 2003), object-based attention took time to develop because it was dependent on the resolution of object-related competition within higher ventral stream areas.

### 5.5.1 Object-based Attentional Effects at the Cellular Level

Object-based attention develops in the model's ventral stream due to a target object-related bias that is assumed to be derived from the sensory response in prefrontal cortex, which builds over time according to a sigmoid function. This allows the timing of the onset of significant object-based effects in V4 and IT to be replicated, beginning from at least 150ms after the onset of the search array (Chelazzi et al., 1993, 1998, 2001). Although no target effect is present in the earlier part of the sensory response in V4 (Chelazzi et al., 2001), average early

responses in IT show a slight target effect that is much less significant than that in the late response (Chelazzi et al., 1998). This effect has been simulated here by the addition of a very small, but sustained, prefrontal bias, which modulates responses in IT from its earliest sensory response. In this simulation, the strongest target effect is still that occurring later in the response (after ~150ms post-stimulus), due to the development of the sigmoidal prefrontal bias. It is suggested that the early effect and more significant late effect observed in the monkeys (Chelazzi et al., 1993, 1998) may be due to more than one type of bias signal, or a bias signal with more than one component, for example prefrontal activity that is not constant over time but has at least two components:

- A sustained mnemonic component, resulting in target-related delay period activity and the small target effect seen in early sensory responses in IT (Chelazzi et al., 1993, 1998)
- A sensory response that differentiates targets from non-targets later in the response (Everling et al., 2002; Rainer et al., 1998) and leads to the significant target effects later in IT and V4 (Chelazzi et al., 1993, 1998, 2001).

Prefrontal cells with shorter response latencies may be able to produce the significant early target effects found in a few cells in IT (see figure 5.5). The prefrontal bias and alternative suggestions for the source of the late bias signal will be discussed further in chapter 8.

In order to simulate the onset of object-based attention in more highly trained monkeys (Chelazzi et al., 2001 compared to 1993), it was necessary for the prefrontal feedback to IT to be effective slightly earlier. This suggests that the timing of the prefrontal response and, hence, feedback to IT, is adapted by learning and familiarity with the searched objects and the task. This could be due to the latency of IT inputs to prefrontal cells being longer in unfamiliar situations. The prediction of a link between prefrontal feedback to IT and learning is tentative in view of the stimulus configuration differences between the two experiments (Chelazzi et al., 1993, 2001). However, in FEF, stronger responses to target as opposed to non-target stimuli occur earlier if the monkey has had long experience of searching for the same target (Bichot, Schall & Thompson, 1996) and this suggests that experience-dependant plasticity of object-related responses is possible in higher stages of cortical processing. Overall, the current work highlights the need for more recordings and modelling of the dynamics in prefrontal cortex. Future modelling in this area should seek to gain a better understanding of the dynamics within the sensory response in prefrontal cortex and how it interacts with IT, and other ventral stream areas, to influence the development of object-based attention.

In addition to the effect of learning on prefrontal feedback to IT, there is a possible effect on the tuning of IT feedback to V4. The most accurate replication of the target object effect in more highly trained monkeys (Chelazzi et al., 2001 compared to 1993) required stronger feedback from IT to V4. Again, this prediction is

tentative due to the differences in the two experiments. However, it is plausible that experience and knowledge of searched objects would lead to more highly tuned feedback within the ventral stream and, hence, stronger target object modulation of responses. Familiarity, in particular with distractors but also the target, does appear to make search more efficient (Greene & Rayner, 2001; Lee & Quessy, 2003; Lubow & Kaplan, 1997; Wang, Cavanagh & Green, 1994); as does familiarity with the search task (Sireteanu & Rettenbach, 2000).

The number of stimuli falling on the retina has a slight effect on overall levels of activity in IT, particularly noticeable in the level of activity of the target object assembly when it wins the competition. This is attributed to increased competition due to the large number of stimuli. This suggests that targets may be less easily identified in very cluttered scenes or that it may take longer for these representations to reach supra-threshold levels in IT, i.e. that reaction times may be longer. Latency to first saccade has been found to increase with greater number of distractors (McSorley & Findlay, 2003).

### 5.5.2 Saccades

The automatic initiation of a saccade in the model is temporally linked to the development of a significant object-based effect in IT, as described in chapter 4. This enables not only the timescale of development of object-based attention to be replicated, but also saccade onset times to mirror those found in the cellular

recordings studied here (Chelazzi et al., 1993, 1998, 2001). This suggests that saccade onset may be directly linked to the development of a significant object-based effect in the ventral stream. Due to its dependence on the development of object-based attention, saccade onset here is also subject to the timing of prefrontal feedback and, possibly, familiarity with searched objects and the task. Therefore, it is predicted that saccade latency may be shorter when prefrontal response latency is short, for example amongst very familiar objects or in a familiar task. Short-term memory of target features, such as colour, does appear to reduce saccade latency (McPeek, Maljkovic & Nakayama, 1999). The effect of varying the strength of prefrontal feedback on scan path behaviour will be described in the next chapter.

### 5.5.3 Conclusion

The content of this chapter relates to the first research goal, presented in chapter 2. The aim was to investigate the concurrent modulation of V4 responses by spatial and object-based attention such that object-based attention, which depends on the resolution of competition between objects in IT, appears later in the V4 response. The simulations described in this chapter have been able to replicate the timescale of these effects seen in single cell recordings. The following summarises the results and contribution of the model at the cellular level, when replicating neurophysiological data.

| Neurophysiological Data Replicated | Results from the Model |
|---|---|
| **Colby et al. (1996) Monkey single cell recording in LIP** | Activity of cells assemblies in LIP whose receptive fields overlap the initial spatial focus of attention (AW) were enhanced due to the requirement to attend at fixation. See figures 4.10 to 4.12 in chapter 4. |
| **Luck et al. (1997) Monkey single cell recording showing spatial modulation of responses in V4, including a baseline modulation** | The activity of V4 cell assemblies whose receptive fields overlap the initial spatial focus of attention (AW) were enhanced from the earliest sensory response at 60ms post-stimulus, and baseline activity before this was also modulated. See figure 5.1 above, and figures 4.11 and 4.12 in chapter 4. |
| **Hopfinger et al. (2000) Human fMRI recording showing increased extrastriate activity due to spatial attention** | |

| | |
|---|---|
| **Motter (1994a,b) Monkey single cell recording of object(feature)-based modulation of responses across V4 from 150-200ms post-stimulus** | Cell assemblies with selectivity for target features (in particular, colour) were enhanced in parallel across V4 from ~150-200ms post-stimulus. See figures 5.7 and 5.8 above, and figures 4.12 and 4.15 in chapter 4. However, the spatial AW facilitated the development of object-based attention such that object-based responses within the AW were enhanced. This links with human ERP data (Anllo-Vento & Hillyard, 1998), which suggests that feature selection is contingent upon prior spatial selection. |
| **Chelazzi et al. (1993, 1998, 2001) Monkey single cell recording of object-based modulation of responses in IT &V4 from 150-200ms post-stimulus** | Object-based effects were produced from ~150ms-200ms in IT and V4 cell assemblies such that, when two stimuli were placed in the same receptive field, the response was determined by which stimulus was attended. See figures 5.3 to 5.10. The time course of this modulation observed in these seminal monkey recordings was reproduced more accurately than by previous models in this area. Also, the time of onset of a saccade was linked to the development of significant object-based effects. This enabled saccade onset times observed during these cellular recordings to be also simulated. A greater number of stimuli within the IT receptive field slightly reduced the response in IT to the target stimulus and this could be responsible for increasing reaction times. |

## Predictions

- Active search involves an initial spatial focus of attention at fixation, which, over time, becomes object-based.
- A prefrontal bias to IT is suggested by biased competition theories (Desimone, 1998) to be the top-down source of target object effects in IT. This has been previously modelled as a steady current injection to the dynamics in IT (Deco, 2001; Deco & Lee, 2002; Rolls & Deco, 2002; Renart et al., 2001; Usher & Niebur, 1996). The results here suggest that this bias needs to consist of more than one component so that its magnitude is not constant over time:
    - A sustained mnemonic component
    - A sensory response that builds over time, in line with normal prefrontal response latencies
- Object-related feedback in the ventral stream (prefrontal to IT; IT to V4) may be tuned by learning, i.e. familiarity with objects and the task.
- The onset of a saccade is dependant upon the development of a significant object-based effect in the ventral stream. Dynamics within the ventral stream may influence saccade-related areas such as parietal cortex, FEF and superior colliculus.
- Due to its proposed dependence on the development of object-based attention, saccade latency may be affected by familiarity with objects and the task.

## Contribution of the Work

With the exception of Gustavo Deco's model (2001; Deco & Lee, 2002), few computational models of both spatial and object-based attention currently exist. Unlike Deco's model, the model presented here has spatial and object-based biases operating concurrently. None of the existing biased competition models specifically address the different time course of spatial and object-based attentional effects and, therefore, the replication of the time course of these effects at the cellular level is novel. It is suggested that the simple static value prefrontal feedback to IT commonly used in biased competition models (e.g. Deco, 2001; Deco & Lee, 2002; Renart et al., 2001; Usher & Niebur, 1996) is not sufficient to accurately model object-based responses in the ventral stream. Also, possible links between object-related feedback and learning have been highlighted.

The cellular level object-based effects presented here are important collectively to producing systems level behaviour that models human and monkey visual search. This systems level behaviour is the subject of the next chapter.

# Chapter 6

# Systems Level Behaviour – Scan Path Simulation Results

The systems level output from the system, i.e. the visual search scan path behaviour, is examined in this chapter. Attentional effects at the cellular level of the model were presented in the previous chapter. Results here show how these cellular level effects, which simulated monkey single cell data, lead to systems level behaviour that qualitatively replicates visual search behaviours in humans and monkeys. The systems level output was introduced in chapter 4, where activity in each of the dynamic modules over time during a fixation was described. Here, the focus is on the search scan paths resulting from this activity and examining the effects of changes of some of the model's parameters on this behaviour. In order to examine these effects, multiple scan path simulations were run over the same image for each particular setting of the parameter. Appendix A3 gives the results of these simulations and the individual subsections of appendix A3 will be referenced here as appropriate.

The object-based effects at the cellular level of the model produce systems level behaviour where the search is guided to locations containing behaviourally relevant features. Due to the relative strength of feature connections from V4 to LIP, the search scan paths produced replicate those found by Motter and Belky (1998b; also Lauria & Strauss, 1975; Scialfa & Joffe, 1998; Williams, 1967; Williams & Reingold, 2001) where colour was most influential in guiding search for a colour-orientation conjunction target. However, in

principle, the outcome of object-based attention at the cellular level allows any behaviourally relevant feature, that is preferentially enhanced due to object-based processes in the ventral stream, to attract attention and guide the search.

## 6.1 Analysing Fixation Position With Respect to Stimuli

Following scan path simulations, the positions of fixations were analysed to determine whether they fell within 1° of the centre point of a stimulus. Specifically, an area of radius 1° around the fixation point was tested for the presence of stimuli in the two maps containing the centre points of the red and green stimuli respectively. If a stimulus was found in this area, the scan path was considered to have targeted this stimulus. If more than one stimulus fell within 1° of the fixation position, the closest stimulus (using a Euclidean distance measure) was considered to be the one fixated. The radius of 1° was chosen to match that used by Motter and Belky (1998a,b), whose scan path results are simulated here. If an area larger than 1° is tested for the presence of stimuli, the scan path appears to fixate less in blank areas of the display and targets more stimuli.

## 6.2 Guidance of Scan Path by Colour

When target object effects develop in the V4 module, representations of target features are enhanced in parallel across the visual field, as found by Motter (1994a,b) and McAdams & Maunsell (2000), although there is a slight spatial enhancement over the initial area of the AW (ERP data suggests an early spatial selection within which attention to features develops: Annlo-Vento & Hillyard, 1996). Figures 4.12 and 4.15 in chapter 4 showed this effect. This target feature information is conveyed to LIP such that colour has a stronger effect than orientation. This results in LIP encoding locations containing the target colour and causing saccades to these stimuli rather than those of the target orientation but non-target colour. The system is able to qualitatively simulate scan path behaviour observed by Motter and Belky (1998b; see also Lauria & Strauss, 1975; Scialfa & Joffe, 1998; Williams, 1967; Williams & Reingold, 2001), where fixations normally landed within 1° of stimuli, rather than in blank areas, and these stimuli tended to be target coloured. Figures 6.1, 6.2 and 6.3 show typical scan paths through relatively dense, sparse and very sparse scenes respectively.

Results in this chapter will show how factors such as the density of stimuli in the retinal image, the strength of object-related feedback in the ventral stream and the relative weight of connection of the V4 features to LIP affect fixation position, i.e. the relative number of fixations on target coloured stimuli compared to those on non-target coloured stimuli or in blank locations. With these parameters set to their

normal values, as shown in figure 6.1, 6.2 and 6.3, a high proportion of fixations land near target coloured stimuli and, in reasonably dense scenes, very few fixations land in blank areas.

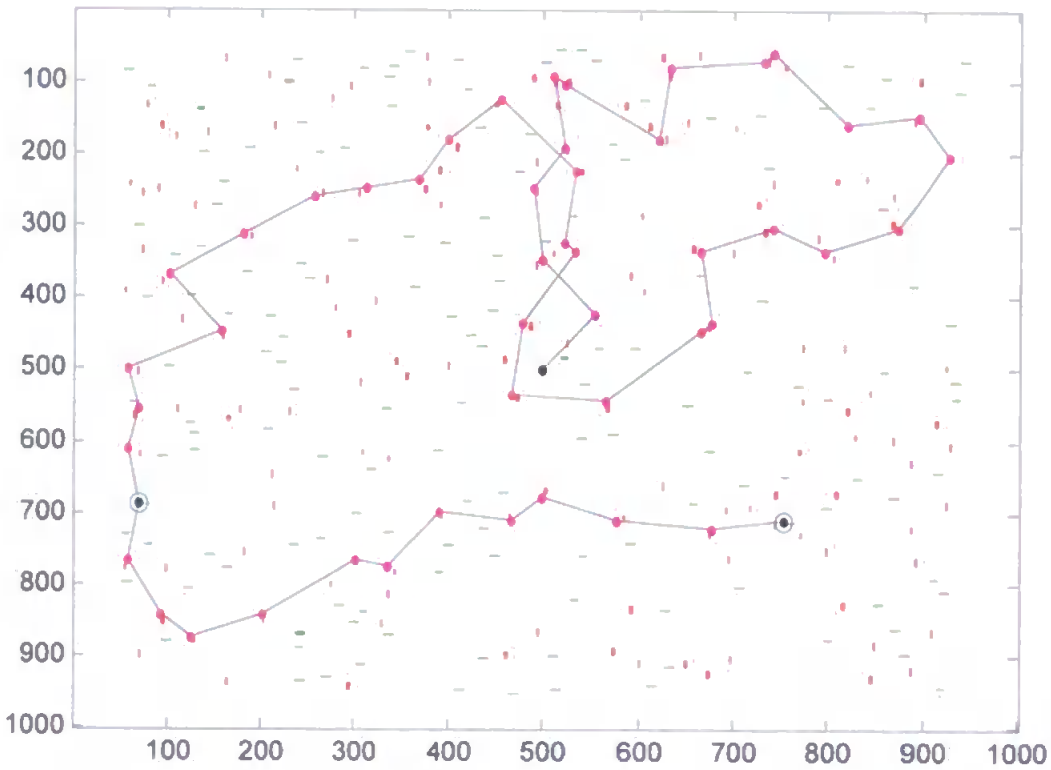



**Figure 6.1 A scan path through a dense scene**

The target is a red stimulus. The scan path tends to select locations near target coloured stimuli. Here, 96% of fixations are within 1° of a target coloured stimulus and are shown as magenta dots. 4% of fixations here are within 1° of a non-target colour (but target orientation) stimulus and are shown with a blue circle around the fixation point. Average saccade amplitude is 7.43°. IT feedback to V4 was set to its stronger value, used for V4 single cell simulations in chapter 5, i.e. parameter $\eta$ in equations 4.4 and 4.6 of chapter 4 is set to 5. For reference throughout, the image name is 1000rh150.

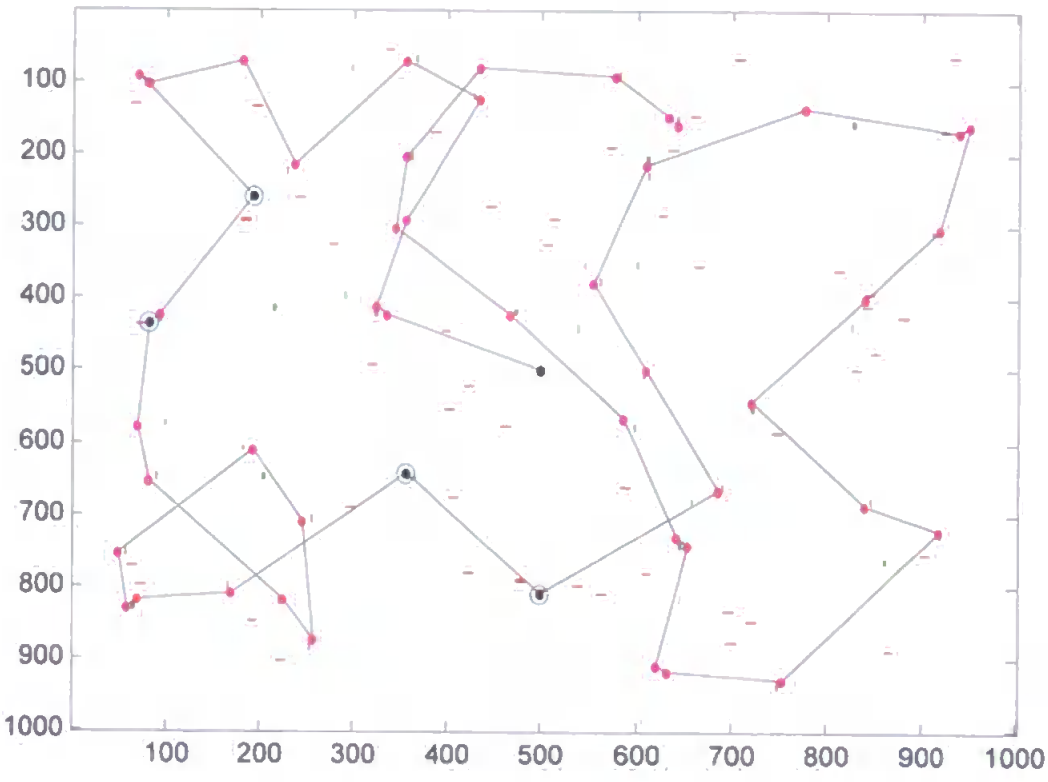



**Figure 6.2 A scan path through a sparse scene**

The target is a green stimulus. 92% of fixations here are within 1° of a target coloured stimulus and are shown as magenta dots. 8% of fixations here are within 1° of a non-target colour stimulus and are shown with a blue circle around the fixation point. Average saccade amplitude here is 12.1°. IT feedback to V4 was again set to its stronger value, used for V4 single cell simulations in chapter 5, i.e. parameter $\eta$ in equations 4.4 and 4.6 of chapter 4 is set to 5. For reference throughout, the image name is 1000gh50.

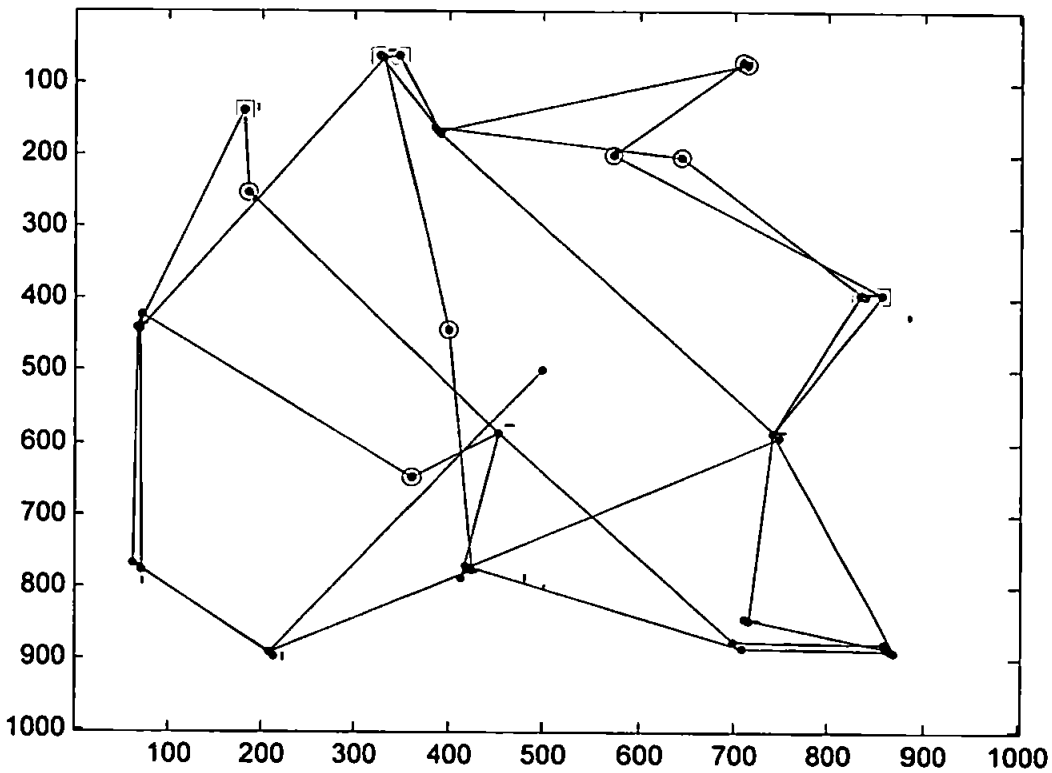



**Figure 6.3 A scan path through a very sparse scene**

The target is a red stimulus. 76% of fixations here are within 1° of a target coloured stimulus and are shown as magenta dots. 14% of fixations here are within 1° of a non-target colour stimulus and are shown with a blue circle around the fixation point. 10% of fixations here are in blank areas of the image and are shown with a yellow square around the fixation point. Average saccade amplitude here is 18.8°. IT feedback to V4 was again set to its stronger value, used for V4 single cell simulations in chapter 5, i.e. parameter $\eta$ in equations 4.4 and 4.6 of chapter 4 is set to 5. For reference throughout, the image name is 1000rv11.

### 6.2.1 Effect on Scan Paths of Varying the Relative Weight of V4 Featural Inputs to LIP

The difference in strength of connection of the V4 features to LIP need only be marginal in order to achieve the priority of colour attracting attention over orientation. Figure 6.4 shows the effect of adjusting the relative connection weights. To produce this analysis multiple scan paths were run over the image 100rh150 shown in figure 6.1 above. Unless otherwise stated, this image was used for all analysis in this chapter. Figure 6.4 shows that when the strength of connection of the colour features is 103% that of the orientation features, there is a priority for colour in guiding the scan path (Lanyon & Denham, 2004b, 2005a).

Figure 6.4 also indicates a priority for orientation guiding the scan path when the connections are equally weighted. This is a computational artefact of the system and is due to slightly unequal outputs of the different filters used in V1 combined with the effect of passing these outputs through a normalising transfer function. This is described in detail in appendix A3.2 but does not affect the normal performance of the system since colour is normally more highly weighted.

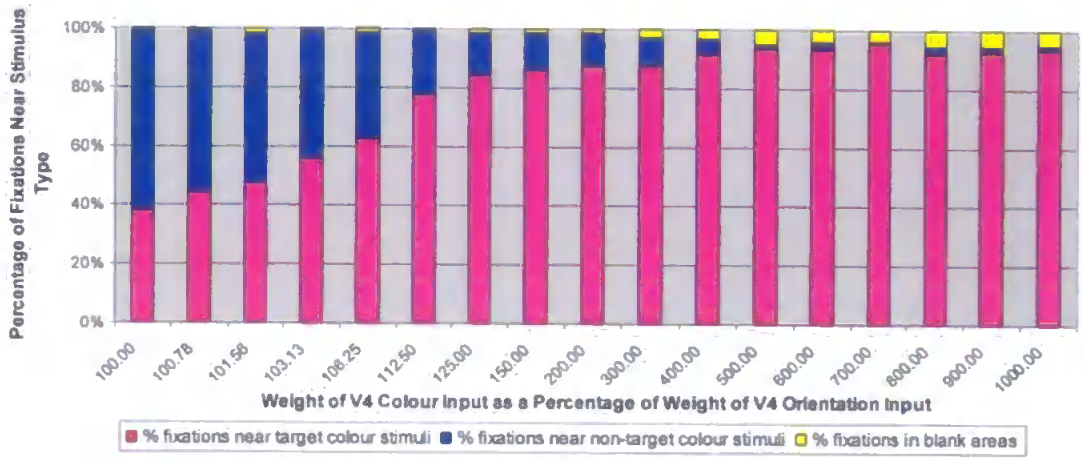



**Figure 6.4 The effect on fixation position of increasing the relative weight of V4 colour feature input to LIP**

When V4 colour features are marginally more strongly connected to LIP than V4 orientation features, the scan path is attracted to target coloured stimuli in preference to stimuli of the target orientation. Fixation positions were averaged over 10 scan paths, each consisting of 50 fixations, over the image shown in Figure 6.1. The original data are given in appendix A3.1.

The dominance for colour guiding the scan path occurs from a relatively small difference in connection weights and increases quickly as this connection is made stronger. Once colour is significantly more strongly connected than orientation (for example, over 150% of the orientation connection strength) performance is fairly stable so that further increases in colour connection produce only slight increases in the dominance of colour in the scan path. Even if colour is connected 10 times more strongly than orientation (1000% bar in figure 6.4), some non-target coloured

locations and blank areas attract the scan path. The normal connection weights used in simulations are set such that colour is 5 times more strongly connected than orientation. This is shown by the 500% bar in figure 6.4, which is in the part of the chart that demonstrates relatively stable performance (i.e. there is little change in fixation statistics with variation in relative connection weight).

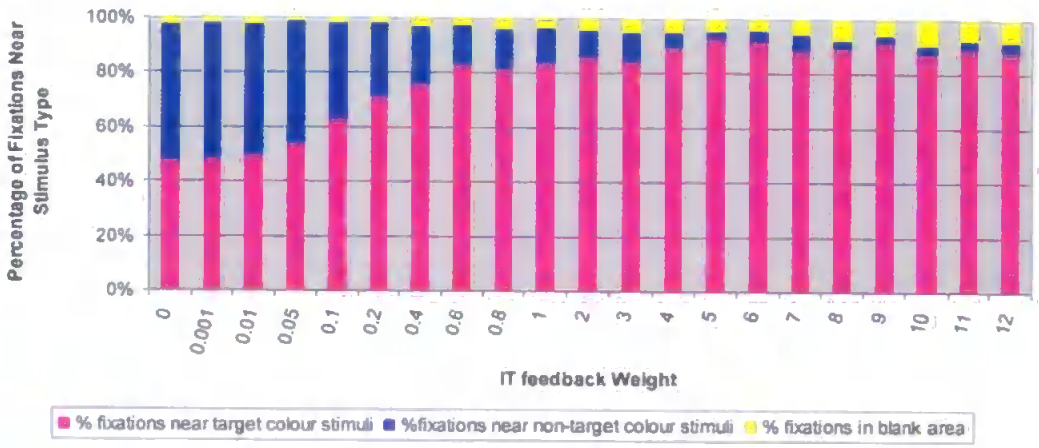
### 6.2.2 Effect on Scan Paths of Strength of Object-Related Feedback in the Ventral Stream

#### 6.2.2.1 Object-Related Feedback and Scan Path Guidance

Object-based attention in the model's ventral stream is crucial to the guidance of the scan path. The object-based effects enhance behaviourally relevant target features in V4 and this information biases competition in LIP so that the scan path is attracted to these locations. These object-based effects are reliant on feedback within the ventral stream, specifically prefrontal feedback to IT and IT feedback to V4. The effect of varying the weight of this feedback on scan path behaviour was examined. Figures 6.5 and 6.6 show that object-based effects are fairly stable over a range feedback weights. However, if the feedback is very weak, object-based attention and the guidance of the scan path are affected (Lanyon & Denham, 2004b, 2005a). This is because weak object-related feedback within the ventral pathway reduces the object-based effect within V4. Object-based attention in V4 enhances target features and suppresses non-target features. Due to the weight of V4 feature

connections to LIP, the effect of object-based attention in the colour feature layer in V4 is most significant to the guidance of the scan path. If ventral stream object-related feedback is weak, non-target colours are almost as active as target colours. This results in target and non-target coloured locations in LIP receiving a similar bias from V4. Thus, more non-target coloured stimuli are able to capture attention. Figures 6.5 and 6.6 show that, when object-based effects are weak, there is an increase in the number of non-target coloured locations capturing attention but not an increase in the number of fixations in blank areas of the display. Although the model is tuned (via the V4 to LIP connection) to be guided by target colour, the principle of these results applies to the capture of attention by any output of object-based processing in the ventral stream, i.e. any behaviourally relevant stimuli.

a.




b.

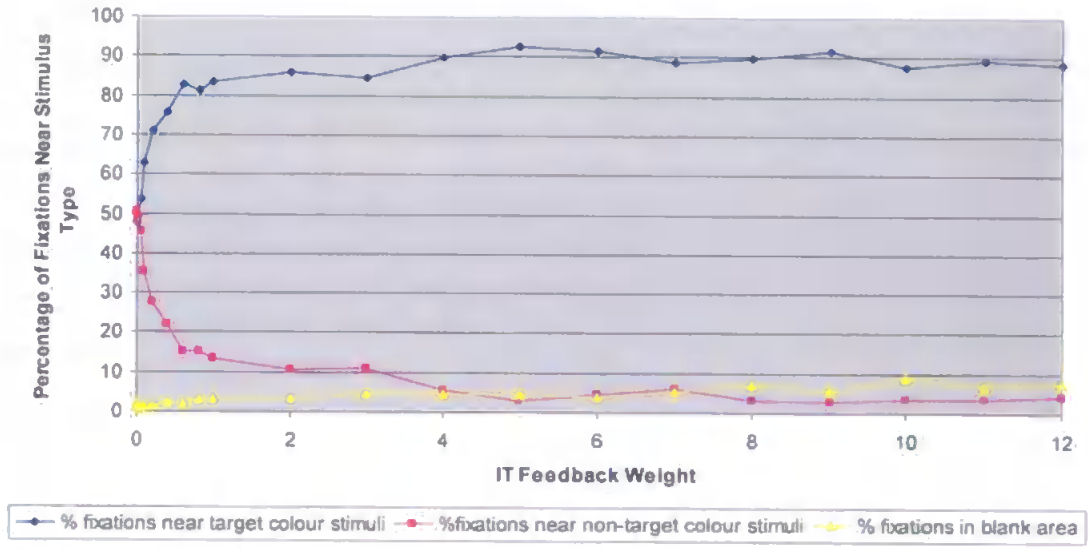



**Figure 6.5 The effect on fixation position of varying the weight of IT feedback to V4**

The weight of IT feedback is given by parameter η in equations 4.4 and 4.6 of chapter 4. The same data is presented in (**a**.) a column chart format, similar to other charts in this chapter, and (**b**.) as a scatter plot. More feedback weights were tested at the lower range compared to the higher range and, therefore, the scatter plot better illustrates that performance is stable over a range of feedback weights. The normal weight of feedback is 5 for scan path simulations. The original data are presented in appendix A3.3. For each parameter setting, average fixation positions were calculated over 10 scan paths, each consisting of 50 fixations, over the image shown in figure 6.1. As the weight of feedback is increased there is a tendency for more target coloured stimuli and less non-target coloured stimuli to be fixated because object-based effects in the ventral stream are stronger.

Replication of single cell data from older and more highly trained monkeys (Chelazzi et al., 2001, compared to Chelazzi et al., 1993) in the previous chapter suggested that IT feedback to V4 might be tuned by learning, i.e. the strength of the IT feedback to V4 related to the amount of the monkeys' training in the task and stimuli. Figure 6.5 shows that the strength of this feedback had a strong effect on the scan path also. As this feedback is strengthened, object-based effects in IT and V4 are both increased (IT receives the effects of object-based attention in V4 as a bottom-up input in addition to V4 receiving object-related feedback from IT) and the scan path is more likely to select target coloured stimuli. The behaviour of the scan path is stable at all higher strengths of IT to V4 feedback. There is a slight increase in the number of fixations in blank areas of the display when IT feedback is very strongly weighted. This may be due to the feedback being inhibitory and becoming overly suppressive at stimulus locations in V4 when the feedback is too strong.

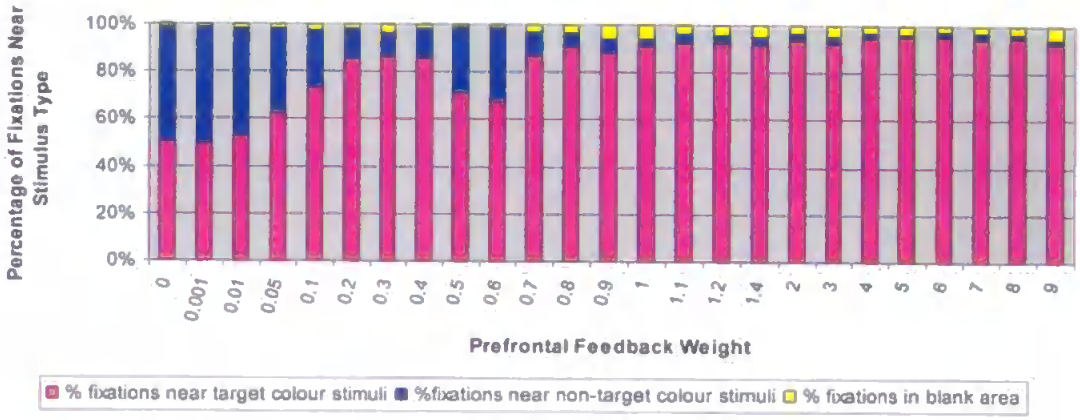



**Figure 6.6 The effect on fixation position of varying the weight of prefrontal feedback to IT**

The original data is presented in appendix A3.4. For each parameter setting, average fixation positions were calculated over 10 scan paths, each consisting of 50 fixations, over the image shown in figure 6.1. As the weight of feedback is increased there is a tendency for more target coloured stimuli and less non-target coloured stimuli to be fixated because object-based effects in the ventral stream are stronger. The normal weight of prefrontal feedback is 1.2.

Similar to the effect of IT feedback to V4, stronger prefrontal feedback to IT produces stronger object-based effects and guidance of the scan path. Performance is stable across higher weights of prefrontal feedback. However, there is a slight decrease in the number of target coloured objects capturing attention when prefrontal feedback weighted around the values 0.5 and 0.6, as shown in figure 6.6. This has been confirmed in subsequent simulations (see appendix A3.4). The reason for this is unclear and is difficult to establish due to the complex dynamical

interactions within the system. However, it is not likely to be a significant finding. Overall, the effect of increasing the weight of prefrontal feedback to IT is to increase the object-based effect in the model's ventral stream and strengthen the guidance of the scan path towards target coloured locations.

In summary, stronger object-related feedback with the model's ventral stream increases selectivity in the search scan path.

#### 6.2.2.2 Object-Related Feedback and Saccade Onset

The strength of object-based attention in the model's ventral stream also has an effect on how quickly a saccade is made, i.e. the duration of the fixation. This is due to the link between the development of a significant object-based effect in IT and the initiation of a saccade. This was examined for the single cell simulations in chapter 5. These simulations revealed two factors that influence saccade onset time. Firstly, reducing the weight of IT feedback to V4 in order to replicate data from less trained monkeys resulted in a slightly longer latency to saccade. More significantly, saccade onset time was dependant on the nature and latency of object-based prefrontal feedback to IT, also suggested to be related to learning. Therefore, saccade onset time may depend on the familiarity with objects in the display and the task.

The same effect was found here for scan path simulations. This is not surprising since the cellular level effects are used to drive the systems level behaviour. Plots of activity in IT during scan path simulations were given in chapter 5 and were very similar to those for the single cell simulations. The effect on saccade onset time of varying the strength of IT feedback to V4 during scan path simulations is shown in figure 6.7. Initially, increasing the weight of object-based feedback from IT to V4 results in the latency to saccade being reduced because the object-based effect in IT is reinforced over time by the feedforward inputs from V4 (Lanyon & Denham, 2004b). However, when feedback is very highly weighted, fixation duration tends to increase. This is thought to be due to the strong suppressive effect that the feedback then has in V4, as discussed above, which means that bottom up inputs to IT are weakened. At these values there is also an increase in fixations in blank areas. So, these effects may combine to reduce the bottom-up input to IT and make IT less able to quickly differentiate targets and non-targets. These higher values of IT feedback (over 6) are never used in simulations (the strongest value used is 5) and the results here indicate that such strong feedback may be beyond the tolerance of the system in that it slightly alters scan path behaviour.

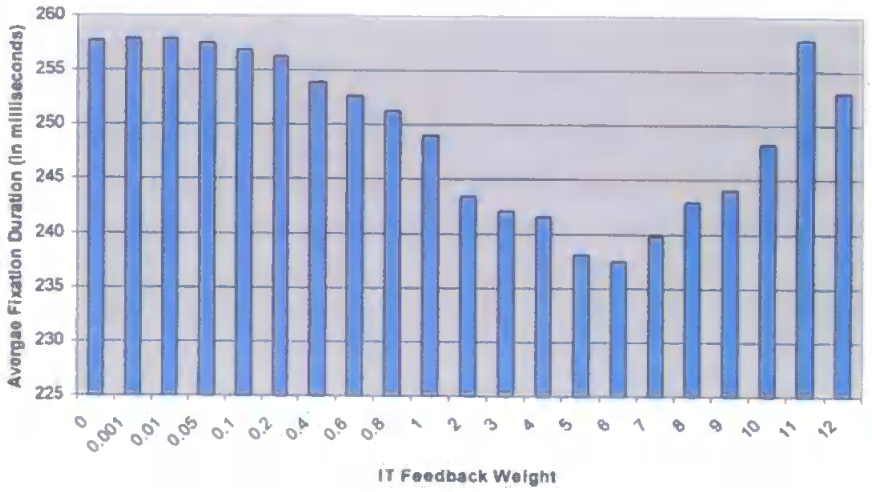



**Figure 6.7 The effect on fixation duration of varying the weight of IT feedback to V4**
The original data is taken from the same simulations used to create figure 6.5. IT feedback is normally weighted at 5 for scan path simulations. It is weighted at 2.5 for certain single cell simulations, as described in chapter 5.

If the weight of the IT feedback is set to the value (5) used to replicate the single cell recordings of Chelazzi et al. (2001), scan path saccades take place about 230-240ms after the onset of the fixation. For the scene shown in figure 6.1, this results in saccades taking place, on average, at 238ms post-stimulus. Chelazzi et al. (2001) found that saccades took place on average 237-240ms post-stimulus, depending on stimulus configuration in relation to the receptive field. Although, the images used here are different to those of Chelazzi et al., this comparison indicates that fixation durations here are within a reasonable range with respect to biological data.

The link between saccade onset and the timing of prefrontal feedback to IT was discussed in chapter 5. The effect of varying the weight of prefrontal feedback to IT was also examined, and the results presented in figure 6.8. In this figure, fixation durations were capped at 350ms (for reasons of processing speed). However, this does not indicate clearly the effect of low weights of prefrontal feedback on fixation duration. A subset of simulations in this range of weights were re-run with the fixation duration capped at 1000ms. These results are shown in figure 6.9, which indicates that there is a steep trend towards increasing fixation duration with decreasing prefrontal feedback. Very weak prefrontal feedback leads to extremely long fixations. This is reasonable since the prefrontal feedback triggers the object-based effects in IT on which saccade onset is based. Figures 6.8 and 6.9 show that there is a critical range of feedback weights over which fixation duration rapidly decreases. Weaker feedback weights cause substantial delay in saccade onset (or, possibly, saccades never taking place). Above the critical range of weights saccade onset time becomes reasonably stable. Thus, the weight of prefrontal feedback to the system is crucial to enable the decision to saccade to be made but, beyond the critical limit, increasing the weight of feedback further has little impact on saccade onset. Therefore, the timing of this feedback (examined in chapter 5) is more critical that its weight in determining fixation duration.

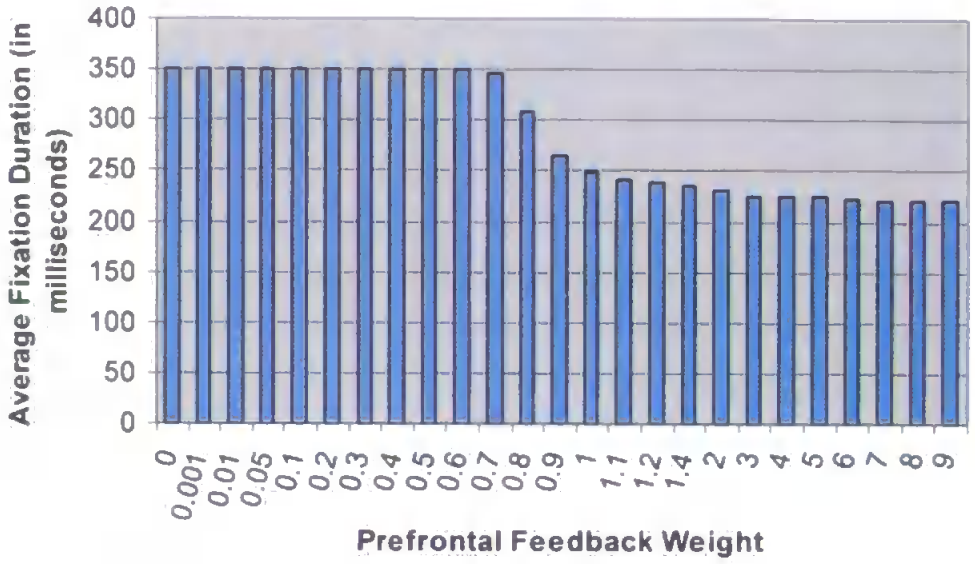



**Figure 6.8 The effect on fixation duration of varying the weight of prefrontal feedback to IT**

The original data is taken from the same simulations used to create figure 6.6. The duration has been capped at 350ms. Prefrontal feedback is normally weighted at 1.2 for all simulations.

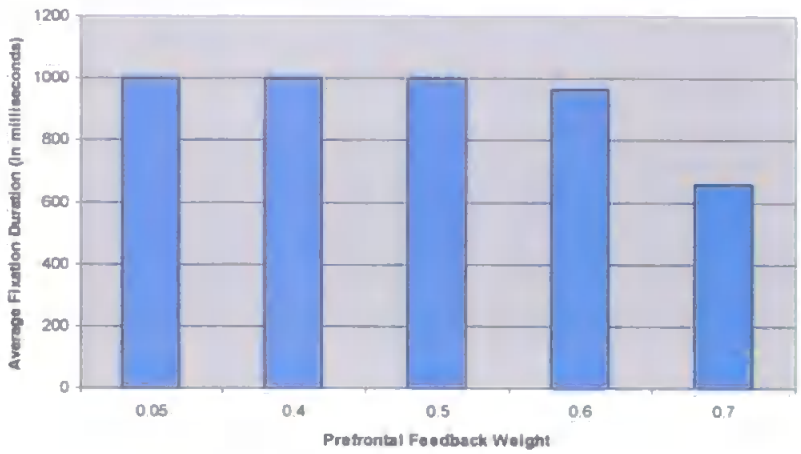



**Figure 6.9 Fixation duration at low weights of prefrontal feedback to IT**

The maximum duration was capped at 1000ms. The original data are shown in appendix A3.4.

The decision to saccade is based on the development of a significant object-based effect developing in IT. Thus, the system only makes a saccade after object-based attention has developed and the target object effect is significant in V4 (this begins from ~150ms post-fixation). As V4 inputs bias competition in LIP, this means that the target for the saccade is a behaviourally relevant stimulus. If saccades were to take place earlier, it is predicted that this guidance of the scan path would be weaker, i.e. less behaviourally relevant stimuli may capture attention. This is because, at that time, activity in LIP at all stimulus locations is approximately equally active, i.e. behavioural relevance is not represented as was shown in figure 4.11 of chapter 4. If a saccade were to take place at this early stage of fixation, the system would be equally likely to choose a target or non-target coloured location as the next fixation point. Therefore, because object-based attention takes some time to develop at the cellular level, it is predicted that saccades taking place after very short fixation durations might tend to select locations that are less behaviourally relevant and, therefore, make the search less efficient. Direct comparison of fixation duration across psychological studies is difficult due to differences in task difficulty etc. However, Hooge & Erkelens (1999) found that search was more selective and efficient when fixation durations were longer. Also, the time course of discrimination of targets versus non-targets in FEF, to which LIP projects (Blatt et al., 1990) is similar to that shown in LIP here and is delayed when distractor-target similarity is increased making search less efficient (Sato & Murthy, 2001). Here, this effect is attributed to featural similarity in the ventral stream leading to less clear object-based effects therein and, hence, weaker discrimination in LIP.

Hence, it is predicted that distractor stimuli (specifically, stimuli not sharing a target feature) may be more likely to capture attention when object-based effects within the ventral stream are weak (possibly due to unfamiliarity with the objects or the task) or have not yet had time to develop fully when the saccade is initiated. In particular, the tendency for the scan path to select only target coloured locations, during search for a colour and form conjunction target, may be weaker when the objects or task are less well known. During search in crowded scenes, as well as in the smaller arrays used for the single cell simulations, fixation duration would be expected to be shorter amongst very familiar objects or in a familiar task. Therefore search may be faster under these conditions. In experiments, search has been found to be facilitated, enabling faster manual reaction times, when distractors are familiar (Lublow & Kaplan, 1997; Reicher, Snyder & Richards, 1976; Richards & Reicher, 1978; Wang, Cavanagh & Greene, 1994). However, Greene & Rayner (2001) suggest that this *familiarity effect* may be due to the span of effective processing (which may be likened to the size of AW here) being wider around familiar distractors, rather than fixation duration being shorter.

Overall, results here indicate that search during a familiar task or with familiar objects may be faster and the scan path would be expected to fixate more target coloured locations than when the task or objects were unfamiliar.

## 6.3 Effect of Scene Density on Saccade Amplitude

Motter and Belky (1998b) found that saccades tend to be shorter in dense scenes compared to sparse scenes. This behaviour can be replicated here, as can be seen by comparison of figures 6.1, 6.2 and 6.3. The average amplitude of saccades in the dense scene was 7.43°, in the sparse scene was 12.1° and was 18.8° in the very sparse scene. Thus, the average saccade amplitude was much larger in the sparse compared to dense scenes (Lanyon & Denham, 2004b). The same sized retinal image was used in all three simulations so stimuli at the same distances were potentially capable of attracting attention. However, the spatial AW is scaled based on local stimulus density such that, when fixation is placed in an area of dense stimuli, the spatial AW is smaller than that when stimuli are sparse. Table 4.1 of chapter 4 showed that AW radius decreases as stimulus density increases. The AW contributes a positive bias to the competition in LIP and results in locations containing target features within the AW being favoured within LIP and being most likely to attract attention. Therefore, it is predicted that saccade amplitude is dependant on the stimulus density in the local area around fixation.

The effect of stimulus density on saccade amplitude is summarised in table 6.1, which shows the average saccade amplitude during scan path simulations over the three images shown in figures 6.1 to 6.3. These images are the same size but contain different amounts of stimuli. The table shows that average saccade amplitude decreases with increasing stimulus density. This result follows from

table 4.1 in chapter 4, which showed that the radius of the AW decreased as stimulus density increased.

| Image (size 1000x1000 pixels) | Average Saccade Amplitude (in Pixels) | Average Saccade Amplitude (in Degrees) |
|---|---|---|
| **1000rv11 – contains 23 bars** | 19 | 2 |
| **1000gh50 – contains 101 bars** | 12 | 1 |
| **1000rh150 – contains 301 bars** | 7 | 0.6 |

**Table 6.1 Saccadic Amplitude**

Summary of the saccadic amplitude from scan paths over three images of differing stimulus density. The data is taken from the simulations given in appendix A3.7 to A3.9 (the weight of the novelty bias was varied in these simulations, but the data shown here uses the normal setting of this parameter, i.e. 0.0009).

The effect of scaling the AW on the basis of stimulus density is best demonstrated with an image that contains uneven stimulus density. Figure 6.10 shows a scan path through an image that contains two patches of dense stimuli in a background of sparser, but still reasonably dense, stimuli. The AW is scaled using local stimulus information, i.e. that within the current retinal view, so that it expands and contracts as it moves around a scene of mixed density. Therefore, the spatial attentional focus zooms in and out as it moves around the scene. This results in smaller saccades in the dense areas of a scene and larger saccades in the sparser

areas. Hence, the scan path is able to investigate the dense, and potentially more interesting, areas of a scene with a series of shorter amplitude saccades, as can be seen in figure 6.10 (Lanyon & Denham, 2005b). Although not investigated here, this is potentially useful for natural scene processing: Uniform areas, such as sky, could be examined quickly with large amplitude saccades, allowing the scan path to concentrate in more detailed and interesting aspects of the scene.

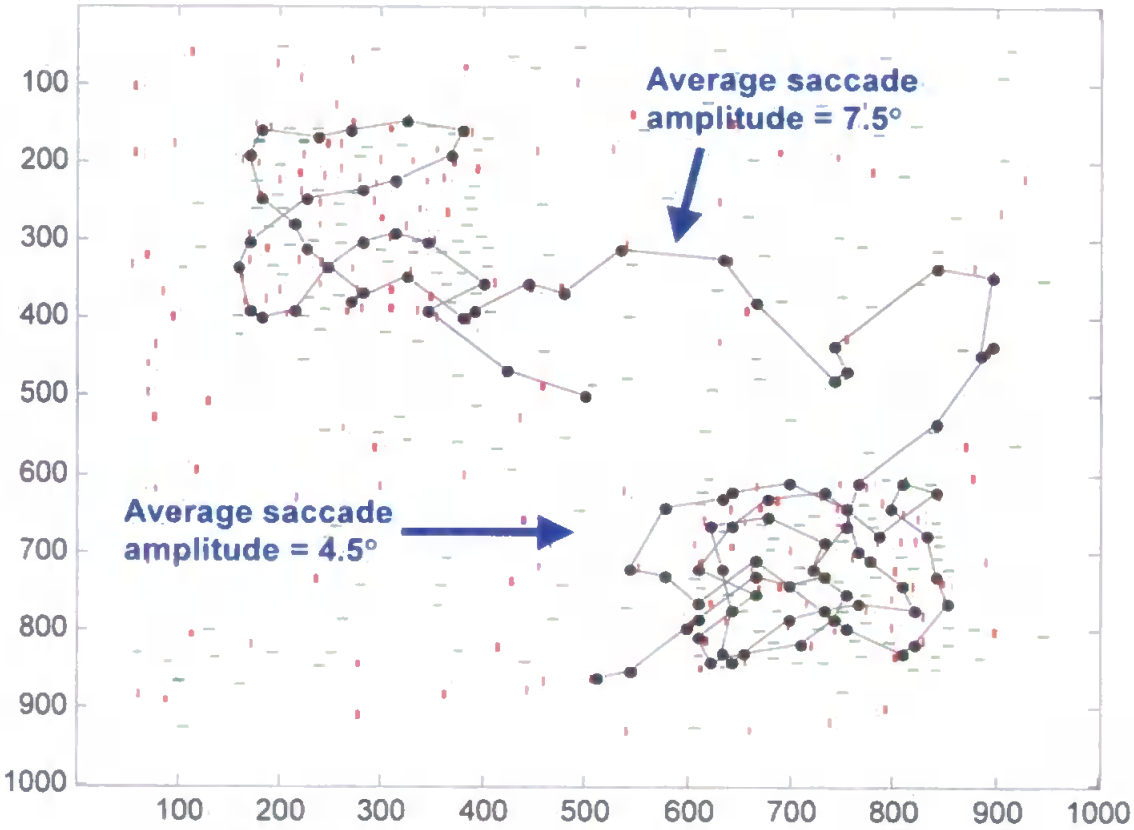



**Figure 6.10 A scan path through a scene of mixed stimulus density**

Stimuli are reasonably dense throughout the scene but there are two patches of slightly more dense stimuli in the bottom right and top left quadrants. The scan path examines these dense areas more thoroughly with a series of shorter saccades (average amplitude 4.5° in dense areas, compared to 7.5° in slightly more sparse areas).

Crucial to this ability to investigate dense areas of the scene with smaller saccades is the update to the novelty map. If too large a region is inhibited in a dense area of stimuli, the scan path is forced to move further afield. Therefore, updates to the novelty map are also based on the size of the AW. This means that a smaller region is inhibited for return in areas of dense stimuli. A larger region is inhibited in sparse areas of the scene so that such uninformative regions are considered to have been inspected. Thus, saccades are less likely to be made to neighbouring locations in these regions so that they are moved through quickly and the scan path is able to reposition itself towards more interesting regions. Figure 6.11 shows the values in the novelty map at the end of the scan path shown in figure 6.10. The area of novelty reduction is smaller in the dense patch of stimuli than the sparse regions to which the scan path is moving at the end, shown by the final couple of large areas of novelty reduction. The novelty map is considered further in a later section about inhibition of return.

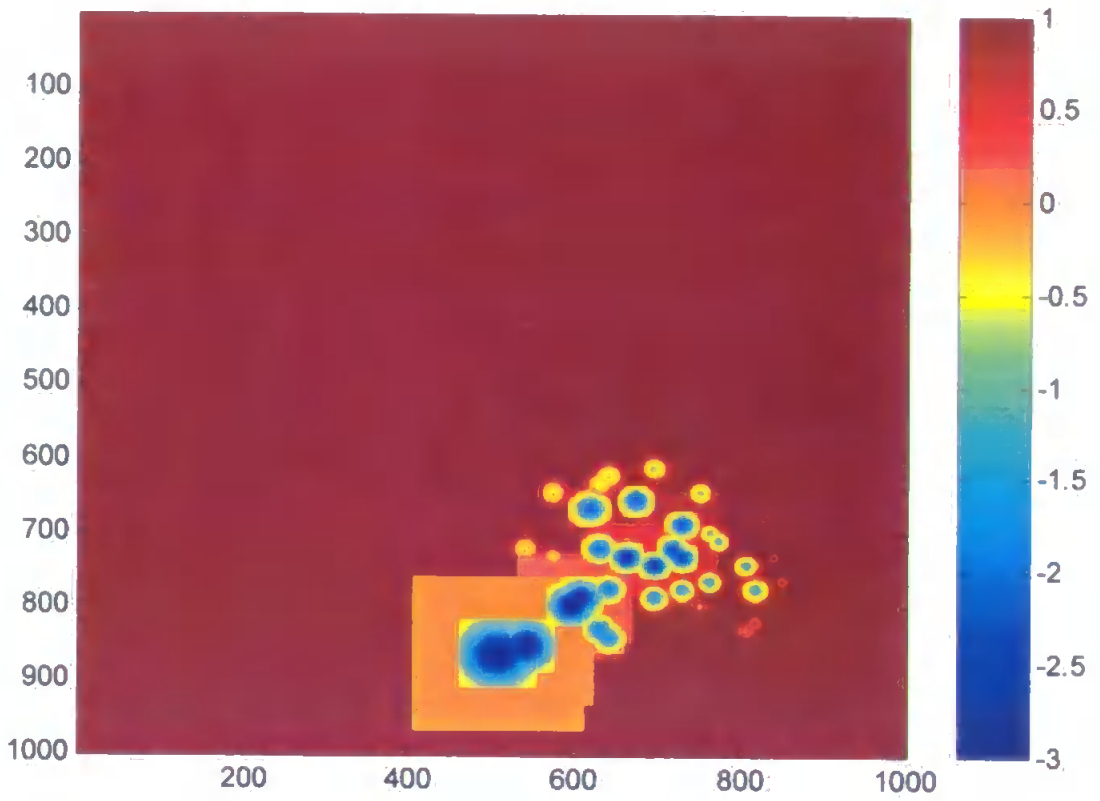



**Figure 6.11 Values in the novelty map after the scan path shown in figure 6.10**

Recently fixated locations in the area of dense stimuli have a small region of reduced novelty. Locations that were fixated early in the scan path have recovered novelty over time. As the scan path leaves the dense area (the final two fixations), the size of the region of reduced novelty is larger than that in the dense area. This is due to the region of reduced of novelty being based on the size of the current AW, which zooms in and out as it moves around the scene. It is necessary for the reduced novelty region to be similarly scaled according to local stimulus density in order to allow detailed inspection of the dense areas.

## 6.4 Retinal Image Size

Provided it has a reasonably large radius, the size of the retinal image does not appear to have a great effect on the nature of stimuli selected by the scan path. Obviously, only stimuli within the current retinal view can be selected as potential saccade targets (because the cortical areas are only aware of these stimuli), but those chosen tend to be target coloured regardless of retinal image size. Thus, guidance of the scan path remains consistent provided the retinal image is of reasonable size. The size at which performance stabilises depends on the density of stimuli in the scene. Figures 6.12 and 6.13 show fixation positions with respect to stimulus colour over a range of retinal image sizes for two images: 1000rh150, the dense image shown in figure 6.1 above and 1000gh50, the sparse image shown in figure 6.2 above. Behaviour is very similar when the retinal radius is set to approximately 10° or more for the dense image, and is set to approximately 20° or more for the sparse image. The original data are given in appendix A3.5 and A3.6. If the retina is very small, say 1° in radius, the scan path may never be able to move beyond its starting position at the centre of the image because no stimuli are available within the retina to be analysed as potential saccade targets. Obviously, this is an extreme, and not biologically plausible, situation.

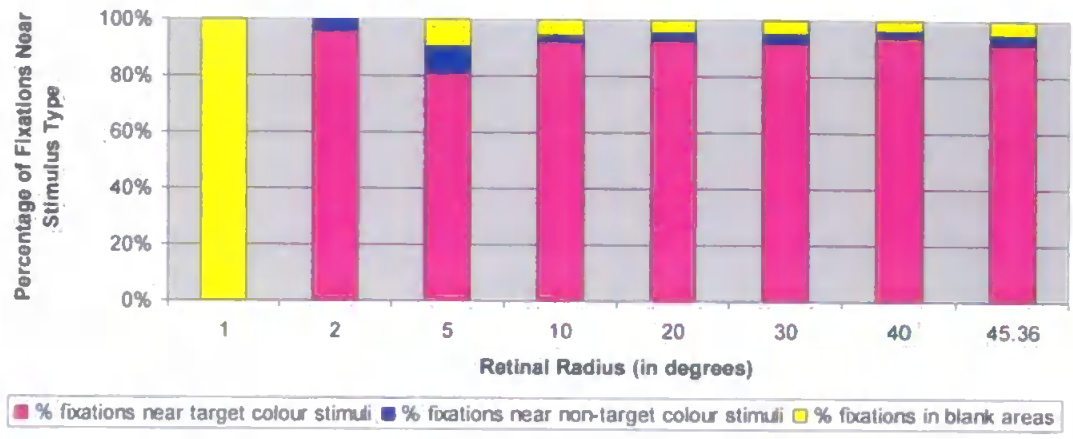



**Figure 6.12 The effect on fixation position of increasing the retinal radius for scan paths over a dense image**

Fixation positions were averaged over 10 scan paths, each consisting of 50 fixations over image 1000rh150. For this image, performance stabilises for retinal radii over ~10°, which is a field of view of 20°. The upper bin is set to the maximum size possible in the original image, i.e. the retina is the same size as the entire image.

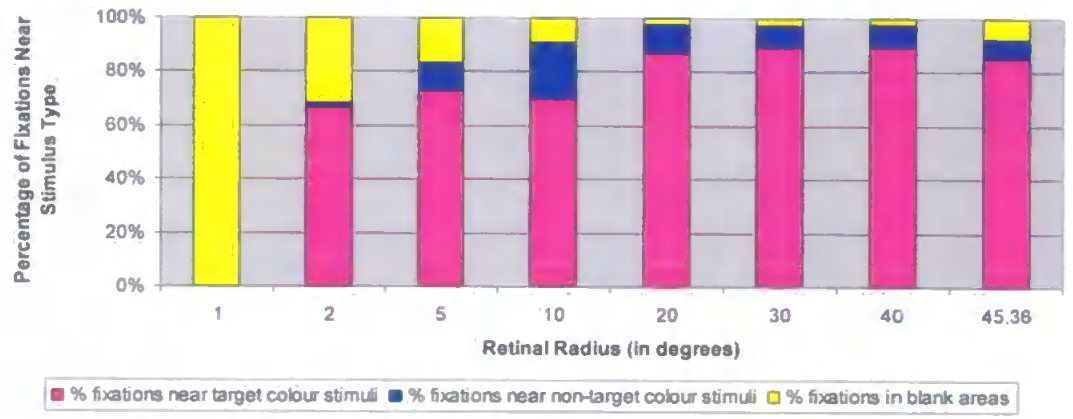



**Figure 6.13 The effect on fixation position of increasing the retinal radius for scan paths over a sparse image**

Fixation positions were averaged over 10 scan paths, each consisting of 50 fixations, over image 1000gh50. For this image, performance stabilises for retinal radii over ~20°, which is a field of view of 40°. Again, the upper bin is set to the maximum size possible in the original image.

Figures 6.14 and 6.15 show that the reason why guidance of the scan path by target colour is less consistent at small retinal radii is that either saccades are unable to move fixation position from its starting position (e.g. 1° radii case) or saccades are very small in amplitude. These small amplitude saccades either move around in blank areas, because of lack of stimuli within the small retina, or shift fixation repetitively between a very few stimuli within the retina.

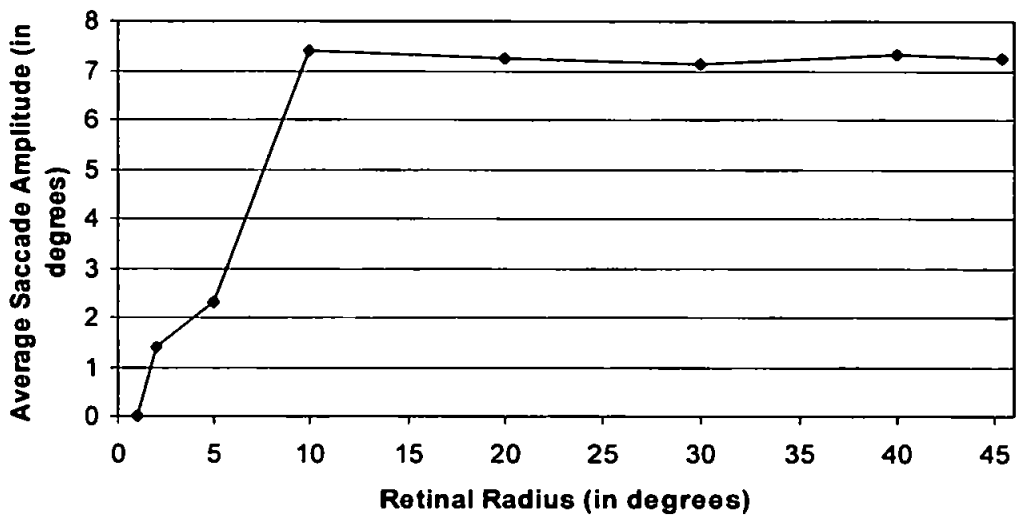



**Figure 6.14 The effect on saccade amplitude of increasing the retinal radius for scan paths over a dense image**

This figure relates to the same simulations as figure 6.12, over image 1000rh150. For this image, a retinal radius of less than ~10° produces scan paths that are unable to explore the scene.

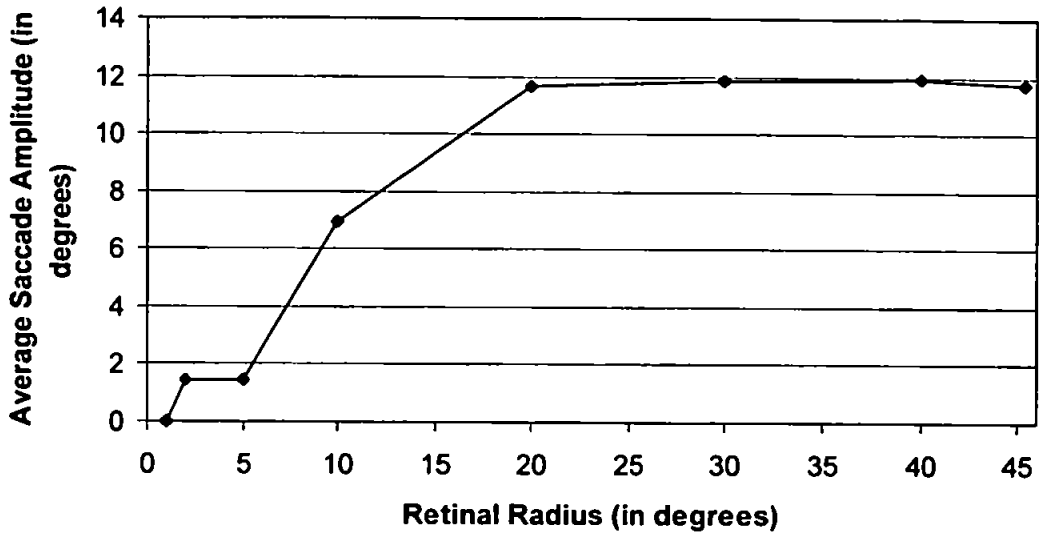



**Figure 6.15 The effect on saccade amplitude of increasing the retinal radius for scan paths over a sparse image**

This figure relates to the same simulations as figure 6.13, over image 1000gh50. For this image, a retinal radius of less than ~20° produces scan paths that are unable to explore the scene.

At small retinal radii, the radius of the AW is restricted to the maximum aperture of the retinal image. Once this restriction is removed, the AW is able to be scaled correctly and its size is consistent despite further increases in retinal image size. So, above the threshold value for the retinal radius, the radius of the AW is consistently scaled according to stimulus density. Figures 6.16 and 6.17 show this effect for the dense and sparse images respectively.

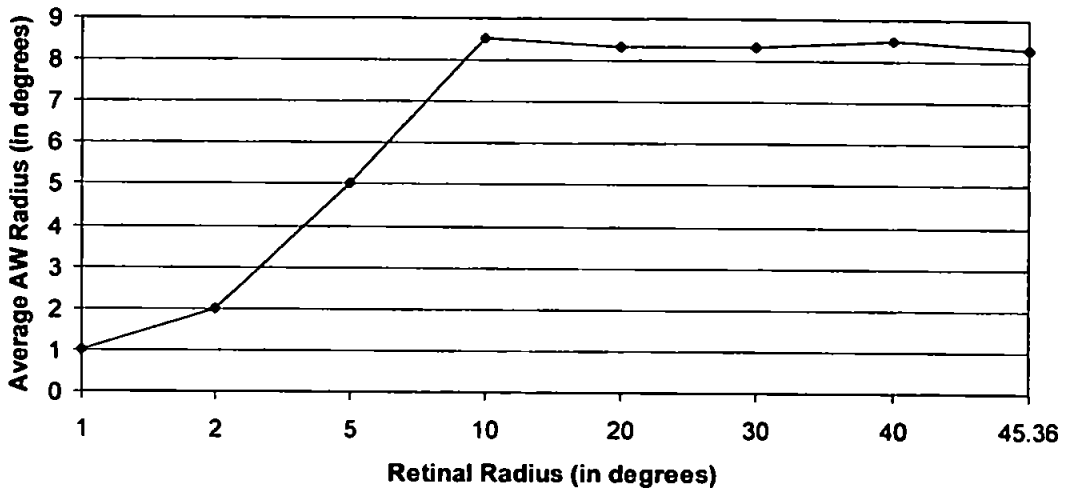



**Figure 6.16 The effect on AW radius of increasing the retinal radius for scan paths over a dense image**

This figure relates to the same simulations as figure 6.12, over image 1000rh150. For this image, the AW cannot be scaled correctly when the retinal radius is less than ~10°.

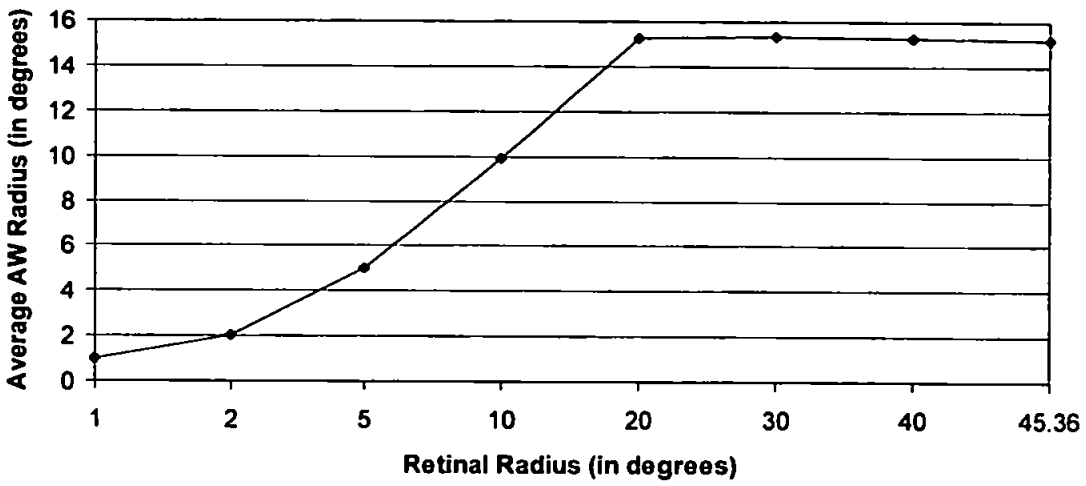



**Figure 6.17 The effect on AW radius of increasing the retinal radius for scan paths over a sparse image**

This figure relates to the same simulations as figure 6.13, over image 1000gh50. For this image, the AW cannot be scaled correctly when the retinal radius is less than ~20°.

Therefore, provided the retina is larger than a threshold size, results are largely independent of retinal radius. Although not systematically analysed across different images, this threshold appears to depend on the density of stimuli in the scene. When the retina is smaller than this threshold, the AW cannot be scaled correctly according to stimulus density and the scan path is unable to explore the scene fully.

Figure 6.18 demonstrates a problem with a small retina image in a sparse scene (Lanyon & Denham, 2004b). The scan path is reasonably successful with large amplitude saccades that explore the scene reasonably well. However, the small retina in such a sparse scene has restricted the scan path by making it unable to reach target coloured stimuli in the bottom left corner of the display. These novel stimulus were not available within the retinal image at nearby fixations so could not be chosen as the next fixation point. When the retinal image is increased in size, more stimuli are available to the retina and this allows the stimuli at the extremity to attract attention, as shown in figure 6.19. In both figures, target coloured stimuli tend to be fixated more than non-target coloured stimuli or blank areas. Thus, altering the retinal image size does not affect the nature of the stimuli that attract attention but it does alter the course of the scan path by changing the extent of the scene available to cortex at each fixation. Therefore, it is predicted that, below a critical size, field of view can have a dramatic effect on the scan path. This could be likened to performing search in real life where the field of view was restricted by looking through a tube.

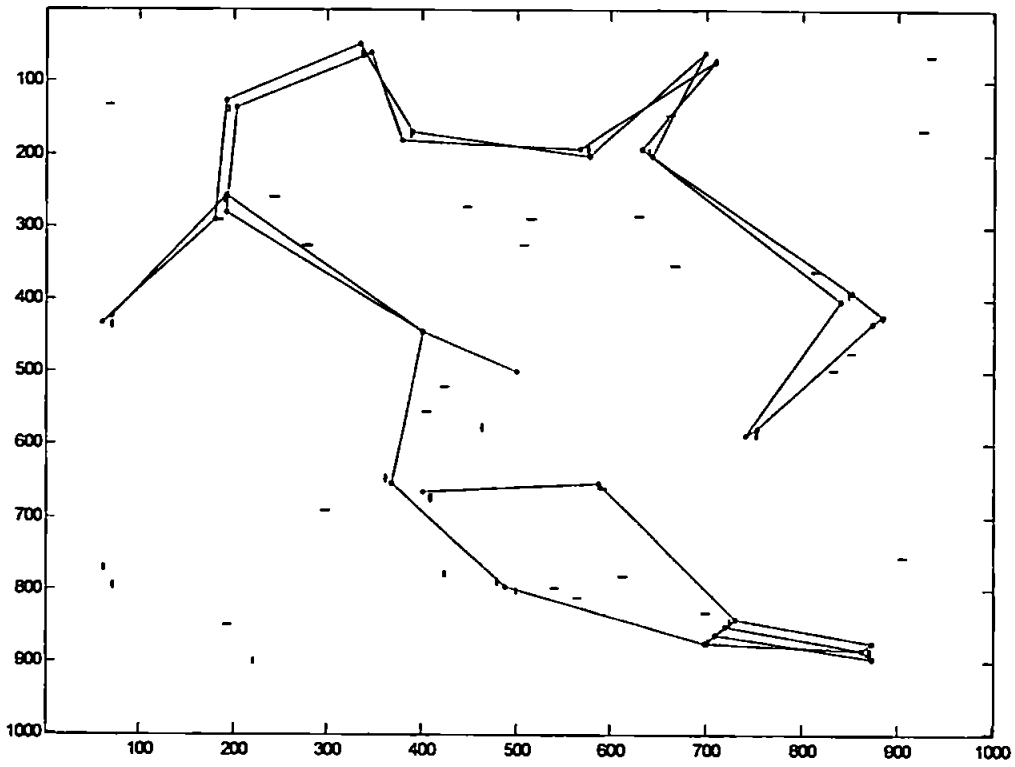



**Figure 6.18 A scan path in a sparse scene where the retinal image size is below the threshold at which stable performance is achieved**

The retinal radius is 20°, which creates a retinal image of size 441x441 pixels, ~40°. In this particular scan path, target coloured stimuli in the bottom left of the image have not been examined because nearby fixations never captured these stimuli within the retina. This problem only occurs in sparse images when the retina image is restricted in size (and is much smaller than that normally available to humans and monkeys).

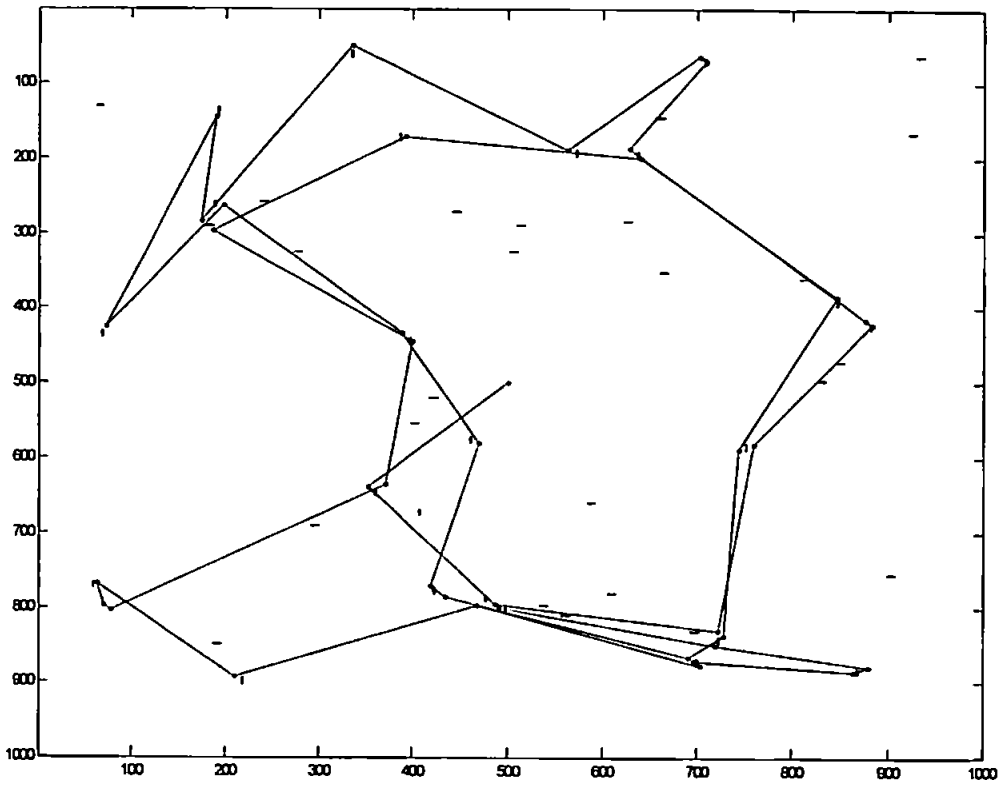



**Figure 6.19 A scan path in a sparse scene where the retinal image size has been increased**

Another scan path over the image in figure 6.18 but with the retina enlarged to 801x801 pixels (~73°). All target coloured stimuli are now examined.

## 6.5 Inhibition of Return in the Scan Path

Implementation of the novelty bias to LIP successfully produces IOR within the scan path that declines linearly over time allowing the eventual re-visiting of sites previously inspected. The ability of the scan path to move around the scene tending to examine novel locations can be seen in the scan paths already presented in this chapter. The novelty map at the end of the scan paths shown in figures 6.1, 6.2 and 6.3 are shown in figures 6.20, 6.21 and 6.22 respectively. Comparison of these figures shows how the novelty map update is related to the size of AW, so that a smaller area receives a reduction of novelty in dense, compared to sparse, scenes (Lanyon & Denham, 2004b). Figure 6.11 above showed that this area scales with local stimulus density so that the novelty update zooms in and out as stimuli become more or less dense during the scan path. This helps the scan path to move through sparser areas with larger saccades but thoroughly examine any dense areas it encounters with a series of small saccades.

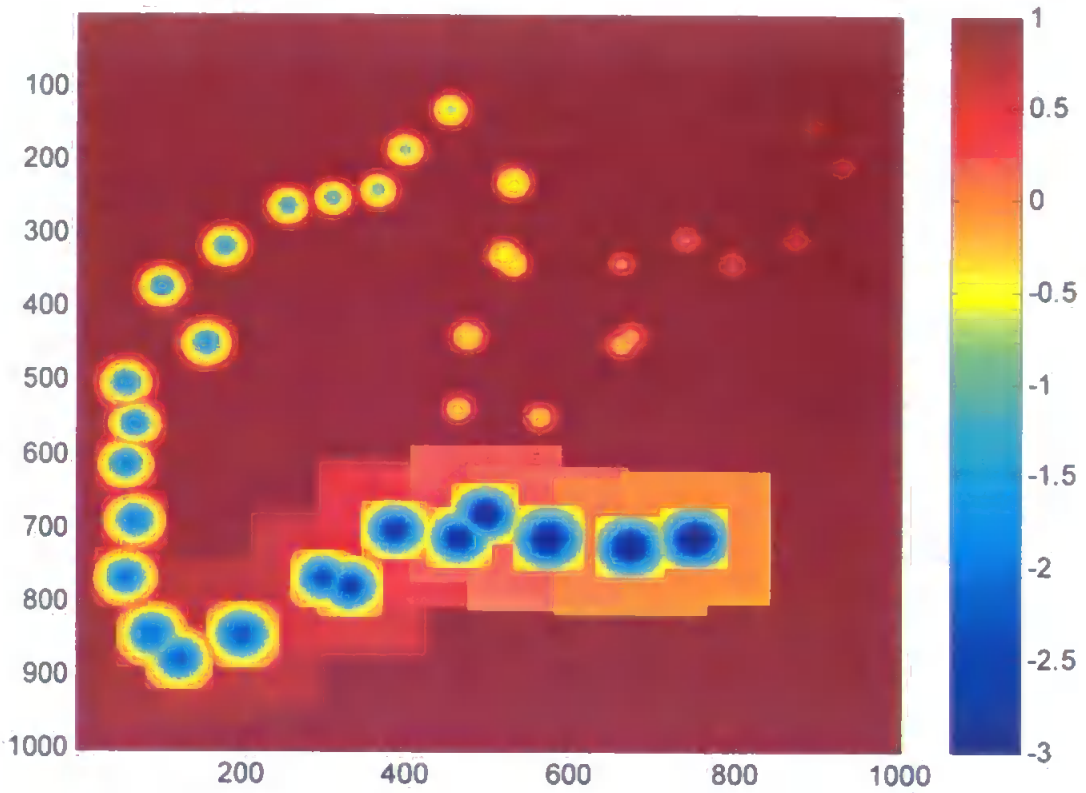



**Figure 6.20 Novelty map in a dense scene**

The novelty map after the scan path shown in figure 6.1. The most recently fixated locations have the lowest novelty. Novelty has recovered at locations fixated earlier in the scan path. The extent of the area of reduced novelty at each fixation is quite small because the stimuli in this scene are relatively dense and, therefore, the AW is narrow.

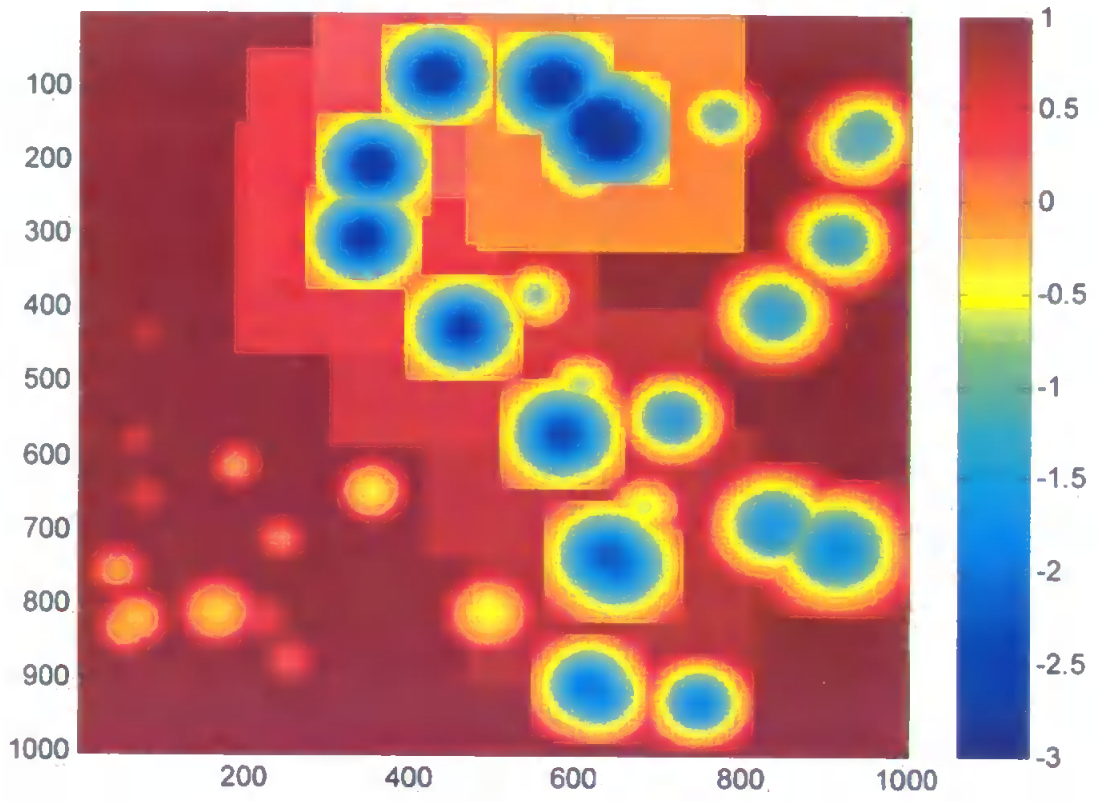



**Figure 6.21 Novelty map in a sparse scene**

The novelty map after the scan path shown in figure 6.2. Again, the most recently fixated locations have the lowest novelty and low values have partially recovered at locations fixated earlier in the scan path. The extent of the area of reduced novelty at each fixation is larger than that shown in figure 6.20 because the stimuli in this scene are more sparsely arranged and, therefore, the AW is wider.

