Stimulus Optimisation and the Properties of Visual Neurons in Temporal Cortex

Robin Edwards

Thesis submitted for the degree of PhD

School of Psychology

University of St. Andrews

July 2002.

Declarations

I, Robin Edwards, hereby certify that this thesis, which is approximately 40,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

date 22/7/02 signature of candidate

I was admitted as a research student in October 1998 and as a candidate for the degree of Doctor of Philosophy in October 1998; the higher study for which this is a record was carried out in the University of St. Andrews between 1998 and 2002.

date 22/7/02 signature of candidate .

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date 22/7/02 signature of supervisor

Thesis Copyright

Unrestricted

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

date 22/7/02 signature of candidate .

Abstract

A fundamental goal of visual neuroscience is to understand the sequence of representations and transformations by which the brain extracts meaningful information from the retinal image. However, feature selectivities of neurons in higher visual areas of temporal cortex (e.g. STS and IT) are complex and poorly understood. This thesis describes a novel computational method for determining effective stimuli for visual neurons. It attempts to optimise the response of a cell by changing stimuli in the estimated direction of the response gradient with respect to stimulus parameters. Results are presented showing that this method produces multiple effective solutions for a simulated complex cell. It may therefore prove more applicable than reverse correlation since it can characterise non-linear neurons.

Due to the high dimensionality of image space, optimisation is more likely to be successful if the search can be restricted to a specific stimulus subspace. To this end, the colour and spatial frequency response of neurons in areas STS and IT was investigated. Theories of coarse to fine processing (e.g. Delorme et al. 2000) suggest that, at the high presentation rates necessary for stimulus optimisation, the response of these neurons may be largely determined by low spatial frequency achromatic aspects of a stimulus.

Recordings were made from 50 neurons in STS/IT to colour, achromatic, false colour and high- and low-pass filtered versions of each cell's preferred stimulus. Differential activity for achromatic and coloured stimuli was evident from response onset; thus the hypothesis that information about object colour is delayed with respect to object form is not supported. Furthermore, no evidence was found to support the idea that high spatial frequencies were delayed with respect to low spatial frequencies. Overall, low-pass achromatic stimuli were not found to be effective stimuli, even at the fastest presentation rate (14ms/image). Thus this thesis finds no evidence for progressive availability of spatial frequency information from coarse to fine.

Acknowledgements

It has been a great pleasure and a privilege to work with my supervisors, Peter Földiák and David Perrett, and I owe them many thanks. Their passion and enthusiasm for science gave me the all the encouragement I needed to complete this thesis.

Thanks are also due to Dengke Xiao, whose good humour and persistence during the experiments ensured that we recorded sufficient data, long after I was ready to give up. I also greatly appreciate the assistance of Mary Latimer, who helped with the histology, and of Christian Keysers, who introduced me to the joys of Mathematica, and wrote some of the software used to perform the data analysis.

Finally, special thanks to Esther, and to my parents, for all their encouragement and support during my years of study at St. Andrews.

I would also like to acknowledge the financial support of the BBSRC, without which this work would not have been possible.

Table of Contents

1. General Introduction

1.1. The Problem: Characterisation of Visual Receptive Fields

A pattern of light falls on the retina and is transformed by the brain into a rich visual world – our perception is of objects and individuals, colours and textures. How does the brain represent the visual world? And what sequence of computations leads to this representation?

Theoretical models of object recognition generally suppose a sequence of computational steps, whereby progressively more complex features are extracted from the retinal image with each stage of processing (e.g. Marr, 1982).

In the primate brain, object recognition is thought to be accomplished by the ventral visual pathway, which carries visual information from V1 (the first area of cortex to receive retinal input) to area TE of the inferior temporal lobe, via areas V2, V4 and TEO. Bilateral removal of inferior temporal cortex results in a severe deficit in learning tasks that require object recognition and discrimination (see reviews in Mishkin, 1982; Dean, 1976; Gross, 1973).

Within the ventral visual pathway, neurons with increasingly complex feature selectivities are found, as visual information progresses along the pathway and away from the retina (Kobatake & Tanaka, 1994). For instance, neurons in the retina and LGN respond to spots of light, while the "simple cells" of V1 respond to bars of a specific orientation. With each stage of processing it is assumed that neurons use the output of the preceding area to extract increasingly complex features from the retinal input. Thus a bar detector in V1 could be formed by combining the output of a set of LGN cells with collinear receptive fields.

This increasing complexity makes an objective description of the feature selectivity of neurons in higher visual areas (i.e. towards the end of the pathway) very difficult. In TE, neurons have been described that respond selectively to the sight of a face, a hand, or complex geometrical objects (Perrett et al., 1982; Gross, 1972; Tanaka et al., 1991). However, the current methods for establishing the stimulus selectivity of these cells rely on the assumptions and luck of the experimenters, and make use of a limited and predetermined set of stimuli. Thus, we can say that a particular stimulus is the most effective of those that we have tested, but this does not allow us to say with any certainty what the actual trigger features of the cell are. Furthermore, an estimated minimum of eight synapses lie between V1 and TE (Oram & Perrett, 1992), which places few constraints on the type of features that cells may be interested in.

At present, we lack a complete and objective description of the feature selectivities of neurons at all but the earliest stages of the visual system. Such a description would not only allow us to better understand how the visual world is represented, but would also provide valuable insights into the general principles of organisation and computation in the brain.

1.2. The Present Thesis: Towards a Systematic Method

This thesis describes work on a systematic method to find the optimal stimulus that could be used to investigate the properties of receptive fields in higher visual areas. In essence, the idea is to use the response of a neuron under study as a feedback signal to guide the computer-based generation of new stimuli. Although this idea is not new (see e.g. ALOPEX, Harth & Tzanakou, 1974), the method we describe is novel, and is designed to be more widely applicable than previous attempts.

A systematic search requires that responses to a large number of candidate stimuli be measured. To this end, we use the rapid serial visual presentation (RSVP) paradigm (Keysers et al., 2001) whereby responses are measured to a rapid and continuous stream of visual images, each presented for as little as 13 milliseconds.

However, a number of studies have suggested (e.g. Sugase et al., 1999; Delorme et al., 2000; Parker et al., 1992) that a serial and feedforward model of the visual system – whereby each visual area processes its input, then passes the result on to the next stage of processing – may be an over simplification. Stimulus representations may not emerge fully formed, rather information might become progressively available in a coarse to fine manner over several tens of milliseconds.

If this is the case, any search technique employing rapid stimulus presentation should take this into account, perhaps by rapidly optimising a coarse stimulus, then slowing down the presentation rate to allow finer levels of detail to be investigated. Therefore, the latter part of this thesis presents two neurophysiological experiments investigating the response of cells in area TE to stimuli with reduced levels of detail, at varying rates of presentation.

1.3. The Organisation of This Thesis

Firstly, the properties of visual neurons that are relevant to a discussion of stimulus optimisation will be reviewed in Chapter 2. Our consideration will be largely centred on the properties of neurons in areas STS and IT of the temporal cortex, since these areas contain visual neurons with complex and poorly understood receptive fields – the main area of study.

Chapter 3 will review the previous attempts to study receptive field properties of

neurons in a systematic way. Although this chapter concentrates on the visual modality, we will also look at relevant work that has been carried out on the other senses, specifically the auditory system. Chapter 4 presents the novel algorithm we have developed for visual stimulus optimisation, how it can be adapted to make use of different image basis sets, and its performance with simulated neurons.

In recognition of the high dimensionality of the problem faced, Chapters 5, 6 and 7 will explore some of the features of neurons in STS and IT that might be used to constrain the search for optimal stimuli. Chapter 5 provides an introduction to this idea and reviews the relevant literature. Chapter 6 describes the results with colour, then Chapter 7 goes on to look at spatial frequency. Finally, Chapter 8 concludes this thesis and gives an overview of the main findings. Appendices at the end of this thesis contain the list of stimuli, histological reconstructions of the positions of cells studied in Chapters 6 & 7, and the computer code, with a commentary, written to support the optimisation algorithm described in Chapter 4.

Some of the findings of this thesis, specifically those relating to the colour and frequency experiments (Chapters 6 & 7) have already been presented at the Society for Neuroscience meeting in New Orleans (2000). Chapter 6 has also been published as a paper in the Journal of Neurophysiology[1]. I also intend to prepare chapter 7 for publication. The findings on contrast sensitivity (Chapter 2) were presented at the Society for Neuroscience meeting in 2001.

---

[1] Edwards, R., Xiao, D-K., Keysers, C., Földiák, P. & Perrett, D. (2003). Color Sensitivity of Cells Responsive to Complex Stimuli in the Temporal Cortex. *Journal of Neurophysiology* 90: 1245-1256.

2. Properties of Cells in Higher Visual Areas

This chapter will discuss in some detail the properties of neurons in higher visual areas of temporal cortex, specifically STS and IT. The complex properties of these cells underlines the need for a systematic means of study, and illustrates some of the constraints that must be built into such a process.

2.1. Anatomy and Connectivity of the Inferior Temporal lobe

The inferior temporal cortex (or IT) of the monkey is located on the inferior convexity of the temporal lobe, extending from the superior temporal sulcus (STS) dorsally to the rhinal sulcus, ventrally and medially. On the basis of cell morphology, the anterior part of IT was considered by von Bonin & Bailey (1947) to be a single cytoarchitectonic division, which they labelled TE. A more posterior area, in the vicinity of the inferior occipital sulcus, was labelled TEO. This nomenclature is still in current usage, and is broadly equivalent to a more recent division between anterior IT (AIT) and posterior IT (PIT) that has been proposed by Tanaka et al. (1991) and based on receptive field properties. TE (or AIT) consists of the anterior two-thirds of IT cortex, with TEO (or PIT) occupying the posterior third; the boundary between TE and TEO lies at the rostral tip of the posterior middle temporal sulcus (PMTS; Boussaoud et al. 1991; Iwai & Mishkin, 1969).

TE lies at the head of the ventral visual pathway (Ungerleider & Mishkin, 1982) and receives visual input from V1 after processing in areas V2, V4 and TEO. Although certain pathways exist that allow visual information to take shortcuts (e.g. from V2 to TEO, from V4 to posterior TE; for refs. see review by Tanaka, 1996), the stepwise projections are most numerous, and the pathway is generally considered to be a set of

serial processing steps, albeit with certain shortcuts.

IT projects to several areas outside visual cortex including the perirhinal cortex and hippocampus, prefrontal cortex, the amygdala and the striatum of the basal ganglia (see Tanaka, 1996). Projections to these areas are more numerous from TE than from any of the earlier stages of the visual pathway.

Anatomically, the anterior part of IT is divided into the middle temporal gyrus (MTG) dorsally and the inferior temporal gyrus (ITG) ventrally by the anterior middle temporal sulcus[2] (AMTS), with these areas largely corresponding to Brodmann's areas 20 and 21 respectively (Brodmann, 1905). This division is also reflected in patterns of connectivity, and several researchers have suggested (see Buckley et al., 1997) the division of TE and TEO into dorsal and ventral subregions, with the part of TE lateral to the AMTS often referred to as TEd, and the part medial to the AMTS (and occupying the ITG) as TEv.