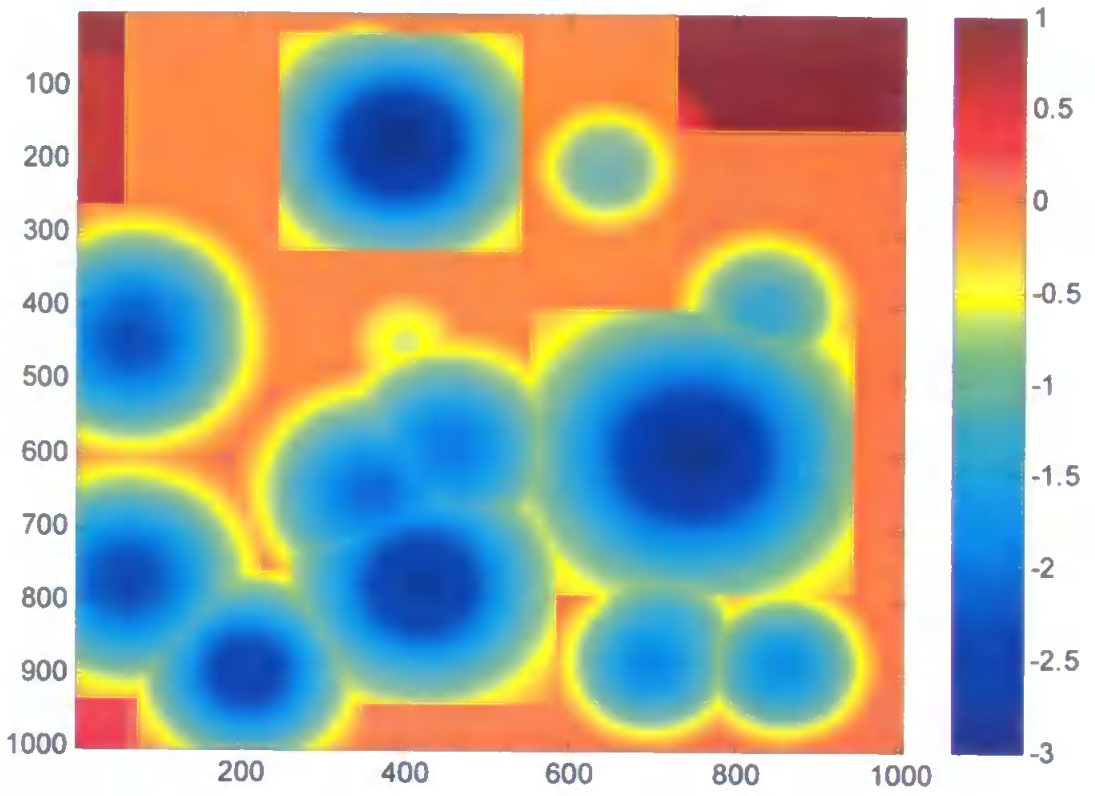



**Figure 6.22 Novelty map in a very sparse scene**

The novelty map after the scan path shown in figure 6.3. The extent of the area of reduced novelty at each fixation is much larger due to the stimuli in this scene being very sparsely arranged.

The nature of inhibition of previously attended locations here is intended to broadly replicate a range of psychophysical data. Figures 6.20 to 6.22 show that during a scan path novelty may be present at multiple locations, as found by Danziger et al. (1998), Snyder and Kingston (2001) and Tipper et al. (1996). Also, the novelty values here recover over time during the course of the scan path, similar to the recovery of inhibited locations over time in psychophysical tasks (Irwin &

Zelinsky, 2002; Snyder & Kingston, 2000). Following the withdrawal of attention from a location, the novelty map is updated such that the reduction in novelty is strongest in the immediate vicinity of the fixation point with the strength of reduction decreasing with distance from the fixation point. This means that saccades are inhibited from returning to this region but the strength of this inhibition declines with distance from the original fixation point, similar to the effect observed by Hooge and Frens (2000).

The total value in the novelty map over the area of a LIP assembly's receptive field is applied as a biasing current that influences the dynamics of the assembly. The bias is applied with a weighting factor (parameter $\eta$ in equation 4.11 of chapter 4), which determines the strength, or importance, of novelty to the search process. The strength of this bias was found to influence the scan path. Appendices A3.7 to A3.9 present data from simulations that examined the effect of varying the strength of this bias on scan path behaviour. These data are summarised in the figures here.

The novelty bias has a direct effect on re-fixation rate, i.e. the proportion of fixations that occur at a location already fixated by the scan path. Figures 6.23 to 6.25 show the effect of varying the weight of the novelty bias for search in a dense, sparse and very sparse scene respectively. The normal parameter setting for the weight of novelty bias is $9.0E^{-4}$ (i.e. 0.0009). If the weight is set less than $9.0E^{-5}$, an order of magnitude less than the normal setting, the tendency for the scan path to re-fixate locations is dramatically increased. Hence, there is an order of magnitude

flexibility in reduction of the weight of novelty feedback in this model. Increasing the weight of the novelty bias beyond its normal setting has little effect on re-fixation rates.

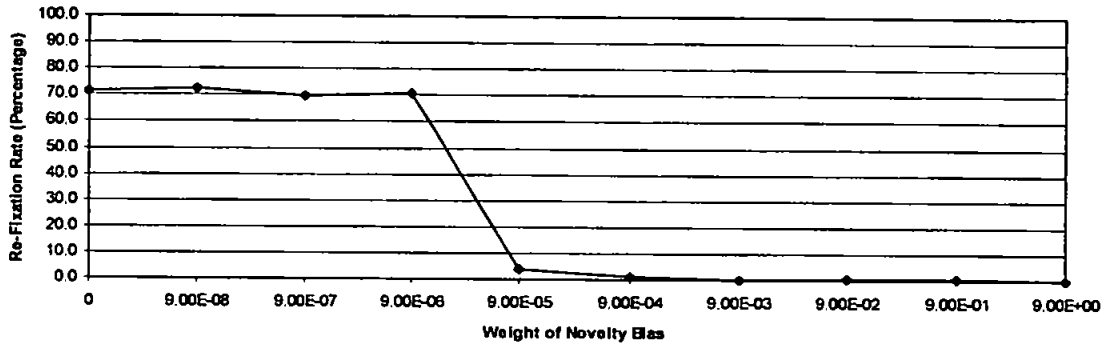



**Figure 6.23 The effect of the varying the novelty bias on re-fixation rate for scan paths in a dense scene**

Shows the average re-fixation rate (expressed as a percentage) found over 10 scan paths, each consisting of 50 fixations, over the image shown in figure 6.1.

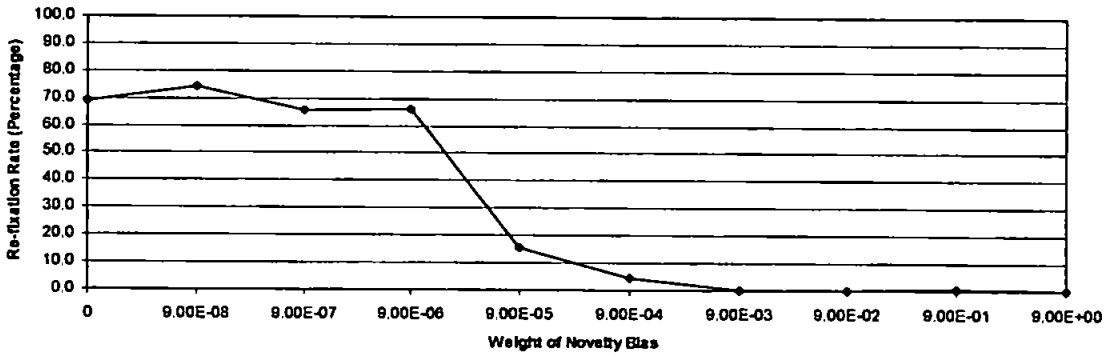



**Figure 6.24 The effect of the varying the novelty bias on re-fixation rate for scan paths in a sparse scene**

Shows the average re-fixation rate (expressed as a percentage) found over 10 scan paths, each consisting of 50 fixations, over the image shown in figure 6.2.

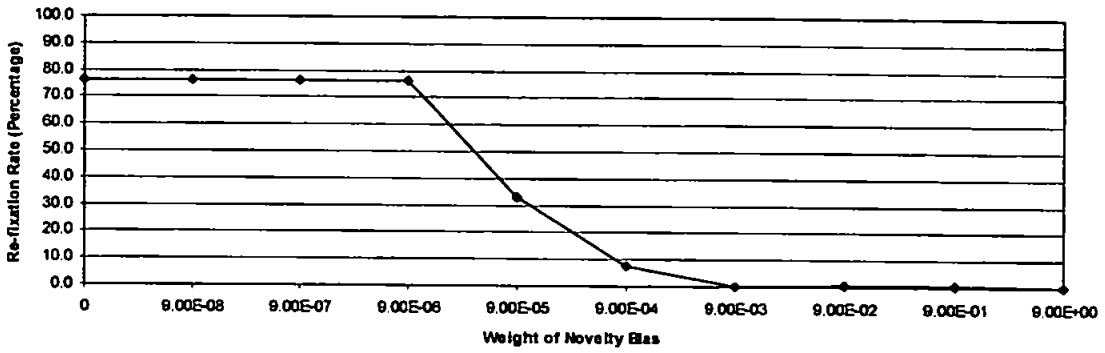



**Figure 6.25 The effect of the varying the novelty bias on re-fixation rate for scan paths in a very sparse scene**

Shows the average re-fixation rate (expressed as a percentage) found over 10 scan paths, each consisting of 50 fixations, over the image shown in figure 6.3.

Increasing the weight of the novelty bias also has the effect of causing the scan path to move further afield, i.e. increasing average saccade amplitude. Figures 6.26 to 6.28 show the effect of varying the weight of the novelty bias on average saccade amplitude. The amplitude increases with an increase in the importance of novelty but this effect saturates at strong novelty bias. This effect has obvious links with the effect on re-fixation rate.

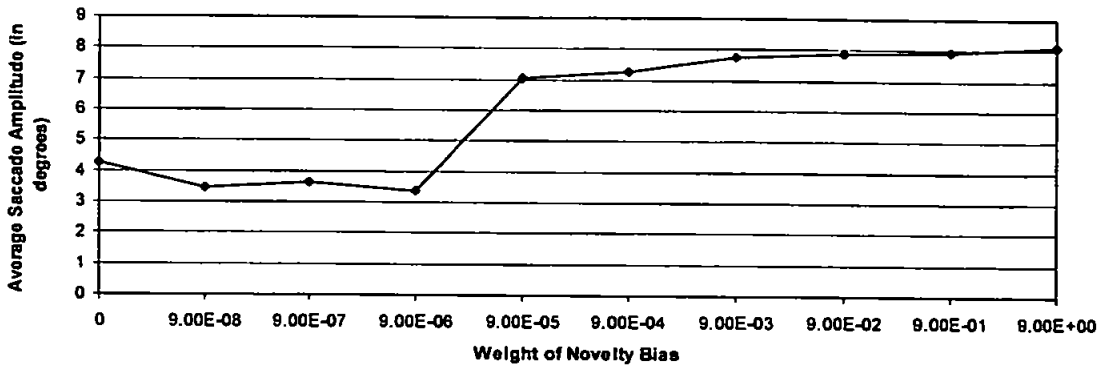



**Figure 6.26 The effect of the varying the novelty bias on average saccade amplitude in a dense scene**

Shows average saccade amplitude from the same simulation series as figure 6.23, over the image shown in figure 6.1.

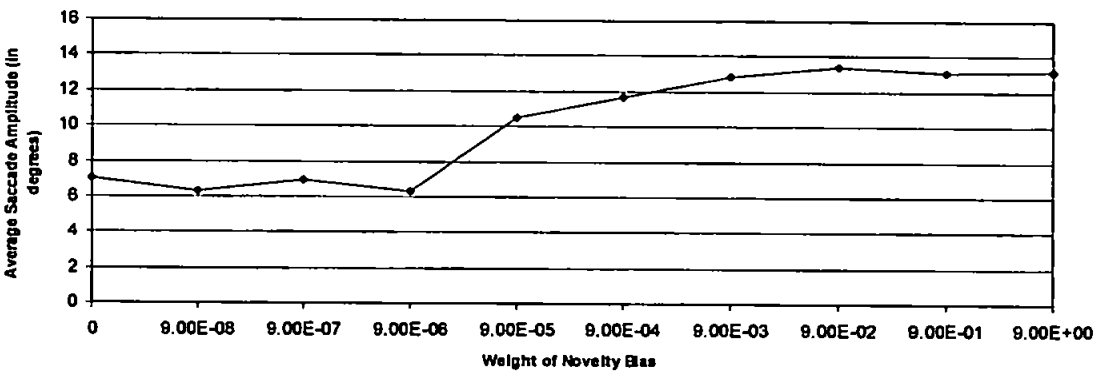



**Figure 6.27 The effect of the varying the novelty bias on average saccade amplitude in a sparse scene**

Shows average saccade amplitude from the same simulation series as figure 6.24, over the image shown in figure 6.2.

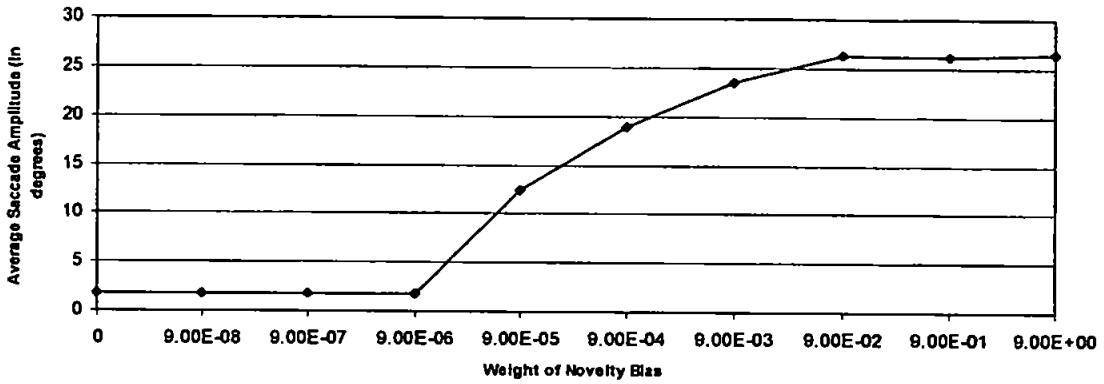



**Figure 6.28 The effect of the varying the novelty bias on average saccade amplitude in a very sparse scene**

Shows average saccade amplitude from the same simulation series as figure 6.25, over the image shown in figure 6.3.

More interestingly, altering the strength of the novelty bias has an effect on the guidance of the scan path. Figures 6.29 to 6.31 show the effect on fixation position of altering the novelty bias for scan paths in a dense, sparse and a very sparse scene respectively. Increasing the weight of the novelty bias to LIP reduces the likelihood of fixating target coloured stimuli and increases the number of fixations landing on non-target coloured stimuli and in blank areas (Lanyon & Denham, 2004b, 2005a). Hence, search becomes less guided. As its importance is increased, novelty has a stronger effect on the competition in LIP than the featural bias from V4. This illustrates the way in which different factors compete for importance in the competition for the capture of attention in LIP. In particular, in sparse scenes, a strong novelty bias dramatically increases the likelihood of fixating in blank areas of the display. However, it should be noted that the number of fixations in blank

areas of the sparse images might be artificially heightened by the fact that a fairly small retinal image was used (40° visual field). Due to this, fixations near the edge of the image have limited access to stimuli due to part of the retina "overflowing" the image and there being very sparse stimuli in the remainder of the retina. However, this trend is still evident in the dense image. This possibly suggests that a search where novelty is the key factor, for example a hasty search of the entire scene, may result in more "wasted" fixations in blank areas.

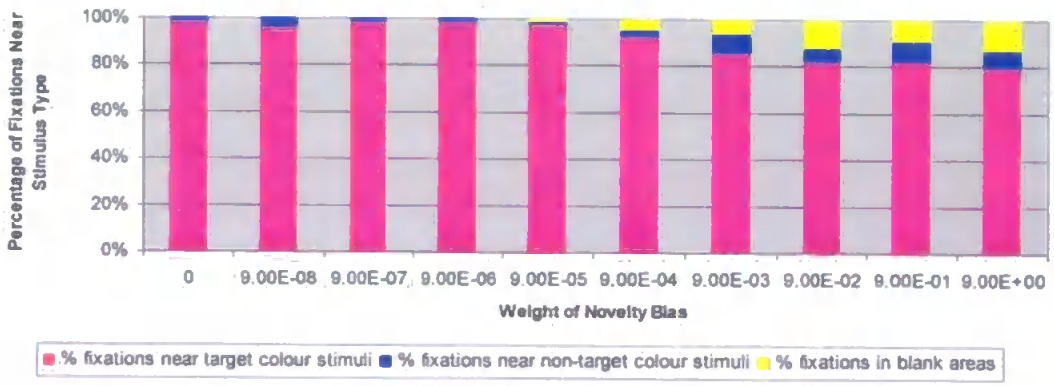



**Figure 6.29 The effect of the varying the novelty bias on fixation position in a dense scene**

Shows average fixation positions from the same simulation series as figure 6.23, over the dense image shown in figure 6.1. As the weight of the novelty bias increases, the number of fixations near target coloured stimuli decreases and fixations are slightly more likely to occur in blank areas of the display than when the novelty bias is weak.

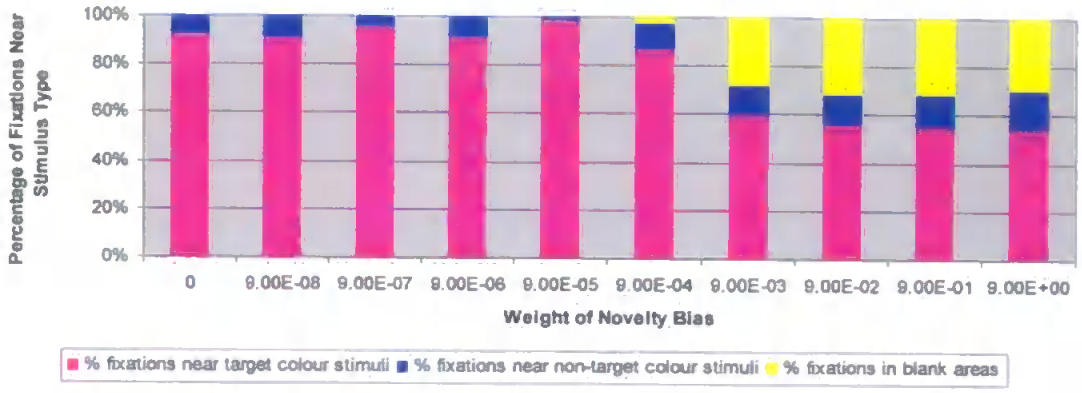



**Figure 6.30 The effect of the varying the novelty bias on fixation position in a sparse scene**

Shows average fixation positions from the same simulation series as figure 6.24, over the sparse image shown in figure 6.2. As the weight of the novelty bias increases, the number of fixations near target coloured stimuli decreases and fixations are more likely to occur in blank areas of the display than when the novelty bias is weak.

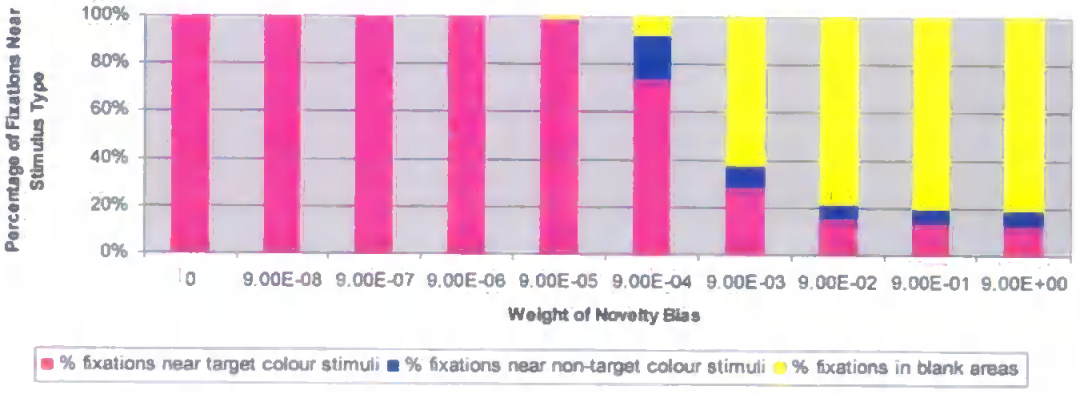



**Figure 6.31 Effect of the varying the novelty bias on fixation position in a very sparse scene.**

Shows average fixation positions from the same simulation series as figure 6.25, over the very sparse image shown in figure 6.3. As the weight of the novelty bias increases, the number of fixations near target coloured stimuli decreases and fixations are much more likely to occur in blank areas of the display than when the novelty bias is weak.

## 6.6 Chapter Summary

The systems level behaviour of the model was examined in this chapter and search scan path behaviour was replicated. The system was able to produce saccades that targeted behaviourally relevant locations in the scene. The feature priority built into the model enabled search to be guided by target colour. Therefore, scan path behaviour was able to replicate behaviour observed by Motter and Belky (1998b) in a monkey psychophysical task using similar stimuli (also Lauria & Strauss, 1975; Scialfa & Joffe, 1998; Williams, 1967; Williams & Reingold, 2001, in human psychophysics)

Systems level results were introduced in chapter 4. Here, the effect of varying various parameters on scan path behaviour was tested. All these parameters have a biological basis and are not computational artefacts. Overall, examination of the effect of altering the model's parameters shows that the model's normal settings are in areas of the parameter space that produce stable behaviour. It is not surprising that these parameters produce the optimum system behaviour, since the system was empirically tuned to these settings. However, this analysis has revealed that the settings are not critical to behaviour. In all cases, small changes to parameter values have almost no effect on the performance of the system. The behaviour remains stable over a range of parameter settings. This is not always the case with other computational models and gives an indication of the robustness of this model.

A summary of the effect of varying the model's parameters is given below.

| Aspect of the Model | Effect of Varying Parameters |
|---|---|
| **Weight of V4 connections to LIP** | A feature priority in the guidance of scan paths is possible with only a very marginal difference in the weight of connection from the various features in V4 to LIP. These results suggest one method by which a feature priority in search scan paths may be generated. |
| **Object-related Feedback in the Ventral Stream** | Effect on Scan Path Guidance<br>Object-related feedback in the model's ventral stream hierarchy creates object-based attention, which is necessary to guide the scan path to behaviourally relevant locations. These object-based effects are stronger when the weight of prefrontal feedback to IT, or IT feedback to V4, is increased. Stronger object-based effects in V4 produce stronger guidance of the search scan path by target colour. In the previous chapter, this object-related feedback was linked with object learning. Therefore, the results here lead to a prediction that familiarity with the search task and objects can increase search selectivity by more strongly guiding the scan path by target features. |

Effect on Saccade Onset

Object-based effects in the ventral stream, specifically those in IT, determine saccade onset. Weaker object-based effects in the ventral stream mean that saccade onset is delayed.

- Prefrontal feedback is crucial to create object-based attention but the weight of this feedback is not critical to saccade onset timing provided that it is above a threshold level.
- Stronger IT feedback decreases fixation duration because of the increased object-based effect in V4, which then feeds forward to IT. However, this effect breaks down at very high levels of IT feedback (probably due to its suppressive effect in V4).

These results lead to a prediction that saccades taking place after a short fixation duration may be less able to target behaviourally relevant locations due to insufficient time during which object-based attention can develop. However, such effects will require less time when object-based feedback is faster (see chapter 5) or stronger, i.e. when the working memory and sensory signal from prefrontal cortex is available earlier or is stronger.

| | |
|---|---|
| **Scene Density & Saccade Amplitude** | Similar to the result found by Motter and Belky (1998b), saccades are shorter in dense scenes than in sparse scenes. Also, the amplitude varies as the scan path moves around a scene of mixed stimulus density. This is due to the scaling of the AW being based on the density of local stimuli, i.e. stimuli within the retinal view.<br>In order to allow the scan path to investigate dense areas with a series of shorter saccades and to move through sparse areas with larger saccades, it was necessary to base the update to the novelty map on the size of the AW. Hence, the novelty update is also scaled based on local stimulus density. This leads to the prediction that a larger area around fixation is inhibited for future saccade return in a sparse scene than in a dense scene. |
| **Retinal Image Size** | The guidance of the scan path by target features is relatively invariant to retinal size provided that the radius is above a certain critical value, which is determined by stimulus density in the scene. However, below this critical value, field of view (retinal size) can have a dramatic effect on the scan path. |

**Inhibition of Return**

IOR, implemented here by means of reductions in values in the novelty map, broadly replicates a range of psychophysical results from visual search. As the strength of the novelty bias to LIP is increased, the following effects are observed:

- A reduction in re-fixation rate, i.e. previously inspected locations are less likely to be re-visited by the scan path.
- A larger average saccade amplitude, i.e. the scan path tends to move further from the current fixation point so that the scan path is encouraged to move further afield.
- A slight reduction in the guidance of the scan path by target colour. There is a tendency, particularly noticeable in sparse scenes, for more blank areas of the display to capture attention. However, the scan path is still primarily guided by target colour. Possibly where novelty is the key factor for search, for example a hasty search of the entire scene, less fixations will gain useful information because more will be "wasted" in blank areas.

## Contribution of the Work

This is the first implementation of biased competition within an active vision paradigm. Cellular level results described in the previous chapter have been used collectively to provide systems level behaviour that mimics psychological data from humans and monkeys. Much of this was outlined in chapter 2. In particular, the search scan path has been guided by target colour in a colour-orientation feature conjunction search task, similar to that described by Motter and Belky (1998b). Thus, it has been possible to implement a feature priority in the guidance of search towards behaviourally relevant stimuli. This was due to the ventral (V4) to dorsal (LIP) stream connection. This suggests that object-based effects in the ventral stream are able to influence cortical and sub-cortical areas involved in the generation of saccades.

The systems level behavioural results are extended in the next chapter to investigate the effect of lesioning parts of the model.

# Chapter 7

# Simulation of Visual Search Behaviour Following Cortical Lesion

The effect on search scan paths of applying lesions to LIP and the orbitofrontal bias (novelty map) is examined in this chapter. The lesioned model is able to replicate the type of visual search behaviour found in some parietal and orbitofrontal patients.

Neurological patients with unilateral lesions to posterior parietal cortex exhibit a visual deficit (although other sensory modalities may also be affected) known as "unilateral neglect", "hemineglect", "hemispatial neglect" or simply "neglect". The main symptom of neglect is the patient's inability to notice objects and events in the hemifield contralateral to their lesion, i.e. the contralesional hemifield. During visual search, neglect patients' fixations tend to be concentrated in the ipsilesional side (e.g. Husain et al., 2001; Mannan, Mort, Hodgson, Kennard, Driver & Husain, 2005). An example scan path from a patient (G.K.) with infarcation of the right inferior parietal lobe (but intact frontal lobe) is shown in figure 7.1. Husain et al. (2001) report that, although G.K. was clearly able to make leftward saccades, he consistently neglected targets on the left of the search array. Also, he frequently re-fixated targets and his re-fixation rate in this study was 13 times higher than that of age-matched normal controls. In a task where he had to click when he fixated a new target, his re-click rate was 34 times higher than that of normals. This indicated a failure to maintain a memory of previously fixated locations across intervening saccades and this problem worsened with increasing numbers of targets in the display.

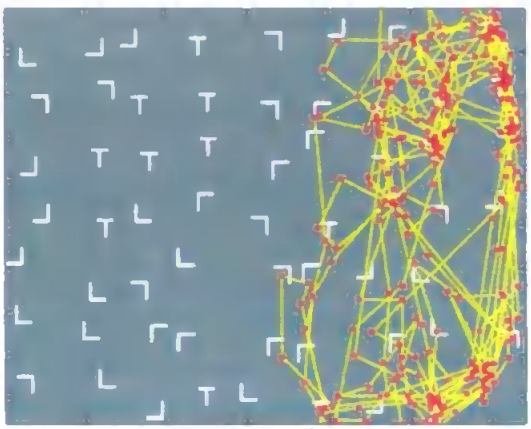


**Figure 7.1 Example scan path from a neglect patient**

Example scan path from patient G.K. who has infarcation of the right inferior parietal lobe (Husain et al., 2001). The search task was to find letter Ts amongst distractor Ls. There is profound neglect of the left side of the array and many re-fixations on the right. Figure from Husain et al. (2001) 'Impaired spatial working memory across saccades contributes to abnormal search in parietal neglect' Brain, 124, 941-952, by permission of Oxford University Press.

Neglect appears to be an attentional, rather than purely sensory, problem because it is context dependent. For example, the phenomenon of "extinction" prevents perception of an item presented contralaterally to the lesion but only when another item is simultaneously presented ipsilaterally. Typically, damage to right parietal cortex is associated with more severe symptoms of neglect than damage to left parietal cortex (Behrmann, Ebert & Black, 2004) and leads to impaired perception of the left visual field. This situation is simulated here.

Some patients exhibit neglect to the contralesional side of objects, indicating some form of object-based attentional deficit. Following right hemispheric lesion, these patients neglect the left side of objects even when it is presented in their right hemifield. However, contralesional saccades are possible to view left-half-only objects presented alone in the left hemifield and this suggests that the problem is not simply an eye movement deficit (Walker, Findlay, Young, Lincoln, 1996; Walker & Young, 1996). Therefore, neglect appears to consist of a perceptual, as well as saccadic, component and its frame of reference is related to context. Thus, it may be 'within-object' based or 'between-object' spatially based (Humphreys & Riddoch, 1994). In fact, recent evidence (Baylis, Baylis & Gore, 2004) suggests that whether the reference frame is scene-based or object-based depends on the requirements of the task, such that the internal (possibly parietal) representation is changed so that neglect applies to the contralesional side of the represented "space". The simulations presented in this chapter address only the tendency for patients' fixations to land in the ipsilesional side of an array during feature conjunction search, i.e. scene-based neglect, and do not address object-based neglect, which is likely to be less relevant within such a task.

The role of a novelty-related bias and associated memory of searched locations across saccades was described in previous chapters in relation to providing inhibition of return in the model's scan paths. It was suggested that this bias originated in orbitofrontal cortex and was applied to the LIP module. In this chapter, in addition to the impaired scanning symptoms of the neglect syndrome, the related poor visuo-spatial memory of locations visited is also examined. Husain et al. (2001) established that not only were neglect

patients' fixations concentrated in one half of the display, but they appeared to be unable to remember locations that had been previously visited during search, so that repeated re-visiting of locations was apparent in the scan paths. When this was tested under a paradigm that required patients to click a key if they thought that they were fixating a new object, patients appeared not to realise that many of their fixations were at previously visited locations (Mannan et al., 2005). Forgetting of previously visited locations appeared to worsen over time, i.e. number of fixations. In the model here, the inhibited locations in the novelty map bias applied to LIP recover linearly over time and this makes it more likely for previously visited locations to be re-visited over the course of several fixations. Such a bias, combined with a parietal lesion, which limits the spatial area available for search, could lead to increased re-visiting of ipsilesional locations. Hodgson et al. (2002) reported that orbitofrontal damage also affected scan paths and, whilst not leading to hemispatial neglect, led to a disorganised search with a high rate of re-fixation. The effect of lesions to LIP and the orbitofrontal bias on rates of re-fixation in the scan path is examined in this chapter.

It is recognised that the lesions here are a simplification of the parietal damage possibly involved in the neglect syndrome. Other, more posterior, areas of parietal cortex may be involved, for example the inferior parietal lobule, which is linked to reflexive saccades (Mort, Perry, Mannan, Hodgson, Anderson, Quest, McRobbie, McBride, Husain & Kennard, 2003). More detailed modelling of parietal cortex is beyond the scope of the current work. This chapter is not intended to be a comprehensive computational investigation of behavioural phenomena relating to the neglect syndrome and re-fixation

but simply provides some initial results that indicate the capacity of the model to replicate behaviour following lesion. As such, the results in this chapter provide further support for the validity of the model and highlight an interesting area for future development.

Data presented in appendix A4.1 from individual scan path simulations, containing 50 and 100 fixations, are discussed here. Figures from single simulations are not necessarily statistical significant. However, presenting this data allows an individual scan path to be shown along with data that relate to it, such as its re-fixation rate. The effects reported here were tested by running multiple simulations (10 scan paths containing 50 fixations each), in the same manner as the simulations reported in chapter 6, and these data are presented in appendix A4.2. The multiple simulations produced similar results to the individual ones and the re-fixation rates from the multiple scan paths are compared to those from the individual scan paths in the appropriate places.

## 7.1 Unilateral LIP Lesion

In order to simulate a unilateral lesion to right posterior parietal cortex, the left half of the LIP module was "lesioned" by setting the activity of cell assemblies in this area to zero. This simulates a lesion to the right hemisphere because the model contains no crossover of visual inputs in the optic chiasm. Therefore, the terms contra- and ipsilesional used here refer to a patients' right hemisphere lesion, as

modelled by a lesion to the left half of the LIP module (or the left half of the novelty map in the simulations described in section 7.2).

Scan path simulations in previous chapters used an active vision approach where attention was moved overtly as the retina moved around the scene. However, for the lesion simulations, the model uses covert attention where the retina is held static and encompasses the entire image. The movement of the AW represents the shift of covert attention and the AW is scaled according to stimulus density in the normal manner. The reason why covert attention is used for these simulations is due to a computational constraint imposed by the fact that the original image is not continuous in the way that the natural world is. When the retina is moved to a location near the image edge, the model's cortical coordinates extend beyond the image and this is dealt with as described in appendix A1.2.3. This results in locations in LIP beyond the extent of the original image receiving no bottom-up stimulus information and, therefore, these locations do not become highly active or attract attention. This produces the correct effect for normal scan path simulations reported previously. However, where the left half of LIP is lesioned and fixation is near the right edge of the original image, the only remaining effective side of LIP (the right half) may extend beyond the original image. This results in LIP being unable to operate effectively and the scan path tending to be unable to move away from the edge area of the image. Therefore, to avoid this computational artefact for these simulations, the system operates by keeping its retina in a fixated position whilst still moving the AW. This means the system operates in a similar manner to

other models that have used a static retina to reproduce attentional scanning behaviour (e.g. Deco & Lee, 2001; Niebur, Itti, Koch, 2001).

Figure 7.2 shows a normal covert scan path through an image, which will be used throughout most of this chapter, where the target is a red vertical bar. Most scan paths in this chapter are allowed to continue for 100 "fixations" where "fixation" under conditions of covert attention refers to the position of the centre of the AW. For this scan path simulation, there were no lesions and the weight of the novelty bias was set to 0.0009, the normal setting used in previous chapters. The novelty map after this scan path is shown in figure 7.3. These figures serve as a comparison to those produced under lesion conditions later.

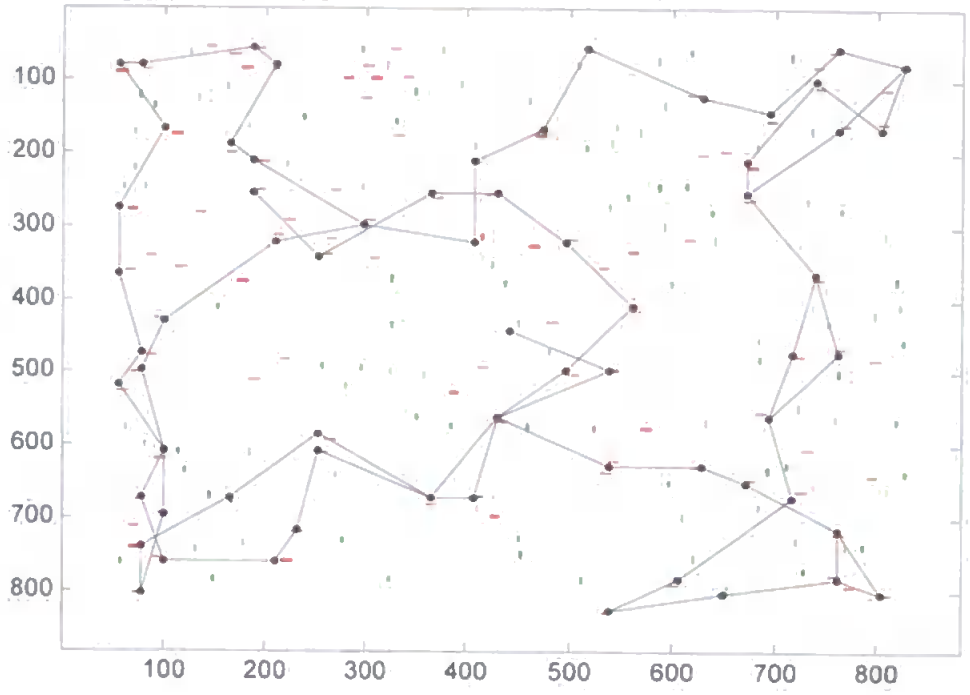



**Figure 7.2 Control condition scan path**

Control condition – covert scan path of 100 "fixations" through image '881rv100'

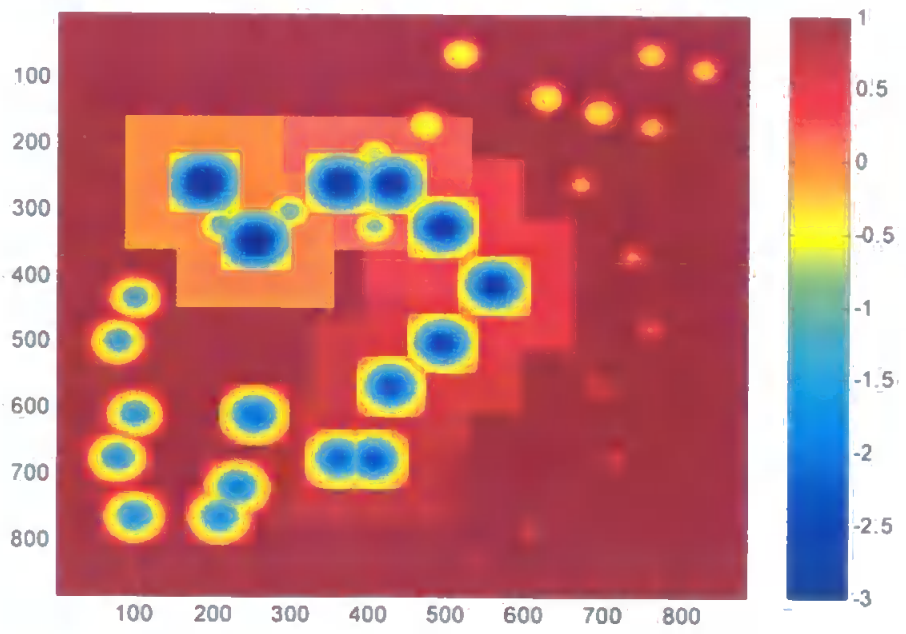



**Figure 7.3 Control condition novelty map**

Control condition - novelty map after the scan path shown in figure 7.2.

In contrast to figure 7.2, figure 7.4 shows a covert scan path through the same image but following a unilateral lesion to the LIP module. The novelty map after this scan path is shown in figure 7.5. Clearly, the LIP lesion has produced scan path behaviour characteristic of neglect, for example that reported by Husain et al. (2001, *figure 2A*; shown in figure 7.1 here) and Mannan et al. (2005, *figure 1*). The focus of attention remains in the right hemifield despite the presence of high novelty in the left hemifield. Target-coloured stimuli in the portion of the scene corresponding to ipsilesional space are thoroughly examined.

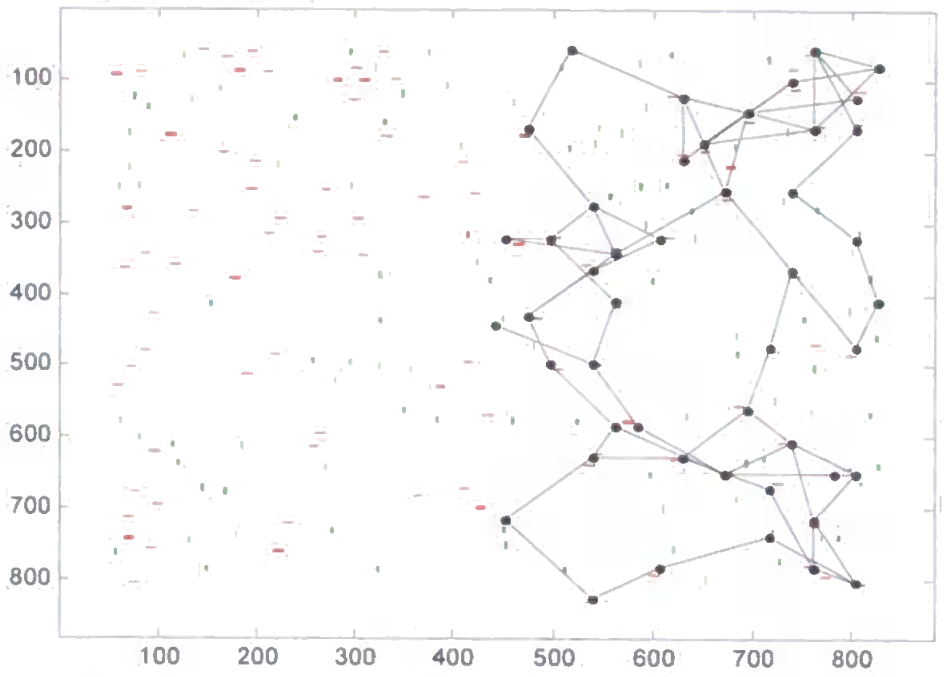



**Figure 7.4 Unilateral LIP lesion scan path**

Covert scan path of 100 "fixations" through image 881rv100 following unilateral LIP lesion

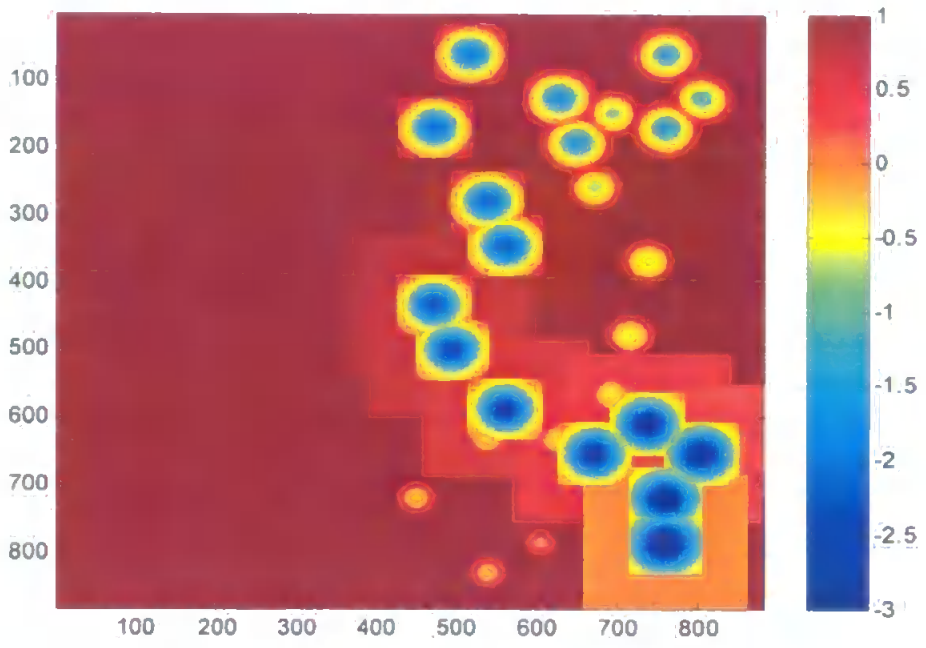



**Figure 7.5 Novelty map after unilateral LIP lesion scan path**

Novelty map after the scan path shown in figure 7.4.

The reason why the scan path is concentrated in one hemifield can be seen by examining the cortical activity. Figure 7.6 shows the activity in IT, LIP and the V4 feature maps after 80ms, 180ms and 225ms of processing at the start of the scan path shown in figure 7.4. At this point attention is allocated to the centre of the image and the AW forms around this location. Note that the AW and stimulus information are not represented in the lesioned half of LIP. This means that attention will be attracted next to a location represented by the unlesioned portion of LIP.

The LIP lesion also affects the level of activity in the ipsilesional side of V4. Due to the representation of the AW in LIP being affected by its lesion, the ipsilesional half of the AW does not produce a spatial attentional enhancement in V4 and only stimuli in the contralesional AW in V4, i.e. those that are not neglected, are enhanced. This is a very interesting effect in the model because it suggests that a spatial deficit may be apparent in V4 cellular activity following posterior parietal lesion, dependant on the spatial position of attention. Contrary to earlier theories that the ventral stream was the neural correlate of conscious perception (Milner & Goodale, 1995), recent imaging and ERP studies have found activity in striate and ventral extrastriate areas relating to extinct stimuli in neglect patients, i.e. those contralesional stimuli that are not consciously perceived (Driver, Vuilleumier, Eimer & Rees, 2001; Rees, Wojciulik, Clarke, Husain, Frith & Driver, 2000, 2002; Vuilleumier, Sagiv, Hazeltine, Poldrack, Swick, Rafal & Gabrieli, 2001).

However, this activity appears to be reduced in extinction compared to when the stimulus is presented in the contralesional hemifield and is not extinguished but is consciously perceived (Driver et al., 2001; Rees et al., 2002; Vuilleumier et al., 2001). Hence, representation of objects that are consciously perceived may require greater ventral stream activity than that for objects not consciously perceived. The decrease in activity in V4 observed in the simulations here for the neglected hemifield may relate to reductions observed in human extrastriate ventral cortex when a stimulus is not consciously perceived. Thus, the difference in V4 activity levels shown, in the two halves of the AW for example, in figure 7.6 may reflect the difference in conscious verses unconscious perception. It is extremely interesting that the model, in particular the cross-stream connection, is able to produce an effect that correlates with human data about conscious perception.

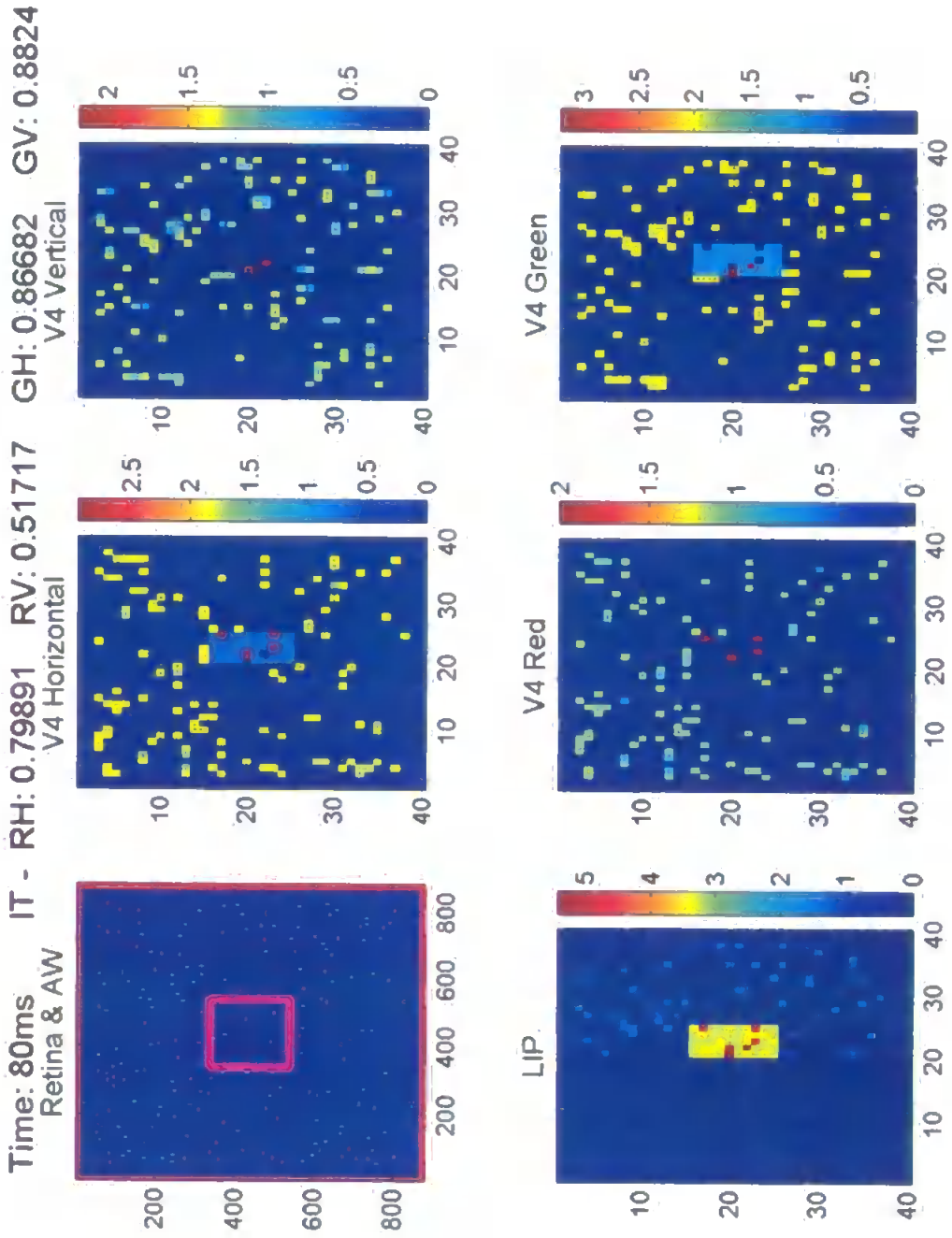



**Figure 7.6a Cortical activity following LIP lesion – 80ms**

Activity in the cortical areas at 80ms post-stimulus (soon after the initial sensory response). The retina extends across the whole image and the AW is shown as the magenta outlined box in the top left sub-plot. Stimuli in V4 falling within the half of the AW affected by the LIP lesion are not subject to spatial attentional enhancement in LIP or V4.

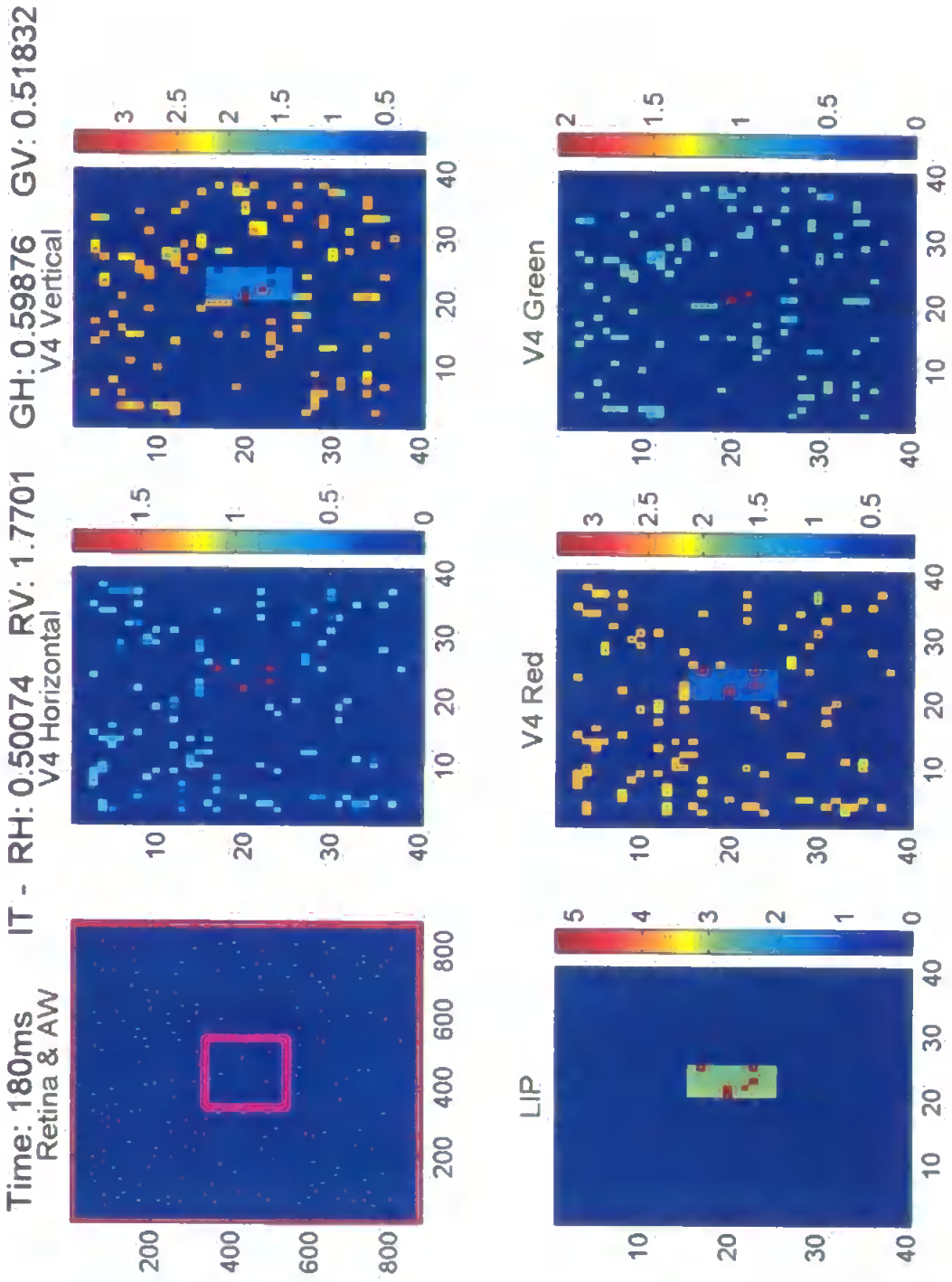



**Figure 7.6b Cortical activity following LIP lesion – 180ms**

Activity in the cortical areas at 180ms post-stimulus, i.e. after the onset of object-based attention.

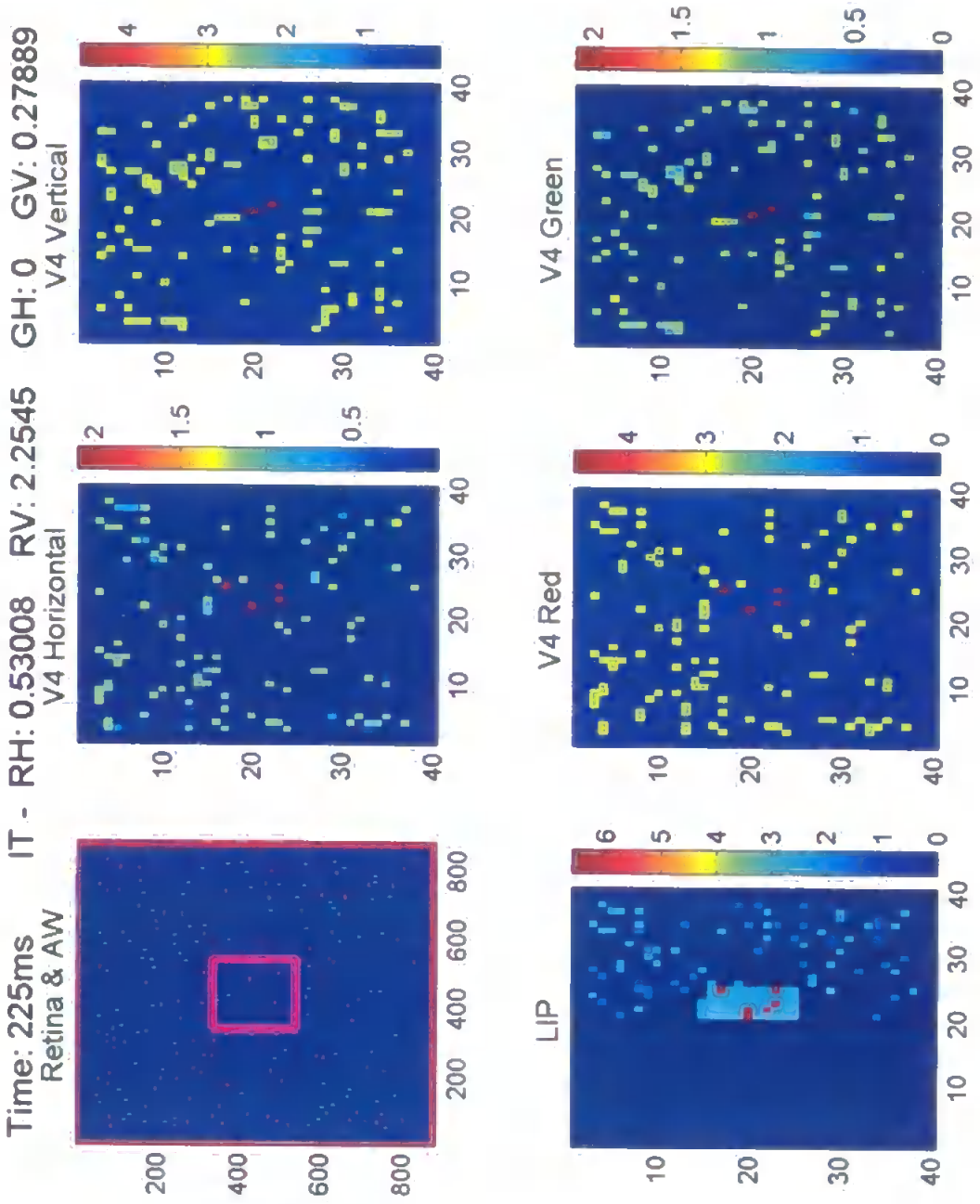



**Figure 7.6c Cortical activity following LIP lesion – 225ms**

Activity in the cortical areas at 225ms post-stimulus. Stimuli falling within the half of the AW affected by the LIP lesion are much more weakly represented in V4 than those on the ipsilesional side. Such differences in representation could relate to conscious awareness.

Figure 7.7 shows the same situation as figure 7.6 (i.e. the lesioned hemisphere in LIP is inactive) but for a point later in the scan path where attention has been allocated to a location in the top right quadrant of the image and retina. This figure shows the effects of covert attention in the model, as opposed to the overt attention simulations used in previous chapters, because the AW is not central to the retina. The AW is positioned fully in the non-neglected portion of the scene and produces a spatial enhancement of responses in LIP and V4. All stimuli in the AW in V4 are enhanced and may, therefore, enter conscious awareness.

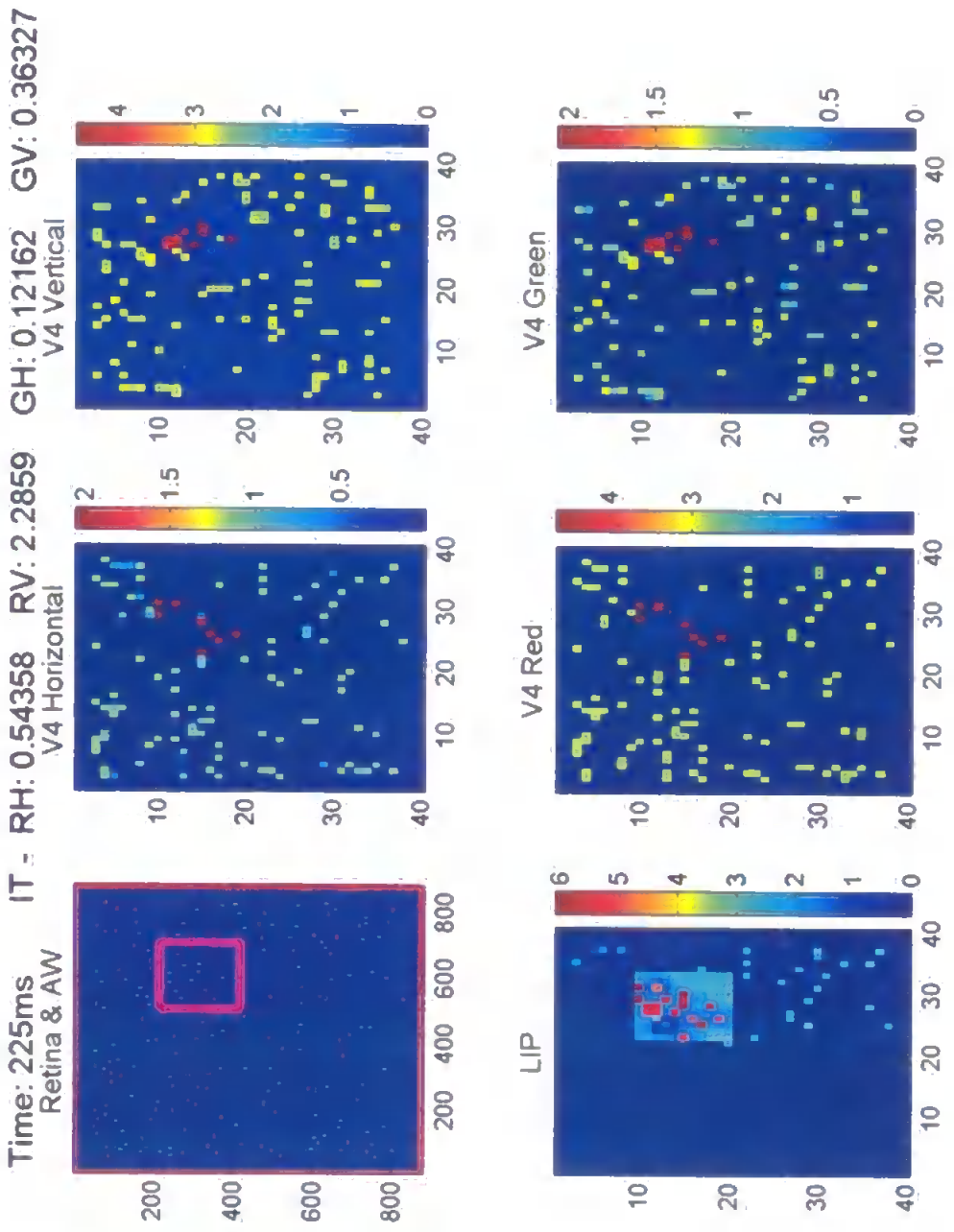



**Figure 7.7 Cortical activity following LIP lesion with attention to an offset location**
Activity 225ms after attention was moved to a location in the top right quadrant of the image during the scan path shown in figure 7.4. The retina extends across the whole image and the AW is shown as the magenta outlined box in the top left sub-plot, being offset from the centre of the retina because the shift of attention was covert. In this case, the AW falls within the unlesioned portion of LIP and spatial attentional modulations are present in LIP and V4.

Parietal lesion patients tend to have a higher than normal re-fixation rate in their scan paths (Husain et al., 2001; Mannan et al., 2005). Lesions to the model's LIP also produce a small increase in re-fixation rate. Figure 7.8 shows a comparison of the re-fixation rate under conditions of unilateral LIP lesion (figure 7.4) compared to the control condition (figure 7.2). The re-fixation rate for the first 50 "fixations" in the scan path is shown in addition to that for the entire scan path. Unsurprisingly, there is a tendency for the re-fixation rate to be higher for longer scan paths in both the control and lesion conditions. The increase in re-fixations under simulated lesion is lower than that reported by Husain et al. (2001), where the re-fixation rate was 13 times higher than that of normals for dense scenes. The increase in re-fixation simulated here is less than that produced by full orbitofrontal lesions, as the next sections will describe.

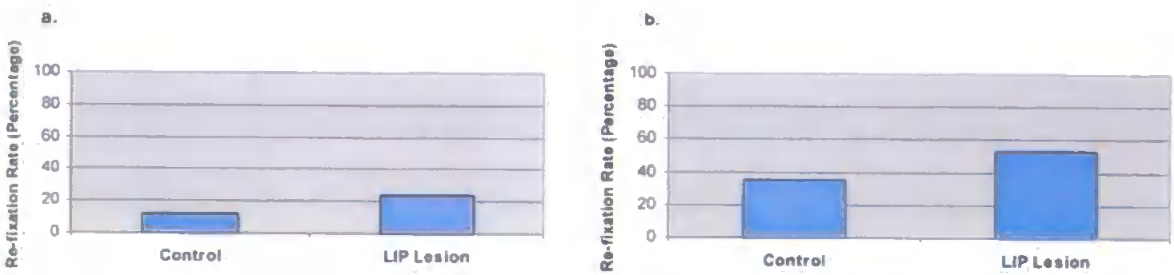



**Figure 7.8 Re-fixation rate following unilateral LIP lesion, compared to control**

a. Re-fixation rates during the first 50 "fixations" of the scan paths shown in figures 7.2 (control) and 7.4 (LIP lesion). Multiple scan path simulations produced average re-fixation rates very similar to these in both the control and lesion condition (see appendix A4.2).

b. Re-fixation rates during the full 100 "fixations" of the scan paths shown in figures 7.2 (control) and 7.4 (LIP lesion).

## 7.2 Orbitofrontal Lesions

There is some evidence to suggest that removal of orbitofrontal cortex is associated with abnormal patterns of eye movements during visual search, a deficit in reward-related IOR and an increased rate of re-fixating previously fixated search locations (Hodgson et al., 2002). Figure 7.9 shows the increase in re-fixation found in the patient study reported by Hodgson et al. (2002). Note that the re-fixation rate is calculated differently from that shown for simulations here, which calculate the percentage of re-fixations as a function of all fixations in the scan path, rather than as a function of the number of stimuli because simulated scan paths were of fixed length. There is no attempt at quantitative replication but the qualitative similarity of the results is obvious.

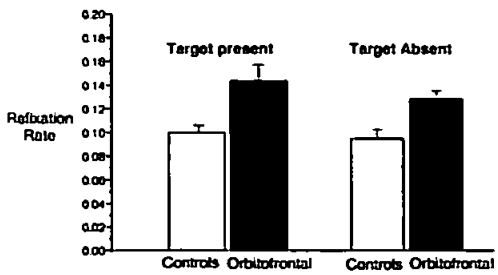



**Figure 7.9 Re-fixation rates for an orbitofrontal patient**

Re-fixation rates for an orbitofrontal patient and controls during a visual search task reported by Hodgson et al. (2002). The patient's mean re-fixation rate of 13.5% is higher than that of normals. The re-fixation rate was calculated by dividing the number of fixations on previously fixated locations by the number of elements in the display and the number of trials in the block. Reprinted from Neuropsychologia, 40(12), Hodgson, T.L. et al., Orbitofrontal cortex mediates inhibition of return, 1891-1901, Copyright (2002), with permission from Elsevier.

To simulate a lesion to orbitofrontal cortex, or its connection with posterior parietal cortex, the values in the novelty map, or the weight of the novelty map bias to LIP, were adjusted to represent three conditions, as follows:

- For a unilateral lesion, the right side of the novelty map was updated as normal but the left side was set to zero (neutral novelty) to simulate a right hemisphere orbitofrontal lesion.
- For a complete lesion, the whole novelty map was set to zero, i.e. neutral novelty, throughout the simulation so that it was ineffective at biasing competition in LIP.
- To simulate conditions where all orbitofrontal output to LIP is weakened, the weight of the novelty bias to LIP was reduced.

It should be noted that the source of the novelty bias, and IOR, here could be a structure other than orbitofrontal cortex. IOR is typically associated with eye movement structures such as superior colliculus (Clohessy et al., 1991; Sapir et al., 1999; Trappenberg, Dorris, Munoz & Klein, 2001) and, therefore, the neural correlate of the simulated lesion is slightly uncertain.

Most simulations presented here use covert attention in order to compare and combine the results of lesions to the novelty map with the LIP lesion simulations presented above. However, the novelty/orbitofrontal lesion simulations do not require covert attention, and the effects of such lesions on the overt attention

simulations presented in previous chapters are discussed towards the end of this section.

### 7.2.1 Unilateral Orbitofrontal Lesion

A lesion to the left hemifield of the novelty map (corresponding to a right hemisphere orbitofrontal lesion) results in the covert scan path tending to be attracted to the right half (ipsilesional hemifield) of the image. A scan path with 50 "fixations" is shown in figure 7.10 and the corresponding novelty map at the end of this scan path is shown in figure 7.11. However, when this scan path was allowed to continue for a further 50 "fixations", attention was drawn into the left half of the image, as shown in figure 7.12. Once there, due to the AW being quite small, attention was not able to move from its leftmost location because the left side of the novelty map contains neutral (zero) novelty and there was no effective novelty bias within the current AW to attract attention to another location. Therefore, there was a difficulty re-orienting attention in contralesional space. The novelty map at the end of the 100 "fixations" is shown in figure 7.13. As attention was unable to re-orient from its leftmost position, the values in the right-hand side of the novelty map have recovered over time to have high novelty again. However, the AW, which results in enhanced activity in LIP, is so far to the left of the image that it does not include these high novelty areas and attention is unable to be captured by them.

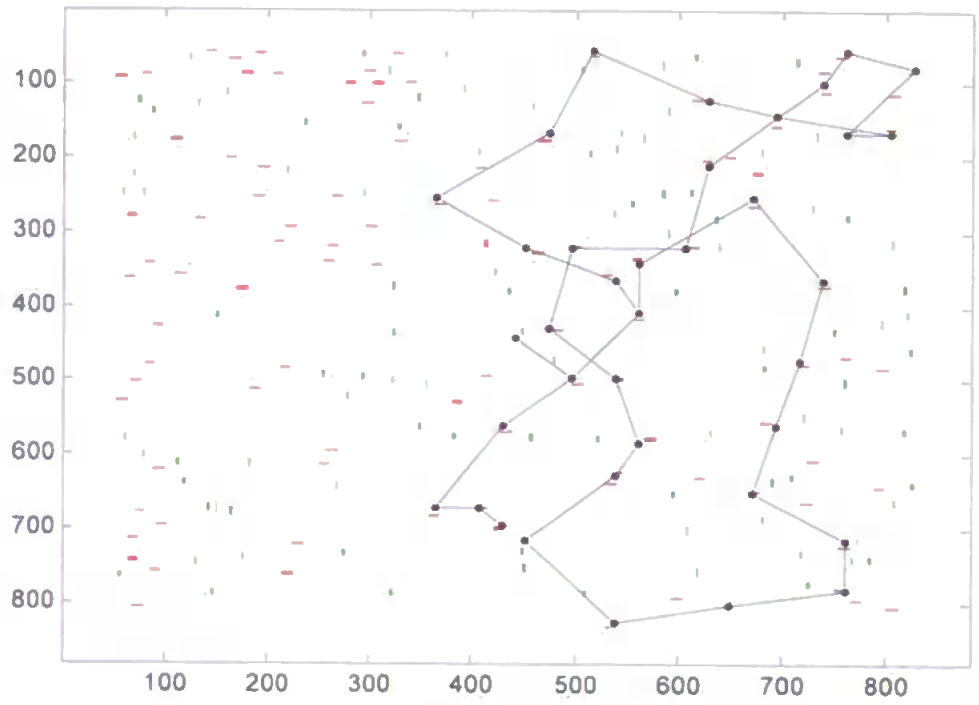



**Figure 7.10 Scan path of 50 "fixations" following a unilateral lesion to the novelty map**

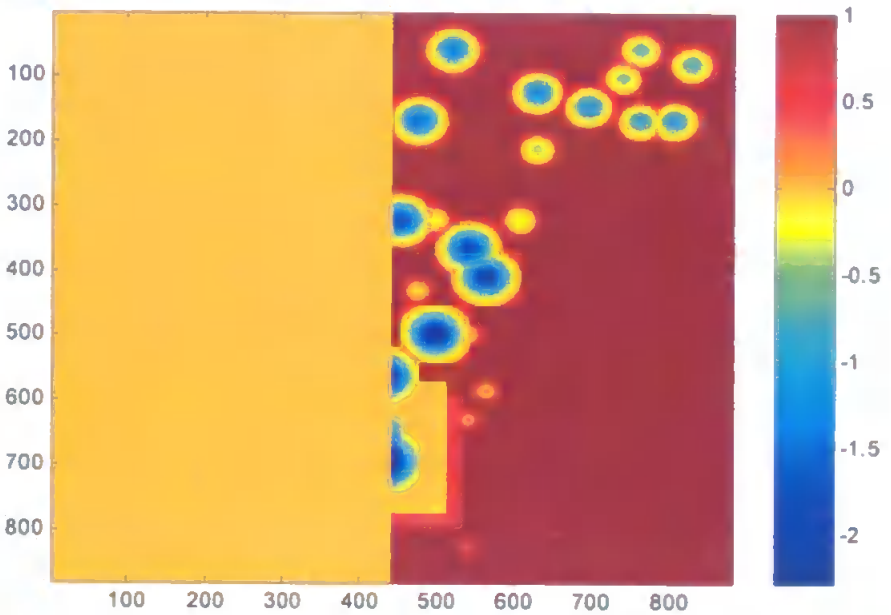



**Figure 7.11 Unilaterally lesioned novelty map at the end of the scan path shown in figure 7.10**

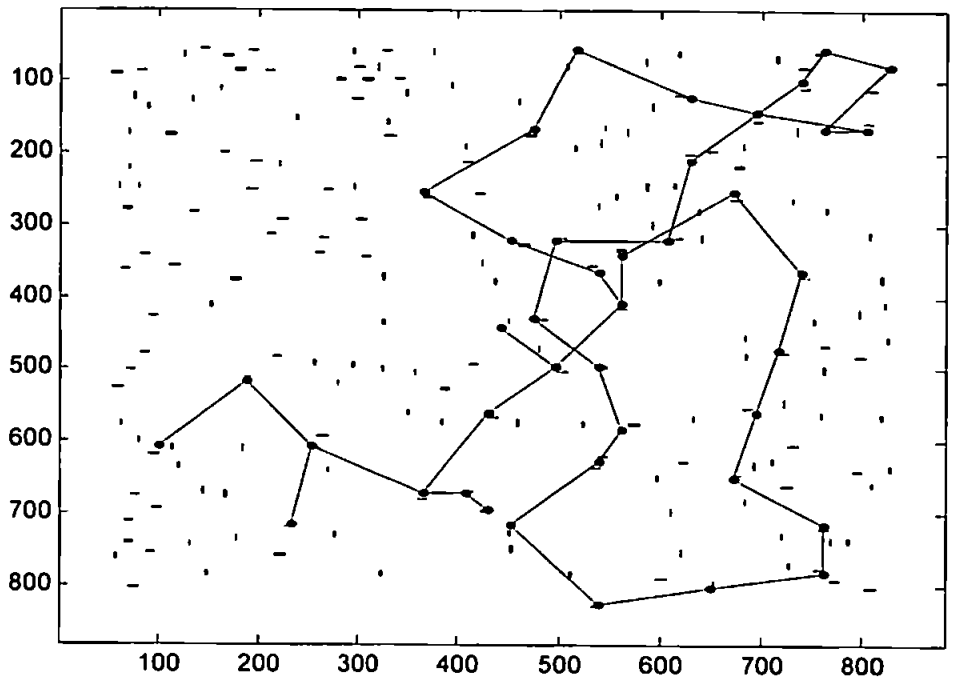



**Figure 7.12 Scan path of 100 "fixations" following a unilateral lesion to the novelty map**

The first 50 "fixations" were shown in figure 7.10. Here the scan path enters the contralesional hemifield but has more difficulty re-orienting attention in this area.

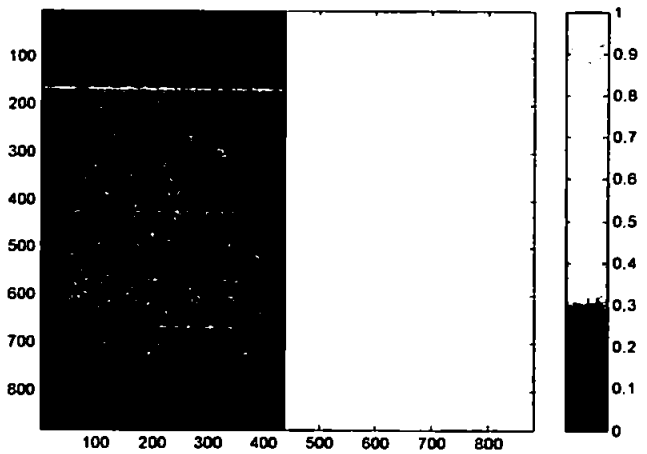



**Figure 7.13 Unilaterally lesioned novelty map at the end of the scan path shown in figure 7.12**

The scan path behaviour following a unilateral lesion to the novelty map (which might correspond to a unilateral lesion to orbitofrontal cortex) was similar to that produced by a unilateral lesion LIP, where attention tended to be attracted to locations in the ipsilesional hemifield. Whereas the LIP lesion produced scan paths that were confined to the ipsilesional hemifield, scan paths were able to move into the contralesional hemifield following a unilateral orbitofrontal lesion. However, there were problems re-orienting attention in the contralesional hemifield.

Figure 7.14 shows the re-fixation rates for the scan paths shown in figures 7.10 and 7.12, compared to the control situation given in section 7.1 above. Taking into account the rate being inflated, in the 100 fixation case, by the inability to re-orient attention in the contralesional hemifield, re-fixation rates are similar to those following the LIP lesion described above. However, the average re-fixation rate presented in appendix A4.2 for 10 scan paths (each consisting of 50 fixations) was higher (42.4% compared to 26%) than that for the individual scan path in figure 7.14a. This is due to several of these scan paths entering the contralesional hemifield. Thus, re-fixation rates here are slightly higher than those following a unilateral LIP lesion described above but this difference is mainly due to an inability to re-orient attention in the contralateral hemifield. Therefore, a unilateral orbitofrontal results in an overall increase in re-fixation and tends to produce scan paths that favour the ipsilesional hemifield.

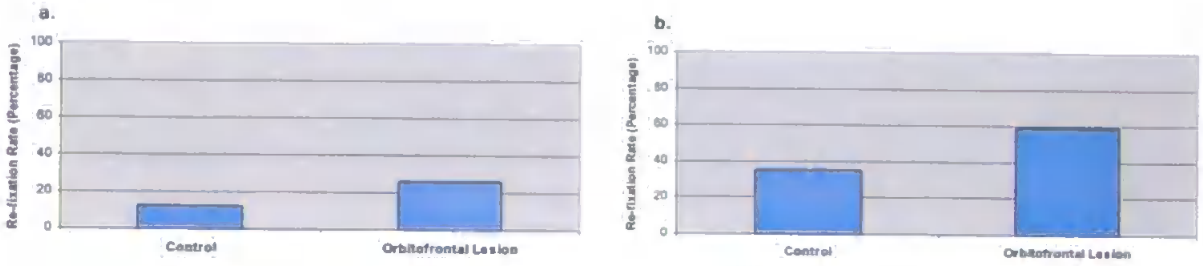



**Figure 7.14 The effect of a unilateral novelty map lesion on the re-fixation rate**

**a.** Re-fixation rates during the first 50 "fixations" of the scan path shown in figures 7.2 (control) and the 50 "fixations" of the scan path shown in figure 7.10 (lesion). Multiple scan path simulations produced an average re-fixation rate in the lesion condition that was much higher (42.4%) than that shown for this scan path (see appendix A4.2).

**b.** Re-fixation rates during the 100 "fixations" of the scan paths shown in figures 7.2 (control) and 7.12 (lesion).

### 7.2.2 Full Orbitofrontal Lesion

In order to simulate a total lesion to the orbitofrontal feedback to LIP, the novelty map was set to zero, which indicates neutral novelty. This means that novelty does not influence the competition for attention in LIP, and IOR within the scan path is limited. A sample scan path, consisting of 100 "fixations" is shown in figure 7.15.

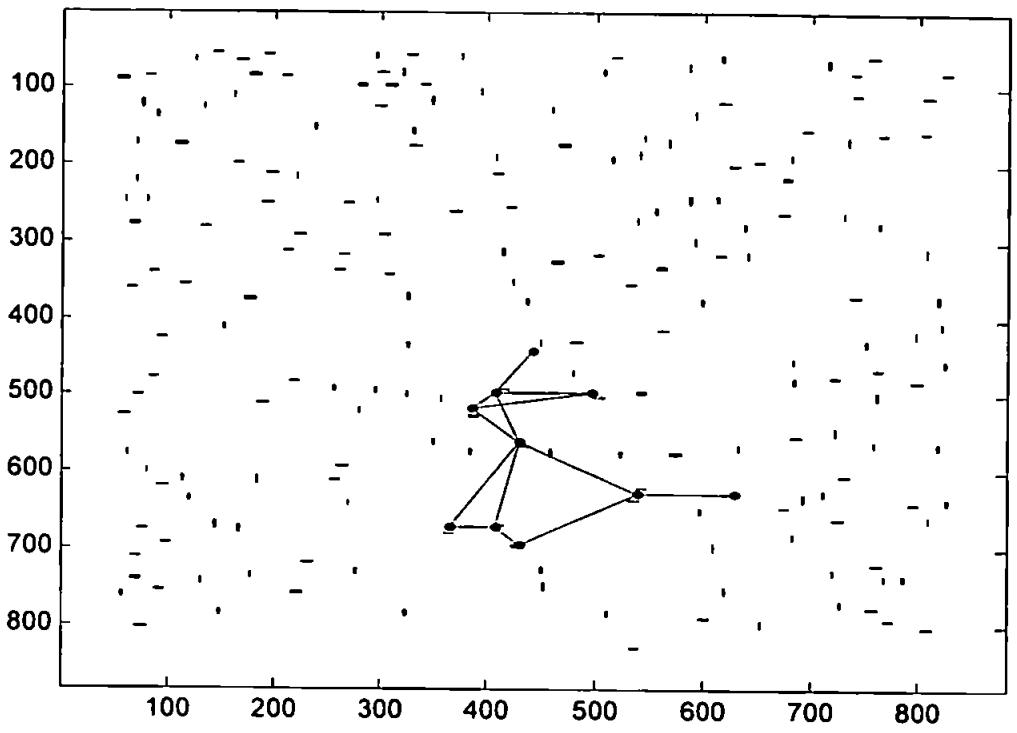



**Figure 7.15 Scan path of 100 "fixations" following full lesion to orbitofrontal cortex**
The novelty map is set entirely to zero (neutral novelty) so that novelty cannot affect the competition in LIP.

Figure 7.15 shows that the lack of the orbitofrontal novelty feedback prevents the scan path widely exploring the scene. Re-fixations are common. Figure 7.16 shows a comparison of re-fixation rates for this type of lesion compared with the control. A complete lesion to the model's orbitofrontal novelty input leads to increased re-fixation in the scan path and this is greater than the increase observed for unilateral lesions. Appendix A4.2 shows average re-fixation rates over 10 scan paths when the orbitofrontal bias is completely removed (as in figure 7.15) and, also, when it is reduced in strength. In accordance with the results in chapter 6, re-

fixation rates increase as the weight of orbitofrontal bias decreases. The model exhibits a higher level of re-fixation and a much more confined scan path than that reported in an orbitofrontal patient study (Hodgson et al., 2002; see figure 7.9 above). This could be due to the model's lack of other prefrontal areas that could provide executive functions such as a search strategy. If the re-fixation rate is calculated in the same way as that given by Hodgson et al. (2002), where the number of re-fixations is divided by the number of stimuli in the scene, the 50 fixation scan path given in figure 7.16a and appendix A4.1 has the following rate:

- (41 re-fixations / 201 stimuli) = 20%

This is only slightly higher than the 13.5% reported by Hodgson et al. (2002).

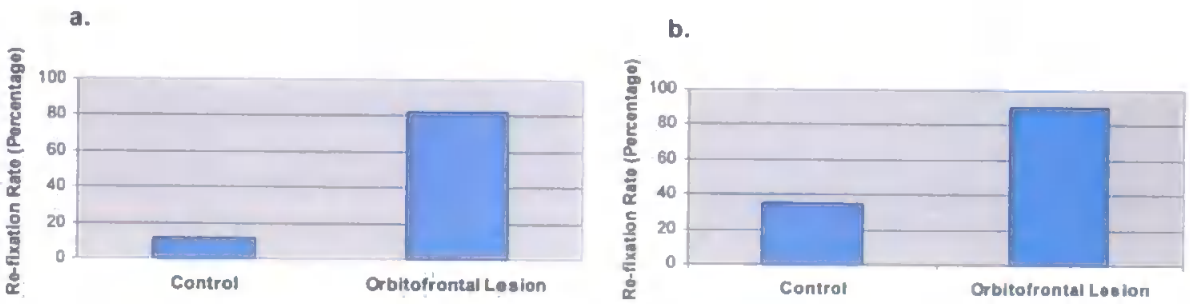



**Figure 7.16 Re-fixation rates following a total orbitofrontal (novelty map) lesion, compared to control**

**a.** Re-fixation rates during the first 50 "fixations" of the scan paths shown in figures 7.2 (control) and 7.15 (lesion). Multiple scan path simulations produced re-fixation rates similar to these in both the control and lesion condition (see appendix A4.2).

**b.** Re-fixation rates during the full 100 "fixations" of the scan paths shown in figures 7.2 (control) and 7.15 (lesion).

### 7.2.3 Full Orbitofrontal Lesion Combined With Unilateral LIP Lesion

A unilateral LIP lesion, as described in section 7.1 above, was combined with a full orbitofrontal lesion, as described in section 7.2.2 above, in order to investigate the effect of a lack of novelty input in addition to the symptoms of neglect observed at 7.1 above. Figure 7.17 shows a scan path under these conditions. This figure can be compared to figure 7.4, which was produced under conditions of unilateral LIP lesion but with normal orbitofrontal feedback. Like that in figure 7.4, the contralateral hemifield is neglected in figure 7.17, due to the LIP lesion. However, in addition to this effect, the scan path is less able to widely investigate the ipsilesional hemifield and is confined to a smaller area of the scene. This is due to weak IOR in the scan path resulting from the lack of a novelty bias. Thus, the combined effect of these lesions is to neglect the contralesional hemifield and to make the re-orienting of attention within the ipsilesional hemifield more difficult. Re-fixation rates, shown in figure 7.18, are similar to those produced by a full orbitofrontal lesion without the LIP lesion.

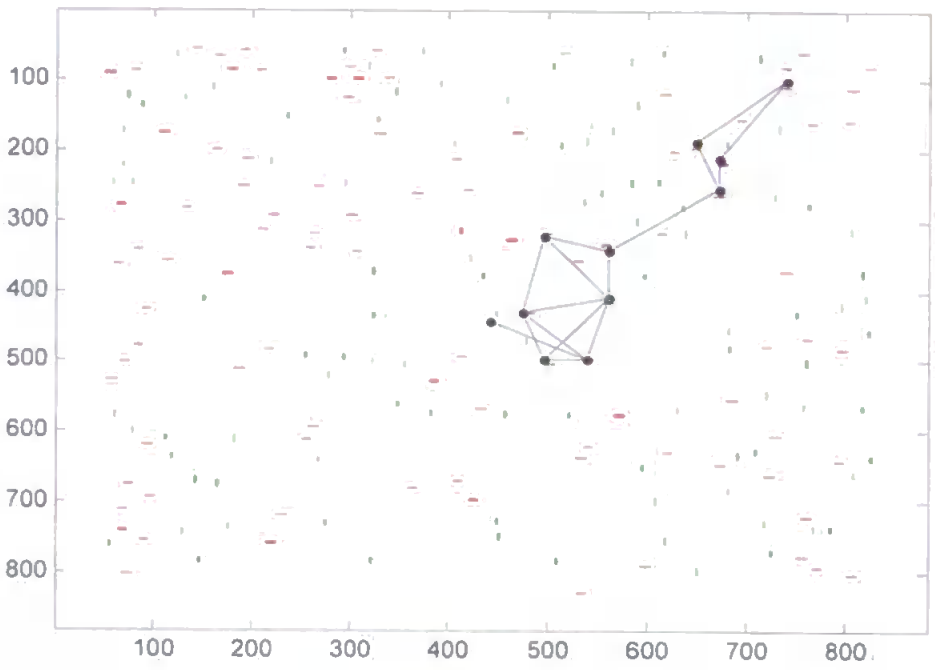



**Figure 7.17 Scan path of 100 "fixations" produced following unilateral LIP lesion and full orbitofrontal lesion**

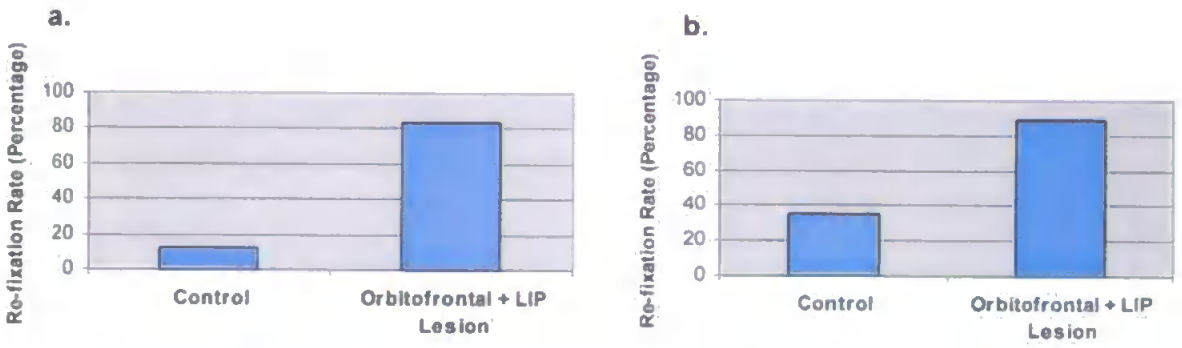



**Figure 7.18 Re-fixation rates following a total orbitofrontal (novelty map) lesion, combined with a unilateral LIP lesion, compared to control**

a. Re-fixation rates during the first 50 "fixations" of the scan paths shown in figures 7.2 (control) and 7.17 (lesion). Multiple scan path simulations produced re-fixation rates similar to these in both the control and lesion condition (see appendix A4.2).

b. Re-fixation rates during the full 100 "fixations" of the scan paths shown in figures 7.2 (control) and 7.17 (lesion).

### 7.2.4 Reduced Orbitofrontal Feedback

Next, the impact of a reduction in the weight of novelty feedback was assessed when combined with a unilateral LIP lesion. The novelty map was updated at each attention shift, as normal, but the weight of its bias to LIP was reduced. This simulates a problem in the connection from orbitofrontal cortex to posterior parietal cortex, so that this bias is less effective at driving competition in posterior parietal cortex, or a partial orbitofrontal lesion.

First, the weight of the novelty bias (parameter $\eta$ in equation 4.11 of chapter 4) was reduced from its normal value of $9e^{-4}$ to $9e^{-7}$. Figure 7.19 shows a scan path produced. Figure 7.20 shows the associated novelty map, which was updated as normal, and figure 7.21 compares re-fixation rates with the control condition. The weakened novelty input has prevented the scan path exploring the scene as widely as normal and re-fixation is more common than in the control situation, but is slightly less common than that with no novelty bias, shown in figure 7.18. The scan path is again confined to the ipsilesional hemifield, due to the LIP lesion.

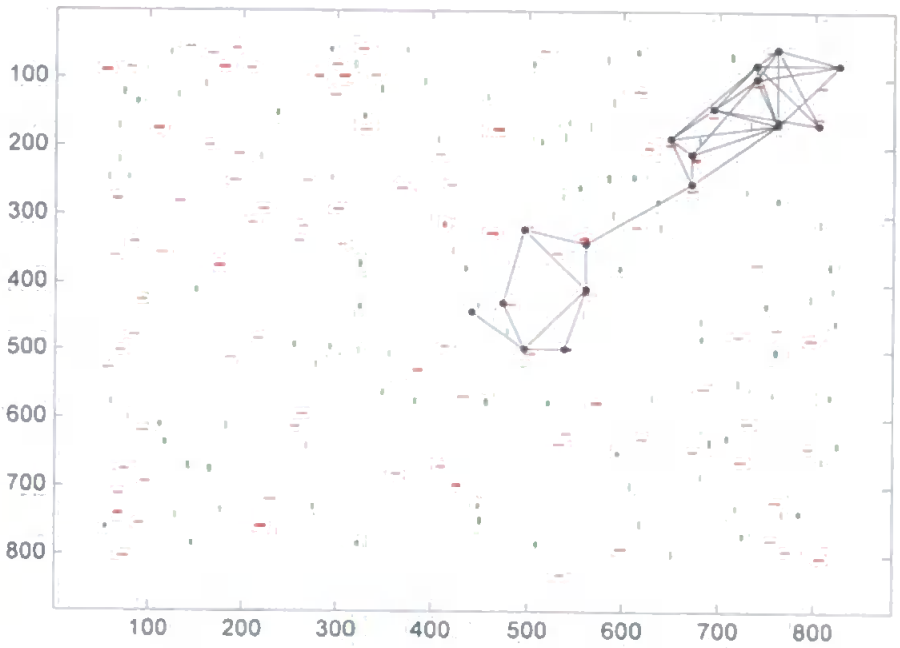



**Figure 7.19 Scan path of 100 "fixations" following unilateral LIP lesion and weight of orbitofrontal novelty bias reduced to $9e^{-7}$**

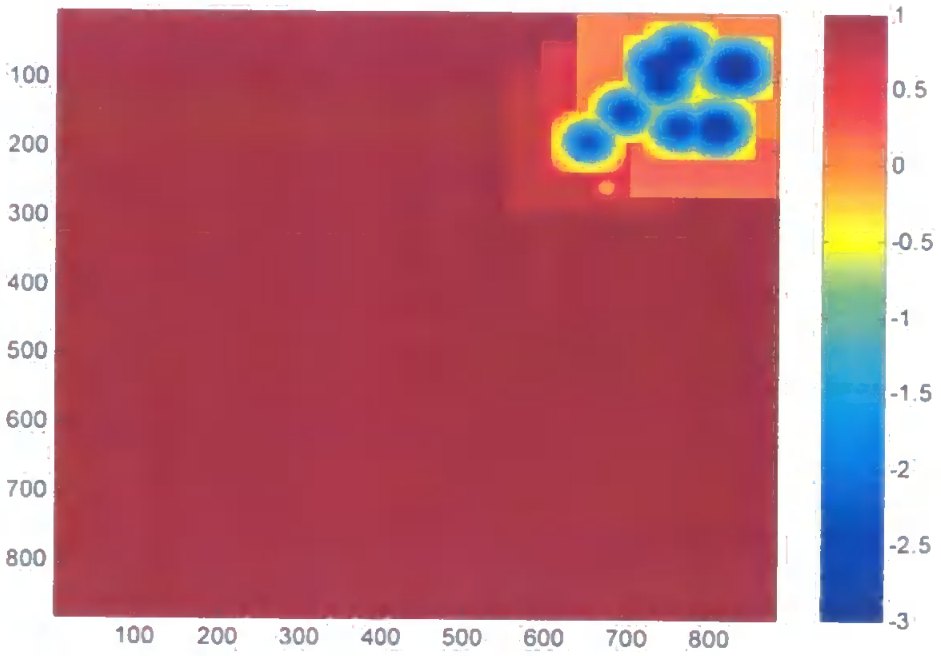



**Figure 7.20 Novelty map at the end of the scan path shown in figure 7.19**

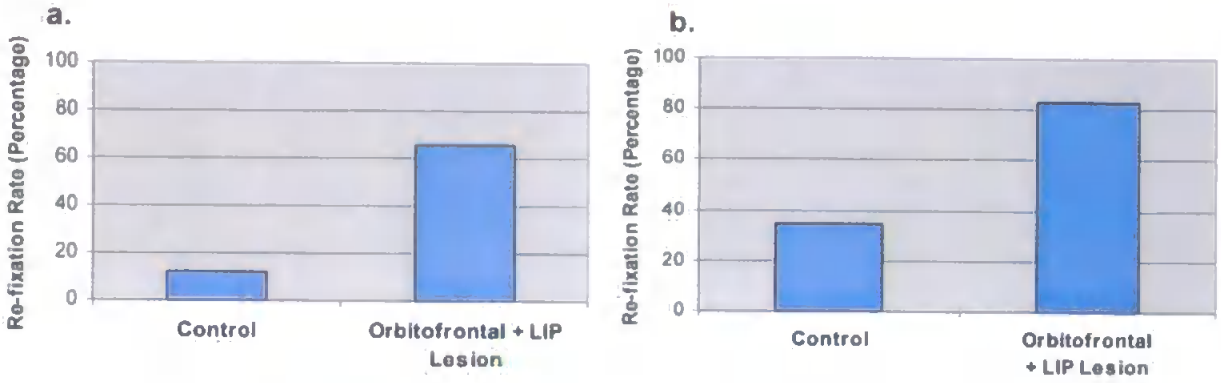



**Figure 7.21 Re-fixation rates following unilateral LIP lesion combined with the weight of orbitofrontal (novelty map) feedback being reduced to $9e^{-7}$, compared to control**

a. Re-fixation rates during the first 50 "fixations" of the scan paths shown in figures 7.2 (control) and 7.19 (lesion). Multiple scan path simulations produced re-fixation rates similar to these in both the control and lesion condition (see appendix A4.2).

b. Re-fixation rates during the full 100 "fixations" of the scan paths shown in figures 7.2 (control) and 7.19 (lesion).

When the weight of the novelty bias was increased to $9e^{-5}$, an order of magnitude weaker than the normal setting ($9e^{-4}$), the scan path produced in figure 7.22 was produced and the associated novelty map is shown in figure 7.23. A comparison of re-fixation rates, compared to the control, is given in figure 7.24.

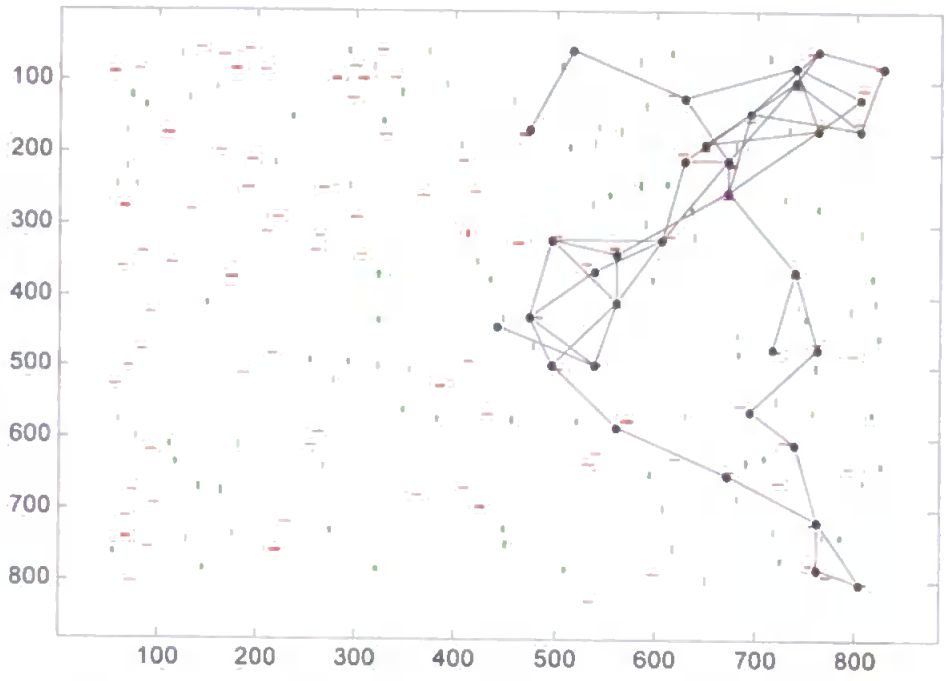



**Figure 7.22 Scan path of 100 "fixations" produced following unilateral LIP lesion and weight of orbitofrontal novelty bias reduced to $9e^{-5}$**

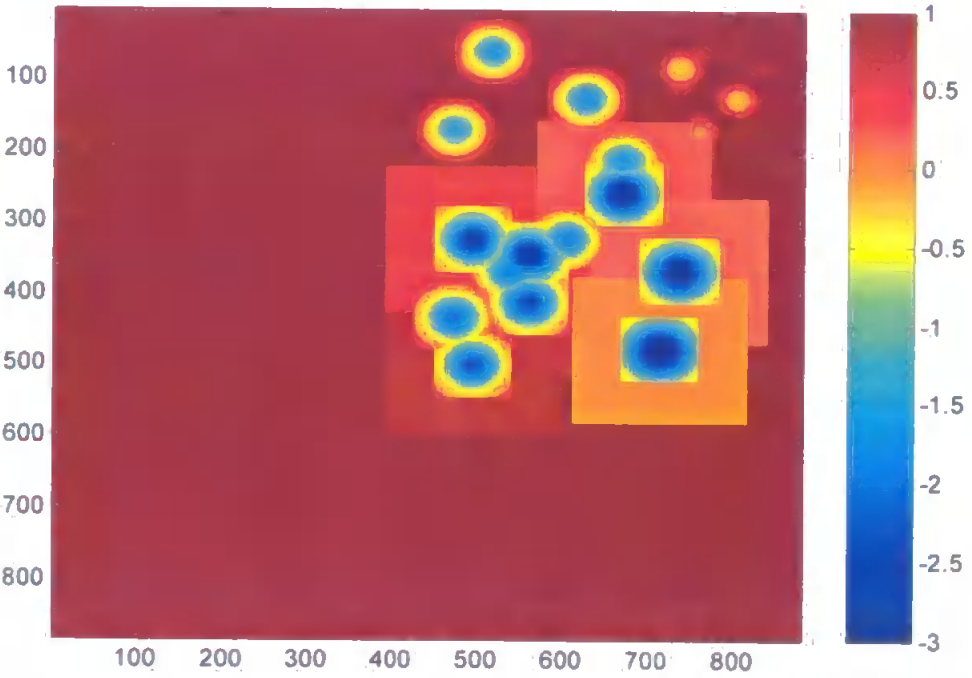



**Figure 7.23 Novelty map at the end of the scan path shown in figure 7.22**

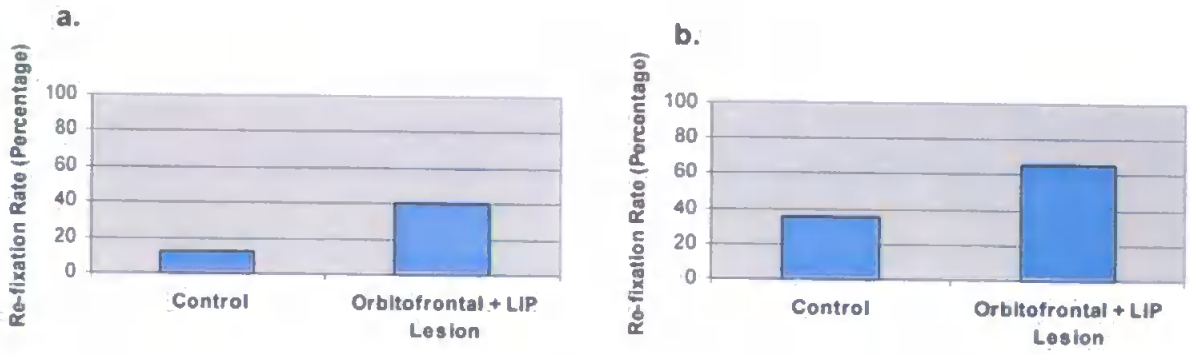



**Figure 7.24 Re-fixation rates following unilateral LIP lesion combined with the weight of orbitofrontal (novelty map) feedback being reduced to $9e^{-5}$, compared to control**

**a.** Re-fixation rates during the first 50 "fixations" of the scan paths shown in figures 7.2 (control) and 7.22 (lesion). Multiple scan path simulations produced re-fixation rates similar to these in both the control and lesion condition (see appendix A4.2).

**b.** Re-fixation rates during the full 100 "fixations" of the scan paths shown in figures 7.2 (control) and 7.22 (lesion).

The scan path shown in figure 7.22 was able to move more extensively throughout the ipsilesional hemifield than those shown in figures 7.17 and 7.19, which have much weaker orbitofrontal feedback. This suggests that some form of inhibition of return bias, possibly from a frontal area such as orbitofrontal cortex, is required to allow neglect patients to examine ipsilesional hemifields extensively. However, this bias may be weaker than normal and still produce reasonable search within the ipsilesional hemifield. A comparison of figures 7.18, 7.21 and 7.24 shows that the re-fixation rate increases as the weight of the novelty feedback decreases. This is

the same effect that was found in chapter 6 for unlesioned overt scan path simulations when the weight of the novelty bias was systematically varied. Results here suggest that scan paths may become more erratic and have high re-fixation rates when the effect of the novelty bias is weaker than normal. In patients, such a reduction could be due to:

- lesions within the source area (such as orbitofrontal cortex),
- the loss of connection from the frontal (source) area to parietal cortex (target),
- the bias being ineffective in parietal cortex due to lesions therein.

Such behaviour has been observed in parietal and orbitofrontal patients (Hodgson et al., 2002; Husain et al., 2001; Mannan et al., 2005).

### 7.2.5 Simulations Using Active Vision

A subset of the orbitofrontal lesion simulations were re-run under conditions of active vision, where the retina was allowed to move around the scene. The retinal size was set to a radius of 220 pixels, as in previous analysis in chapter 6. The same image was used as that for the covert simulations, in order to compare results with those above. For comparison with normal active vision conditions, figures 7.25 and 7.26 show a scan path and novelty map for a simulation without lesions.

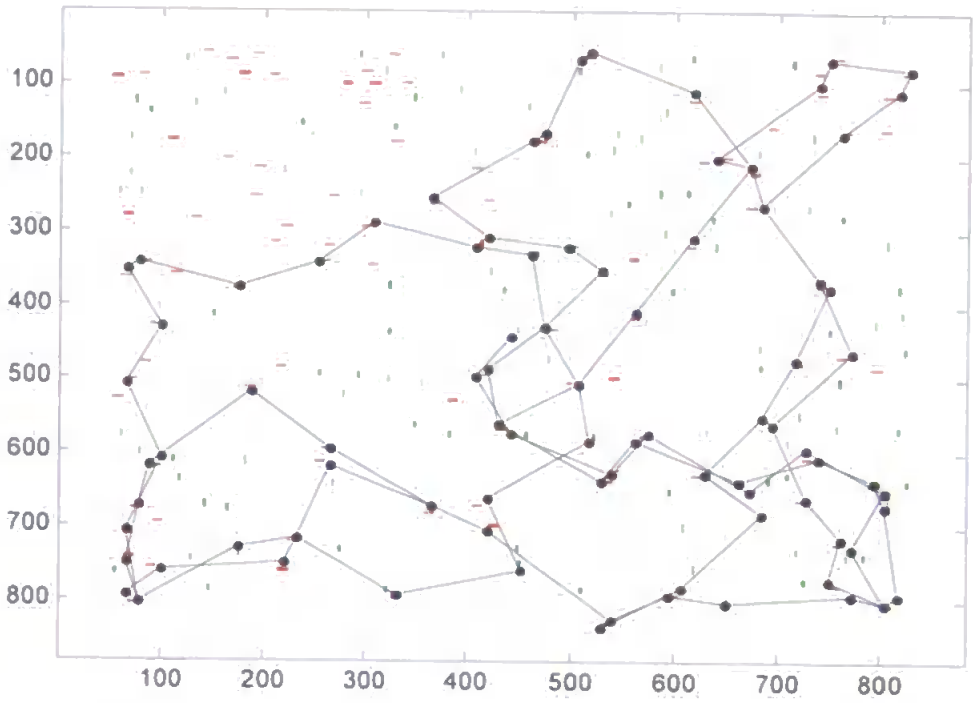



**Figure 7.25 Scan path of 100 fixations under normal active visual search conditions without lesion**

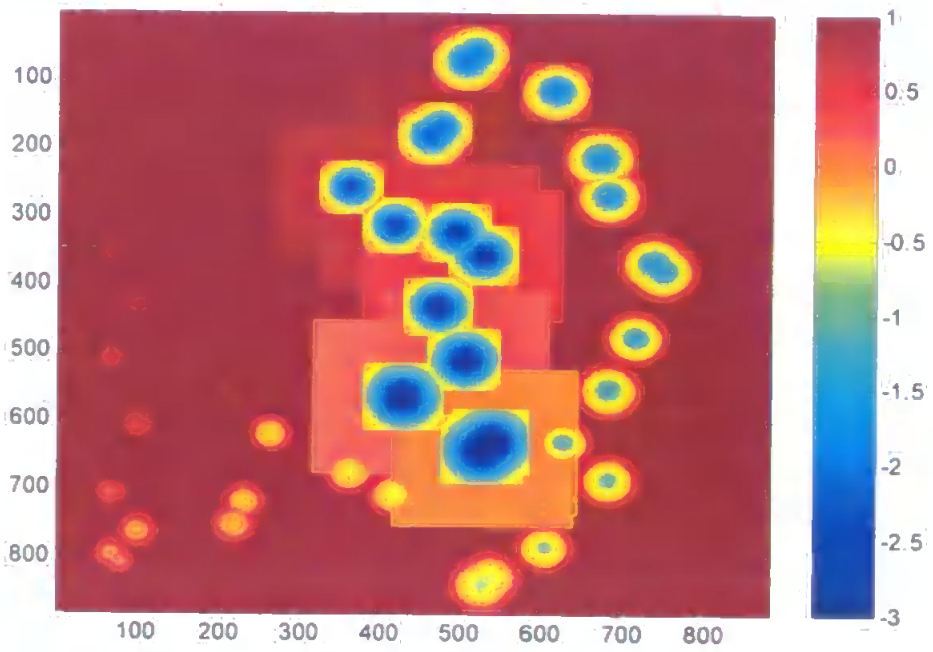



**Figure 7.26 Novelty map at the end of the scan path shown in figure 7.25**

### 7.2.5.1 Unilateral Orbitofrontal Lesion Under Active Search

Figures 7.27 to 7.30 show that, like those under covert attention conditions, unilateral orbitofrontal lesion (i.e. unilateral lesion of the novelty map) resulted in scan paths that tended to favour the ipsilesional hemifield.

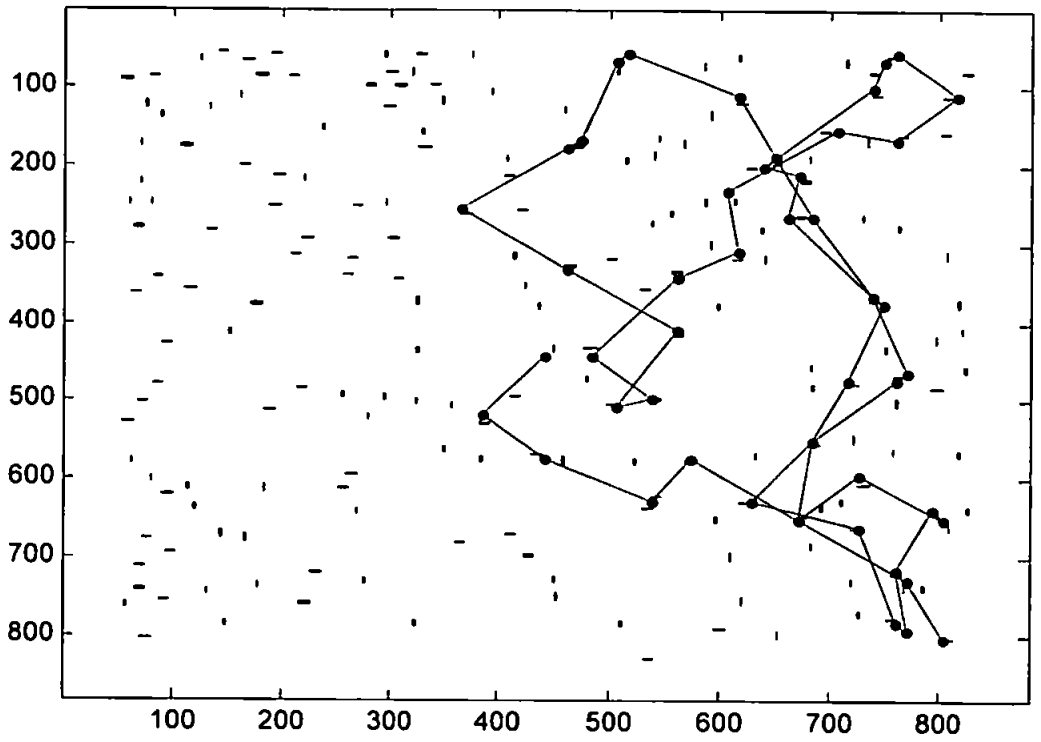



**Figure 7.27 Scan path of 50 fixations following a unilateral lesion to the novelty map under conditions of overt attention**

The active visual search scan path neglects the contralesional hemifield.

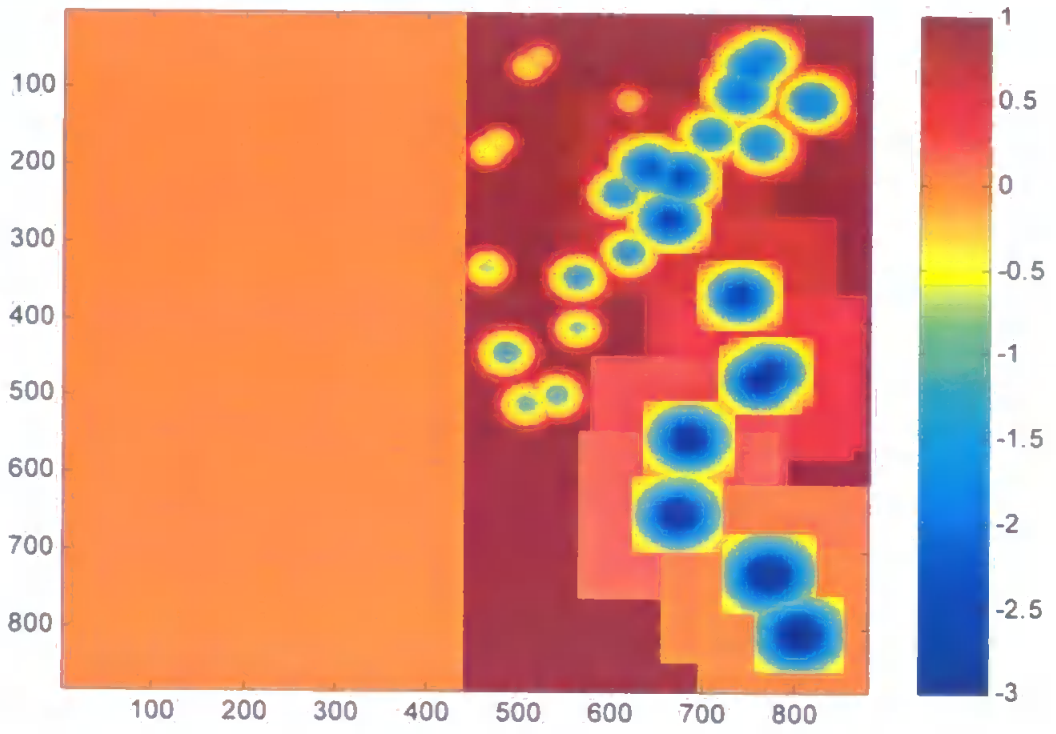



**Figure 7.28 Unilaterally lesioned novelty map at the end of the scan path shown in figure 7.27**

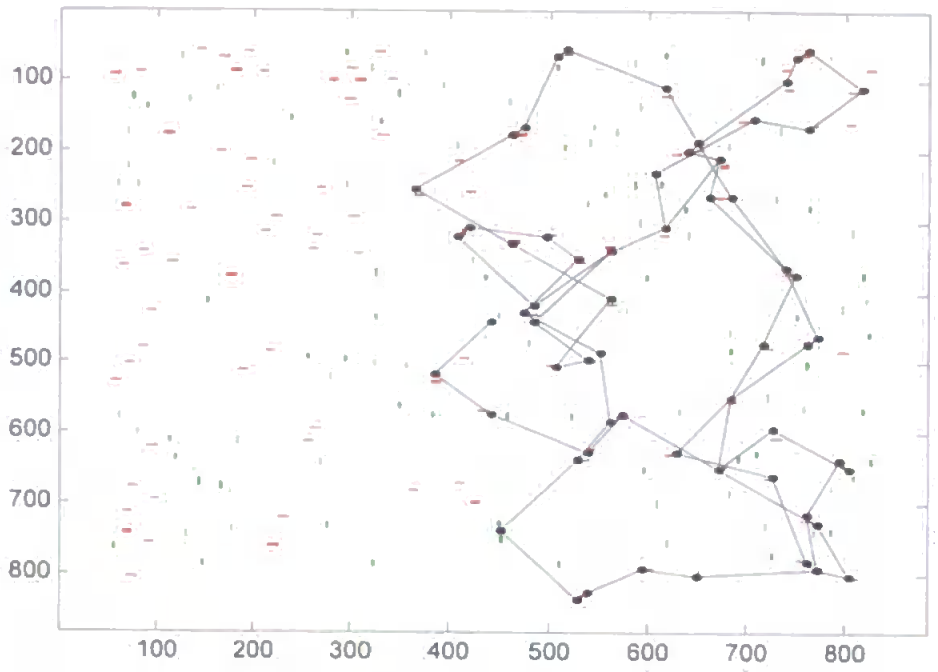



**Figure 7.29 Scan path of 100 fixations following a unilateral lesion of the novelty map under conditions of overt attention**

The first 50 fixations were shown in figure 7.27.

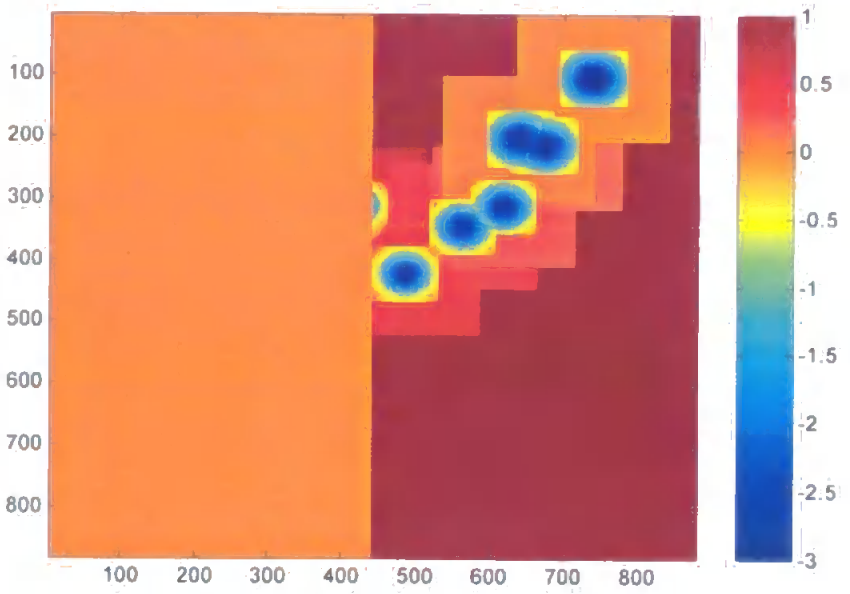



**Figure 7.30 Unilaterally lesioned novelty map at the end of the scan path shown in figure 7.29**

Under overt, as well as covert search, the unilateral orbitofrontal lesion led to an increase in re-fixation, as shown in figure 7.31. Multiple scan path simulations produced an average re-fixation rate that was higher (16.4% compared to 6%) than that shown for the individual 50 fixation scan path simulation shown here. However, like that described in section 7.2.1 for the covert simulations, this effect is due to the scan path entering the contralesional hemifield and being less able to re-orient attention therein.

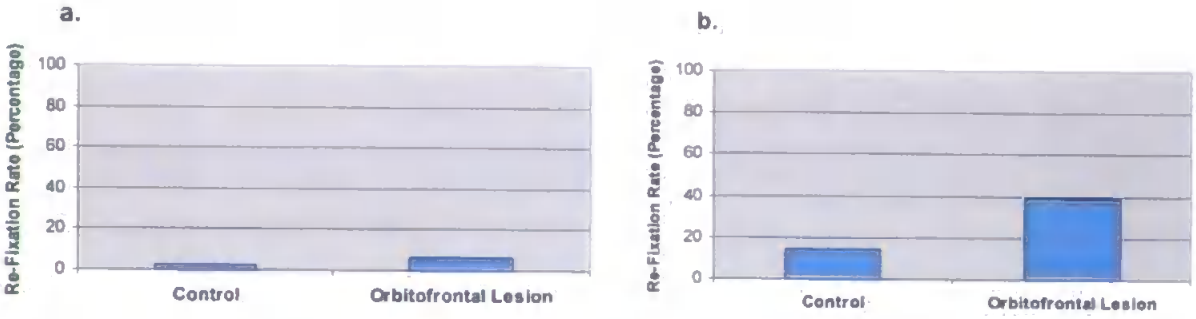



**Figure 7.31 Re-fixation rates following unilateral orbitofrontal lesion, compared to control, under active search**

**a.** Re-fixation rates during the first 50 fixations of the scan path shown in figure 7.25 (control) and the 50 fixation scan path shown in figure 7.27 (lesion). Multiple scan path simulations produced an average re-fixation rate in the lesion condition that was higher (16.4%) than that shown for this scan path (see appendix A4.2).

**b.** Re-fixation rates during the 100 fixations of the scan paths shown in figures 7.25 (control) and 7.29 (lesion).

### 7.2.5.2 Full Orbitofrontal Lesion Under Active Search

Consistent with the results found for covert search, a full orbitofrontal lesion (novelty map containing zeros) led to restricted scan paths compared to the control condition, as shown in figure 7.32 and high re-fixation rates, as shown in figure 7.33.

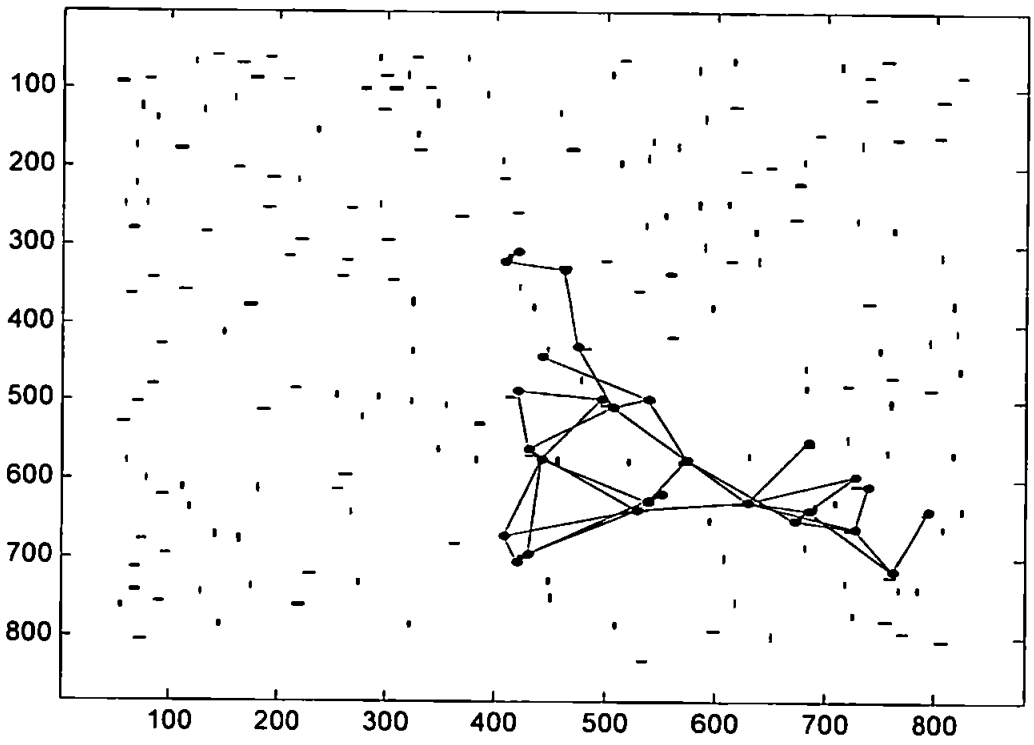



**Figure 7.32 Scan path of 100 fixations following a full lesion to the novelty map under conditions of overt attention**

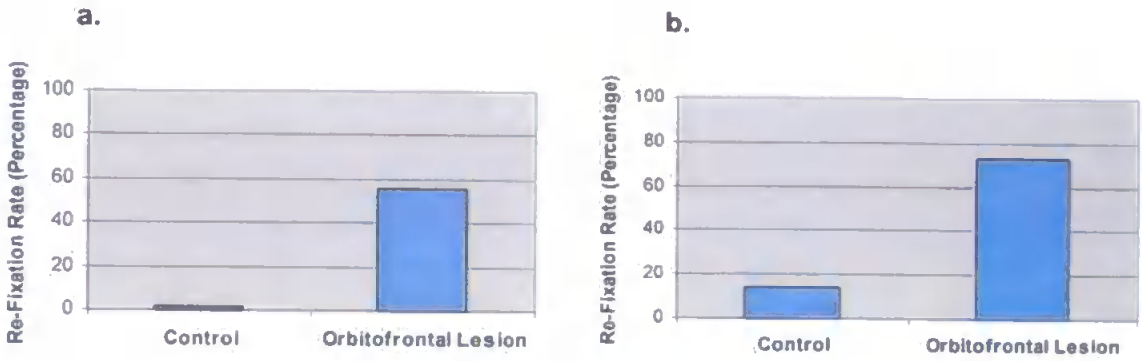



**Figure 7.33 Re-fixation rates following full orbitofrontal lesion, compared to control, under active search**

**a.** Re-fixation rates during the first 50 fixations of the scan path shown in figures 7.25 (control) and 7.32 (lesion). Multiple scan path simulations produced an average re-fixation rate in the lesion condition that was higher (70.8%) than that shown for this scan path (see appendix A4.2).

**b.** Re-fixation rates during the 100 fixations of the scan paths shown in figures 7.25 (control) and 7.32 (lesion).

The effect of varying the weight of the novelty bias under overt attention was described in chapter 6 using statistics collected from multiple scan paths over three images containing dense, sparse and very sparse stimulus density respectively. For each image, the average results from 10 scan paths, each consisting of 50 fixations, were reported. Therefore, this was a more systematic study than that shown here. Consistent with the results for covert and overt search here, reference to figures 6.23 to 6.25 shows that increasing the weight of the novelty bias from zero to its normal value has the effect of reducing re-fixation rates. In particular, orbitofrontal feedback that is weighted more than one order of magnitude less than the normal

weight of $9e^{-4}$ produces very high re-fixation rates for overt search. This means that damage to the orbitofrontal feedback will result in increased re-fixation in search scan paths. However, increasing the strength of novelty bias beyond its normal value appears to have little effect in reducing re-fixations.

Overall, covert search appears to produce slightly more re-fixation than overt search (data from appendix A4.2):

- 12.8% (covert) compared to 4% (overt) average re-fixation in control conditions
- 42.4% compared to 16.4% average re-fixation when the orbitofrontal bias is unilaterally lesioned.
- 87.2% compared to 70.8% average re-fixation when the orbitofrontal bias is completely lesioned.

This effect could be due to the static nature of the bottom-up inputs under conditions of covert search. A strong set of stimuli available within a receptive field may be capable of capturing attention many times whereas a stimulus positions move across different receptive fields under overt conditions. In real biological situations such an effect in covert search would be eliminated by microsaccadic eye movements.

## 7.3 Other Models of Visual Neglect

Deco and Rolls (2002) use a simplified version of Deco's model (Deco, 2001; Deco & Lee, 2002; Rolls & Deco, 2002), which was discussed in chapter 2, to simulate lesions leading to visual neglect behaviour during visual search. The model has connections between V1 and a posterior parietal module (PP) defined by Gaussian connection weights. Importantly, local lateral connections were introduced in the V1 and PP modules in this version of the model. These provide inhibitory input from surrounding locations. Parietal lesion is simulated by reducing activity in half of the PP module using a gradient factor to simulate gradually increased impairment towards the patient's left. The effect of the lateral inhibition in combination with the gradient of impairment allows the model to simulate object-based neglect of the left side of individual objects. Also, when two objects are viewed, the distance of their separation (in relation to the extent of lateral inhibition) determines whether the right side of both objects are "perceived" or whether only the right side of the right-most object is "perceived". If connecting lines join the two objects, only the right-most area of the combined object is then "perceived". These are quite remarkable simulation results, being based on the neurobiology and mirroring neurological symptoms. However, a very small difference in activity in the PP module determines which areas of space are "perceived" are which are not. Therefore, this threshold appears to be very well tuned to the specific stimuli configurations used and it is unclear how well the model would be able to generalise. Also, it is not clear that a difference in activity

in posterior parietal cortex indicates conscious perception. The simulations described in this chapter show that lesions to the LIP module of the model presented here lead to reduced activity in the ventral stream, which is linked to conscious perception.

The Selective Attention for Identification (SAIM: Heinke & Humphreys, 2003) model has been used to replicate a range of behaviours found in parietal patients. In SAIM, graded lesions to the selection network (which is linked to the pulvinar but could have possible parietal correlates) result in either spatial or object-based neglect, or a combination of both. The reason for this is the way the selection network maps inputs into the Focus of Attention (FOA). This mapping performs a transformation from a retinal coordinate system into an object-based system. Different effects are produced depending whether the network is lesioned vertically or horizontally because this affects mappings from the left/right side of the visual field or to the left/right side of the FOA respectively. It is interesting that the model is able to produce object-based as well as spatial neglect through these spatially-based lesions. Object-based neglect occurs following horizontal lesion of the selection network so that the left hand side of objects are neglected wherever they are presented in the visual field. Also, objects with higher pixel mass on the left hand side of the object suffer less neglect than objects with higher pixel mass on the right hand side. The authors support this finding with evidence from patient studies.

SAIM is able to simulate extinction under conditions of vertical lesion and combined horizontal and vertical lesion. The vertical lesion simulations of extinction required limiting the time for which the stimuli were presented to the system. The stimulus in the "good" hemifield was perceived and then the bottom-up input was removed before the system was able to re-orient its FOA to the other stimulus. Therefore, it is not clear whether this effect was due to the lesion or removal of inputs. However, inputs are similarly removed in patient studies. Also, the lesion did increase the time that the system took to shift its attention to the neglected stimulus and evidence for this effect in patients is cited by the authors. Some of the simulated effects are due to the non-neglected stimuli becoming misaligned within the FOA and failing to match their object template. This may be a valid prediction of biological problems but could also be a computational artefact. The neurobiological identity of the mappings and lesions described in SAIM are not fully articulated because the aim of the model was not to provide complete neural plausibility. Also, the input is restricted to very small images, which have limited spatial positions for stimuli. However, the model captures many aspects of selective processing in human vision and provides a range of impressive simulations of both spatial and object-based neglect. The results from the combined vertical and horizontal lesions of the selection network suggest that simple retinotopic spatial lesions alone may not be capable of explaining the range of disorders associated with neglect and extinction.

## 7.4 Chapter Summary

Under conditions of lesion, the simulated scan path behaviour mimicked that seen in neurological case studies of visual neglect, where scan paths were attracted to the ipsilesional hemifield. Such neglect behaviour here follows a unilateral lesion to the LIP module, which also leads to a slight increase in re-fixation in the scan path. This is consistent with reports of scan path behaviour from parietal lesion patients (e.g. Husain et al., 2001; Mannan et al., 2005) who tended to exclusively scan the ipsilesional hemifield and also failed to remember previously searched locations so that they made repeated saccades back to them.

Cases of object-based neglect (e.g. Walker et al., 1996) have not been simulated here. Lesions within the model's ventral stream may be a possible method of simulating such behaviour. An intermediate retinotopic stage representing posterior inferior temporal area TEO could be added in the ventral hierarchy below the invariant representation in IT, which represents anterior inferior temporal area TE. Lesion to this module could provide a spatial deficit in object perception. Also, the model currently lacks local lateral connections within the ventral stream. Such connections may be important for grouping of features into object shapes. The addition of such connections was found to be important in simulating object-based neglect in a similar model (Deco & Rolls, 2002). Also, simulation of a LIP lesion under conditions of overt search could produce interesting results. Under such a simulation, the contralesional side of whatever was currently fixated (within

the retina) and attended (within the AW) would be neglected and this frame of reference would move with the position of the retina. Such an overt search simulation would be possible if the computational constraint relating to the image edge were removed by assuming that the image was continuous (replicating information across the edge boundary) or preventing fixation near the edge.

Simulations with unilateral damage to the novelty map, reflecting a unilateral orbitofrontal lesion, also exhibited scan path behaviour characteristic of neglect. In some cases, following such a lesion, the scan path managed to enter the contralesional hemifield. This contrasted with behaviour following a unilateral lesion to the model's LIP module, where the scan path was not able to leave the ipsilesional hemifield. However, following unilateral orbitofrontal lesion, search within the contralesional hemifield suffered problems re-orienting attention.

In the simulations presented here, the lesioned half of LIP was completely destroyed. However, if this region had been only partially lesioned or was subject to a smaller reduction in its activity, the scan path would still have been biased towards the ipsilesional hemifield but may have been able to enter the contralateral hemifield for some fixations, as is sometimes observed in patients (e.g. Mannan et al., 2005). Further simulations could also adopt the approach of graded lesion, similar to that used by Deco and Rolls (2002) and Heinke and Humphreys (2003), so that the severity of the deficit increases further to the left extremity of the visual field.

Damage to the orbitofrontal novelty feedback had a more substantial effect on increasing re-fixation rates than damage to the LIP module, and tended not to allow the scan path to reach some novel stimuli. The reported increase in re-fixation rate for a parietal patient (Husain et al., 2001) is larger than that for an orbitofrontal patient (Hodgson et al., 2002). Comparison across different patient studies is difficult due to the different experimental paradigms used. However, damage to both cortical areas individually is reported to lead to significantly increased re-fixation rates (Hodgson et al., 2002; Husain et al., 2001).

Here, simulated scan paths under conditions of severely reduced orbitofrontal novelty feedback were very restricted and erratic. Although scan paths reported in the orbitofrontal patient study (Hodgson et al., 2002) were less restricted than those reported here, they were disorganised, in addition to having a significantly increased rate of re-fixation. The simulations tended to produce higher re-fixation rates under conditions of orbitofrontal lesion than that reported by Hodgson et al. This could be due to the extent of the lesion modelled here (a complete orbitofrontal lesion) or a lack of inclusion of other prefrontal cognitive functions to provide a high-level executive search strategy.

Re-fixation rates in a parietal patient (Husain et al., 2001) were higher than those reported here following unilateral LIP lesion alone. This could be due to more extensive lesion in the patient affecting other parietal areas and, possibly,

impairment in other cognitive function. When orbitofrontal damage was combined here with a unilateral LIP lesion, scan paths were confined to the ipsilesional hemifield, were quite restricted and often re-visited the same locations.

Following unilateral damage to the LIP module, attentional modulation within the model's V4 module was affected when the AW extended into the neglected hemifield. This meant that stimuli in the neglected hemifield did not receive a spatial attentional benefit and these representations were less active in V4 than those in the AW ipsilateral to the LIP lesion. It has been suggested that reduced activity in ventral stream areas for extinguished stimuli, compared to normally perceived stimuli, may be the neural correlate of unconscious versus conscious perception (Driver et al., 2001; Rees et al., 2002; Vuilleumier et al., 2001). The results here suggest that a parietal lesion can have a direct impact on reducing the level of activity in the ventral stream and this effect could cause the lack of conscious perception of stimuli in the neglected hemifield. Further work could examine the effect of the reduced activity in V4 upon object representations in IT to explore whether the model's representations could give some insight into conscious perception of objects.

The current lesion simulations provide interesting initial results and indicate a capability to replicate further lesion data with this model. Future work, in relation to neglect scan paths, could include simulation of LIP lesions under overt search and more selective lesioning of LIP to provide only a slight directional bias to the

competition for attentional capture. Also, the effects of stimulus density under lesion conditions could be examined further to test the impact of AW scaling combined with spatial deficits in LIP. The effects of parietal lesions on representations in the ventral stream would be a very interesting area to pursue further, in order to investigate the difference between conscious and unconscious representations of features and objects in V4 and IT respectively.

A summary of the effects of the simulated lesions described in this chapter is presented on the following page.

**Summary of the Effects of Lesions To the Model**

| Lesion | Effect |
|---|---|
| **Unilateral parietal (LIP) Lesion** | ▪ Symptoms of unilateral neglect in scan paths<br>▪ Lower activity in V4 contralateral to parietal lesion<br>▪ Small increase in re-fixation rates |
| **Unilateral orbitofrontal lesion** | ▪ Symptoms of unilateral neglect in scan paths but less severe than unilateral parietal lesion<br>▪ Small increase in re-fixation rates<br>▪ Difficulty re-orienting attention when fixation is contralateral to lesion |
| **Full orbitofrontal lesion** | ▪ Substantially increased re-visiting of locations in the scan path<br>▪ Restricted scan path exploration |

# Chapter 8

# Discussion

This work used a computational cognitive neuroscience approach to examine processes of visual attention during active visual search. The resulting computational model is biologically plausible and is able to replicate a range of neurophysiological data at the cellular level, and psychological data at the systems level. In common with other similar models (Deco et al, 2002; Rolls & Deco, 2002), the work described here seeks to unify different levels of neuroscience: Neurobiological, neurophysiological and psychophysical.

The following main goals of this research were discussed in chapter 2:

- **To investigate computationally whether and how spatial and object-based attention can operate concurrently in the ventral stream and produce attentional effects over time courses similar to those found in neurophysiological studies of area V4.**
- **To investigate whether object-based effects occurring in parallel across V4 (Motter 1994a,b) can be used to bias spatio-featural representations in LIP towards behaviourally relevant locations and, hence, guide the attentive scan path. In addition, to test the feasibility of using this route to provide a priority for certain features in attracting the scan path.**

The first goal, which relates to the cellular level behaviour of the model, was successfully addressed by the results discussed in chapters 4 and 5. The second goal, which relates to the systems level behaviour of the model, was successfully addressed by results in chapters 4 and 6. Therefore, the computational model that has been developed and tested has fulfilled the aims of the work. That is not to say that the model cannot be improved upon and its weaknesses and areas for further development are discussed later.

The model is based on the biased competition hypothesis (Desimone & Duncan 1995; Duncan et al., 1997; Duncan & Humphreys 1989, 1992). Previous computational models of biased competition have evolved from early small-scale models (e.g. Renart et al., 2001; Reynolds et al., 1999; Usher & Niebur, 1996) to larger system-level models with more interacting modules (Deco, 2001; Deco & Lee, 2002; de Kamps & Van der Velde, 2001; Hamker, 1998; van der Velde & de Kamps, 2001). The model developed by Deco (2001; Deco & Lee, 2002) has been extremely influential in this field and has been able to replicate a range of experimental data (see Rolls & Deco, 2002). This model was discussed in chapter 2 and its contribution was to prove that it was possible to produce both spatial and object-based attentional effects from system dynamics in which object features were processed in parallel across the scene. The model presented here extends this line of research to an active vision paradigm where a spatial bias, related to the eye movement, is present concurrently with a target object bias. In common with Deco's model, the work here adopts the emerging approach in cognitive and computational modelling where no divide between pre-attentive and attentive phases is assumed. Instead one phase gradually changes into the next as the dynamics of the system evolve. Comparison with Deco's model has been made at various points throughout this thesis. Other models in the field are

now briefly reviewed for the purpose of comparison with the current work. However, this is not intended to be a comprehensive review of modelling work in this area. Shipp (2004) provides a recent interesting review of current cognitive and computational models and compares similarities in the frameworks.

## 8.1 Other Models in the Field

In the work here, the LIP module provided a saliency map not based purely on bottom-up saliency but also on top-down behavioural relevance. There is no clear distinction between pre-attentive and attentive stages in this model and the representation in LIP here is dynamically updated by attentional information so that it becomes an attentive (or behaviourally-relevant) saliency map and not simply a pre-attentive pop-out representation. A bottom-up saliency map was first suggested by Koch and Ullman (1985) as a means by which parallel pop-out search might be implemented in biological systems. However, the neurobiological identity of the map was not fully specified. The thalamus has long been implicated in the control of spatial attention (Crick, 1984). In addition, LIP (Kusunoki et al., 2000), FEF (Thompson & Schall, 2000), superior colliculus (Kustov & Robinson, 1996) and even area V1 (Li, 2002) have all been implicated in providing saliency map functions and may cooperate in this. In his review, Shipp (2004) discusses the contribution of the thalamus, specifically the pulvinar nucleus (Robinson & Petersen, 1992), and the more extended neural architecture controlling the spatial focus of attention. Drawing upon ideas in current models of attention (Deco at al.,

2002; Heinke & Humphreys, 2003; Koch & Ullman, 1985; Itti & Koch, 2000; Treisman, 1998; Wolfe, 1994), he suggests the Real Neural Architecture model, shown in figure 8.1, which is not implemented computationally. This model consists of a distributed saliency system in which the pulvinar acts as the central hub saliency map providing the attentional spotlight and communicating with superior colliculus, ventral stream areas and, indirectly, with LIP (shown as PEF: Parietal Eye Field) and FEF. This is broadly compatible with the model presented here since the current work does not suggest that LIP is alone in providing a saliency map function. The LIP module can be considered part of the distributed saliency system in which the pulvinar is a possible source of the spatial bias to LIP and V4.

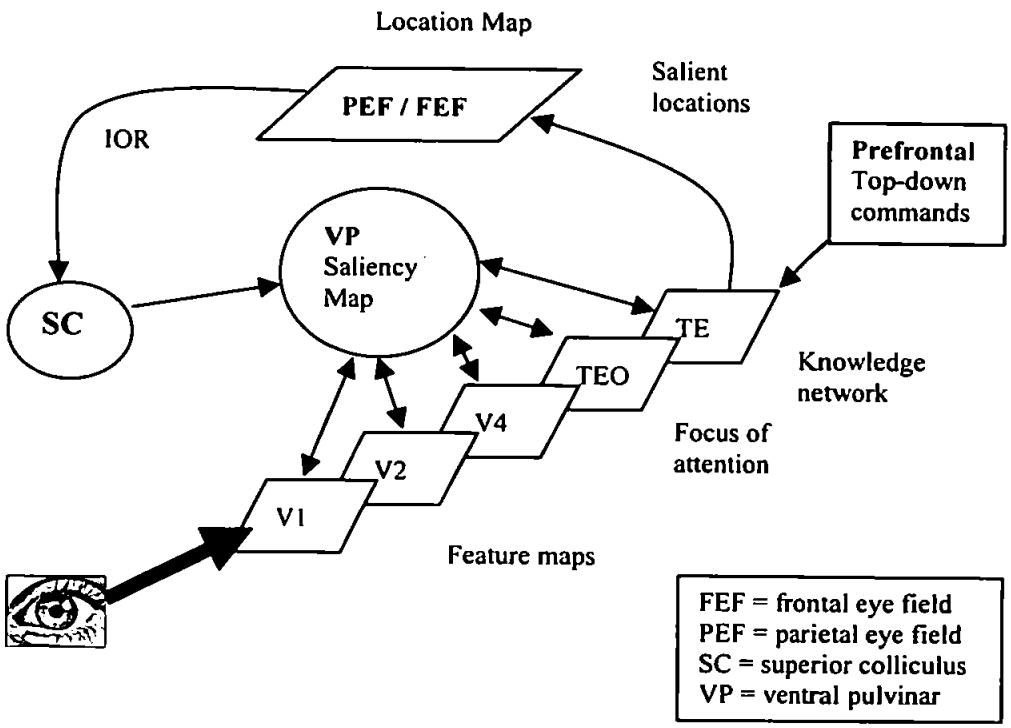



**Figure 8.1 Shipp's Real Neural Architecture Model**

Adapted from Trends in Cognitive Sciences, 8(5), Shipp, S., The brain circuitry of attention, 223-229, Copyright (2004), with permission from Elsevier

The pre-attentive saliency map suggested by Koch and Ullman (1985) was implemented by Itti and Koch (2000). In Itti and Koch's model the focus of attention is governed entirely by bottom-up factors. Low-level features (colour, intensity and orientation) are extracted from the image and centre-surround difference-of-gaussian processing provides within-feature competition that results in the formation of a conspicuity map for each feature, i.e. the most conspicuous location for each feature is determined. The conspicuity maps are then linearly summed to create a saliency map, in which integrate-and-fire neurons with strong global inhibition provide a winner-take-all (WTA) competition to determine the winning location. The saliency map is also subject to inhibitory feedback from the WTA array so that the most salient is subsequently inhibited to provide inhibition of return. This model can be used on a range of images to focus attention serially on the most bottom-up salient locations. Its functionality is, however, limited by a lack of top-down attention control such as that presented in the work here. However, a very recent and large extension to the Itti-Koch model allows task knowledge to bias the processing of features so that task information influences the capture of attention (Navalpakkam & Itti, 2005). In terms of the work presented here, competition in LIP could be extended to allow further bottom-up factors to influence the capture of attention and this is discussed later. It would be possible also to include more top-down task-related information to bias the competition between features in the model's ventral stream towards task-relevant features in, perhaps, a more general manner than the object-specific bias currently does.

The model presented here is not the first model in which both bottom-up and top-down factors have been used to influence the capture of attention. In earlier chapters this work was contrasted with similar models by Wolfe (1994) and Deco (2001; Deco et al., 2002; Deco & Lee, 2002; Rolls & Deco, 2002). The current work, of course, draws much inspiration from Deco's model and similarities and differences were commented upon in chapter 2. Similarities with Wolfe's (1994) Guided Search 2.0 model were noted in chapters 2 and 4. In particular, the representation in the LIP module can be likened to Wolfe's *'activation map'*. The retinotopic representation in LIP here is guided by information from the ventral stream (V4) feature maps. These feature maps resolve bottom-up and top-down influences in a similar manner to Wolfe's *'feature maps'*. His input channels correspond to the feature detection performed here in the retina and V1.

The model here could be extended to include more bottom-up saliency factors in the competition for attentional capture in LIP so that the interplay between the endogenous and exogenous capture of attention could be examined in detail. The model has been designed to allow an extension for lower resolution magnocellular information to be processed ahead of the higher resolution parvocellular information. LIP is able to respond very quickly to new onsets (Bisley, Krishna & Goldberg, 2004). Low-resolution form information relating to sudden onset or motion could be conveyed rapidly to LIP here. Then, during its early response, LIP would act as a bottom-up saliency map and allocate the focus of attention to a particular location. Later, the object-based effects from the V4 bias would provide the top-down attentional response in LIP. The early saliency effect in LIP would

provide a spatial enhancement in V4 at those locations (due to the LIP bias to V4) so that slower bottom-up high-resolution information at these locations would be enhanced. This is similar in concept to the *'blob map'* utilised by Olshausen et al. (1993) in their dynamic routing circuit and they suggested posterior parietal cortex as the neural substrate for this. The blob map is a low-resolution description of the most salient locations in the scene and is used to focus and scale the attentional window within which object information is collected. Information is selected within this window as a result of control neurons dynamically adapting weighted connections (synaptic strengths) in order to set a band of open connections so that information within a windowed region of V1 is selectively routed to higher cortical areas. This band of open connections determines the size of the attentional focus. The spatial bias projecting onto the ventral stream is similar to, although much less exclusive than, the window of open connections in Olshausen's model. Possibly, the neural substrate of the Olshausen's control neurons, hypothesised to be the pulvinar, could be similar to that of the spatial bias in the current work.

Another model that utilises the concept of control neurons for directing information flow is that by Tsotsos (1993, 1995), where feedback provides selective tuning of connections through a cascade of winner-take-all networks. A form of biased competition was implemented in this model before this concept had been promoted in the neurophysiological literature. In Tsotsos' model, the target object is selected at the top level of processing on the basis of the bottom-up saliency of incoming stimulus information and, possibly, top-down biasing for certain features or locations. Feedback throughout the network leads to those feedforward paths that

do not contribute to the target location being selectively pruned in a manner that results in an 'inhibitory beam' around the selected target. Again, this is slightly more restrictive for non-attended stimuli entering cortical processing than the model developed here, which adopts a more parallel processing approach albeit with a spatial selective bias.

In their influential Selective Attention for Identification Model (SAIM), Heinke and Humphreys (2003) adopted a similar approach to that of Olshausen et al. (1993) in that a selection network controls which inputs are routed to higher levels. However, SAIM contains top-down control in the form of a knowledge network that stores memories of objects. The fundamental aim of the original SAIM was to provide translation invariant object recognition. Bottom-up processing of visual information is carried out in the *contents network* and the *selection network* contains the control neurons that modulate the contents network. Top-down knowledge in the *knowledge network* modulates activity in the selection network such that competition between positions in this network is biased. Hence, a form of biased competition is also implemented in this model, where competition in the selection network is spatial and in the knowledge network is object-specific. However, in contrast to the work here, SAIM is not biologically constrained in its implementation and models at a higher, more conceptual, level. However, some similarity can be drawn between the modules in SAIM and the neurophysiology modelled here as follows:

- Contents network ≈ the early to intermediate ventral stream (V1-V4), featural representation
- Knowledge network ≈ IT, object representation
- Selection network ≈ the dorsal stream, providing a focus of attention by selectively gating the contents network/ventral stream. The authors relate this area to the pulvinar, as suggested by Olshausen et al. (1993), but lesions to this part of the network produce symptoms of visual neglect, which is associated with parietal damage (see chapter 7).

Proximity constraints built into the selection network mean that the bottom-up capture of attention is biased towards large objects with pixel densities placed around their centre of mass. Thus, bottom-up attentional capture is efficient for large, densely packed stimuli that might tend to easily pop-out of the scene. A newer version of SAIM (Heinke & Humphreys, *submitted*) includes a feature extraction stage that provides input to the contents network. As a computational artefact of the system, this version favours heterogeneous representations of features so that, for example, oblique lines are more likely to capture attention than horizontal or vertical lines. However, SAIM is able to replicate much psychophysical and lesion data. An interesting feature of this model, compared to the work discussed here, is that inhibition of return is mediated by two mechanisms. The first is a spatial mechanism that inhibits the selection network and the second is inhibition of the selected object in the knowledge network. The later leads to inhibition of objects that formerly occupied a selected location, even

if they subsequently move. This idea could be adapted here to inhibit objects in the IT module. However, the object recognition component of the model here, specifically the feature binding, needs to be strengthened first.

The work here does not address issues of oscillatory neuronal activity in relation to attentional selection. It has been suggested that oscillatory behaviour could be used to 'temporally tag' features in order to attend to them and possible bind them into percepts (Crick & Koch, 1990; Singer & Gray, 1995). This suggestion has been implemented computationally by Niebur, Koch and Rosin (1993). Their model suggests that signals in the "tagged" pathway compete with those in the "untagged" pathway. Signals are "tagged" at the V1 stage (under an assumption of sub-cortical structures, such as pulvinar or superior colliculus, being the source of attentional modulation) and inhibitory interneurons in V4 act as bandpass filters that respond best to the "tagged" frequencies. Competition in V4 is biased by means of the inhibitory interneurons so that those cells responding to the selected feature win over cells that prefer the unattended feature. This results in V4 cells that are selective for an unattended stimulus only firing significantly at stimulus onset. Oscillatory models are plausible because neurons in the fronto-parieto-temporal attentional network appear to synchronise their activity during attentionally demanding tasks (Gross, Schmitz, Schnitzler, Kessler, Shapiro, Hommel & Schnitzler, 2004). However, this subject is not within the scope of the current work and is not explored further here. Other investigators, for example Borisyuk and Kazanovich (2004) are currently pursuing the modelling of oscillatory behaviour in the area of visual attention.

In common with many cognitive models (Koch & Ullman, 1985; Itti & Koch, 2000; Treisman, 1998; Wolfe, 1994), the model presented here deals primarily with attentional guidance. Models by Deco (2001; Deco & Lee, 2002, Deco et al., 2002; Rolls & Deco, 2002), van der Velde and de Kamps (2001; de Kamps & van der Velde, 2001), and Heinke and Humphreys (2003) also include aspects of object recognition, which is currently a weak area of the model presented here due to issues of feature binding, which will be discussed later.

The behaviour of the model, which was presented in previous chapters, is now reviewed.

## 8.2 Cellular Level Behaviour of the Model

At the cellular level, the work here was novel in examining the times of onset of spatial and object-based attention in detail so that these effects were replicated over time courses that reflect neurophysiological data (Anllo-Vento & Hillyard, 1996; Anllo-Vento et al., 1998; Chelazzi et al., 1993, 1998, 2001; Colby et al., 1996; Hillyard & Anllo-Vento, 1998; Luck et al., 1997; Motter, 1994 a, b).

### 8.2.1 Spatial Attentional Effects

Following fixation, the spatial Attention Window (AW) provides an early spatial effect in V4 which, like that found by Luck et al. (1997), affects early responses. This is due to a spatial bias to V4 from LIP, which is subject to a bias related to the eye movement and new fixation position, possibly from FEF (Moore & Armstrong, 2003). The spatial bias being applied during visual search as a result of the new eye position is a novel contribution of the model. The allocation of spatial attention to the location to which the eyes have arrived is supported by a number of psychophysical findings (Hoffman & Subramiam, 1995; Shepherd, Findlay & Hockey, 1986) showing that there is a spatial benefit at the location to which a saccade is directed compared to other locations, even if subjects are told to attend elsewhere.

During a fixation, attention in area V4 develops dynamically over time from an initial spatial window to become object-based. This is a different approach to that of Deco (2001) and Deco and Lee (2002), where spatial and object attention occur in two distinct modes of operation: object recognition or visual search. Here, the spatial bias is able to cause an early attentional effect (Anllo-Vento & Hillyard, 1996; Anllo-Vento et al., 1998; Hillyard & Anllo-Vento, 1998; Luck et al., 1997; Moore & Armstrong, 2003) but object-based attention takes time to develop because it is dependent on the resolution of object-related competition within higher ventral stream areas, such as IT. There is evidence from single cell recordings in V4 (McAdams and Maunsell, 2000) and the middle temporal (MT)

area (Treue & Martinez Trujillo, 1999) to support the idea that feature-based attentional effects may be additively combined with modulations based on spatial location. Motter (1994a,b) suggested that features are enhanced in parallel throughout the visual field in V4. The spatial effect here does not prevent object-based attention developing in parallel outside the spatial AW but it provides an enhancement of responses within the spatial aperture. In Motter's experiment stimuli were not placed at large eccentricities and were sparse, which would lead to a large spatial AW in this model. The spatial AW gives a possible explanation why targets are most easily detected within a zone around fixation, scaled according to stimulus density (Motter & Belky, 1998a).

### 8.2.2 Object-Based Attentional Effects

Object-based attention develops in the model's ventral stream due to a target object-related bias from prefrontal cortex. The strength of the sensory signal from prefrontal cortex builds over time in a biologically plausible manner (Everling et al., 2002; Miller et al., 1996; Hasegawa et al., 2000) according to a sigmoid function. This differs from the constant value prefrontal signal, used in some other models (e.g. Deco, 2001; Deco & Lee, 2002, Usher & Niebur, 1996, Renart et al., 2001), that produces a steadily increasing early target effect in IT. The static value prefrontal bias does not fully capture the difference in the significance of the object-based effects in the early and late portions of the response in IT and V4, where the most significant object-based effects begin from at least 150ms after the onset of the search array (Chelazzi et al., 1993, 1998, 2001). The type of signal

modelled here has been able to replicate these effects in detail in IT and V4 (see chapter 5; Lanyon & Denham, 2005a, 2005b). This is the major contribution of the model at the cellular level.

Although no target effect is present in the earlier part of the sensory response in V4 (Chelazzi et al., 2001), average early responses in IT show a slight target effect that is much less significant than that in the late response (Chelazzi et al., 1998). This effect was simulated here by the addition of a very small, but sustained, prefrontal bias, which modulates responses in IT from its earliest response. In this simulation, the strongest target effect still occurred later in the response (after ~150ms post-stimulus), due to the development of the sigmoidal prefrontal bias (Lanyon & Denham, 2005a). It is suggested that the early effect and more significant late effect observed in the monkeys (Chelazzi et al., 1993, 1998) may be due to more than one type of bias signal, or a bias signal with more than one component, for example:

- **A sustained mnemonic component**, resulting in target-related delay period activity and the small target effect seen in early sensory responses in IT (Chelazzi et al., 1993, 1998).
- **A sensory response** that differentiates targets from non-targets later in the response (Everling et al., 2002; Rainer et al., 1998) and leads to the significant target effects later in IT and V4 (Chelazzi et al., 1993, 1998, 2001).

Prefrontal cells with shorter response latencies may be able to produce the significant early target effects found in a few cells in IT, as discussed in chapter 5.

An alternative suggestion is that the source of the late bias signal is an area other than prefrontal cortex. Hamker (2003, 2004) provides a model whereby a spatial signal, attributed to FEF, and relating to the forthcoming eye movement "re-enters" ventral stream processing (e.g. area V4). Such a re-entry signal provides a means of modulating responses in a spatially specific manner late in the response but would not necessarily cause the target effect (an object-based modulation) to be increased at that time. The results presented here suggest that the source of the late bias should be a target object-related signal. However, in vivo there are many complex inter-cortical connections influencing the dynamics and it is possible that a late spatial re-entry signal could contribute towards the effect observed. Further experiments using techniques that allow accurate temporal recording of responses (such as single cell and ERP studies) should attempt to link responses in ventral stream areas, such as IT and V4, with areas such as FEF, ventrolateral prefrontal cortex, parietal cortex and, possibly, dorsolateral prefrontal cortex.

In order to simulate the onset of object-based attention in more highly trained monkeys (Chelazzi et al., 2001 compared to 1993), it was necessary for prefrontal feedback to IT to be effective slightly earlier and for IT feedback to V4 to be stronger. For the latter, it is plausible that experience and knowledge of searched objects would lead to more highly tuned feedback within the ventral stream (IT to V4; V4 to V2 etc.) and, hence, stronger target object modulation of responses. Familiarity, in particular with distractors but also the target, does appear to make search more efficient (Greene & Rayner, 2001; Lee & Quessy, 2003; Lubow & Kaplan, 1997; Wang et al., 1994); as does familiarity with the search task

(Sireteanu & Rettenbach, 2000). The result also suggests that the timing of the prefrontal feedback is adapted by learning and familiarity with the searched objects and the task. This is also plausible, partly because the timing of a strong prefrontal response is likely to be affected by the latency of IT inputs to prefrontal cells, which would be affected by novelty. The current work highlights the need for more recordings and modelling of the dynamics in prefrontal cortex. Future modelling in this area should seek to gain a better understanding of the dynamics within the sensory response in prefrontal cortex and how it interacts with IT, and other ventral stream areas, to influence the development of object-based attention.

When there were a greater number of stimuli within the retina, the increased competition resulted in slightly lower activity in IT, particularly noticeable in the level of activity of the winning target object assembly. This suggests that targets may be less easily identified in very cluttered scenes or that it may take longer for these representations to reach supra-threshold levels in IT, i.e. that reaction times may be longer. Latency to first saccade has been found to increase with greater number of distractors (McSorley & Findlay, 2003).

### 8.2.3 Lateral Connections and Perceptual Grouping

There is currently no attempt to model bottom-up grouping effects and how these might affect search strategies. At low levels of the ventral stream hierarchy, such as V1, where local form features are detected, contextual influences beyond the classical receptive field are known to modulate responses (Ito & Gilbert, 1999).

Grouping is likely to be important to the development of object-based attention, for example causing collinear facilitation of activity in V1 for cells with similar orientation preference, which might be useful in extending activity across an attended object (Roelfsema et al., 1998). Such influences have been modelled (e.g. Grossberg & Raizada, 2000; Li, 2002) using feedback and lateral connections to achieve Gestalt-like perceptual grouping. Such grouping is likely to be important in object segregation and identification as well as being influential for object-based attentional effects.

This type of grouping appears to be learnt after birth. In comparison to visual feedforward connections, which are present at birth, feedback and local lateral connections only develop significantly post-natally (Burkhalter 1993; Anker 1977; Hevner, 2000) on the basis of visual experience (Hirsh 1985, Luhmann, Martinez Millan & Singer, 1986; Toyama, Komatsu, Yamamoto, Kurotani & Yamada, 1991; Kovacs, Kozma, Feher & Benedek, 1999). Thus, young infants are unable to segregate objects on the basis of static Gestalt principles that adults use effortlessly (Arterberry & Yonas, 2000; Bower, 1974; Craton, 1996; Hicks & Richards, 1998; Kellman & Spelke, 1983; Kellman et al., 1987; Mak & Vera, 1999; Spelke, et al., 1993, 1995) and the object-based attention processes in the young brain may be very different from that of the adult.

The current model does not include lateral connectivity in V1 and V4 but this could be added in a similar manner to that modelled by Grossberg and Raizada (2000) or Li (2002). The model has been designed and implemented to accommodate such

an extension. It would be a reasonably straightforward process to add a local connectivity component into dynamic equation 4.4 of chapter 4, for example. Inputs from local assemblies could be convolved with a kernel to provide iso-orientation facilitation or suppression, for example an elongated Gaussian of appropriate orientation could be used to provide inputs from neighbouring assemblies of similar orientation preference.

Alternatively, these lateral connections could be tuned through experience of moving stimuli if a module representing MT were to be added to the model. This idea was discussed in chapter 4. Responses to collinear movement are effectively phase shifted in MT, which produces a form of perceptual grouping (Chavane et al., 2000). MT feedback to parvocellular V1 arrives before LGN input to these areas, projects onto the same layers in V1 (4B & 6) that provide the feedforward pathway (Lamme et al., 1998) and affects the earliest response of V1 (Hupe et al., 2001). Thus, MT could feedback information about grouped objects to V1 in order to influence parvocellular grouping within the ventral stream.

The effects of perceptual grouping would allow not only object-based attention effects in overlapping objects to be examined but also the effect of grouping of items during search such that identical distractors may group together to allow easier detection of the target. The Search via Recursive Rejection model (Humphreys & Müller, 1993) specifically addresses this issue. The addition of lateral connections may also be crucial for the simulation of object-based neglect (discussed in chapter 7). Hence, investigation of such local effects could provide a

very useful extension of the model and allow further simulation of object-based attention, grouping in visual search and neglect. The addition of a component that learns objects from experience of moving stimuli would be an extremely novel and informative investigation.

## 8.3 Systems Level Behaviour of the Model

At the systems level, the model produces active vision scan path behaviour during search for a colour-orientation feature conjunction target. Effects at the cellular level of the model, in particular the object-based effects, collectively provide the systems level behaviour.

Evidence for the strategy adopted during conjunction search is somewhat mixed. In some cases, it appears that the smallest group of elements in a display is used to search through (the distractor-ratio effect). However, in many studies it appears that one of the target's features is used to guide search (Bichot and Schall, 1999; Findlay, 1997; Luria and Strauss, 1975; Motter & Belky, 1998b; Scialfa and Joffe, 1998; Williams, 1967; Williams and Reingold, 2001). Hence, there is variability in the strategy used across different experiments and displays. Sobel and Cave (2002) examined the effects of display properties and found evidence for a strong bottom-up effect on search strategy. In large densely packed arrays or when features are highly discriminable (e.g. perpendicular orientations) search was conducted through the smallest group of distractors defined by a particular feature (the

distractor-ratio effect). However, when a feature dimension had less ability to distinguish the target value from distractor values (e.g. closely similar orientations), search proceeded through the group of stimuli defined by another of the target's features.

Here, bottom-up effects were not fully modelled and the work focused more on the top-down guidance of search. Therefore, not all psychophysical investigations of search behaviour can be accounted for by the current version of the model. A useful future extension would be to allow the strategy adopted by the model to be flexible depending on the current context (extracted from display properties) and task. Properties relating to density and discriminability could be extracted from the scene. The model already goes some way towards examining scene density in its fast magnocellular pathway. Currently features are highly discriminable, but the addition of further orientations and colours could allow the impact of feature discriminability to be examined.

Despite these limitations, the model was able to demonstrate top-down guidance of search scan paths towards locations containing target features and avoid blank areas of the display (Lanyon & Denham, 2004b, 2005a, 2005b). Under the stimulus conditions and top-down search requirements used by Motter and Belky (1998b), the system was able to produce scan paths that replicate the active search behaviour found in that experiment, where the majority of fixations landed near (within 1° of) target coloured stimuli. This behaviour is also similar to that found in humans

under similar conditions (Scialfa & Joffe, 1998; Williams & Reingold, 2001). Such ability has potential application in computer vision applications.

### 8.3.1 Scaling the AW

A spatial bias to LIP creates the spatial AW. This bias is assumed to result from magnocellular processing, which is influential in the serial deployment of attention (Cheng et al., 2004). The size of the AW is calculated on the basis of the density of stimuli within the retina so that it contracts in dense scenes compared to sparse ones. The bias provides a selective enhancement of ventral stream activity for the processing of stimuli within the AW. This could provide an explanation why receptive fields in IT appear to contract in dense natural scenes compared to when measured for stimuli placed in plain backgrounds (Rolls, Aggelopoulos & Zheng, 2003).

The outputs of low-resolution orientation detection in V1 were used for the purpose of scaling the AW. It is plausible for an early cortical area to be involved in this function because the cortical representation of image density appears to set the framework for target detection so that detection is invariant to set size and eccentricity when the cortical density of stimuli is held constant (Motter and Holsapple, 2000).

AW scaling was shown, in chapter 6, to be invariant to the size of the retina. Above a critical size, which is dependent upon the density of stimuli in the scene, the radius of the retina has little effect on scan path behaviour.

The size of the AW has an effect on saccade amplitude due to the excitatory bias to the competition for attention in LIP. This means that saccades are shorter within dense stimulus arrays than in sparse arrays, as found in psychophysics (Motter & Belky, 1998b). The AW is scaled based on local stimulus information and this means that the focus of attention zooms in and out, according to stimulus density, as it moves around the scene (Lanyon & Denham, 2004a). The size of saccades also varies accordingly allowing dense patches of stimuli to be investigated most thoroughly with a series of smaller saccades. Therefore, it is predicted that saccade amplitude is dependant on the stimulus density in the local area around fixation. Again, this behaviour has potential application in computer vision.

### 8.3.2 Guiding the Scan Path

As biased competition leads to object-based attention in the model's ventral steam, feature-based effects become apparent in V4. Over time, target features are enhanced across V4 and non-target features are suppressed (Lanyon & Denham, 2004b). In the case of colour, locations containing non-target colours are suppressed such that the scene is segmented by colour as if viewed through a colour-attenuating filter (Motter 1994 a, b). As this object-based effect develops in V4, it is able to influence the scan path due to its connection with LIP. Featural

information from retinotopically organised V4 biases the spatial competition in LIP such that this information becomes integrated into the saliency map encoded in LIP, and LIP is able to represent the locations of behaviourally relevant stimuli most strongly (Colby & Goldberg, 1999; Gottlieb et al., 1998; Kusunoki et al., 2000). Models, such as those described by Deco (2001; Deco & Lee, 2002), Hamker (1998) and Niebur et al. (2001), normally treat bottom-up features (colour, luminance and orientation) equally when forming a saliency map or equivalent device and do not give any special bias to colour or other feature dimensions in attracting attention. However, there is evidence to suggest that surface features such as colour can play a special role in attracting the scan path during active visual search (Motter & Belky, 1998b; Lauria & Strauss, 1975; Scialfa & Joffe, 1998; Williams, 1967; Williams & Reingold, 2001). The results here (chapter 6; Lanyon & Denham, 2004b) suggest that certain feature types, or maps (Treisman, 1988; Wolfe, 1994; Wolfe et al., 1989), can be given a priority in attracting the scan path due to the relative strength of connection to LIP from V4 cells sensitive to these features. Such interaction may be direct from V4 to LIP to create a coarse spatial resolution targeting of the scan path but may be supplemented by V1, which does not currently form part of the dynamic portion of the system, to allow finer spatial precision due to its smaller receptive fields.

The difference in the relative weight of connection from the various feature types in V4 to LIP need only be marginal in order to achieve the effect of feature priority (Lanyon & Denham, 2004b, 2005a; see chapter 6). The strength of connection to LIP from feature encoding cells in V4 may be learnt during early development such

that certain features become more important than others in attracting attention. The relative weight of these connections could be malleable on the basis of task requirements or stimulus-related information, such as the proportion of each distractor type present in the retinal image (Bacon & Egeth, 1997; Shen, Reingold, & Pomplum, 2000, 2003). If the model were to be extended to allow its search strategy to be adapted according to context and task, as discussed earlier, these connections could also be influenced by the chosen strategy. This could provide an interesting line of investigation. However, the priority for features such as colour may be difficult to override by stimulus information (Folk & Remington, 1999) or task demands (Motter & Belky, 1998b).

The V4 to LIP connection provides one (not exclusive) method by which a feature priority may be implemented. Such a "cross-stream" connection could be relevant in the case of a feature like colour that seems to have a strong influence on search under various conditions. The ability to prioritise features for search is a novel element in this work and has potential application in computer vision. However, the V4 to LIP connection could potentially convey any outcome of object-based attention that develops within the ventral stream so that equal top-down priority could be given to all features. It is suggested that object-based attention and object-related feedback signals in the ventral stream are important in general for efficient search so that fixations land near behaviourally relevant stimuli that contain target features.

Further bottom-up saliency effects that could influence feature priorities are not currently addressed. Other top-down cognitive factors may also be important. Task-relevant information held in working memory in prefrontal cortex could be used to prime a particular feature map (e.g. colour; a similar effect to that seen in human imaging: Chawla et al., 1999) to be more active than others, thus giving these features an advantage in biasing LIP. An extension to the model to include further bottom-up saliency representations in LIP is discussed later. However, during search where bottom-up saliency is equal across stimuli, the current model shows that it is possible to implement a feature priority by "cross-stream" interaction between the ventral and dorsal streams. It is predicted that, if the connections to LIP from the ventral stream (particularly V4 in relation to colour search) were lost, the scan path would be less able to target locations containing suitable stimuli, for example stimuli of the correct colour.