However, although the ITG was previously thought to be included within the area broadly designated as TE, it is now thought (Buckley et al., 1997) that the lateral edge of perirhinal cortex (Brodmann's areas 35 and 36) may actually lie near the fundus of the AMTS (i.e. at the MTG-ITG boundary). For this reason, this thesis shall adopt the following conventions for clarity when referring to these areas of anterior inferotemporal cortex–

---

[2] These two gyri are, somewhat confusingly, collectively referred to as the Inferior Temporal Gyrus by some authors (e.g. Baylis et al., 1987).

(i) The area of cortex extending laterally and dorsally from the fundus of the AMTS to the fundus of the STS (i.e. TEd) will be referred to as TE.

(ii) Cortex extending medially from the fundus of the AMTS to the rhinal sulcus (i.e. TEv) will be referred to as perirhinal cortex (PRh).

TE has been subdivided into many smaller areas labelled TEa, TEm, TE3, TE2 and TE1 (Seltzer & Pandya, 1978). These areas run roughly parallel to the gyri, with TEa and TEm lying inside and on the ventral lip of the STS, followed by areas TE3–1 across the MTG and ITG. Of the areas outside the STS, TE3 is most posterior and projects to the more anterior areas TE2 and TE1. Functional differences between these areas have been investigated by Baylis et al. (1987), who found some differentiation between receptive field types and latencies, particularly between areas TEa, TEm and

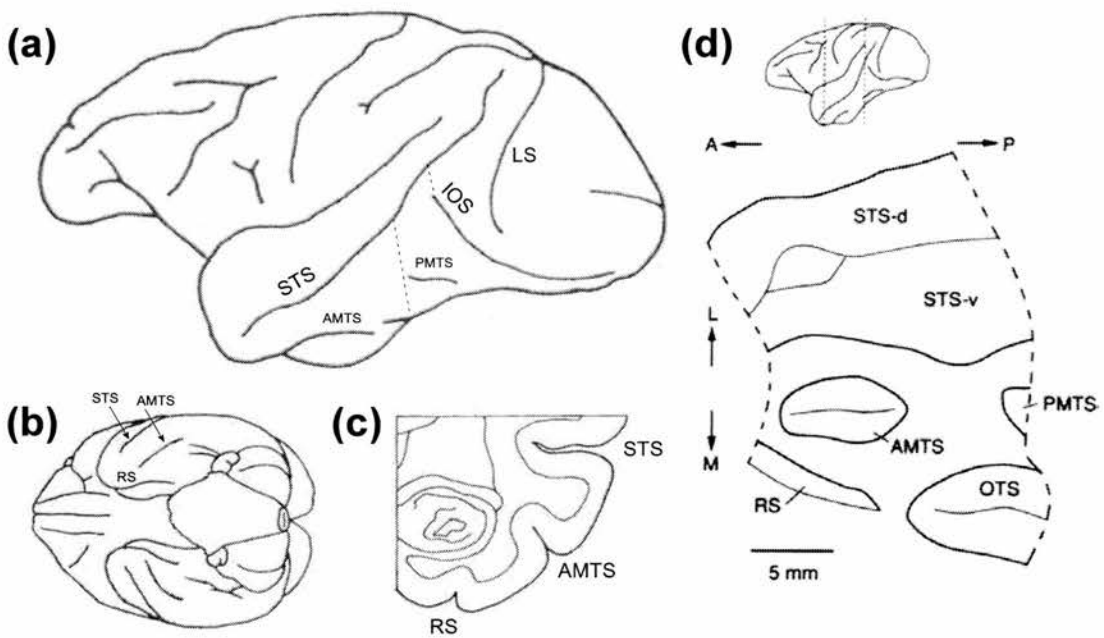


figure 2.1 – Anatomy of the macaque temporal lobe. *(a)* lateral view, with dotted lines indicating boundaries between TE, TEO and V4 *(b)* ventral view *(c)* coronal section *(d)* unfolded map of anterior inferior temporal cortex. *STS* – superior temporal sulcus, *AMTS* – anterior middle temporal sulcus, *PMTS* – posterior middle temporal sulcus, *RS* – rhinal sulcus, *OTS* – occipito-temporal sulcus, *IOS* – inferior occipital sulcus, *LS* – luncate sulcus. [adapted from Kobatake & Tanaka (1994); Higuchi & Miyashita (1996); Kobatake et al. (1998); Komatsu & Ideura (1993)].

TE3 and the more anterior areas TE2 and TE1. However, since most studies do not report recording location to this degree of accuracy, this terminology will not be used in this thesis.

2.1.1. Behavioural Effects of IT Lesions

Studies carried out over the past half-century indicate that lesions of monkey anterior inferotemporal cortex lead to deficits in learning visual discriminations, and loss of those acquired prior to surgery (reviewed in Gross, 1973; Dean, 1976). The deficits reported include discriminations of 3-D objects, 2-D shapes, and along the dimensions of hue, brightness, orientation and size. Impairment was found to be strictly visual, with IT lesioned animals unimpaired in learning discriminations of audition, touch or olfaction. However, monkeys with anterior IT lesions do show some residual visual discrimination ability, and the severity of a deficit tends to depend on the difficulty of the task (i.e. how long it takes a normal monkey to learn the discrimination).

The effect of IT lesions on visual recognition has also been investigated (Mishkin, 1982). Monkeys were trained in a delayed non-matching to sample (DNMS) task, where they were rewarded for choosing a novel object which did not match a previously shown sample object. Animals with bilateral lesions of anterior IT were found to be significantly impaired at relearning the task, with an abrupt increase in impairment as the delay between sample presentation and choice went up. In contrast, animals with lesions of posterior IT were found to perform as well as controls, except where the number of objects to be remembered was also increased, which led to a fall in performance of this group (though their performance remained superior to that of TE lesioned animals).

Mishkin (1982) concluded that TE was critical in the recognition of previously seen objects, acting as a storehouse, where central representations of visual stimuli are formed and stored. TEO was necessary for shape perception, but unnecessary for the recognition task, so long as the number of objects was low and could be adequately discriminated on the basis of other features, such as colour or size.

Is IT largely concerned with visual discrimination (i.e. perception), or with recognition (i.e. learning or memory)? The early lesion studies described above generally destroyed the whole of anterior IT cortex, including both the middle- and inferior temporal gyri (i.e. TEd and TEv/PRh). However, a recent study (Buckley et al., 1997) has provided evidence of a functional double dissociation between these two areas, which has helped clarify anterior inferotemporal cortex's role in the discrimination and recognition processes.

Buckley et al. (1997) trained monkeys in two tasks: a hue discrimination task, where subjects were required to choose green from an array of nine equiluminant colour patches, and a recognition task, based on the DMNS task of Mishkin (1982). The animals were then given lesions of either ITG (PRh group) or TEd (MTG group) and, after recovery, tested again with the two tasks. In the colour discrimination task, the MTG group showed a large impairment in post-operative learning, whereas the PRh group was not impaired. In contrast, the MTG group was only mildly impaired in the DNMS task, but the PRh group showed a large impairment.

The study reveals that there are functional differences within inferotemporal cortex, with TEd more important in the perception of visual form (specifically colour, in this

study[3]), and PRh playing some role in the visual learning and memory of stimuli, allowing objects to be recognised as familiar. However, a more precise description of the function of these areas remains an important goal of future studies.

2.2. Feature Selectivity of IT neurons

The receptive fields of IT neurons (particularly those of area TE) are well described as complex – cells have been found that maximally respond to the outline of a hand, certain views of particular faces, and complex geometrical figures, with specificity for boundary, texture and colour (Gross et al., 1972; Perrett et al., 1982; Tanaka et al., 1991). This section will first review the properties of shape selective responses, investigated in some detail by Tanaka and colleagues in area TE, and will then go on to describe the properties of face-selective cells that have been discovered using somewhat different methodologies in TE and STS.

2.2.1. Shape Selective Responses in TE

The stimulus selectivity of single units in the temporal lobe was first investigated by Gross et al. (1969, 1972) recording from the posterior and central regions of IT. They noted that the receptive fields of TE neurons were large, always included the fovea, and frequently extended across the midline into the ipsilateral hemifield. They also concluded that their standard stimulus set (consisting of slits, spots of light and checkerboards) was less than ideal, with many neurons failing to respond to it, preferring instead to respond uniquely to specific outlines such as the shadow of a hand, or three-dimensional objects placed in front of the animal.

---

[3] A more complete account of studies investigating the effect of inferotemporal lesions on colour discrimination is presented in chapter 5.

This study led to a reductive technique where many objects were first presented to each cell in order to find effective stimuli. Then, cardboard cut-outs were made of effective stimuli in order to determine what particular features were critical to excite the cell (Desimone et al., 1984).

This method has been further developed by Tanaka and colleagues (Tanaka et al., 1991; Fujita et al., 1992; Kobatake & Tanaka, 1994; Tanaka, 1996), and remains the most effective and systematic way to determine the property of these cells –

(i) A large number of pictures and three-dimensional stimuli are presented to the animal.

(ii) Images of effective stimuli are digitised and presented on a computer monitor.

(iii) The image of the most effective stimulus is simplified step by step, until the minimum combination of features that produces the maximal activation of the cell is found.

(iv) After reduction, the features of the simplified stimulus are modified so that the selectivity of the cell can be further examined.

Using an early version of the technique outlined above, Tanaka et al. (1991) examined the selectivity of cells in anterior and posterior regions of IT (TE and TEO) and classified cells into three broad categories of primary (activated by slits or spots by adjusting size, colour or orientation), texture (activated by simple textures such as stripes or dots) and elaborate (requiring a particular shape, or combination of shape with colour or texture). Although cells of each category were present in both regions,

there was an abrupt change in their distribution, with elaborate cells predominantly found in TE, and primary cells mainly found in TEO. An example of two different elaborate cells from this study is presented in figure 2.2.

A later study (Kobatake & Tanaka, 1994) further examined the complexity of stimulus selectivity along the ventral stream, using the same methods to compare V2, V4, TEO and TE. Neurons in V2 were found to respond equally well to simple stimuli (such as coloured bars and dots) when compared the more complex stimuli produced using the reductive technique. However, about half of the cells in V4 and TEO, and the majority of cells in TE required the complex features that had been found, producing little or no response to the simpler stimuli. The authors concluded that there is a systematic increase in complexity along the ventral stream, with ever more elaborate feature detectors created from the output of previous steps and, in parallel, a progressive increase in the area over which these stimuli will provoke a response (i.e. the receptive field size).