Simulations showed that weaker object-based effects in the ventral stream led to the scan path being more likely to land on non-target coloured stimuli. These effects were linked to the weakening of the prefrontal bias to IT or IT feedback to V4, and it was suggested that these feedback signals could be tuned by learning. This leads to a prediction that familiarity with the search task and objects can increase search selectivity by more strongly guiding the scan path by target features. Although the model is tuned (via the V4 to LIP connection) to be guided by target colour, the principle of these results applies to the capture of attention by any output of object-based processing in the ventral stream, i.e. any behaviourally relevant stimuli.

### 8.3.3 Inhibition of Return

The active vision implementation of the model raised issues in relation to a memory trace of locations already visited. This was discussed in detail in chapter 4, with additional information in appendix A2.2.

Inhibition of saccade return (Hooge & Frens, 2000) was implemented by a scene-based map of novelty value that provided a bias to competition in LIP so that activity therein was modulated by the potential reward of locations (Platt & Glimcher, 1999). This is a simplification of coordinate transformations that might occur in parietal cortex (Colby & Goldberg, 1999) but is linked to the loss of spatial memory for locations previously searched that is associated with parietal damage (Husain et al., 2001; Shimozaki et al., 2003). In chapter 4 (Lanyon & Denham, 2004b), it was suggested that parietal cortex could receive this novelty or reward-related feedback from frontal areas, such as orbitofrontal cortex (Hodgson et al., 2002), so that competition for attentional capture is biased towards novel locations during search. Damage to either parietal or orbitofrontal areas could affect the scan path and lead to increased re-fixation rates, as was shown in chapter 7.

The implementation of IOR here was able to broadly replicate a range of psychophysical data so that inhibition was present at multiple locations (Danzinger et al., 1998; Snyder & Kingston, 2001; Tipper et al., 1996), was strongest in the immediate vicinity of the fixation point and decreased with distance from the fixation point (Hooge & Frens, 2000). The magnitude of the effect decreased

approximately linearly from its largest value at the most recently searched location so that and several previous locations were affected (Irwin & Zelinski, 2002; Snyder & Kingston, 2000).

In order to allow the scan path to investigate dense areas of a scene with a series of shorter saccades and to move through sparse areas with larger saccades, it was necessary to base the update to the novelty map on the size of the AW. Thus, the region inhibited following the withdrawal of attention from a location was scaled according to local stimulus density. This leads to the prediction that a larger area around fixation is inhibited for future saccade return in a sparse scene than in a dense scene.

If the weight of the novelty bias was very weak, re-fixations in the scan path were common. Increasing the weight of the novelty bias beyond its normal value tended to increase the amplitude of saccades resulting in the next fixation to being further from the current fixation point so that the scan path was encouraged to move further afield. Highly weighted novelty also slightly reduced the selectivity of the search process. The scan path was less likely to fixate target coloured stimuli and, in sparse scenes, there were significantly more fixations landing in blank areas of the display (Lanyon & Denham, 2004b, 2005a,b; see chapter 6). Hence, a search where novelty is a key factor, for example a hasty search of the entire scene due to time constraints, may result in more "wasted" fixations in blank areas.

The requirement for a novelty map in world/head-based co-ordinates as a memory trace of locations visited suggests that visual search cannot be totally amnesic (Horowitz & Wolfe, 1998) and supports the idea of memory across saccades (Dodd et al, 2003; Irwin & Zelinski, 2002; Mitchell & Zipser, 2001). The involvement of frontal and parietal areas in this form of IOR may explain why orbitofrontal and parietal patients display increased revisiting of locations in their scan paths (Husain et al., 2001; Hodgson et al., 2002).

### 8.3.4 Latency to Saccade and Saccade Amplitude

The model was developed under a hypothesis that the development of object-based effects in the ventral stream can influence the cortical structures responsible for decisions about saccades (e.g. parietal cortex, FEF, and superior colliculus). Having replicated the onset times for spatial and object-based effects at the cellular level, the automatic onset of a saccade was temporally linked to the development of a significant object-based effect in IT. Allowing a preset time for motor preparation, saccade onset times seen by Chelazzi et al. (1993, 1998, 2001) were able to be replicated. This suggests that the link between the development of significant object-based effects in the ventral stream and the onset of a saccade during search is valid.

As saccade onset depends on the development of object-based attention in the ventral stream, there is a prediction that saccades taking place after short fixation durations may be less able to target behaviourally relevant locations due to

insufficient time during which object-based attention can develop. Hooge & Erkelens (1999) found that search was more selective and efficient when fixation durations were longer. However, it is predicted that such effects will require less time when object-based feedback is faster or stronger. This is because saccade onset is subject to the timing of prefrontal feedback and strength of object-related feedback in the ventral stream. The effects of these factors and their link with familiarity of searched objects and the task were examined in chapters 5 and 6. It is predicted that saccade latency may be shorter amongst very familiar objects or in a familiar task and, in particular, when prefrontal response latency is short. Short-term memory of target features, such as colour, does appear to reduce saccade latency (McPeek et al., 1999).

It is predicted that distractor stimuli (specifically, stimuli not sharing a target feature) may be more likely to capture attention when object-based effects within the ventral stream are weak (possibly due to unfamiliarity with the objects or the task) or have not yet had time to develop fully when the saccade is initiated. In particular, the tendency for the scan path to select only target coloured locations, during search for a colour and form conjunction target, may be weaker when the objects or task are less well known. In experiments, search has been facilitated, enabling faster reaction times, when distractors are familiar (Lublow & Kaplan, 1997; Reicher et al., 1976; Richards & Reicher, 1978; Wang et al., 1994).

Overall, results here indicate that search during a familiar task or with familiar objects may be faster and the scan path would be expected to fixate more target coloured locations than when the task or objects were unfamiliar.

## 8.4 Bottom-Up Versus Top Down Capture of Attention

Bottom-up and top-down factors influencing the capture of attention were resolved through competition in both the ventral (V4) and dorsal (LIP) streams. However, the model currently works mainly as a top-down guided system and could be developed further in terms of bottom-up saliency analysis for the reflexive capture of overt attention. One idea for allowing the rapid transmission of magnocellular information to LIP was mentioned above.

Bottom-up capture of attention has been previously examined in other models, such as that described by Itti and Koch (2000). Li (2002) provides a model of V1, which suggests that V1 may operate as a saliency map for the bottom-up saliency-based capture of attention. Bottom-up saliency has a context dependent property and the addition of lateral connections in the V1 and V4 modules of the model would probably be necessary to provide a bottom-up saliency effect.

V1 has direct connections to superior colliculus and this route could be responsible for reflexive saccades, whereas the route via parietal cortex to FEF and superior colliculus suggested here appears to be linked with voluntary saccades. Voluntary

saccades that are linked to top-down information take longer to generate than saccades that operate on the basis of bottom-up saliency information (van Zoest, Donk & Theeuwes, 2004). Note that this is explained well by the time course of top-down object-based effects captured in the model. Bottom-up saliency effects via a direct route from V1 to superior colliculus or to a parietal area would be able to be processed more quickly to generate fast reflexive saccades where a highly salient stimulus captured attention in a bottom-up manner. Saccades appear to activate different regions of parietal cortex when they are voluntary compared to when they are reflexive (Mort, Perry, Mannan, Hodgson, Anderson, Quest, McRobbie, McBride, Husain & Kennard, 2003). More extensive modelling of parietal cortex in this regard, and in relation to coordinate transformations across eye movements would be useful in this field generally. Within the existing model, further bottom-up factors in the competition for attentional capture could be used to bias LIP since it is neurally plausible that LIP is involved in the representation of bottom-up stimulus saliency (Colby & Goldberg, 1999; Gottlieb et al., 1998; Kusunoki et al., 2000). This also has the advantage of allowing top-down and bottom-up factors to compete for the capture of attention within a single module. The biased competition approach allows additional factors to be easily added to bias the competition within LIP. Hence, there is much scope for future work to investigate the interplay between the top-down and bottom-up capture of attention using existing modules or by adding additional modules such as superior colliculus. Superior colliculus has very large receptive fields and may be used, for example, for the coarse localisation of express saccades.

## 8.5 The Resolution of Spatial Attention and Saccade Targets

The spatial bias applied to LIP and transmitted to V4 operates by increasing activity for neurons whose receptive fields overlap with the attention window. This means that the precision of spatial attention is limited to the size of the V4 receptive field. The addition of a V1 module would allow a finer spatial precision for spatial attention within the ventral stream. This would allow certain locations within the V4 cell's receptive field to receive increased bottom-up activation, thus encouraging the cell's response to be determined by the stimulus at the attended location. However, Intriligator and Cavanagh (2001) have shown that the spatial resolution of the spatial attentional focus is coarse and the current model is plausible in this regard.

The centre of the receptive field of the LIP assembly that is most active at the time of a saccade is chosen as the end point of the saccade. Locations in LIP are activated by featural inputs from V4, which receives bottom-up inputs from V1 over each receptive field area. Therefore, the size of V4 and LIP receptive fields (currently set to 3°) determines the spatial precision of the localisation of saccade end points. Receptive field sizes in LIP tend to be larger than those modelled here (averaging 12.8° wide during fixation, with 67% being less than 15° wide: Ben Hamed et al., 2001). However, increasing the size of the modelled receptive fields in V4 and LIP reduces spatial accuracy for fixation point localisation. There is some evidence in psychophysics that saccade control is only coarsely localised (Findlay, 1997) such that many fixations tend to land in the region neighbouring a

target, including landing on neighbouring distractors. Also, when double targets occur in adjacent locations, saccades often land in intermediate positions, appearing to "average" the target locations. These "averaged" saccades tend to avoid locations around non-targets. This would be explained here by non-target suppression, as a result of object-based attention, in the ventral stream. It has been suggested that the averaging could be due to the large receptive fields in superior colliculus producing this "global effect" (Findlay, 1982, 1997) whereby separate stimuli are treated globally by the saccadic system. The spatial precision of stimuli identified within the ventral stream may be lost as information is conveyed to parietal cortex and onwards to superior colliculus for saccadic processing. However, in addition to the inputs from parietal cortex that are modelled here, saccadic systems in FEF and superior colliculus may receive more spatially precise inputs (from V1 to superior colliculus, for example), which allow better precision in saccade end-point selection. These extensive networks of eye movement control are not fully modelled here. However, the representation in the modelled LIP demonstrates the principle of how the ventral stream may inform areas associated with attentional capture and guide the eyes accordingly.

## 8.6 Feature Binding & Target Detection

The model is currently focused on the guidance of visual search rather than object recognition. A weakness in the current system is its inability to recognise the target object and terminate the search scan path. Ideally, the target object in IT should be more active under the condition when it is both primed by prefrontal feedback and receives bottom-up inputs from the target stimulus than when it receives the prefrontal priming but does not have a target stimulus in its receptive field. However, setting this detection threshold has proved difficult. This is a new issue in this area, raised by the active vision implementation. Previous biased competition models (e.g. Deco, 2001) had static retinas that encompassed the entire scene, which always contained the target. In common with these models, the prefrontal bias here is a simple current injection applied in all cases. However, the sigmoidal element of the bias represents a sensory response to the target object. This signal would only be expected when the relevant cells in prefrontal cortex received bottom-up inputs relating to the target object, i.e. when it is present in the retina. The lack of a prefrontal sensory module in the model means that this current injection is applied to IT in all cases. If the prefrontal dynamics were explicitly modelled, it may be easier to detect the presence of the target in IT.

Another problem that hinders target detection is a weakness in feature binding across feature dimensions in V4. Under the current manually set feedforward connections from V4 to IT, illusory conjunctions can occur within the IT receptive field, which covers the entire retina. For example, the red vertical object assembly

in IT receives input from all red and vertical assemblies in V4. These assemblies may have been activated by red horizontal and green vertical stimuli respectively and, therefore, there is a risk of an illusory conjunction of red and vertical at the level of the IT receptive field. Note that this problem was not overcome by a similar biased competition model (Deco, 2001; Deco & Lee, 2002, Rolls & Deco, 2002, chapter 9), which had only one feature dimension in its original version. When extended to two feature dimensions, the IT module was removed and replaced with a direct current injection (Rolls & Deco, 2002, chapter 10). The current work was not primarily concerned with accuracy in feature binding. However, the bias signal from LIP to V4 illustrates the concept of a parietal source of binding in the ventral stream (Corbetta et al., 1995; Shafritz et al., 2002. Under this model, damage to parietal cortex would affect the binding across feature dimensions but the binding of form into shape, which takes place within the ventral stream hierarchy, would be remain intact. This problem has been noticed in parietal patients (Friedman-Hill et al., 1995; Humphreys et al., 2000). The level of spatial precision for feature binding is the size of the V4 receptive field. This is because the excitatory inputs from LIP to V4 affect the binding. The introduction of a V1 module connected with LIP would allow binding at the precision of a V1 receptive field, up to 1°.

If the issues of feature binding were resolved, the model could be further enhanced to allow identification of more types of feature in V1 and V4. For example, detection of a wider range of orientations and colours would allow the system to deal with natural images. The model's scan path behaviour would have very useful

applications in natural image processing. Also, lesion simulations could provide a computational explanation of specific binding errors found in parietal patients (Friedman-Hill et al., 1995; Humphreys et al., 2000).

## 8.7 Serial Versus Parallel Search

This model adopts a biased competition approach but includes a serial spatial component, relating to the eye movement. The spatial AW bias is just one of a number of biases affecting competition within the system. This approach suggests that there is scope to reconcile the differing views of attention. The serial spatial spotlight view (Crick, 1984; Helmholtz, 1867; Treisman, 1982, 1988, 1991, 1992, 1998; Treisman & Gelade 1980; Treisman & Gormican 1988) suggests that space is an important feature that is selected early. This was accommodated by the spatial bias, which has an early spatial attention effect. The alternative view that all features are processed in parallel and objects compete for access to short term memory (Desimone & Duncan 1995; Duncan & Humphreys 1989, 1992; Duncan et al., 1997) was accommodated by the feature/object biased competition processing within the model's ventral stream.

Models of attention that rely on a spatial spotlight (Eriksen & St. James, 1986; Posner, 1980) generally assume that feature search is a result of broad attentional distribution whereas conjunction search is thought to require a narrower focus. This explains why search times for simple feature search (where the target has a

unique feature, such as a different colour) are largely invariant to search array size, whereas times in conjunction search (where the target is defined by a unique combination of two or more features) increase with number of distractors. The model's AW is currently scaled according to low-resolution form information detected in V1. However, scaling the AW on the basis of a target feature, such as colour, could produce an interesting result. In the case where there was just one stimulus of unique colour, the AW would be very large due to the lack of stimuli of this colour being present and, therefore, density being low. In this *feature search* example, the wide AW would lead to all locations having an excitatory spatial bias. This would lead to fast parallel processing with the target popping out and being identified immediately. In the conjunction search example, there would be several stimuli of the same colour and the AW would be much narrower. Therefore, it would need to serially scan several distractors leading to a slower search process. This feature-based AW scaling could be performed on the basis of the target colour feature map in V1 or the retina. The detection of colour information at a low spatial resolution was already accommodated in the model in chapter 3.

Therefore, the approach of scaling of the AW offers one explanation for the dichotomy of serial versus parallel search times found in psychophysical experiments. The current work provides a basic estimation algorithm for AW scaling but this could be extended to include the influence of basic scene characteristics, such as features important to the search. Broad information such as the "gist" of a scene may be attained rapidly and be available to AW scaling. In real world scenes, the magnocellular monochromatic scene density information

suggested here might be a useful device for scaling our normal attention as we move around the world.

## 8.8 Divided Attention: Multiple Spatial "Spotlights"

The current implementation has a single "spotlight" but the model has been designed such that attention (in its covert form) could be divided into multiple "spotlights". In this situation of divided attention, the spatial bias to LIP would target more than one location. This would lead to multiple locations in LIP and V4 becoming enhanced and being equally likely to attract the next fixation (subject to other factors, such as stimulus properties and IOR input at these locations). Enhancement of activity in more than one location in extrastriate cortex has been observed in humans under conditions of divided attention (McMains & Somers, 2004). Under such conditions the spatial bias to LIP may be prefrontal in origin rather than eye movement related. This bias could alternatively be applied directly to V4 from prefrontal cortex.

## 8.9 Lesioning the Model

The effect of simple lesions to the parietal and orbitofrontal parts of the model were examined. These provided further proof of the validity of the model by simulating the type of scan path behaviour seen in such patients (Husain et al., 2001; Hodgson et al., 2002). Scan paths under simulated unilateral parietal lesion neglected the contralesional hemifield and had higher rates of re-fixation than normal. This replicated behaviours seen in patients with a spatial neglect syndrome (e.g. Husain et al., 2001). Scan paths under simulated unilateral orbitofrontal lesion favoured the ipsilesional hemifield and featured much re-fixation. Within the contralesional hemifield, there was a difficulty re-orienting attention. Full orbitofrontal lesion led to scan paths that tended not to explore the scene fully and had much higher rates of re-fixation. An orbitofrontal patient also had high rates of re-fixation and erratic scan paths (Hodgson et al., 2002). However, the patient's scan paths tended to explore the scene more fully than the simulated ones. This could be due to other prefrontal functions, not modelled here, that provide a search strategy. As the neural correlate of the novelty bias is not totally clear, the simulated lesion to the novelty map could possibly relate to an eye movement structure such as superior colliculus (Clohessy et al., 1991; Sapir et al., 1999; Trappenberg et al., 2001) rather than orbitofrontal cortex. Although the correlate of the novelty bias is less clear than that for the LIP lesion, both types of simulation produced interesting and novel results.

Of particular note, the simulation of a unilateral parietal lesion led to reduced activity in the ventral stream: Activity in the neglected hemifield of V4 was significantly lower than that in the normal hemifield. It is suggested that a spatial deficit may be found in V4 cellular activity following posterior parietal lesion, dependant on the spatial position of attention. Here, lower ventral stream activity was associated with stimuli that were neglected and not consciously perceived. Such an effect has been found in human imaging of extrastriate cortex (Driver et al., 2001; Rees et al., 2002; Vuilleumier et al., 2001). Therefore, a correlate of conscious versus unconscious perception has been found in the model as a by-product of the simulated parietal lesion.

There is much scope for further simulation of lesions within the model.

## 8.10 Conclusion

This is a biologically plausible model of active visual search that accurately replicates single cell data in V4 and IT in addition to reproducing search scan path behaviour found in psychophysical experiments. This shows that the accurate replication of data at the cellular level can produce system level behaviour that mimics that observed in humans (Scialfa & Joffe, 1998; Williams & Reingold, 2001) and monkeys (Motter & Belky, 1998b). This provides a valuable insight into the link between cellular and system level behaviour in natural systems. It also highlights the

fact that biological modelling can lend much to active vision research for practical computer vision applications.

The time course of object-based effects at the cellular level have been accurately replicated for the first time. The nature of the prefrontal bias required to replicate these effects was more complex that that used by previous similar models. It was suggested that further modelling of the dynamics within prefrontal cortex and its interaction with the ventral stream would be useful. Also, such a model would be able to overcome the problems of target detection identified here.

The cellular-level object-based effects were used to generate systems level behaviour that replicated a range of psychological data. The time course of the object-based effect could provide insight into behaviours such as those reported by van Zoest et al. (2004), where top-down target knowledge affected longer latency saccades but not short latency ones, which were driven by bottom-up information only.

The model is also novel in having concurrent spatial and object-based biases affecting the systems dynamics and linking the spatial effect to the eye movement. The model goes some way to reconcile parallel and serial views of attention. Like the model by Deco (2001), it also has no divide between pre-attentive and attentive stages because these evolve gradually from the system's dynamics.

Scaling the spatial AW enabled dense areas of a scene to be searched more thoroughly with a series of shorter saccades and sparse areas to be moved through

rapidly with larger amplitude saccades. Such characteristics may have future potential application in computer vision and surveillance, particularly with the addition of further stimulus-related saliency factors. The scaled AW also provided a possible explanation for the contracting of receptive fields in IT during search in naturally cluttered scenes (Rolls et al., 2003).

Top-down and bottom-up influences were resolved within the model's ventral stream. A "cross-stream" connection from the ventral to the dorsal stream implemented a novel method of guiding the search scan path to behaviourally relevant locations. This connection was able to provide a priority in the types of features guiding search.

It is suggested that object-based attention and object-related feedback signals in the ventral stream, in addition to the "guidance" connections from the ventral to dorsal stream, are important in general for efficient search. The selectivity of search and the time of saccade onset were dependant upon the development of object-based attention in the model's ventral stream. It is suggested that object-based effects in the ventral stream may influence the cortical and sub-cortical structures responsible for decisions about saccades (e.g. parietal cortex, FEF, and superior colliculus). It is predicted that familiarity with the search task and objects can increase search selectivity by more strongly guiding the scan path by target features. Familiarity may also decrease fixation durations leading to faster saccade onset.

This work is the first active vision implementation of biased competition and raises issues not previously encountered by biased competition models. It was necessary to maintain a memory trace of locations previously visited in the scan path but no longer available in retinal coordinates. This supports theories suggesting the use of memory across saccades (Dodd et al., 2003; Irwin & Zelinski, 2002; Mitchell & Zipser, 2001) rather than search being totally amnesic (Horowitz & Wolfe, 1998). The implementation of the novelty map bias to LIP was consistent with patient studies suggesting a parietal cortex involvement in the spatial memory for locations visited across saccades (Husain et al., 2001). The neural correlate of the novelty map is not clear but orbitofrontal cortex was suggested as a possible source of such a bias signal on the basis of abnormal search behaviour in an orbitofrontal patient (Hodgson et al., 2002).

It was important that the region inhibited for IOR scaled with local stimulus density in order to allow the scan path to investigate dense areas thoroughly. It is, therefore, predicted that a larger area around fixation is inhibited for future saccade return in a sparse scene than in a dense scene. It is also predicted that the importance of novelty in the search process affects the selectively of eye movements so that a search where novelty is a key factor may result in more "wasted" fixations in blank areas.

The model is currently primarily driven by top-down target information. Such guidance of the search scan path may be of use in applications such as web-based information retrieval, robotic vision and surveillance. The competition for attention within the system can be extended to include further bottom-up stimulus and top-

down mnemonic and cognitive biases. Thus, a biased competition model of the type described here provides an extendable framework within which additional factors can be added to compete for the capture of attention (Folk & Remington, 1999; Kim & Cave, 1999; Remington, Johnson & Yantis, 1992).

The simulation of lesions and comparison with patient data provided further verification of the model. This is an area that could be explored further. In particular, the simulation of visual neglect suggested that the model has the capability to provide insights into the neural correlates of the conscious perception of stimuli.

# Appendix A1

# Supplementary Equations

## A1.1 Random Image Creation

Image creation is based on a random process described here. The user is asked to specify the size of the image, the orientation and colour of the bar that will act as the visual search target and the number of distractor bars to be generated.

The size of the bars is chosen to conform to those used by Motter and Belky (1998a,b) in a monkey active visual search task, where bars of one of two colours and one of two orientations subtended 1° x 0.25° of visual angle. Here, 1° of visual angle is correlated with a distance of 11 pixels. This is due to the size of the kernels used for V1 processing (see chapter 3 of main text), where the high-resolution kernels (used for the majority of processing) are 11x11 pixels. It is assumed that an average V1 cell's receptive field covers approximately 1° of visual angle (0.25-1° near the fovea, up to 1.3° in the periphery: Rolls and Deco, 2002). Therefore, the bars are of length 11 pixels and of width 3 pixels.

An equal number of each distractor type is generated (based on the number entered by the user) so, for example where the target is a red vertical, there will be an equal number of red horizontals and green verticals present.

For a reasonably uniform distribution of bars (i.e. not when patches of dense stimuli are required in order to test scenes of mixed stimulus density), the placement of the bars is calculated as follows:

The starting position of a bar (startPosI,startPosJ) is calculated by:

$$startPosI = \max ( round ( rand * ( M-l-E) ) , 1 )$$

$$startPosJ = \max ( round ( rand * ( N-w-E ) ) , 1 )$$

where:

rand is a random number generator in the interval (0,1)

round ensures an integer is returned

M,N is the size of the image being generated

$l$ is the length of the bar, 11 pixels

$w$ is the width of the bar, 3 pixels

E is the distance around the image periphery in which bars will not be drawn.

To minimise edge effects during retinal and V1 stimulus and restricted retinal image information during the scan path, the image generation process avoids placing bars too close to the edge of the original image. The distance E, is set to 50 pixels. In addition to being subtracted above to ensure that stimuli are not placed in the right or bottom extremities of the image, the starting position may be incremented by E if it is less than this distance from the image's top and left edges.

The system aims to draw bars such that they do not overlap or touch each other. This is achieved by examining the local region for bars already drawn. The area to be examined is defined by:

For vertical bars:

$\Gamma$ ( max ( startPosI – r , 1 ) : min ( startPosI + *l* + r , M ) , max ( startPosJ – r , 1 ) : min ( startPosJ + *w* + r , N ) ) )

For horizontal bars:

$\Gamma$ ( max ( startPosI – r , 1 ) : min ( startPosI + *w* + r , M ) , max ( startPosJ - r , 1 ) : min ( startPosJ + *l* + r , N ) ) )

Where:

$\Gamma$ is the image

r is the local region that is examined: 10 pixels distance around the new bar

If another bar exists in this region, the current bar is not drawn and another random position is sought.

The image generation is done within two featural images:

$\Gamma^{red}$ containing the red-coloured bars, and

$\Gamma^{green}$ containing the green-coloured bars

Both images are included in the local region search above.

The centre point of every bar is stored for the purpose of calculating the position of fixations with respect to bars (see chapter 6 and appendices 3 and 4, for these statistics). This means that fixation position is able to be compared to the centre point of the bar in a similar manner to that described by Motter and Belky (1998a,b).

Finally, these are combined to form $\Gamma^{grey}$, which contains the greyscale information for both sets of bars.

Figure A1.1 shows an example image, of size 1000x1000 pixels, with a red horizontal bar as the target, 150 green horizontal and 150 red vertical distractors.

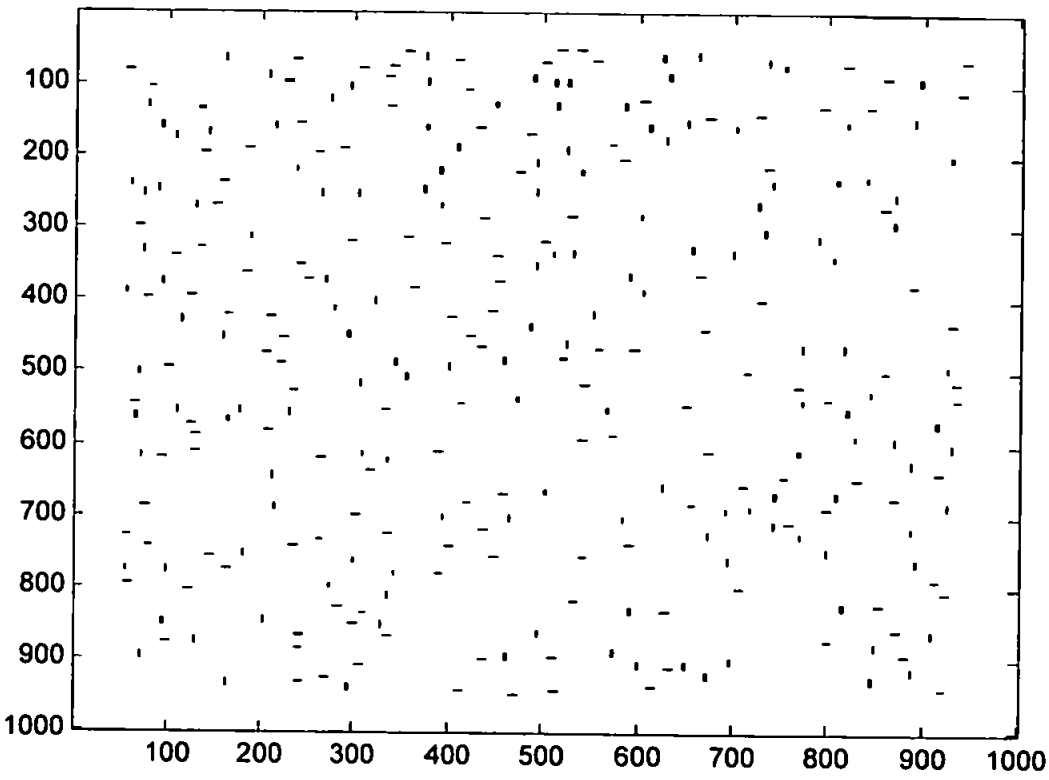



**Figure A1.1 Example Image Created by the System**

## A1.2 Active Vision Coordinate Systems

Three co-ordinate systems operate within the model:

- The original image/scene
- The retinal image
- The retinotopic arrangement of cell assemblies in V4 and LIP. These cell assemblies in combination have receptive fields that cover the entire retinal image. However, there are fewer assemblies than positions in the retinal image

due to convergent inputs. If these receptive fields overlapped more than at present (currently an overlap of one V1 neuron), this would increase the number of assemblies in V4 and LIP. This would allow greater spatial precision but would substantially increase the processing time at every fixation.

The retina and V1 share the original image's coordinate system because there is a cell for every pixel position. Convergence of inputs over the V1 receptive field could have been used to reduce the number of cells in V1 and, thus, reduce the amount of processing performed in this module. However, as this is done as pre-processing, processing time is not an issue and keeping the same number of V1 cells as pixels in the original image limits the number of co-ordinate systems to the three above, rather than introducing another for V1.

### A1.2.1 Fixation Positioning

Initially the fixation point is placed at centre of the original image and a retinal image is formed around this area. Subsequent fixation points are determined by the most active LIP cell assembly and are assumed to be central to the receptive field of the most active assembly when considered in terms of the retinal or original image co-ordinates. Hence, subsequent fixation points are transformed from cortical co-ordinates to retinal image and original image co-ordinates.

The most active cell assembly in LIP is given by:

$CorticalMax^{ij}$ = find (max(Y))

Where:

Y is the activity in LIP, described in chapter 4.

The find command finds the location in LIP that contains the stated activity

### A1.2.2 Converting cortical co-ordinates to retinal co-ordinates

The cortical maximum position is converted to a location in the retinal image first by finding the starting point of the cell assembly's receptive field and then taking the mid-position in the cell's receptive field. The extent of a cell assembly's receptive field is also required when establishing the V1 inputs to higher cortical areas.

The extent of a cell's receptive field in original image (and V1) co-ordinates is given by:

$RFstart^{i} = ((cellAssembly^{i} - 1) \times (RFsize - RFoverlap)) + 1$

$RFstart^{j} = ((cellAssembly^{j} - 1) \times (RFsize - RFoverlap)) + 1$

Where:

$cellAssembly^{ij}$ is the position of the cell assembly in the cortical area. This is equal to $CorticalMax^{ij}$ when the procedure is used to establish the next fixation position from the LIP cell assemblies

RFsize is the spread of the receptive field in V1 coordinates (normally set to 23)

RFoverlap is the amount of overlap of receptive fields of adjacent cell assemblies, in V1 coordinates (normally set to 1)

In order to find the next fixation point, the centre of the cell assembly's receptive field is found by:

$$retFix^{i} = \min(RFstart^{i} + \text{round}(RFsize / 2) - 1, m)$$

$$retFix^{j} = \min(RFstart^{j} + \text{round}(RFsize / 2) - 1, n)$$

where m and n are the dimensions of the retinal image, given below

The general transformation from retinal to original image co-ordinates is also given below.

### A1.2.3 Forming the Retinal Image

For implementation simplicity, the retinal image and AW are implemented as square or rectangular matrices. However, terminology relating to circles ("diameter" and "radius") is used to reflect the more circular biological counterpart.

The current fixation point in the original image is given by $[fix^i, fix^j]$

The size of the original image is given by M, N.

The radius of the retinal image is given by retRad so that the extent of the retinal image is (retRad * 2) + 1

Therefore, the dimensions of the retinal image (m,n) are set as follows:

m = (retRad * 2) +1

n = (retRad * 2) +1

and the fixation point in retinal co-ordinates becomes

[retRad + 1 , retRad + 1], i.e. the centre of the retinal image

The retinal image projected onto the original image extends from [imStartI , imStartJ] to the point [imEndI , imEndJ]

Where:

$imStartI = \max ( fix^i - retRad, 1 )$

$imEndI = \min ( fix^i + retRad, M )$

$imStartJ = \max ( fix^j - retRad, 1 )$

imEndJ = min ( $fix^j$ + retRad, N )

and M, N are the dimensions of the original image

In cases where this results in the retinal image extending past the limits of the original image, the retinal/V1 inputs to the higher cortical areas are restricted such that no bottom-up inputs are available in the excess area.

The retinal/V1 inputs to the higher cortical areas are formed from $I_{ijrk+c}$, for all orientation and colour features, within the retinal image as follows:

Ret(dataAreaStartI:dataAreaEndI,dataAreaStartJ:dataAreaEndJ,R,K+C) =
I(imStartI:imEndI,imStartJ:imEndJ,R,K+C);

Where:

The left-hand expression is given in retinal image co-ordinates and the right-hand expression is given in original image/V1 co=ordinates.

R is the maximum number of spatial frequencies

K+C is the maximum number of features (K orientations plus C colours)

The dataArea expressions deal with the situation where the retinal image "overflows" the original image (and, hence, V1) data. In this case, information is only passed up to the dynamic cortical areas where it is

available from the original image (the dataArea). The remainder of Ret is set to zeros. In this case the dataArea dimensions are set as follows:

dataAreaStartI = m - (imEndI - imStartI);

dataAreaEndI = dataAreaStartI + (imEndI - imStartI);

dataAreaStartJ = n - (imEndJ - imStartJ);

dataAreaEndJ = dataAreaStartJ + (imEndJ - imStartJ);

The same method is used to extract values from the novelty map relating to the current retinal image. This region is then available to LIP so that each assembly can access the novelty values over the area of its receptive field only.

### A1.2.4 Scaling the Cortical Areas According to the Size of the Retinal Image

The number of cell assemblies in areas V4 and LIP is set according to the size of the retinal image. For each area the number of assemblies is set as follows:

$neurons^i$ = floor ( (m-RFsize) / (RFsize – RF overlap) ) + 1

$neurons^j$ = floor ( (n-RFsize) / (RFsize – RF overlap) ) + 1

Where:

$neurons^i$ is the number of rows of assemblies and

$neurons^j$ is the number of columns of assemblies

### A1.2.5 Establishing the Extent of the AW in the retinal image

The region of the retinal image relating to the AW is given by:

Ret (retAWstartI: retAWendI, retAWstartJ: retAWendJ)

where the limits are given in terms of retinal image co-ordinates by:

$$retAWstartI = \max ( retFix^{i} - AWrad , 1 )$$

$$retAWendI = \min ( retFix^{i} + AWrad , m )$$

$$retAWstartJ = \max ( retFix^{j} - AWrad , 1 )$$

$$retAWendJ = \min ( retFix^{j} + AWrad , n )$$

AWrad being the radius of the AW, calculated at each fixation, as described in chapter 4.

### A1.2.6 Retinal Image to Original Image Co-Ordinate Transformation

To convert retinal image co-ordinates ($retIm^{i}$,$retIm^{j}$) to original image co-ordinates ($Im^{i}$,$Im^{j}$) when the current fixation point in the retinal image is ($retFix^{i}$,$retFix^{j}$) and in the original image is ($fix^{i}$, $fix^{j}$):

$$Im^{i} = fix^{i} - retFix^{i} + retIm^{i}$$

$$Im^{j} = fix^{j} - retFix^{j} + retIm^{j}$$

As the size of the retinal image is fixed and may overspill the original image, the following adjustment is required in order to prevent reference beyond the extent of the original image:

$Im^i = \min( \max ( fix^i - retFix^i + retIm^i , 1 ) , M )$

$Im^j = \min( \max ( fix^j - retFix^j + retIm^j , 1 ) , N )$

### A1.2.7 Forming the cortical AW - Retinal Image to Cortical Assemblies Co-Ordinate Transformations

LIP cell assemblies in the following area will receive a spatial bias due to the AW:

Y(corticalAWstartI : corticalAWendI , corticalAWstartJ : corticalAWendJ)

Where:

The cell population whose receptive field is the first to end after the start of the AW in retinal image co-ordinates is given by:

$$corticalAWstart^i = \varpi\left[\frac{retAWstartI - RFsize}{RFsize - RFoverlap} + 1\right]$$

$$\text{corticalAWstart}^j = \varpi\left[\frac{retAWstartJ - RFsize}{RFsize - RFoverlap} + 1\right]$$

And, the cell population whose receptive field is the last to start before the end of the AW in retinal image co-ordinates is given by:

$$\text{corticalAWend}^i = \xi\left[\frac{retAWendI - 1}{RFsize - RFoverlap} + 1\right]$$

$$\text{corticalAWend}^j = \zeta\left[\frac{retAWendJ - 1}{RFsize - RFoverlap} + 1\right]$$

Where:

$\varpi$ = max ( ceil (x), 1 )

$\xi$ = min ( floor (x) , $\text{neurons}^i$ )

$\zeta$ = min ( floor (x) , $\text{neurons}^j$ )

and $\text{neurons}^i$ is the number of rows of cell assemblies in the cortical area; $\text{neurons}^j$ is the number of columns of cell assemblies in the cortical area

# Appendix A2

# Supplementary Information

## A2.1 System Performance – Run Time Speed

Neuronal assemblies in V4, LIP and IT are modelled using differential equations, as described in chapter 4. Each of these is implemented as a difference equation with discrete time steps. The system is relatively insensitive to the size of time step used but is normally run in steps of 1ms for single cell simulations and 5ms for the scan path simulations (in order to increase the speed of the simulation). Running in Matlab v6.5 using a Pentium(4) 3.06GHz processor with 2GB of RAM, the system takes less than 0.5 second to process a 240ms fixation at 1ms steps for a small retinal image (covering 23x23pixels) used for monochromatic single cell simulations. For chromatic scan path simulations, the system normally requires approximately 13 seconds to process an average fixation lasting 240ms, in 5 ms steps (normal image size is 1000x1000 pixels and there are 20 x 20 V4 assemblies in each feature layer for a retinal image size of 441x441 pixels). Time can be reduced by limiting the size of the retinal image or increasing the step size. Also, the model would be significantly faster if it were programmed in a third-generation language, such as C, rather than Matlab.

## A2.2 Compensation for Saccades

The issue of how biological systems cope with large-field motion and change in the co-ordinate system induced by a saccade was introduced in chapter 4. This appendix provides more information about the suppression of inputs during a saccade and how perceptual constancy is achieved.

### A2.2.1 Saccadic Suppression

It is unclear what information is available within cortex during a saccade but there appears to be a suppression of visual inputs during the eye movement in order to maintain perception of a stable visual image. Typically this saccadic suppression precedes the eye movement by 30-60ms, lasts 120-180ms, and is followed by a 100-150ms facilitation (see Zhu & Lo, 1996, for review). Early theories about the suppression suggested that saccades were accompanied by a 'collorary discharge' or 'efference copy' of the motor signal and that this was used to cancel the associated image motion. In its earliest form, the idea became less plausible with the realisation that motion signals are processed in specialised brain areas (e.g. areas MT and MST), which are separate to those for form perception, and could not be simply annulled by a contrary displacement signal. However, the idea of the efference copy is still investigated as an important extraretinal source of information for perceptual localisation, being supplemented by other mechanisms to achieve space constancy (for example, see Bridgeman, 1995). Investigation of contrast sensitivity for gratings flashed briefly during saccades revealed that low

spatial frequencies (that predominate natural scenes) become suppressed during a saccade, whereas high spatial frequencies do not seem to need to be suppressed. This is probably because, when the retina is moving, it lacks the temporal resolution to detect these high frequencies (Bridgeman, van der Heijden & Velichkovsky, 1996). The magnocellular pathway, which operates at low spatial frequency, appears to be suppressed during the saccade (Anand & Bridgeman, 2002; Thiele, Henning, Kubischik & Hoffmann, 2002; and see Ross, Morrone, Goldberg & Burr, 2001, and Zhu & Lo, 1996, for reviews) and this may temporarily dampen the sensation of motion sufficient to ignore the saccade-related motion. Furthermore, neurons in motion areas MT and MST are inhibited during saccades and, in some cases, direction sensitivity may be altered (Theile, Henning, Kubischik & Hoffman, 2002). Regional cerebral blood-flow is decreased in striate (V1), extrastriate (V2, V4 etc.) and parietal cortex during a saccade, indicating saccadic suppression of activity in these regions (Paus, Marrett, Worsley and Evans, 1995). Direct stimulation of the deep layers of the superior colliculus in rabbits has identified a possible inhibitory circuit for saccadic suppression that extends from the deep layers of the superior colliculus to the central lateral nucleus of the thalamus, on to the thalamic reticular nucleus and then the LGN (Zhu & Lo, 1995; 1996; 1998). Eye movements also affect responses in the LGN in monkeys (Reppas, Usrey & Reid, 2002; Ramcharan, Gnadt & Sherman, 2001) and cats (Lee & Malpeli, 1998). Such thalamic changes may act as a trigger to adjust responses in cortex.

It is worth noting a further issue about neural changes relating to saccades. In addition to the saccades described throughout this thesis, unconscious microsaccadic eye movements (of $\leq 2^{\circ}$), together with slow drifts, take place around the fixation point. Such microsaccades are not modelled here but they do have an impact on thalamic and cortical responses, causing bursts of firing in the LGN (Martinez-Conde, Macknik & Hubel, 2002) and V1 (Martinez-Conde, Macknik & Hubel, 2000). The importance of microsaccades in sustaining perception is demonstrated by stabilising an image on the retina (i.e. compensating for the microsaccades with image motion). In this circumstance, the image appears to fade. The neural correlates of microsaccadic suppression are controversial and may be different to those for larger saccades (see Martinez-Conde et al., 2002; and Ross et al, 2001, for discussions).

Clearly, saccadic suppression is a complex area that is not yet fully understood. Modelling of intra-saccadic processing here is simplified due to the lack of a full representation of the magnocellular pathway. In the model, saccades are instantaneous and this avoids the issue of needing to eliminate the motion sensed during the eye movement. The dynamic cortical areas (IT, LIP and V4) are reset to baseline random values at the start of the subsequent fixation to reflect saccadic suppression of magnocellular, and possibly parvocellular inputs, in addition to the cortical dynamics during the saccade. The issue of pre- and post-saccadic spatial constancy is discussed further next and has links with the discussion about inhibition of return in chapter 4, in the context of a spatial memory for locations already inspected in the scene.

### A.2.2.2Spatial Constancy

Saccades occur, on average, three times a second. Despite these shifts in retinal image, we perceive the world as stable, whereas comparable external image motion produces a sense of instability (see Ross et al, 2001, for review). Visual information immediately after the saccade appears to be important in recalibrating the system. Although sensitivity to object displacement is weak during the saccade, if the same object is blanked following its displacement during the saccade and then reappears some times after the end of the saccade, detection of a displacement is unimpaired (Deubel, Schneider & Bridgeman, 1996). If the object remains visible for even a short time after the saccade, the perceptual benefit is lost. The displacement detection improves with longer periods of blanking (up to about 300ms) but most of the benefit occurs within the first 80ms after saccade end. In addition, if only one of two objects is blanked, it is more often perceived as having moved regardless of which of the two objects actually moved (Deubel, Bridgeman & Schneider, 1998). Therefore, continuously visible objects are preferentially perceived as being stable and may act as a spatial reference for the post-saccadic re-mapping of visual space.

Around the time of a saccade, representation of the visual world appears to be subject to some compression. The perceived position of a target flashed before, during and after a saccade is not its veridical position but changes over a 200ms period, starting less than 100ms before the eyes start to move and reaching maximal

effect about the time of the onset of the saccade (see Ross et al, 2001, for review). Before a saccade, objects appear to be displaced in the direction of the saccade, suggesting an anticipatory shift, which may be slower than the saccade itself because there is a rebound effect once the eyes start to move. However, such errors in localisation are not simply a compensation for eye movement because they do not always occur in the direction of the saccade. Ross, Morrone and Burr (1997) showed that the size and sign of such errors depend on the position of the target in the visual field. Rather than being simply displaced in the direction of the saccade, objects tend to converge towards the saccadic target. This results in a "compression" of the visual world that distorts scenes immediately before the start of a saccade, and also shifts it in the same direction and with the same amplitude as the saccade. The timecourse of perceptual compression is similar to that for saccadic suppression, beginning more than 50ms before the start of the saccade (see Ross et al., 2001).

Following a saccade, retinotopic representations in LIP shift into the new co-ordinate system based on the post-saccadic centre of gaze and this is replicated in the model here. However, as suggested above, the changes in cortical representation around the time of a saccade are extremely complex. Just prior to the saccade the spatial properties of receptive fields in LIP change (Ben Hamed, Duhamel, Bremmer & Graf, 1996) and many neurons respond (~80ms) before the saccade to salient stimuli that will enter their receptive fields after the saccade (Duhamel, Colby & Goldberg, 1992). The response of these "predictive" cells may be interpreted in areas such as superior colliculus and frontal eye field, along with

that from neurons that respond only to what is currently in their receptive fields. This combined response could contribute to the compression of visual space allowing a smooth shift of coordinates (Ross et al., 2001). Neurons in V4 (Moore, 1999; Tolias, Moore, Smirnakis, Tehovnik, Siapas & Schiller, 2001) and IT (Sheinberg & Logothetis, 2001) also exhibit complex changes in receptive field profiles immediately prior to the onset of a saccade, and begin to respond to objects that are to be the target of the forthcoming saccade. This may further facilitate the integration of perceptual information across saccades. Such "predictive re-mapping" (Duhamel et al., 1992) of the visual scene is not modelled here and assemblies respond only to what is currently in their receptive fields. These complex issues of pre- and post-saccadic spatial constancy and perceptual stability are not fully addressed because they may rely on further modelling of the parietal cortex. Other parietal areas, such as the ventral intraparietal area (VIP), encode the location of objects relative to the head or body, invariant to eye position (Bremmer, Schlack, Duhamel, Graf & Fink, 2001; Duhamel, Bremmer, BenHamed & Graf, 1997). Coordinate transformations between different spatial representations in the parietal cortex may be necessary to maintain a stable perception of the world. The importance of parietal cortex in such transformations is highlighted by lesion studies, for example, some patients with right hemisphere parietal cortex damage are unable to compensate for the change in retinal position (following leftward saccades) during a double-step saccade task[1] to a target (Duhamel, Goldberg, FitzGibbon, Sirigu & Grafman, 1992).

[1] In such a double-step task, two targets are briefly presented but are extinguished prior to the eye movements so that programming of the saccade to the second target must take into account the change in eye position resulting from the saccade to the first target.

Some neural network models have explored the issue of saccadic re-mapping (Andersen & Zipser, 1988; Mazzoni, Andersen & Jordan, 1991; Quaia, Optican & Goldberg, 1998; Zipser & Andersen, 1988). For example, Andersen and Zipser (1988) used the position of the eye in its orbit to modulate retinotopic responses but Quaia et al. (1998) provide a detailed model of LIP and LIP/FEF interaction, which accounts for target re-mapping using only retinal coordinates across saccades. Research is now shedding light on the nature of transsaccadic memory, with recent evidence suggesting that the shape of an entire object can be transferred across the saccade and that the memory is selective for attended objects (Khayat, Spekreijse & Roelfsema, 2004). Such investigations will provide inspiration for future computational modelling in this area.

Here, the issues of spatial re-mapping, constancy and perceptual stability are left to specialised models that could investigate changes in representation around the time of a saccade (Ross et al., 1997), memory for targets across saccades (Findlay et al., 2001; McPeek, Skavenski & Nakayama, 2000), the possible use of visual markers for co-ordinate transform (Deubel et al., 1998), along with the associated issues of suppression of magnocellular (Anand & Bridgeman, 2002; Thiele et al., 2002; and see Ross et al., 2001, and Zhu & Lo, 1996, for reviews) and possibly parvocellular cortical inputs just prior to and during a saccade.

# Appendix A3

## Scan Path Statistics

## Appendix A3.1 - Effect of Varying V4 to LIP Connection Weights

Figures averaged over 10 scan paths, of 50 fixations each, over image 1000rh150

| Workspace Name | Weight of Colour as % Orient. Weight | % Fixations Near Target Colour Stimuli | % Fixations Near Non-Target Colour Stimuli | % Fixations in Blank Areas | Average Saccade Amplitude |
|---|---|---|---|---|---|
| 1000rh150-50fixesRetRad220V4toLIP2pt4-10scan | 100.00 | 38.16326531 | 61.02040816 | 0.816326531 | 7.795931964 |
| 1000rh150-50fixesRetRad220LIPo2pt39c2pt4-10scan | 100.78 | 44.48979592 | 54.69387755 | 0.816326531 | 7.603046239 |
| 1000rh150-50fixesRetRad220LIPo2pt38c2pt42-10scan | 101.56 | 47.55102041 | 50.81632653 | 1.632653061 | 7.687832075 |
| 1000rh150-50fixesRetRad220LIPo2pt36c2pt44-10scan | 103.13 | 56.12244898 | 43.26530612 | 0.612244898 | 7.626742035 |
| 1000rh150-50fixesRetRad220LIPo2pt33c2pt47-10scan | 106.25 | 62.44897959 | 36.32653061 | 1.224489796 | 7.673429146 |
| 1000rh150-50fixesRetRad220LIPo2pt26c2pt52-10scan | 112.50 | 77.55102041 | 21.63265306 | 0.816326531 | 7.650844104 |
| 1000rh150-50fixesRetRad220LIPo2pt13c2pt67-10scan | 125.00 | 84.08163265 | 14.89795918 | 1.020408163 | 7.765283472 |
| 1000rh150-50fixesRetRad220LIPo1pt92c2pt88-10scan | 150.00 | 85.91836735 | 12.85714286 | 1.224489796 | 7.419689766 |
| 1000rh150-50fixesRetRad220LIPo1pt6c3pt2-10scan | 200.00 | 87.14285714 | 11.63265306 | 1.224489796 | 7.509915805 |
| 1000rh150-50fixesRetRad220LIPo1pt2c3pt6-10scan | 300.00 | 87.75510204 | 9.387755102 | 2.857142857 | 7.604535321 |
| 1000rh150-50fixesRetRad220LIPo01p96c3pt84-10scan | 400.00 | 91.63265306 | 5.102040816 | 3.265306122 | 7.295288876 |
| 1000rh150-50fixesRetRad220LIPo0pt8c4-10scan | 500.00 | 93.26530612 | 1.632653061 | 5.102040816 | 7.241178834 |
| 1000rh150-50fixesRetRad220LIPo0pt6857c4pt1142-10scan | 600.00 | 93.46938776 | 2.244897959 | 4.285714286 | 7.024286954 |
| 1000rh150-50fixesRetRad220LIPo0pt6c4pt2-10scan | 700.00 | 95.51020408 | 0.612244898 | 3.87755102 | 7.00107098 |
| 1000rh150-50fixesRetRad220LIPo0pt5333c4pt2664-10scan | 800.00 | 91.83673469 | 2.857142857 | 5.306122449 | 7.10024876 |
| 1000rh150-50fixesRetRad220LIPo0pt48c4pt32-10scan | 900.00 | 92.85714286 | 1.632653061 | 5.510204082 | 7.231113556 |
| 1000rh150-50fixesRetRad220LIPo0pt43636c4pt3636-10scan | 1000.00 | 93.26530612 | 1.632653061 | 5.102040816 | 7.178617169 |

Normal parameter values shown in orange

## A3.2 Supplementary Information – V4 to LIP Connection Weights

When the V4 feature connections to LIP are equally weighted, there is a priority for target orientation, rather than target colour, to capture attention. This is due to the combined effect of the transfer function in converting V4 assembly activity into discharge and being used to normalise V1 convergent input into V4. An example best demonstrates the effect.

The retinal image and AW for the first (central) fixation on image 1000rh150, which has a red horizontal target, are shown in figure A3.1. Position (599,533) is chosen as the next fixation point. This location is close to a green stimulus, i.e. a non-target colour distractor. This location is chosen because the LIP assembly at (15,12) is most active at the time of the saccade. All locations in this appendix are given in matrix coordinates.

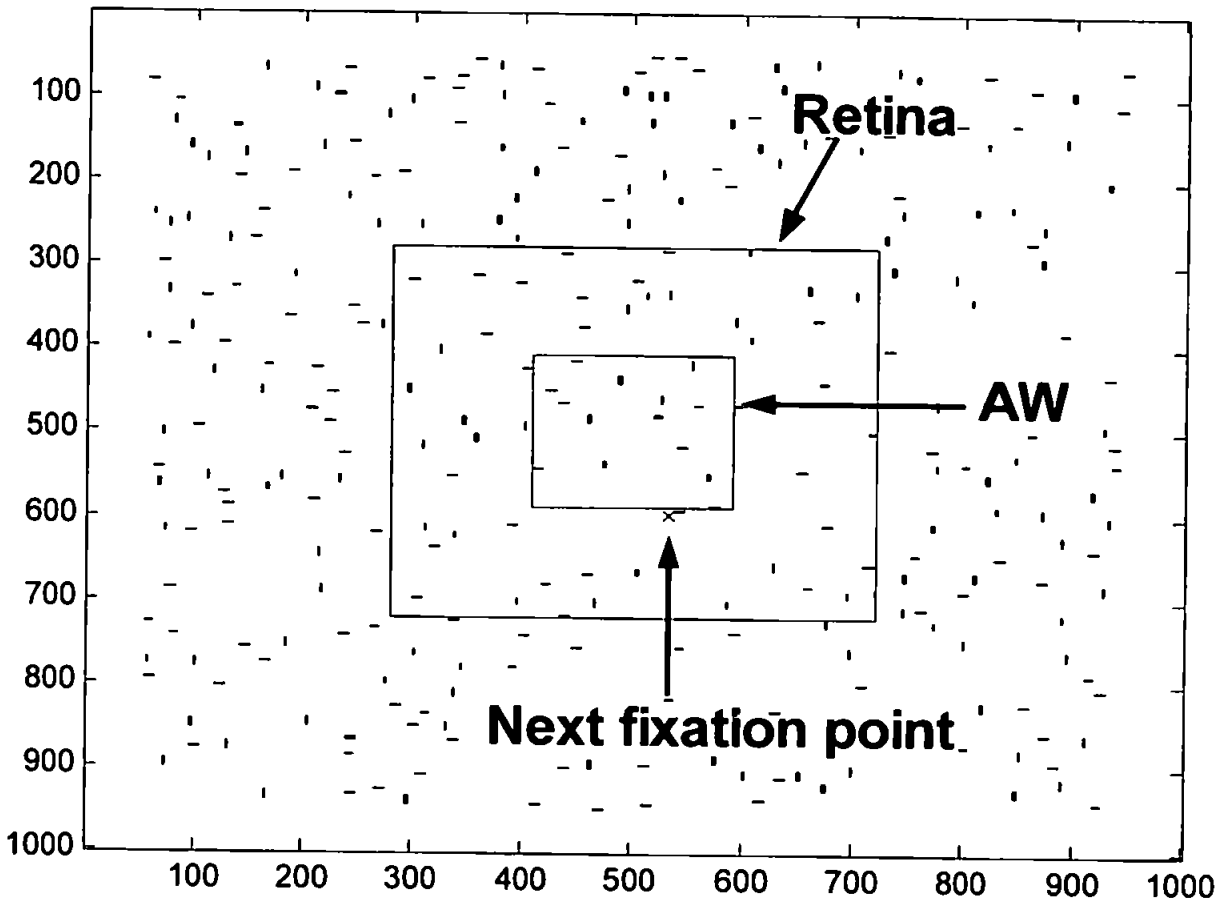



**Figure A3.1 Original Image With Retina and AW**

The next fixation point, near a green stimulus, is shown as a magenta cross.

Figure A3.2 shows the activity in the dynamic modules at the time of the saccade. The target features (red and horizontal) appear to be more highly active than non-target features (green and vertical).

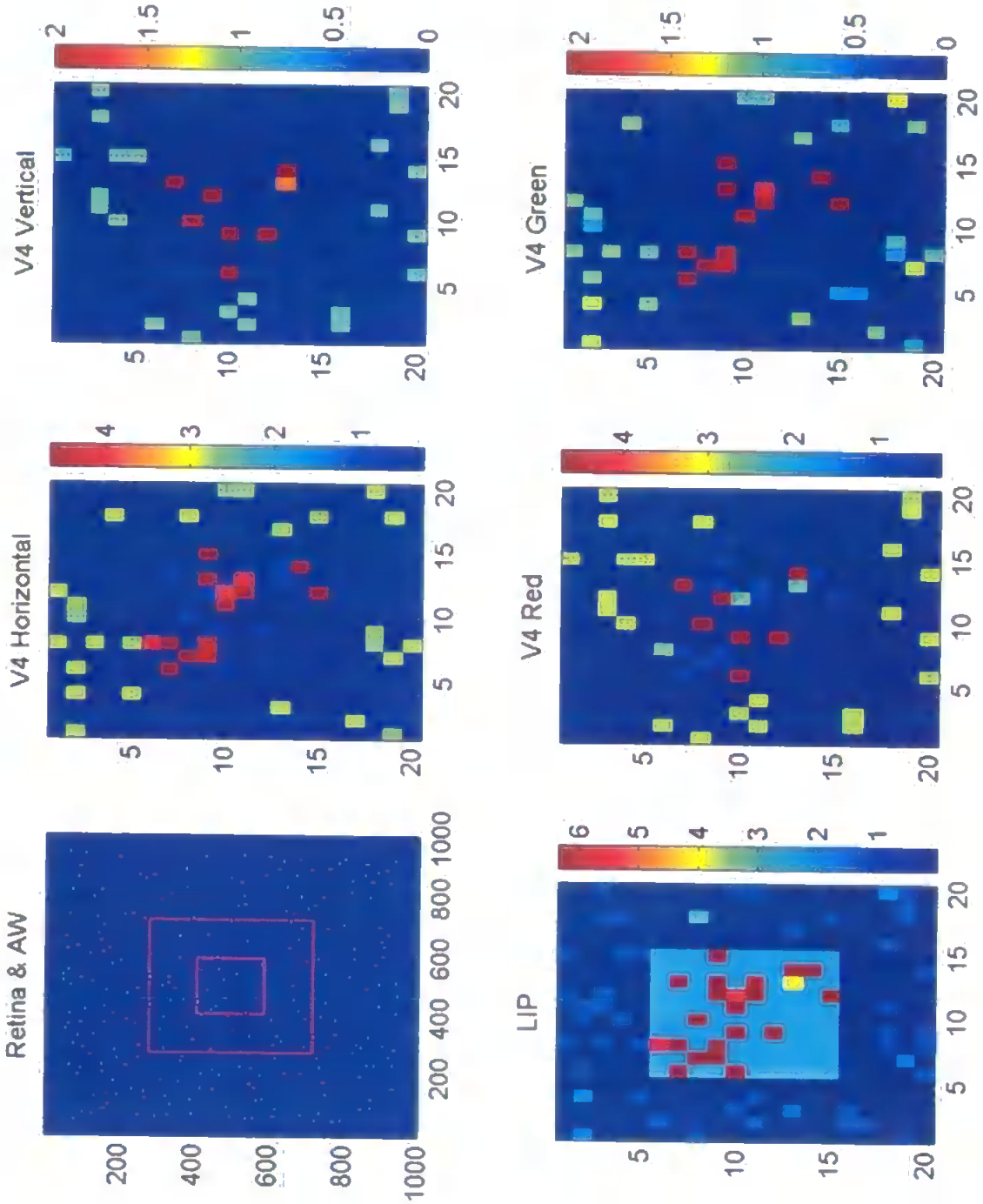



Figure A3.2 Cortical Activity at the Time of the Saccade

The activity in V4 at the chosen location (15,12) at the time of the saccade was compared to that at red stimuli locations (7,13) and (10,9) to ascertain why a green stimulus location was chosen as the next fixation location. Table A3.1 shows V4 activity at these locations. All results are rounded to 4 decimal places.

| | Cortical location (15,12)-green horizontal | Cortical location (7,13)-red vertical | Cortical location (10,9)-red vertical |
|---|---|---|---|
| Activity in V4 horizontal | 4.6239 | 0.7662 | 0.7633 |
| Activity in V4 vertical | 0 | 1.9880 | 2.0030 |
| Activity in V4 red | 0.7708 | 4.7211 | 4.7198 |
| Activity in V4 green | 2.0673 | 0 | 0 |
| Total activity | 7.4621 | 7.4753 | 7.4861 |

**Table A3.1 Activity in V4 at the Time of the Saccade**

In addition to the effects of detecting bottom-up stimulus information at each location, the V4 activity is modulated by the top-down object-based feedback from IT, which tends to increase activity in red and horizontal assemblies and decrease activity in green and vertical assemblies. The red (target colour) locations are marginally more active than the green location but the difference is small.

LIP receives activity from V4 after it has been passed through the transfer function, as described in chapter 4. Table A3.2 gives the output (discharge) from the transfer function for the activities given in table A3.1.

| | Cortical location (15,12)-green horizontal | Cortical location (7,13)-red vertical | Cortical location (10,9)-red vertical |
|---|---|---|---|
| Discharge - V4 horizontal | 0.7689 | 0 | 0 |
| Discharge - V4 vertical | 0 | 0.5882 | 0.5908 |
| Discharge - V4 red | 0 | 0.7730 | 0.7729 |
| Discharge - V4 green | 0.5984 | 0 | 0 |
| Total discharge | 1.3673 | 1.3612 | 1.3637 |

**Table A3.2 Transfer Function Output for Activity in V4 at the Time of the Saccade**

Although, the red locations in V4 were more active than the green locations, passing the activity through the transfer function and summing the discharges has resulted in the green location now producing a greater input to LIP. This is because the discharge from the V4 green assembly at (15,12) is larger than that from the V4 vertical assemblies at either (7,13) or (10,9). Thus, the reason why the green location wins the competition in LIP is that the green activity and discharge in V4 at (15,12) is slightly higher than the vertical discharge at (7,13) and (10,9). This is due to the fact that the V1 input to V4 colour assemblies over the V4 receptive field is slightly larger than the V1 input to V4 orientation assemblies. This is because of

the response function thresholds chosen to normalise the different types of V1 filter outputs.

The inputs from V1 to V4 (the convergent output from V1 having been passed through the response function) at these locations are shown in tables A3.3 to A3.5.

| | |
|---|---|
| Horizontal | 0.9487 |
| Vertical | 0 |
| Red | 0 |
| Green | 0.9714 |

**Table A3.3 V1 input at V4 location (15,12)**

| | |
|---|---|
| Horizontal | 0 |
| Vertical | 0.9541 |
| Red | 0.9751 |
| Green | 0 |

**Table A3.4 V1 input at V4 location (7,13)**

| | |
|---|---|
| Horizontal | 0 |
| Vertical | 0.9541 |
| Red | 0.9744 |
| Green | 0 |

**Table A3.5 V1 input at V4 location (10,9)**

Thus, this effect is a computational artefact of the system due mainly to the normalisation of V1 inputs over the V4 receptive field that results in marginally higher inputs from V1 colour-selective cells than orientation-selective cells. This difference results from the necessarily different filters used for colour and orientation detection in V1. If V1 inputs to V4 were consistently identical in magnitude, the scan path would be expected to have no preference for target coloured and non-target coloured locations when V4 inputs to LIP were equally weighted.

## Appendix A3.3 - Effect of Varying Weight of IT Feedback to V4

Figures averaged over 10 scan paths, of 50 fixations each, over image 1000rh150

| Workspace Name | Feedback weight | % Fixations Near Target Colour Stimuli | % Fixations Near Non-Target Colour Stimuli | % Fixations in Blank Areas | Average Saccade Amplitude | Average Fixation Duration (max 350ms) |
|---|---|---|---|---|---|---|
| 1000rh150-50fixesRetRad220ITfb0-10scan | 0 | 47.75510204 | 50.6122449 | 1.632653061 | 7.684317017 | 257.80 |
| 1000rh150-50fixesRetRad220ITfb0pt001-10scan | 0.001 | 48.16326531 | 50.6122449 | 1.224489796 | 7.684810956 | 257.94 |
| 1000rh150-50fixesRetRad220ITfb0pt01-10scan | 0.01 | 49.3877551 | 48.7755102 | 1.836734694 | 7.861759768 | 257.94 |
| 1000rh150-50fixesRetRad220ITfb0pt05-10scan | 0.05 | 53.67346939 | 45.51020408 | 0.816326531 | 7.732871774 | 257.53 |
| 1000rh150-50fixesRetRad220ITfb0pt1-10scan | 0.1 | 63.06122449 | 35.51020408 | 1.428571429 | 7.749112311 | 256.84 |
| 1000rh150-50fixesRetRad220ITfb0pt2-10scan | 0.2 | 71.02040816 | 27.55102041 | 1.428571429 | 7.811793992 | 256.24 |
| 1000rh150-50fixesRetRad220ITfb0pt4-10scan | 0.4 | 75.71428571 | 21.83673469 | 2.448979592 | 7.531005883 | 253.94 |
| 1000rh150-50fixesRetRad220ITfb0pt6-10scan | 0.6 | 82.85714286 | 15.10204082 | 2.040816327 | 7.664460153 | 252.71 |
| 1000rh150-50fixesRetRad220ITfb0pt8-10scan | 0.8 | 81.42857143 | 15.30612245 | 3.265306122 | 7.480105082 | 251.23 |
| 1000rh150-50fixesRetRad220ITfb1-10scan | 1 | 83.46938776 | 13.46938776 | 3.06122449 | 7.461476742 | 249.05 |
| 1000rh150-50fixesRetRad220ITfb2-10scan | 2 | 85.71428571 | 10.6122449 | 3.673469388 | 7.567808226 | 243.41 |
| 1000rh150-50fixesRetRad220ITfb2pt5-10scan | 2.5 | 83.87755102 | 10.6122449 | 5.510204082 | 7.493609436 | 242.22 |
| 1000rh150-50fixesRetRad220ITfb3-10scan | 3 | 84.28571429 | 11.02040816 | 4.693877551 | 7.343795604 | 242.12 |
| 1000rh150-50fixesRetRad220ITfb4-10scan | 4 | 89.79591837 | 5.510204082 | 4.693877551 | 7.2743471 | 241.53 |
| 1000rh150-50fixesRetRad220ITfb5-10scan | 5 | 92.65306122 | 2.857142857 | 4.489795918 | 7.241732505 | 238.09 |
| 1000rh150-50fixesRetRad220ITfb6-10scan | 6 | 91.63265306 | 4.489795918 | 3.87755102 | 7.203961684 | 237.47 |
| 1000rh150-50fixesRetRad220ITfb7-10scan | 7 | 88.57142857 | 6.12244898 | 5.306122449 | 7.167778205 | 239.80 |
| 1000rh150-50fixesRetRad220ITfb8-10scan | 8 | 89.59183673 | 3.265306122 | 7.142857143 | 6.967975926 | 242.89 |
| 1000rh150-50fixesRetRad220ITfb9-10scan | 9 | 91.42857143 | 2.857142857 | 5.714285714 | 7.175675038 | 244.03 |
| 1000rh150-50fixesRetRad220ITfb10-10scan | 10 | 87.55102041 | 3.469387755 | 8.979591837 | 7.02691061 | 248.30 |
| 1000rh150-50fixesRetRad220ITfb11-10scan | 11 | 89.3877551 | 3.469387755 | 7.142857143 | 7.441139681 | 257.92 |
| 1000rh150-50fixesRetRad220ITfb12-10scan | 12 | 88.16326531 | 4.285714286 | 7.551020408 | 7.004929583 | 253.06 |

Normal parameter values shown in orange

## Appendix A3.4 - Effect of Varying Weight of Prefrontal Feedback to IT

Figures averaged over 10 scan paths, of 50 fixations each, over image 1000rh150

| Workspace Name | Feedback weight | % Fixations Near Target Colour Stimuli | % Fixations Near Non-Target Colour Stimuli | % Fixations in Blank Areas | Average Saccade Amplitude | Average Fixation Duration (max 350ms) |
|---|---|---|---|---|---|---|
| 1000rh150-50fixesRetRad220FixDurPF0-10scan | 0 | 50.20408163 | 48.57142857 | 1.224489796 | 7.671714004 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt001-10scan | 0.001 | 49.79591837 | 48.7755102 | 1.428571429 | 7.642090352 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt01-10scan | 0.01 | 52.44897959 | 45.91836735 | 1.632653061 | 7.56802096 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt05-10scan | 0.05 | 62.44897959 | 35.71428571 | 1.836734694 | 7.646538539 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt1-10scan | 0.1 | 73.87755102 | 23.87755102 | 2.244897959 | 7.290534972 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt2-10scan | 0.2 | 85.10204082 | 12.85714286 | 2.040816327 | 7.384596044 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt3-10scan | 0.3 | 86.53061224 | 10.20408163 | 3.265306122 | 7.206485149 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt4-10scan | 0.4 | 85.91836735 | 12.44897959 | 1.632653061 | 7.393628921 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt5-10scan | 0.5 | 71.42857143 | 27.34693878 | 1.224489796 | 7.542802301 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt6-10scan | 0.6 | 67.95918367 | 30.81632653 | 1.224489796 | 7.708296305 | 349.97 |
| 1000rh150-50fixesRetRad220FixDurPF0pt7-10scan | 0.7 | 87.14285714 | 10 | 2.857142857 | 7.515457781 | 345.89 |
| 1000rh150-50fixesRetRad220FixDurPF0pt8-10scan | 0.8 | 90.81632653 | 5.918367347 | 3.265306122 | 7.215101503 | 307.83 |
| 1000rh150-50fixesRetRad220FixDurPF0pt9-10scan | 0.9 | 88.57142857 | 5.306122449 | 6.12244898 | 7.175793116 | 264.84 |
| 1000rh150-50fixesRetRad220FixDurPF1-10scan | 1 | 91.02040816 | 3.06122449 | 5.918367347 | 7.273903415 | 248.29 |
| 1000rh150-50fixesRetRad220FixDurPF1pt1-10scan | 1.1 | 92.44897959 | 3.87755102 | 3.673469388 | 7.307176615 | 241.77 |
| 1000rh150-50fixesRetRad220FixDurPF1pt2-10scan | 1.2 | 92.65306122 | 3.06122449 | 4.285714286 | 7.258737804 | 238.16 |
| 1000rh150-50fixesRetRad220FixDurPF1pt4-10scan | 1.4 | 91.63265306 | 3.87755102 | 4.489795918 | 7.038522097 | 234.57 |
| 1000rh150-50fixesRetRad220FixDurPF2-10scan | 2 | 93.67346939 | 2.653061224 | 3.673469388 | 7.357461733 | 229.97 |
| 1000rh150-50fixesRetRad220FixDurPF3-10scan | 3 | 92.44897959 | 2.857142857 | 4.693877551 | 7.166745356 | 225.02 |
| 1000rh150-50fixesRetRad220FixDurPF4-10scan | 4 | 94.48979592 | 2.653061224 | 2.857142857 | 7.338723476 | 225.00 |
| 1000rh150-50fixesRetRad220FixDurPF5-10scan | 5 | 95.10204082 | 1.632653061 | 3.265306122 | 7.248030493 | 225.00 |
| 1000rh150-50fixesRetRad220FixDurPF6-10scan | 6 | 95.30612245 | 2.448979592 | 2.244897959 | 7.276873885 | 223.02 |
| 1000rh150-50fixesRetRad220FixDurPF7-10scan | 7 | 94.28571429 | 2.653061224 | 3.06122449 | 7.227680597 | 220.32 |
| 1000rh150-50fixesRetRad220FixDurPF8-10scan | 8 | 94.69387755 | 2.040816327 | 3.265306122 | 7.428864945 | 220.02 |
| 1000rh150-50fixesRetRad220FixDurPF9-10scan | 9 | 92.44897959 | 1.632653061 | 5.918367347 | 7.262053971 | 220.00 |

Normal parameter values shown in orange

## Appendix 3.4 - Effect of Varying Weight of Prefrontal Feedback to IT
(Continued)

**Re-Run of simulations for subset of feedback values, to confirm trends:**

| Workspace Name | Feedback weight | % Fixations Near Target Colour Stimuli | % Fixations Near Non-Target Colour Stimuli | % Fixations in Blank Areas | Average Saccade Amplitude | Average Fixation Duration (max 350ms) |
|---|---|---|---|---|---|---|
| 1000rh150-50fixesRetRad220FixDurPF0pt1c-10scan | 0.1 | 76.32653061 | 21.2244898 | 2.448979592 | 7.522952781 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt2c-10scan | 0.2 | 82.24489796 | 15.30612245 | 2.448979592 | 7.423739559 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt3c-10scan | 0.3 | 86.93877551 | 11.02040816 | 2.040816327 | 7.3615561 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt4c-10scan | 0.4 | 89.3877551 | 8.163265306 | 2.448979592 | 7.180862309 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt5b-10scan | 0.5 | 73.67346939 | 24.28571429 | 2.040816327 | 7.790698452 | 350 |
| 1000rh150-50fixesRetRad220FixDurPF0pt6b-10scan | 0.6 | 69.18367347 | 29.3877551 | 1.428571429 | 7.587623665 | 349.97 |
| 1000rh150-50fixesRetRad220FixDurPF0pt7b-10scan | 0.7 | 88.57142857 | 9.591836735 | 1.836734694 | 7.34456581 | 346.26 |

**Re-Run of simulations for subset of feedback values, with maximum fixation duration increased:**

| Workspace Name | Feedback weight | % Fixations Near Target Colour Stimuli | % Fixations Near Non-Target Colour Stimuli | % Fixations in Blank Areas | Average Saccade Amplitude | Average Fixation Duration (max 1000ms) |
|---|---|---|---|---|---|---|
| 1000rh150-50fixesRetRad220FixDur1000PF0pt05-10scan | 0.05 | 68.97959184 | 29.59183673 | 1.428571429 | 7.609854659 | 1000 |
| 1000rh150-50fixesRetRad220FixDur1000PF0pt4-10scan | 0.4 | 83.26530612 | 15.30612245 | 1.428571429 | 7.621883322 | 1000 |
| 1000rh150-50fixesRetRad220FixDur1000PF0pt5-10scan | 0.5 | 68.16326531 | 30.40816327 | 1.428571429 | 7.824837122 | 1000 |
| 1000rh150-50fixesRetRad220FixDur1000PF0pt6-10scan | 0.6 | 74.28571429 | 23.06122449 | 2.653061224 | 7.732951004 | 965.94 |
| 1000rh150-50fixesRetRad220FixDur1000PF0pt7-10scan | 0.7 | 91.42857143 | 6.326530612 | 2.244897959 | 7.159217488 | 658.08 |

## Appendix A3.5 - Effect of Varying Retinal Radius Over Dense Scene

Figures averaged over 10 scan paths, of 50 fixations each, over image 1000rh150

| Workspace Name | Retinal Radius (in pixels) | Retinal Radius (in degrees) | % Fixations Near Target Colour Stimuli | % Fixations Near Non-Target Colour Stimuli | % Fixations in Blank Areas | Average Saccade Amplitude | Average AW Radius (pixels) | Average AW Radius (degrees) |
|---|---|---|---|---|---|---|---|---|
| 1000rh150-50fixesRetRad11-10scan | 11 | 1 | 0 | 0 | 100 | 0 | 11 | 1 |
| 1000rh150-50fixesRetRad22-10scan | 22 | 2 | 95.91836735 | 4.081632653 | 0 | 1.41421356 | 22 | 2 |
| 1000rh150-50fixesRetRad55-10scan | 55 | 5 | 80.81632653 | 9.387755102 | 9.79591837 | 2.30732727 | 55 | 5 |
| 1000rh150-50fixesRetRad110-10scan | 110 | 10 | 92.04081633 | 2.448979592 | 5.51020408 | 7.41307316 | 93.892 | 8.54 |
| 1000rh150-50fixesRetRad220lTfb5-10scan | 220 | 20 | 92.65306122 | 2.857142857 | 4.48979592 | 7.2417325 | 91.66 | 8.33 |
| 1000rh150-50fixesRetRad330-10scan | 330 | 30 | 91.83673469 | 3.265306122 | 4.89795918 | 7.12418702 | 91.478 | 8.32 |
| 1000rh150-50fixesRetRad440-10scan | 440 | 40 | 93.67346939 | 2.448979592 | 3.87755102 | 7.32864745 | 93.174 | 8.47 |
| 1000rh150-50fixesRetRad499-10scan | 499 | 45.36 | 91.42857143 | 3.469387755 | 5.10204082 | 7.25748318 | 90.95 | 8.27 |

Pearson correlation coefficient for average saccade amplitude & average AW radius (all retinal radii shown) = 0.98090689

Pearson correlation coefficient for average saccade amplitude & average AW radius (retinal radii of >= 10 deg) = 0.81485867

## Appendix A3.6 - Effect of Varying Retinal Radius Over Sparse Scene

Figures averaged over 10 scan paths, of 50 fixations each, over image 1000gh50

| Workspace Name | Retinal Radius (in pixels) | Retinal Radius (in degrees) | % Fixations Near Target Colour Stimuli | % Fixations Near Non-Target Colour Stimuli | % Fixations in Blank Areas | Average Saccade Amplitude | Average AW Radius (pixels) | Average AW Radius (degrees) |
|---|---|---|---|---|---|---|---|---|
| 1000gh50-50fixesRetRad11-10scan | 11 | 1 | 0 | 0 | 100 | 0 | 11 | 1 |
| 1000gh50-50fixesRetRad22-10scan | 22 | 2 | 66.73469388 | 1.632653061 | 31.632653 | 1.414213562 | 22 | 2 |
| 1000gh50-50fixesRetRad55-10scan | 55 | 5 | 72.85714286 | 10.40816327 | 16.734694 | 1.459817431 | 55 | 5 |
| 1000gh50-50fixesRetRad110-10scan | 110 | 10 | 69.79591837 | 21.02040816 | 9.1836735 | 6.92179422 | 108.768 | 9.89 |
| 1000gh50-50fixesRetRad220Nov0pt0009-10scan | 220 | 20 | 86.93877551 | 11.02040816 | 2.0408163 | 11.66269134 | 167.698 | 15.25 |
| 1000gh50-50fixesRetRad330-10scan | 330 | 30 | 89.18367347 | 8.163265306 | 2.6530612 | 11.8698447 | 169.012 | 15.36 |
| 1000gh50-50fixesRetRad440-10scan | 440 | 40 | 89.18367347 | 8.571428571 | 2.244898 | 11.93360997 | 167.836 | 15.26 |
| 1000gh50-50fixesRetRad499-10scan | 499 | 45.36 | 85.10204082 | 7.346938776 | 7.5510204 | 11.68440375 | 166.724 | 15.16 |

Pearson correlation coefficient for Av saccade amplitude & Av AW radius (all retinal radii shown) = 0.9911917

Pearson correlation coefficient for Av saccade amplitude & Av AW radius (retinal radii of >=20 deg) = 0.6042034

## Appendix A3.7 - Effect of Varying Weight of Novelty Bias In a Dense Scene

Figures averaged over 10 scan paths, of 50 fixations each, over image 1000rh150

| Workspace Name | Weight of Novelty Bias | % Fixations Near Target Colour Stimuli | % Fixations Near Non-Target Colour Stimuli | % Fixations in Blank Areas | % Fixations in New Location | % Re-Fixations |
|---|---|---|---|---|---|---|
| 1000rh150-50fixesRetRad220Nov0-10scan | 0 | 98.7755102 | 1.224489796 | 0 | 28.8 | 71.2 |
| 1000rh150-50fixesRetRad220Nov0pt000009-10scan | 0.00000009 | 96.12244898 | 3.87755102 | 0 | 27.6 | 72.4 |
| 1000rh150-50fixesRetRad220Nov0pt000009-10scan | 0.0000009 | 98.97959184 | 1.020408163 | 0 | 30.6 | 69.4 |
| 1000rh150-50fixesRetRad220Nov0pt000009-10scan | 0.000009 | 98.97959184 | 1.020408163 | 0 | 29.2 | 70.8 |
| 1000rh150-50fixesRetRad220Nov0pt00009-10scan | 0.00009 | 97.34693878 | 2.040816327 | 0.612244898 | 96.2 | 3.8 |
| 1000rh150-50fixesRetRad220lTfb5-10scan | 0.0009 | 92.65306122 | 2.857142857 | 4.489795918 | 99.0 | 1.0 |
| 1000rh150-50fixesRetRad220Nov0pt009-10scan | 0.009 | 85.91836735 | 8.775510204 | 5.306122449 | 99.8 | 0.2 |
| 1000rh150-50fixesRetRad220Nov0pt05-10scan | 0.05 | 80.81632653 | 8.367346939 | 10.81632653 | 100.0 | 0.0 |
| 1000rh150-50fixesRetRad220Nov0pt09-10scan | 0.09 | 81.83673469 | 6.326530612 | 11.83673469 | 99.4 | 0.6 |
| 1000rh150-50fixesRetRad220Nov0pt5-10scan | 0.5 | 78.97959184 | 8.163265306 | 12.85714286 | 99.6 | 0.4 |
| 1000rh150-50fixesRetRad220Nov0pt9-10scan | 0.9 | 82.04081633 | 9.183673469 | 8.775510204 | 99.4 | 0.6 |
| 1000rh150-50fixesRetRad220Nov1-10scan | 1 | 77.34693878 | 10 | 12.65306122 | 99.6 | 0.4 |
| 1000rh150-50fixesRetRad220Nov2-10scan | 2 | 80.20408163 | 7.551020408 | 12.24489796 | 99.4 | 0.6 |
| 1000rh150-50fixesRetRad220Nov3-10scan | 3 | 81.02040816 | 10 | 8.979591837 | 99.6 | 0.4 |
| 1000rh150-50fixesRetRad220Nov4-10scan | 4 | 79.3877551 | 7.959183673 | 12.65306122 | 99.6 | 0.4 |
| 1000rh150-50fixesRetRad220Nov5-10scan | 5 | 82.44897959 | 6.326530612 | 11.2244898 | 99.6 | 0.4 |
| 1000rh150-50fixesRetRad220Nov6-10scan | 6 | 82.04081633 | 9.591836735 | 8.367346939 | 99.6 | 0.4 |
| 1000rh150-50fixesRetRad220Nov7-10scan | 7 | 78.57142857 | 9.795918367 | 11.63265306 | 100.0 | 0.0 |
| 1000rh150-50fixesRetRad220Nov8-10scan | 8 | 75.71428571 | 10 | 14.28571429 | 98.6 | 1.4 |
| 1000rh150-50fixesRetRad220Nov9-10scan | 9 | 79.79591837 | 7.755102041 | 12.44897959 | 100.0 | 0.0 |
| 1000rh150-50fixesRetRad220Nov10-10scan | 10 | 83.06122449 | 7.755102041 | 9.183673469 | 99.6 | 0.4 |
| 1000rh150-50fixesRetRad220Nov11-10scan | 11 | 84.28571429 | 7.346938776 | 8.367346939 | 99.6 | 0.4 |
| 1000rh150-50fixesRetRad220Nov12-10scan | 12 | 75.71428571 | 12.04081633 | 12.24489796 | 100.0 | 0.0 |

Normal parameter values shown in orange

## Appendix 3.7 - Effect of Varying Weight of Novelty Bias In a Dense Scene
(Continued)

| Workspace Name | Weight of Novelty Bias | Average Saccade Amplitude | Average AW Radius (pixels) | Average Fixation Duration |
|---|---|---|---|---|
| 1000rh150-50fixesRetRad220Nov0-10scan | 0 | 4.251003008 | 95.64 | 237.78 |
| 1000rh150-50fixesRetRad220Nov0pt000009-10scan | 0.00000009 | 3.435604381 | 95.674 | 237.86 |
| 1000rh150-50fixesRetRad220Nov0pt000009-10scan | 0.0000009 | 3.639787234 | 94.428 | 237.74 |
| 1000rh150-50fixesRetRad220Nov0pt000009-10scan | 0.000009 | 3.340248145 | 96.47 | 237.69 |
| 1000rh150-50fixesRetRad220Nov0pt00009-10scan | 0.00009 | 7.013703691 | 92.864 | 237.72 |
| 1000rh150-50fixesRetRad220lTfb5-10scan | 0.0009 | 7.241732505 | 91.66 | 238.09 |
| 1000rh150-50fixesRetRad220Nov0pt009-10scan | 0.009 | 7.76836961 | 92.162 | 2.38E+02 |
| 1000rh150-50fixesRetRad220Nov0pt05-10scan | 0.05 | 8.182743835 | 94.232 | 2.40E+02 |
| 1000rh150-50fixesRetRad220Nov0pt09-10scan | 0.09 | 7.883669296 | 92.812 | 2.40E+02 |
| 1000rh150-50fixesRetRad220Nov0pt5-10scan | 0.5 | 8.077268015 | 93.146 | 2.41E+02 |
| 1000rh150-50fixesRetRad220Nov0pt9-10scan | 0.9 | 7.913897199 | 92.544 | 2.40E+02 |
| 1000rh150-50fixesRetRad220Nov1-10scan | 1 | 7.978554851 | 92.526 | 2.40E+02 |
| 1000rh150-50fixesRetRad220Nov2-10scan | 2 | 8.135738884 | 95.224 | 2.41E+02 |
| 1000rh150-50fixesRetRad220Nov3-10scan | 3 | 7.998275997 | 93.136 | 2.41E+02 |
| 1000rh150-50fixesRetRad220Nov4-10scan | 4 | 8.065560717 | 94.756 | 2.40E+02 |
| 1000rh150-50fixesRetRad220Nov5-10scan | 5 | 8.168042423 | 93.018 | 2.41E+02 |
| 1000rh150-50fixesRetRad220Nov6-10scan | 6 | 7.869835051 | 92.276 | 2.40E+02 |
| 1000rh150-50fixesRetRad220Nov7-10scan | 7 | 8.030390402 | 91.778 | 2.40E+02 |
| 1000rh150-50fixesRetRad220Nov8-10scan | 8 | 8.094264436 | 92.776 | 2.41E+02 |
| 1000rh150-50fixesRetRad220Nov9-10scan | 9 | 8.102649699 | 93.318 | 2.40E+02 |
| 1000rh150-50fixesRetRad220Nov10-10scan | 10 | 7.661822983 | 91.19 | 2.41E+02 |
| 1000rh150-50fixesRetRad220Nov11-10scan | 11 | 8.006570471 | 94.008 | 2.40E+02 |
| 1000rh150-50fixesRetRad220Nov12-10scan | 12 | 8.006213221 | 93.138 | 2.40E+02 |

Normal parameter values shown in orange

## Appendix A3.8 - Effect of Varying Weight of Novelty Bias in a Sparse Scene

Figures averaged over 10 scan paths, of 50 fixations each, over image 1000gh50

| Workspace Name | Weight of Novelty Bias | % Fixations Near Target Colour Stimuli | % Fixations Near Non-Target Colour Stimuli | % Fixations in Blank Areas | % Fixations in New Location | % Re-Fixations |
|---|---|---|---|---|---|---|
| 1000gh50-50fixesRetRad220Nov0-10scan | 0 | 92.65306122 | 7.346938776 | 0 | 31.0 | 69.0 |
| 1000gh50-50fixesRetRad220Nov0pt000009-10scan | 0.00000009 | 91.42857143 | 8.571428571 | 0 | 25.8 | 74.2 |
| 1000gh50-50fixesRetRad220Nov0pt000009-10scan | 0.0000009 | 95.91836735 | 4.081632653 | 0 | 34.6 | 65.4 |
| 1000gh50-50fixesRetRad220Nov0pt000009-10scan | 0.000009 | 92.04081633 | 7.959183673 | 0 | 34.2 | 65.8 |
| 1000gh50-50fixesRetRad220Nov0pt00009-10scan | 0.00009 | 98.7755102 | 1.020408163 | 0.204081633 | 84.8 | 15.2 |
| 1000gh50-50fixesRetRad220Nov0pt0009-10scan | 0.0009 | 86.93877551 | 11.02040816 | 2.040816327 | 95.8 | 4.2 |
| 1000gh50-50fixesRetRad220Nov0pt00495-10scan | 0.00495 | 66.73469388 | 15.51020408 | 17.75510204 | 98.2 | 1.8 |
| 1000gh50-50fixesRetRad220Nov0pt009-10scan | 0.009 | 59.79591837 | 12.65306122 | 27.55102041 | 100.0 | 0.0 |
| 1000gh50-50fixesRetRad220Nov0pt05-10scan | 0.05 | 56.73469388 | 12.04081633 | 31.2244898 | 99.8 | 0.2 |
| 1000gh50-50fixesRetRad220Nov0pt09-10scan | 0.09 | 56.32653061 | 12.85714286 | 30.81632653 | 100.0 | 0.0 |
| 1000gh50-50fixesRetRad220Nov0pt5-10scan | 0.5 | 54.08163265 | 13.67346939 | 32.24489796 | 99.6 | 0.4 |
| 1000gh50-50fixesRetRad220Nov0pt9-10scan | 0.9 | 54.89795918 | 13.87755102 | 31.2244898 | 99.6 | 0.4 |
| 1000gh50-50fixesRetRad220Nov1-10scan | 1 | 54.08163265 | 12.44897959 | 33.46938776 | 99.2 | 0.8 |
| 1000gh50-50fixesRetRad220Nov2-10scan | 2 | 53.46938776 | 14.28571429 | 32.24489796 | 99.2 | 0.8 |
| 1000gh50-50fixesRetRad220Nov3-10scan | 3 | 52.85714286 | 13.46938776 | 33.67346939 | 100.0 | 0.0 |
| 1000gh50-50fixesRetRad220Nov4-10scan | 4 | 56.93877551 | 11.63265306 | 31.42857143 | 99.0 | 1.0 |
| 1000gh50-50fixesRetRad220Nov5-10scan | 5 | 58.57142857 | 11.63265306 | 29.79591837 | 99.0 | 1.0 |
| 1000gh50-50fixesRetRad220Nov6-10scan | 6 | 55.10204082 | 14.48979592 | 30.40816327 | 99.6 | 0.4 |
| 1000gh50-50fixesRetRad220Nov7-10scan | 7 | 56.12244898 | 13.06122449 | 30.81632653 | 99.0 | 1.0 |
| 1000gh50-50fixesRetRad220Nov8-10scan | 8 | 56.32653061 | 14.69387755 | 28.97959184 | 98.8 | 1.2 |
| 1000gh50-50fixesRetRad220Nov9-10scan | 9 | 54.48979592 | 16.12244898 | 29.3877551 | 99.8 | 0.2 |
| 1000gh50-50fixesRetRad220Nov10-10scan | 10 | 55.10204082 | 11.63265306 | 33.26530612 | 98.8 | 1.2 |
| 1000gh50-50fixesRetRad220Nov11-10scan | 11 | 56.12244898 | 14.28571429 | 29.59183673 | 98.8 | 1.2 |
| 1000gh50-50fixesRetRad220Nov12-10scan | 12 | 55.10204082 | 11.83673469 | 33.06122449 | 99.6 | 1.4 |

Normal parameter values shown in orange

## Appendix 3.8 - Effect of Varying Weight of Novelty Bias in a Sparse Scene
(Continued)

| Workspace Name | Weight of Novelty Bias | Average Saccade Amplitude | Average AW Radius (pixels) | Average Fixation Duration |
|---|---|---|---|---|
| 1000gh50-50fixesRetRad220Nov0-10scan | 0 | 7.003043315 | 1.78E+02 | 239.89 |
| 1000gh50-50fixesRetRad220Nov0pt000009-10scan | 0.00000009 | 6.297520846 | 1.76E+02 | 239.66 |
| 1000gh50-50fixesRetRad220Nov0pt000009-10scan | 0.0000009 | 6.91335453 | 1.70E+02 | 239.21 |
| 1000gh50-50fixesRetRad220Nov0pt000009-10scan | 0.000009 | 6.344367527 | 1.82E+02 | 240.12 |
| 1000gh50-50fixesRetRad220Nov0pt00009-10scan | 0.00009 | 10.46519479 | 1.68E+02 | 238.46 |
| 1000gh50-50fixesRetRad220Nov0pt0009-10scan | 0.0009 | 11.66269134 | 1.68E+02 | 239.35 |
| 1000gh50-50fixesRetRad220Nov0pt00495-10scan | 0.00495 | 12.94988565 | 1.67E+02 | 240.39 |
| 1000gh50-50fixesRetRad220Nov0pt009-10scan | 0.009 | 12.88058567 | 1.63E+02 | 2.40E+02 |
| 1000gh50-50fixesRetRad220Nov0pt05-10scan | 0.05 | 13.2325724 | 1.64E+02 | 2.41E+02 |
| 1000gh50-50fixesRetRad220Nov0pt09-10scan | 0.09 | 13.43756226 | 165 | 2.42E+02 |
| 1000gh50-50fixesRetRad220Nov0pt5-10scan | 0.5 | 13.25458817 | 1.65E+02 | 2.42E+02 |
| 1000gh50-50fixesRetRad220Nov0pt9-10scan | 0.9 | 13.10664982 | 1.65E+02 | 2.42E+02 |
| 1000gh50-50fixesRetRad220Nov1-10scan | 1 | 13.30842803 | 1.66E+02 | 2.42E+02 |
| 1000gh50-50fixesRetRad220Nov2-10scan | 2 | 13.52374765 | 1.64E+02 | 2.41E+02 |
| 1000gh50-50fixesRetRad220Nov3-10scan | 3 | 13.60500015 | 1.65E+02 | 2.42E+02 |
| 1000gh50-50fixesRetRad220Nov4-10scan | 4 | 13.25181625 | 1.65E+02 | 2.42E+02 |
| 1000gh50-50fixesRetRad220Nov5-10scan | 5 | 13.18180173 | 1.66E+02 | 2.42E+02 |
| 1000gh50-50fixesRetRad220Nov6-10scan | 6 | 13.32287026 | 1.65E+02 | 2.42E+02 |
| 1000gh50-50fixesRetRad220Nov7-10scan | 7 | 13.29002384 | 1.64E+02 | 2.41E+02 |
| 1000gh50-50fixesRetRad220Nov8-10scan | 8 | 13.37635284 | 1.66E+02 | 2.43E+02 |
| 1000gh50-50fixesRetRad220Nov9-10scan | 9 | 13.19503244 | 1.66E+02 | 2.43E+02 |
| 1000gh50-50fixesRetRad220Nov10-10scan | 10 | 13.30440535 | 1.66E+02 | 2.42E+02 |
| 1000gh50-50fixesRetRad220Nov11-10scan | 11 | 13.41374148 | 1.68E+02 | 2.42E+02 |
| 1000gh50-50fixesRetRad220Nov12-10scan | 12 | 13.3410216 | 1.66E+02 | 2.42E+02 |

Normal parameter values shown in orange

## Appendix A3.9 - Effect of Varying Weight of Novelty Bias in a Very Sparse Scene

Figures averaged over 10 scan paths, of 50 fixations each, over image 1000rv11

| Workspace Name | Weight of Novelty Bias | % Fixations Near Target Colour Stimuli | % Fixations Near Non-Target Colour Stimuli | % Fixations in Blank Areas | % Fixations in New Location | % Re-Fixations |
|---|---|---|---|---|---|---|
| 1000rv11-50fixesRetRad220Nov0-10scan | 0 | 100 | 0 | 0 | 24.0 | 76.0 |
| 1000rv11-50fixesRetRad220Nov0pt000009-10scan | 0.00000009 | 100 | 0 | 0 | 24.0 | 76.0 |
| 1000rv11-50fixesRetRad220Nov0pt000009-10scan | 0.0000009 | 100 | 0 | 0 | 24.0 | 76.0 |
| 1000rv11-50fixesRetRad220Nov0pt000009-10scan | 0.000009 | 100 | 0 | 0 | 24.0 | 76.0 |
| 1000rv11-50fixesRetRad220Nov0pt00009-10scan | 0.00009 | 99.18367347 | 0 | 0.816326531 | 67.2 | 32.8 |
| 1000rv11-50fixesRetRad220Nov0pt0009-10scan | 0.0009 | 73.87755102 | 18.97959184 | 7.142857143 | 92.8 | 7.2 |
| 1000rv11-50fixesRetRad220Nov0pt00495-10scan | 0.00495 | 32.44897959 | 12.24489796 | 55.30612245 | 99.4 | 0.6 |
| 1000rv11-50fixesRetRad220Nov0pt009-10scan | 0.009 | 28.36734694 | 9.591836735 | 62.04081633 | 100.0 | 0.0 |
| 1000rv11-50fixesRetRad220Nov0pt05-10scan | 0.05 | 14.08163265 | 6.12244898 | 79.79591837 | 99.6 | 0.4 |
| 1000rv11-50fixesRetRad220Nov0pt09-10scan | 0.09 | 15.51020408 | 6.326530612 | 78.16326531 | 99.6 | 0.4 |
| 1000rv11-50fixesRetRad220Nov0pt5-10scan | 0.5 | 11.83673469 | 6.734693878 | 81.42857143 | 99.8 | 0.2 |
| 1000rv11-50fixesRetRad220Nov0pt9-10scan | 0.9 | 13.46938776 | 6.12244898 | 80.40816327 | 99.6 | 0.4 |
| 1000rv11-50fixesRetRad220Nov1-10scan | 1 | 12.44897959 | 6.326530612 | 81.2244898 | 100.0 | 0.0 |
| 1000rv11-50fixesRetRad220Nov2-10scan | 2 | 12.24489796 | 6.326530612 | 81.42857143 | 100.0 | 0.0 |
| 1000rv11-50fixesRetRad220Nov3-10scan | 3 | 12.44897959 | 6.326530612 | 81.2244898 | 99.8 | 0.2 |
| 1000rv11-50fixesRetRad220Nov4-10scan | 4 | 13.26530612 | 5.510204082 | 81.2244898 | 100.0 | 0.0 |
| 1000rv11-50fixesRetRad220Nov5-10scan | 5 | 13.26530612 | 8.163265306 | 78.57142857 | 100.0 | 0.0 |
| 1000rv11-50fixesRetRad220Nov6-10scan | 6 | 12.65306122 | 5.918367347 | 81.42857143 | 99.8 | 0.2 |
| 1000rv11-50fixesRetRad220Nov7-10scan | 7 | 14.28571429 | 6.93877551 | 78.7755102 | 100.0 | 0.0 |
| 1000rv11-50fixesRetRad220Nov8-10scan | 8 | 14.08163265 | 5.306122449 | 80.6122449 | 99.8 | 0.2 |
| 1000rv11-50fixesRetRad220Nov9-10scan | 9 | 12.44897959 | 6.734693878 | 80.81632653 | 100.0 | 0.0 |
| 1000rv11-50fixesRetRad220Nov10-10scan | 10 | 10.81632653 | 5.510204082 | 83.67346939 | 100.0 | 0.0 |
| 1000rv11-50fixesRetRad220Nov11-10scan | 11 | 13.46938776 | 6.734693878 | 79.79591837 | 100.0 | 0.0 |
| 1000rv11-50fixesRetRad220Nov12-10scan | 12 | 14.08163265 | 5.510204082 | 80.40816327 | 100.0 | 0.0 |

Normal parameter values shown in orange

## Appendix 3.9 - Effect of Varying Weight of Novelty Bias in a Very Sparse Scene
(Continued)

| Workspace Name | Weight of Novelty Bias | Average Saccade Amplitude | Average AW Radius (pixels) | Average Fixation Duration |
|---|---|---|---|---|
| 1000rv11-50fixesRetRad220Nov0-10scan | 0 | 1.835287065 | 3.03E+02 | 239.71 |
| 1000rv11-50fixesRetRad220Nov0pt000009-10scan | 0.00000009 | 1.835287065 | 3.03E+02 | 239.56 |
| 1000rv11-50fixesRetRad220Nov0pt000009-10scan | 0.0000009 | 1.835287065 | 3.03E+02 | 239.63 |
| 1000rv11-50fixesRetRad220Nov0pt000009-10scan | 0.000009 | 1.835287065 | 3.03E+02 | 239.42 |
| 1000rv11-50fixesRetRad220Nov0pt00009-10scan | 0.00009 | 12.39881725 | 3.26E+02 | 239.81 |
| 1000rv11-50fixesRetRad220Nov0pt0009-10scan | 0.0009 | 18.88022391 | 3.45E+02 | 240.80 |
| 1000rv11-50fixesRetRad220Nov0pt00495-10scan | 0.00495 | 21.74977915 | 3.30E+02 | 240.39 |
| 1000rv11-50fixesRetRad220Nov0pt009-10scan | 0.009 | 23.53632711 | 3.32E+02 | 2.39E+02 |
| 1000rv11-50fixesRetRad220Nov0pt05-10scan | 0.05 | 26.20567622 | 3.32E+02 | 2.39E+02 |
| 1000rv11-50fixesRetRad220Nov0pt09-10scan | 0.09 | 26.26795437 | 3.31E+02 | 2.40E+02 |
| 1000rv11-50fixesRetRad220Nov0pt5-10scan | 0.5 | 26.49544534 | 3.29E+02 | 2.41E+02 |
| 1000rv11-50fixesRetRad220Nov0pt9-10scan | 0.9 | 26.11866506 | 3.30E+02 | 2.40E+02 |
| 1000rv11-50fixesRetRad220Nov1-10scan | 1 | 26.22471676 | 3.29E+02 | 2.40E+02 |
| 1000rv11-50fixesRetRad220Nov2-10scan | 2 | 25.88365033 | 3.29E+02 | 2.40E+02 |
| 1000rv11-50fixesRetRad220Nov3-10scan | 3 | 26.96904865 | 3.30E+02 | 2.39E+02 |
| 1000rv11-50fixesRetRad220Nov4-10scan | 4 | 26.1047434 | 3.29E+02 | 2.40E+02 |
| 1000rv11-50fixesRetRad220Nov5-10scan | 5 | 26.89931845 | 3.28E+02 | 2.41E+02 |
| 1000rv11-50fixesRetRad220Nov6-10scan | 6 | 26.39442767 | 3.30E+02 | 2.40E+02 |
| 1000rv11-50fixesRetRad220Nov7-10scan | 7 | 26.10968128 | 3.28E+02 | 2.41E+02 |
| 1000rv11-50fixesRetRad220Nov8-10scan | 8 | 26.47471723 | 3.29E+02 | 2.40E+02 |
| 1000rv11-50fixesRetRad220Nov9-10scan | 9 | 26.44106964 | 3.29E+02 | 2.41E+02 |
| 1000rv11-50fixesRetRad220Nov10-10scan | 10 | 26.51859772 | 3.30E+02 | 2.40E+02 |
| 1000rv11-50fixesRetRad220Nov11-10scan | 11 | 26.22979777 | 3.32E+02 | 2.41E+02 |
| 1000rv11-50fixesRetRad220Nov12-10scan | 12 | 26.24276007 | 3.28E+02 | 2.41E+02 |

Normal parameter values shown in orange

# Appendix A4

## Scan Path Statistics After Lesion

## Appendix A4.1 - Lesion Simulations
### (Single Scan Path Simulations)

| Workspace Name | Active or Covert | LIP Lesion | Novelty Weight | No. of Fixations | No. of fixations at new locations | %age fixations at new locations | No. of re-fixations | %age re-fixation |
|---|---|---|---|---|---|---|---|---|
| **Normal-Control** | | | | | | | | |
| 881rv100-50fixesNoLesionCovert | Covert | N | 0.0009 | 50 | 44 | 88 | 6 | 12 |
| 881rv100-100fixesNoLesionCovert | Covert | N | 0.0009 | 100 | 65 | 65 | 35 | 35 |
| 881rv100-50fixesNoLesionActive | Active | N | 0.0009 | 50 | 49 | 98 | 1 | 2 |
| 881rv100-100fixesNoLesionActive | Active | N | 0.0009 | 100 | 86 | 86 | 14 | 14 |
| **Unilateral LIP Lesion** | | | | | | | | |
| 881rv100-50fixesLesionLIPCovert | Covert | Y | 0.0009 | 50 | 38 | 76 | 12 | 24 |
| 881rv100-100fixesLesionLIPCovert | Covert | Y | 0.0009 | 100 | 47 | 47 | 53 | 53 |
| **Full Orbitofrontal Lesion** | | | | | | | | |
| 881rv100-50fixesOrbitoFullLesionCovert | Covert | N | 0 | 50 | 9 | 18 | 41 | 82 |
| 881rv100-100fixesOrbitoFullLesionCovert | Covert | N | 0 | 100 | 10 | 10 | 90 | 90 |
| **Unilateral Orbitofrontal Lesion** | | | | | | | | |
| 881rv100-50fixesOrbitoLeftLesionCovert | Covert | N | 0 & 0.0009 | 50 | 37 | 74 | 13 | 26 |
| 881rv100-100fixesOrbitoLeftLesionCovert | Covert | N | 0 & 0.0009 | 100 | 41 | 41 | 59 | 59 |
| **Orbitofrontal Active Vision Simulations** | | | | | | | | |
| 881rv100-50fixesOrbitoFullLesionActive | Active | N | 0 | 50 | 22 | 44 | 28 | 56 |
| 881rv100-100fixesOrbitoFullLesionActive | Active | N | 0 | 100 | 27 | 27 | 73 | 73 |
| 881rv100-50fixesOrbitoLeftLesionActive | Active | N | 0 & 0.0009 | 50 | 47 | 94 | 3 | 6 |
| 881rv100-100fixesOrbitoLeftLesionActive | Active | N | 0 & 0.0009 | 100 | 61 | 61 | 39 | 39 |
| **Full Orbitofrontal & Unilateral LIP Lesions** | | | | | | | | |
| 881rv100-50fixesOrbitoFullLesionPlusLIPLesionCovert | Covert | Y | 0 | 50 | 8 | 16 | 42 | 84 |
| 881rv100-100fixesOrbitoFullLesionPlusLIPLesionCovert | Covert | Y | 0 | 100 | 11 | 11 | 89 | 89 |
| 881rv100-50fixesOrbitoFullReduced0pt0000009PlusLIPLesionCovert | Covert | Y | 0.0000009 | 50 | 17 | 34 | 33 | 66 |
| 881rv100-100fixesOrbitoFullReduced0pt0000009PlusLIPLesionCovert | Covert | Y | 0.0000009 | 100 | 17 | 17 | 83 | 83 |
| 881rv100-50fixesOrbitoFullReduced0pt00009PlusLIPLesionCovert | Covert | Y | 0.00009 | 50 | 30 | 60 | 20 | 40 |
| 881rv100-100fixesOrbitoFullReduced0pt00009PlusLIPLesionCovert | Covert | Y | 0.00009 | 100 | 34 | 34 | 66 | 66 |

## Appendix A4.2 - Lesion Simulations

### (Multiple Simulations - averaged over 10 scan paths of 50 fixations each)

| Workspace Name | Active or Covert | LIP Lesion | Novelty Weight | No. of Fixations in each scan path | Av. No. of fixations at new locations | Av. %age fixations at new locations | Av. No. of re-fixations | Av. %age re-fixation |
|---|---|---|---|---|---|---|---|---|
| **Normal-Control** | | | | | | | | |
| 881rv100-50fixesCovertNoLesion-10scan | Covert | N | 0.0009 | 50 | 43.6 | 87.2 | 6.4 | 12.8 |
| 881rv100-50fixesActiveNoLesion-10scan | Active | N | 0.0009 | 50 | 48 | 96 | 2 | 4 |
| **Unilateral LIP Lesion** | | | | | | | | |
| 881rv100-50fixesCovertLIPLesion-10scan | Covert | Y | 0.0009 | 50 | 37.7 | 75.4 | 12.3 | 24.6 |
| **Full Orbitofrontal Lesion** | | | | | | | | |
| 881rv100-50fixesCovertWholeOrbitoLesion-10scan | Covert | N | 0 | 50 | 6.4 | 12.8 | 43.6 | 87.2 |
| 881rv100-50fixesCovertOrbitoReduced0pt0000009-10scan | Covert | N | 0.0000009 | 50 | 8 | 16 | 42 | 84 |
| 881rv100-50fixesCovertOrbitoReduced0pt00009-10scan | Covert | N | 0.00009 | 50 | 39.5 | 79 | 10.5 | 21 |
| **Unilateral Orbitofrontal Lesion** | | | | | | | | |
| 881rv100-50fixesCovertLeftOrbitoLesion-10scan | Covert | N | 0 & 0.0009 | 50 | 28.8 | 57.6 | 21.2 | 42.4 |
| **Orbitofrontal Active Vision Simulations** | | | | | | | | |
| 881rv100-50fixesActiveOrbitoFullLesion-10scan | Active | N | 0 | 50 | 14.6 | 29.2 | 35.4 | 70.8 |
| 881rv100-50fixesActiveOrbitoLeftLesion-10scan | Active | N | 0 & 0.0009 | 50 | 41.8 | 83.6 | 8.2 | 16.4 |
| **Full Orbitofrontal & Unilateral LIP Lesions** | | | | | | | | |
| 881rv100-50fixesCovertWholeOrbitoPlusLIPLesion-10scan | Covert | Y | 0 | 50 | 11.2 | 22.4 | 38.8 | 77.6 |
| 881rv100-50fixesCovOrbReduced0pt0000009PlusLIPLesion-10scan | Covert | Y | 0.0000009 | 50 | 14 | 28 | 36 | 72 |
| 881rv100-50fixesCovOrbReduced0pt00009PlusLIPLesion-10scan | Covert | Y | 0.00009 | 50 | 28.2 | 56.4 | 21.8 | 43.6 |

# Glossary of Terms

**Cortical Regions**

| | |
|---|---|
| AIP | Anterior intraparietal area, in parietal cortex |
| FEF | Frontal eye field |
| LGN | Lateral geniculate nucleus |
| LIP | Lateral intraparietal area, in parietal cortex |
| MIP | Medial intraparietal area, in parietal cortex |
| MST | Medial superior temporal area |
| MT | Middle temporal area |
| TE | An area in anterior IT cortex |
| TEO | An area in posterior IT cortex |
| V1 | Visual area 1, Primary Visual cortex; also known as striate cortex |
| V2 | Visual area 2, in extrastriate cortex |
| V3 | Visual area 3, in extrastriate cortex |
| V4 | Visual area 4, in extrastriate cortex |
| IT | Inferior temporal cortex |
| VIP | Ventral intraparietal area, in parietal cortex |

## Neurological Recording Techniques

| | |
|---|---|
| BOLD | Blood oxygen level dependant response used in fMRI |
| ERP | Event-related potentials |
| fMRI | Functional magnetic resonance imaging |
| PET | Positron emission tomography |

## Miscellaneous

| | |
|---|---|
| AW | Attention window |
| DOOG | Difference of offset Gaussian (see chapter 3) |
| HCI | Human-computer interaction |
| IOR | Inhibition of return |
| Retinal image | Subset of the original image falling on the retina and available to the cortical modules |
| Retinal view | same as retinal image |

# List of References

Aivar, M.P., Hayhoe, M.M., Chizk, C.L. & Mruczek, R.E.B. (2005) Spatial memory and saccadic targeting in a natural task. Journal of Vision, 5(3), 117-193

Anand, S., & Bridgeman, B. (2002) An unbiased measure of the contributions of chroma and luminance to saccadic suppression of displacement. Experimental Brain Research, 142(3), 335-341

Andersen, R. A., & Zipser, D. (1988) The role of the posterior parietal cortex in coordinate transformations for visual-motor integration. Canadian Journal of Physiology & Pharmacology, 66(4), 488-501

Anker, R. L. (1977) The prenatal development of some of the visual pathways in the cat. Journal of Computational Neuroscience, 173(1),185-204

Anllo-Vento, L. & Hillyard, S. A. (1996) Selective attention to the color and direction of moving stimuli: electrophysiological correlates of hierarchical feature selection. Perception & Psychophysics, 58(2), 191-206

Anllo-Vento, L., Luck, S. J. & Hillyard, S. A.(1998) Spatio-temporal dynamics of attention to color: Evidence from human electrophysiology. Human Brain Mapping, 6(4), 216-238

Arterberry, M. E. & Yonas, A. (2000) Perception of three-dimensional shape specified by optic flow by 8-week-old infants. Perception & Psychophysics, 62(3), 550-556

Bacon, W. J., & Egeth, H. E. (1997) Goal-directed guidance of attention: evidence form conjunctive visual search. Journal of Experimental Psychology: Human Perception & Performance, 23(4), 948-961

Baylis, G.C., Baylis, L.L. & Gore, C.L. (2004) Visual neglect can be scene-based or object-based depending on task representation. Cortex, 40(2), 237-246

Behrmann, M., Ebert, P. & Black, S.E. (2004) Hemispatial neglect and visual search: A large scale analysis. Cortex, 40(2), 247-263

Behrmann, M., Zemel, R. S., & Mozer, M. C. (1998) Object-based attention and occlusion: Evidence from normal participants and a computational model. Journal of Experimental Psychology: Human Perception and Performance, 24(4), 1011-1036

Ben Hamed, S., Duhamel, J-R., Bremmer, F. & Graf, W. (1996) Dynamic changes in visual receptive field organisation in the macaque lateral intraparietal area (LIP) during saccade preparation. Society of Neuroscience Abstracts, 22, 635.12

Ben Hamed, S., Duhamel, J-R., Bremmer, F. & Graf, W. (2001) Representation of the visual field in the lateral intraparietal area of macaque monkeys: a quantitative receptive field analysis. Experimental Brain Research, 140, 127-144

Ben Hamed, S., Duhamel, J. R., Bremmer, F. & Graf, W. (2002) Visual receptive field modulation in the lateral intraparietal area during attentive fixation and free gaze. Cerebral Cortex, 12, 234-245

Bertera, J.H. & Rayner, K. (2000) Eye movements and the span of the effective stimulus in visual search. Perception & Psychophysics, 62(3), 576-585

Bichot, N.P. & Schall, J.D. (1999) Effects of similarity and history on neural mechanisms of visual selection. Nature Neuroscience, 2(6), 549-554

Bichot, N.P., Schall, J.D. & Thompson, K.G. (1996) Visual feature selectivity in frontal eye fields induced by experience in mature macaques. Nature, 381, 697-699

Bisley, J.W. & Goldberg, M.E. (2003) Neuronal activity in the lateral intraparietal area and spatial attention. Science, 299, 81-86

Bisley, J.W., Krishna, B.S. & Goldberg, M.E. (2005) A rapid and precise on-response in posterior parietal cortex. Journal of Neuroscience, 24(8):1833-8

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000) Tracking an object through feature space. Nature, 408

Blatt, G.L., Andersen, R.A. & Stoner, G.R. (1990) Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (LIP) in the macaque. Journal of Computational Neurology, 299, 421-445

Borisyuk, R.M. & Kazanovich, Y.B. (2004) Oscillatory model of attention-guided object selection and novelty detection. Neural Networks, 17(7), 899-915

Bower, T. G. R. (1974) Development in Infancy. San Francisco, CA, USA, W. H. Freeman

Brady, N. & Field, D.J. (2000) Local contrast in natural images: normalisation and coding efficiency. Perception, 29(9):1041-55

Bremmer, F., Schlack, A., Duhamel, J.R., Graf, W. & Fink, G.R. (2001) Space coding in primate posterior parietal cortex. Neuroimage, 14(1 Pt 2), S46-51

Brefczynski, J. & DeYoe, E. A. (1999) A physiological correlate of the 'spotlight' of visual attention. Nature Neuroscience, 2, 370-374

Bricolo, E., Gianesini, T., Fanini, A., Bundesen, C. & Chelazzi, L. (2002) Serial attention mechanisms in visual search: a direct behavioural demonstration. Journal of Cognitive Neuroscience, 14 (7), 980-993

Bridgeman, B. (1995) A review of the role of efference copy in sensory and oculomotor control systems. Annals of Biomedical Engineering, 23(4), 409-422

Bridgeman, B., Vanderheijden, A.H.C., Velichkovsky, B.M. (1996) Relationship of saccadic suppression to space constancy – response. Behavioural and Brain Sciences, 19(3), 533

Burkhalter, A. (1993) Development of forward and feedback connections between areas V1 and V2 of human visual cortex. Cerebral Cortex, 3(5),476-87

Burt, P J (1988) Smart sensing within a pyramid vision machine. Proceedings of the IEEE, 76(8), 1006-1015

Bushnell, M. C., Goldberg, M. E. & Robinson, D. L. (1981) Behavioural enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. Journal of Neurophysiology, 46(4), 755-772

Cave, K. R. & Wolfe, J. M. (1990) Modeling the role of parallel processing in visual search. Cognitive Psychology, 22, 225-271

Chavane, F., Monier, C., Bringuier, V., Baudot, P., Borg-Graham, L., Lorenceau, J. & Fregnac, Y. (2000) The visual cortical association field: a Gestalt concept or a psychophysiological entity? Journal of Physiology-Paris, 94(5-6), 333-342

Chawla, D., Rees, G. & Friston, K. J. (1999) The physiological basis of attentional modulation in extrastriate visual areas. Nature Neuroscience, 2(7), 671-676

Chelazzi, L, Duncan, J, Miller, E.K. (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. Journal of Neurophysiology, 80, 2918-2940

Chelazzi, L., Miller, E.K., Duncan, J. & Desimone, R. (1993) A neural basis for visual search in inferior temporal cortex. Nature, 363, 345-347

Chelazzi, L., Miller, E.K., Duncan, J., & Desimone, R. (2001) Responses of neurons in macaque area V4 during memory-guided visual search. Cerebral Cortex, 11, 761-772

Cheng, A., Eysel, U.T. & Vidyasagar, T.R. (2004) The role of the magnocellular pathway in serial deployment of visual attention. European Journal of Neuroscience, 20(8):2188-92

Clark, V. P., & Hillyard, S. A (1996) Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. Journal of Cognitive Neuroscience, 8(5), 387-402

Clohessy, A. B., Posner, M. I., Rothbart, M. K. & Vecera, S. P. (1991) The development of inhibition of return in early infancy. Journal of Cognitive Neuroscience, 3, 345-350

Colby, C. L., Duhamel, J. R. & Goldberg, M. E. (1996) Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. Journal of Neurophysiology, 76(5), 2841-2852

Colby, C. L. & Goldberg, M. E. (1999) Space and attention in parietal cortex. Annual Review of Neuroscience, 22, 319-349

Compte, A., Constantinidis, C., Tegner, J., Raghavachari, S., Chafee, M.V., Goldman-Rakic, P.S. & Wang, X.J. (2003) Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. Journal of Neurophysiology, 90(5), 3441-3454

Connor, C.E., Gallant, J.L., Preddie, D.C., & Van Essen, D.C. (1996) Responses in area V4 depend on the spatial relationship between stimulus and attention. Journal of Neurophysiology, 75, 1306-1308

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000) Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nature Neuroscience, 3, 292-297

Corbetta, M., Miezin, F.M., Shulman, G.L. & Petersen, S.E. (1993) A PET study of visuospatial attention. The Journal of Neuroscience, 13, 1202-1226

Corbetta, M., Shulman, G.L., Miezin, F.M. & Petersen, S.E. (1995) Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. Science, 270, 802-805

Craik, K.J.W (1940) The effect of adaptation on subjective brightness. Proceedings of the Royal Society of London, Series B, 128, 232-247

Craton, L. G. (1996) The development of perceptual completion abilities: infant's perception of stationary, partially occluded objects. Child Development, 67(3), 890-904

Crick, F. (1984) Function of the thalamic reticular complex: The searchlight hypothesis. Proceedings of the National Academy of Science of the USA, 81, 4586-4590

Crick, F., & Koch, C. (1990) Towards a neurobiological theory of consciousness. Seminars in the Neurosciences, 2, 263-275

Croner, L. J., & Albright, T. D. (1997). Image segmentation enhances discrimination of motion in visual noise. Vision research, 37(11), 1415-1427

Damasio, A. R. (1996) The somatic marker hypothesis and the possible functions of the prefrontal cortex. Philosophical Transactions of the Royal Society London B, 356, 1413-1420.

Danziger, S., Kingstone, A. & Snyder, J. J. (1998) Inhibition of return to successively stimulated locations in a sequential visual search paradigm. Journal of Experimental Psychology: Human Perception and Performance, 24(5), 1467-1475

Deco, G. (2001) Biased Competition Mechanisms for Visual Attention. In Emergent Neural Computational Architectures Based on Neuroscience (Wermter, S., Austin, J. & Willshaw, D.), 114-126, Heidelberg, Springer

de Frockert, J. W., Rees, G., Frith, C. D. & Lavie, N. (2001) The role of working memory in visual selective attention. Science, 291, 1803-1806

de Kamps, M. & van der Velde, F. (2001) Using a recurrent network to bind form, color and position into a unified percept. Neurocomputing, 38-40, 523-528

Deco, G. & Lee, T.S. (2002) A unified model of spatial and object attention based on inter-cortical biased competition. Neurocomputing, 44-46, 775-781

Deco, G., Pollatos, O. & Zihl, J. (2002) The time course of selective visual attention: theory and experiments. Vision Research, 42, 2925-2945

Desimone, R. (1998) Visual attention mediated by biased competition in extrastriate visual cortex. Philosophical Transactions of the Royal Society of London B, 353, 1245-1255

Desimone, R., & Duncan, J. (1995) Neural mechanisms of selective visual attention. Annual Review of Neuroscience, 18, 193-222

Desimone, R., & Schein, S. J. (1987) Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form. Journal of Neurophysiology, 57(3), 835-868

Deubel, H., Bridgeman, B. & Schneider, W.X. (1998) Immediate post-saccadic information mediates space constancy. Vision Research, 38(20), 3147-3159

Deubel, H., Schneider, W.X. & Bridgeman, B. (1996) Postsaccadic target blanking prevents saccadic suppression of image displacement. Vision Research, 36(7):985-96

Dodd, M.D., Castel, A.D. & Pratt, J. (2003) Inhibition of return with rapid serial shifts of attention: Implications for memory and visual search. Perception & Psychophysics, 65(7), 1126-1135

Driver, J., Vuilleumier, P., Eimer, M. & Rees, G. (2001) Functional magnetic imaging and evoked potential correlates of conscious and unconscious vision in parietal extinction patients. Neuroimage, 14, S68-S75

Deubel, H., Schneider, W.X. & Bridgeman B. (1996) Postsaccadic target blanking prevents saccadic suppression of image displacement. Vision Research, 36, 985-996

Deubel, H., Bridgeman, B. & Schneider, W.X. (1998) Immediate post-saccadic information mediates space constancy. Vision Research, 38(20), 3147-3159

De Weerd, P., Peralta, M.R. 3rd, Desimone, R. & Ungerleider, L.G. (1999) Loss of attentional stimulus selection after extrastriate cortical lesions in macaques. Nature Neuroscience, 2(8), 753-758

de Valois, R.L. & de Valois, K.K. (1988) Spatial Vision. New York, Oxford, Oxford University Press

Duhamel, J.R., Bremmer, F., BenHamed, S. & Graf, W. (1997) Spatial invariance of visual receptive fields in parietal cortex neurons. Nature, 389(6653), 845-848

Duhamel J.-R., Colby C.L., & Goldberg M.E. (1992) The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, *255*, 90–92.

Duhamel J.-R., Goldberg M.E., FitzGibbon E.J., Sirigu A., & Grafman J. (1992). Saccadic dysmetria in a patient with a right frontoparietal lesion: the importance of corollary discharge for accurate spatial behavior. *Brain*, *115*, 1387–1402.

Duncan, J. (1984) Selective attention and the organisation of visual information. Journal of Experimental Psychology: General, 113, 501-517

Duncan, J (2001) An Adaptive Coding Model of Neural Function in Prefrontal Cortex. Nature Reviews Neuroscience, 2, 820- 829

Duncan, J. & Humphreys, G. W. (1989) Visual search and stimulus similarity. Psychological Review, 96(3), 433-458.

Duncan, J. & Humphreys, G. W. (1992) Beyond the search surface: Visual search and attentional engagement. Journal of Experimental Psychology: Human Perception & Performance, 18, 578-588

Duncan, J., Humphreys, G. W. & Ward R (1997) Competitive brain activity is visual attention. Current Opinion in Neurobiology, 7(2), 255-261

Eriksen, C. W., & St. James, J. D. (1986) Visual attention within and around the field of focal attention: A zoom lens model. Perception and Psychophysics, 40, 225-240

Everling, S., Tinsley, C. J., Gaffan, D. & Duncan, J. (2002) Filtering of neural signals by focused attention in the monkey prefrontal cortex. Nature Neuroscience, 5(7), 671-676

Fantz, R. L. (1964) Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. Science, 146, 668-670

Findlay, J.M. (1982) Global visual processing for saccadic eye movements. Vision Research, 22(8), 1033-45

Findlay, J.M. (1997) Saccade target selection during visual search. Vision Research, 37(5), 617-631

Findlay, J.M., Brown, V., & Gilchrist, I.D. (2001) Saccade target selection in visual search: the effect of information from the previous fixation. Vision Research, 41(1), 87-95

Fink, G. R., Dolan, R.J., Halligan, P. W., Marshall, J. C., & Frith, C. D. (1997) Space-based and object-based visual attention: shared and specific neural domains. Brain, 120, 2013-2028

Folk, C. L., & Remington, R. (1999) Can new objects override attentional control settings? Perception & Psychophysics, 61(4), 727-739

Freedman, D.J., Riesenhuber, M., Poggio, T. & Miller, E.K. (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. The Journal of Neuroscience, 23(12), 5235-5246

Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995) Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. Science, 269(5225), 853-855

Gerstner, W. (2000) Population dynamics of spiking neurons: Fast transients, asynchronous states, and locking. Neural Computation, 12, 43-89

Ghandi, S. P., Heeger, D. J. & Boynton, G. M. (1999) Spatial attention affects brain activity in human primary visual cortex. Proceedings of the National Academy of Science of the USA, 96, 3314-3319

Ghose, G.M. & Ts'O, D.Y. (1997) Form processing modules in primate area V4. The Journal of Neurophysiology, 77(4), 2191-2196

Gilchrist, I.D. & Harvey, M. (2000) Refixation frequency and memory mechanisms in visual search. Current Biology, 10(19), 1209-1212

Gilchrist, I.D., Heywood, C.A. & Findlay, J.M. (1999) Saccade selection in visual search: evidence for spatial frequency specific between-item interactions. Vision Research, 39(7), 1373-1383

Giesbrecht, B., Woldorff, M.G., Song, A.W. & Mangun, G.R. (2003) Neural mechanisms of top-down control during spatial and feature attention. Neuroimage, 19, 496-512

Gottlieb, J. P., Kusunoki, M. & Goldberg, M .E. (1998) The representation of visual salience in monkey parietal cortex. Nature, 391, 481-484

Greene, H.H. & Rayner, K. (2001) Eye movements and familiarity effects in visual search. Vision Research, 41, 3763-3773

Gross, J., Schmitz, F., Schnitzler, I., Kessler, K., Shapiro, K., Hommel, B. & Schnitzler, A. (2004) Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. Proceedings of the National Academy of Science of the USA, 101(35), 13050-13055

Grossberg, S. & Raizada, R. D. S. (2000) Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. Vision Research, 40, 1413-1432

Haenny, P. E. & Schiller, P. H. (1988) State dependent activity in monkey visual cortex. I. Single cell activity in V1 and V4 on visual tasks. Experimental Brain Research, 69, 225-244

Haenny, P. E., Maunsell, J. H. R. & Schiller, P. H. (1988) State dependent activity in monkey visual cortex. II. Retinal and extraretinal factors in V4. Experimental Brain Research, 69: 245-259

Hamker, F.H. (1998) The role of feedback connections in task-driven visual search. In: Connectionist Models in Cognitive Neuroscience: Proceedings of the 5th Neural Computation and Psychology Workshop (NCPW'98). Eds. von Heinke, D., Humphreys, G.W., & Olson, A. Springer-Verlag, London

Hamker, F.H. (2003) The reentry hypothesis: linking eye movements to visual perception. Journal of Vision, 3, 808-816

Hamker, F.H. (2004) A dynamic model of how feature cues guide spatial attention. Vision Research, 44, 501-521

Hasegawa, R.P., Matsumoto, M., & Mikami, A. (2000) Search target selection in monkey prefrontal cortex. The Journal of Neurophysiology, 84, 1692-1696

Helmholtz, H. v. (1867) Handbuch der physiologishen Optik, Voss, Leipzig

Heinke, D. & Humphreys, G.W. (2003) Attention, spatial representation, and visual neglect: simulating emergent attention and spatial memory in the selective attention for identification model (SAIM). Psychological Review, 110, 29-87

Heinke, D. & Humphreys, G.W. Top-down guidance of visual search: A computational account. *Submitted*

Heuer, H.W. & Britten, K.H. (2002) Contrast dependence of response normalisation in area MT of the rhesus macaque. Journal of Neurophysiology, 88, 3398-3408

Hevner, R. F. (2000) Development of connections in the human visual system during fetal mid-gestation: a DiL-tracing study. Journal of Neuropathology & Experimental Neurology, 59(5),385-92

Hicks, J. M., & Richards, J. E. (1998) The effects of stimulus movement and attention on peripheral stimulus localisation by 8- to 26-week old infants. Infant Behaviour & Development, 21(4), 571-589

Hillyard, S. A. & Anllo-Vento, L. (1998) Event-related brain potentials in the study of visual selective attention. Proceedings of the National Academy of Science of the USA, 95, 781-787

Hinkle, D.A. & Connor, C.E. (2001) Disparity tuning in macaque area V4. Neuroreport, 12(2), 365-369

Hirsch, H. V. (1985) The role of visual experience in the development of cat striate cortex. Cellular Molecular Neurobiology, 5(1-2), 103-121

Hodgson, T. L., Mort, D., Chamberlain, M. M., Hutton, S. B., O'Neill, K. S. & Kennard, C. (2002) Orbitofrontal cortex mediates inhibition of return. Neuropsychologia, 40(12), 1891-1901

Hoffman, J.E. & Subramaniam, B. (1995) The role of visual attention in saccadic eye movements. Perception & Psychophysics, 57(6), 787-795

Hooge, I. T. & Frens, M. A. (2000) Inhibition of saccade return (ISR): Spatio-temporal properties of saccade programming. Vision Research, 40(24), 3415-3426

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000) The neural mechanisms of top-down attentional control. Nature Neuroscience, 3(3), 284-291

Hooge, I.T.C. & Erkelens, C.J. (1999) Peripheral vision and oculomotor control during visual search. Vision Research, 39, 1567-1575

Hooge, I. T. & Frens, M. A. (2000) Inhibition of saccade return (ISR): Spatio-temporal properties of saccade programming. Vision Research, 40(24), 3415-3426

Horowitz, T. S., & Wolfe, J. M. (1998) Visual search has no memory. Nature, 394, 575-577

Humphreys, G.W. & Müller, H.J. (1993) SEarch via Recursive Rejection (SERR): A connectionist model of visual search. Cognitive Psychology, 25, 43-110

Humphreys, G.W. & Riddoch, M.J. (1994) Attention to within-object and between-object spatial representations: Multiple sites for visual selection. Cognitive Neuropsychology, 11(2), 207-241

Humphreys, G. W., Cinel, C., Wolfe, J., Olson, A., & Klempen, N. (2000) Fractionating the binding process: neuropsychological evidence distinguishing binding of form from binding of surface features. Vision Research, 40, 1569-1596

Hupe, J. M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R. & Bullier, J. (2001) Feedback connections act on the early part of the responses in monkey visual cortex. Journal of Neurophysiology, 85(1), 134-145

Husain, M., Mannan, S., Hodgson, T., Wojciulik, E., Driver, J. & Kennard, C. (2001) Impaired spatial working memory across saccades contributes to abnormal search in parietal neglect. Brain, 124, 941-952

Intriligator, J. & Cavanagh, P. (2001) The spatial resolution of visual attention. Cognitive Psychology, 43(3), 171-216

Irwin, D.E., & Zelinsky, G.J. (2002) Eye movements and scene perception: Memory for things observed. Perception & Psychophysics, 64(6), 882-895

Ito, M., & Gilbert, C. D. (1999) Attention modulates contextual influences in the primary visual cortex of alert monkeys. Neuron, 22, 593-604

Itti, L, & Koch, C. (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research, 20, 1489-1506

James, W. (1890) The Principles of Psychology, Dover, New York

Kandel, E. R., Schwartz, J. H. & Jessell, T. M. (1991) Principles of Neural Science 3rd Edition, London, Prentice Hall International (UK) Ltd

Kastner, S., Pinsk, M., De Weerd, P., Desimone, R. & Ungerleider, L. (1999) Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron, 22, 751-761

Kaufmann, F. (1995) Development of motion perception in early infancy. European Journal of Pediatrics, 154(9 Supplement 4), S48-53

Kellman, P. J., & Spelke, E. S. (1983) Perception of partly occluded objects in infancy. Cognitive Psychology, 15, 483-524

Kellman, P. J., Gleitman, H., & Spelke, E. S. (1987) Object and observer motion in the perception of object by infants. Journal of Experimental Psychology: Human Perception & Performance, 13(4), 586-593

Khayat, P.S., Spekreijse, H. & Roelfsema, P.R. (2004) Visual information transfer across eye movements in the monkey. Vision Research, 44(25), 2901-2917

Kim, M-S. & Cave, K.R. (1999) Top-down and bottom-up attentional control: On the nature of interference from a salient distractor. Perception & Psychophysics, 61(6), 1009-1023

Klein, R.M. & MacInnes, W.J. (1999) Inhibition of return is a foraging facilitator in visual search. Psychological Science, 10, 346-352

Koch, C. & Ullman, S. (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology, 4, 219-227 (1985)

Kovacs, I., Kozma, P., Feher, A. & Benedek, G. (1999) Late maturation of visual spatial integration in humans. Proceedings of the National Academy Science of the USA, 96(21), 12204-9

Kustov, A.A. & Robinson, D.L. (1996) Shared neural control of attentional shifts and eye movements. Nature, 384(6604), 74-77

Kusunoki, M, Gottlieb, J., & Goldberg, M. E. (2000) The lateral intraparietal area as a salience map: the representation of abrupt onset, stimulus motion, and task relevance. Vision Research, 40, 1459-1468

Lamme, V. A. F., Super, H., & Spekreijse, H. (1998) Feedforward, horizontal and feedback processing in visual cortex. Current Opinion in Neurobiology, 8, 529-535

Lanyon, L. J., & Denham, S.L. (2004a) A biased competition computational model of spatial and object-based attention mediating active visual search. Neurocomputing (Special Issue: Computational Neuroscience: Trends in Research 2004), 58-60C, 655-662

Lanyon, L. J., & Denham, S.L. (2004b) A model of active visual search with object-based attention guiding scan paths. Neural Networks (Special Issue: Vision & Brain), 17(5-6), 873-897

Lanyon, L. J., & Denham, S.L. (2005a) A model of object-based attention that guides active visual search to behaviourally relevant locations. Lecture Notes in Computer Science (Special Issue: WAPCV 2004; L Paletta et al., Eds), Vol. 3368, 42-56

Lanyon, L.J. & Denham, S.L. (2005b) A Model of Spatial & Object-Based Attention for Active Visual Search. Modelling Language, Cognition and Action: Proceedings of the 9th Neural Computation & Psychology Workshop. World Scientific, Progress in Neural Processing, 16, 239-248

Lauria, S.M. & Strauss, M.S. (1975) Eye movements during search for coded and uncoded targets. Perception & Psychophysics, 17, 303-308

Law, M. B., Pratt, J. & Abrams, R. A. (1995) Color-based inhibition of return. Perception & Psychophysics, 57(3), 402-408

Le, T. H., Pardo, J. V. & Hu, X. (1998) 4T-fMRI study of nonspatial shifting of selective attention: cerebellar and parietal contributions. Journal of Neurophysiology. 79, 1535-1548

Lee, T. S. (1996) Image representation using 2D Gabor wavelets. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18, 959-971

Lee, D. & Malpeli, J.G. (1998) Effects of saccades on the activity of neurons in the cat lateral geniculate nucleus. Journal of Neurophysiology, 79(2):922-36

Lee, D. & Quessy, S. (2003) Visual search is facilitated by scene and sequence familiarity in rhesus monkeys. Vision Research, 43(13):1455-63

Lerner, Y. L., Hendler, T., Ben-Bashat, D, Harel, M., & Malach, R. (2001) A hierarchical axis of object processing stages in the human visual cortex. Cerebral Cortex, 11, 287-297

Li, Zhaoping. (2002) A saliency map in primary visual cortex. Trends in Cognitive Sciences, 6(1), 9-16

Logothetis, N. K. (1998) Object vision and visual awareness. Current Opinion in Neurobiology, 8, 536-544

Lublow, R.E. & Kaplan, O. (1997) Visual search as a function of type of prior experience with target and distractor. Journal of Experimental Psychology: Human Perception and Performance, 23(1), 14-24

Luck, S.J., Chelazzi, L., Hillyard, S.A., & Desimone, R. (1997) Neural mechanisms of spatial attention in areas V1, V2 and V4 of macaque visual cortex. Journal of Neurophysiology, 77, 24-42

Luck, S. J., Fan, S. & Hillyard, S. A. (1993) Attention-related modulation of sensory-evoked brain activity in a visual-search task. Journal of Cognitive Neuroscience, 5(2), 188-195

Luck, S. J. & Hillyard, S. A. (1995) The role of attention in feature detection and conjunction discrimination: an electrophysiological analysis. International Journal of Neuroscience, 80(1-4), 281-297.

Luhmann, H. J., Martinez Millan, L. & Singer, W. (1986) Development of horizontal intrinsic connections in cat striate cortex. Experimental Brain Research, 63(2), 443-8

Lumer, E D., Friston, K. J. & Rees, G. (1998) Neural correlates of perceptual rivalry in the human brain. Science, 280(5371), 1930-1934.

Luria, S & Strauss, M (1975) Eye movements during search for coded and uncoded targets. Perception & Psychophysics, 17, 303-308

Lynch, J.C., Graybiel, A.M. & Lobeck, L.J. (1985) The differential projection of two cytoarchitectonic subregions of the inferior parietal lobule of macaque upon the deep layers of the superior colliculus. Journal of Comparative Neurology, 235, 241-254

Mak, B. S. K. & Vera, A. H. (1999) The role of motion in children's categorization of objects. Cognition, 71, B11-B21

Mannan, S.K., Mort, D.J., Hodgson, T.L., Kennard, C., Driver, J. & Husain, M. (2005) Revisiting previously searched locations in visual neglect: Role of right parietal and frontal lesions in misjudging old locations as new. Journal of Cognitive Neuroscience, 17(2), 340-354

Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., Wong, E. C., Hinrichs, H., Heinze, H. J. & Hillyard, S. A. (1999) Involvement of striate and extrastriate visual cortical areas in spatial attention. Nature Neuroscience, 2(4), 364-369

Martinez-Conde, S., Macknik, S.L., Hubel, D.H. (2000) Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. Nature Neuroscience, 3(3), 251-258

Martinez-Conde, S., Macknik, S.L., Hubel, D.H. (2002) The function of bursts of spikes during visual fixation in the awake primate lateral geniculate nucleus and primary visual cortex. Proceedings of the National Academy of Science of the USA, 99(21), 13920-13925

Mazzoni, P., Andersen, R. A., & Jordan, M. I. (1991) A more biologically plausible learning rule than backpropagation applied to a network model of cortical area 7a. Cerebral Cortex, 1(4), 293-307

McAdams, C.J., & Maunsell, J.H.R. (2000) Attention to both space and feature modulates neuronal responses in macaque area V4. Journal of Neurophysiology, 83(3), 1751-1755

McMains, S.A. & Somers, D.C. (2004) Multiple spotlights of attentional selection in human visual cortex. Neuron, 42(4), 677-86

McPeek R.M., Maljkovic V. & Nakayama K. (1999) Saccades require focal attention and are facilitated by a short-term memory system. Vision Research, 39(8), 1555-66

McPeek, R.M., Skavenski, A.A., & Nakayama, K. (2000) Concurrent processing of saccades in visual search. Vision Research, 40(18), 2499-2516

McSorley, E. & Findlay, J.M. (2003) Saccade target selection in visual search: Accuracy improves when more distractors are present. Journal of Vision, 3, 877-892

Merigan, W. H. (1996) Basic visual capacities and shape discrimination after lesions of extrastriate area V4 in macaques. Visual Neuroscience, 13, 51-60

Miller, E.K. & Asaad, W.F. (2002) The prefrontal cortex: conjunction and cognition. In: J. Grafman (Ed), Handbook of Neuropsychology, 2nd Edition, Vol. 7: The Frontal Lobes. Elsevier Science

Miller, E.K., Erickson, C. A. & Desimone, R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. The Journal of Neuroscience, 16(16), 5154-5167

Miller, E., Gochin, P., & Gross, C. (1993) Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. Brain Research, 616, 25-29

Milner, A. D. & Goodale, M. A. (1995) The Visual Brain In Action. Oxford University Press.

Mitchell, J. F., Stoner, G. R., Fallah, M., & Reynolds, J. H. (2003) Attentional selection of superimposed surfaces cannot be explained by modulation of the gain of color channels. Vision Research, 43, 1323-1325

Mitchell, J. & Zipser, D. (2001) A model of visual-spatial memory across saccades. Vision Research, 41, 1575-1592

Moore, T. (1999) Shape representation and visual guidance of saccadic eye movements. Science, 285, 1914-1917

Moore, T. & Armstrong, K. M. (2003) Selective gating of visual signals by microstimulation of frontal cortex. Nature, 421, 370-373

Moran J. & Desimone R. (1985) Selective attention gates visual processing in the extrastriate cortex. Science, 229, 782-784

Mort, D.J., Perry, R.J., Mannan, S.K., Hodgson, T.L., Anderson, E., Quest, R., McRobbie, D., McBride, A., Husain, M. & Kennard, C. (2003) Differential cortical activation during voluntary and reflexive saccades in man. Neuroimage, 18(2), 231-246

Motter, B.C. (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. Journal of Neurophysiology, 70(3), 909-919

Motter, B.C. (1994a) Neural correlates of attentive selection for color or luminance in extrastriate area V4. The Journal of Neuroscience, 14(4), 2178-2189

Motter, B.C. (1994b) Neural correlates of feature selective memory and pop-out in extrastriate area V4. The Journal of Neuroscience, 14(4), 2190-2199

Motter, B.C. & Belky, E.J. (1998a) The zone of focal attention during active visual search. Vision Research, 38(7), 1007-1022

Motter, B.C., & Belky, E.J. (1998b) The guidance of eye movements during active visual search. Vision Research, 38(12), 1805-1815

Motter, B.C., & Holsapple, J.W. (2000) Cortical image density determines the probability of target discovery during active search. Vision Research, 40(10-12), 1311-1322

Mounts, J. R. W., & Melara, R. D. (1999) Attentional selection of objects or features: Evidence from a modified search task. Perception & Psychophysics, 61(2), 322-341

Muller, J.R., Philiastides, M.G. & Newsome, W.T. (2005) Microstimulation of the superior colliculus focuses attention without moving the eyes. Proceedings of the National Academy of Science, U.S.A., 102(3), 524-529

Nakamura, H., Gattass, R., Desimone, R. & Ungerleider, L. G. (1993) The modular organization of projections from areas V1 and V2 to areas V4 and TEO in macaques. The Journal of Neuroscience,13(9), 3681-3691

Navalpakkam, V. & Itti, L. (2005) Modeling the influence of task on attention. Vision Research, 45(2):205-31

Neisser, U. (1967) Cognitive Psychology. Appleton-Century-Crofts

Niebur, E., Koch, C., & Rosin, C. (1993) An oscillation-based model for the neuronal basis of attention. Vision Research, 33, 2789-2802

Niebur, E., Itti, L., & Koch, E. (2001) Controlling the focus of visual selective attention. In: Models of Neural Networks IV (van Hemmen, J.L., Cowan, J., & Domany, E. Eds), 247-276, New York, Springer Verlag

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999) fMRI evidence for objects as the units of attentional selection. Nature, 401, 584-587

Ohzawa, I., Sclar, G. & Freeman, R.D. (1985) Contrast gain control in the cat's visual system. Journal of Neurophysiology, 54(3):651-67

Olshausen, B.A., Anderson, C.H., & Van Essen, D.C. (1993) A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. The Journal of Neuroscience, 13, 4700-4719

Paus, T., Marrett, S., Worsley, K.J. & Evans, A.C. (1995) Extraretinal modulation of cerebral blood flow in the human visual cortex – implications for saccadic suppression. Journal of Neurophysiology, 74(5), 2179-2183

Pasupathy, A. & Connor, C. E. (1999) Responses to contour features in macaque area V4. Neurophysiology, 82(5), 2490-2502

Pinilla, T., Cobo, A., Torres, K., & Valdes-Sosa, M. (2001) Attentional shifts between surfaces: Effects on detection and early brain potentials. Vision Research, 41, 1619-1630

Platt, M.L. & Glimcher, P.W. (1999) Neural correlates of decision variables in parietal cortex. Nature, 400, 233-238

Posner, M. I. (1980) Orienting of attention. Quarterly Journal of Experimental Psychology, 32, 3-26

Posner, M., & Peterson, S.E. (1990) The attention system of the human brain. Annual Review of Neuroscience, 13, 25-42

Posner, M. I., Walker, J.A., Friedrich, F.J. & Rafal, R. D. (1984) Effects of parietal injury on covert orienting of attention. The Journal of Neuroscience, 4, 1863-1874

Quaia, C., Optican, L. M., & Goldberg, M. E. (1998) The maintenance of spatial accuracy by the perisaccadic remapping of visual receptive fields. Neural Networks, 11(7-8), 1229-1240

Rainer, G., Asaad, W.F. & Miller, E.K. (1998) Selective representation of relevant information by neurons in the primate prefrontal cortex. Nature, 393, 577-579

Rainer, G. & Miller, E. K. (2002) Timecourse of object-related neural activity in the primate prefrontal cortex during a short-term memory task. European Journal of Neuroscience, 15(7), 1244-1254

Ramcharan, E.J., Gnadt, J.W., & Sherman, S.M.(2001) The effects of saccadic eye movements on the activity of geniculate relay neurons in the monkey. Visual Neuroscience, 18(2), 253-258

Rao, S. C., Rainer, G., & Miller, E. K. (1997) Integration of what and where in the primate prefrontal cortex. Science, 276, 821-824

Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C. & Driver, J. (2000) Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. Brain, 123, 1624-1633

Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C. & Driver, J. (2002) Neural correlates of conscious and unconscious vision in parietal extinction. Neurocase, 5(8), 387-393

Reicher, G. M., Snyder, C. R. R. & Richards, J. T. (1976) Familiarity of background characters in visual scanning. Journal of Experimental Psychology: Human Perception and Performance, 2, 522-530

Remington, R. W., Johnson, J. C., & Yantis, S. (1992) Involuntary attentional capture by abrupt onsets. Perception & Psychophysics, 51(3), 279-290

Renart, A., Moreno, R., de la Rocha, J., Parga, N., & Rolls, E.T. (2001) A model of the IT-PF network in object working memory which balanced persistent activity and tuned inhibition. Neurocomputing, 38-40, 1525-1531

Reppas, J.B., Usrey, W.M., & Reid, R.C. (2002) Saccadic eye movements modulate visual responses in the lateral geniculate nucleus. Neuron, 35(5), 961-974

Reynolds, J. H., Alborzian, S, Stoner, G. R. (2003) Exogenously cued attention triggers competitive selection of surfaces. Vision Research, 43(1), 59-66

Reynolds, J. H., Chelazzi, L. & Desimone, R. (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. The Journal of Neuroscience, 19(5), 1736-1753

Richards, J. T. & Reicher, G. M. (1978) The effect of background familiarity in visual search. Perception & Psychophysics, 23, 499-505

Riesenhuber, M. & Poggio, T. (1999) Hierarchical models of object recognition in cortex. Nature Neuroscience, 2(11), 1019-1025

Riesenhuber, M. & Poggio, T. (2000) Models of object recognition. Nature Neuroscience Supplement, 3, 1199-1204

Rizzolatti, G., Riggio, L., Dascola, I. & Umilta C. (1987) Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. Neuropsychologia, 25(1A), 31-40

Robinson, D. L., Bowman, E. M. & Kertzman, C. (1995) Covert orienting of attention in macaques. II. Contributions of parietal cortex. Journal of Neurophysiology, 74(2), 698-712

Robinson, D.L. & Petersen, S.E. (1992) The pulvinar and visual salience. Trends in Neuroscience, 15(4):127-32

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998) Object-based attention in the primary visual cortex of the macaque monkey. Nature, 395, 376-381

Rolls, E.T., Aggelopoulos, N.C. & Zheng, F. (2003) The receptive fields of inferior temporal cortex neurons in natural scenes. The Journal of Neuroscience, 23(1), 339-348

Rolls, E. & Deco, G. (2002) Computational Neuroscience of Vision, Oxford, Oxford University Press

Rolls, E.T., Xiang, J.Z. & Franco, L. (2005) Object, space, and object-space representations in the primate hippocampus. Journal of Neurophysiology, Mar 23 e-publication ahead of print

Romo, R., Brody, C.D., Hernandez, A. & Lemus, L. (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. Nature, 399(6735), 470-473

Ross, J., Morrone, M.C. & Burr, D.C. (1997) Compression of visual space before saccades. Nature, 386(6625), 598-601

Ross, J., Morrone, M.C., Goldberg, M.E. & Burr, D.C. (2001) Changes in visual perception at the time of saccades. Trends in Neurosciences, 24(2), 113-121

Rybak, I. A., Gusakova, V. I., Golovan, A. V., Podladchikova, L. N., & Shevtsova, N. A. (1998) A model of attention-guided visual perception and recognition. Vision Research, 38, 2387-2400

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002) Global effects of feature-based attention in human visual cortex. Nature Neuroscience, 5(7), 631-632

Sapir, A., Soroker, N., Berger, A. & Henik, A. (1999) Inhibition of return in spatial attention: Direct evidence for collicular generation. Nature Neuroscience, 2, 1053-1054

Sapir, A., Rafal, R. & Henik A. (2002) Attending to the thalamus: inhibition of return and nasal-temporal asymmetry in the pulvinar. Neuroreport, 13(5), 693-7

Sapir, A., Soroker, N., Berger, A. & Henik, A. (1999) Inhibition of return in spatial attention: Direct evidence for collicular generation. Nature Neuroscience, 2, 1053-1054

Sato, T. & Murthy, A. (2001) Search efficiency but not response interference affects visual selection in frontal eye field. Neuron, 30, 583-591

Schall, J. D. (2002) The neural selection and control of saccades by the frontal eye field. Philosophical Transactions of the Royal Society of London B Biological Science, 357, 1073-1082

Schein, S.J. & Desimone, R. (1990) Spectral properties of V4 neurons in the macaque. The Journal of Neuroscience, 10(10), 3369-89

Scheinberg, D.L. & Logothetis, N.K. (2001) Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. The Journal of Neuroscience, 21(4), 1340-1350

Schiller, P.H. & Lee, K. (1991) The role of the primate extrastriate area V4 in vision. Science, 251(4998),1251-1253

Schmidt, W.C. (1996) 'Inhibition of return' without visual input. Neuropsychologia, 34(10), 943-952

Schneider, W.X. (1999) Visual-spatial working memory, attention, and scene representation: A neuro-cognitive theory. Psychological Research, 62, 220-236

Scialfa, C.T., & Joffe, K. M. (1998) Response times and eye movements in feature and conjunction search as a function of target eccentricity. Perception & Psychophysics, 60(6), 1067-1082

Scudder, C.A., Kaneko, C.R.S. & Fuchs, A.F. (2002) The brainstem burst generator for saccadic eye movements: A modern synthesis. Experimental Brain Research, 142, 439-462

Serences, J.T., Schwarzbach, J., Courtney, S.M., Golay, W. & Yantis, S. (2004) Control of object-based attention in human cortex. Cerebral Cortex, 14, 1346-1357

Sereno, A.B. & Maunsell, J.H.R. (1995) Spatial and shape selective sensory and attentional effects in neurons in the macaque lateral intraparietal cortex (LIP). Investigative Ophthalmology & Visual Science – Supplement, 36(4), S692, 3178-3165

Shafritz, K. M., Gore, J. C., & Marois, R. (2002) The role of the parietal cortex in visual feature binding. Proceedings of the National Academy of Science U.S.A., 99(16), 10917-10922

Shen, J., Reingold, E. M., & Pomplum, M. (2000) Distractor ratio influences patterns of eye movements during visual search. Perception, 29(2), 241-250

Shen, J., Reingold, E. M., & Pomplum, M. (2003) Guidance of eye movements during conjunctive visual search: the distractor ratio effect. Canadian Journal of Experimental Psychology, 57(2), 76-96

Shepherd, M., Findlay, J.M. & Hockey, R.J. (1986) The relationship between eye movements and spatial attention. The Quarterly Journal of Experimental Psychology, 38A, 475-491

Shimozaki, S.S., Hayhoe, M.M., Zelinsky, G.J., Weinstein, A., Merigan, W.H. & Ballard, D.H. (2003) Effect of parietal lobe lesions on saccade targeting and spatial memory in a naturalistic visual search task. Neuropsychologia, 41(10), 1365-86

Shipp, S. (2004) The brain circuitry of attention. Trends in Cognitive Sciences, 8(5), 223-229

Shipp, S. & Zeki, S. (1995) Segregation and convergence of specialised pathways in macaque monkey visual cortex. Journal of Anatomy, 187(3), 547-562

Shulman, G. L., Ollinger, J. M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Petersen, S. E. & Corbetta, M. (1999) Areas involved in encoding and applying directional expectations to moving objects. The Journal of Neuroscience, 19(21), 9480-9496

Singer, W. & Gray, C. M. (1995) Visual feature integration and the temporal correlation hypothesis. Annual Review of Neuroscience, 18, 55-586

Sireteanu R. & Rettenbach R. (2000) Perceptual learning in visual search generalizes over tasks, locations, and eyes. Vision Research, 40(21), 925-49

Smith, D.R.R. & Derrington, A.M. (1996) What is the denominator for contrast normalisation. Vision Research, 36(23), 3759-3766

Snyder, J. J. & Kingstone, A. (2000) Inhibition of return and visual search: How many separate loci are inhibited? Perception & Psychophysics, 62(3), 452-458

Snyder, J.J. & Kingstone, A. (2001) Inhibition of return at multiple locations in visual search: When you see it and when you don't. Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 54(4), 1221-1237

Sobel, K.V. & Cave, K.R. (2002) Roles of salience and strategy in conjunction search. Journal of Experimental Psychology: Human Perception and Performance, 28(5), 1055-1070

Somers, D. C., Dale, A. M., Seiffert, A. E. & Tootell, R. B. H. (1999) Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. Proceedings of the National Academy of Science U.S.A., 96, 1663-1668

Spelke, E. S., Breinlinger, K., Jacobson, K., & Phillips, A. (1993) Gestalt relations and object perception: A developmental study. Perception, 22, 1483-1501

Spelke, E. S., Breinlinger, K., Macomber, J. & Jacobson, K. (1992) Origins of knowledge. Psychological Review, 99, 605-632

Spelke, E. S., Kestenbaum, R., Simons, D., & Wein, D. (1995) Spatiotemporal continuity, smoothness of motion and object identity in infancy. British Journal of Developmental Psychology, 13, 113-142

Their, P. & Andersen, R.A. (1998) Electrical microstimulation distinguishes distinct saccade-related areas in the posterior parietal cortex. Journal of Neurophysiology, 80, 1713-1735

Thiele, A., Henning, P., Kubischik, M. & Hoffmann, K.P. (2002) Neural mechanisms of saccadic suppression, Science, 295(5564), 2460-2462

Thompson, K.G. & Schall, J.D. (2000) Antecedents and correlates of visual detection and awareness in macaque prefrontal cortex. Vision Research, 40(10-12), 1523-38

Tipper, S. P., Driver, J. & Weaver, B. (1991) Object-centred inhibition of return of visual attention. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 43A, 289-298

Tipper, S. P., Weaver, B. & Watson, F. L. (1996) Inhibition of return to successively cued spatial locations: Commentary on Pratt and Abrams (1995). Journal of Experimental Psychology: Human Perception and Performance, 22(5), 1289-1293

Tolias, A. S., Moore, T., Smirnakis, S. M., Tehovnik, E. J., Siapas, A. G., Schiller, P. H. (2001) Eye movements modulate visual receptive fields of V4 neurons. Neuron, 29, 757-767

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999) Top-down signal from prefrontal cortex in executive control of memory retrieval. Nature, 401(6754), 699-703

Toth, L. J. & Assad, J. A. (2002) Dynamic coding of behaviourally relevant stimuli in parietal cortex. Nature, 415, 165-168

Toyama, K., Komatsu, Y., Yamamoto, N., Kurotani, T., & Yamada, K. (1991) In vitro approach to visual cortical development and plasticity. Neuroscience Research, 12(1), 57-71

Trappenberg, T. P., Dorris, M. C., Munoz, D. P. & Klein, R. M. (2001) A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. Journal of Cognitive Neuroscience, 13, 256-271

Treue, S., & Martinez Trujillo, J. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. Nature, 339(6736), 575-579

Treisman, A. (1982) Perceptual grouping and attention in visual search for features and for objects. Journal of Experimental Psychology: Human Perception and Performance, 8, 194-214

Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. Quarterly Journal of Experimental Psychology, 40, 201-237

Treisman, A. (1991) Search, similarity, and integration of features between and within dimensions. Journal of Experimental Psychology: Human Perception and Performance, 17(3), 652-676

Treisman, A. (1992) Spreading suppression or feature integration? A reply to Duncan and Humphreys (1992). Journal of Experimental Psychology: Human Perception and Performance, 18(2), 589-593

Treisman, A. (1998) Feature binding, attention and object perception. Philosophical Transactions of the Royal Society of London B, 353, 1295-1306

Treisman, A. & Gelade, G. (1980) A feature integration theory of attention. Cognitive Psychology, 12, 97-136

Treisman, A. & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. Psychological Review, 95, 15-48.

Tsotsos, J.K. (1993) An inhibitory beam for attentional selection. (1993) In: Spatial Vision: Proceedings of the 1991 York Conference on Spatial Vision in Humans and Robots (Harris, L., & Jenkin, M. eds), 313-331, Cambridge University Press, Cambridge, UK

Tsotsos, J.K., Culhane, S.M., Wai, W.Y.K., Lai, Y., Davis, N., & Nuflo, F. (1995) Modeling visual attention via selective tuning. Artificial Intelligence, 78, 507-545

Ungerleider, L. G. (1995) Functional brain imaging studies of cortical mechanisms for memory. Science, 270, 769-775

Ungerleider, L.G., & Mishkin, M. (1982) Two cortical visual systems. In: Analysis of Visual Behaviour (Ingle, D.J., Goodale, M.A., & Mansfield, R.W.J. Eds), 549-586, MA, MIT Press

Usher, M., & Niebur, E. (1996) Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. Journal of Cognitive Neuroscience, 8(4), 311-327

Valdes-Sosa, M., Bobes, M.A., Rodriguez, V., & Pinilla, T. (1998) Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. Journal of Cognitive Neuroscience, 10(1), 137-151

Valdes-Sosa, M., Cobo, A. & Pinilla, T. (2000) Attention to object files defined by transparent motion. Journal of Experimental Psychology: Human Perception & Performance, 26(2), 488-505

van der Velde, F. & de Kamps, M. (2001) From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation. Journal of Cognitive Neuroscience, 13(4), 479-491

van Zoest, W., Donk, M. & Theeuwes, J. (2004) The role of stimulus-driven and goal-driven control in saccadic visual selection. Journal of Experimental Psychology: Human Perception & Performance, 30(4), 746-59

Vickery, T.J., King, L-W. & Jiang, Y. (2005) Setting up the target template in visual search. Journal of Vision, 5(1), 81-92; http://journalofvision.org/5/1/8/, doi:10.1167/5.1.8.

Vuilleumier, P., Sagiv, N., Hazeltine, E., Poldrack, R.A., Swick, D., Rafal, R.D. & Gabrieli, J.D.E. (2001) Neural fate of seen and unseen faces in visuospatial neglect: A combined event-related functional MRI and event-related potential study. Proceedings of the National Academy of Science of the USA, 98(6), 3495-3500

Walker, R., Findlay, J.M., Young, A.W. & Lincoln, N.B. (1996) Saccadic eye movements in object-based neglect. Cognitive Neuropsychology, 13(4), 569-615

Walker, R. & Young, A.W. (1996) Object-based neglect: An investigation of the contributions of eye movements and perceptual completion. Cortex, 32(2), 279-295

Wallis, G., & Rolls, E. T. (1997) Invariant face and object recognition in the visual system. Progress in Neurobiology, 51, 167-194

Wang, Q., Cavanagh, P. & Greene, M. (1994) Familiarity and pop-out in visual search. Perception & Psychophysics, 56(5), 495-500

De Weerd, P., Peralta, M.R. 3rd, Desimone, R., Ungerleider, L.G. (1999) Loss of attentional stimulus selection after extrastriate cortical lesions in macaques. Nature Neuroscience, 2(8), 753-758

Wilcox, T. (1999) Object individuation: infants' use of shape, size, pattern, and color. Cognition, 72(2),125-66

Williams, D. E., & Reingold, E. M. (2001) Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude. Psychonomic Bulletin & Review, 8(3), 476-488

Williams, L.G. (1967) The effects of target specification on objects fixated during visual search. Acta Psychologica, 27, 355-360

Wilson, H., & Cowan, J. (1972) Excitatory and inhibitory interaction in localized populations of model neurons. Biophysics Journal, 12, 1-24

Wolfe, J. M. (1994) Guided Search 2.0: a revised model of visual search. Psychonomic Bulletin & Review, 1, 202-238

Wolfe, J. M. & Cave, K. R. (1989) Deploying visual attention: The guided search model. In T. Troscianko & A. Blake (Eds), AI and the eye (pp. 7910). Chichester, U.K. Wiley

Wolfe, J. M., Cave, K. R. & Franzel, S. L. (1989) Guided search: An alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human Perception & Performance, 15, 419-433

Wolfe, J. M., Horowitz, T.S., Kenner, N., Hyle, M. & Vasan, N. (2004) How fast can you change your mind? The speed of top-down guidance in visual search. Vision Research, 44, 1411-1426

Woodman, G.F., Vogel, E.K. & Luck, S.J. (2001) Visual search remains efficient when visual working memory is full. Psychological Science,12(3),219-24

Xu, F. (1999) Object individuation and object identity in infancy: The role of spatiotemporal information, object property information, and language. Acta Psychologica, 102, 113-136

Zelinsky, G.J. (1996) Using eye saccades to assess the selectivity of search movements. Vision Research, 36(14), 2177-2187

Zelinsky, G.J., Dickinson, C.A. & Chen, X. (2003) Evidence for memory and order in visual search tasks. Abstracts of the Psychonomic Society, 44th Annual Meeting, 218

Zeki, S. (1993) A Vision of the Brain. Blackwell Scientific Publications, Oxford, UK

Zhu, J.J., & Lo, F.S. (1995) Physiological properties of the output neurons in the deep layers of the superior colliculus of the rabbit. Brain Research Bulletin, 38(5), 495-505

Zhu, J.J., & Lo, F.S. (1996) Time course of inhibition induced by a putative saccadic suppression circuit in the dorsal lateral geniculate nucleus of the rabbit. Brain Research Bulletin, 41(5), 281-291

Zhu, J.J., & Lo, F.S. (1998) Control of recurrent inhibition of the lateral posterior-pulvinar complex by afferents from the deep layers of the superior colliculus of the rabbit. Journal of Neurophysiology, 80(3), 1122-1131

Zipser, D., & Andersen, R. A. (1988) A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. Nature, 331(6158), 679-684

# Selected Published Papers

This section contains the following papers:

- Lanyon, L. J., & Denham, S.L. (2004) A model of active visual search with object-based attention guiding scan paths. **Neural Networks (Special Issue: Vision & Brain)**, 17(5-6), 873-897

- Lanyon, L. J., & Denham, S.L. (2004) A biased competition computational model of spatial and object-based attention mediating active visual search. **Neurocomputing (Special Issue: Computational Neuroscience: Trends in Research 2004)**, 58-60C, 655-662

- Lanyon, L. J., & Denham, S.L. (2005) A model of object-based attention that guides active visual search to behaviourally relevant locations. **Lecture Notes in Computer Science (Special Issue: WAPCV 2004; L Paletta et al., Eds)**, 3368, 42-56 [Publisher's reprint not yet available]

- Lanyon, L.J. & Denham, S.L. (2005) A model of spatial & object-based attention for active visual search. **Modelling Language, Cognition and Action: Proceedings of the 9th Neural Computation & Psychology Workshop. World Scientific, Progress in Neural Processing**, 16, 239-248 [Publisher's reprint not available]

- Lanyon, L.J. & Denham, S.L. (2005) A model of object-based attention that guides active visual search to behaviourally relevant locations. **Proceedings of the 2nd International Workshop on Attention & Performance in Computer Vision**, Prague, Czech Republic [Publisher's reprint not available]

Available online at www.sciencedirect.com
SCIENCE DIRECT
Neural Networks 17 (2004) 873–897
Neural Networks
www.elsevier.com/locate/neunet

# A model of active visual search with object-based attention guiding scan paths

Linda J. Lanyon*, Susan L. Denham

*Centre for Theoretical and Computational Neuroscience, University of Plymouth, Drakes Circus, Plymouth, Devon PL4 8AA, UK*

Received 7 October 2003; revised 30 March 2004; accepted 30 March 2004

**bstract**

When a monkey searches for a colour and orientation feature conjunction target, the scan path is guided to target coloured locations in reference to locations containing the target orientation [Vision Res. 38 (1998b) 1805]. An active vision model, using biased competition, is le to replicate this behaviour. As object-based attention develops in extrastriate cortex, featural information is passed to posterior parietal rtex (LIP), enabling it to represent behaviourally relevant locations [J. Neurophysiol. 76 (1996) 2841] and guide the scan path. Attention olves from an early spatial effect to being object-based later in the response of the model neurons, as has been observed in monkey single ll recordings. This is the first model to reproduce these effects with temporal precision and is reported here at the systems level allowing the plication of psychophysical scan paths.


*ywords:* Visual attention; Biased competition; Active visual search; Mean field population approach

## Introduction

### *l. Biased competition*

Vision attention literature has been strongly influenced cently by the biased competition hypothesis (Desimone, 98; Desimone & Duncan, 1995; Duncan & Humphreys, 89; Duncan, Humphreys, & Ward, 1997). There is much idence from monkey single cell recordings to support the pothesis (Chelazzi, Miller, Duncan, & Desimone, 1993, 01; Miller, Gochin, & Gross, 1993; Moran & Desimone, 85; Motter, 1993, 1994a,b; Reynolds, Chelazzi, & simone, 1999) and, gradually, models have been veloped to test the theory computationally. Early models ded to be small scale with only a few interacting units eynolds et al., 1999; Usher & Niebur, 1996). More cently, systems level models have begun to emerge (Deco Lee, 2002; De Kamps & Van der Velde, 2001; Hamker, 98). However, there has been no systems level modelling the biased competition hypothesis for active visual search, where retinal inputs change as the focus of attention is shifted. This model addresses this issue; replicating attentional effects in inferior temporal (IT) region and extrastriate visual area 4 (V4), with temporal precision, to qualitatively reproduce scan paths recorded from monkeys carrying out active search for a feature conjunction target. The guidance of scan paths is carried out by a novel object-based 'cross-stream' interaction between extrastriate areas in the ventral pathway leading to temporal cortex and the dorsal pathway leading to parietal cortex (Milner & Goodale, 1995; Ungerleider & Mishkin, 1982).

Biased competition suggests that neuronal responses are determined by competitive interactions that are subject to a number of biases, such as 'bottom-up' stimulus information and 'top-down' cognitive requirements. Important in the theory is the idea that a working memory template of the target object can bias competition between objects and features such that the target object is given a competitive advantage and other objects are suppressed. At the single cell level, studies in IT (Chelazzi et al., 1993; Miller et al., 1993) and V4 (Chelazzi et al., 2001; Moran & Desimone, 1985; Motter, 1993, 1994a,b; Reynolds et al., 1999) show that a high response to a preferred stimulus (i.e. a stimulus that causes a strong response from the cell) presented in

Corresponding author. Tel.: +44-1752-2325-67; fax: +44-1752-2333-

*E-mail address:* linda.lanyon@plymouth.ac.uk (L.J. Lanyon).

3-6080/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
10.1016/j.neunet.2004.03.012

the cell's receptive field is attenuated by the addition of a second non-preferred stimulus (i.e. a stimulus that causes only a weak response when presented alone in the receptive field). Responses are eventually determined by which of the two stimuli are attended. When the preferred stimulus is attended, the response approaches that when this stimulus is presented alone. If the non-preferred stimulus is attended, responses are severely suppressed, despite the presence of the preferred stimulus in the receptive field. Neurons from modules representing areas IT and V4 from the model presented here have been able to replicate such effects at the cellular level (Lanyon & Denham, submitted). Here, we examine the systems level behaviour of the model in more detail when replicating the nature of search scan paths observed by Motter and Belky (1998b) who found most fixations landed within 1° of stimuli (only 20% fell in blank areas of the display despite the use of very sparse displays) and these stimuli tended to be target coloured (75% of fixations landed near target coloured stimuli and only 5% near non-target coloured stimuli).