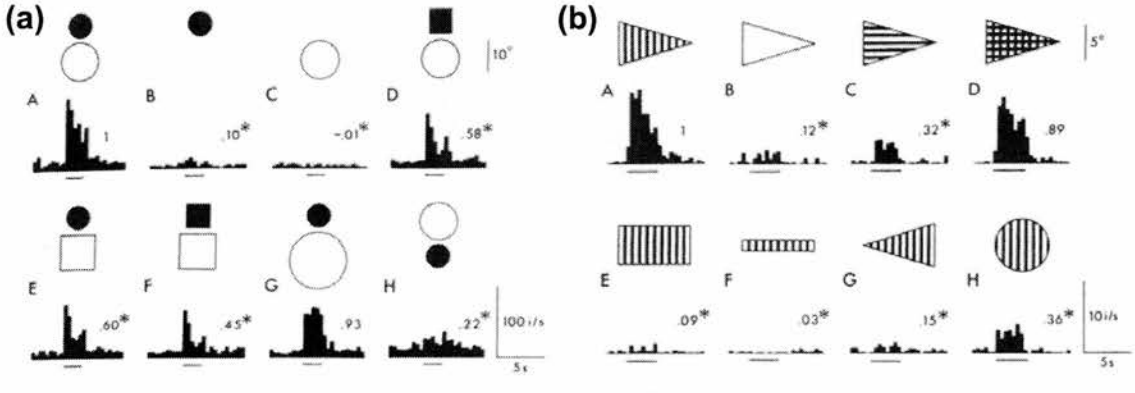


figure 2.2 – The visual selectivity of two TE neurons, from the study of Tanaka et al. (1991). The magnitude of the response relative to the best stimulus is shown by a number above the PSTH, with an asterisk if the difference was significant. Stimulus presentation is indicated by a bar under the PSTH. *(a)* This cell required the presence of a dark disc above a light disc. Neither feature presented alone produces a response. *(b)* This cell required the combination of triangular shape and a vertically oriented texture.

However, there are rather obvious flaws in determining visual selectivity using the systematic reduction technique, specifically – how can we be sure that the initial set of stimuli will contain features sufficient to drive a particular cell? Is the simplified image a true reflection of a cell's underlying selectivity, or are there additional features which, if added or substituted, would produce an even bigger response? Furthermore, is it realistic to assume that a complete description of a cell can be provided by a single set of features, or does the cell in fact code for multiple combinations of features, of which we have only found one? Despite these objections, the method has proved very effective in finding combinations of features that are, at the least, sufficient to drive cells in TE (and other higher visual areas), and the findings are important in providing constraints to inform the design of more methodical and systematic techniques to determine stimulus selectivity.

Two important properties of TE neurons are made apparent by the responses of neurons shown in figure 2.2. Firstly –

1. The responses of TE neurons are non-linear. i.e. the response to features A & B presented together is not the sum of the responses to features A or B, when presented alone.

[boxed paragraphs indicate observations with implications for optimisation]

This principle is very clearly illustrated by the cell shown on the left of figure 2.2. There is a large response to a dark circle above a light circle, but the response to either circle presented alone is close to baseline. A second, very important, property is that –

2. The response of a TE neuron is not all-or-nothing, but increases gradually as a stimulus becomes closer to an 'optimal' set of features.

The existence of a gradient around an optimal stimulus is necessary for any optimisation technique, and this is can be seen in the cells illustrated in figure 2.2,

with both cells producing reduced responses to stimuli that possess some, but not all, of the features necessary to maximally drive the cell.

2.2.2. Constancy of Shape Selective Response

(a) Position (Retinal Location)

As previously noted, neurons in TE have large receptive fields, and stimulus selectivity has been found to be largely constant across these fields (Tanaka, 1996; Gross et al., 1985). Thus TE cells can respond selectively to a stimulus, regardless of its specific location on the retina, providing a basis for perceptual equivalence across retinal translation. An example of a cell displaying this behaviour is shown in figure 2.3(c).

(b) Size

Stimulus selectivity also remains constant over changes in the size of a stimulus, as shown in figure 2.3(a). The tolerance of individual cells varies widely, with some cells continuing to respond strongly to a stimulus over a range of sizes covering four octaves, however the majority of cells are more narrowly tuned to size (Tanaka, 1996; Tanaka et al., 1991). Despite these differences in the overall breadth of tuning, cells tend to have an optimal size of stimulus, that produces the maximal response.

(c) Orientation

In contrast to the wide tuning curves found for position and size, TE cells are very strongly selective for the orientation of a stimulus, with a sharp decline in response as an object is rotated (in the picture plane) away from a cell's preferred orientation. The responses of 8 TE cells, which were typical of the study of Tanaka et al. (1991) are

shown in figure 2.4.

(d) Contrast and Boundary Definition

Figure 2.3(b) shows how some TE cells can maintain selectivity over changes in stimulus contrast (inversion in the example illustrated)[4]. Shape selective responses have also been found which are maintained whether the boundary of the stimulus is defined by luminance, motion (by opposing movement of random dot patterns), and texture (Vogels & Orban, 1996).

(e) Colour

There have been a mixture of reports as to the degree of colour selectivity exhibited by TE neurons. This topic is discussed in a later chapter and is just briefly reviewed here for completeness. Some studies (e.g. Tanaka et al., 1991) suggested that only a small proportion of cells are selective for colour in addition to shape. However, other reports have shown that a majority of cells do have a degree of preference for particular colours, at least within the population of cells that can be driven by relatively simple stimuli (Komatsu et al., 1992; Komatsu & Ideura, 1993). Two examples of cells from the later study, one showing both shape and colour selectivity, the other showing only shape selectivity, can be seen in figure 2.5.

3. While TE neurons are selective for the orientation, and often colour, of a stimulus, they can show invariance for size, position and the means by which the stimulus boundary is defined. However, there may be optimal settings for these parameters which produce the greatest response.

[4] In a later section we shall see that response latency, as well as magnitude, can be altered by stimulus contrast. This has the potential to be problematic for stimulus optimisation procedures.

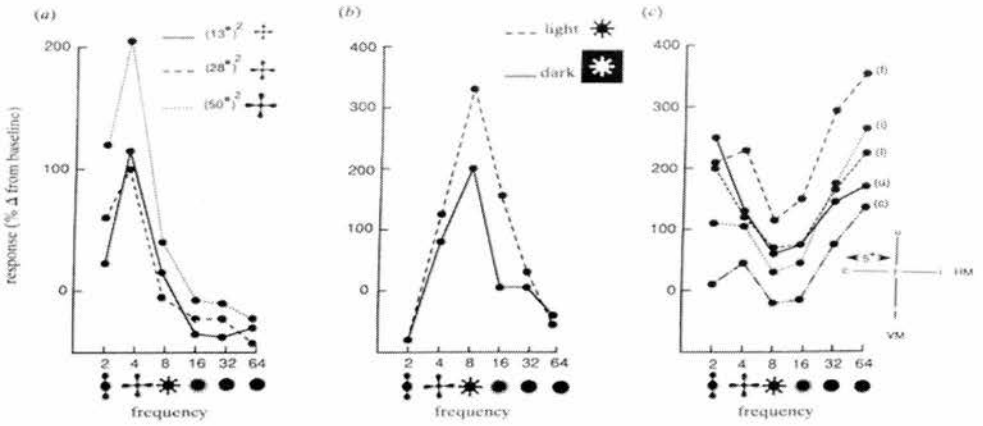


figure 2.3 – Response of three inferotemporal neurons, showing constancy of shape selectivity over changes in *(a)* size, *(b)* contrast and *(c)* retinal location. Key to symbols: f – fovea, c – contralateral field, i – ipsilaterial field u – upper, l – lower. From Gross et al. (1985).

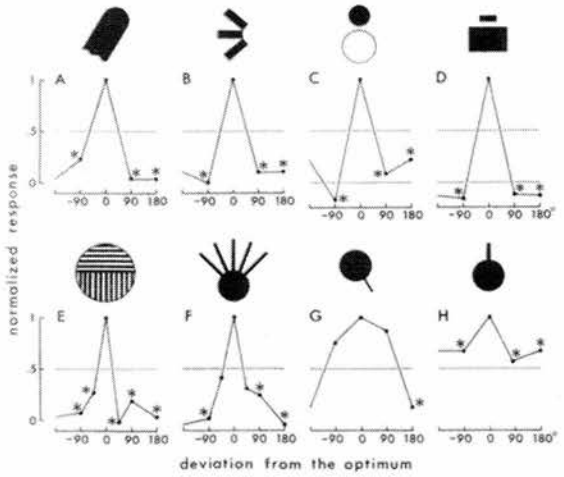


figure 2.4 – The effect of stimulus orientation on the response of 8 neurons in anterior IT. Responses are normalised to the maximum response (to best stimulus at its optimal orientation). The majority of cells are strongly tuned for orientation, with only G and H showing a degree of response invariance. From Tanaka et al. (1991).

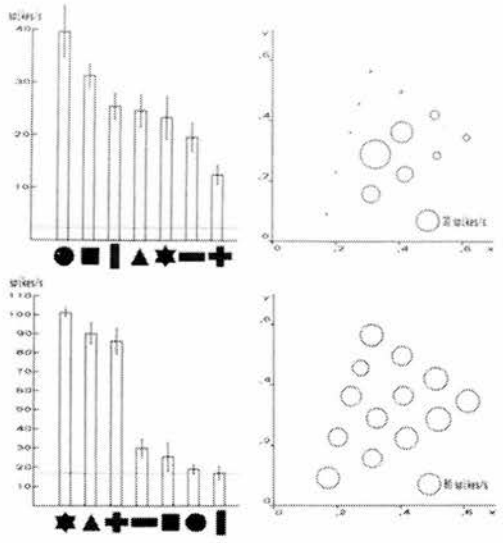


figure 2.5 – The responses of two anterior IT neurons to shape (left) and colour of the best shape (right). Colour responses are plotted in CIE colour space, with the size of circle indicating the magnitude of response. The top neuron is selective for both shape and its colour, the lower neuron is only selective for shape. From Komatsu & Ideura (1993).

2.2.3. Face Selective Responses in STS/TE

Gross (1972) made the first discovery of cells that responded selectively to faces in the temporal cortex. These cells are treated separately here because they may form a distinct population, and there is an extensive body of literature (perhaps more extensive than that examining pattern selectivity in IT more generally) which has concentrated exclusively on the properties of these cells.

Face cells are defined as those cells that produce responses to faces that are at least twice as large as to any other stimulus tested, such as gratings, geometric stimuli and three-dimensional objects (Perrett et al., 1982; Baylis et al., 1985). Though face cells have been found in throughout the STS and TE, there is some evidence for differential properties between the two regions, which will be discussed below.

(a) Viewer- and Object-Centred coding in STS

Face cells generally exhibit a preference for a particular view of the head (e.g. a front view or profile view) with the response falling as the head is rotated away from the preferred view (Perrett et al., 1982, 1991). These cells are termed "viewer-centred", because their response to a particular head or individual is not invariant, but depends on the position of the viewer relative to the head or individual being viewed. A small number of cells produce bimodal responses (e.g. to both the left and right profile views), and a smaller number still have been found that respond to all views of a head, without distinguishing between particular views. This latter category is termed "object-centred" since they would produce an invariant response to an individual irrespective of viewing position. Examples of the responses of uni- and bimodal viewer-centred cells, and a object-centred cell are shown in figure 2.6.

It has been suggested that object-centred representations could be built from the outputs of a number of viewer-centred cells, and an analysis of latency data (Perrett et al., 1992) indicates that this may be the case, with slightly higher latencies found for cells that show view-invariant (i.e. object-centred) properties.

4. If a cell responds to an object (e.g. a head) irrespective of view, where particular views can share little or nothing in terms of retinal input, then multiple runs or multiple solutions would be required for optimisation to produce an adequate description of the cell.

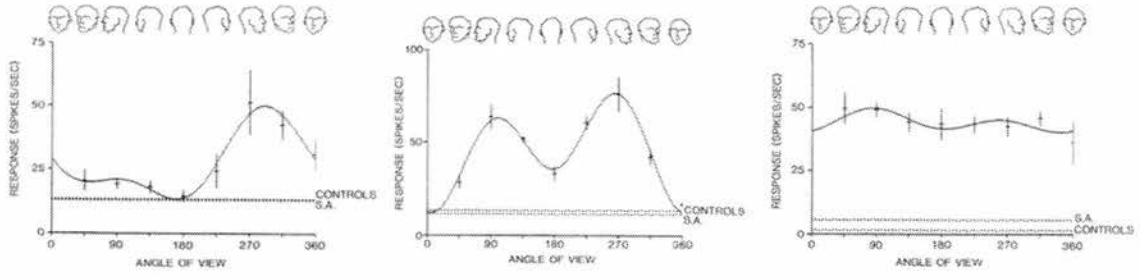


figure 2.6 – Viewer-centred (*left*) and object-centred (*right*) coding of faces in STS. Bimodal responses, such as that shown (*middle*) may reflect intermediate stages of face processing. Dotted lines at the bottom of each graph show spontaneous activity, and that to control (i.e. non-face) stimuli. From Perrett et al. (1991).

(b) Response to Elements and Structural Configuration of a Face

Are these cells responding to the face as a whole, or perhaps just some element (e.g. the eyes) that might reduce to a more simple geometric description of the selectivity? This has been investigated to some extent (Perrett et al., 1982; Rolls et al., 1985) and a continuum of selectivities found. For example, Perrett et al. (1982) reports cells that require only a single feature (the eyes), any one of several features or, in some cases, require all of the features of the face in order to respond. This study also found that the relative positioning of facial elements is important, with jumbled faces leading to a significant reduction in the response.

(c) Invariance to Position, Size, Colour, Contrast and Orientation

Although the spatial arrangement of features within a face is important, face cells show generalisation for their preferred view across retinal translation (Desimone et al., 1984), size and contrast (Rolls & Baylis, 1986) and colour[5] (Perrett et al., 1982) – though, like the shape-selective responses described earlier, cells may have preferred settings for these properties that will evoke a maximal response.

Neurons in TE are selective for shape orientation in the picture plane and this orientation sensitivity has also been found for cells responsive to faces in TE (Tanaka et al., 1991). In STS, however, responses that are *invariant* to changes in orientation have been described (Perrett et al., 1982; Ashbridge et al., 2000). This may reflect the location of TE at an earlier stage in a pathway that leads to object-centred responses, since TE is known to project to the STS (Seltzer & Pandya, 1978).

(d) Population Coding of Identity in STS and TE

Although Perrett et al. (1984, 1991) found that the majority of cells in the dorsal bank of the STS were not sensitive to the identity of a face, other studies have indicated that a large proportion of cells do code for facial identity, particularly those that were recorded from the ventral bank of the STS (part of TE) and the inferior temporal gyrus (Baylis et al., 1985; Hasselmo et al., 1989a).

---

[5] The effect of colour on the responses of neurons in IT to faces (and other complex stimuli) is investigated more fully in a later chapter.

Baylis et al. (1985) found that cells had relatively broad tuning for identity, with few cells that responded exclusively to views of only one individual. They suggested that identity was coded by an ensemble of neurons, with different relative responses to each member of a set of faces.

Although they only found a small number of cells showing this behaviour, Perrett et al. 1991) noted that both viewer- and object-centred cells could produce differential responses to individuals.

Knowing something about the population coding being used by a set of neurons offers the potential to restrict optimisation to a small subset of image space. For example, if we are confident that cells in a particular area are coding for facial identity, we could carry out optimisation on a set of parameters designed to be sufficient to span this 'face' space (i.e. with parameters for eye separation, colour, etc). Obviously, this is moving away from the ideal of an generalised image search (i.e. sufficient to find the optimal stimuli for *any* cell) but limiting the search space by this approach may allow optimisation techniques to be applied in a to a more tractable domain.

5. Knowledge of population coding may allow optimisation to be restricted to a more limited image subspace, improving the chances of success.

(e) Neurons Responsive to Movement and Gestures in STS

A minority of neurons in STS produce no response to static faces, and respond only if the face is moving e.g. towards or away from the viewer, or if it is rotated into profile (Perrett et al., 1985; Jeeves et al., 1983). This was further investigated in some detail by Hasselmo et al. (1989b), who tested cells with a variety of different head movements. A small number of neurons were found that responded selectively to

ventral flexion (head lowering – a gesture important in social interaction), dorsal flexion, rotation and translation of the head (e.g. from left to right).

Furthermore, cells responsive to heads undergoing ventral flexion were found to respond specifically, even when the view of the head was inverted. This suggests that these cells operate in a object-centred manner, since retinal movement vectors relating to flexion would be reversed by the inversion.

(f) Differential Populations in STS and TE

Some of the studies presented above had led to the suggestion that face cells in TE code largely for facial identity, whereas the populations found in STS may be more involved in the processing of expressions, movements, gestures and gaze (Rolls, 1992; Perrett et al., 1992; Hasselmo et al., 1989a).

It seems likely that cells in STS largely operate in an object-centred space, with networks within TE (and from TE to STS) largely responsible for combining the outputs of neurons with appropriate selectivity to make this possible. On this basis –

6. Optimisation techniques relying on the manipulation of two-dimensional images are more likely to find success in TE and TEO, whereas investigations of STS may require a methodology based on dynamic manipulation of three-dimensional objects.

(g) Are Face Cells Really a Special Case?

In a later section we shall see that neurons with some very similar properties to face cells (e.g. view invariance) are found in TE after exposure to novel three-dimensional shapes (a comparison of figures 2.6 and 2.16 is instructive). This suggests that face

cells may simply be the result of a more general form of learning that allows the brain to discriminate between similar objects.

This could be problematic for optimisation, since face cells in STS that respond to specific movements and gestures show that dynamic properties of the visual world can also be integrated into stimulus selectivity, where this aids discrimination. This could occur if inputs representing constituent parts of a gesture are activated simultaneously (the range of latencies found in STS and IT would make this possible even without specific delay or mnemonic mechanisms) or if neurons can distinguish the order of activation of their inputs.

More generally, it is therefore possible that stimulus selectivity in IT might also involve a dynamic component, if this was relevant and could play a role in helping the visual system to discriminate between objects or behaviours in a social context. Neurons with such properties might go unnoticed with standard stimulus test batteries, and could perhaps account for many of the neurons currently classed as visually unresponsive.

2.3. Response Dynamics

Up to this point, we have assumed that visual neurons have a fixed latency and that the response to a particular stimulus is fully described by a simple measure of the total number of spikes emitted in the time window following stimulus presentation. In this section we shall challenge this assumption, reviewing a number of studies that have investigated the temporal dynamics of visual responses in IT.

2.3.1. Persistence and Masking

Visual persistence is the effect whereby the duration of the response of a visual cell to a stimulus outlasts the duration of the actual stimulus presentation itself. This phenomenon is found throughout the visual system, up to and including IT, where Rolls & Tovée (1994) described responses lasting 200-300ms to very briefly (16ms) presented faces (see figure 2.7a).

If visual stimuli are represented by particular neurons (or subsets of neurons) in IT, and these cells continue to respond to a stimulus long after it has disappeared from view, then it is interesting to consider what happens when two stimuli are presented in rapid succession. Perhaps both stimuli will be simultaneously represented, with the firing of two different sets of neurons overlapping in time; alternatively, the activity of neurons responding to the first stimulus could be suppressed by the firing of other cells responding to the second stimulus, perhaps by means of reciprocal inhibitory connections between pattern selective units.

This question was answered by backward-masking studies, where the presentation of an effective stimulus is closely followed by a masking stimulus, with varying stimulus onset asynchrony (SOA). In general these studies have shown that the presence of a masking stimulus significantly shortens the response to an effective stimulus, both with face-selective cells in STS (Rolls & Tovée, 1994 – figure 2.7b-e), and cells responding to shapes in TE (Kovács et al., 1995), with the response duration becoming shorter and shorter as the SOA is reduced. Despite this, responses still outlast stimulus presentation time, and information analysis (Rolls et al., 1999) shows that significant amounts of information about the stimulus seen are still available even

with an SOA of 20ms.

Generally, the mask stimulus is chosen to be one that is not effective at driving the cell under study. Interesting, therefore, is the final experiment conducted by Rolls & Tovée (1994) where masking was investigated using an *effective* stimulus as a mask. As might be expected, perhaps, two separate responses were seen at higher SOAs. However, when SOA was reduced, the standard masking effect was seen, with the response to the target stimulus being shortened and reduced, such that the overall response was smaller than that to the target stimulus presented alone.

These studies indicate that the duration of response to an effective stimulus (and therefore the overall spike count and total information available) can be shortened by following stimuli, irrespective of whether these subsequent stimuli are also effective

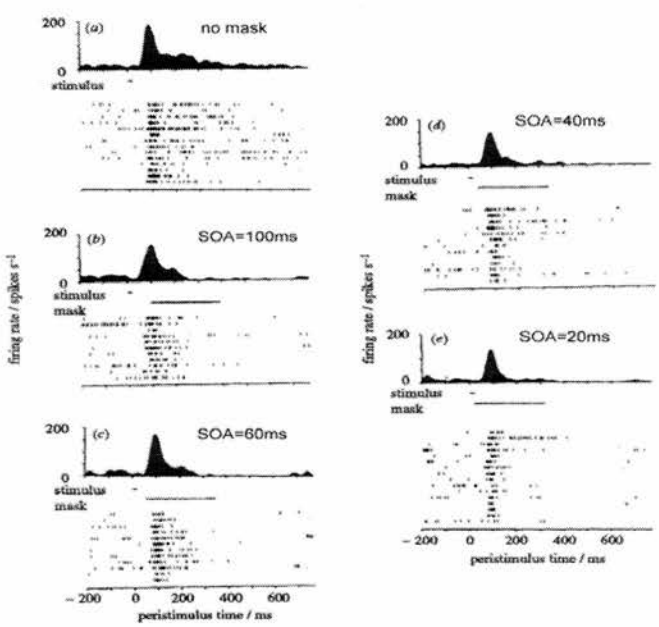


figure 2.7 – Persistence and backward masking. Effective stimuli were either presented alone *(a)* or followed by a mask *(b)-(e)* In this case, the mask stimulus was ineffective at driving the cell by itself. Stimulus onset asynchrony (SOA) indicates the amount of time between stimulus and mask presentation. It can be seen that, without a mask, the response is several hundred milliseconds longer than the stimulus duration. Presentation of the mask shortens the response to the effective stimulus. However, even with an SOA of 20ms *(e)* the cell's response is still much longer than the actual presentation time of the effective stimulus. Adapted from Rolls & Tovée (1994).

at driving the cell or not. This is very surprising, since we might have assumed that two effective stimuli following each other closely would provoke a greater (or longer) response that the presentation of a single stimulus alone. This finding does not rest easily with the idea that masking is due to lateral inhibitory effects between cells with different pattern selectivities (Rolls & Tovée, 1994), though it can be made to do so if we assume short-term depression of incoming connections resulting in a reduced response to the presentation of the second stimulus, alongside inhibition from neighbouring cells that are also (at least partially) tuned to aspects of that stimulus.