### *1.2. Visual attention*

There has been debate over the issue of whether visual attention operates purely as a spatial 'spotlight' (Crick, 1984; Helmholtz, 1867; Treisman, 1982) or is more complex, operating in an object-based manner. The evidence for object-based attention has been convincing and growing in the psychophysical literature (Blaser, Pylyshyn, & Holcombe, 2000; Duncan, 1984), from functional magnetic resonance imaging (O'Craven, Downing, & Kanwisher, 1999), event-related potential recordings (Valdes-Sosa, Bobes, Rodriguez, & Pinilla, 1998; Valdes-Sosa, Cobo, & Pinilla, 2000) and from single cell recordings (Chelazzi et al., 1993, 2001; Roelfsema, Lamme, & Spekreijse, 1998). However, there is no doubt that attention can produce spatially specific effects also (Bricolo, Gianesini, Fanini, Bundesen, & Chelazzi, 2002; Connor, Callant, Preddie, & Van Essen, 1996). In the lateral intraparietal area (LIP), an anticipatory spatial enhancement of responses has been recorded from single cells (Colby, Duhamel, & Goldberg, 1996) and has been seen in imaging of the possible human homologue of LIP (Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000; Hopfinger, Buonocore, & Mangun, 2000; Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999). Spatial effects have been recorded in single cells in area V4 in advance of the sensory response and have then modulated the earliest stimulus-invoked response at 60 ms post-stimulus (Luck, Chelazzi, Hillyard, & Desimone, 1997). However, object-based effects have not been recorded until much later in the response, from ~150 ms in IT and V4 (Chelazzi et al., 1993, 2001).

The model presented here has been used (Lanyon & Denham, submitted) to suggest that spatial attention is available earlier than object-based attention, at least in area V4, because the latter relies on the resolution of competition between objects in IT. The model was able to combine spatial and object-based attention at both the single cell and systems level in order to reproduce attention effects seen in single cells in V4 (Chelazzi et al., 2001; Luc et al., 1997) and IT (Chelazzi et al., 1993) with temporal accuracy. Here, we use these attentional effects to produ biologically plausible active vision behaviour.

There is evidence to suggest that an eye movement m be linked with a spatial enhancement of responses in ar V4 because microstimulation of the frontal eye field (FE which is involved in the allocation of attention and eyes locations in the scene (Schall, 2002), results in responses V4 that are spatially modulated (Moore & Armstron 2003). In addition to anticipatory spatial increases activity being found in LIP (Colby et al., 1996), poster parietal cortex in general is implicated in the control of bo spatial and object-based attention (Corbetta et al., 200 Corbetta, Shulman, Miezin, & Petersen, 1995; Fink, Dol Halligan, Marshall, & Frith, 1997; Hopfinger et al., 200 Martinez et al., 1999; Posner, Walker, Friedrich, & Raf 1984; Robinson, Bowman, & Kertzman, 1995). Therefo the model assumes that FEF provides a spatial bias to V4 circuitry in LIP. Thus, a spatial bias is applied directly fr FEF to LIP, and LIP then biases V4. The source of the b to LIP could also be dorsolateral prefrontal cortex, whi has connections with parietal cortex (Blatt, Andersen, Stoner, 1990), or pulvinar.

### *1.3. Visual search behaviour*

When a visual target contains a simple feature that absent from distractors, it tends to effortlessly 'pop-o from the scene. However, when a target is defined by conjunction of features, the search takes longer and appe to require a serial process, which has been suggested to the serial selection of spatial locations to which attentio allocated (Treisman, 1982; Treisman & Gelade, 1980).

Active visual search involves the movement of eyes, it is presumed attention (Hoffman & Subramaniam, 199 to locations to be inspected. The resultant series of poi where the eyes fixate form a scan path. This differs from more commonly modelled covert search where the attent focus is shifted but eye position and, hence, retinal input held constant. During active search for a feature conjunct target, it seems that colour (or luminance) is more influen on the scan path than form features, such as orientation monkeys (Motter & Belky, 1998b), and in humans (Sci & Joffe, 1998; Williams & Reingold, 2001). When numbers of each distractor type are equal, this preference target coloured locations over locations containing target orientation seems robust, even when the task is bia towards orientation discrimination (Motter & Bel 1998b). Colour appears to segment the scene and gu the scan path in a manner that resembles *guided sea* (Wolfe, 1994; Wolfe, Cave, & Franzel, 1989). Even ab

nsets can be ineffective in overriding a top-down ttentional set for colour and capturing attention in an xogenous manner (Folk & Remington, 1999). However, hen distractor types are not in equal proportion in the isplay, saccadic selectivity may be biased towards the eature dimension that has the fewest distractors sharing this eature with the target (Bacon & Egeth, 1997; Shen, eingold, & Pomplum, 2000, 2003). In other words, if here are fewer distractors sharing the target's orientation an its colour, the scan path may be drawn target rientation distractors. This is known as the distractor-atio effect. Displays used in simulations here replicate ose used by Motter and Belky (1998a,b) and have equal umbers of each type of distractor.

### 4. Representation of behaviourally relevant locations

Posterior parietal area LIP encodes the locations of ehaviourally relevant features (Colby et al., 1996), perhaps cting as a 'salience map' (Colby & Goldberg, 1999; ottlieb, Kusunoki, & Goldberg, 1998; Kusunoki, Gottlieb, Goldberg, 2000). Although responses in LIP are not ependant on motor response (Bushnell, Goldberg, & obinson, 1981), it has projections to superior colliculus ynch, Graybiel, & Lobeck, 1985) and FEF (Blatt et al., 990), which are thought to be involved in generating ccades, and direct electrical stimulation of LIP elicits ccades (Their & Andersen, 1998). Thus, LIP is likely to rm part of the cortical network responsible for selecting ssible targets for saccades and, in this model, activity in P is used to decide the next location to be fixated.

In the model LIP, featural information is integrated thin a spatial competitive framework in order to represent e most behaviourally relevant locations in the scene olby et al., 1996) and guide the scan path accordingly.

## The model

The model focuses on intermediate stages of visual ocessing with attention arising as an emergent property of e dynamics of the system within modules representing V4, and LIP. This dynamic portion of the model uses mean ld population dynamics (Section A.3), where represen-ion is at the level of cell populations, known as pools or semblies. Before the dynamic portion of the model, tural 'pre-processing' is carried out by two modules presenting the retina (and lateral geniculate nucleus, LGN, the thalamus) and V1. These modules do not form part of e dynamic portion of the system, for reasons of mputational simplicity and speed, and because the aim this model is to focus on attention operating in rastriate areas and beyond. However, attention is found V1 (Brefczynski & DeYoe, 1999; Ghandi, Heeger, & ynton, 1999; Motter, 1993; Roelfsema et al., 1998; ners, Dale, Seiffert, Tootell, & Functional, 1999; but see

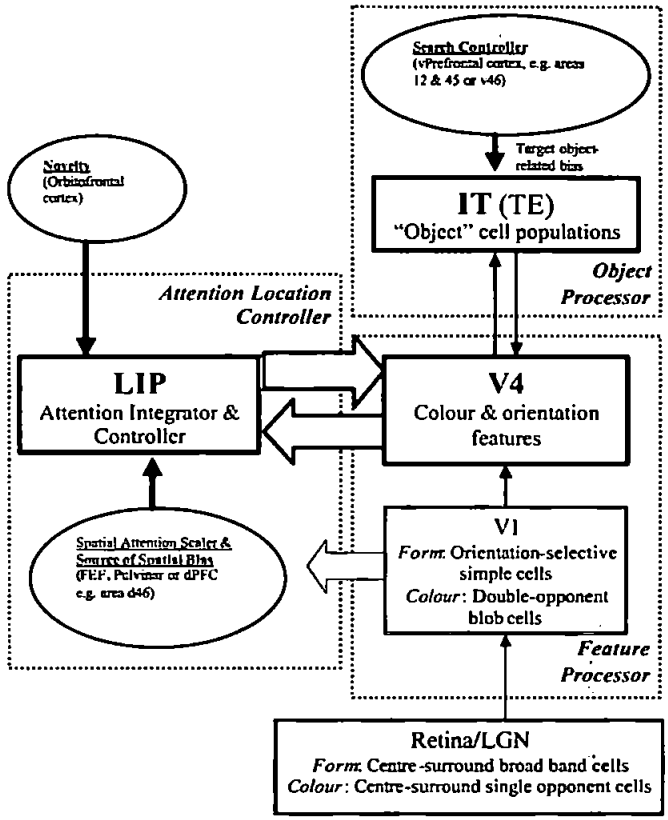



Fig. 1. An overview of the model.

Clark & Hillyard, 1996, and Martinez et al., 1999 for ERP evidence against an early attentional modulation in V1) and the V1 module could be added to the dynamic part of this model in order to replicate attentional effects therein.

An overview of the model is given in Fig. 1 and is formally defined in Appendix A. Fig. 2 shows the inhibition and biases that influence competition in each of the dynamic modules. External biases, in the form of direct current injections, from prefrontal cortex and FEF provide information relating to the target object and a spatial bias relating to the newly fixated location, respectively. These top-down biases, in addition to the bottom-up stimulus information from the retina and V1, serve to influence the competitive interactions within V4, IT and LIP.

To detect form, the retina performs simple centre-surround processing simulating retinal ganglion broadband cells, as modelled by Grossberg and Raizada (2000). A similar approach is adopted for the concentric single opponent cells that process the colour information. Within V1, simple cells detect orientation using a difference-of-offset-Gaussian approach, also described by Grossberg and Raizada (2000). This approach was found to be more accurate for detecting oriented edges than a Gabor filter. To process colour, double-opponent V1 cells are modelled, as described in Section A.2.2.

The ventral stream (V1, V4 and IT) operates as a feature/object processing hierarchy with receptive field sizes increasing towards IT and encoded features becoming

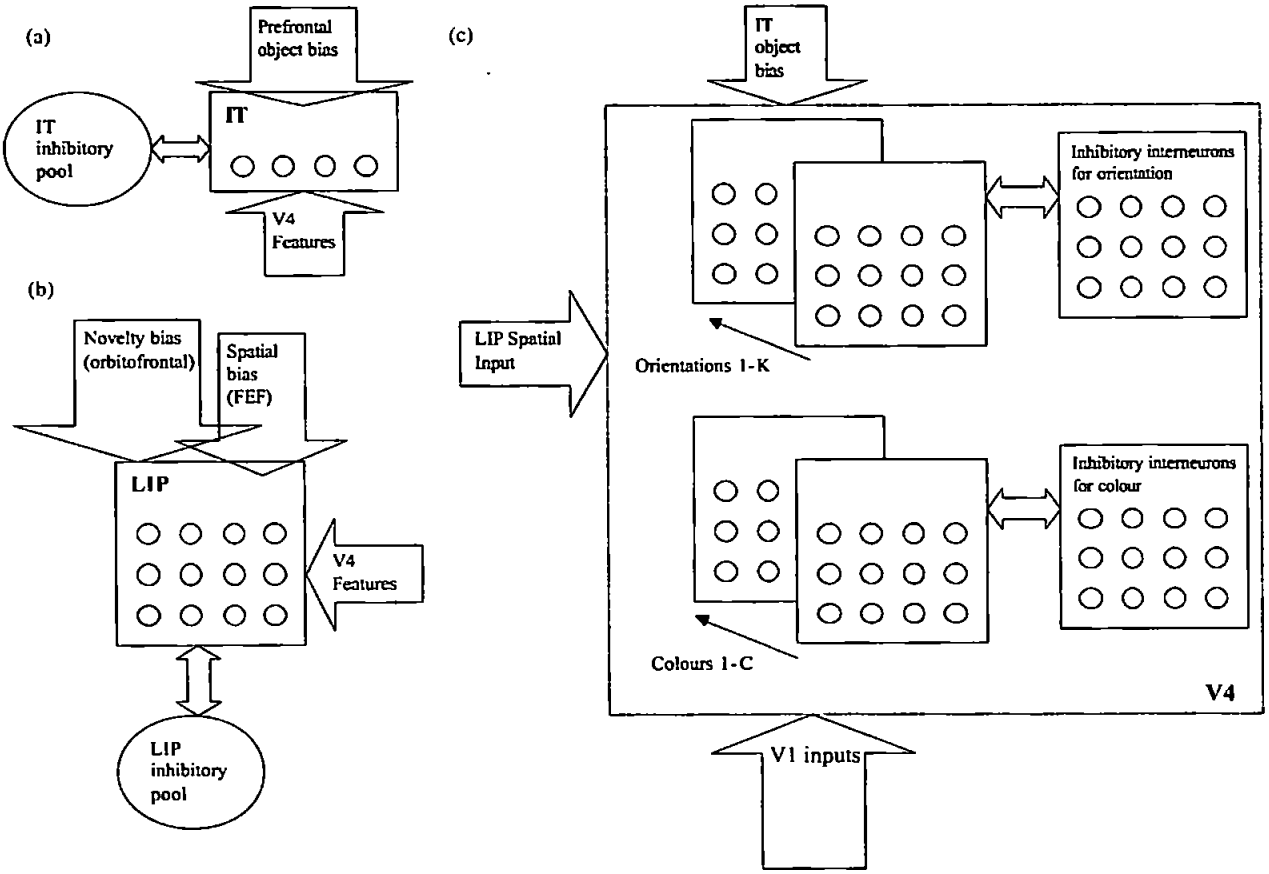



Fig. 2. The inhibition and biases that influence competition in each of the dynamic modules (a) IT; (b) LIP; (c) V4.

more complex. Anterior areas in IT, such as area TE, are not retinotopic but encode objects in an invariant manner (Wallis & Rolls, 1997) and the IT module here represents such encoding. An inhibitory interneuron pool mediates competition between objects in IT. V1 and V4 are retinotopic and process both colour and orientation. A V1 cell for every feature exists at every pixel position in the image. V4 receives convergent inputs from V1 over the area of its receptive field and V4 receptive fields overlap by one V1 neuron. V4 is arranged as a set of feature 'layers' encoding each feature in a retinotopic manner. Each feature belongs to a feature type, i.e. colour or orientation. V4 neurons are known to be functionally segregated (Ghose & Ts'O, 1997) and the area is involved in the representation of colour as well as form (Zeki, 1993).

LIP provides a retinotopic spatio-featural map that is used to control the spatial focus of attention and fixation. Locations in LIP compete with one another and the centre of the receptive field of the assembly with the highest activity is chosen as the next fixation point. LIP is able to integrate featural information in its spatial map due to its connection with area V4. Competition between locations in LIP is mediated by an inhibitory interneuron pool.

Stimuli consist of vertical and horizontal, red and green bars. During active search for an orientation and colour conjunction target, distractors differ from the target in one feature dimension only. Stimuli of this type were chosen mirror those used by Motter and Belky (1998b), in order t the active vision scan paths from this experiment could simulated. The size of the V1 orientation and colour filt described in Section A.2, determines the number of pix that represent 1° of visual angle, since V1 receptive fie tend to cover no more than about 1° (Wallis & Rolls, 199 This size is then used to scale the stimuli to be 1 × 0.25° used by Motter and Belky (1998a,b).

The model operates in an active vision manner moving its retina around the image so that its view of world is constantly changing. Most visual attention mod (Deco, 2001; Deco & Lee; 2002; Niebur, Itti, & Koch, 20 have a static retina. Here, cortical areas receive differ bottom-up input from the retina and V1 at each fixation. retinal image is the view of the scene entering the reti and cortical areas, at any particular fixation. From information within the retinal image, the system has select its next fixation point in order to move its retina. size of the retinal image is variable for any partic simulation but is normally set to 441 pixels, which equ to approximately 40° of visual angle. This is smaller t our natural vision but stable performance across a rang retinal image sizes is possible (the only restriction being a very small retina tends to lead to a higher proportio fixations landing in blank areas of very sparse scenes du

ack of stimuli within the limited retina) due to the ormalisation of inputs to IT, described in Section A.3.2. Cortical areas V4 and LIP are scaled dependant on the size f the retinal image so that larger retinal images result in ore assemblies in V4 and LIP and, thus, longer processing mes during the dynamic portion of the system. For a 40° tinal window, V4 consists of 20 × 20 assemblies in each eature layer, and processing in 5 ms steps for typical xations lasting ~240 ms (saccade onset is determined by e system) takes approximately 12 s to run in Matlab using Pentium (4) PC with a 3.06 GHz processor and 1 GB of AM. For monochromatic single cell simulations the ystem takes 0.5 s to process a 240 ms fixation at 1 ms eps using a small retinal image (covering 23 × 23 pixels).

*1. Spatial attention*

Following fixation an initially spatial attention window W) is formed. The aperture of this window is scaled cording to coarse resolution information reflecting local imulus density (Motter & Belky, 1998a), which is ssumed to be conveyed rapidly by the magnocellular thway to parietal cortex, including LIP, and other possibly volved areas, such as FEF. Alternatively, this information ay be conveyed sub-cortically, and superior colliculus or lvinar could be the source of this spatial effect in LIP and 4. All other information within the system is assumed to derived from the parvocellular pathway. Thus, during a an path, the size of the AW is dynamic, being scaled cording to stimulus density found around any particular xation point (see Lanyon and Denham (2004a) for further tails). Attention gradually becomes object-based over ne, as competition between objects and features is solved. Object-based attention is not constrained by the atial AW but is facilitated within it. Thus, object-based tention responses are strongest within the AW causing a mbined attentional effect, as found by McAdams and aunsell (2000) and Treue and Martinez Trujillo (1999).

The initial spatial AW is implemented as a spatial bias ovided from FEF to LIP, which results in a spatial tention effect in LIP that is present in anticipation of mulus information (Colby et al., 1996; Corbetta et al., 00; Hopfinger et al., 2000; Kastner et al., 1999). A nnection from LIP to V4 allows a spatial attentional effect be present in V4 as found in many studies (Connor et al., 96; Luck et al., 1997). Spatial attention in the model V4 semblies appears as an increase in baseline firing in vance of the stimulus information and as a modulatory fect on the stimulus-invoked response beginning at 60 ms, as found in single cells by Luck et al. (1997) d reported in Lanyon and Denham (submitted). Spatial ention in V4 provides a facilitatory effect to object-based ention within the AW. The excitatory connection from P to V4 also serves to provide feature binding at the olution of the V4 receptive field.

*2.2. Object-based attention*

Object-based attention operates within the ventral stream of the model (V4 and IT) in order that features belonging to the target object are enhanced and non-target features are suppressed. A connection from the retinotopic ventral stream area (V4) to the parietal stream of the model allows the ventral object-based effect to influence the representation of behaviourally relevant locations in the LIP module.

The prefrontal object-related bias to IT provides a competitive advantage to the assembly encoding the target object such that, over time, this object wins the competition in IT. Attention appears to require working memory (De Frockert, Rees, Frith, & Lavie, 2001) and, due to its sustained activity (Miller, Erickson, & Desimone, 1996), prefrontal cortex has been suggested as the source of a working memory object-related bias to IT. Other models have implemented such as bias (Deco & Lee, 2002; Renart, Moreno, de la Rocha, Parga, & Rolls, 2001; Usher & Niebur, 1996). Here, the nature of this bias resembles the responses of so-called 'late' neurons in prefrontal cortex, whose activity builds over time and tends to be highest late in a delay period (Rainer & Miller, 2002; Romo, Brody, Hernandez, & Lemus, 1999). It is modelled with a sigmoid function, as described in Section A.3.2. This late response also reflects the time taken for prefrontal neurons to distinguish between target and non-target objects, beginning 110–120 ms after the onset of stimuli at an attended location (Everling, Tinsley, Gaffan, & Duncan, 2002).

IT provides an ongoing object-related bias to V4 that allows target features to win local competitions in V4 such that these features become the most strongly represented across V4. Each assembly in IT provides an inhibitory bias to features in V4 that do not relate to the object encoded by it. As the target object becomes most active in IT, this results in suppression of non-target features in V4. These object-based effects appear in IT and V4 later in their response, from ~150 ms post-stimulus, as was found in single cell recordings (Chelazzi et al., 1993, 2001; Motter, 1994a,b). The result of object-based attention in V4 is that target features are effectively 'highlighted' in parallel across the visual field (McAdams & Maunsell, 2000; Motter, 1994a,b).

*2.3. The scan path*

In order to carry out effective visual search, brain areas involved in the selection of possible target locations should be aware of object-based effects that are suppressing non-target features in the visual scene (as represented in the retinotopic visual areas). This model suggests that object-based effects occurring in area V4 are able to influence the spatial competition for next fixation location in LIP. It is this cross-stream interaction between the ventral and dorsal visual streams (Milner & Goodale, 1995; Ungerleider & Mishkin, 1982) that allows visual search to select

appropriate stimuli for examination. Thus, object-based attention, at least at the featural level, may be crucial for efficient search.

As object-based attention becomes effective within the ventral stream, the parallel enhancement of target features across V4 results in LIP representing these locations as being salient, due to its connection from V4. The weight of connection from the V4 colour assemblies to LIP is slightly stronger than that from the V4 orientation assemblies. This gives target coloured locations an advantage in the spatial competition in LIP such that target coloured locations are represented as more behaviourally relevant and, hence, tend to be most influential in attracting the scan path, as found by Motter and Belky (1998b). The difference in strength of connection of the V4 features to LIP need only be marginal in order to achieve this effect. Fig. 3 shows the effect of adjusting the relative connection weights. The strength of these connections could be adapted subject to cognitive requirement or stimulus-related factors, such as distractor ratios (Bacon & Egeth, 1997; Shen, Reingold, & Pomplum, 2000, 2003). However, with the proportion of distractor types equal, Motter and Belky (1998b) found that orientation was unable to override colour even during an orientation discrimination task. This suggests that the bias towards a stronger colour connection could be learnt during development and be less malleable to task requirement.

### 2.4. Saccades

Single cell recordings (Chelazzi et al., 1993, 2001) provide evidence that saccade onset may be temporally linked to the development of significant object-based effects, with saccades taking place ~70–80 ms after a significant effect was observed in either IT or V4. In the model, saccades are linked to the development of a significant object-based effect in IT. The effect is deemed to be significant when the most active object assembly twice as active as its nearest rival (such a quantitati difference is reasonable when compared to the recordings Chelazzi et al., 1993) and a saccade is initiated 70 ms late reflecting motor preparation latency.

It is unclear what information is available within cort during a saccade but evidence suggests that magnocellul inputs are suppressed during this time (Anand & Bridg man, 2002; Thiele, Henning, Kubischik, & Hoffmann, 200 and see Ross, Morrone, Goldberg, and Burr (2001) and Z and Lo (1996) for reviews). In the model, saccades a instantaneous but the dynamic cortical areas (IT, LIP a V4) are randomised at the start of the subsequent fixation order to reflect saccadic suppression of magnocellular a possibly parvocellular inputs, and cortical dynamics duri the saccade.

### 2.5. Inhibition of return

As an integrator of spatial and featural informati (Colby et al., 1996; Gottlieb et al., 1998; Toth & Assa 2002), LIP provides the inhibition of saccade return (Hoo & Frens, 2000) mechanism required here to prevent the sc path returning to previously inspected sites. Inhibitory aft effects once attention is withdrawn from an area demonstrated in classic inhibition of return (IOR) studi (Clohessy, Posner, Rothbart, & Vecera, 1991; Posner et 1984) and may be due to oculomotor processes, possib linked with the superior colliculus (Sapir, Soroker, Berg & Henik, 1999; Trappenberg, Dorris, Munoz, & Kle 2001) or a suppressive after-effect of high activity a previously attended location. In a model with a static reti suppression of the most active location over time by speci IOR input (Itti & Koch, 2000; Niebur et al., 2001) through self-inhibition is possible. Here, such a proc within LIP could lead to colour-based IOR (Law, Pratt,

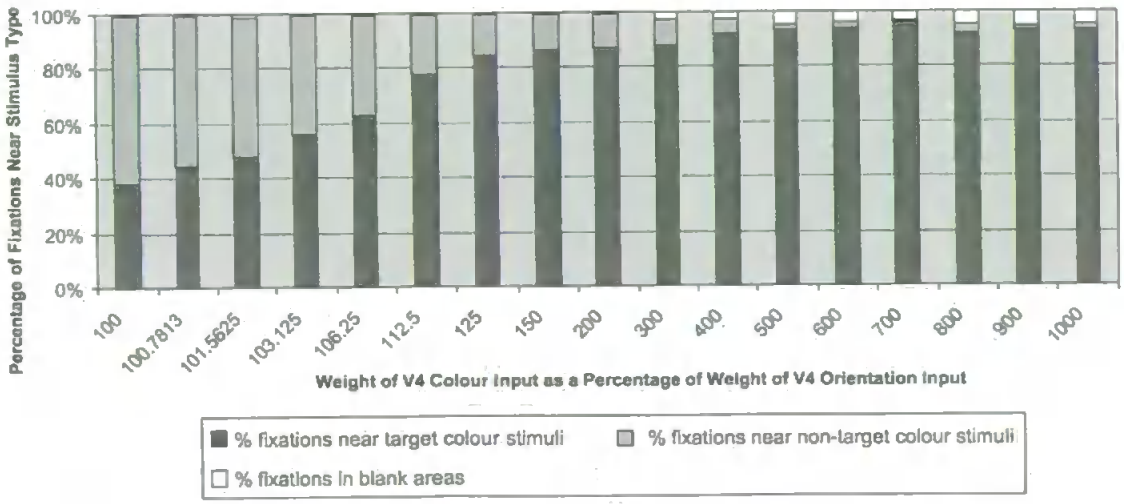



Fig. 3. The effect on fixation position of increasing the relative weight of V4 colour feature input to LIP. When V4 colour features are marginally more stro connected to LIP than V4 orientation features, the scan path is attracted to target coloured stimuli in preference to stimuli of the target orientation. Fixa positions were averaged over 10 scan paths, each consisting of 50 fixations, over the image shown in Fig. 6a.

Abrams, 1995) due to the suppression of the most active locations. When the retina is moving, such inhibition is inadequate because there is a need to remember previously visited locations across eye movements and there may be a requirement for a head- or world-centred mnemonic representation. This is a debated issue with there being some evidence to suggest that humans use very little memory during search (Horowitz & Wolfe, 1998; Woodman, Vogel, & Luck, 2001). However, even authors advocating 'amnesic search' (Horowitz & Wolfe, 1998) do not preclude the use of higher-level cognitive and mnemonic processes for efficient active search. Parietal damage is linked to the inability to retain a spatial working memory of searched locations across saccades so that locations are repeatedly re-fixated (Husain et al., 2001) and computational modelling has suggested that units with properties similar to those found in LIP could contribute to visuospatial memory across saccades (Mitchell & Zipser, 2001). Thus, it is plausible for the model LIP to be involved in IOR.

Also, IOR seems to be influenced by recent event/reward associations linked with orbitofrontal cortex (Hodgson et al., 2002). In the model, the potential reward of a location is linked to its novelty, i.e. whether a location has previously been visited in the scan path and, if so, how recently. Competition in LIP is biased by the 'novelty' of each location with the possible source of such a bias being frontal areas, such as orbitofrontal cortex. A mnemonic map of novelty values is constructed in a world- or head-centred co-ordinate frame of reference and converted into retinotopic co-ordinates when used in LIP. Initially, every location in the scene has a high novelty, but when fixation (and, thus, attention) is removed from an area, all locations that fall within the spatial AW have their novelty values reduced. At the fixation point the novelty value is set to the lowest value. In the immediate vicinity of the fixation point (the area of highest acuity and, therefore, discrimination ability, in a biological system) the novelty is set to low values that gradually increase, in a Gaussian fashion, with distance from the fixation point (Hooge & Frens, 2000). All locations that fall within the AW, but are not in the immediate vicinity of the fixation point, have their novelty set to a neutral value. Novelty is allowed to recover linearly with time. This allows IOR to be present at multiple locations, as has been found in complex scenes (Danziger, Kingstone, & Snyder, 1998; Snyder & Kingstone, 2001; Tipper, Weaver, & Watson, 1996) where the magnitude of the effect decreases approximately linearly from its largest value at the most recently searched location so that at least five previous locations are affected (Irwin & Zelinsky, 2002; Snyder & Kingstone, 2000).

The use of such a scene-based map (in world- or head-centred coordinates) reflects a simplification of processing that may occur within parietal areas, such as the ventral intraparietal area, where receptive fields range from retinotopic to head-centred (Colby & Goldberg, 1999) or area 7a, where neurons respond to both the retinal location of the stimulus and eye position in the orbit. Such a mapping may be used to keep track of objects across saccades and this concept has already been explored in neural network models (Andersen & Zipser, 1988; Mazzoni, Andersen, & Jordan, 1991; Quaia, Optican, & Goldberg, 1998; Zipser & Andersen, 1988). Representations in LIP are retinotopic such that, after a saccade the representation shifts to the new co-ordinate system based on the post-saccadic centre of gaze. Just prior to a saccade the spatial properties of receptive fields in LIP change (Ben Hamed, Duhamel, Bremmer, & Graf, 1996) and many LIP neurons respond (~80 ms) before the saccade to salient stimuli that will enter their receptive fields after the saccade (Duhamel, Colby, & Goldberg, 1992). In common with most models of this nature (Deco & Lee, 2002), such 'predictive re-mapping' of the visual scene is not modelled here. At this time, it is left to specialised models to deal with the issue of pre- and post-saccadic spatial constancy, involving changes in representation around the time of a saccade (Ross, Morrone, & Burr, 1997), memory for targets across saccades (Findlay, Brown, & Gilchrist, 2001; McPeek, Skavenski, & Nakayama, 2000) and the possible use of visual markers for co-ordinate transform (Deubel, Bridgeman, & Schneider, 1998), along with the associated issue of suppression of magnocellular (Anand & Bridgeman, 2002; Thiele, Henning, Kubischik, & Hoffmann, 2002; and see Ross, Morrone, Goldberg, and Burr (2001) and Zhu and Lo (1996) for reviews) and possibly parvocellular cortical inputs just prior to and during a saccade.

## 3. Results

### *3.1. Dynamics of object-based attention result in representation of behaviourally relevant locations in LIP*

Fig. 4 shows the activity in V4, IT and LIP at different times during the first fixation on an image. The outer box plotted on the image in Fig. 4a represents the retinal image and the inner box represents the AW. Initially, spatial attention within the AW modulates representations in LIP and V4. Object assemblies in IT are all approximately equally active. Later in the response (from ~150 ms post-fixation), object-based attention develops and the target object becomes most active in IT, whereas distractor objects are suppressed. Features belonging to the red vertical target are enhanced at the expense of the non-target features across V4 (Motter, 1994a,b). Responses in V4 are modulated by both spatial attention and object/feature attention (Anllo-Vento & Hillyard, 1996; McAdams & Maunsell, 2000). This occurs because V4 is subject to both a spatial bias, from LIP, and an object-related bias, from IT. Both inputs are applied to V4 throughout the temporal processing at each fixation. However, the spatial bias results in an earlier

spatial attention effect in V4 than the object effect because the latter is subject to the development of significant object-based attention in IT resulting from the resolution of competition therein.

Once these object-based effects are present in V4, LIP able to represent the locations that are behavioural relevant, i.e. the red coloured locations, as possible sacca targets. Prior to the onset of object-based attention

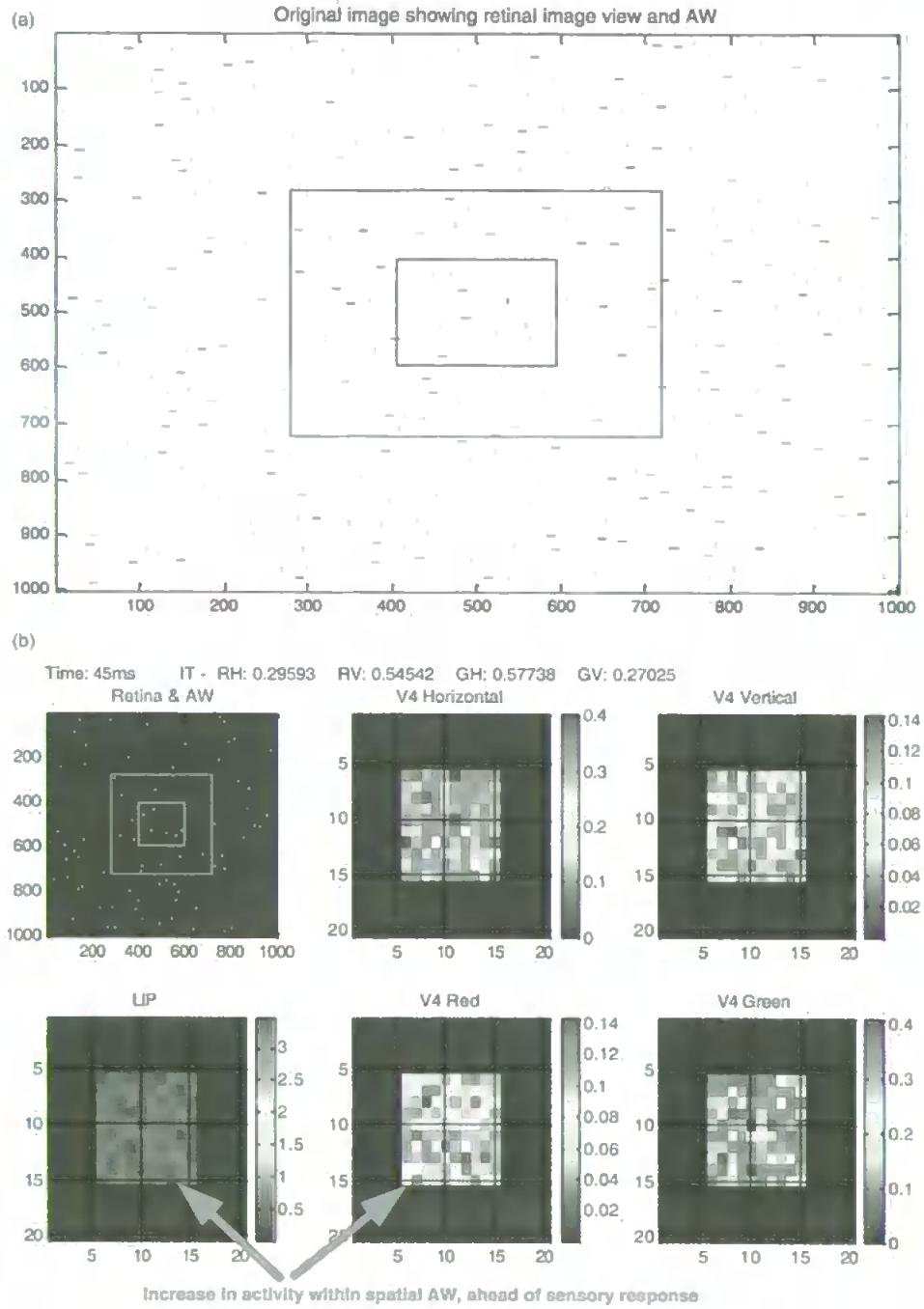



Fig. 4. (a) A complete scene with the retinal image shown as the outer of two boxes plotted. Within the retinal image the initial spatial AW, shown as the i box, is formed. This AW is scaled according to local stimulus density. (N.B. When figures containing red and green bars are viewed in greyscale print, th appears as a darker grey than the green.) (b) The activity within the cortical areas 45 ms after the start of the fixation. This is prior to the sensory response i (at 60 ms). However, there is an anticipatory elevation in activity level within the spatial AW in LIP and V4, as seen in single cell studies of LIP (Colby e 1996) and V4 (Luck et al., 1997). (c) The activity within the cortical areas 70 ms after the start of the fixation. At this time, spatial attention modulates respo in V4 and LIP. LIP represents both red and green stimulus locations. (d) The activity 180 ms after the start of the fixation. Object-based attention significantly modulated responses in IT and V4. LIP represents the location of target coloured (red) stimuli more strongly than non-target coloured (gr stimuli. Object-based effects are still subject to a spatial enhancement in V4 within the AW. (For interpretation of the references to colour in this figure leg the reader is referred to the web version of this article.)

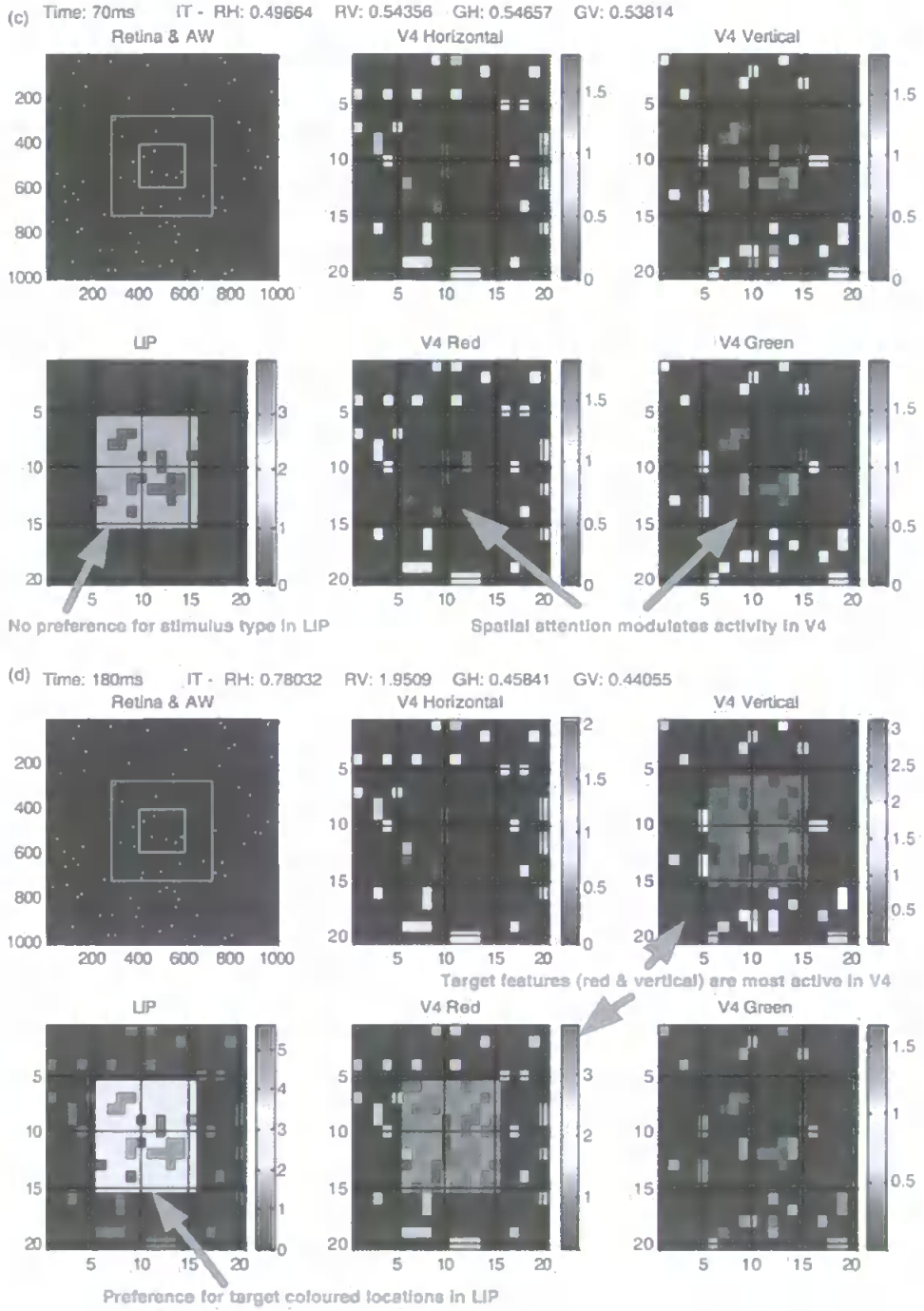



Fig. 4 (*continued*)

muli are approximately equally represented in LIP (Fig. ) and a saccade at this time would select target and non-rget coloured stimuli with equal probability. Saccade set is determined by the development of object-based fects in IT and, by the time this has occurred later in the sponse (Fig. 4d), the target coloured locations in LIP have come more active than non-target coloured locations. erefore, the saccade tends to select a target coloured cation. Increased fixation duration has been linked with re selective search (Hooge & Erkelens, 1999) and this would be explained, in this model, by the time taken to develop object-based effects in the ventral stream and convey this information to LIP.

### 3.2. Inhibition in V4

The nature of inhibition in V4 affects the representation of possible target locations in LIP. If V4 is implemented with a common inhibitory pool for all features or per feature type (i.e. one for colours and one for orientations), this

results in a normalising effect similar to a winner-take-all process, whereby high activity in any particular V4 assembly can strongly suppress other assemblies. V4 assemblies receive excitatory input from LIP, as a result of the reciprocal connection from V4 to LIP. When LIP receives input from all features in V4 as a result of two different types of stimulus (e.g. a red vertical bar and a horizontal green bar) being within a particular V4 receptive field, these feature locations in V4 will become most highly active and may suppress other locations too strongly. Thus, the common inhibitory pool tends to favour locations that contain two different stimuli within the same receptive field. Previous model that have used a common inhibitory pool (Deco, 2001; Deco & Lee, 2002) may not have encountered this problem because only one feature type was encoded. Therefore, in this model currently, the requirement is that features within a particular feature type should compete locally so, for example, an assembly selective for red stimuli competes with an assembly at the same retinotopic location but selective for green stimuli. Thus, inhibitory interneuro assemblies in V4 exist for every retinotopic location in V and for each feature type. This is shown in Fig. 2c.

This type of local inhibition results in the performanc shown in Fig. 5, which records the V4 colour assemblies ar LIP at two different time steps during a search for a re target. Within the receptive field of the V4 assemblies (matrix coordinate) location [6,13] both a red horizontal b and a green vertical bar are present in the retinal image ar are within the spatial AW. During the initial stimulu related response from 60 ms post-fixation, the red and t green selective assemblies at this location are approximate equally active. These are amongst the strongest V assemblies at this point because they are in receipt bottom-up stimulus information and are within the A positively biased by LIP. LIP represents the locations of a strong featural inputs from V4 and has no preference f stimulus colour at this point. However, by 70 ms po stimulus strong competitive effects, due to there being tv

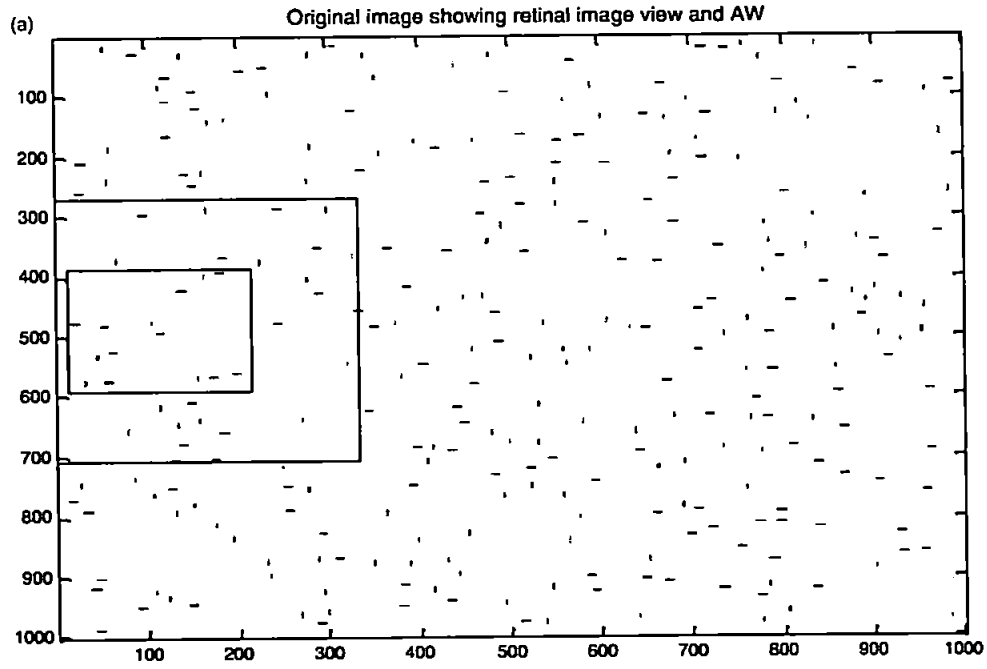



Fig. 5. The dynamic effect when a V4 receptive field includes two different stimuli. This image also shows what happens when fixation is close to the edg the original image (a computational problem not encountered in the real world) and the retinal image extends beyond the image. In this case the cortical a receive no bottom-up stimulus information for the area beyond the original image but other cortical processing remains unaltered. As LIP represents locations of behaviourally relevant stimuli, the next fixation is never chosen as a location beyond the extent of the original image. (a) The extent of the re image (outer box) and AW are shown for a fixation, forced to be at this location. (b) The activity within the cortical areas 70 ms after the start of the fixat Within the receptive field of V4 assemblies at position [6,13] there is both a red horizontal and a green vertical bar. Therefore, all V4 assemblies at this posi receive stimulus input. At the time of the initial sensory response assemblies at this position were as active as other assemblies in receipt of bottom-up stim information. However, by 70 ms competition between the colours at this location, and between the orientations at this location, has caused the activity he be slightly lower than that of other assemblies that encode the features of a single stimulus within their receptive field. This is due to there being competition in the single stimulus case. The addition of the second stimulus drives responses down, as has been found in single cell recordings such as Chel et al. (2001). At this time all V4 assemblies are approximately equally active and there is no preference for the target object's features. (c) The activity wi the cortical areas 200 ms after the start of the fixation when object-based attention has become significant. The red and vertical (target feature) assembli [6,13] (in matrix coordinates) have become more active than the non-target features, which are suppressed, despite each receiving 'bottom-up' information A plot of the activity of the V4 assemblies over time at position [6,13]. From ~ 150 ms object-based effects suppress non-target features. The activity o vertical orientation and red colour selective assemblies are shown as solid lines and that of the horizontal orientation and green colour selective assembli dotted lines.)

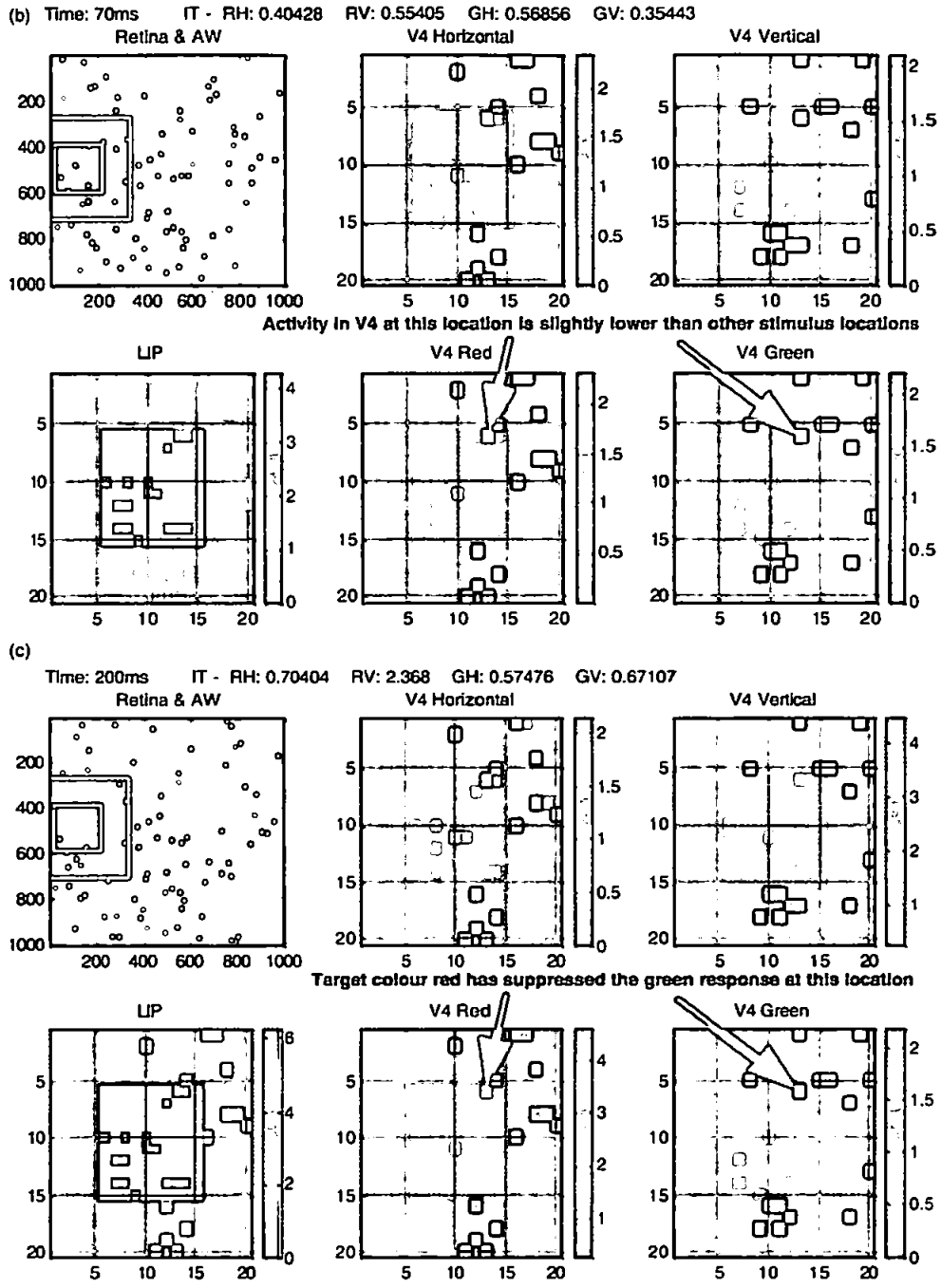
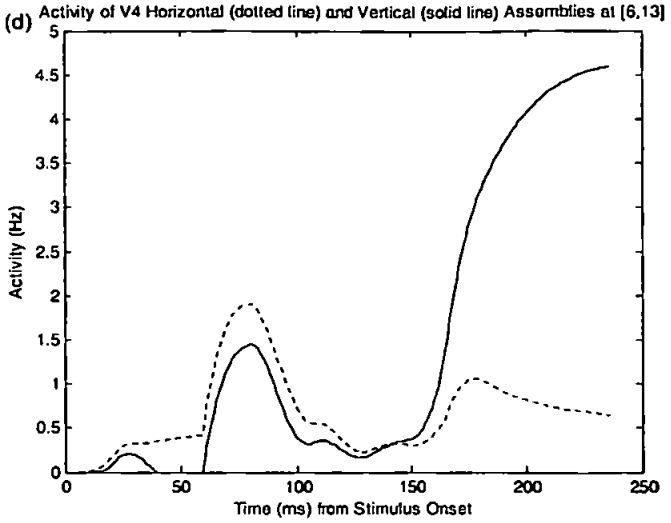



Fig. 5 (*continued*)

muli within the receptive field, result in overall responses each V4 feature at this location being lower than at cations that contain only one stimulus, i.e. the addition of econd stimulus lowers responses of the cell to the referred mulus, as found in single cell studies such as Chelazzi al. (2001). By 200 ms, object-based attention has become fective in V4 and the green assembly in V4 has become nificantly suppressed compared to the red assembly at the ne location. The representation in LIP is now biased wards the representation of the locations of target coloured stimuli and it is one of these locations that will be chosen as the next fixation point.

### 3.3. The scan path

The system produces scan paths that are qualitatively similar to those found by Motter and Belky (1998b) where fixations normally land within 1° of orientation-colour feature conjunction stimuli, rather than in blank areas, and these stimuli tend to be target coloured. Fig. 6a and b shows



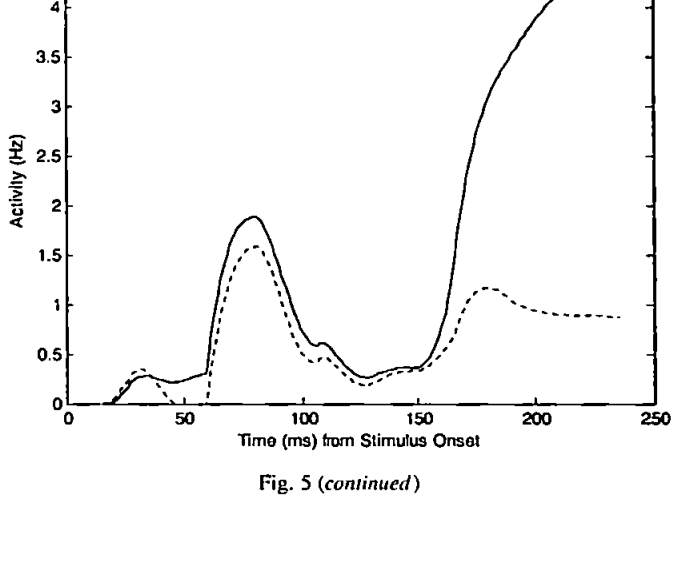




Fig. 5 (*continued*)

typical scan paths through relatively dense and sparse scenes, respectively.

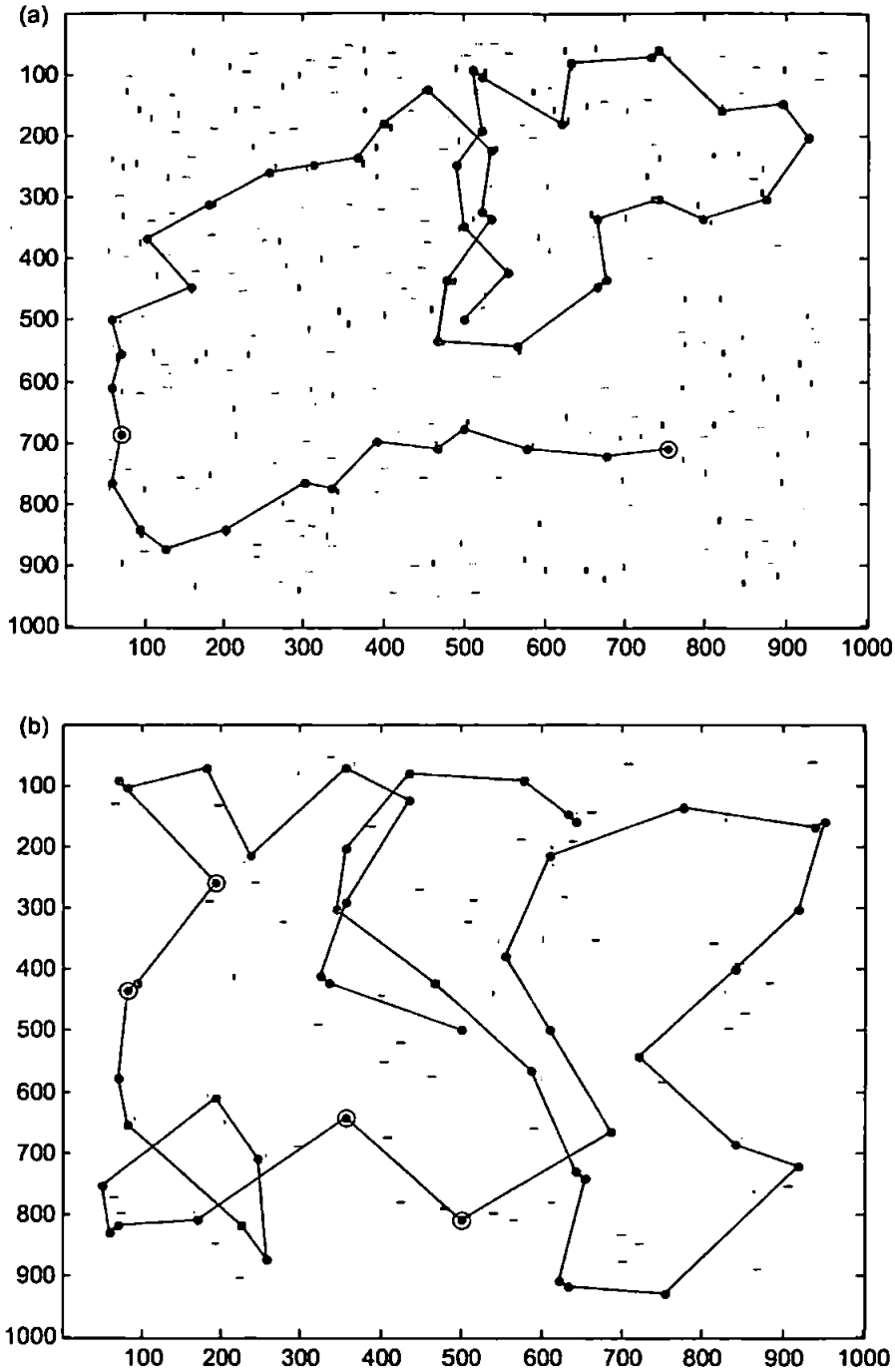
Weaker object-related feedback within the ventral pathway (prefrontal to IT; IT to V4) reduces the object-based effect within V4 and this results in more non-target coloured stimuli being able to capture attention. Replication of single cell data (described by Lanyon & Denham, submitted) from older and more highly trained monkeys (Chelazzi et al., 2001, compared to Chelazzi et al., 1993) suggested that IT feedback to V4 is tuned by learning, i.e. the strength of the IT feedback to V4 related to the amount of the monkeys' training in the task and stimuli. The strength of this feedback also affects the scan path. When the feedback is strengthened, object-based effects in IT and V4 are increased and the scan path is more likely to select target coloured stimuli. The effect of varying the weight of IT feedback to V4 on fixation position is shown in Fig. 7. Reducing the strength of object-related feedback from IT to V4, which may reflect learning, results in the capture of more non-target coloure stimuli within the scan path. Motter and Belky (1998 found that 5% of fixations landed within 1° of non-targ coloured stimuli during active search. Hence, the mod predicts that distractor stimuli (specifically, stimuli n sharing a target feature) may be more likely to captu attention when object-based effects within the ventr stream are weak (possibly due to unfamiliarity with t objects or the task) or have not yet had time to develop ful when the saccade is initiated. In particular, the tendency f the scan path to select only target coloured locations, durir search for a colour and form conjunction target, may weaker when the objects or task are less well known.

In addition, the strength of the novelty bias affects t scan path. Increasing the weight of the novelty bias to L slightly reduces the likelihood of fixating target colour stimuli and, in sparse scenes, increases the number fixations landing in blank areas of the display, as shown Fig. 8. This possibly suggests that a search where novelty the key factor, for example a hasty search of the enti scene, may result in more 'wasted' fixations in blank area

### 3.4. Saccade onset

Lanyon and Denham (submitted) replicated sacca onset times for single cell recordings in IT (Chelazzi et a 1993) and V4 (Chelazzi et al., 2001) and found that, due the link with the development of object-based attentio saccade onset can be delayed by two factors. First reducing the weight of IT feedback to V4 in order replicate data from less trained monkeys resulted in slightly longer latency to saccade. More significant saccade onset time was dependant on the nature and laten of object-based prefrontal feedback to IT, also suggested be related to learning. Thus, saccade onset time may depe on the familiarity with objects in the display and the tas

The same effect is found here for scan path simulatio As saccade onset is determined by the timing of obje based attention in IT, a typical plot of activity in IT durin scan path fixation is given in Fig. 9. The effect on sacc onset time of varying the strength of IT feedback to during scan path simulations is shown in Fig. 10. Stron object-based feedback from IT to V4 (and, similar prefrontal feedback to IT) results in the latency to sacc being reduced because the object-based effect in IT reinforced over time by the feedforward inputs from V4 the weight of the IT feedback (parameter $\eta$ in Eqs. (A and (A22)) is set to the value (5) used by Lanyon Denham (submitted) to replicate the single cell recordi of Chelazzi et al. (2001), scan path saccades take pl about 230–240 ms post-stimulus (i.e. after fixation). For scene used in Fig. 6a, this results in saccades taking pl on average, at 238 ms post-stimulus. Chelazzi et al. (20 found that saccades took place on average 237–240 post-stimulus, depending on stimulus configuration relation to the receptive field.




g. 6. Scan paths that tend to select locations near a target coloured stimulus. Here, IT feedback to V4 is strong (parameter $\eta$ in Eqs. (A20) and (A22) is set to Fixations are shown as (i) magenta dots: within 1° of a target coloured stimulus; (ii) blue circles: within 1° of a non-target colour stimulus. (a) A scan path ough a dense scene. The target is a red stimulus. 95.92% of fixations are within 1° of a target coloured stimulus. 4.08% of fixations are within 1° of a non-get colour stimulus. Average saccade amplitude = 7.43°. (b) A scan path through a sparse scene. The target is a green stimulus. 91.84% of fixations are hin 1° of a target coloured stimulus. 8.16% of fixations are within 1° of a non-target colour stimulus. Average saccade amplitude = 12.13°. (For erpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Thus, the model tentatively predicts that during search in owded scenes, as well as in the smaller arrays used for the gle cell simulations, fixation duration may be shorter nongst very familiar objects or in a familiar task. erefore, search may be faster under these conditions. In experiments, search has been found to be facilitated, enabling faster manual reaction times, when distractors are familiar (Lubow & Kaplan, 1997; Reicher, Snyder, & Richards, 1976; Richards & Reicher, 1978; Wang, Cavanagh, & Green, 1994). However, Greene and Rayner (2001)

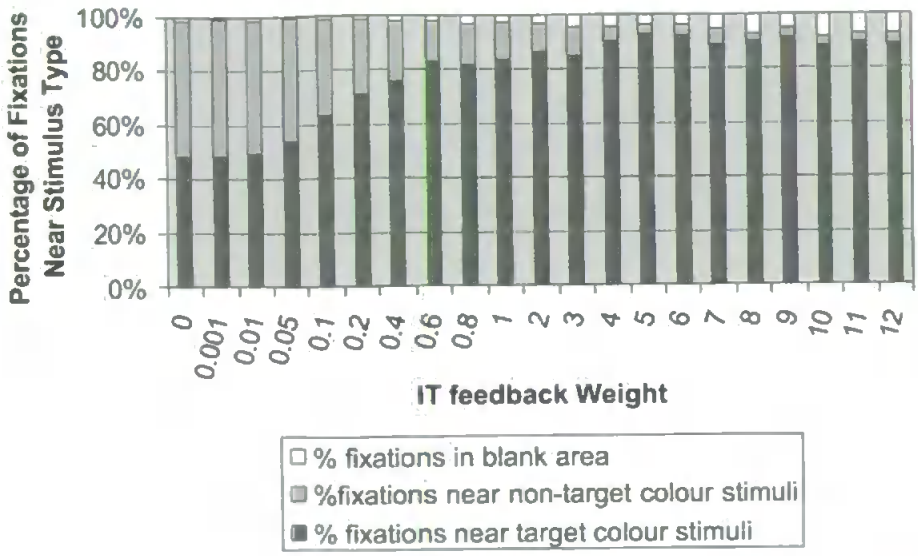



Fig. 7. The effect of the weight of IT to V4 feedback on fixation position. Fixation positions were averaged over 10 scan paths, each consisting of 50 fixatio over the image shown in Fig. 6a. Fixations are considered to be near a stimulus when they are within 1° of the stimulus. As the weight of feedback is increas there is a tendency for more target coloured stimuli and less non-target coloured stimuli to be fixated because object-based effects in the ventral stream stronger.

suggest that this *familiarity effect* may be due to the span of effective processing (which may be likened to the size of AW here) being wider around familiar distractors, rather than fixation duration being shorter.

### 3.5. Effect of scene density on saccade amplitude

As Motter and Belky (1998b) found, saccades tend to be shorter in dense scenes compared to sparse scenes. Fig. 6 shows this effect (also, large amplitude saccades are shown in a sparse scene in Fig. 11). The same sized retinal image was used for both the simulations in Fig. 6 in order that stimuli at the same distances could attract attention. However, the spatial AW is scaled based on local stimulus density such that, when fixation is placed in an area of den stimuli, the spatial AW is smaller than that when stimuli a sparse. The AW contributes a positive bias to t competition in LIP and results in locations containing targ features within the AW being favoured within LIP a being most likely to attract attention. Thus, the mod predicts that saccade amplitude is dependant on the stimul density in the local area around fixation.

Fig. 11a shows a scan path that has been unable to rea target coloured stimuli in the bottom left corner of t display. This was due to the constraint imposed by the si of the retinal image in this case (441 × 441 pixe equivalent to approximately 40° of visual angle). Wh the retinal image is increased to 801 × 801 pix

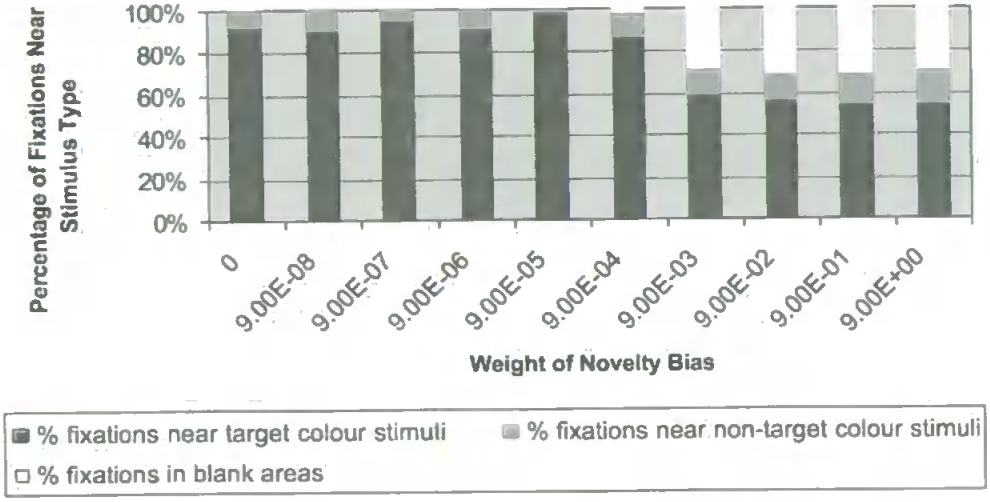
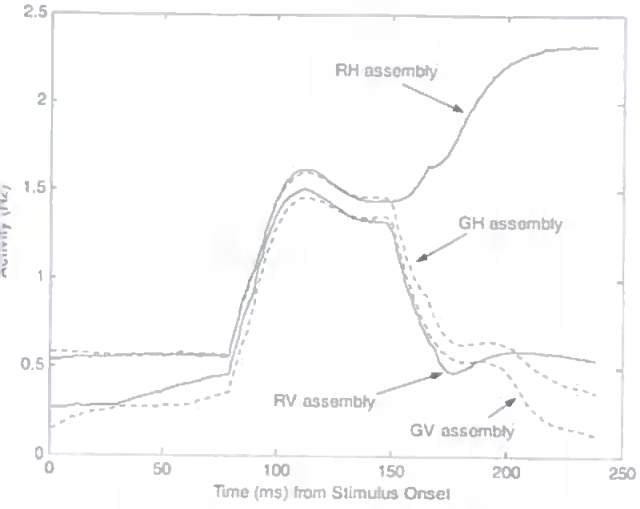
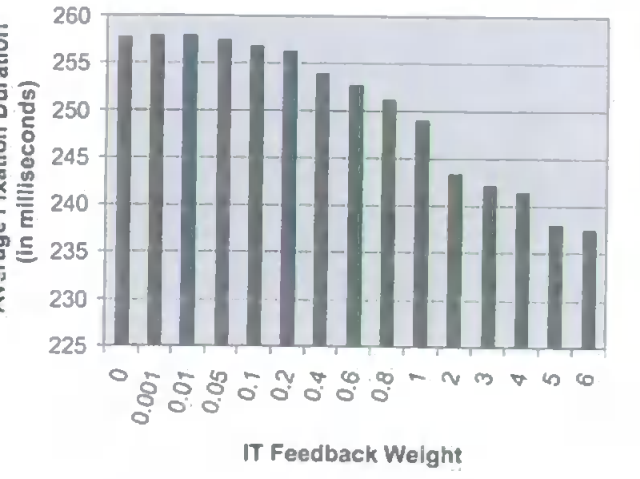



Fig. 8. The effect of varying the weight of the novelty bias on fixation position in a relatively sparse scene. Fixation position was averaged over 10 scan p each consisting of 50 fixations, over the image shown in Fig. 6b. Fixations are considered to be near a stimulus when they are within 1° of the stimulus. A weight of the novelty bias increases, the number of fixations near target coloured stimuli decreases and fixations are more likely to occur in blank areas o display.




g. 9. Plot of activity of the four object assemblies in IT (RH: encodes red orizontal bar; RV: encodes red vertical bar; GH: encodes green horizontal ar; GV: encodes green vertical bar) over time during one fixation. The itial sensory response begins from 80 ms post-stimulus (after fixation tablished). At that time all objects are approximately equally active. owever, from about 150 ms post-stimulus the target object (a red orizontal bar, RH) becomes most active. Significant object-based attention established and a saccade is initiated approximately 240 ms after the set of the stimulus. The saccade is indicated by the cessation of the plot nes. (For interpretation of the references to colour in this figure legend, the ader is referred to the web version of this article.)

quivalent to approximately 73° of visual angle), more imuli are available to the retina and this allows the stimuli the extremity to attract attention, as shown in Fig. 11b. ltering the retinal window size does not affect the nature of e stimuli that attract attention (in this case, target coloured imuli) but does alter the course of the scan path. Thus, the odel predicts that field of view can have a dramatic effect the scan path.




. 10. The effect on saccade onset of varying the weight of IT object-ted feedback to V4. Saccade onset times were averaged over 10 scan hs, each consisting of 50 fixations, over the image shown in Fig. 6a. As strength of this feedback increases, the latency to saccade is reduced, i.e. tion duration shortens.

### 3.6. Inhibition of return in the scan path

Implementation of the novelty bias to LIP successfully produces IOR within the scan path that declines over time allowing the eventual re-visiting of sites previously inspected. Plots of the novelty values after the scan paths shown in Fig. 6 are shown in Fig. 12. These figures show novelty being present at multiple locations (Danziger et al., 1998; Snyder & Kingston, 2001; Tipper et al., 1996), recovering over time (Irwin & Zelinsky, 2002; Snyder & Kingston, 2000), being strongest in the immediate vicinity of the fixation point and decreasing with distance from the fixation point (Hooge & Frens, 2000). Novelty scales with the size of the AW so that a smaller area around fixation receives reduced novelty in a dense scene than in a sparse scene. In a scene containing patches of different stimulus densities, the size of AW and, hence, novelty update is based on the local density. This allows the scan path to move through sparser areas with larger saccades but thoroughly examine any dense areas it encounters (Lanyon & Denham, 2004a).

## 4. Discussion and conclusion

This appears to be the first model that has both spatial and object-based attention occurring over different time courses (that accord with cellular recordings: Chelazzi et al., 1993, 2001; Luck et al., 1997) and with the scan path able to locate behaviourally relevant stimuli due to the creation of a retinotopic salience-type map based on the development of object-based attention. The model allows spatial attention and elements of early spatial selection theories (Treisman, 1982; Treisman & Gelade, 1980) to co-exist with object-based attention and the 'parallel' biased competition framework (Desimone, 1998; Desimone & Duncan, 1995; Duncan & Humphreys, 1989; Duncan et al., 1997). No specific gating of signals (Olshausen, Anderson, & Van Essen, 1993) is required to achieve the initial spatial AW but all attentional effects arise as a result of different biases that influence the dynamics of the system.

The model concentrates on object-based effects observed in monkey single cell studies (Chelazzi et al., 1993, 2001; Luck et al., 1997; Moran & Desimone, 1985). Such effects at this level may be important in leading to the object-based effects observed psychophysically (Blaser et al., 2000; Duncan, 1984), in imaging (O'Craven et al., 1999) and evoked potentials (Valdes-Sosa et al., 1998, 2000). The top-down object bias leads to feature-based attention across V4, as observed in monkey single cell recordings in the area (Motter, 1994a,b). Human imaging has also shown that attending to a feature increases activity in areas that encode that feature type (Chawla, Rees, & Friston, 1999; Saenz, Buracas, & Boynton, 2002). Target objects are selectively enhanced in IT in a bottom-up manner, due to the enhancement across its feature types, and top-down, due to

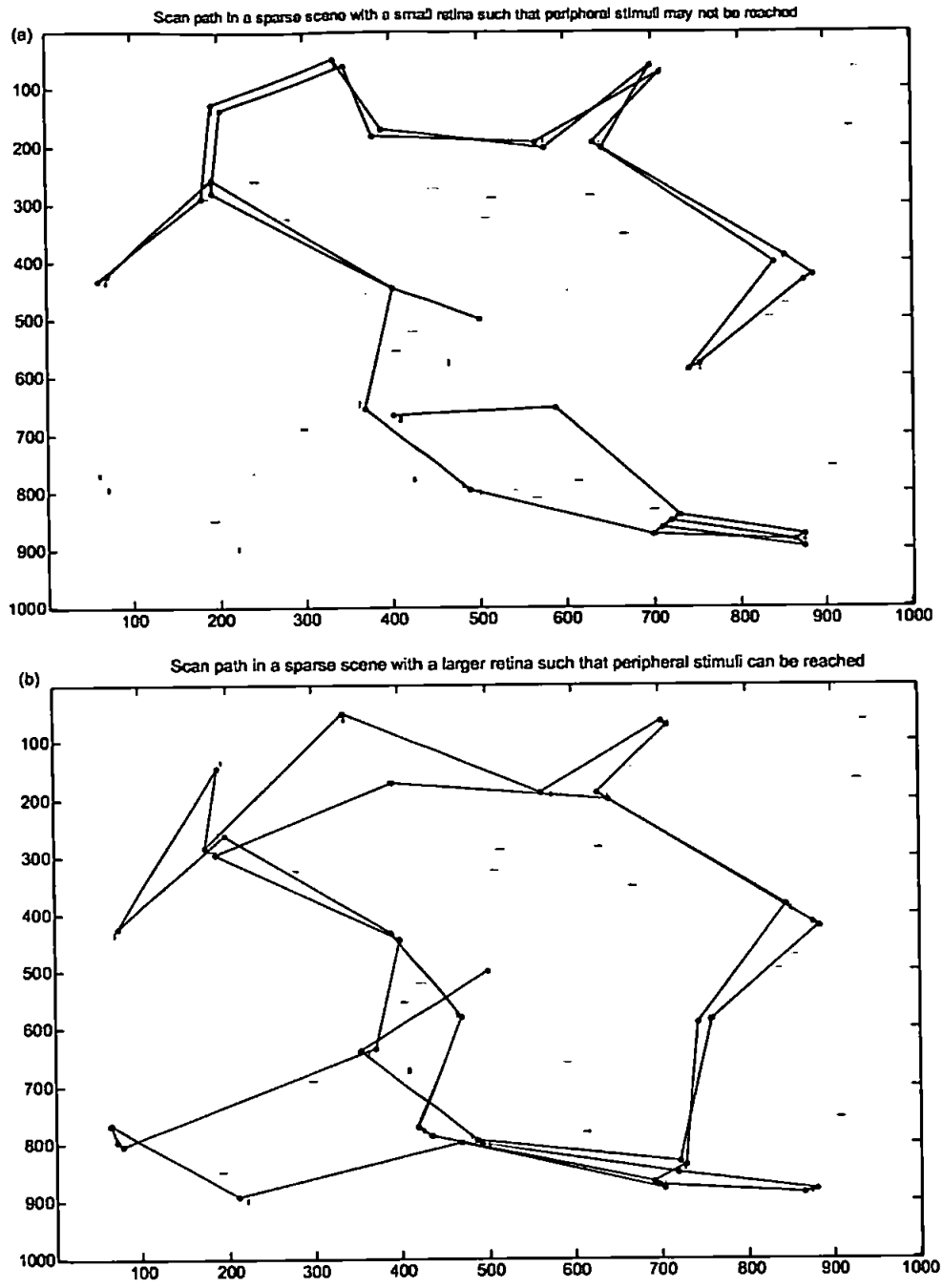
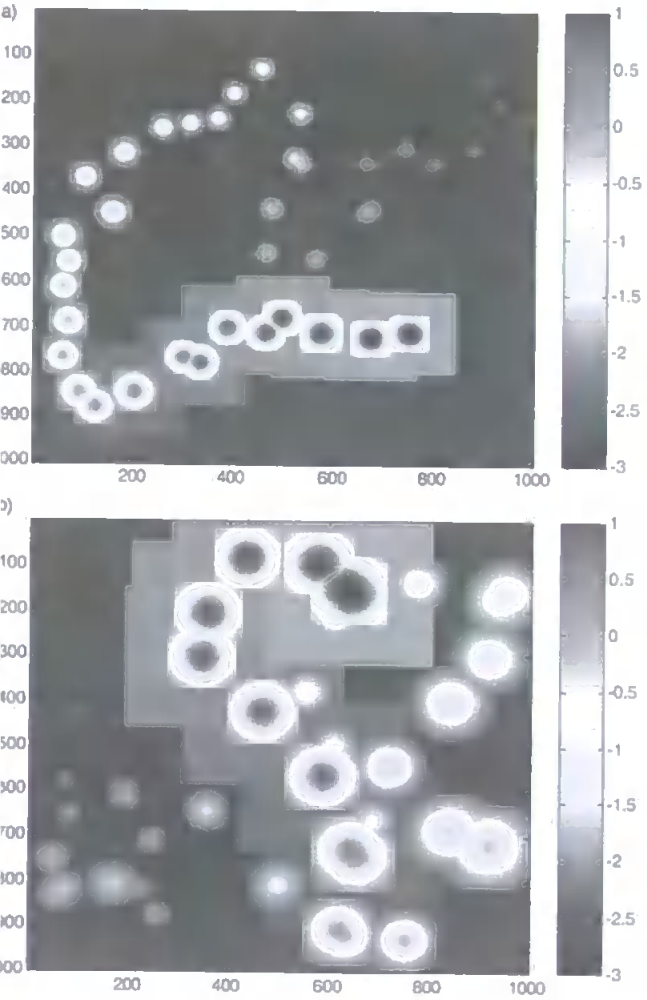



Fig. 11. Scan paths on a sparse image. (a) The retina is restricted to 441 × 441 pixels (~40°). This results in target coloured stimuli in the bottom left of image not being available to the retina from the nearby fixation point and, thus, not being reached by the scan path. Note, this problem only occurs in sp images with the retina image restricted to a smaller size than is normally available to humans and monkeys. (b) The retina is enlarged to be 801 × 801 pi (~73°). All target coloured stimuli are now examined.

the working memory bias. Many object-based attention experimental findings are confounded by issues relating to whether the selection is object-based, feature-based or surface-based. In psychophysics, object-based attention typically refers to the performance advantage given to all features belonging to the same object (but not to those of a different object at the same location). However, Mounts and Melara (1999) have suggested that not all features bound to objects are the subject of object-based attention and tha possibly earlier feature-based attention operates beca performance benefits were found only when the feat (colour or orientation) to be discriminated was the same in which the target 'pop-out' occurred. Confusion resu from the difficulty in designing an experimental paradi that eliminates all but object-based attentional effects., example, the rapid serial object transformation paradi




g. 12. The values remaining in the novelty map of the scene after a scan th. Novelty is present at multiple locations (Danziger et al., 1998; Snyder Kingston, 2001; Tipper et al., 1996), recovers over time (Irwin & linsky, 2002; Snyder & Kingston, 2000) and is strongest in the mediate vicinity of the fixation point, decreasing with distance from e fixation point (Hooge & Frens, 2000). (a) Novelty in a dense scene after e scan path shown in Fig. 6a. (b) Novelty in a sparse scene after the scan th shown in Fig. 6b.

itroduced by Valdes-Sosa et al., 1998, 2000, and used by itchell, Stoner, Fallah, & Reynolds, 2003, Pinilla, Cobo, rres, & Valdes-Sosa, 2001, and Reynolds, Alborzian, & oner, 2003) uses two transparent surfaces containing dots different colours and directions of motion and has been ported in the object-based attention literature. However, it s been shown to produce surface-based effects rather than ject- or feature-based effects (Mitchell et al., 2003). A mber of complex neural mechanisms are likely to nstitute the range of so-called object-based attention fects observed in various experimental paradigms and not are reflected in the current model. One factor not fully dressed in the current version is the neural correlate of rceptual grouping, according to Gestalt principles, to ally group features into objects. For example, lateral nnections in early stages of visual processing, such as those modelled by Grossberg and Raizada (2000), may result in collinear enhancements leading to object-based effects such as those observed by Roelfsema et al. (1998) in V1. In the current model, such local lateral interactions could be added to V1 and V4 by means of an excitatory input from similar feature selective cells in the local area. Such an addition would allow local feature grouping to be performed, which could assist the separation of overlapping objects. However, arguably, depth information may be involved in this process in order to create separable surfaces. Whilst not addressing all issues of object-based attention this model makes a valuable contribution by unifying object-based and spatial effects at the single cell level and using the resultant activity to drive search behaviour.

The model is also novel in that certain features have priority in driving the search scan path, as has been found to be the case for colour in monkeys (Motter & Belky, 1998b) and humans (Sciala & Joffee, 1998; Williams & Reingold, 2001) during certain types of search. Object-based attention, as a result of biased competition, leads to feature-based attention in V4. For example, non-target colours are suppressed such that the scene is segmented by colour as if viewed through a colour-attenuating filter (Motter, 1994a, b). The development of object-based attention across V4 was shown to influence the scan path due to the integration of featural information in the spatial competition in LIP. The nature of connections from V4 to LIP allowed colour to have priority over orientation in driving the scan path but only when object-based effects were strong enough in V4. This may be viewed as priority being given to certain feature maps (Treisman, 1988; Wolfe, 1994; Wolfe, Cave, & Franzel, 1989). Previous models (Deco & Lee, 2002; Hamker, 1998; Niebur et al., 2001) have tended not to have a particular featural bias in attracting attention; for example, when forming a 'salience map'. The map produced here in LIP integrates the output of object-based attention in the model's ventral stream and produces the featural bias by means of a slight difference in connection strength from V4 features to LIP. The relative weight of these connections may be malleable on the basis of task requirements or stimulus-related information, such as the proportion of each distractor type present in the retinal image. Thus, the distractor-ratio effect (Bacon & Egeth, 1997; Shen et al., 2000, 2003) could be achieved through the ratio of featural connection strengths from V4 to LIP, allowing the feature type that has the lowest proportion of distractors to bias competition in LIP most strongly and, therefore, attract the scan path to these distractor locations. However, other process may contribute to this effect. For example, normalisation of total activity within each V4 feature layer could allow features occurring rarely in the retinal image to achieve higher activity in V4 so that the locations of stimuli possessing a rarely occurring target feature become most active in LIP. The primary goal of the current work was to replicate behaviour observed by Motter and Belky (1998b), where distractor types were equal and colour tended to

influence the scan path most strongly despite task requirements.

The 'cross-stream' interaction between V4 and LIP is an important feature of this model and it is predicted that loss of the connection from the ventral stream to LIP, due to lesion, would result in deficits in the ability for the scan path to locate behaviourally relevant stimuli. In order to the object-based effect in V4 to bias LIP correctly in the current architecture, it was necessary for competition in V4 to be local. Given this, object-based attention in V4 allowed LIP to be able to map the behaviourally relevant locations in the scene (Colby et al., 1996) and accurately control the search scan path. Therefore, the model predicts that ventral stream feedforward connections to parietal cortex may determine search selectivity. The model also predicts that the latency to saccade is dependant on the development of object-based attention, which relies on strong object-related feedback within the ventral stream. Tentatively, the strength of such feedback may be related to learning and familiarity with stimuli and the task. The connection from LIP to V4 serves to bind features across feature types at the resolution of the V4 receptive field. Therefore, it is predicted that the loss of this connection would cause cross-feature type binding errors (such as binding form and colour) whilst leaving intact the binding of form into object shape within the ventral stream. Such an effect has been observed in a patient with bilateral parietal lesions (Humphreys, Cinel, Wolfe, Olson, & Klempen, 2000).

Further factors affecting the scan path included the field of view (the size of the retinal image), local stimulus density (because this determines the size of the AW, which influences saccade amplitude), and the importance of novelty (in the form of inhibition of previously inspected locations).

Future work includes extending the factors influencing competition in LIP so that the model is able to represent further bottom-up stimulus-related factors and, possibly, further top-down cognitive factors in the competition for attentional capture. As such, the model may be able to replicate psychophysical data relating to attentional capture under competing exogenous and endogenous influences, and have practical application in computer vision.