7. The response to a stimulus can outlast the presentation time by hundreds of milliseconds; however, response persistence can be cut short by subsequent stimuli, whether or not these stimuli are effective in driving the cell by themselves.

2.3.2. Rapid Serial Visual Presentation (RSVP)

RSVP can be viewed as a more extreme version of the masking paradigm, where rather than simply being following by a single masking image, stimuli are instead embedded in a continuous stream of images, which are rapidly presented, one after another, at the same location on a screen. While RSVP has been used elsewhere for behavioural studies (Chun & Potter, 1995; Potter & Levy, 1969), it was first developed and applied to neurophysiological experiments by Keysers et al. (2001), in order to allow the selectivity of visual neurons to be investigated with large numbers of different stimuli.

The paradigm is outlined in figure 2.8, and has been compared to normal viewing conditions, where a saccade occurs approximately 3 times a second, resulting in a series of snapshots of the visual world, each falling on the same region of the retina.

Because modern computer monitors can present a new image approximately every 14ms, RSVP could provide a powerful technique for exhaustively investigating the properties of visual neurons, with the potential to test almost one hundred novel stimuli per second.

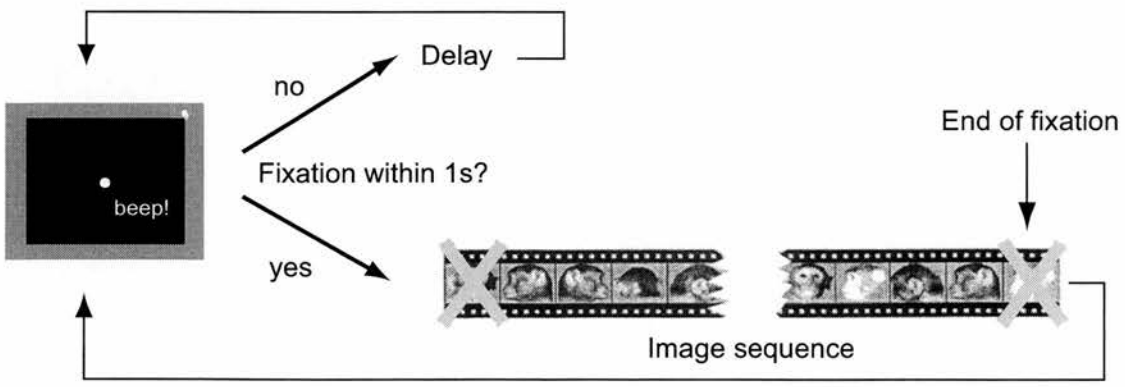


figure 2.8 – RSVP paradigm. Eye position is monitored, and presentation begins when the subject fixates on a dot presented in the centre of the screen. This is followed by a rapid and continuous sequence of images (in the centre of the monitor), which ends only when the subject breaks fixation. Responses relating to the first and last images in the fixation period are discarded. Redrawn from Keysers (2000).

2.3.3. Response Properties Under RSVP

Under RSVP conditions, a cell is not simply responding to a single stimulus presented in isolation, but instead to a randomised stream of images. The response to a particular stimulus can be established by lining up each section of the continuous recording which contains an occurrence of the stimulus, and represents the systematic response to that stimulus against a background of the activity evoked by all the stimuli.

Keysers et al. (2001) tested IT neurons with 8 stimuli which were chosen to span the response range of the neuron under study. They found that the majority of cells (65%) were able to discriminate between stimuli to a significant level, even at the fastest presentation rate of 14ms per image. This can be seen in figure 2.9(a), where stimulus rank is plotted against presentation rate for the population of cells tested. It is noteworthy that, although selectivity is preserved at higher rates, the range spanned by the responses decreased dramatically.

The response duration to an effective stimulus in an RSVP sequence was found to typically last 60ms more than the duration of the stimulus itself, and this effect was robust across the presentation rates tested. This can be compared to the masking results described earlier, where reducing the SOA resulted in shortened responses, but not below a certain lower limit to response duration, which was also around 60ms (Rolls & Tovée, 1994). This has implications for using the RSVP technique for stimulus optimisation, because the response of cell at any given point in time cannot be simply and directly attributed to a particular stimulus in the sequence, particularly if the presentation rate is high. If the set of stimuli contains a mixture of effective and ineffective stimuli, then a solution is provided by averaging the response over

multiple presentations (as described earlier). Other potential solutions to the response sampling problem are addressed in chapter 4.

Introducing gaps between stimuli leads to response persistence, such that responses are similar to those that would be obtained had the stimulus remained on the screen (Keysers, 2000). This was unpredictable from the masking results described earlier, and implies that SOA rather than stimulus duration effects the persistence of visual responses in IT. An example of this effect is shown in figure 2.9(b).

8. SOA, rather than stimulus duration, affects the response magnitude and duration.

9. The response duration to stimuli presented in RSVP sequences equals SOA + 60ms.

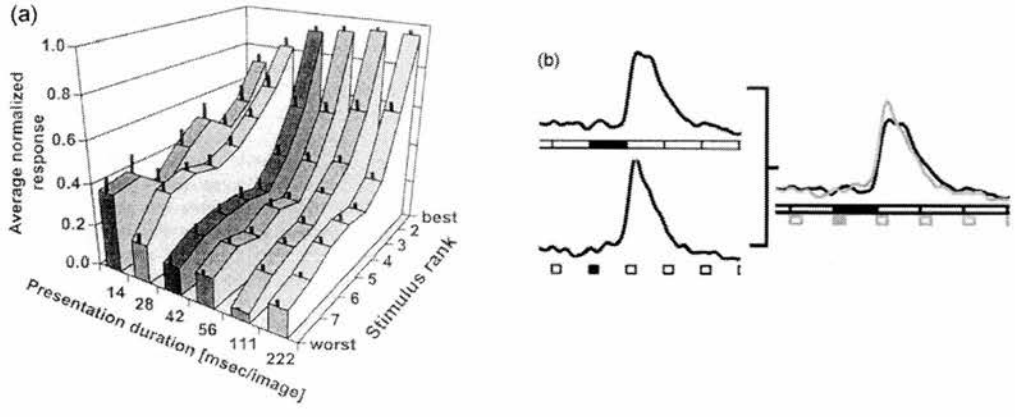


figure 2.9 – *(a)* Response as a function of presentation rate and stimulus rank order (ranking is based on responses during the slowest presentation rate condition). Responses were normalised for each presentation rate then averaged across the population. Black vertical lines indicate the standard error. It can be clearly seen that stimulus selectivity is maintained, even at the fastest presentation rate. From Keysers et al. (2001).

(b) Responses to an effective stimulus in an RSVP sequence, with *(bottom)* and without *(top)* gaps. The SOA was equal in both cases (111ms). Responses are shown superimposed on the right, with the gap condition in grey and the no-gap condition in black. Filled rectangles indicate the stimulus of alignment, empty rectangles indicate other randomised stimuli in the sequence. The response duration is equal in both cases. This effect was replicated at twice the presentation rate (SOA=55ms, not shown). From Keysers (2000).

2.3.4. Temporal Encoding in Pattern-Selective Responses

Is it reasonable to assume that the response latency of a cell is fixed, with different patterns producing responses that vary only in magnitude? Perhaps not – Perrett et al. (1998) suggest at least one means by which evidence accumulation could lead to differences in latency within a single cell. In this framework, objects presented in unfamiliar views might be expected to only partially activate the subset of neurons tuned to the object's features. This would lead to increased response latencies in postsynaptic cells, since more time would be required for them to integrate sufficient activity to reach threshold. In a later section we shall see that this effect almost certainly occurs in conditions of reduced stimulus contrast.

Richmond et al. (1987) report both latency differences in response onset for different Walsh patterns (figure 2.10a) and varying response dynamics, with different patterns of excitation and inhibition, depending on the stimulus (figure 2.10b). In the second paper of this series (Richmond & Optican, 1987) principal component analysis was used to quantify these temporal effects. While the first principal component was

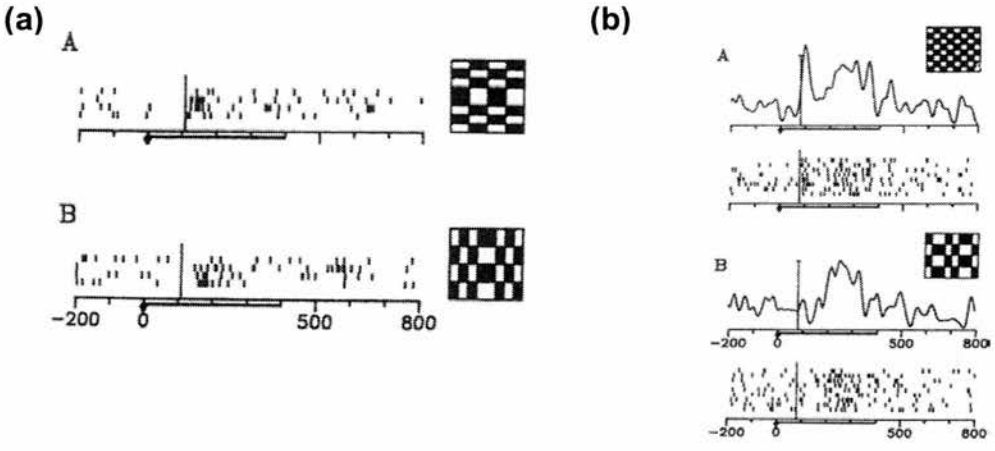


figure 2.10 – *(left)* Latency difference in response onset in one neuron to two different patterns. Line below the axis represents stimulus presentation. Measured latencies were 120ms and 150ms to A and B respectively. *(right)* Differing temporal dynamics in the response of one neuron to two different stimuli. Notice how the response to stimulus A shows an early excitatory peak which is entirely absent from the response in B. From Richmond et al. (1987).

highly correlated with the spike count, higher principal components were also dependent on stimulus, indicating that the distribution of spikes, as well as their number, depends on stimulus features. Information analysis (Optican & Richmond, 1987) showed that, while about half of the information was contained in the magnitude of the response, the other half was accounted for by temporal effects within the response.

In many ways, this is not entirely surprising. Neurons with a range of latencies are found throughout the visual system, so it would seem reasonable to expect that a neuron with complex pattern selectivity in IT might undergo distinct patterns of excitation and inhibition as separate individual features of a stimulus are signalled by different pre-synaptic cells of varying latency.