A further issue that could be addressed is the constancy of the scene across saccades. It may be possible to model changes in receptive field profiles in LIP (Ben Hamed et al., 1996) and V4 (Tolias et al., 2001) immediately before saccade onset in order to provide 'predictive re-mapping' of stimuli within LIP (Duhamel et al., 1992). This could provide further insight into the neural basis of the novelty bias necessary for IOR. The model has highlighted deficits in the knowledge of the neural correlates of inhibition of saccade return during active visual search. Further work is needed to establish the link between scene-based inhibition (or a scene-based memory of locations already inspected) and retinotopic areas, such as LIP, that are involved in selecting saccadic targets.

In conclusion, this is a biologically plausible model tha is able to replicate a range of experimental evidence for bot spatial and object-based attentional effects. Modellin visual attention in a biologically constrained manner withi an active vision paradigm has raised issues, such as memor for locations already visited in the scene, not addressed b models with static retinas. The model is extendable t include other factors in the competition for the capture o attention and should be of interest in many fields includin neurophysiology, neuropsychology, computer vision an robotics.

## Appendix A

### A.1. Retina

Colour processing in the model focuses on the red green channel with two colour arrays, $I^{\text{red}}$ and $I^{\text{green}}$, simplification of the output of the medium and lon wavelength retinal cones, being input to the retin ganglion cells. References to red and green througho this paper refer to long and short wavelengths. T greyscale image, $I^{\text{grey}}$, used for form processing, is composite of the colour arrays and provides luminan information.

#### A.1.1. Form processing in the retina

At each location in the greyscale image, retinal gangli broad-band cells perform simple centre-surround proc sing, according to Grossberg and Raizada (2000) as follow

*On-centre, off-surround broadband cells:*

$$u^{+} = I_{ij}^{\text{grey}} - \sum_{pq} G_{pq}(i,j,\sigma_1) I_{pq}^{\text{grey}} \quad \text{(A}$$

*Off-centre, on-surround broadband cells:*

$$u^{-} = -I_{ij}^{\text{grey}} + \sum_{pq} G_{pq}(i,j,\sigma_1) I_{pq}^{\text{grey}} \quad \text{(A}$$

where $G_{pq}(i,j,\sigma_1)$ is a two-dimensional Gaussian kern given by:

$$G_{pq}(i,j,\sigma_1) = \frac{1}{2\pi\sigma_1^2} \exp\left( -\frac{1}{2\sigma_1^2}((p-i)^2 + (q-j)^2) \right) \quad \text{(A}$$

The Gaussian width parameter is set to: $\sigma_1 = 1$.

These broadband cells provide luminance inputs to interblob simple cells that are orientation selective.

#### A.1.2. Colour processing in the retina

Retinal concentric single opponent cells process col information as follows.

*Red on-centre, off-surround concentric single-opponent cells:*

$$I_{ij}^{\text{redON}} = I_{ij}^{\text{red}} - \sum_{pq} G_{pg}(i,j,\sigma_1) I_{pq}^{\text{green}} \quad \text{(A4)}$$

*Red off-centre, on-surround concentric single-opponent cells:*

$$I_{ij}^{\text{redOFF}} = -I_{ij}^{\text{red}} + \sum_{pq} G_{pg}(i,j,\sigma_1) I_{pq}^{\text{green}} \quad \text{(A5)}$$

*Green on-centre, off-surround concentric single-opponent cells:*

$$I_{ij}^{\text{greenON}} = I_{ij}^{\text{green}} - \sum_{pq} G_{pg}(i,j,\sigma_1) I_{pq}^{\text{red}} \quad \text{(A6)}$$

*Green off-centre, on-surround concentric single-opponent cells:*

$$I_{ij}^{\text{greenOFF}} = -I_{ij}^{\text{green}} + \sum_{pq} G_{pg}(i,j,\sigma_1) I_{pq}^{\text{red}} \quad \text{(A7)}$$

These concentric single-opponent cells provide colour-specific inputs to V1 double-opponent blob neurons.

### .2. V1

The V1 module consists of K + C neurons at each location in the original image, so that neurons detect K orientations and C colours. At any fixation, V1 would only process information within the current retinal image. However, in this model, the entire original image is 'pre-processed' by V1 in order to save computational time during the active vision component of the system. As V1 is not dynamically updated during active vision, this does not alter the result. Only those V1 outputs relating to the current retinal image are forwarded to V4 during the dynamic active vision processing.

#### .2.1. Form processing in V1

For orientation detection, V1 simple and complex cells are modelled as described by Grossberg and Raizada (2000), with the distinction that two spatial resolutions are calculated here. Simple cells detect oriented edges using a difference-of-offset-Gaussian (DOOG) kernel.

The right- and left-hand kernels of the simple cells are given by

$$_{ijk} = \Sigma_{pq}([u_{pq}^{+}]_+ - [u_{pq}^{-}]_+)[D_{pqij}^{(lk)}]_+ \quad \text{(A8)}$$

$$_{ijk} = \Sigma_{pq}([u_{pq}^{-}]_+ - [u_{pq}^{+}]_+)[-D_{pqij}^{(lk)}]_+ \quad \text{(A9)}$$

here

$u^+$ and $u^-$ are the outputs of the retinal broadband cells above;

$[x]_+$ signifies half-wave rectification, i.e.

$$[x]_+ = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

and the oriented DOOG filter $D_{\text{pqij}}^{(\text{lk})}$ is given by

$$D_{pqij}^{(lk)} = G_{pq}(\text{i} - \delta\cos\theta, \text{j} - \delta\sin\theta, \sigma_2) - G_{pq}(\text{i} + \delta\cos\theta, \text{j} + \delta\sin\theta, \sigma_2) \quad \text{(A10)}$$

where

$\delta = \sigma_2/2$ and $\theta = \pi(k-1)/K$, where $k$ ranges from 1 to $2K$, $K$ being the total number of orientations (2 is used here).

$\sigma_2$ is the width parameter for the DOOG filter, set as below.

$r$ is the spatial frequency octave (i.e. spatial resolution), such that

$r = 1$ and $\sigma_2 = 1$ for low resolution processing, used in the magnocellular (or sub-cortical) pathway for scaling the AW;

$r = 2$ and $\sigma_2 = 0.5$ for high resolution processing, used in the parvocellular pathway, which forms the remainder of the model

The direction-of-contrast sensitive simple cell response is given by

$$S_{rijk} = \gamma[R_{rijk} + L_{rijk} - |R_{rijk} - L_{rijk}|]_+ \quad \text{(A11)}$$

$\gamma$ is set to10.

The complex cell response is invariant to direction of contrast and is given by

$$I_{rijk} = S_{rijk} + S_{rij(k+K)} \quad \text{(A12)}$$

where $k$ ranges from 1 to $K$.

The value of the complex cells, $I_{rijk}$, over the area of the current retinal image, is input to V4.

#### *A.2.2. Colour processing in V1*

The outputs of LGN concentric single-opponent cells (simplified to be the retinal cells here) are combined in the cortex in the double-opponent cells concentrated in the blob zones of layers 2 and 3 of V1, which form part of the parvocellular system. The outputs of blob cells are transmitted to the thin stripes of V2 and from there to colour-specific neurons in V4. For simplicity, V2 is not included in this model.

Double-opponent cells have a centre-surround antagonism and combine inputs from different single-opponent cells as follows:

*Red on-centre portion:*

$$\omega_{ij}^{\text{redON}} = \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{\text{redON}} + \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{\text{greenOFF}} - \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{\text{redOFF}} - \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{\text{greenON}} \quad \text{(A13)}$$

*Red off-surround portion:*

$$\omega_{ij}^{\text{redOFF}} = \sum_{pq} G_{pq}(i,j,\sigma_2)\nu_{ij}^{\text{redOFF}} + \sum_{pq} G_{pq}(i,j,\sigma_2)\nu_{ij}^{\text{greenON}} - \sum_{pq} G_{pq}(i,j,\sigma_2)\nu_{ij}^{\text{redON}} - \sum_{pq} G_{pq}(i,j,\sigma_2)\nu_{ij}^{\text{greenOFF}} \quad \text{(A14)}$$

*Green on-centre portion:*

$$\omega_{ij}^{\text{greenON}} = \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{\text{greenON}} + \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{\text{redOFF}} - \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{\text{greenOFF}} - \sum_{pq} G_{pq}(i,j,\sigma_1)\nu_{ij}^{\text{redON}} \quad \text{(A15)}$$

*Green off-surround portion:*

$$\omega_{ij}^{\text{greenOFF}} = \sum_{pq} G_{pq}(i,j,\sigma_2)\nu_{ij}^{\text{greenOFF}} + \sum_{pq} G_{pq}(i,j,\sigma_2)\nu_{ij}^{\text{redON}} - \sum_{pq} G_{pq}(i,j,\sigma_2)\nu_{ij}^{\text{greenON}} - \sum_{pq} G_{pq}(i,j,\sigma_2)\nu_{ij}^{\text{redOFF}} \quad \text{(A16)}$$

where

$\sigma_1 = 1, \quad \sigma_2 = 2$

The complete red-selective blob cell is given by

$$I_{ijc^1} = \gamma[\omega_{ij}^{\text{redON}} - \omega_{ij}^{\text{redOFF}}]_+ \quad \text{(A17)}$$

The complete green-selective blob cell is given by

$$I_{ijc^2} = \gamma[\omega_{ij}^{\text{greenON}} - \omega_{ij}^{\text{greenOFF}}]_+ \quad \text{(A18)}$$

where

$\gamma = 0.2$. This scales the output of V1 blob cells to be consistent with that of the orientation-selective cells.
$c^1 = K + 1$. This represents the position of the first colour input to V4 (i.e. red).
$c^2 = K + 2$. This represents the position of the second colour input to V4 (i.e. green).

The blob cell outputs over the area of the current retinal image are input to V4.

### A.3. Dynamic cortical modules

The dynamic cortical modules follow a similar approach to Deco (2001) and Deco and Lee (2002), and are modelled using mean field population dynamics (Gerstner, 2000), as used by Usher and Niebur (1996), in which an average ensemble activity is used to represent populations, c assemblies, of neurons with similar encoding properties. is assumed that neurons within the assembly receive simil external inputs and are mutually coupled. Modelling at th neuronal assembly level is inspired by observations tha several brain regions contain populations of neurons tha receive similar inputs and have similar properties, and it is suitable level at which to produce a systems level mod such as this (see Gerstner (2000) and Rolls and Deco (200 for further details of this modelling approach). Populatio averaging does not require temporal averaging of th discharge of individual cells and, thus, the response of th population may be examined over time, subject to the size the time step used in the differential equations within th model. The response function, which transforms curre (activity within the assembly) into discharge rate, is give by the following sigmoidal function that has a logarithm singularity

$$F(x) = \frac{1}{T_r - \tau \log\left(1 - \frac{1}{\tau x}\right)} \quad \text{(Gerstner, 2000)} \quad \text{(A1}$$

where

$T_r$, the absolute refractory time, is set to 1 ms
$\tau$ is the membrane time constant (where $1/\tau$ determin the cell's firing threshold). The threshold is normally s to half the present maximum activity in the layer (in V this is half the maximum within the feature type in ord that the object-based attention differences betwe features are not lost due to the normalising effect this function).

LIP contains the same number of neurons as V4 and h reciprocal connections with both the orientation and colo layers in V4. The size of V4 and LIP is determined by t size of the retinal image, which is flexible, but is norma set to a radius of 220 pixels for simulations of the scan pa If monochromatic stimuli are used (normally only for sing cell simulations), V4 contains only orientation selecti assemblies.

#### A.3.1. V4

V4 consists of a three-dimensional matrix of pyrami cell assemblies. The first two dimensions repres the retinotopic arrangement and the other represents individual feature types: Orientations and colours. In latter dimension, there are K + C cell assemblies, as sho in Fig. 2c. The first K assemblies are each tuned to one o orientations. The next C cell assemblies are each tuned t particular colour. Two orientations (vertical and horizon and two colours (red and green) are normally used h Two sets of inhibitory interneuron pools exist: One mediates competition between orientations and the ot mediates competition between colours. V4 receives c vergent input from V1 over the area of its receptive fi

with a latency of 60 ms to reflect normal response latencies (Luck et al., 1997). In order to simulate the normalisation of inputs occurring during retinal, LGN and V1 processing, the V1 inputs to V4 are normalised by passing the convergent inputs to each V4 assembly through the response function at Eq. (A19) with its threshold set to a value equivalent to an input activity for approximately half a stimulus within its receptive field.

*A.3.1.1. Form processing in V4.* The output from the V1 simple cell process, $I_{ijk}$, for each position $(i,j)$ at orientation $k$, provides the bottom-up input to orientation selective pyramidal assemblies in V4 that evolve according to the following dynamics

$$\tau_1 \frac{\delta}{\delta t} W_{ijk}(t) = -W_{ijk}(t) + \alpha F(W_{ijk}(t)) - \beta F(W_{ij}^{IK}(t)) + \chi \sum_{pq} I_{pqk}(t) + \gamma Y_{ij}(t) + \eta \sum_{m} B_{W_{ijk}X_m} \cdot F(X_m(t)) + I_0 + \nu \quad \text{(A20)}$$

where

$\tau_1$ is set to 20 ms

$\alpha$ is the parameter for excitatory input from other cells in the pool, set to 0.95

$\beta$ is the parameter for inhibitory interneurons input, set to 10

$I_{pqk}$ is the input from the V1 simple cell edge detection process at all positions within the V4 receptive field area $(p,q)$, and of preferred orientation $k$

$\chi$ is the parameter for V1 inputs, set to 4

$Y_{ij}$ is the input from the posterior parietal LIP module, reciprocally connected to V4

$\gamma$ is the parameter for LIP inputs, set to 3

$X_m$ is the feedback from IT cell populations via weight $B_{W_{ijk}X_m}$, described later

$\eta$ is the parameter representing the strength of object-related feedback from IT (Fig. 7); normally set to 5, but set to 2.5 for simulation of single cell recordings in IT (Chelazzi et al., 1993),

$I_0$ is a background current injected in the pool, set to 0.25

$\nu$ is additive noise, which is randomly selected from a uniform distribution on the interval (0,0.1)

The dynamic behaviour of the associated inhibitory pool for orientation-selective cell assemblies in V4 is given by

$$\frac{\delta}{\delta t} W_{ij}^{IK}(t) = -W_{ijk}^{IK}(t) + \lambda \sum_{k} F(W_{ijk}(t)) - \mu F(W_{ij}^{IK}(t)) \quad \text{(A21)}$$

where

$\lambda$ is the parameter for pyramidal cell assembly input, set to 1

$\mu$ is the parameter for inhibitory interneuron input, set to 1

Over time, this results in local competition between different orientation selective cell assemblies.

*A.3.1.2. Colour processing in V4.* The output from the V1 simple cell process, $I_{ijc}$, for each position $(i,j)$ and colour $c$, provides the bottom-up input to colour selective pyramidal assemblies in V4 that evolve according to the following dynamics

$$\tau_1 \frac{\delta}{\delta t} W_{ijc}(t) = -W_{ijc} + \alpha F(W_{ijc}(t)) - \beta F(W_{ij}^{IC}(t)) + \chi \sum_{pq} I_{pqc}(t) + \gamma Y_{ij}(t) + \eta \sum_{m} B_{W_{ijc}X_m} \cdot F(X_m(t)) + I_0 + \nu \quad \text{(A22)}$$

where

$I_{pqc}$ is the input from the V1 blob cells at all positions within the V4 receptive field area $(p,q)$, and of preferred colour $c$.

$X_m$ is the feedback from IT cell populations via weight $B_{W_{ijc}X_m}$, described later.

The remaining terms are the same as those in Eq. (A20).

The dynamic behaviour of the associated inhibitory pool for colour-selective cell assemblies in V4 is given by:

$$\tau_1 \frac{\delta}{\delta t} W_{ij}^{IC}(t) = -W_{ijc}^{IC}(t) + \lambda \sum_{c} F(W_{ijc}(t)) - \mu F(W_{ij}^{IC}(t)) \quad \text{(A23)}$$

Parameters take the same values as those in Eq. (A21).

Over time, this results in local competition between different colour selective cell assemblies.

*A.3.2. IT*

Neuronal assemblies in IT are assumed to represent anterior IT (for example, area TE) where receptive fields cover the entire retinal image and populations encode invariant representations of objects. The model IT encodes all possible objects, i.e. feature combinations, and receives feedforward feature inputs from V4 with a latency of 80 ms to reflect normal response latencies (Wallis & Rolls, 1997). V4 inputs to IT are normalised by dividing the total input to each IT assembly by the total number of active (i.e. non-zero) inputs. IT also feeds back an object bias to V4. The strength of these connections is given by the following weights, which are set by hand (to $-1$ or 0, as appropriate, for inhibitory feedback, although the model may also be implemented with excitatory feedback; 0, $+1$) to represent prior object learning. These simple matrices reflect the type of weights that would be achieved through Hebbian learning without the need for a lengthy learning procedure (such as Deco, 2001), which is not the aim of this work. The result is

that the connections that are active for excitatory feedback (or inactive for inhibitory feedback) are those features relating to the object.

*V4 cell assemblies to IT (feedforward):*

$A_{X_m W_{ijz}}$

*IT to V4 cell assemblies (feedback):*

$B_{W_{ijz} X_m}$

where $z$ indicates orientation, $k$, or colour, $c$.

The pyramidal cell assemblies in IT evolve according to the following dynamics

$$\tau_1 \frac{\delta}{\delta t} X_m(t) = -X_m(t) + \alpha F(X_m(t)) - \beta F(X^I(t)) + \chi \sum_{ijk} A_{X_m W_{ijk}} \cdot F(W_{ijk}(t)) + \chi \sum_{ijc} A_{X_m W_{ijc}} \cdot F(W_{ijc}(t)) + \gamma P_M^v(t) + I_0 + \nu \quad \text{(A24)}$$

where

- $\beta$ is the parameter for inhibitory interneuron input, set to 0.01
- $W_{ijk}$ is the feedforward input from V4 relating to orientation information, via weight $A_{X_m W_{ijk}}$
- $W_{ijc}$ is the feedforward input from V4 relating to colour information, via weight $A_{X_m W_{ijc}}$
- $\chi$ is the parameter for V4 inputs, set to 2.5
- $P_M^v$ is the object-related feedback current from ventrolateral prefrontal cortex, injected directly into this pool

This feedback is sigmoidal over time as follows:

$$P_M^v = -1/(1 + \exp(\tau_{sig} - t)) \quad \text{(A25)}$$

where $t$ = time (in milliseconds) and
$\tau_{sig}$ is the point in time where the sigmoid reaches half its peak value, normally set to 150 ms for scan path simulations.
$\gamma$ is the parameter for the object-related bias, set to 1.2

The remaining terms and parameters are evident from previous equations.

The dynamic behaviour of the associated inhibitory pool in IT is given by

$$\tau_1 \frac{\delta}{\delta t} X^I(t) = -X^I(t) + \lambda \sum_m F(X_m(t)) - \mu F(X^I(t)) \quad \text{(A26)}$$

where

- $\lambda$ is the parameter for pyramidal cell assembly input, set to 3
- $\mu$ is the parameter for inhibitory interneuron input, set to 1

*A.3.3. LIP*

The pyramidal cell assemblies in LIP evolve according t the following dynamics

$$\tau_1 \frac{\delta}{\delta t} Y_{ij}(t) = -Y_{ij}(t) + \alpha F(Y_{ij}(t)) - \beta F(Y^I(t)) + \chi \sum_k F(W_{ijk}(t)) + \varepsilon \sum_c F(W_{ijc}(t)) + \gamma P_{ij}^d(t) + \eta \sum_{pq} Z_{pq} + I_0 + \nu \quad \text{(A2}$$

where

- $\beta$ is the parameter for inhibitory input, set to 1
- $W_{ijk}$ is the orientation input from V4 for orientation $k$, location $(i,j)$
- $W_{ijc}$ is the colour input from V4 for colour $c$, at locatio $(i,j)$
- $\varepsilon > \chi$ so that colour-related input from V4 is strong than orientation-related input. Set to $\chi = 0.$ $\varepsilon = 4$.
- $P_{ij}^d$ is the top-down bias, i.e. spatial feedback curre from FEF, dorsolateral prefrontal cortex or pul nar, injected directly into this pool when there is requirement to attend to this spatial location. He when fixation is established, this spatial bias applied and the spatial AW is formed.
- $\gamma$ is the parameter for the spatial top-down bias, set 2.5
- $Z_{pq}$ is the bias from the area, $pq$, of the novel map (which is the size of the original image, that represents the receptive field. Area represents the size of the LIP receptive field.
- $\eta$ is the parameter for the novelty bias, normally to 0.0009 (Fig. 8)

In order to attract the scan path to target colour locations:

The remaining terms are evident from previo equations.

The dynamic behaviour of the associated inhibitory po in LIP is given by

$$\tau_1 \frac{\delta}{\delta t} Y^I(t) = -Y^I(t) + \lambda \sum_{ij} F(Y_{ij}(t)) - \mu F(Y^I(t)) \quad \text{(A}$$

where

- $\lambda$ is the parameter for pyramidal cell assembly inp set to 1;
- $\mu$ is the parameter for inhibitory interneuron inp set to 1.

## References


Anand, S., & Bridgeman, B. (2002). An unbiased measure of the contributions of chroma and luminance to saccadic suppression of displacement. *Experimental Brain Research, 142*(3), 335–341.

Andersen, R. A., & Zipser, D. (1988). The role of the posterior parietal cortex in coordinate transformations for visual-motor integration. *Canadian Journal of Physiology and Pharmacology, 66*(4), 488–501.

Anllo-Vento, L., & Hillyard, S. A. (1996). Selective attention to the color and direction of moving stimuli: Electrophysiological correlates of hierarchical feature selection. *Perception and Psychophysics, 58*(2), 191–206.

Bacon, W. J., & Egeth, H. E. (1997). Goal-directed guidance of attention: Evidence form conjunctive visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23*(4), 948–961.

Ben Hamed, S., Duhamel, J. R., Bremmer, F., & Graf, W. (1996). Dynamic changes in visual receptive field organisation in the macaque lateral intraparietal area (LIP) during saccade preparation. *Society of Neuroscience, Abstracts Part 2*, 1619.

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, 408.

Blatt, G. L., Andersen, R. A., & Stoner, G. R. (1990). Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (LIP) in the macaque. *Journal of Computational Neurology, 299*, 421–445.

Brefczynski, J., & DeYoe, E. A. (1999). A physiological correlate of the spotlight of visual attention. *Nature Neuroscience, 2*, 370–374.

Bricolo, E., Gianesini, T., Fanini, A., Bundesen, C., & Chelazzi, L. (2002). Serial attention mechanisms in visual search: A direct behavioural demonstration. *Journal of Cognitive Neuroscience, 14*(7), 980–993.

Bushnell, M. C., Goldberg, M. E., & Robinson, D. L. (1981). Behavioural enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. *Journal of Neurophysiology, 46*(4), 755–772.

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience, 2*(7), 671–676.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature, 363*, 345–347.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cerebral Cortex, 11*, 761–772.

Clark, V. P., & Hillyard, S. A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *Journal of Cognitive Neuroscience, 8*(5), 387–402.

Cohessy, A. B., Posner, M. I., Rothbart, M. K., & Vecera, S. P. (1991). The development of inhibition of return in early infancy. *Journal of Cognitive Neuroscience, 3*, 345–350.

Colby, C. L., Duhamel, J. R., & Goldberg, M. E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *Journal of Neurophysiology, 76*(5), 2841–2852.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience, 22*, 319–349.

Connor, C. E., Callant, J. L., Preddie, D. C., & Van Essen, D. C. (1996). Responses in area V4 depend on the spatial relationship between stimulus and attention. *Journal of Neurophysiology, 75*, 1306–1308.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience, 3*, 292–297.

Corbetta, M., Shulman, G. L., Miezin, F. M., & Petersen, S. E. (1995). Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science, 270*, 802–805.

Crick, F. (1984). Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Science, USA, 81*, 4586–4590.

Danziger, S., Kingstone, A., & Snyder, J. J. (1998). Inhibition of return to successively stimulated locations in a sequential visual search paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 24*(5), 1467–1475.

Deco, G. (2001). Biased competition mechanisms for visual attention in a multimodular neurodynamical system. In S. Wermter, J. Austin, & D. Willshaw (Eds.), *Emergent neural computational architectures based on neuroscience: Towards neuroscience-inspired computing* (pp. 114–126). Heidelberg: Springer.

Deco, G., & Lee, T. S. (2002). A unified model of spatial and object attention based on inter-cortical biased competition. *Neurocomputing, 44–46*, 775–781.

De Frockert, J. W., Rees, G., Frith, C. D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science, 291*, 1803–1806.

De Kamps, M., & Van der Velde, F. (2001). Using a recurrent network to bind form, color and position into a unified percept. *Neurocomputing, 38–40*, 523–528.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London B, 353*, 1245–1255.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193–222.

Deubel, H., Bridgeman, B., & Schneider, W. X. (1998). Immediate post-saccadic information mediates space constancy. *Vision Research, 38*(20), 3147–3159.

Duhamel, J.-R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science, 255*, 90–92.

Duncan, J. (1984). Selective attention and the organisation of visual information. *Journal of Experimental Psychology: General, 113*, 501–517.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433–458.

Duncan, J., Humphreys, G. W., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology, 7*, 255–261.

Everling, S., Tinsley, C. J., Gaffan, D., & Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nature Neuroscience, 5*(7), 671–676.

Findlay, J. M., Brown, V., & Gilchrist, I. D. (2001). Saccade target selection in visual search: The effect of information from the previous fixation. *Vision Research, 41*(1), 87–95.

Fink, G. R., Dolan, R. J., Halligan, P. W., Marshall, J. C., & Frith, C. D. (1997). Space-based and object-based visual attention: Shared and specific neural domains. *Brain, 120*, 2013–2028.

Folk, C. L., & Remington, R. (1999). Can new objects override attentional control settings? *Perception and Psychophysics, 61*(4), 727–739.

Gerstner, W. (2000). Population dynamics of spiking neurons: Fast transients, asynchronous states, and locking. *Neural Computation, 12*, 43–89.

Ghandi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Science, USA, 96*, 3314–3319.

Ghose, G. M., & Ts'O, D. Y. (1997). Form processing modules in primate area V4. *The Journal of Neurophysiology, 77*(4), 2191–2196.

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature, 391*, 481–484.

Greene, H. H., & Rayner, K. (2001). Eye movements and familiarity effects in visual search. *Vision Research, 41*, 3763–3773.

Grossberg, S., & Raizada, R. D. S. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research, 40*, 1413–1432.

Hamker, F. H. (1998). The role of feedback connections in task-driven visual search. In D. Von Heinke, G. W. Humphreys, & A. Olson (Eds.), *Connectionist models in cognitive neuroscience: proceedings of the fifth neural computation and psychology workshop (NCPW'98)* (pp. 252–261). London: Springer.

Helmholtz, H. v. (1867). *Handbuch der physiologishen optik*. Leipzig: Voss.

Hodgson, T. L., Mort, D., Chamberlain, M. M., Hutton, S. B., O'Neill, K. S., & Kennard, C. (2002). Orbitofrontal cortex mediates inhibition of return. *Neuropsychologia, 1431*, 1–11.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception and Psychophysics, 57*(6), 787–795.

Hooge, I. T., & Erkelens, C. J. (1999). Peripheral vision and oculomotor control during visual search. *Vision Research, 39*(8), 1567–1575.

Hooge, I. T., & Frens, M. A. (2000). Inhibition of saccade return (ISR): Spatio-temporal properties of saccade programming. *Vision Research, 40*(24), 3415–3426.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience, 3*(3), 284–291.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature, 394*, 575–577.

Humphreys, G. W., Cinel, C., Wolfe, J., Olson, A., & Klempen, N. (2000). Fractionating the binding process: neuropsychological evidence distinguishing binding of form from binding of surface features. *Vision Research, 40*, 1569–1596.

Husain, M., Mannan, S., Hodgson, T., Wojciulik, E., Driver, J., & Kennard, C. (2001). Impaired spatial working memory across saccades contributes to abnormal search in parietal neglect. *Brain, 124*, 941–952.

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception and Psychophysics, 64*(6), 882–895.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 20*, 1489–1506.

Kastner, S., Pinsk, M., De Weerd, P., Desimone, R., & Ungerleider, L. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22*, 751–761.

Kusunoki, M., Gottlieb, J., & Goldberg, M. E. (2000). The lateral intraparietal area as a salience map: The representation of abrupt onset, stimulus motion, and task relevance. *Vision Research, 40*, 1459–1468.

Lanyon, L. J., & Denham, S. L. (2004a). A biased competition computational model of spatial and object-based attention mediating active visual search. *Neurocomputing*, in press.

Lanyon, L. J., & Denham, S. L. A biased competition model of spatial and object-based attention mediating active visual search. Submitted for publication.

Law, M. B., Pratt, J., & Abrams, R. A. (1995). Color-based inhibition of return. *Perception and Psychophysics, 57*(3), 402–408.

Lubow, R. E., & Kaplan, O. (1997). Visual search as a function of type of prior experience with targets and distractor. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 14–24.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial attention in areas V1, V2 and V4 of macaque visual cortex. *Journal of Neurophysiology, 77*, 24–42.

Lynch, J. C., Graybiel, A. M., & Lobeck, L. J. (1985). The differential projection of two cytoarchitectonic subregions of the inferior parietal lobule of macaque upon the deep layers of the superior colliculus. *Journal of Comparative Neurology, 235*, 241–254.

Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., Wong, E. C., Hinrichs, H., Heinze, H. J., & Hillyard, S. A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience, 2*(4), 364–369.

Mazzoni, P., Andersen, R. A., & Jordan, M. I. (1991). A more biologicall plausible learning rule than backpropagation applied to a networ model of cortical area 7a. *Cerebral Cortex, 1*(4), 293–307.

McAdams, C. J., & Maunsell, J. H. R. (2000). Attention to both space an feature modulates neuronal responses in macaque area V4. *Journal Neurophysiology, 83*(3), 1751–1755.

McPeek, R. M., Skavenski, A. A., & Nakayama, K. (2000). Concurre processing of saccades in visual search. *Vision Research, 40*(18 2499–2516.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisn of visual working memory in prefrontal cortex of the macaque. *T Journal of Neuroscience, 16*(16), 5154–5167.

Miller, E., Gochin, P., & Gross, C. (1993). Suppression of visual respons of neurons in inferior temporal cortex of the awake macaque by additi of a second stimulus. *Brain Research, 616*, 25–29.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxfor Oxford University Press.

Mitchell, J. F., Stoner, G. R., Fallah, M., & Reynolds, J. H. (200 Attentional selection of superimposed surfaces cannot be explained modulation of the gain of color channels. *Vision Research, 4* 1323–1325.

Mitchell, J., & Zipser, D. (2001). A model of visual-spatial memory acr saccades. *Vision Research, 41*, 1575–1592.

Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals microstimulation of frontal cortex. *Nature, 421*, 370–373.

Moran, J., & Desimone, R. (1985). Selective attention gates visu processing in the extrastriate cortex. *Science, 229*, 782–784.

Motter, B. C. (1993). Focal attention produces spatially selecti processing in visual cortical areas V1, V2, and V4 in the presence competing stimuli. *Journal of Neurophysiology, 70*(3), 909–919.

Motter, B. C. (1994a). Neural correlates of attentive selection for color luminance in extrastriate area V4. *The Journal of Neuroscience, 14*( 2178–2189.

Motter, B. C. (1994b). Neural correlates of feature selective memory a pop-out in extrastriate area V4. *The Journal of Neuroscience, 14*( 2190–2199.

Motter, B. C., & Belky, E. J. (1998a). The zone of focal attention duri active visual search. *Vision Research, 38*(7), 1007–1022.

Motter, B. C., & Belky, E. J. (1998b). The guidance of eye moveme during active visual search. *Vision Research, 38*(12), 1805–1815.

Mounts, J. R. W., & Melara, R. D. (1999). Attentional selection of obje or features: Evidence from a modified search task. *Perception Psychophysics, 61*(2), 322–341.

Niebur, E., Itti, L., & Koch, E. (2001). Controlling the focus of vis selective attention. In J. L. Van Hemmen, J. Cowan, & E. Dom (Eds.), *Models of neural networks IV* (pp. 247–276). New Yo Springer.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evide for objects as the units of attentional selection. *Nature, 401*, 584–5

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). neurobiological model of visual attention and invariant patt recognition based on dynamic routing of information. *The Journa Neuroscience, 13*, 4700–4719.

Pinilla, T., Cobo, A., Torres, K., & Valdes-Sosa, M. (2001). Attenti shifts between surfaces: Effects on detection and early brain potenti *Vision Research, 41*, 1619–1630.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Eff of parietal injury on covert orienting of attention. *Journal Neuroscience, 4*, 1863–1874.

Quaia, C., Optican, L. M., & Goldberg, M. E. (1998). The maintenanc spatial accuracy by the perisaccadic remapping of visual recep fields. *Neural Networks, 11*(7/8), 1229–1240.

Rainer, G., & Miller, E. K. (2002). Time course of object-related ne activity in the primate prefrontal cortex during a short-term mem task. *European Journal of Neuroscience, 15*(7), 1244–1254.

Reicher, G. M., Snyder, C. R. R., & Richards, J. T. (1976). Familiarity of background characters in visual scanning. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 522–530.

Renart, A., Moreno, R., de la Rocha, J., Parga, N., & Rolls, E. T. (2001). A model of the IT-PF network in object working memory which includes balanced persistent activity and tuned inhibition. *Neurocomputing, 38–40*, 1525–1531.

Reynolds, J. H., Alborzian, S., & Stoner, G. R. (2003). Exogenously cued attention triggers competitive selection of surfaces. *Vision Research, 43*(1), 59–66.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *The Journal of Neuroscience, 19*(5), 1736–1753.

Richards, J. T., & Reicher, G. M. (1978). The effect of background familiarity in visual search. *Perception and Psychophysics, 23*, 499–505.

Robinson, D. L., Bowman, E. M., & Kertzman, C. (1995). Covert orienting of attention in macaques. II. Contributions of parietal cortex. *Journal of Neurophysiology, 74*(2), 698–712.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395*, 376–381.

Rolls, E., & Deco, G. (2002). *Computational neuroscience of vision*. Oxford, UK: Oxford University Press.

Romo, R., Brody, C. D., Hernandez, A., & Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature, 399*(6735), 470–473.

Ross, J., Morrone, M. C., & Burr, D. C. (1997). Compression of visual space before saccades. *Nature, 386*, 598–601.

Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences, 24*(2), 113–121.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience, 5*(7), 631–632.

Sapir, A., Soroker, N., Berger, A., & Henik, A. (1999). Inhibition of return in spatial attention: Direct evidence for collicular generation. *Nature Neuroscience, 2*, 1053–1054.

Schall, J. D. (2002). The neural selection and control of saccades by the frontal eye field. *Philosophical Transactions of the Royal Society of London B Biological Science, 357*, 1073–1082.

Scialfa, C. T., & Joffe, K. M. (1998). Response times and eye movements in feature and conjunction search as a function of target eccentricity. *Perception and Psychophysics, 60*(6), 1067–1082.

Shen, J., Reingold, E. M., & Pomplun, M. (2000). Distractor ratio influences patterns of eye movements during visual search. *Perception, 29*(2), 241–250.

Shen, J., Reingold, E. M., & Pomplun, M. (2003). Guidance of eye movements during conjunctive visual search: The distractor ratio effect. *Canadian Journal of Experimental Psychology, 57*(2), 76–96.

Snyder, J. J., & Kingstone, A. (2000). Inhibition of return and visual search: How many separate loci are inhibited? *Perception and Psychophysics, 62*(3), 452–458.

Snyder, J. J., & Kingstone, A. (2001). Inhibition of return at multiple locations in visual search: When you see it and when you don't. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 54*(4), 1221–1237.

Somers, D. C., Dale, A. M., Seiffert, A. E., & Tootell, R. B. H. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proceedings of the National Academy of Science, USA, 96*, 1663–1668.

Sommer, P., & Andersen, R. A. (1998). Electrical microstimulation distinguishes distinct saccade-related areas in the posterior parietal cortex. *Journal of Neurophysiology, 80*, 1713–1735.

Thiele, A., Henning, P., Kubischik, M., & Hoffmann, K. P. (2002). Neural mechanisms of saccadic suppression. *Science, 295*(5564), 2460–2462.

Tipper, S. P., Weaver, B., & Watson, F. L. (1996). Inhibition of return to successively cued spatial locations: Commentary on Pratt and Abrams (1995). *Journal of Experimental Psychology: Human Perception and Performance, 22*(5), 1289–1293.

Toth, L. J., & Assad, J. A. (2002). Dynamic coding of behaviourally relevant stimuli in parietal cortex. *Nature, 415*, 165–168.

Tolias, A. S., Moore, T., Smirnakis, S. M., Tehovnik, E. J., Siapas, A. G., Schiller, P. H. (2001). Eye movements module visual receptive fields of V4 neurons. Neuron, 29, 757–767.

Trappenberg, T. P., Dorris, M. C., Munoz, D. P., & Klein, R. M. (2001). A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience, 13*, 256–271.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 194–214.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology, 40*, 201–237.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Treue, S., & Martinez Trujillo, J. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 339*(6736), 575–579.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. W. J. Mansfield (Eds.), *Analysis of visual behaviour* (pp. 549–586). Cambridge, MA: MIT Press.

Usher, M., & Niebur, E. (1996). Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *Journal of Cognitive Neuroscience, 8*(4), 311–327.

Valdes-Sosa, M., Bobes, M. A., Rodriguez, V., & Pinilla, T. (1998). Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. *Journal of Cognitive Neuroscience, 10*(1), 137–151.

Valdes-Sosa, M., Cobo, A., & Pinilla, T. (2000). Attention to object files defined by transparent motion. *Journal of Experimental Psychology: Human Perception and Performance, 26*(2), 488–505.

Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology, 51*, 167–194.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and popout in visual search. *Perception and Psychophysics, 56*, 495–500.

Williams, D. E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude. *Psychonomic Bulletin and Review, 8*(3), 476–488.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review, 1*, 202–238.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 419–433.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science, 12*(3), 219–224.

Zeki, S. (1993). *A vision of the brain*. Oxford, UK: Blackwell.

Zhu, J. J., & Lo, F. S. (1996). Time course of inhibition induced by a putative saccadic suppression circuit in the dorsal lateral geniculate nucleus of the rabbit. *Brain Research Bulletin, 41*(5), 281–291.

Zipser, D., & Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature, 331*(6158), 679–684.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Neurocomputing 58–60 (2004) 655–662
NEUROCOMPUTING
www.elsevier.com/locate/neucom

# A biased competition computational model of spatial and object-based attention mediating active visual search

Linda J. Lanyon*, Susan L. Denham

*Centre for Theoretical & Computational Neuroscience, University of Plymouth, Drakes Circus, Plymouth PL4 8AA, UK*

**Abstract**

Colour is more important than form in driving the scan path during active visual feature conjunction search. A model shows how feature-based attention in V4 provides information to LIP allowing it to represent behaviourally relevant locations and attract attention to target coloured locations. Attentional effects in V4 develop such that an initial spatial attention focus, scaled according to stimulus density, becomes object-based later in the response. Thus, this is the first model to replicate, with temporal precision, different attentional effects observed in single-cell studies of V4 and IT, in addition to the systems level scan path behaviour.


*Keywords:* Visual attention; Visual search; Biased competition; Object-based attention; Saccades

## 1. Introduction

The biased competition hypothesis [5] is currently a very topical theory used to explain attention effects observed in single-cell studies of the primate visual system (e.g. [1,2,15,16]). Biased competition suggests that neuronal responses are determined by competitive interactions that are subject to a number of biases, such as "bottom-up" stimulus information and "top-down" cognitive requirements, in particular a working memory template of the target object from prefrontal cortex. When more than one stimulus is present within a cell's receptive field, the response of the cell is determined by which of the stimuli are attended. Biased competition models have recently

* Corresponding author. Tel.: +44-1752-23-25-67; fax: +44-1752-23-33-49.
*E-mail address:* linda.lanyon@plymouth.ac.uk (L.J. Lanyon).

0925-2312/$ - see front matter 
doi:10.1016/j.neucom.2004.01.110

developed from small scale (e.g. [20]) to more complex systems level models (e.g. [4]). Here, we describe a model for *active* visual search, where retinal inputs change at each fixation as the focus of attention moves. The model is able to simulate both spatial and object-based attentional modulations seen in single-cell studies of inferior temporal (IT) region [1], and extrastriate area V4 [2,12,16]. This appears to be the first model to accurately replicate the time courses of these different attentional modulations. At the systems level, the model is able to reproduce scan path behaviour found [18] in monkeys performing feature conjunction search, where fixations tended to land near stimuli of the target colour. Target orientation did not appear to influence the scan path. The model achieves this by means of a "cross-stream" interaction between extrastriate areas in the ventral pathway leading to temporal cortex and the dorsal pathway leading to parietal cortex.

A detailed description of the model's simulation of single cell recordings in IT [1] and V4 [2,12,16] is given by Lanyon and Denham [10] and a subsequent examination of the scan path under different scene conditions is provided by Lanyon and Denham [11]. Here, we investigate the effect of the aperture of the initial spatial attention focus being scaled when stimulus density varies across the scene. We also consider the sensitivity of the model to the weighting of V4 connections to parietal area LIP and the strength of novelty feedback.

## 2. The model

The model focuses on emergent attentional effects in intermediate stages of the visual system and Fig. 1 shows the areas modelled. Featural "pre-processing" is provided by retinal and V1 modules, but these do not form part of the dynamics of the main active vision portion of the system. Attention arises as an emergent property of this portion of the system in modules representing V4, IT and the lateral intraparietal area (LIP). These modules use mean field population dynamics [6], where representation is at the level of cell populations, known as pools or assemblies. The model operates by moving its retina around the scene. The size of the dynamic cortical areas is dependant on the size of the retina, which can be adjusted for each simulation. The model's ventral stream (running from V1 to V4 and IT) operates as a feature/object processing hierarchy with receptive field sizes increasing towards IT and encoded features becoming more complex, such that objects are encoded in an invariant manner in IT. In V1 and V4 both orientation and colour features are encoded. LIP encodes the locations of behaviourally relevant features [3], perhaps acting as a "salience map" [7] and may form part of the cortical network responsible for selecting possible targets for saccades. In the modelled LIP, spatial locations compete and activity is used to decide the next location to be fixated. Featural information from V4 is integrated within the spatial competitive framework in LIP and this allows it to represent behaviourally relevant locations most strongly, and guide the scan path accordingly.

Object-based attention has not been observed in V4 until at least 150 ms post-stimulus [2,16]. However, spatial attention seems able to module responses from the onset of the stimulus-related response at $\sim$ 60 ms post-stimulus, and even enhances baseline

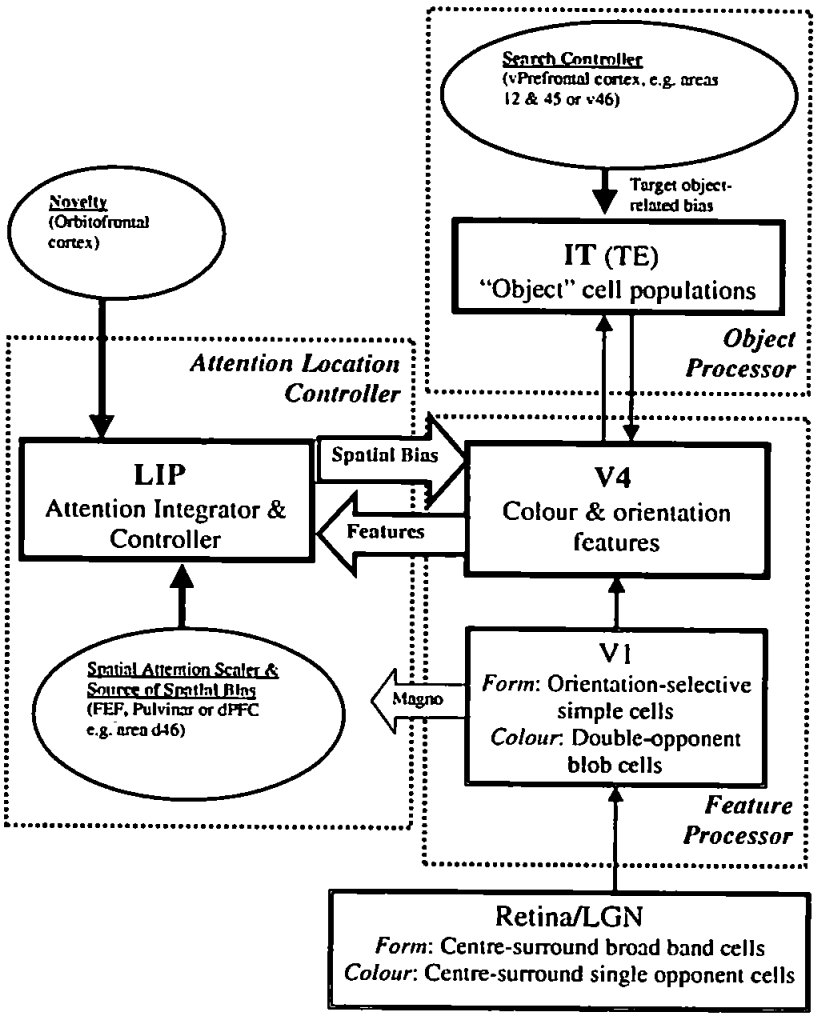



Fig. 1. Overview of the model.

firing in advance of the stimulus-related response [12]. In the model, both effects are simulated at the appropriate times by means of two different types of bias to V4. The spatial enhancement is due to a spatial bias to competition in LIP (where spatial attention is also found: e.g. [3,9]), which then biases V4. The source of the spatial bias is suggested to be the frontal eye field (FEF), because microstimulation of FEF leads to spatial modulation of V4 [14]. The spatial bias results in the creation of an initially spatial *attention window* (AW). The aperture of this window is scaled according to coarse resolution information reflecting local-stimulus density [17], which is assumed to be conveyed rapidly by the magnocellular pathway to parietal cortex. All other

information within the system is assumed to be derived from the parvocellular pathway. Thus, during a scan path, the size of the AW is dynamic, being scaled according to stimulus density found around any particular fixation point. The following gives the radius of the AW and is inspired by the psychophysical findings of Motter and Belky [17]:

$$\mathrm{AW}_{\mathrm{rad}} = \min\left[\mathrm{Round}\left[\sqrt{\frac{m*n}{d}}\,2\right]\,\min(m,n)\right], \tag{1}$$

where $d$ is the number of non-zeros in $f$ below, $m,n$ are the dimensions of the retinal image:

$$f = \psi\left(\sum_{k=1}^{K} I_{ijrk}\right), \tag{2}$$

where $I_{ijrk}$ is the output of V1 simple cells, selective to $K$ orientations, over the area of the retinal image, $r$ is the spatial resolution of the orientation information; set to the lowest resolution detected, $\psi$ is a function that removes the lowest 95% of activity and reduces to zero the activity at points where a neighbour has been found within a Euclidean distance equal to the length of the stimuli (simple bar stimuli are used).

Over time, attention gradually becomes object-based. Object-based attention develops in parallel across the cortical region and is not constrained by the spatial AW. However, object-based attention is facilitated by spatial attention such that responses are strongest within the AW, resulting in a combined attentional effect [13]. It is suggested that object-based attention takes longer than spatial attention to develop because it is reliant on the resolution of competition between objects and features. Object-based attention results from an object-related bias from prefrontal cortex (responses of prefrontal "late" neurons [19] are modelled) to IT. Objects in IT provide a "top-down" bias to features in V4, such that features not related to the object are suppressed, i.e. a feature-based attention effect. Under conditions of visual search involving a colour (or luminance-defined) target, object-based attention in V4 results in target coloured (or target luminance) stimuli becoming effectively "highlighted" across the scene in V4 [16] and this enables V4 to bias LIP towards representing these locations most strongly. Object-based effects also operate within the orientation feature dimension in V4 but the connection from colour features in V4 to LIP is slightly stronger than the connection from orientation features in V4. This allows the scan path to be drawn to target coloured locations across the scene.

Inhibition of return (IOR) in the scan path is implemented by means of a frontal bias to LIP related to the *novelty* of the location. IOR has been suggested to be influenced by recent event/reward associations linked to orbitofrontal cortex [8], which is a possible source of the bias.

## 3. Simulation results

During a scan path, next fixation points tend to be chosen from within the current AW because this area of LIP is subject to a spatial enhancement of activity. For

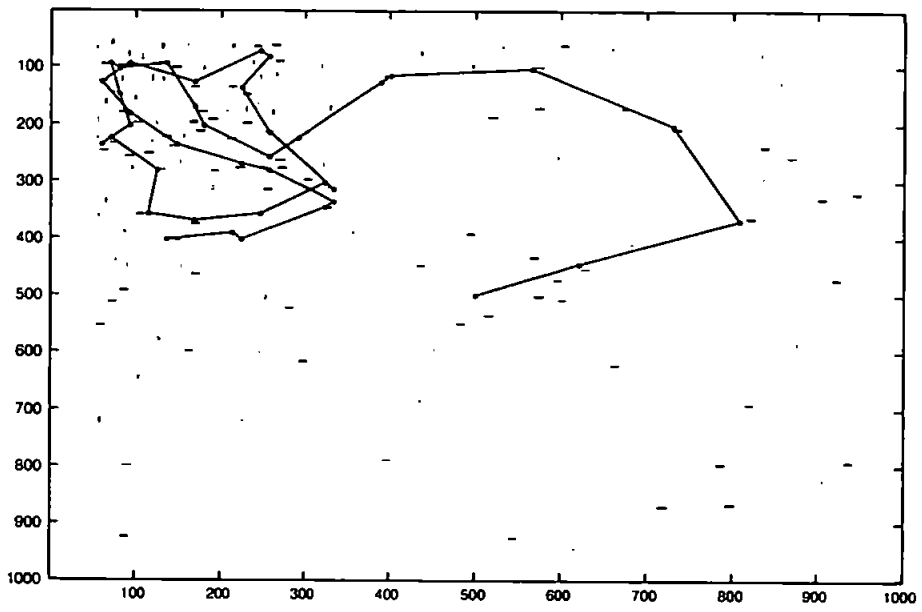



Fig. 2. Scan path over a scene with varying stimulus density. Attention is attracted to the locations of target coloured stimuli (in this case the target colour is green, shown here as black; the non-target colour is red, shown as grey). From an initial fixation point in the centre of the image, large saccades are made within the sparse area of the scene but subsequent saccades within the dense area are shorter.

any given retinal image size, the AW is much smaller for dense scenes than that within sparse scenes. This results in small saccades being made in dense images and larger saccades in sparse scenes. The AW is dynamically scaled according to local information as the retina moves around the scene. Therefore, when stimulus density varies across the scene, the AW changes size during the scan path so that dense areas are examined more thoroughly by a series of smaller saccades, as shown in Fig. 2. The AWs for two of the fixations in contrasting areas of this scan path are shown in Fig. 3.

Spatial attention has an early effect on V4 responses and subsequent responses are modulated by the combined effect of the spatial AW and object-based attention [13]. Object-based attention in V4 is able to influence activity in LIP such that the scan path is attracted to areas containing the target colour, as shown in Fig. 2. The amount of non-target coloured locations attracting attention is influenced by the weight of the novelty bias to LIP, the weight of connections from V4 to LIP and the strength of object-related feedback from IT to V4. The latter is suggested by Lanyon and Denham [10] to be affected by training, i.e. familiarity with objects and the task. The relative weight of connection from colour and orientation assemblies in V4 to LIP has an effect on the number of non-target coloured locations attracting attention. However, target coloured locations continue to dominate the scan path even when the colour

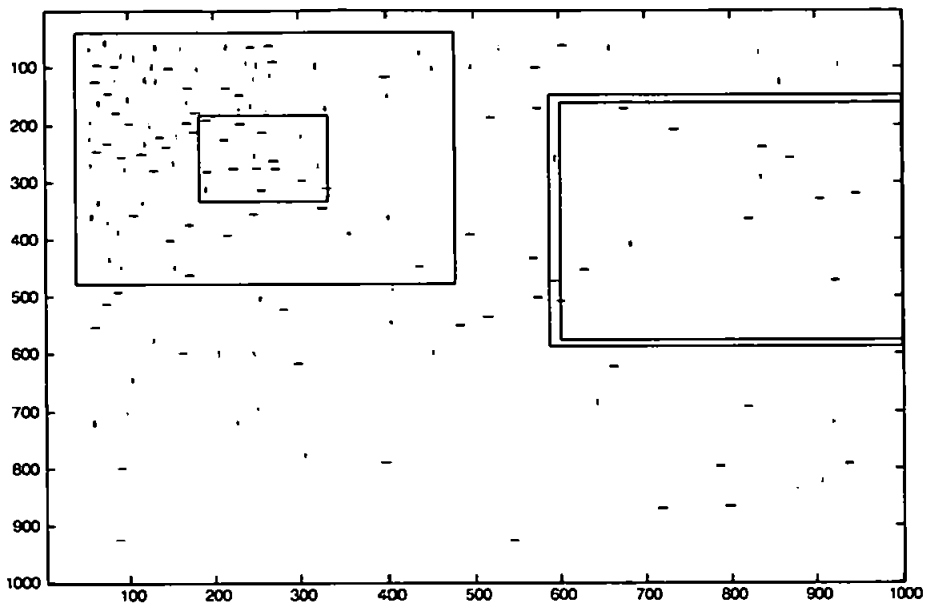



Fig. 3. Shows the difference in aperture of the AW for two points during the scan path shown in Fig. 2. The retinal image is shown as the outer of the plotted boxes. In the dense stimulus region, the AW (inner boxes plotted within each retinal image box) is much smaller than that in the sparse region.

connection is only very slightly stronger than the orientation connection. The strength of novelty bias to LIP appears to have a large effect on the scan path. If the novelty bias to LIP is very strong, it does not allow LIP to discriminate any stimuli within the vicinity of a previous fixation and this tends to reduce the choice of stimulus locations for the next fixation. In very sparse scenes, this can result in blank areas of the display becoming more active in LIP than stimulus locations. Therefore, when the novelty bias is set to a very high value, the scan path tends to select more blank areas (20% of fixations were found in blank areas in [18]).

## 4. Conclusion

This model has been used to simulate attentional effects at the single-cell level, with accurate onset times for both spatial and object-based attention in IT and V4 [10]. At the behavioural level, it has also been able to simulate visual search scan paths found in monkeys performing feature conjunction search [18]. IT feedback to V4 allows object-based attention to develop therein and this influences spatial competition in LIP such that the scan path is attracted to behaviourally relevant locations. Here, we find that, in addition to other factors, the scan path is influenced by the strength of novelty feedback to LIP. The novelty bias may reflect a cognitive strategy to seek

new locations and could reflect a previous event-reward association [8]. During search within a scene where stimulus density varies, large saccades tend to traverse sparse areas but much shorter saccades take place within dense regions, allowing detailed investigation of the area.

## References


[1] L. Chelazzi, E.K. Miller, J. Duncan, R. Desimone, A neural basis for visual search in inferior temporal cortex, Nature 363 (1993) 345–347.

[2] L. Chelazzi, E.K. Miller, J. Duncan, R. Desimone, Responses of neurons in macaque area V4 during memory-guided visual search, Cereb. Cortex 11 (2001) 761–772.

[3] C.L. Colby, J.R. Duhamel, M.E. Goldberg, Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area, J. Neurophysiol. 76 (5) (1996) 2841–2852.

[4] G. Deco, T.S. Lee, A unified model of spatial and object attention based on inter-cortical biased competition, Neurocomp. 44–46 (2002) 775–781.

[5] R. Desimone, J. Duncan, Neural mechanisms of selective visual attention, Ann. Rev. Neurosci. 18 (1995) 193–222.

[6] W. Gerstner, Population dynamics of spiking neurons: fast transients, asynchronous states, and locking, Neural Comput. 12 (2000) 43–89.

[7] J.P. Gottlieb, M. Kusunoki, M.E. Goldberg, The representation of visual salience in monkey parietal cortex, Nature 391 (1998) 481–484.

[8] T.L. Hodgson, D. Mort, M.M. Chamberlain, S.B. Hutton, K.S. O'Neill, C. Kennard, Orbitofrontal cortex mediates inhibition of return, Neuropsychologia 431 (2002) 1–11.

[9] S. Kastner, M. Pinsk, P. De Weerd, R. Desimone, L. Ungerleider, Increased activity in human visual cortex during directed attention in the absence of visual stimulation, Neuron 22 (1999) 751–761.

[10] L.J. Lanyon, S.L. Denham, A biased competition model of spatial and object-based attention mediating active visual search, Vision Res. (2004), submitted for publication.

[11] L.J. Lanyon, S.L. Denham, A model of active visual search with object-based attention guiding scan paths, Neural Networks (2004), submitted for publication.

[12] S.J. Luck, L. Chelazzi, S.A. Hillyard, R. Desimone, Neural mechanisms of spatial attention in areas V1, V2 and V4 of macaque visual cortex, J. Neurophysiol. 77 (1997) 24–42.

[13] C.J. McAdams, J.H.R. Maunsell, Attention to both space and feature modulates neuronal responses in macaque area V4, J. Neurophysiol. 83 (3) (2000) 1751–1755.

[14] T. Moore, K.M. Armstrong, Selective gating of visual signals by microstimulation of frontal cortex, Nature 421 (2003) 370–373.

[15] J. Moran, R. Desimone, Selective attention gates visual processing in the extrastriate cortex, Science 229 (1985) 782–784.

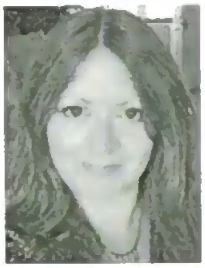
[16] B.C. Motter, Neural correlates of attentive selection for color or luminance in extrastriate area V4, J. Neurosci. 14 (4) (1994) 2178–2189.

[17] B.C. Motter, E.J. Belky, The zone of focal attention during active visual search, Vision Res. 38 (7) (1998) 1007–1022.

[18] B.C. Motter, E.J. Belky, The guidance of eye movements during active visual search, Vision Res. 38 (12) (1998) 1805–1815.

[19] R. Romo, C.D. Brody, A. Hernandez, L. Lemus, Neuronal correlates of parametric working memory in the prefrontal cortex, Nature 399 (6735) (1999) 470–473.

[20] M. Usher, E. Niebur, Modeling the temporal dynamics of IT neurons in visual search: a mechanism for top-down selective attention, J. Cog. Neurosci. 8 (4) (1996) 311–327.



**Linda Lanyon** has a 1st class honours degree in Computing & Informatics and comes from a background in computer systems development, most recently acting as IT Advisor to a major UK e-government project. She is currently completing the final year of her PhD, from which this work is drawn, at the Centre for Theoretical & Computational Neuroscience, University of Plymouth, UK.

**Dr Sue Denham** obtained her PhD in 1995 from the University of Plymouth where she holds the post of Principal Lecturer in the Centre for Theoretical & Computational Neuroscience. Her research interests lie in developing theoretical and computational models of perception and learning, with particular interest in the representation of timbre and the segregation and grouping of sounds within natural acoustic environments.

# A Model of Object-Based Attention That Guides Active Visual Search to Behaviourally Relevant Locations

Linda Lanyon and Susan Denham

Centre for Theoretical and Computational Neuroscience,
University of Plymouth, U.K
linda.lanyon@plymouth.ac.uk

**Abstract.** During active visual search for a colour-orientation conjunction target, scan paths tend to be guided to target coloured locations (Motter & Belky, 1998). An active vision model, using biased competition, is able to replicate this behaviour. At the cellular level, the model replicates spatial and object-based attentional effects over time courses observed in single cell recordings in monkeys (Chelazzi et al., 1993, 2001). The object-based effect allows competition between features/objects to be biased by knowledge of the target object. This results in the suppression of non-target features (Chelazzi et al., 1993, 2001; Motter, 1994) in ventral "what" stream areas, which provide a bias to the spatial competition in posterior parietal cortex (LIP). This enables LIP to represent behaviourally relevant locations (Colby et al., 1996) and attract the scan path. Such a biased competition model is extendable to include further "bottom-up" and "top-down" factors and has potential application in computer vision.

## 1 Introduction

The biased competition hypothesis [1][2][3][4] is currently very influential in the study of visual attention at many levels ranging from neurophysiological single cell studies to brain imaging, evoked potentials and psychophysics. The hypothesis suggests that responses of neurons, such as those that encode object features, are determined by competitive interactions. This competition is subject to a number of biases, such as "bottom-up" stimulus information and "top-down" cognitive requirements, for example a working memory template of the target object during visual search. Building on early small-scale models using this hypothesis, e.g.[5][6], systems level models, e.g. [7][8][9], have been able to replicate a range of experimental data. However, there has been no systems level modelling of the biased competition hypothesis for active visual search, where retinal inputs change as the focus of attention is shifted. The model presented here uses biased competition and adopts an active vision approach such that, at any particular fixation, its cortical modules receive retinal inputs relating to a portion of the entire scene. This allows the model to process a smaller area of the image at any fixation and provides a realistic model of normal everyday search.

Traditionally, visual attention was thought to operate as a simple spatial "spotlight" [10][11][12] that acted to enhance responses at a particular location(s) and it is clear that attention can produce spatially specific modulation of neuronal responses, e.g.[13][14],

L. Paletta et al. (Eds.): WAPCV 2004, LNCS 3368, pp. 42–56, 2005.

and behaviour [15]. However, many studies in psychophysics [16][17], functional magnetic resonance imaging [18], event-related potential recordings [19][20] and single cell recordings [21][22][23] provide convincing evidence for attention operating in a more complex object-based manner. However, these findings are confounded by issues relating to whether the selection is object-based, feature-based or surface-based. Here we use the term object-based in the sense that it refers an attentional modulation of cellular responses on the basis of the features of a target object, as used by [7][21][23]. The term feature-based is also appropriate. Many different mechanisms may be responsible for the range of object-based effects seen experimentally and not all are addressed here. For example, there is no attempt to provide Gestalt-like perceptual grouping, but an extension to the model's feature processing modules to allow local lateral connectivity across similar features could produce such an effect.

Event-related potential [24] and single cell recordings suggest that spatial attention may be able to modulate responses earlier than object-based attention. For example, in monkey lateral intraparietal area (LIP), an early spatial enhancement of responses has been recorded from single cells in anticipation of the stimulus [25] and has been seen in imaging of the possible human homologue of LIP [26][27][28]. In ventral pathway area V4, which encodes form and colour features [29], spatial modulation of baseline responses has been recorded from single cells in advance of the sensory response and the earliest stimulus-invoked response at 60ms post-stimulus was also spatially modulated [14]. However, object-based effects in V4 appear to take longer to be resolved and have been recorded from approximately 150ms post-stimulus [23][30][31]. Such modulation, leading to the response of the cell being determined by which object within its receptive field is attended, is known as the 'target effect' [23]. In inferior temporal cortex (IT), object-based modulation of responses during a delay period following a cue stimulus has been recorded prior to the onset of the response relating to the search array [32]. However, the initial sensory response invoked by the search array is generally not significantly modulated by the target object compared to the strong target effect developing from 150-200ms post-onset [21][32]. The model presented here replicates these spatial and object-based effects, at the neuronal level, with temporal precision. Such effects at the neuronal level allow the model, at the systems level, to produce the search behaviour similar to that seen in humans [33][34] and monkeys [35].

A "cross-stream" interaction between the ventral "what" pathway, responsible for the encoding of features and objects and leading from primary visual cortex (V1) through V2 and V4 to temporal cortex, and the dorsal "where" pathway, leading from V1 to parietal cortex [36][37], allows the search scan path to be drawn to behaviourally relevant locations rather than less relevant stimulus locations or blank areas. Object-based attention in the model's ventral pathway biases the dorsal module (LIP) to represent the location of target features most strongly [25][38]. In addition, this connection allows certain features to have priority in attracting attention. In particular, the model is able to reproduce the scan path behaviour of monkeys searching for a colour-orientation feature conjunction target [35] where fixations tended to land within 1 ° of stimuli (only 20% fell in blank areas of the display) and these stimuli tended to be target coloured (75% of fixations landed near target coloured stimuli and only 5% near non-target coloured

stimuli). The use of object-based attention within the model to guide the scan path is the focus of this paper.

## 2 The Model

Fig. 1 shows a schematic overview of the model, which is formally described in [39]. It consists of a retina and V1, which detect form and colour features in a retinotopic manner, and three dynamic modules representing areas V4, IT and LIP, which are modelled using a mean field approach [40], in a similar manner to the model in [7], where the level of representation is a population of cells with similar properties, known as a cell assembly. The size of the retina can be varied for any particular simulation because the size of the cortical modules is scaled according to retina size. Thus, a larger retina may be used where biologically valid predictions are to be made, or a smaller retina where processing speed needs to be optimised (the only restriction being that a very small retina tends to lead to a higher proportion of fixations landing in blank areas of very sparse scenes due to lack of stimuli within the limited retina). For reasons of computational simplicity and speed, V1 is not included in the dynamic portion of the system and there is

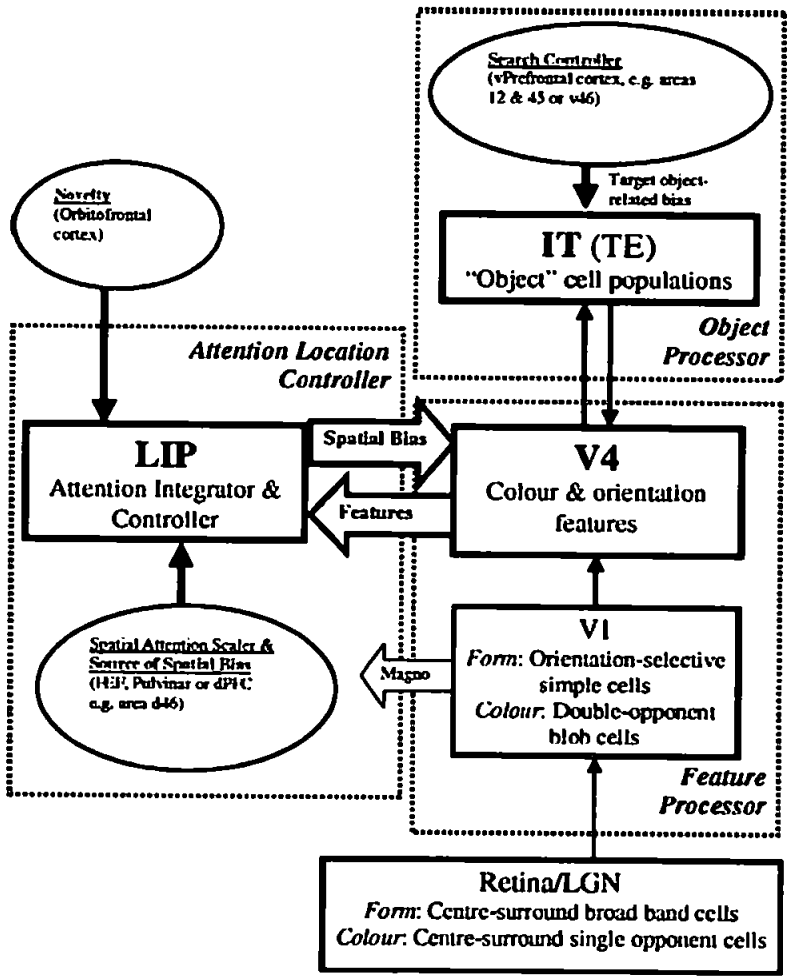



**Fig. 1.** Overview of the model

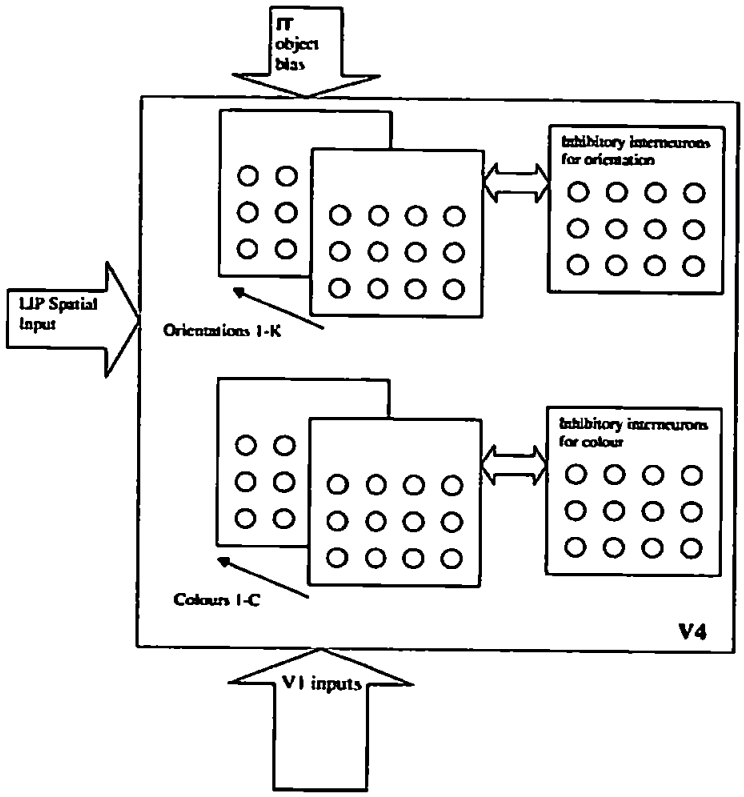



**Fig. 2.** Competition in V4

no attempt in this version of the model to reproduce attentional effects therein, although such effects have been observed [22][41]. Intermediate stages of visual processing, such as area V4, are the focus of this model. The V1 module contains a neuron at every pixel location and features detected in V1 are fed forward to V4 in a convergent manner over the V4 receptive field area. Thus, V4 is retinotopic with biologically plausible [42] larger receptive fields than found in V1 (the size of filters used in V1 determines the ratio of pixels to degrees of visual angle). Stimuli are coloured oriented bars, chosen to match the feature conjunctions used by [35], in order to reproduce the effects seen in that psychophysical experiment. V4 neurons are known to be functionally segregated [43] and the V4 module encodes colour and orientation in separate feature layers, a common modelling simplification. Different features of the same type at the same location compete via inhibitory interneuron assemblies (i.e. different colours within the same V4 receptive field area compete and different orientations within the same V4 receptive field area compete), as shown in fig. 2. This provides the necessary competition between features within the same receptive field that has been found in single cell recordings [44][14][21][23]. V4 featural information is fed forward to activate non-retinotopic invariant object populations in IT. Thus, the model's ventral pathway (V1, V4 and IT) operates as a feature/object processing hierarchy with receptive field sizes increasing towards IT, where receptive fields span the entire retina. It is known that anterior areas of IT, such as area TE, are not retinotopic but have large receptive fields and encode objects in an invariant manner [42].

Dorsal pathway module LIP contains a cell assembly for every retinotopic location in V4. Each LIP assembly is reciprocally connected to the assembly in each feature layer in V4 at that location. LIP provides a spatio-featural map that is used to control the spatial focus of attention and, hence, fixation. Although responses in monkey LIP have been found to be independent of motor response [45], LIP is thought to be involved in selecting possible targets for saccades and has connections to superior colliculus and the frontal eye field (FEF), which are involved in saccade generation. Furthermore, direct electrical stimulation of LIP has elicited saccades [46]. Competition in the model LIP is spatial and the centre of the receptive field of the most active assembly at the end of the fixation is chosen as the next fixation point. During fixation a spatial bias creates an early spatial attention window (AW), which is scaled according to local stimulus density (on the basis of low resolution orientation information assumed to be conveyed rapidly to parietal cortex through the magnocellular pathway), as described by [47]. This reflects the scaling of the area around fixation within which targets can be detected [48]. This spatial bias to LIP creates an anticipatory spatial attention effect, as has been found in single cell recordings in this area [25]. LIP provides a spatial bias to V4 and this creates an early spatial attention effect therein, modulating baseline firing before the onset of the stimulus evoked response and, subsequently, modulating this response from its onset at ~60ms post-stimulus, as found by [14]. The source of the spatial bias could be FEF because stimulation of this area leads to a spatially specific modulation of neuronal response in V4 [49]. The spatial bias from LIP to V4 also provides some binding across feature types in V4.

Within V4, the early spatial focus of attention becomes object-based over time. This is due to the fact that V4 is subject to two concurrent biases: A spatial bias from LIP and an object-related bias from IT. Thus, object-based attention evolves in the model's ventral pathway due to its object-related feedback biases. The development of object-based attention in V4 is dependent on the resolution of competition between objects in IT. In addition to the feed forward inputs from V4, competition between objects in IT is biased by a feedback current from prefrontal cortex, which is assumed to hold a working memory template of the target object. IT feeds back an inhibitory bias to V4 that suppresses features not related to the object. Over time, the target object is able to win the competition in IT, and then target features become enhanced whilst non-target features are suppressed across V4, as has been found in single cell recordings in monkeys [30][31] over the same time course. Once a significant object-based effect is established in IT, a saccade is initiated 70ms later, to reflect motor preparation, and replicate the timing of saccade onset following the target object effect seen in single cell recordings in IT [21] and V4 [23].

Following this object-based "highlighting" of target features in parallel across V4 [30][31], the featural inputs from V4 to LIP provide a bias that results in locations containing target-related features becoming most active in LIP. Thus, LIP is able to represent the locations of salient, behaviourally relevant stimuli [25][38]. Prior to the onset of object-based attention in the ventral stream (which occurs ~150ms post-stimulus, as found by [21][23][32], the initial sensory response in LIP tends to represent locations containing stimuli equally, but more strongly than blank areas. The connection from V4 to LIP is set such that colour features provide a slightly stronger bias than orientation

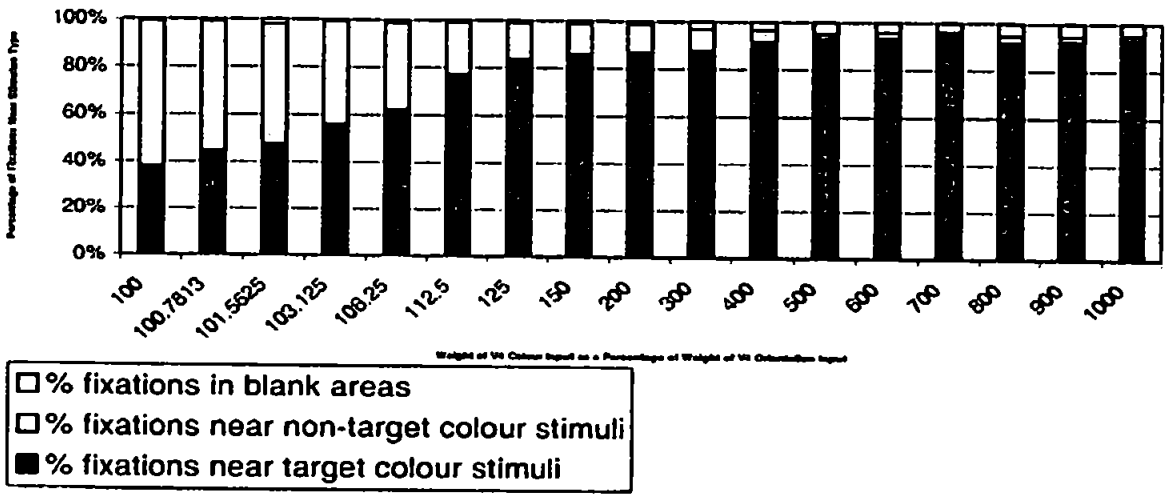



**Fig. 3.** The effect of increasing the relative weight of V4 colour feature input to LIP. When V4 colour features are marginally more strongly connected to LIP than V4 orientation features, the scan path is attracted to target coloured stimuli in preference to stimuli of the target orientation

features. Thus, LIP tends to represent locations containing target coloured stimuli more strongly than locations containing the target orientation but non-target colour. Therefore, colour is given a priority in guiding the scan path, as found by [35]. The difference in strength of connection of the V4 features to LIP need only be marginal in order to achieve this effect. Fig. 3 shows the effect of adjusting the relative connection weights.

The use of this active vision paradigm with a moving retina means that inhibition of return (IOR) in the scan path cannot be implemented by inhibition of previously active cortical locations in retinotopic coordinates. Parietal damage is linked to an inability to retain a spatial working memory of searched locations across saccades so that locations are repeatedly re-fixated [50]. Here, IOR in the scan path is provided by a novelty-related bias to LIP. The source of such a bias could be a frontal area, such as orbitofrontal cortex, which is linked with event/reward associations and, when damaged, affects IOR [51]. A world/head-based map reflects the novelty of all locations in the scene and, hence, their potential reward. Initially, the novelty of all locations in the scene is high. Then, when attention is withdrawn from a location, locations in the immediate vicinity of the fixation point have their novelty reduced to a low value that climbs in a Gaussian fashion with distance from fixation such that peripheral locations within the AW have neutral novelty. Novelty recovers over time such that IOR is present at multiple locations [52][53][54], the magnitude of the effect decreases approximately linearly from its largest value at the most recently searched location and several previous locations are affected [55][56].

## 3 Results

### 3.1 Neuronal Level Results

At the neuronal level, attentional effects within V4 evolve from being purely spatial to being object-based, over time courses found in single cell recordings from monkeys [14][23][30][31]. The enhancement of target features across V4 is particularly important to the guidance of the scan path and fig. 4 shows the development of object-based

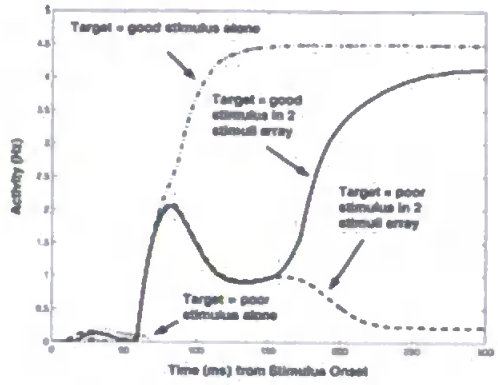



**Fig. 4.** Comparison of V4 single assembly response: Shows the replication of [23] figure 5 from array onset until the time of the saccade. Here, saccades have been suppressed but, for the two stimuli case, would have occurred at ~235-240ms post-stimulus, the same onset time as found by [23]. A good stimulus for the cell (i.e. one that causes a high response) presented alone in the receptive field produces a high response: Top line. When a poor stimulus is presented alone the cell's response is suppressed: Bottom line. When both stimuli are simultaneously presented within the receptive field, the high response normally invoked by the good stimulus is reduced due to the competing presence of the poor stimulus. However, once object-based effects begin from ~150ms, the response is determined by which of the two stimuli is the target, with activity tending towards that of the single stimulus case for the target object. The 2 stimuli case when the target is a poor stimulus for the cell shows the suppression of the non-target responses (i.e. the cell's response to the good stimulus) that enables V4 to represent target stimuli most strongly and bias LIP to guide the scan path to these locations

attention from ~150ms post-stimulus in V4, which mirrors that recorded by [23]. This effect in V4 is due to object-related feedback in IT and depends on the development of the target object effect in IT, which is shown in fig. 5. Fig. 5 replicates the time course of the development of a significant target object effect in the sensory response, as found by [21] and [32]. Figs. 4 and 5 replicate data from the onset of the sensory response (the start of the fixation in the model) and do not address the maintenance of cue-related activity during a delay period, which is found in IT [21][32]. In some IT cells there is a slight modulation of early sensory responses on the basis of the target object [32]. However, this is insignificant compared to the target object effect later in the response and has been suggested [32] to be due to a continuation of elevated firing during the delay interval. Although, sustained activity in IT during a delay period is not specifically modelled here, it is possible to reproduce a weak early target object modulation, in addition to the significant target effect beginning ~150ms post-stimulus. The figure on the right of fig. 5 shows this effect and is produced by using a slightly more complex form of prefrontal feedback to IT (combining a small constant excitatory feedback, representing sustained delay period activity, with a climbing inhibitory feedback). Further discussion about the replication of both spatial and object-based attentional effects, and accurate onset timing, from monkey single cell data in IT [21][32] and V4 [14][23] has been reported in more detail elsewhere and is beyond the scope of this paper. However, replication of data from older and more highly trained monkeys [23] compared to [21] suggested that IT feedback to V4 might be tuned by learning.

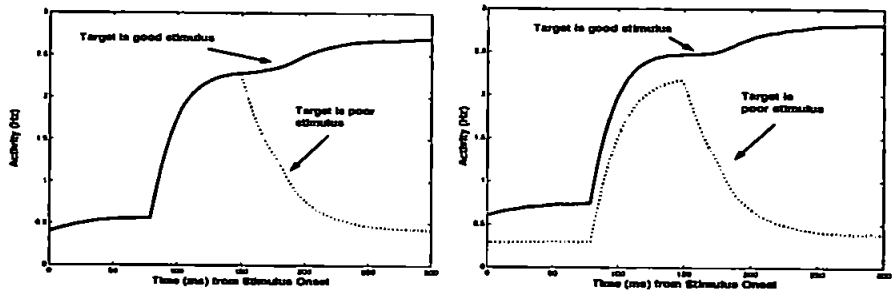



**Fig. 5.** Target object effect in IT: The plot on left shows the replication of [21] figure 3a from array onset until the time of the saccade. Two stimuli (a good stimulus for the cell and a poor one) are simultaneously presented within the cell's receptive field. The initial sensory response is not affected by target object selection but from ~150ms post-stimulus the response is determined by which of the two objects is the search target. If the target is the poor stimulus, responses are significantly suppressed. The plot on the right shows the same situation but includes a slight target effect early in the sensory response, as found in some cells. This replicates [32] figure 7a from array onset until the time of the saccade. A more complex prefrontal feedback to IT, including a sustained excitatory component, is used here. As in the left-hand plot, the significant target object effect begins ~150ms post-stimulus

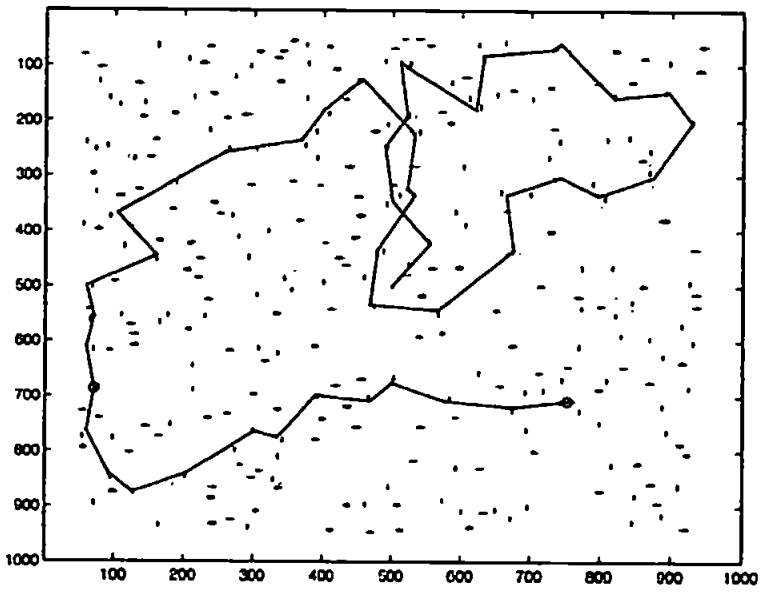



**Fig. 6.** Scan path through a dense scene: The target is a red stimulus. Fixations are shown as i) magenta dots - within 1 ° of a target coloured stimulus; 96% of fixations; ii) blue circles - within 1 ° of a non-target colour stimulus; 4% of fixations. Average saccade amplitude = 7.4 °. (N.B. When figures containing red and green bars are viewed in greyscale print, the red appears as a darker grey than the green)

## 3.2 Scan Path Results

As found in a similar psychophysical task [35], scan paths are attracted to target coloured locations. This is shown in fig. 6 where the target is a red horizontal bar and fig. 7, where the target is a green horizontal bar. Saccades tend to be shorter in dense scenes compared to sparse scenes. This is because the AW is scaled according to stimulus density and it

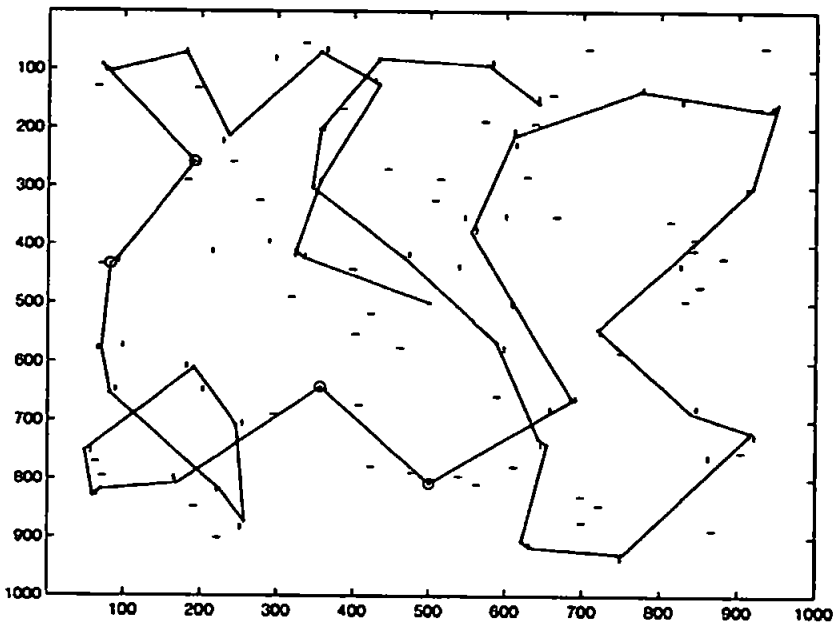



**Fig. 7.** Scan path through a sparse scene: The target is a green stimulus. Fixations are shown as i) magenta dots- within 1 ° of a target coloured stimulus; 92% of fixations; ii) blue circles - within 1 ° of a non-target colour stimulus; 8% of fixations. Average saccade amplitude = 12.1 °

contributes a positive bias to the competition in LIP, meaning that next fixation points tend to be chosen from within the AW. The AW scaling is based on local stimulus density. Therefore, the AW expands and contracts as it moves around a scene of mixed density, resulting in smaller saccades in the dense areas and larger saccades in sparse areas. Updates to the novelty map are also based on the size of the AW and this means that a smaller region is inhibited for return in areas of dense stimuli. Thus, the scan path is able to investigate the dense, and potentially more interesting, areas of a scene with a series of shorter amplitude saccades, as shown in fig. 8. This is potentially useful for natural scene processing, where uniform areas such as sky could be examined quickly with large amplitude saccades, allowing the scan path to concentrate in more detailed and interesting aspects of the scene.