However, a potential problem with temporal encoding is the necessarily large amount of time it takes to decode the information. e.g. Richmond et al. (1987) describe variations in the spike train taking place over hundreds of milliseconds. Yet the latency of pattern selective responses in IT are often under 100ms, which allows only about 15ms at each stage of cortical processing before output is available at the next stage (Rolls 1991, 1992; Oram & Perrett, 1992), and we have already seen that pattern selective responses in IT can be as short as 60ms.

Later studies (Tovée et al., 1993; Tovée & Rolls, 1995), using a set of faces as stimuli, have questioned the amount of information carried by second and subsequent principal components. They found that the first principal component (i.e. firing rate) carries by far the most information, with the second and third principal components accounting for only small proportions. Furthermore, the first 20 or 50ms is sufficient

to give a reasonable proportion (~85%) of the information available in the firing rate over 400ms.

However, it is also clear that subtle differences in pattern selectivity might be revealed if the window of analysis is varied over the time course of the response.

10. The majority of stimulus information is carried by the first principal component of the response correlated with stimulus identity i.e. the spike count.

2.3.5. Contrast Effects on Latency

Response latency has been found to increase as stimulus contrast decreases in a number of visual areas, including the retina, LGN and V1. This effect is responsible for a number of visual illusions, including the Pulfrich pendulum illusion (Pulfrich, 1922).

Earlier we saw that the response magnitude of face cells in STS decreases as stimulus contrast is reduced. However, experiments using the RSVP technique[6] have revealed that stimulus contrast also has a striking effect on the latency of cells in STS and IT (see figure 2.11), with a halving of stimulus contrast resulting in an increase in response latency of 50-100ms. This effect was found at both rapid (55ms per stimulus) and slower (332 ms per stimulus) rates of presentation, with similar shifts in response latency at the two speeds.

---

[6] This work (Oram, in preparation) was carried out in parallel with the experiments contained in this thesis, and a table of cells included in this contrast experiment, along with reconstructions of cell locations have been included in appendices 2 & 3.

The change in latency cannot be simply explained as a consequence of the reduced response magnitude, because responses to the full contrast versions of intermediate stimuli were often smaller than those to reduced contrast versions of the effective stimulus, yet occurred at earlier latencies.

This effect has the potential to be problematic for the design of stimulus optimisation experiments, because it could cause the response of a low contrast stimulus to be incorrectly attributed to a following higher contrast image. It would be interesting to know whether high contrast stimuli (whether effective or not in themselves) could mask the delayed response to a lower contrast stimulus. Although the experiment included a presentation rate of 55ms, low contrast effective stimuli would, in some cases, be followed by low contrast ineffective stimuli, which might not be sufficient to mask the response. On the other hand, they would also be sometimes followed by a

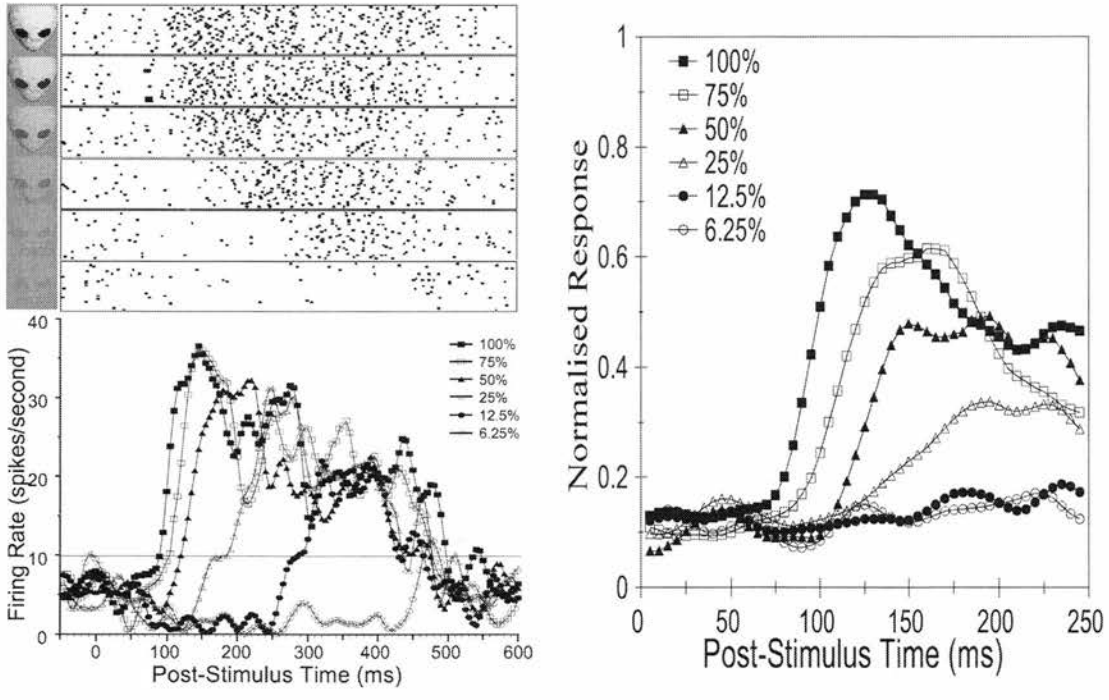


figure 2.11 – Contrast effects on latency. *(left)* Rastergrams and PSTH showing the effect of stimulus contrast on a single neuron in IT. As stimulus contrast is reduced, there is a dramatic increase in response latency. *(right)* Population results for 21 neurons. Individual responses were normalised to the peak response for the maximum contrast stimulus, then averaged. Reducing stimulus contrast leads to a reduction in the maximum firing rate, and an increase in response latency [M.Oram, forthcoming].

high-contrast ineffective stimuli, which might mask the response, but not produce a response by themselves in the cell. The averaging of these two effects could partially account for decreasing responsive magnitude as the contrast of the stimulus of alignment is reduced.

It is also interesting to ask if this finding might be related to the temporal dynamics of the response waveform described earlier. A stimulus (e.g. a face) may contain areas of relatively high and low contrast, resulting temporal differences in the activation of feature detectors earlier in the visual system. This could result in temporally dispersed patterns of excitation and inhibition when these features are integrated further along the visual system, such as in the response of a face cell in STS.

11. Response latency increases in STS and IT as stimulus contrast is reduced. This effect is relatively large, and can result in latency shifts of hundreds of milliseconds.

2.3.6. Summary

Responses outlast stimulus duration by 200-300 milliseconds. This persistence is curtailed by subsequent stimulus presentation in masking and RSVP paradigms. However, even at very high presentation rates, responses are still typically 60ms longer than stimulus duration, which presents a sampling problem which must be dealt with.

There is some evidence that, within a single cell, different stimuli can evoke responses with different latencies. However, analysis has shown that the majority of information about stimulus identity is carried in the overall rate of firing. Manipulating the contrast of a stimulus results in a large effect on response latency. This finding will need to be incorporated into the design of stimulus optimisation experiments.

2.4. Plasticity, Learning and Memory in Visual IT Neurons

A role for IT in visual learning and memory is suggested both by the lesion studies described earlier, and by its location at the end of the ventral visual processing pathway, with close anatomical links to areas implicated in the function of memory. This section reviews studies showing mnemonic effects in IT neurons, considering the implications this may have for stimulus optimisation experiments.

2.4.1. Short-Term Memory

Inferotemporal visual neurons are not adequately described as simple visual filters – a number of studies (Miller & Desimone, 1994; Miller et al., 1991; Baylis & Rolls, 1987) have shown that, in many cells, their response is a joint function of both the immediate visual input and short-term memory traces.

Two parallel short term influences have been described (Miller & Desimone, 1994): an automatic mechanism, relating to visual experience of the immediate past, and an active system which is coupled to working memory. Different subsets of IT neurons appear to participate in each of these mechanisms, and they are discussed in turn below.

(a) Automatic System

The automatic system has been termed "adaptive mnemonic filtering" (Miller et al., 1991) describing an effect whereby the response of a visually driven neuron is changed (generally suppressed) on repeat presentations of a stimulus.

Baylis & Rolls (1987) found this effect in 26% of visually responsive neurons in STS & TE, with "familiar" presentations of a stimulus evoking a different response to the "novel" presentation, in a serial recognition task. However, they found the effect to

be short-lived, failing to span even one intermediate stimulus in the majority of cases. They note that the effect was also present in an untrained animal, with the second presentation of a stimulus generally leading to a suppression of the response, suggesting that the effect is not task-related. However, this was not studied in detail.

A longer lasting effect was found by Miller et al. (1991) using a delayed match to sample (DMS) task. Monkeys were trained to indicate when one of a sequence of test stimuli matched a previously seen sample stimulus. They found that responses to test stimuli which matched the sample were significantly lower than responses to non-matching test items (figure 2.12a). This effect was seen in about 50% of cells. Moreover, this difference in response between matching and non-matching stimuli was found to span several intervening items (figure 2.12b). The authors suggested that IT neurons act as filters which preferentially pass on information about aspects of the visual world which are novel or have not been recently seen. It is not clear why this

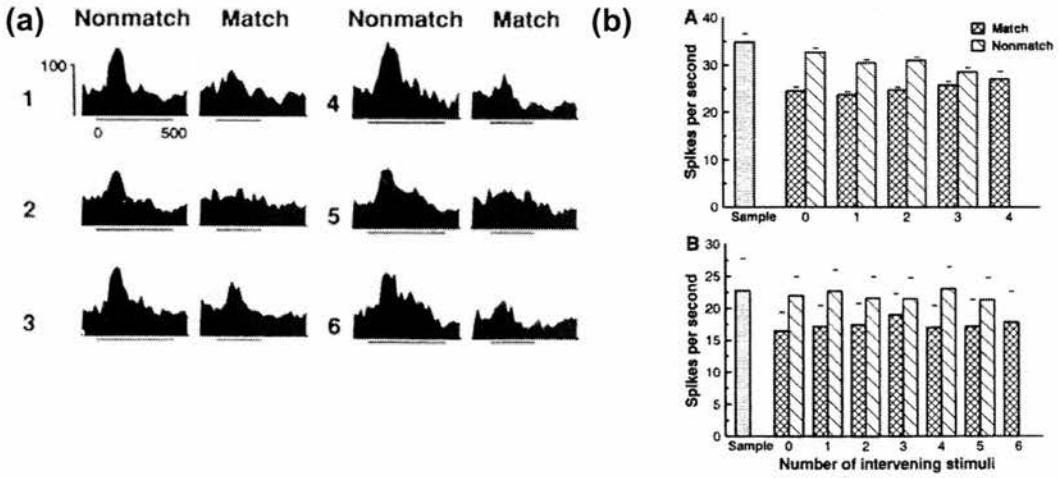


figure 2.12 – From Miller et al. (1991). *(a)* The responses of a single IT neuron to six test items, which matched the sample in some trials (left) and did not match on others (right). There were one to four intervening test presentations between the sample and the test item shown. *(b)* The average response to matching and non-matching test items with the number of intervening stimuli indicated along the x-axis. This figure includes only test items where there a significant difference was found in the response to matching and non-matching items.

result differed from the earlier experiment of Baylis & Rolls (1987), but it may be related to differences between the tasks and recording location.

The above experiment could not determine whether the suppression of response related to a passive mechanism involving stimulus novelty, or was due to active influences of the sample stimulus held in working memory. However, a further study by Miller & Desimone (1994) with a variant of the task has allowed these factors to be distinguished.

In addition to the conventional trials in the DMS task, where none of the test stimuli matched each other (e.g. sample "A" followed by test items "BCDEA"), Miller & Desimone (1994) introduced a second type of trial, which was termed "ABBA", where repeated non-matching test items were shown. The task allows changes in the response of a cell to be attributed to either simple repetition, or the effect of a sample held in working memory. It is outlined in figure 2.13.

Of the cells showing both stimulus selectivity and mnemonic effects (about 50% of the ventral IT neurons tested), roughly two-thirds showed suppression of the response

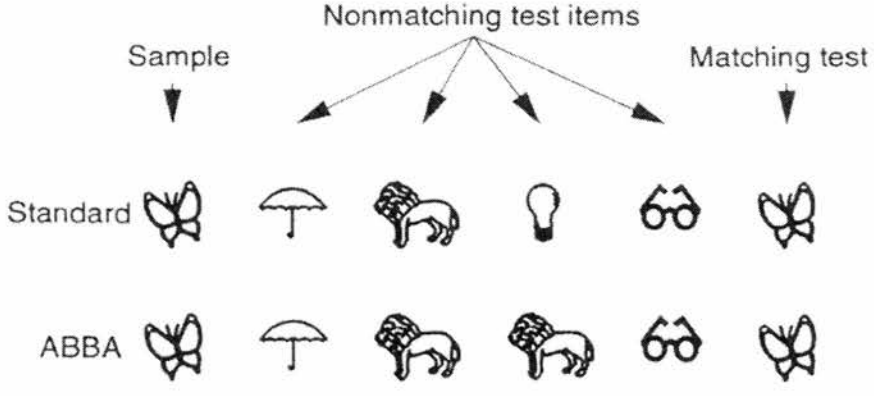


figure 2.13 – The modified DMS task used in Miller & Desimone (1994), with standard trials shown at the top, and the additional ABBA trials shown at the bottom. Repeated non-match trials had either zero (as pictured) or one intervening trial (i.e. ABCBA).

to match stimuli, as in earlier studies. However, the ABBA trials revealed that this suppression also occurred with the repeat presentation of the non-matching stimuli, with no significant difference found between the two (illustrated in the left panels of figure 2.14). Thus responses were suppressed with stimulus repetition, whether or not this was relevant to the task.

Automatic suppression of familiar stimuli may be inherent within IT visual networks, perhaps caused by depression of synapses after repeated use. However, the remaining third of mnemonic cells in this study showed a very different pattern of behaviour, which instead related to the sample held in working memory. This is described below.

(b) Active System

The majority of cells studied by Miller & Desimone (1994) showed response suppression to stimulus repetition, regardless of whether a stimulus matched the sample in memory. However, one-third of the cells behaved quite differently, with *enhanced* responses to test stimuli matching the sample. In contrast to the suppressed cells, this enhancement was specific to matching test items, as it did not occur with repeated presentation of non-match test items (as shown in the right panels of figure 1.14). This effect also spanned several stimuli, uniquely identifying the stimulus matching the sample held in memory.

The active system appears to allow the response of IT neurons to be biased by the contents of working memory such that they give a potentiated response to a matching stimulus, when it occurs. This appears more likely than enhancement by a memory-related feedback signal, since the increased response is present simultaneously with the onset of visually evoked activity (see figure 2.14, bottom right).

These studies reveal at least two parallel short term mnemonic mechanisms acting on different subsets of IT cells. One system results in a reduction of the response to stimuli that have recently be seen, and could be viewed as short-term habituation to a stimulus; the other enhances the response to a stimulus if it matches one held in working memory. The effects of both systems appear to be relatively weak, modifying, rather than completely overriding the stimulus selectivity of a cell.

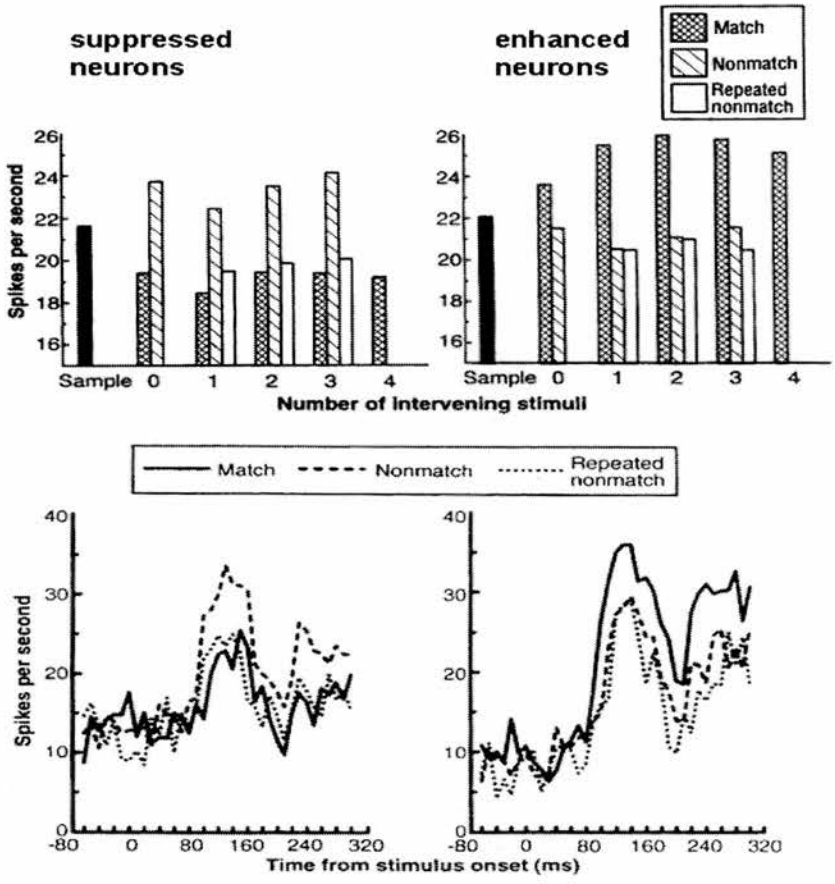


figure 2.14 – From Miller & Desimone (1994). *Left* – suppressed neurons, *Right* – enhanced neurons. The top charts show the population average response to matching, non-matching and repeated non-matching stimuli. The lower charts show the population response across time.

12. Habituation occurs to visual stimuli in at least some cells in IT. Working memory can potentiate responses to relevant stimuli. Both effects, however, are relatively weak.

2.4.2. Experience-Dependent Plasticity in Receptive Fields

The previous section considered how both working memory and recent visual experience can bias the responsiveness of IT cells in the short-term, acting in conjunction with stimulus selectivity to determine the response. This section examines whether stimulus selectivity itself is modified as a result of visual experience.

The receptive fields of cells in monkey striate cortex (V1) have been known for some time to be shaped as a result of visual experience, at least in the early stages of development (Hubel et al, 1977). Does the visual system continue to remain plastic? And to what extent is the stimulus selectivity of neurons in a high-order area (such as anterior IT) also shaped by visual experience, particularly in the adult animal? A number of studies have recently provided clear evidence that experience dependent plasticity does indeed occur, at least for a subset of the cells present within AIT cortex.

Kobatake et al. (1998) looked to see what changes occurred in TE of adult monkeys that had been trained over a period of months to discriminate among members of a class of simple geometric shapes in a DMS task. When training was complete, they measured the response of neurons to a large range of visual stimuli, including the training set. It was found that those stimuli that had been part of the training set evoked a higher response than any other stimuli in 25% of TE neurons in the trained monkeys, compared to only 5% of the cells in untrained monkeys.

The tuning of cells was found to be relatively broad (illustrated by the example cell shown in figure 2.15a), with several of the training stimuli often evoking significant

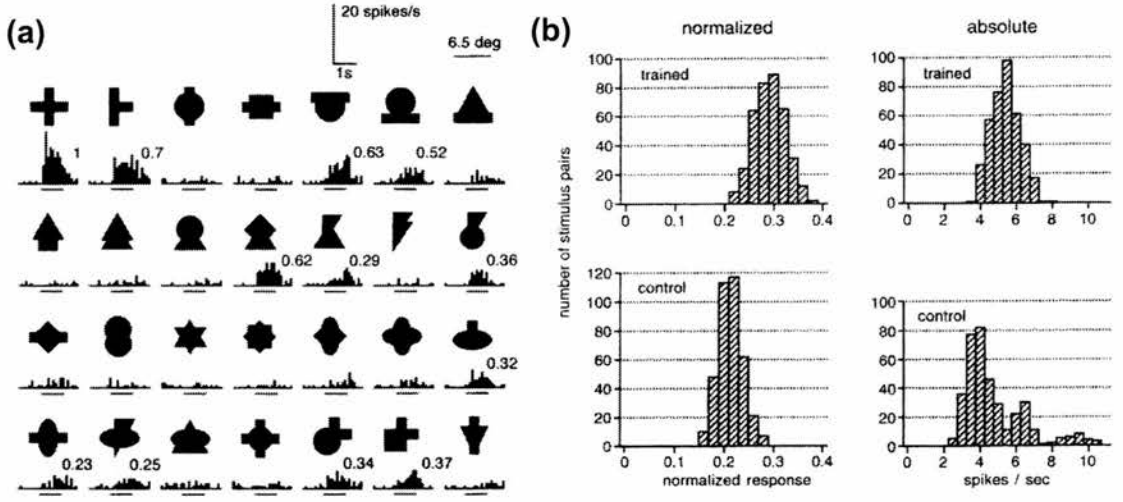


figure 2.15 – Population coding of learned shapes, from Kobatake et al (1998). *(a)* The 28 training stimuli used in the experiment and the response of a single neuron to each shape. The numbers show response magnitudes relative to the maximum for those responses that were statistically significant ($p<0.05$). *(b)* Distributions of the distances between 2 training stimuli in the space spanned by responses in trained (top) and control (bottom) animals. *Left* – distances calculated from responses that were first normalised to the maximum response for a cell. *Right* – distances were calculated from absolute responses. In both cases there is a shift towards larger distances between stimuli after training.

responses in a single cell, suggesting that stimuli were represented on the basis of a population code. The effectiveness with which this code could differentially represent the training set of stimuli was measured by calculating the distances[7] between stimuli in the space spanned by the responses of the TE cells, in both the trained and control animals. This showed that the distances between stimuli in trained animals were consistently larger than in the controls, demonstrating that there were not only more neurons representing these stimuli, but that the responses of these cells were also more effective in spanning the stimulus space (see figure 2.15b).