Increasing the weight of the novelty bias to LIP slightly reduces the likelihood of fixating target coloured stimuli and, in sparse scenes, increases the number of fixations landing in blank areas of the display. Fig. 9 shows this effect. Thus, a search where novelty is the key factor, for example a hasty search of the entire scene due to time constraints, may result in more "wasted" fixations in blank areas.

Weaker object-related feedback within the ventral pathway (prefrontal to IT; IT to V4) reduces the object-based effect within V4 and this results in more non-target coloured stimuli being able to capture attention. Saccade onset is also delayed. The effect of the weight of IT feedback to V4 on fixation position is shown in fig. 10. Replication of single cell data suggested that IT feedback to V4 could be tuned by learning and experience. Therefore, search during a familiar task or with familiar objects may be faster and the scan path would be expected to fixate more target coloured locations than when the task or objects were unfamiliar.

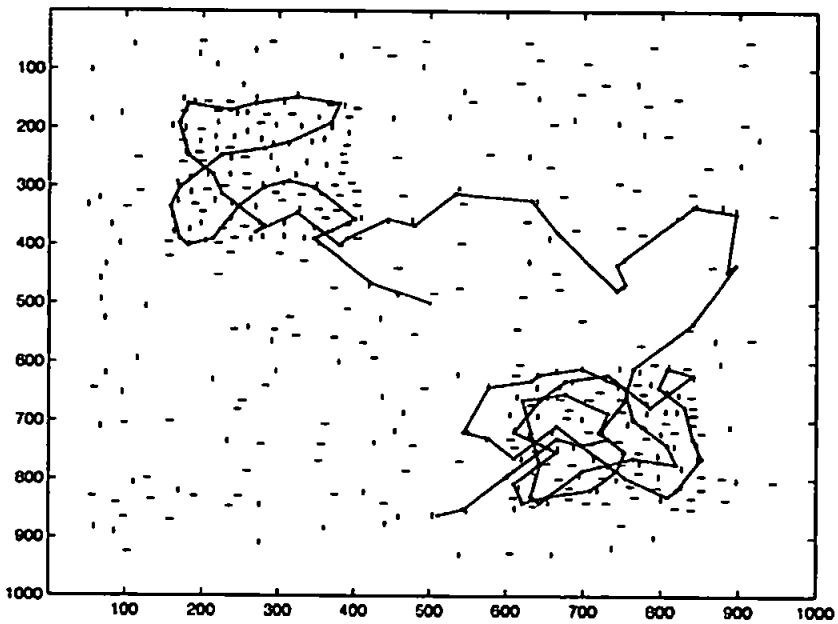



**Fig. 8.** Scan path through a mixed density scene: the scan path examines the dense areas more thoroughly with a series of shorter saccades (average amplitude 4.5 ° in dense areas, compared to 7.5 ° in slightly more sparse areas)

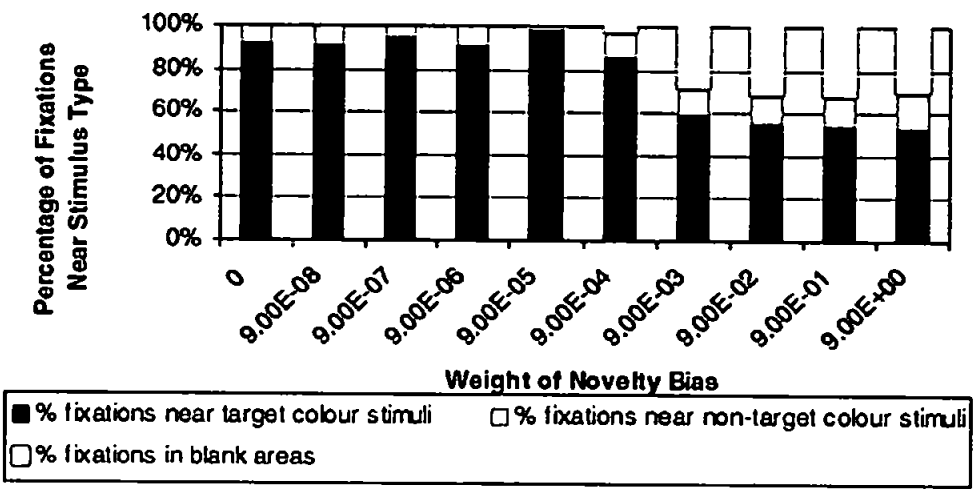



**Fig. 9.** Effect of the novelty bias in a sparse scene: Shows average fixation positions over 10 scan paths, each consisting of 50 fixations, over the image shown in fig. 7. Fixations are considered to be near a stimulus when they are within 1 ° of a stimulus centre. As the weight of the novelty bias increases, the number of fixations near target coloured stimuli decreases and fixations are more likely to occur in blank areas of the display

## 4 Discussion

This model is able to reproduce data from monkeys at both the neuronal [14][21] [23][25][30][31][32] and behavioural [35][48] levels. At the neuronal level, this is the first model to show attentional effects within V4 evolving from being purely spatial to being object-based, over time courses found in single cell recordings from monkeys

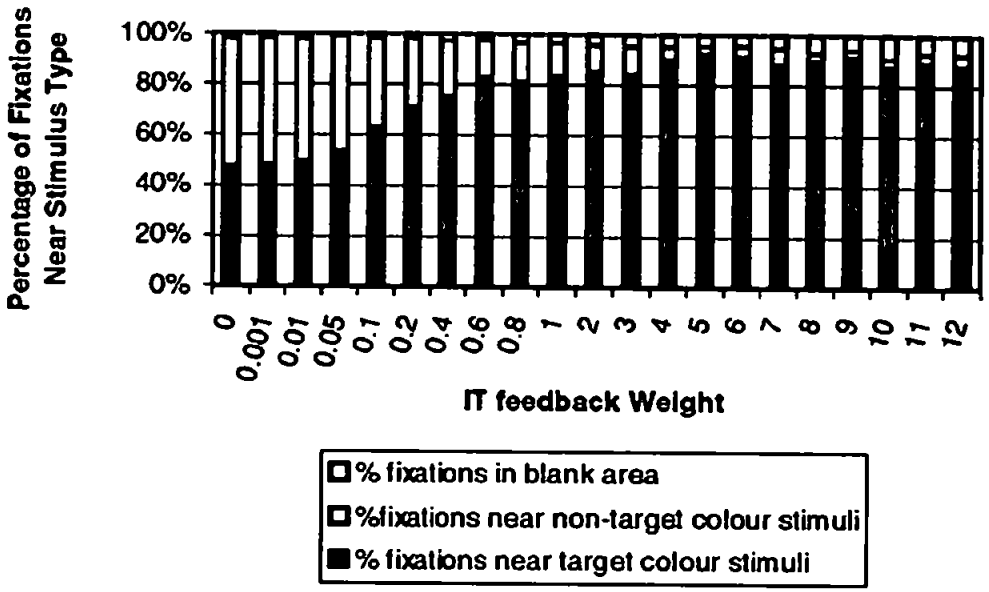



**Fig. 10.** Effect of weight of IT feedback to V4: Shows average fixation positions over 10 scan paths, each consisting of 50 fixations, over the image shown in fig. 6. Fixations are considered to be near a stimulus when they are within 1 ° of a stimulus centre. As the weight of feedback is increased there is a tendency for more target coloured stimuli and less non-target coloured stimuli to be fixated because object-based effects in the ventral stream are stronger

[14][23]. Object-based effects in IT also accurately reproduce those recorded in monkey IT during visual search [21][32]. Single cell recordings from monkeys [21][23] suggest that saccade onset may be temporally linked to the development of significant object-based effects in ventral pathway cortical areas. Here, saccade onset is linked to the development of a significant effect in IT. This leads to saccade onset times that are able to replicate those found by [21][23].

Of potential future use in practical applications such as video surveillance and robot vision is the ability to use these object-based effects to guide an active vision scan path. As a result of the "top-down" cognitive biasing of competition within the model (by prefrontal feedback to the model's ventral pathway), saccades are guided to behaviourally relevant locations that contain a potential target stimulus. Thus, object-based attention (in the ventral pathway) is able to bias spatial competition (within the dorsal pathway), enabling the LIP module to be a spatio-featural integrator that can control the scan path. The model is extendable to include other factors in the competition in LIP. For example, further stimulus-related factors, such as luminance or motion, could contribute to increase the salience of a location. Thus, the model is extendable to include other endogenous and exogenous factors in the competition for the capture of attention; an issue currently more widely explored in psychophysics, e.g. [57], than modelling, but of much practical relevance in future computer vision.

Selectivity in the guidance of eye movements during visual search, giving a preference for locations containing target features, has been found in a number of studies (e.g. [33][34][35][58]; but see [59]). Under several conditions with simple stimuli, colour (or luminance) provides stronger guidance than orientation in monkeys [35] and humans [33][34] (but see [58] for equal preference for shape and colour). At the behavioural

level, the model is able to prioritise the features that guide the scan path, such that this tendency for colour to dominate search can be reproduced. The priority of colour in guiding the scan path is achieved by a marginal difference in the relative strength of connection from V4 feature types to LIP. The relative strength of these connections may be malleable on the basis of cognitive requirement or stimulus-related factors, such as distractor ratios [60]. However, under conditions of equal proportions of distractor types, colour was found to guide the scan path even when the task was heavily biased towards orientation discrimination [35]. This suggests that such connections may be fixed during development and may not be easily adapted to task requirements. The model highlights the possible role for feedforward connections from the ventral stream to parietal cortex in determining search selectivity, and the need for further investigation of such connections in lesion studies. The model makes further predictions beyond the scope of this paper relating to the effect of damage to the LIP/V4 connections on binding of information across feature dimensions.

The requirement for a novelty map in world/head-based co-ordinates as a memory trace of locations visited suggests that visual search cannot be totally amnesic [61] and supports the idea of memory across saccades [56][62][63]. The involvement of frontal and parietal areas in this form of IOR may explain why orbitofrontal and parietal patients display increased revisiting of locations in their scan paths [50][51].

The ability of this model to reproduce active visual search behaviour found in humans and monkeys whilst accurately replicating neurophysiological data at the single cell and brain region levels, suggests that biological models can lend much to active vision research for practical computer vision applications.

## References


1. Duncan, J., Humphreys, G.W.: Visual search and stimulus similarity. Psych. Rev. Vol. 96(3) (1989) 433–458
2. Desimone, R., Duncan, J.: Neural mechanisms of selective visual attention. Ann. Rev. Neuroscience. Vol. 18 (1995) 193–222
3. Duncan, J., Humphreys, GW., Ward, R.: Competitive brain activity in visual attention. Current Opinion in Neurobiology. Vol. 7 (1997) 255–261
4. Desimone, R.: Visual attention mediated by biased competition in extrastriate visual cortex. Phil. Trans. Royal Soc. London B. Vol. 353 (1989) 1245–1255
5. Reynolds, J.H., Chelazzi, L., Desimone, R.: Competitive mechanisms subserve attention in macaque areas V2 and V4. J. Neuroscience. Vol. 19(5) (1999) 1736–1753
6. Usher, M., Niebur, E.: Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. J. Cognitive Neuroscience. Vol. 8(4) (1996) 311–327
7. Deco, G., Lee, T.S.: A unified model of spatial and object attention based on inter-cortical biased competition. Neurocomputing. Vol. 44-46 (2002) 775–781
8. De Kamps, M., Van der Velde, F.: Using a recurrent network to bind form, color and position into a unified percept. Neurocomputing. Vol. 38-40 (2001) 523–528
9. Hamker, F.H.: A dynamic model of how feature cues guide spatial attention. Vision Research. Vol. 44 (2004) 501–521
10. Helmholtz, H: Handbuch der physiologishen Optik, Voss, Leipzig (1867)
11. Treisman, A.: Perceptual grouping and attention in visual search for features and for objects. J. Exp. Psych.: Human Perc. & Perf. Vol. 8 (1982) 194–214

12. Crick, F.: Function of the thalamic reticular complex: The searchlight hypothesis. Proc. Nat. Ac. Sci., USA. Vol. 81 (1984) 4586-4590
13. Connor, C.E., Callant, J.L., Preddie, D.C., Van Essen, D.C.: Responses in area V4 depend on the spatial relationship between stimulus and attention. J. Neurophysiology. Vol. 75 (1996) 1306–1308
14. Luck, S.J., Chelazzi, L., Hillyard, S.A., Desimone, R.: Neural mechanisms of spatial attention in areas V1, V2 and V4 of macaque visual cortex. J. Neurophysiology. Vol. 77 (1997) 24–42
15. Bricolo, E., Gianesini, T., Fanini, A., Bundesen, C., Chelazzi, L.: Serial attention mechanisms in visual search: a direct behavioural demonstration. J. Cognitive Neuroscience. Vol. 14(7) (2002) 980–993
16. Duncan, J.: Selective attention and the organisation of visual information. J. Exp. Psych.: General. Vol. 113 (1984) 501–517
17. Blaser, E., Pylyshyn, Z.W., Holcombe, A.O.: Tracking an object through feature space. Nature. Vol. 408(6809) (2000) 196–199
18. O'Craven, K.M., Downing, P.E., Kanwisher, N.: fMRI evidence for objects as the units of attentional selection. Nature. Vol. 401 (1999) 584–587
19. Valdes-Sosa, M., Bobes, M.A., Rodriguez, V., Pinilla, T.: Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. J. Cognitive Neuroscience. Vol. 10(1) (1998) 137–151
20. Valdes-Sosa, M., Cobo, A., Pinilla, T.: Attention to object files defined by transparent motion. J. Exp. Psych.: Human Perc. & Perf. Vol. 26(2) (2000) 488–505
21. Chelazzi, L., Miller, E.K., Duncan, J., Desimone, R.: A neural basis for visual search in inferior temporal cortex. Nature. Vol. 363 (1993) 345–347
22. Roelfsema, P.R., Lamme, V.A.F., Spekreijse, H.: Object-based attention in the primary visual cortex of the macaque monkey. Nature. Vol. 395 (1998) 376–381
23. Chelazzi, L., Miller, E.K., Duncan, J., Desimone, R.: Responses of neurons in macaque area V4 during memory-guided visual search. Cerebral Cortex. Vol. 11 (2001) 761–772
24. Hillyard, S.A., Anllo-Vento, L.: Event-related brain potentials in the study of visual selective attention, Proc. Nat. Acad, Sci. USA. Vol. 95 (1998) 781–787
25. Colby, C.L., Duhamel, J.R., Goldberg, M.E.: Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. J. Neurophysiology. Vol. 76(5) (1996) 2841–2852
26. Kastner, S., Pinsk, M., De Weerd, P., Desimone, R., Ungerleider, L.: Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron. Vol. 22 (1999) 751–761
27. Corbetta, M., Kincade, J.M., Ollinger, J.M., McAvoy, M.P., Shulman, G.L.: Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nature Neuroscience Vol. 3 (2000) 292–297
28. Hopfinger, J.B., Buonocore, M.H., Mangun, G.R.: The neural mechanisms of top-down attentional control. Nature Neuroscience. Vol. 3(3) (2000) 284–291
29. Zeki, S.: A Vision of the Brain, Blackwell Scientific Publications, Oxford, UK. (1993)
30. Motter, B.C.: Neural correlates of attentive selection for color or luminance in extrastriate area V4. J. Neuroscience. Vol. 14(4) (1994) 2178–2189
31. Motter, B.C.: Neural correlates of feature selective memory and pop-out in extrastriate area V4. J. Neuroscience. Vol 14(4) (1994) 2190–2199
32. Chelazzi, L., Duncan, J., Miller, E.K.: Responses of neurons in inferior temporal cortex during memory-guided visual search. J. Neurophysiology. Vol 80 (1998) 2918–2940
33. Scialfa, C.T., Joffe, K.M.: Response times and eye movements in feature and conjunction search as a function of target eccentricity. Perception & Psychophysics. Vol. 60(6) (1998) 1067–1082

34. Williams, D.E., Reingold, E.M.: Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude. Psychonomic Bulletin & Review. Vol. 8(3) (2001) 476–488
35. Motter, B.C., Belky, E.J.: The guidance of eye movements during active visual search. Vision Research. Vol. 38(12) (1998) 1805–1815
36. Ungerleider, L.G., Mishkin, M.: Two cortical visual systems. In: Ingle, D.J., Goodale, M.A., Mansfield, R.W.J. (eds.): Analysis of Visual Behaviour. MIT Press, MA (1982) 549–586
37. Milner, A.D., Goodale, M.A.: The Visual Brain In Action. Oxford University Press (1995)
38. Gottlieb, J.P., Kusunoki, M., Goldberg, M.E.: The representation of visual salience in monkey parietal cortex. Nature. Vol. 391 (1998) 481–484
39. Lanyon, L.J., Denham, S.L.: A Model of Active Visual Search with Object-Based Attention Guiding Scan Paths. Neural Networks, Special Issue: Vision and Brain. Vol. 17 (2004) 873–897
40. Gerstner, W.: Population dynamics of spiking neurons: Fast transients, asynchronous states, and locking. Neural Computing. Vol. 12 (2000) 43–89
41. Brefczynski, J., DeYoe, E.A.: A phvsiological correlate of the 'spotlight' of visual attention. Nature Neuroscience. Vol. 2 (1999) 370-374
42. Wallis, G., Rolls, E.T.: Invariant face and object recognition in the visual system. Prog. Neurobiology. Vol. 51 (1997) 167–194
43. Ghose, G.M., Ts'O, D.Y.: Form processing modules in primate area V4. J. Neurophysiology. Vol. 77(4) (1997) 2191–2196
44. Moran, J., Desimone, R.: Selective attention gates visual processing in the extrastriate cortex. Science. Vol. 229 (1985) 782–784
45. Bushnell, M.C., Goldberg, M.E., Robinson, D.L.: Behavioural enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. J. Neurophysiology. Vol. 46(4) (1981) 755–772
46. Their, P., Andersen, R.A.: Electrical microstimulation distinguishes distinct saccade-related areas in the posterior parietal cortex. J. Neurophysiology. Vol. 80 (1998) 1713–1735
47. Lanyon, L. J., Denham, S. L.: A biased competition computational model of spatial and object-based attention mediating active visual search. Neurocomputing. Vol 58–60C (2004) 655–662
48. Motter, B.C., Belky, E.J.: The zone of focal attention during active visual search", Vision Research. Vol. 38(7) (1998) 1007–1022
49. Moore, T., Armstrong, K.M.: Selective gating of visual signals by microstimulation of frontal cortex. Nature. Vol. 421 (2003) 370–373
50. Husain, M., Mannan, S., Hodgson, T., Wojciulik, E., Driver, J., Kennard, C. Impaired spatial working memory across saccades contributes to abnormal search in parietal neglect. Brain. Vol. 124 (2001) 941–952
51. Hodgson, T.L., Mort, D., Chamberlain, M.M., Hutton, S.B., O'Neill, K.S., Kennard, C.: Orbitofrontal cortex mediates inhibition of return. Neuropsychologia. Vol. 431 (2002) 1–11
52. Tipper, S.P., Weaver, B., Watson, F.L.: Inhibition of return to successively cued spatial locations: Commentary on Pratt and Abrams (1995). J. Exp. Psych.: Human Perc. & Perf. Vol. 22(5) (1996) 1289–1293
53. Danziger, S., Kingstone, A., Snyder, J.J.: Inhibition of return to successively stimulated locations in a sequential visual search paradigm. J. Exp. Psych.: Human Perc. & Perf. Vol. 24(5) (1998) 1467–1475
54. Snyder, J.J., Kingstone, A.: Inhibition of return at multiple locations in visual search: When you see it and when you don't. Q. J. Exp. Psych. A: Human Exp. Psych. Vol. 54(4) (2001) 1221–1237
55. Snyder, J.J., Kingstone, A.: Inhibition of return and visual search: How many separate loci are inhibited? Perception & Psychophysics. Vol. 62(3) (2000) 452–458

56. Irwin, D.E., Zelinsky, G.J.: Eye movements and scene perception: Memory for things observed. Perception & Psychophysics. Vol. 64(6) (2002) 882–895
57. Kim, M-S., Cave, K.R.: Top-down and bottom-up attentional control: On the nature of interference from a salient distractor. Perception & Psychophysics. Vol. 61(6) (1999) 1009–1023
58. Findlay, J.M.: Saccade target selection during visual search. Vision Research, Vol. 37(5) (1997) 617–631
59. Zelinski, G.J.: Using eye saccades to assess the selectivity of search movements. Vision Research. Vol. 36(14) (1996) 2177–2187
60. Shen, J., Reingold, E.M., Pomplum, M.: Distractor ratio influences patterns of eye movements during visual search. Perception. Vol. 29(2) (2000) 241–250
61. Horowitz, T.S., Wolfe, J.M.: Visual search has no memory. Nature. Vol. 394 (1998) 575–577
62. Mitchell, J., Zipser, D.: A model of visual-spatial memory across saccades. Vision Research. Vol. 41 (2001) 1575–1592
63. Dodd, M.D., Castel, A.D., Pratt, J.: Inhibition of return with rapid serial shifts of attention: Implications for memory and visual search. Perception & Psychophysics. Vol. 65(7) (2003) 1126–1135

# A MODEL OF SPATIAL AND OBJECT-BASED ATTENTION FOR ACTIVE VISUAL SEARCH

LINDA LANYON SUSAN DENHAM

*Centre for Theoretical & Computational Neuroscience, University of Plymouth, Drakes Circus, Plymouth, PL4 8AA, U.K.*
*linda.lanyon@plymouth.ac.uk*

The *biased competition hypothesis* is of particular interest in the visual attention literature currently. We describe the first computational model to apply biased competition in active vision. At the cellular level, the model simulates both spatial and object-based attentional effects over time courses seen in single cell recordings in ventral stream areas. Such effects at the cellular level lead to systems level behaviour that replicates that observed during active visual search for orientation and colour conjunction targets, where the scan path is guided to target coloured locations in preference to locations containing the target orientation or blank areas.

## 1. Introduction

This chapter discusses a computational model of visual attention that has been previously described formally, with results presented at the systems (Lanyon & Denham, 2004b) and cellular (Lanyon & Denham, 2005) levels. Here, we review the contribution of the model at both levels.

Traditional views of visual attention included the simple spatial "spotlight", which is a serial spatial process that enhances responses at a particular location(s). A spatial enhancement is corroborated by cellular recordings in monkeys (Connor, Callant, Preddie & Van Essen, 1996, Luck, Chelazzi, Hillyard & Desimone, 1997). However, many single cell recordings (e.g. Chelazzi, Miller, Duncan & Desimone, 1998; 2001) as well as studies in psychophysics (e.g. Blaser, Pylyshyn & Holcombe, 2000), functional magnetic resonance imaging (e.g. O'Craven, Downing & Kanwisher 1999) and event-related potential recordings (e.g. Valdes-Sosa, Bobes, Rodriguez & Pinilla, 1998) have provided convincing evidence that attention is also able to operate in a more complex object-based manner. Object-based modulation of the responses of single cells leads to the response of the cell being determined by which object within its receptive field is attended. This effect is known as the 'target effect' and results in the cell's receptive field appearing to contract around the object that is the target for visual search when more than one object is present in the cell's receptive field (Chelazzi et al., 1998, 2001). This effect is explained by the

*biased competition hypothesis* (Desimone & Duncan, 1995), which is currently very influential in the study of visual attention. Biased competition suggests that responses of neurons encoding features of visual objects are determined by competitive interactions. It is suggested that there is competition at each level of the ventral "what" stream (Ungerleider & Mishkin, 1982) hierarchy. This competition is subject to a number of biases, such as "bottom-up" stimulus information and "top-down" cognitive requirements, for example a working memory template of the target object during visual search. Early small-scale models proved the feasibility of the hypothesis (e.g. Usher & Niebur, 1996) and recent larger-scale models (e.g. Deco, 2001; Hamker, 2004) have been successful in replicating a range of experimental data. Lanyon and Denham (2004b) presented, for the first time, a systems level biased competition model for active visual search, where retinal inputs change as the focus of attention is shifted, as happens during normal everyday search.

Event-related potential (Hillyard & Anllo-Vento, 1998) and single cell recordings suggest that spatial attention may be able to modulate responses earlier in the stimulus-related response than object-based attention. In ventral pathway area V4, which encodes form and colour features, spatial attention affects the earliest stimulus-invoked response at 60ms post-stimulus and modulates baseline responses before this (Luck et al., 1997). However, object-based effects in V4 appear to take longer to be resolved and have been recorded from approximately 150ms post-stimulus (Chelazzi et al., 2001; Motter, 1994). In inferior temporal cortex (IT), baseline responses during a delay period following a cue and the initial sensory response invoked by the search array are slightly modulated by the target object, but the most significant target effect begins from ~150-200ms post-onset (Chelazzi et al., 1998). Lanyon and Denham (2005) replicated the development of these attentional effects in IT and V4, over the same time course as those observed by Chelazzi et al. (1998, 2001) and Motter (1994). Such effects at the neuronal level allow this model to produce systems level behaviour where scan paths are attracted to behaviourally relevant locations, in particular replicating search behaviour observed in humans (Williams & Reingold, 2001) and monkeys (Motter & Belky, 1998b) where attention was preferentially attracted to locations containing target colour.

## 2. The Model

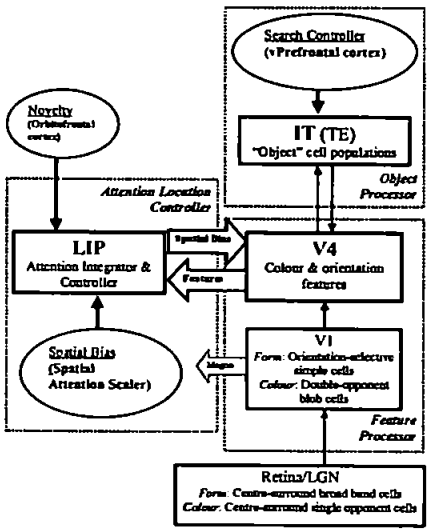



Figure 1. Overview of the model

Figure 1 shows that the model consists of a retina and V1, which detect form and colour features in a retinotopic manner, and three dynamic modules representing areas V4, IT and the lateral intraparietal area (LIP). The model adopts a similar framework to that described by Deco (2001). However, object and spatial biases operate concurrently here to influence the system dynamics and produce attentional effects at the cellular level over time courses observed in single cell recordings. Also in contrast to Deco's model, the retina here moves around the scene in an active vision manner, although a covert mode is also available. The dynamic modules are modelled using a mean field approach where the level of representation is a population of cells with similar properties, known as a cell assembly. Currently, V1 is not included in the dynamic portion of the system and we do not replicate attentional effects therein but, instead, focus on intermediate stages of visual processing, such as area V4. V1 acts as a feature detector, with a neuron for each feature at every pixel location, and features are fed forward to V4 in a convergent manner over the V4 receptive field area. V4 is retinotopic with biologically plausible (Wallis & Rolls, 1997) larger receptive fields than found in V1. Stimuli are coloured oriented bars, chosen to match the feature conjunctions used by Motter & Belky (1998b), in order to reproduce the effects seen in that psychophysical experiment. V4 neurons are functionally segregated (Ghose & Ts'O, 1997) and, like other similar models (e.g. Hamker, 2004), the V4 module encodes colour and

orientation in separate feature layers. In contrast to Deco (2001), competition is local, mediated by inhibitory interneurons, and operates between different features of the same type at the same location. Features within the same receptive field compete, as has been found in single cell recordings (e.g. Chelazzi et al., 1998, 2001; Luck et al., 1997; Moran & Desimone, 1985). V4 featural information feeds forward to activate non-retinotopic invariant object populations in IT. Thus, the model's ventral pathway (V1, V4 and IT) operates as a feature/object processing hierarchy with receptive field sizes increasing towards IT, where receptive fields span the entire retina. Anterior areas of IT, such as area TE, are not retinotopic but have large receptive fields and encode objects in an invariant manner (Wallis & Rolls, 1997).

Dorsal pathway module LIP contains a cell assembly for every retinotopic location in V4 and competition is spatial so that LIP is used to control the spatial focus of attention. Each LIP assembly is reciprocally connected to the assemblies in every feature layer in V4 at that location. During fixation, a spatial bias, linked to the eye movement and possibly originating in the frontal eye field, creates an early spatial attention window (AW). The AW is scaled according to local stimulus density (Lanyon & Denham, 2004a) on the basis of low-resolution orientation information assumed to be conveyed rapidly through the magnocellular pathway (Cheng, Eysel & Vidyasagar, 2004). This reflects the scaling of the area around fixation within which targets can be detected (Motter & Belky, 1998a). This spatial bias to LIP creates an anticipatory spatial attention effect, as has been found in single cell recordings in this area (Colby, Duhamel & Goldberg, 1996). LIP provides a spatial bias to V4 and this creates an early spatial attention effect therein, modulating baseline firing before the onset of the stimulus evoked response and, subsequently, modulating this response from its onset at ~60ms post-stimulus, as found by Luck et al. (1997).

Within V4, the early spatial focus of attention becomes object-based over time. V4 is subject to two concurrent biases: A spatial bias from LIP and an object-related bias from IT. Object-based attention evolves in the model's ventral pathway due to its object-related feedback biases. The development of object-based attention in V4 is dependent on the resolution of competition between objects in IT. In addition to the feed forward inputs from V4, competition between objects in IT is biased by a feedback current from prefrontal cortex, which is assumed to hold a working memory template of the target object. IT feeds back an inhibitory bias to V4 that suppresses features not related to the object. Over time, the target object is able to win the competition in IT, and then target features become enhanced whilst non-target features are suppressed across V4, as has been found in single cell recordings in monkeys

(Motter, 1994) over the same time course. Once a significant object-based effect is established in IT (determined simply by the activity of the strongest IT assembly being at least twice as active as the second strongest assembly), a saccade is initiated 70ms later, to reflect motor preparation, and in accordance with the relationship between object effects and saccade onset recorded IT and V4 (Chelazzi et al., 1998, 2001).

Following this object-based "highlighting" of target features in parallel across V4, the featural inputs from V4 to LIP provide a bias that results in locations containing target-related features becoming most active in LIP. Thus, LIP is able to represent the locations of salient, behaviourally relevant stimuli (Colby et al., 1996; Gottlieb, Kusunoki & Goldberg, 1998). The connection from V4 to LIP is set such that colour features provide a marginally stronger bias than orientation features. Thus, LIP tends to represent locations containing target coloured stimuli more strongly than locations containing the target orientation but non-target colour. Therefore, colour is given a priority in guiding the scan path, as found by Motter & Belky (1998b).

Inhibition of return (IOR) in the scan path cannot be implemented by inhibition of previously active cortical locations in retinotopic coordinates because the retina moves around the scene. A map containing the relative novelty of locations in the scene provides a bias to LIP and, therefore LIP mediates IOR. Parietal damage is linked to an inability to retain a spatial working memory of searched locations across saccades so that locations are repeatedly re-fixated (Husain, Mannan, Hodgson, Wojciulik, Driver, J. & Kennard, 2001). The source of such a bias could be a frontal area, such as orbitofrontal cortex, which is linked with reward association and, when damaged, affects IOR (Hodgson, Mort, Chamberlain, Hutton, O'Neill & Kennard, 2002).

## 3. Results

### 3.1. *Cellular Level Results*

At the cellular level, attentional effects within our model's V4 evolve from being purely spatial to being object-based, over similar times to those seen in monkey cell recordings (Chelazzi et al., 2001; Luck et al., 1997; Motter, 1994). The enhancement of target features across V4, which is dependant upon the development of the target object effect in IT, facilitates the guidance of the scan path. Object-based attention in our model's IT similarly develops over time (Chelazzi et al., 1998) and Figure 2 shows the development of these effects in V4 and IT. Replication of data from older and more highly trained monkeys

(Chelazzi et al., 2001) compared to (Chelazzi et al., 1998) suggested that IT feedback to V4 might be tuned by learning (Lanyon & Denham, 2005).

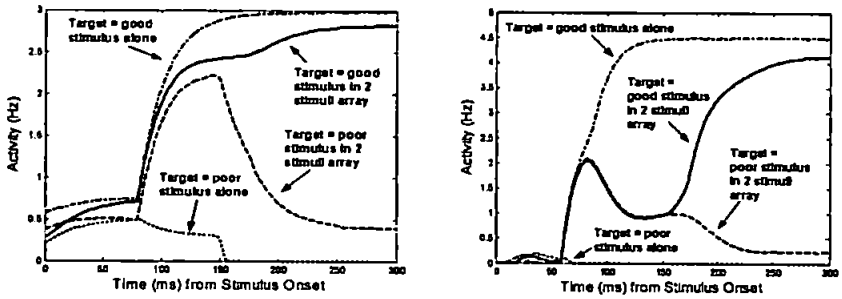



Figure 2. Left: Replicates Chelazzi et al. (1998, *figure 12a*) IT recording. Right: Replicates Chelazzi et al. (2001, *figure 5*) V4 recording. Shows the responses of assemblies in IT and V4 from array onset until the time of the saccade under 4 conditions in separate simulations. When a good stimulus for the cell (i.e. one that causes a high response) is presented alone in the receptive field a high response is produced: Top line. When a poor stimulus is presented alone, the cell's response is suppressed: Bottom line. The most interesting situation occurs when both stimuli are simultaneously presented within the receptive field: Middle two lines. The initial high response normally invoked by the good stimulus is reduced due to the competing presence of the poor stimulus. However, once object-based effects become significant from ~150ms, the response is determined by which of the two stimuli is the target, with activity tending towards that of the single stimulus case for the target object. When the target is the poor stimulus the response is suppressed despite the presence of the good stimulus in the receptive field. This effect enables V4 to represent target stimuli most strongly and bias LIP to guide the scan path to these locations. Here, saccades have been suppressed but, for example, for the two stimuli case in V4, would have occurred at ~235-240ms post-stimulus, which is similar to the time found by Chelazzi et al. (2001).

### 3.2. *Systems Level Results*

Due to the weighted connections from V4 to LIP, scan paths are attracted to target coloured locations, as was found by Motter & Belky (1998b). Saccades tend to be shorter in dense scenes compared to sparse scenes because the AW is scaled according to local stimulus density. This attentional focus zooms in and out as it moves around a scene of mixed density, resulting in smaller saccades in the dense areas and larger saccades in sparse areas. Updates to the novelty map are also based on the size of the AW and this means that a smaller region is inhibited for return in areas of dense stimuli. Thus, the scan path is able to investigate the dense, and potentially more interesting, areas of a scene with a series of shorter amplitude saccades, as shown in Figure 3. Increasing the weight (i.e. importance) of the novelty bias to LIP slightly reduces the likelihood of fixating target coloured stimuli and, in sparse scenes, increases the number of fixations landing in blank areas of the display (Lanyon & Denham, 2004b).

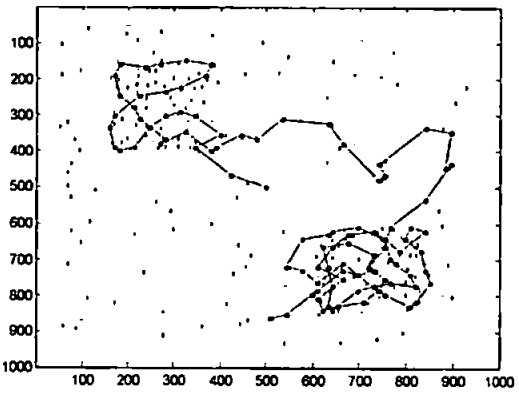



Figure 3. Scan path through a mixed density scene where the target colour is red: Average amplitude of saccades in the dense areas is 4.5°, compared to 7.5° in the slightly more sparse areas. (N.B. In grayscale, the red bars appear as a darker gray than the green bars.)

Weaker object-related feedback within the ventral pathway (prefrontal to IT; IT to V4) reduces the object-based effect in IT and V4, resulting in saccade onset being delayed and more non-target coloured stimuli being able to capture attention. Replication of single cell data suggested that IT feedback to V4 could be tuned by learning and experience. Therefore, search within familiar objects or tasks may be faster and the scan path would be expected to fixate more target coloured locations than when the task or objects were unfamiliar.

## 4. Conclusion

This model shows that attentional effects modeled at the cellular level, in order to replicate data from monkeys, are able to give rise to systems level search behaviour similar to that seen under certain conditions in humans and monkeys. At the neuronal level, the model shows attentional effects within V4 evolving from being purely spatial to being object-based, over time courses found in single cell recordings from monkeys (Chelazzi et al., 2001; Luck et al., 1997). Object-based effects in IT also accurately reproduce those recorded in monkey IT during visual search (Chelazzi et al., 1998). Saccade onset times, linked to the development of a significant effect in IT, are very similar to those found during the money cell recordings in IT and V4 (Chelazzi et al., 1998, 2001). Replication of the monkey single cell data suggested a possible influence of learning upon the development of object-based attention and, hence, saccade onset and the behavioural relevance of locations visited by the scan path (Lanyon and Denham, 2005).

Object-based effects at the cellular level were used to guide the search scan path to behaviourally relevant locations. Prefrontal feedback provides a top-down cognitive bias towards the target object. This leads to the target object in IT winning the competition and, consequently, target features being enhanced across V4. The V4 bias to posterior parietal area LIP causes it to represent locations containing target features most strongly and allow attention to be attracted to these behaviourally salient locations. Thus, similar to Wolfe's (1994) *activation map*, LIP acts as a retinotopic "saliency map" (Gottlieb et al., 1998) combining information from the ventral stream, which has resolved top-down and bottom-up influences. Parietal cortex, the frontal eye field (FEF), the pulvinar nucleus of the thalamus and superior colliculus (SC) have all been implicated in providing saliency map functions and may cooperate in this. Shipp (2004) reviewed the neural architecture controlling the spatial focus of attention in relation to current models of attention. His Real Neural Architecture model suggests a distributed saliency system in which the pulvinar acts as the central hub saliency map providing the attentional spotlight and communicating with SC, ventral stream areas and, indirectly, with LIP and FEF. In the context of the model presented here, the LIP module can be considered part of the distributed saliency system and the pulvinar could be a possible source of the spatial bias to LIP and V4.

Representations in the model LIP are retinotopic and based on the current fixation point. The current model does not fully address issues of saccadic re-mapping of the scene and associated saliency. Cortical transformations and compression of space around the time of a saccade are very complex issues and are discussed in relation to this model in Lanyon and Denham (2004b). A scene-based novelty bias, possibly from orbitofrontal cortex, to LIP provided inhibition of return in the scan path. This suggests that parietal areas may be involved in transforming a memory trace of locations visited in the scene into retinotopic coordinates. The connections from V4 to LIP suggest one possible mechanism by which different features could gain priority in attracting attention. We allowed colour to have priority in order to replicate its strong influence on attentional capture found by Motter and Belky (1998) using the same stimuli. Our stimuli were equally salient in terms of bottom-up contribution. However, a feature type having strong bottom-up salience (represented by high activity in V1 and V4) could be capable of overriding the in-built bias from the V4 to LIP connection, since the difference in connection strength is marginal. In future, the model could be extended to allow LIP to represent bottom-up saliency, for example luminance or motion, more strongly than currently, possibly by allowing V1 to feedforward to LIP directly. The saliency map could be further

enhanced by additional modules, representing the pulvinar and SC, participating in the system dynamics. In summary, the biased competition approach presented in this model provides an extensible framework that, for example, allows further endogenous and exogenous factors to compete for the capture of attention in LIP.

## References


Blaser E., Pylyshyn Z.W., Holcombe, A.O. (2000) Tracking an object through feature space. *Nature*, 408(6809), 196-199

Cheng A., Eysel U.T., Vidyasagar T.R. (2004) The role of the magnocellular pathway in serial deployment of visual attention. *European Journal of Neuroscience*, 20, 2188-2192

Chelazzi L., Duncan J., Miller E.K. (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophysiology*, 80, 2918-2940

Chelazzi L., Miller E.K., Duncan J., Desimone R. (2001) Responses of neurons in macaque area V4 during memory-guided visual search. *Cerebral Cortex*, 11, 761-772

Colby C.L., Duhamel J.R., Goldberg M.E. (1996) Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *J. Neurophysiology*, 76(5), 2841-2852

Connor C.E., Callant J.L., Preddie D.C., Van Essen D.C. (1996) Responses in area V4 depend on the spatial relationship between stimulus and attention. *J. Neurophysiology*, 75, 1306-1308

Deco G. (2001) Biased competition mechanisms for visual attention in a multimodular neurodynamical system. In: Wermter, S, Austin, J. and Willshaw, D. (Eds), *Emergent Neural Computational Architectures Based on Neuroscience: Towards Neuroscience-Inspired Computing.* Heidelberg: Springer-Verlag. (pp. 114-126)

Desimone R., Duncan J. (1995) Neural mechanisms of selective visual attention. *Ann. Rev. Neuroscience*, 18, 193-222

Ghose G.M., Ts'O D.Y. (1997) Form processing modules in primate area V4. *J. Neurophysiology*, 77(4), 2191-2196

Gottlieb J.P., Kusunoki M., Goldberg M.E. (1998) The representation of visual salience in monkey parietal cortex. *Nature*, 391, 481-484

Hamker F.H. (2004) A dynamic model of how feature cues guide spatial attention. *Vision Research*, 44, 501-521

Hillyard S.A., Anllo-Vento L. (1998) Event-related brain potentials in the study of visual selective attention. *Proc. Nat. Acad, Sci. USA*, 95, 781-787

Hodgson T.L., Mort D., Chamberlain M.M., Hutton S.B., O'Neill K.S., Kennard C. (2002) Orbitofrontal cortex mediates inhibition of return. *Neuropsychologia*, 431, 1-11

Husain, M., Mannan, S., Hodgson, T., Wojciulik, E., Driver, J., Kennard, C. Impaired spatial working memory across saccades contributes to abnormal search in parietal neglect. Brain. Vol. 124 (2001) 941-952

Lanyon L.J., Denham S.L. (2004a) A biased competition computational model of spatial and object-based attention mediating active visual search. *Neurocomputing*, 58-60C, 655-662

Lanyon L.J., Denham, S.L. (2004b) A Model of Active Visual Search with Object-Based Attention Guiding Scan Paths. *Neural Networks*, 17, 873-897

Lanyon L.J., Denham S.L. (2005) A biased competition computational model of spatial and object-based attention. *Submitted*

Luck S.J., Chelazzi L., Hillyard S.A., Desimone R. (1997) Neural mechanisms of spatial attention in areas V1, V2 and V4 of macaque visual cortex. *J. Neurophysiology*, 77, 24-42

Moran J., Desimone R. (1985) Selective attention gates visual processing in the extrastriate cortex. *Science*, 229, 782-784

Motter B.C. (1994) Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J. Neuroscience*, 14(4), 2178-2189

Motter B.C., Belky E.J. (1998a) The zone of focal attention during active visual search, *Vision Research*, 38(7), 1007-1022

Motter, B.C., Belky, E.J. (1998b) The guidance of eye movements during active visual search. *Vision Research*. Vol. 38(12) (1998) 1805-1815

O'Craven K.M., Downing P.E., Kanwisher N. (1999) fMRI evidence for objects as the units of attentional selection. *Nature*, 401, 584-587

Shipp, S (2004) The brain circuitry of attention. *Trends Cog. Sci.*, 8(5), 223-230

Ungerleider L.G., Mishkin M. (1982) Two cortical visual systems. In Ingle D.J., Goodale M.A., Mansfield R.W.J. (Eds.), *Analysis of Visual Behaviour*. MA: MIT Press. (pp. 549-586)

Usher M., Niebur E. (1996) Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *J. Cognitive Neuroscience*, 8(4), 311-327

Valdes-Sosa M., Bobes M.A., Rodriguez V., Pinilla T. (1998) Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. *J. Cognitive Neuroscience*, 10(1), 137-151

Wallis G., Rolls E.T. (1997) Invariant face and object recognition in the visual system. *Prog. Neurobiology*, 51, 167-194

Williams D.E., Reingold E.M. (2001) Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude. *Psychonomic Bulletin & Review*, 8(3), 476-488

Wolfe, J.M. (1994) Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202-238

# A Model of Object-Based Attention That Guides Active Visual Search to Behaviourally Relevant Locations

Linda Lanyon & Susan Denham
*Centre for Theoretical & Computational Neuroscience, University of Plymouth, U.K.*

## Abstract

*During active visual search for a colour and orientation feature conjunction target, the scan path is guided to target coloured locations in preference to locations containing the target orientation or blank areas (Motter & Belky, 1998). An active vision model, using biased competition, is able to replicate this behaviour. Attention evolves from an early spatial effect to being object-based later in the response of the model neurons. Competition between features/objects is biased by knowledge of the target object, allowing non-target features to be suppressed over a time course seen in single cell recordings (Chelazzi et al., 1993, 2001). As this object-based attention develops, featural information is passed from extrastriate cortex (V4) to posterior parietal cortex (LIP), biasing spatial competition therein, enabling it to represent behaviourally relevant locations (Colby et al., 1996) and attract the scan path. Connections to LIP allow features to have different priorities in attracting attention. Thus, at the single cell level, this model replicates spatial and object-based attentional effects with temporal precision and, at the systems level, replicates psychophysical scan paths allowing attention to be drawn to behaviourally relevant locations. Such a "top-down" biased model is extendable and has potential future application in surveillance and robot guidance.*

## 1. Introduction

The *biased competition hypothesis* [1,2,3,4] is currently very influential in the study of visual attention at many levels ranging from neurophysiological single cell studies to brain imaging, evoked potentials and psychophysics. The hypothesis suggests that responses of neurons, such as those that encode object features, are determined by competitive interactions. This competition is subject to a number of biases, such as "bottom-up" stimulus information and "top-down" cognitive requirements, for example a working memory template of the target object during visual search. Building on early small-scale models using this hypothesis [e.g. 5,6], systems level models [e.g. 7,8,9] have been able to replicate a range of experimental data. However, there has been no systems level modelling of the biased competition hypothesis for active visual search, where retinal inputs change as the focus of attention is shifted. The model presented here uses biased competition and adopts an active vision approach such that, at any particular fixation, its cortical modules receive retinal inputs relating to a portion of the entire scene. This allows the model to process a smaller area of the image at any fixation and provides a realistic model of normal everyday search.

Traditionally, visual attention was thought to operate as a simple spatial "spotlight" [e.g. 10,11,12] that acted to enhance responses at a particular location(s) and it is clear that attention can produce spatially specific modulation of neuronal responses [e.g. 13,14] and behaviour [e.g. 15]. However, many studies in psychophysics [e.g. 16,17], functional magnetic resonance imaging [e.g. 18], event-related potential recordings [e.g. 19,20] and single cell recordings [e.g. 21,22,23,24] provide convincing evidence for attention operating in a more complex object-based manner. However, these findings are confounded by issues relating to whether the selection is object-based, feature-based or surface-based. Here we use the term object-based in the sense that it refers an attentional modulation of cellular responses on the basis of the features of a target object, as used by [7,21,23]. The term feature-based is also appropriate. Many different mechanisms may be responsible for the range of object-based effects seen experimentally and not all are addressed here. For example, there is no attempt to provide Gestalt-like perceptual grouping, but an extension to the model's feature processing modules to allow local lateral connectivity across similar features could produce such an effect.

Event-related potential [25] and single cell recordings suggest that spatial attention may be able to modulate responses earlier than object-based attention. For example, in monkey lateral intraparietal area (LIP), an early spatial enhancement of responses has been recorded from single cells in anticipation of the stimulus [24] and has been seen in imaging of the possible human homologue of LIP [26,27,28]. In ventral pathway area V4, which encodes form and colour features [e.g. 29], spatial modulation of baseline

responses has been recorded from single cells in advance of the sensory response and the earliest stimulus-invoked response at 60ms post-stimulus was also spatially modulated [14]. However, object-based effects in V4 appear to take longer to be resolved and have been recorded from approximately 150ms post-stimulus [23,30,31]. Such modulation, leading to the response of the cell being determined by which object within its receptive field is attended, is known as the 'target effect' [23]. In inferior temporal cortex (IT), delay period baseline responses are modulated but initial sensory responses reflect all stimuli within the receptive field without target preference. Later in the response, the target effect develops, similar to that in V4, being recorded after 150-200ms post-onset of the search array [21,32]. The model presented here replicates these spatial and object-based effects, at the neuronal level, with temporal precision. Such effects at the neuronal level allow the model to produce the search behaviour similar to that seen in humans [33,34] and monkeys [35] at the systems level.

A "cross-stream" interaction between the ventral "what" pathway, responsible for the encoding of features and objects and leading from primary visual cortex (V1) through V2 and V4 to temporal cortex, and the dorsal "where" pathway, leading from V1 to parietal cortex [36,37], allows the search scan path to be drawn to behaviourally relevant locations rather than less relevant stimulus locations or blank areas. Object-based attention in the model's ventral pathway biases the dorsal module (LIP) to represent the location of target features most strongly [24,38]. In addition, this connection allows certain features to have priority in attracting attention. In particular, the model is able to reproduce the scan path behaviour of monkeys searching for a colour-orientation feature conjunction target [35] where fixations tended to land within 1° of stimuli (only 20% fell in blank areas of the display) and these stimuli tended to be target coloured (75% of fixations landed near target coloured stimuli and only 5% near non-target coloured stimuli). The use of object-based attention within the model to guide the scan path is the focus of this paper.

## 2. The Model

Figure 1 shows a schematic overview of the model. It consists of a retina and V1, which detect form and colour features in a retinotopic manner, and three dynamic modules representing areas V4, IT and LIP, which are modelled using a mean field approach [39], in a similar manner to the model in [7], where the level of representation is a population of cells with similar

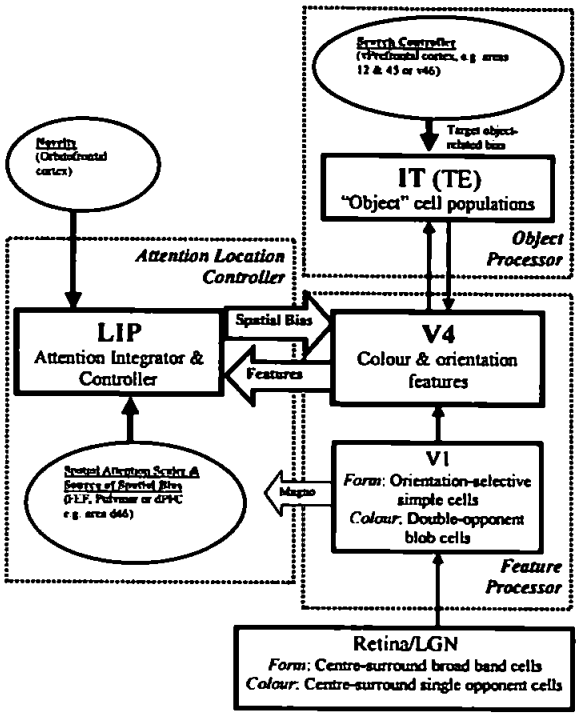



**Figure 1. Overview of the model**

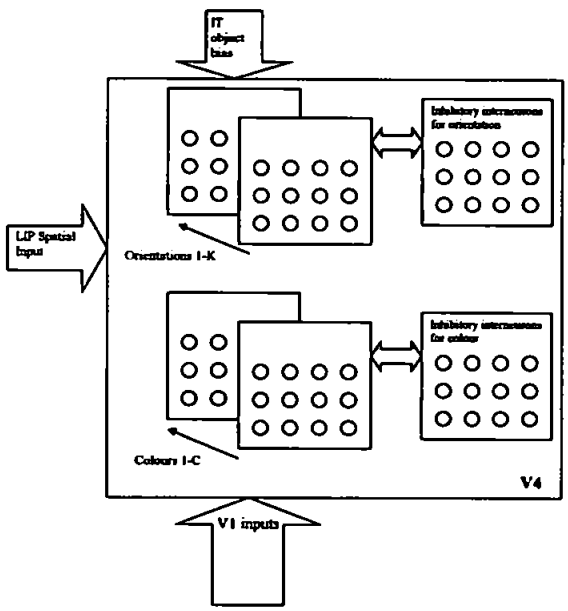



**Figure 2. Competition in V4**

properties, known as a cell assembly. The size of the retina can be varied for any particular simulation because the size of the cortical modules is scaled according to retina size. Thus, a larger retina may be used where biologically valid predictions are to be made, or a smaller retina where processing speed needs to be optimised (the only restriction being that a very small retina tends to lead to a higher proportion of fixations landing in blank areas of very sparse scenes due to lack of stimuli within the limited retina). For reasons of computational simplicity and speed, V1 is not included in the dynamic portion of the system and there is no attempt in this version of the model to reproduce attentional effects therein, although such effects have been observed [e.g. 22,40]. Intermediate stages of visual processing, such as area V4, are the focus of this model. V1 contains a neuron at every pixel location and features detected in V1 are fed forward to V4 in a convergent manner over the V4 receptive field area. Thus, V4 is retinotopic with biologically plausible [e.g. 41] larger receptive fields than found in V1 (the size of filters used in V1 determines the ratio of pixels to degrees of visual angle). Stimuli are coloured oriented bars, chosen to match the feature conjunctions used by [35], in order to reproduce the effects seen in that psychophysical experiment. V4 neurons are known to be functionally segregated [42] and the V4 module encodes colour and orientation in separate feature layers, a common modelling simplification. Different features of the same type at the same location compete via inhibitory interneuron assemblies (i.e. different colours within the same V4 receptive field area compete and different orientations within the same V4 receptive field area compete), as shown in figure 2. This provides the necessary competition between features within the same receptive field that has been found in single cell recordings [e.g. 43,14,21,23]. V4 featural information is fed forward to activate non-retinotopic object populations in IT. Thus, the model's ventral pathway (V1, V4 and IT) operates as a feature/object processing hierarchy with receptive field sizes increasing towards IT, where receptive fields span the entire retina. It is known that anterior areas of IT, such as area TE, are not retinotopic but have large receptive fields and encode objects in an invariant manner [41].

Dorsal pathway module LIP contains a cell assembly for every retinotopic location in V4. Each LIP assembly is reciprocally connected to the assembly in each feature layer in V4 at that location. LIP provides a spatio-featural map that is used to control the spatial focus of attention and, hence, fixation. Although responses in monkey LIP have been found to be independent of motor response [44], LIP is thought to be involved in selecting possible targets for saccades and has connections to superior colliculus and the frontal eye field (FEF), which are involved in saccade generation. Furthermore, direct electrical stimulation of LIP has elicited saccades [45]. Competition in the model LIP is spatial and the centre of the receptive field of the most active assembly at the end of the fixation is chosen as the next fixation point. During fixation a spatial bias creates an early spatial attention window (AW), which is scaled according to local stimulus density (on the basis of low resolution orientation information assumed to be conveyed rapidly to parietal cortex through the magnocellular pathway). This reflects the scaling of the area around fixation within which targets can be detected [46]. This spatial bias to LIP creates an anticipatory spatial attention effect, as has been found in single cell recordings in this area [24]. LIP provides a spatial bias to V4 and this creates an early spatial attention effect therein, modulating baseline firing before the onset of the stimulus evoked response and, subsequently, modulating this response from its onset at ~60ms post-stimulus, as found by [14]. The source of the spatial bias could be FEF because stimulation of this area leads to a spatially specific modulation of neuronal response in V4 [47]. The spatial bias from LIP to V4 also provides some binding across feature types in V4.

Within V4, the early spatial focus of attention becomes object-based over time. This is due to the fact that V4 is subject to two concurrent biases: A spatial bias from LIP and an object-related bias from IT. Thus, object-based attention evolves in the model's ventral pathway due to its object-related feedback biases. The development of object-based attention in V4 is dependent on the resolution of competition between objects in IT. In addition to the feed forward inputs from V4, competition between objects in IT is biased by a feedback current from prefrontal cortex, which is assumed to hold a working memory template of the target object. IT feeds back an inhibitory bias to V4 that suppresses features not related to the object. Over time, the target object is able to win the competition in IT, and then target features become enhanced whilst non-target features are suppressed across V4, as has been found in single cell recordings in monkeys [30,31] over the same time course. Once a significant object-based effect is established in IT, a saccade is initiated 70ms later, to reflect motor preparation, and replicate the timing of saccade onset following the target object effect seen in single cell recordings in IT [21] and V4 [23].

Following this object-based "highlighting" of target features in parallel across V4 [30,31], the featural inputs from V4 to LIP provide a bias that results in locations containing target-related features becoming most active in LIP. Thus, LIP is able to represent the

locations of salient, behaviourally relevant stimuli [24,38]. Prior to the onset of object-based attention in the ventral stream (which occurs ~150ms post-stimulus, as found by [21,23,32]), the initial sensory response in LIP tends to represent locations containing stimuli equally, but more strongly than blank areas. The connection from V4 to LIP is set such that colour features provide a slightly stronger bias than orientation features. Thus, LIP tends to represent locations containing target coloured stimuli more strongly than locations containing the target orientation but non-target colour. Therefore, colour is given a priority in guiding the scan path, as found by [35]. The difference in strength of connection of the V4 features to LIP need only be marginal in order to achieve this effect. Figure 3 shows the effect of adjusting the relative connection weights.

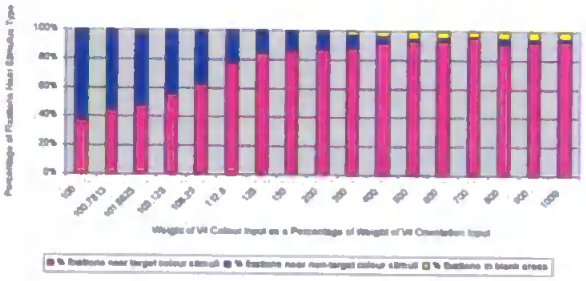


**Figure 3. Effect of increasing the relative weight of V4 colour feature input to LIP**

Figure 3 shows that, when V4 colour features are marginally more strongly connected to LIP than V4 orientation features, the scan path is attracted to target coloured stimuli in preference to stimuli of the target orientation.

The use of this active vision paradigm with a moving retina means that inhibition of return (IOR) in the scan path cannot be implemented by inhibition of previously active cortical locations in retinotopic coordinates. Parietal damage is linked to an inability to retain a spatial working memory of searched locations across saccades so that locations are repeatedly re-fixated [48]. Here, IOR in the scan path is provided by a novelty-related bias to LIP. The source of such a bias could be a frontal area, such as orbitofrontal cortex, which is linked with event/reward associations and, when damaged, affects IOR [49]. A world/head-based map reflects the novelty of all locations in the scene and, hence, their potential reward. Initially, the novelty of all locations in the scene is high. Then, when attention is withdrawn from a location, locations in the immediate vicinity of the fixation point have their novelty reduced to a low value that climbs in a Gaussian fashion with distance from fixation such that peripheral locations within the AW have neutral novelty. Novelty recovers over time such that IOR is present at multiple locations [50,51,52], the magnitude of the effect decreases approximately linearly from its largest value at the most recently searched location and several previous locations are affected [53,54].

## 3. Results

### 3.1. Neuronal Level Results

At the neuronal level, attentional effects within V4 evolve from being purely spatial to being object-based, over time courses found in single cell recordings from monkeys [14,21,23,30,31]. The replication of both spatial and object-based attentional effects, and accurate onset timing, from monkey single cell data in IT [21,32] and V4 [14,23] has been reported in more detail elsewhere and is beyond the scope of this paper. However, the enhancement of target features across V4 is important to the guidance of the scan path and figure 4 shows the development of object-based attention from ~150ms post-stimulus in V4, which mirrors that recorded by [23]. Replication of data from older and more highly trained monkeys [23] compared to [21] suggested that IT feedback to V4 is tuned by learning.

### 3.2. Scan Path Results

As found in a similar psychophysical task [35], scan paths are attracted to target coloured locations. This is shown in figure 5 where the target is a red horizontal bar and figure 6, where the target is a green horizontal bar. Saccades tend to be shorter in dense scenes compared to sparse scenes. This is because the AW is scaled according to stimulus density and it contributes a positive bias to the competition in LIP, meaning that next fixation points tend to be chosen from within the AW. The AW scaling is based on local stimulus density. Therefore, the AW expands and contracts as it moves around a scene of mixed density, resulting in smaller saccades in the dense areas and larger saccades in sparse areas. Thus, the scan path is able to investigate the dense, and potentially more interesting, areas of a scene with a series of shorter amplitude saccades, as shown in figure 7. This is potentially useful for natural scene processing, where uniform areas such as sky could be examined quickly with large amplitude saccades, allowing the scan path to concentrate in more detailed and interesting aspects of the scene.

Increasing the weight of the novelty bias to LIP slightly reduces the likelihood of fixating target coloured stimuli and, in sparse scenes, increases the

number of fixations landing in blank areas of the display. Figure 8 shows this effect. Thus, a search where novelty is the key factor, for example a hasty search of the entire scene due to time constraints, may result in more "wasted" fixations in blank areas.

Weaker object-related feedback within the ventral pathway (prefrontal to IT; IT to V4) reduces the object-based effect within V4 and this results in more non-target coloured stimuli being able to capture attention. Saccade onset is also delayed. The effect of the weight of IT feedback to V4 on fixation position is shown in figure 9. Replication of single cell data suggested that IT feedback to V4 could be tuned by learning and experience. Therefore, search during a familiar task or with familiar objects may be faster and the scan path would be expected to fixate more target coloured locations than when the task or objects were unfamiliar.

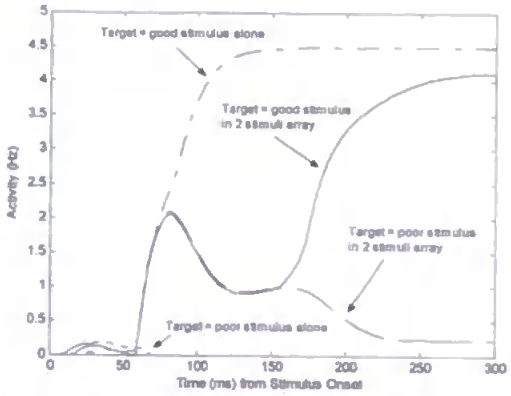



**Figure 4. Comparison of V4 single assembly response**

Shows the replication of [23] *figure 5* from array onset until the time of the saccade. Here, saccades have been suppressed but, for the two stimuli case, would have occurred at ~235-240ms post-stimulus, the same onset time as found by [23]. A good stimulus for the cell (i.e. one that causes a high response) presented alone in the receptive field produces a high response: Top line. When a poor stimulus is presented alone the cell's response is suppressed: Bottom line. When both stimuli are simultaneously presented within the receptive field, the high response normally invoked by the good stimulus is reduced due to the competing presence of the poor stimulus. However, once object-based effects begin from ~150ms, the response is determined by which of the two stimuli is the target, with activity tending towards that of the single stimulus case for the target object. The 2 stimuli case when the target is a poor stimulus for the cell shows the suppression of the non-target responses (i.e. the cell's response to the good stimulus) that enables V4 to represent target stimuli most strongly and bias LIP to guide the scan path to these locations.

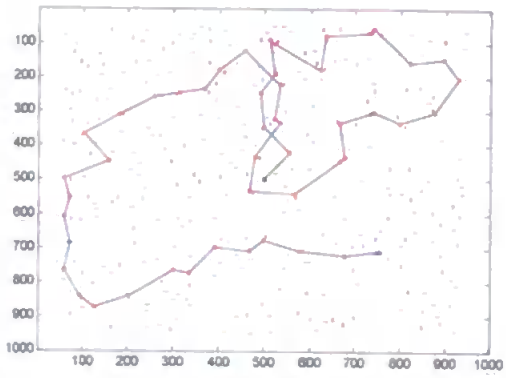


**Figure 5. Scan path through a dense scene**

The target in figure 5 is a red stimulus. Fixations are shown as i) magenta – within 1° of a target coloured stimulus; 96% of fixations; ii) blue – within 1° of a non-target colour stimulus; 4% of fixations. Average saccade amplitude = 7.4°.

## 4. Discussion

This model is able to reproduce data from monkeys at both the neuronal [e.g. 14,21,23,24,30,31,32] and behavioural [35,46] levels. At the neuronal level, this is the first model to show attentional effects within V4 evolving from being purely spatial to being object-based, over time courses found in single cell recordings from monkeys [14,23]. Object-based effects in IT also accurately reproduce those recorded in monkey IT during visual search [21]. Single cell recordings from monkeys [21,23] suggest that saccade onset may be temporally linked to the development of significant object-based effects in ventral pathway cortical areas. Here, saccade onset is linked to the development of a significant effect in IT. This leads to saccade onset times that are able to replicate those found by [21,23].

Of potential future use in practical applications such as video surveillance and robot vision is the ability to use these object-based effects to guide an active vision scan path. As a result of the "top-down" cognitive biasing of competition within the model (by prefrontal feedback to the model's ventral pathway), saccades are guided to behaviourally relevant locations that contain a potential target stimulus. Thus, object-based attention (in the ventral pathway) is able to bias spatial competition (within the dorsal pathway), enabling the LIP module to be a spatio-featural integrator that can control the scan path. The model is extendable to include other factors in the competition in LIP. For example, further stimulus-related factors, such as luminance or motion, could contribute to

increase the salience of a location. Thus, the model is extendable to include other endogenous and exogenous factors in the competition for the capture of attention; an issue currently more widely explored in psychophysics [e.g. 55] than modelling, but of much practical relevance in future computer vision.

Selectivity in the guidance of eye movements during visual search, giving a preference for locations containing target features, has been found in a number of studies [e.g. 33,34,35,36; but see 57]. Under several conditions with simple stimuli, colour (or luminance) provides stronger guidance than orientation in monkeys [35] and humans [33,34; but see 56 for equal

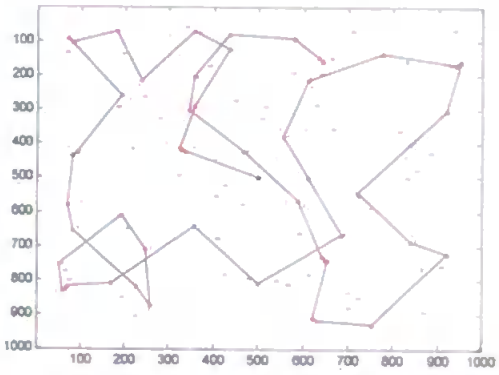



**Figure 6. Scan path through a sparse scene**

The target in figure 6 is a green stimulus. Fixations are shown as i) magenta – within 1° of a target coloured stimulus; 92% of fixations; ii) blue – within 1° of a non-target colour stimulus; 8% of fixations. Average saccade amplitude = 12.1°.

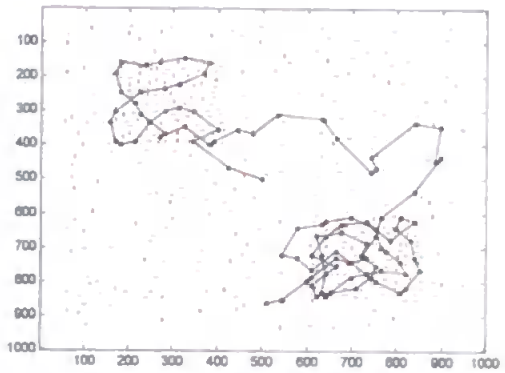



**Figure 7. Scan path through a mixed density scene**

In figure 7, the scan path examines the dense areas more thoroughly with a series of shorter saccades (average amplitude 4.5° in dense areas, compared to 7.5° in slightly more sparse areas).

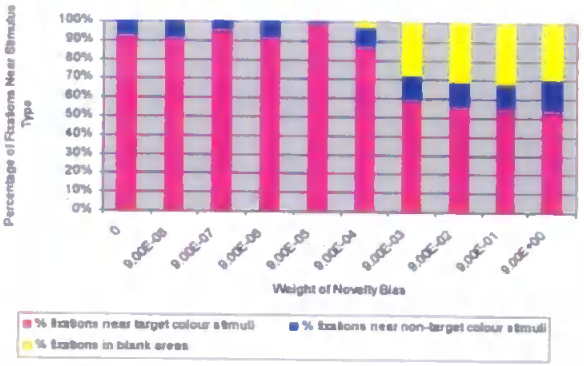



**Figure 8. Effect of novelty bias in a sparse scene**

Figure 8 shows average fixation positions over 10 scan paths, each consisting of 50 fixations, over the image shown in figure 6. Fixations are considered to be near a stimulus when they are within 1° of the stimulus. As the weight of the novelty bias increases, the number of fixations near target coloured stimuli decreases and fixations are more likely to occur in blank areas of the display.

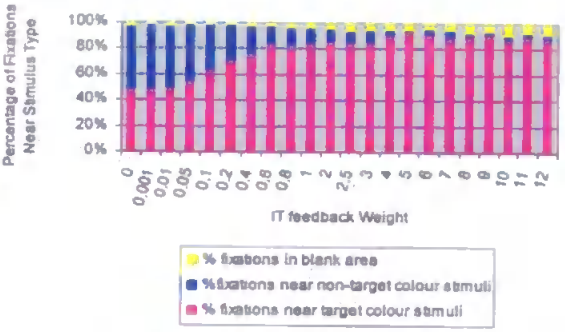



**Figure 9. Effect of weight of IT feedback to V4**

Figure 9 shows average fixation positions over 10 scan paths, each consisting of 50 fixations, over the image shown in figure 5. Fixations are considered to be near a stimulus when they are within 1° of the stimulus. As the weight of feedback is increased there is a tendency for more target coloured stimuli and less non-target coloured stimuli to be fixated because object-based effects in the ventral stream are stronger.

preference for shape and colour]. At the behavioural level, the model is able to prioritise the features that guide the scan path, such that this tendency for colour to dominate search can be reproduced. The priority of colour in guiding the scan path is achieved by a marginal difference in the relative strength of connection from V4 feature types to LIP. The relative

strength of these connections may be malleable on the basis of cognitive requirement or stimulus-related factors, such as distractor ratios [e.g. 58]. However, under conditions of equal proportions of distractor types, colour was found to guide the scan path even when the task was heavily biased towards orientation discrimination [35]. This suggests that such connections may be fixed during development and may not be easily adapted to task requirements. The model highlights the possible role for feedforward connections from the ventral stream to parietal cortex in determining search selectivity, and the need for further investigation of such connections in lesion studies. The model makes further predictions beyond the scope of this paper relating to the effect of damage to the LIP/V4 connections on binding of information across feature dimensions.

The requirement for a novelty map in world/head-based co-ordinates as a memory trace of locations visited suggests that visual search cannot be totally amnesic [59] and supports the idea of memory across saccades [54,60,61]. The involvement of frontal and parietal areas in this form of IOR may explain why orbitofrontal and parietal patients display increased revisiting of locations in their scan paths [48,49].

The ability of this model to reproduce active visual search behaviour found in humans and monkeys whilst accurately replicating neurophysiological data at the single cell and brain region levels, suggests that biological models can lend much to active vision research for practical computer vision applications.

## 9. References


[1] J. Duncan, and G.W. Humphreys, "Visual search and stimulus similarity", *Psych. Rev.*, 1989, vol. 96(3), pp. 433-458

[2] R. Desimone, and J. Duncan, "Neural mechanisms of selective visual attention", *Ann. Rev. Neurosci.*, 1995, vol. 18, pp. 193-222

[3] J. Duncan, G.W. Humphreys, and R. Ward, "Competitive brain activity in visual attention", *Curr. Op. Neurobio.*, 1997, vol. 7, pp. 255-261

[4] R. Desimone, "Visual attention mediated by biased competition in extrastriate visual cortex", *Phil. Trans. Royal Soc. London B*, 1989, vol. 353, pp. 1245-1255

[5] J. H. Reynolds, L. Chelazzi, and R. Desimone, "Competitive mechanisms subserve attention in macaque areas V2 and V4 , *J. Neurosci.*, 1999, vol. 19(5), pp. 1736-1753

[6] M. Usher, E. Niebur, "Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention", *J. Cog. Neurosci.*, 1996, vol. 8(4) pp. 311-327

[7] G. Deco, T.S. Lee, "A unified model of spatial and object attention based on inter-cortical biased competition", *Neurocomp.*, 2002, vol. 44-46, pp. 775-781

[8] M. de Kamps, and F. Van der Velde, "Using a recurrent network to bind form, color and position into a unified percept". *Neurocomp.*, 2001, vol. 38-40, pp. 523-528

[9] F.H. Hamker, "A dynamic model of how feature cues guide spatial attention", *Vis. Res.*, 2004, Vol. 44, pp. 501-521

[10] H. Helmholtz, *Handbuch der physiologishen Optik*, Voss, Leipzig, 1867

[11] A. Treisman, "Perceptual grouping and attention in visual search for features and for objects", *J. Exp. Psych.: Human Perc. & Perf.*, 1982, vol. 8, pp. 194-214

[12] F. Crick, "Function of the thalamic reticular complex: The searchlight hypothesis", *Proc. Nat. Ac. Sci., USA*, 1984, vol. 81, pp. 4586-4590

[13] C.E. Connor, J.L. Callant, D.C. Preddie, and D.C. Van Essen, "Responses in area V4 depend on the spatial relationship between stimulus and attention". *J. Neurophysio.*, 1996, vol. 75, pp. 1306-1308

[14] S.J. Luck, L. Chelazzi, S.A. Hillyard, and R. Desimone, "Neural mechanisms of spatial attention in areas V1, V2 and V4 of macaque visual cortex", *J. Neurophysiol.*, 1997, vol. 77, pp. 24-42

[15] E. Bricolo, T. Gianesini, A. Fanini, C. Bundesen, and L. Chelazzi, "Serial attention mechanisms in visual search: a direct behavioural demonstration", *J. Cog. Neurosci.*, 2002, vol. 14 (7), pp. 980-993

[16] J. Duncan, "Selective attention and the organisation of visual information", *J. Exp. Psych.: General*, 1984, vol. 113, pp. 501-517

[17] E. Blaser, Z.W. Pylyshyn, and A.O. Holcombe, "Tracking an object through feature space". *Nature*, 2000, vol. 408(6809), pp. 196-199

[18] K.M. O'Craven, P.E. Downing, and N. Kanwisher, "fMRI evidence for objects as the units of attentional selection", *Nature*, 1999, vol. 401, pp. 584-587

[19] M. Valdes-Sosa, M.A. Bobes, V. Rodriguez, and T. Pinilla, "Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials", *J. Cog. Neurosci.*, 1998, vol. 10(1), pp. 137-151

[20] M. Valdes-Sosa, A. Cobo, and T. Pinilla, "Attention to object files defined by transparent motion". *J. Exp. Psych.: Human Perc. & Perf.*, 2000, vol. 26(2), pp. 488-505

[21] L. Chelazzi, E.K. Miller, J. Duncan, R. Desimone, "A neural basis for visual search in inferior temporal cortex", *Nature*, 1993, vol. 363, pp. 345-347

[22] P.R. Roelfsema, V.A.F. Lamme, and H. Spekreijse, "Object-based attention in the primary visual cortex of the macaque monkey", *Nature*, 1998, vol. 395, pp. 376-381

[23] L. Chelazzi, E.K. Miller, J. Duncan, R. Desimone, "Responses of neurons in macaque area V4 during memory-guided visual search", *Cereb. Cortex*, 2001, vol. 11, pp. 761-772

[24] C.L. Colby, J.R. Duhamel, M.E. Goldberg, "Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area", *J. Neurophysiol.*, 1996, vol. 76(5), pp. 2841-2852

[25] S.A. Hillyard and L. Anllo-Vento, "Event-related brain potentials in the study of visual selective attention, *Proc. Nat. Acad. Sci. USA*, 1998, vol 95, pp.781-787

[26] S. Kastner, M. Pinsk, P. De Weerd, R. Desimone, L. Ungerleider, "Increased activity in human visual cortex during directed attention in the absence of visual stimulation", *Neuron*, 1999, vol. 22 , pp. 751-761

[27] M. Corbetta, J.M. Kincade, J.M. Ollinger, M.P. McAvoy, and G.L. Shulman, "Voluntary orienting is dissociated from target detection in human posterior parietal cortex", *Nat. Neurosci.*, 2000, vol. 3, pp. 292-297

[28] J.B. Hopfinger, M.H. Buonocore, and G.R. Mangun, "The neural mechanisms of top-down attentional control", *Nat. Neurosci.*, 2000, vol. 3(3), pp. 284-291

[29] S. Zeki, *A Vision of the Brain*, Blackwell Scientific Publications, Oxford, UK, 1993

[30] B.C. Motter, "Neural correlates of attentive selection for color or luminance in extrastriate area V4 , *J. Neurosci.* 1994, vol. 14(4), pp. 2178-2189

[31] B.C. Motter, "Neural correlates of feature selective memory and pop-out in extrastriate area V4 , *J. Neurosci.* 1994, vol. 14(4), pp. 2190-2199

[32] L. Chelazzi, J. Duncan, E.K. Miller, "Responses of neurons in inferior temporal cortex during memory-guided visual search", *J. Neurophysiol*, 1998, vol. 80, pp. 2918-2940

[33] C.T. Scialfa, and K.M. Joffe, "Response times and eye movements in feature and conjunction search as a function of target eccentricity", *Perc. & Psychophys.*, 1998, vol. 60(6), pp. 1067-1082

[34] D.E. Williams, and E.M. Reingold, "Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude", *Psych. Bull. & Rev.*, 2001, vol. 8(3), pp. 476-488

[35] B.C. Motter, E.J. Belky, "The guidance of eye movements during active visual search", *Vis. Res.*, 1998, vol. 38(12), pp. 1805-1815

[36] L.G. Ungerleider, and M. Mishkin, "Two cortical visual systems". In: D. J. Ingle, M. A. Goodale, & R. W. J. Mansfield (Eds.), *Analysis of Visual Behaviour*, pp. 549-586, MIT Press, MA, 1982

[37] A.D. Milner, and M.A. Goodale, *The Visual Brain In Action*, Oxford University Press, 1995

[38] J.P Gottlieb, M. Kusunoki, M.E. Goldberg, "The representation of visual salience in monkey parietal cortex", *Nature*, 1998, vol. 391, pp. 481-484

[39] W. Gerstner, "Population dynamics of spiking neurons: Fast transients, asynchronous states, and locking", *Neural Comput.*, 2000, vol. 12, pp. 43-89

[40] J. Brefczynski, and E.A. DeYoe, "A phvsiological correlate of the 'spotlight' of visual attention". *Nat. Neurosci.*, 1999, vol. 2, pp. 370-374

[41] G. Wallis, and E.T. Rolls, "Invariant face and object recognition in the visual system", *Prog. Neurobio.*, 1997, vol. 51, pp. 167-194

[42] G.M Ghose and D.Y Ts'O, "Form processing modules in primate area V4 , *J. Neurophysiol*, 1997, vol. 77(4), pp. 2191-2196

[43] J. Moran, R. Desimone R, "Selective attention gates visual processing in the extrastriate cortex", *Science*, 1985, vol. 229, pp. 782-784

[44] M.C. Bushnell, M.E. Goldberg, and D.L. Robinson, "Behavioural enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention", *J. Neurophysiol.*, 1981, vol. 46(4), pp. 755-772

[45] P. Their, and R.A. Andersen, "Electrical microstimulation distinguishes distinct saccade-related areas in the posterior parietal cortex", *J Neurophysiol*, 1998, vol. 80, pp. 1713-1735

[46] B.C. Motter, E.J. Belky, "The zone of focal attention during active visual search", *Vis. Res.*, 1998, vol. 38(7), pp. 1007-1022

[47] T. Moore, K.M. Armstrong, "Selective gating of visual signals by microstimulation of frontal cortex", *Nature*, 2003, vol. 421, pp. 370-373

[48] M. Husain, S. Mannan, T. Hodgson, E. Wojciulik, J. Driver, and C. Kennard, "Impaired spatial working memory across saccades contributes to abnormal search in parietal neglect", *Brain*, 2001, vol. 124, pp. 941-952

[49] T.L. Hodgson, D. Mort, M.M. Chamberlain, S.B. Hutton, K.S. O'Neill, C. Kennard, "Orbitofrontal cortex mediates inhibition of return", *Neuropsychologia*, 2002, vol. 431, pp. 1-11

[50] S.P. Tipper, B. Weaver, and F.L. Watson, "Inhibition of return to successively cued spatial locations: Commentary on Pratt and Abrams (1995)", *J. Exp. Psych.: Human Perc. & Perf.*, 1996, vol. 22(5), pp. 1289-1293

[51] S. Danziger, A. Kingstone, and J.J. Snyder, "Inhibition of return to successively stimulated locations in a sequential visual search paradigm", *J. Exp. Psych.: Human Perc. & Perf.*, 1998, vol. 24(5), pp. 1467-1475

[52] J.J. Snyder, and A. Kingstone, "Inhibition of return at multiple locations in visual search: When you see it and when you don't", *Q. J. Exp. Psych. A: Human Exp. Psych.*, 2001, vol. 54(4), pp. 1221-1237

[53] J.J. Snyder, and A. Kingstone, "Inhibition of return and visual search: How many separate loci are inhibited?", *Perc. & Psychophys.*, 2000, vol. 62(3), pp. 452-458

[54] D.E. Irwin, and G.J. Zelinsky, "Eye movements and scene perception: Memory for things observed", *Perc. & Psychophys.*, 2002, vol. 64(6), pp. 882-895

[55] M-S. Kim, and K.R. Cave, "Top-down and bottom-up attentional control: On the nature of interference from a salient distractor", *Perc. & Psychophys.*, 1999, vol. 61(6), pp. 1009-1023

[56] J.M. Findlay, "Saccade target selection during visual search", *Vis. Res.*, 1997, vol. 37(5), pp. 617-631

[57] G.J. Zelinski, "Using eye saccades to assess the selectivity of search movements", *Vis. Res.*, 1996, vol. 36(14), pp. 2177-2187

[58] J. Shen, E.M. Reingold, and M. Pomplum, "Distractor ratio influences patterns of eye movements during visual search", *Perception*, 2000, vol. 29(2), pp. 241-250

[59] T.S. Horowitz, and J.M. Wolfe, "Visual search has no memory", *Nature*, 1998, vol. 394, pp. 575-577

[60] J. Mitchell, and D. Zipser, "A model of visual-spatial memory across saccades", *Vis. Res.*, 2001, vol. 41, pp. 1575-1592

[61] M.D. Dodd, A.D. Castel, and J. Pratt, "Inhibition of return with rapid serial shifts of attention: Implications for memory and visual search", *Perc. & Psychophys.*, 2003, vol. 65(7), pp. 1126-1135