An earlier study by Logothetis et al. (1995) trained monkeys (again over a period of months) to recognise views of two classes of 3-dimensional novel shapes (wire frame

---

[7] Each stimulus was represented as a point in n-dimensional space, with each axis describing the response to the stimulus by one cell. The distance between any two stimuli was simply the Euclidean distance between the points.

and amoeboid objects). Although no controls were present in this study, a large proportion of TE cells in the trained animals (28.5%) were again found to respond significantly more to these novel objects that to any other visual stimulus tested. It was also noted that the percentage of cells responding to stimuli belonging to a particular class of object was correlated with the amount of training that each animal had received in that class.

Of the cells responding best to target objects, the majority were found to be view-selective, with the response decaying sharply as the object was rotated away from the preferred viewing angle. Different cells showed different preferred viewing angles, often for the same target object. A smaller number of cells had 2 preferred views, spaced 180° apart, and a single cell was found which showed completely view-independent responses (illustrative tuning curves from this study are shown in figure 2.16). This suggests that neurons in TE can build view-invariant responses to *learned* objects in perhaps the same way that has been suggested for view-invariant face responses (Perrett, 1984; Rolls, 1992), by associating the outputs of neurons that respond to different views of the same object. It is not clear from the above studies whether TE neurons built representations for novel objects simply as a result of frequent exposure to these stimuli over a period of time (i.e. a passive mechanism), or if the task-based training itself is vital. Two further studies provide somewhat contradictory views on this point.

Booth and Rolls (1998) allowed monkeys to play with novel 3D objects in their cages over several months, and then recorded from visual neurons in STS and IT, presenting 4 different digitised views of each object. Although the monkeys were not explicitly

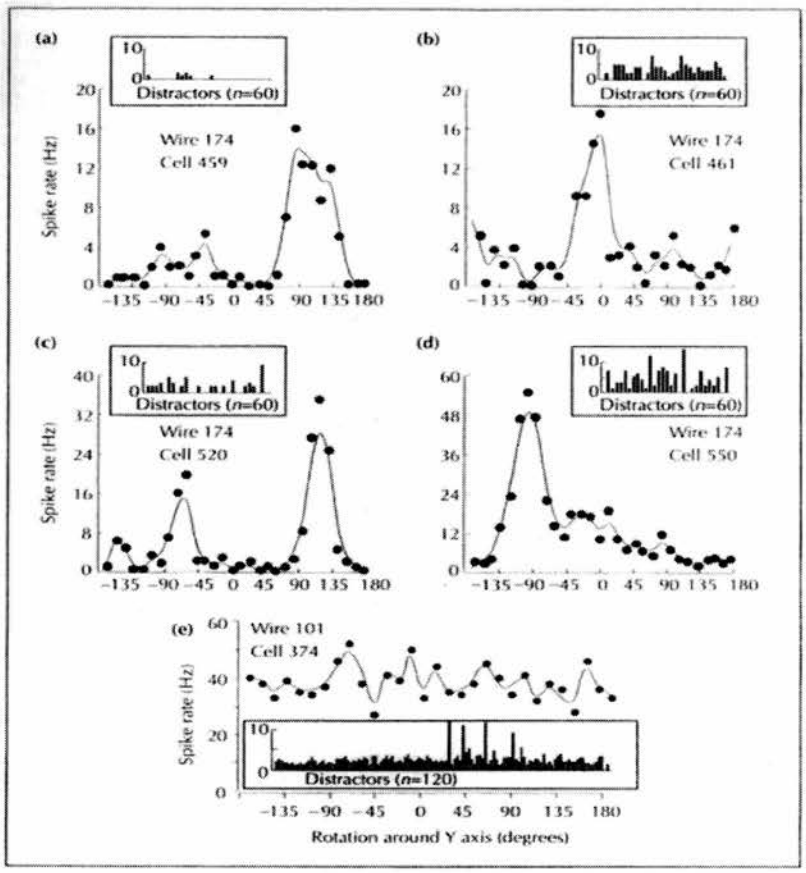


figure 2.16 – Responses to learned novel objects from Logothetis et al (1995). *(a)–(d)* View-dependent responses to learned objects and *(e)* a neuron showing a view-independent response. The responses to the learned objects are considerably greater than to the distractor stimuli in each case.

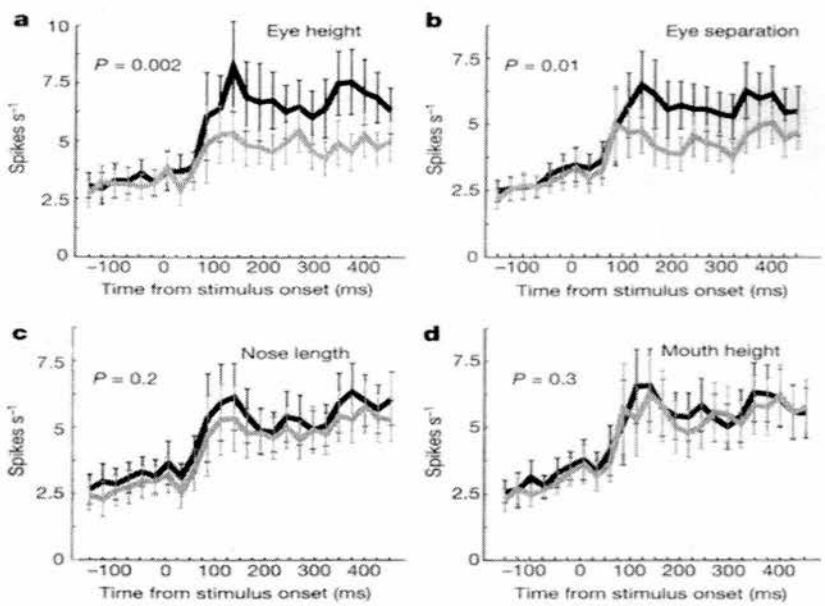


figure 2.17 – Categorisation training influences feature selectivity, from Sigala & Logothetis (2002). Population average response to best (black line) and worst (grey line) values of each of the four stimulus features. *(a)* & *(b)* diagnostic features. *(c)* & *(d)* non-diagnostic features.

trained with the objects, they found 14% of neurons responded in a view-invariant way to one or more of the cage objects, but not to other visual stimuli tested. In common with Kobatake et al. (1998), some neurons were found to respond to more than one object, again suggesting a population code.

In contrast, Sigala & Logothetis (2002) found that the feature selectivity of inferotemporal neurons can be shaped by explicit training in a categorisation task. Monkeys learned to classify sets of stimuli (line drawings of faces and fish), where each set had four variable features, two of which were diagnostic of category, two of which were not. They examined neurons in TE that had significantly different responses across at least one feature of the stimulus set, and discovered that the majority (about 75%) of these cells were tuned to one or both of the diagnostic features, but not to the non-diagnostic features. Population averages for the face set of stimuli (shown in figure 2.17) indicate that neuronal selectivity is shaped by the stimulus features that are *most relevant to the task* that the animal is performing.

The above studies show that a significant proportion of neurons in IT (specifically area TE) are not fixed in their feature selectivity and can be recruited to represent stimulus spaces that are novel and relevant to the adult animal. They can be described as having "tunable receptive fields" (Logothetis et al., 1995), and representation appears to be on the basis of a population code. The features coded by individual neurons appear to range from partial stimulus features (Kobatake et al., 1998; Sigala & Logothetis, 2002) to view-invariant responses to complete objects (Booth & Rolls, 1998), consistent with the pattern of feature-selectivity found more generally in IT neurons.

Table 2.1 – Proportion of neurons showing learning-related selectivity

Study	*Proportion in trained monkeys*	*Proportion in controls*	*Explicit training*	*Stimuli*	*Recording site*
Kobatake et al. (1998)	25%	5%	Yes	Simple geometric shapes	TE
Logothetis et al. (1995)	28.5%	-	Yes	Novel 3D objects	TE
Booth & Rolls (1998)	14%[a]	-	No	Novel 3D objects	STS/TE

[a] criteria was a greater response to all views of object i.e. view-invariant.

Neurons with tunable receptive fields do not appear to be confined to a specific sub-area of TE, since Booth & Rolls (1998) found neurons responding to learned objects from the lower bank of the STS, across the MTG to the AMTS. Cell position was not rigorously reported in the other experiments described, but there is some agreement over the proportion of cells showing plasticity, at around 20% (see table 2.1).

While the timescale over which plasticity occurs is not clear from the experiments described above, some other studies have looked at the effect of repeated exposure to stimuli over the course of a recording session (Rolls et al., 1989; Miller et al., 1991; Li et al. 1993). These have found changes in neuronal response which may be related to the initial phases of a plastic change.

The DMS task has already been described in relation to short-term memory, where it was found that the response to a test stimulus was suppressed if it matched a previously shown sample. However, this suppression has also been investigated (Li et al., 1993) across trials, since the monkey was repeatedly exposed to the same set of stimuli over many trials. It was found that, superimposed on the match suppression described earlier, there was a progressive decline in response in about a third of neurons, as stimuli become more familiar.

This decline can be seen to be greatest when only a small number of trials intervene between repeat presentation of a stimulus (figure 2.18). Furthermore, the response of neurons was found to rebound with the introduction of novel sets of stimuli, suggesting that the decline in response may be due to synaptic depression specific to a stimulus-related subset of the incoming connections.

The recording site in this study was close to the rhinal sulcus, and probably not within area TE proper, so it is not clear whether this occurs in visual neurons throughout IT cortex. An earlier study by Rolls et al. (1989) recorded from face-selective cells in STS and TE, repeatedly presenting a set of novel faces until they became familiar. They found that about a quarter of neurons changed their relative responses to the faces in the set, but in this case the changes that occurred happened very rapidly, largely between the first and second presentations of a stimulus. This effect was not found when already familiar faces were repeatedly presented.

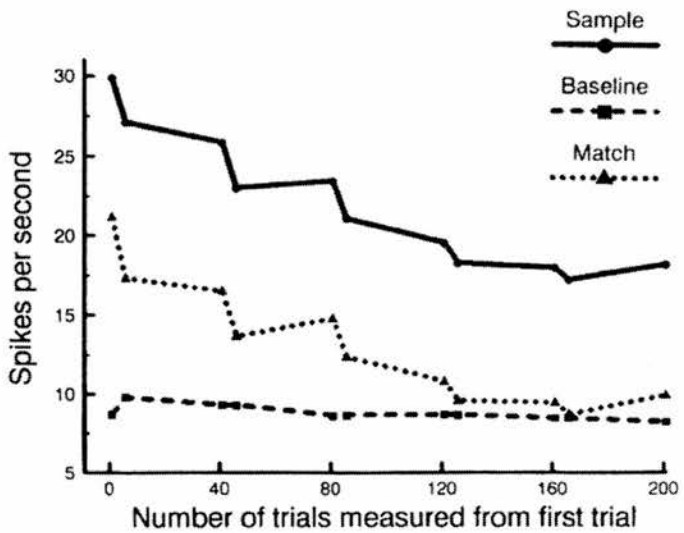


figure 2.18 – Average response of 25 neurons that showed a significant decline in response to a stimulus across DMS trials (from Li et al., 1993). The responses to sample and match stimuli are shown separately. Both show a progressive decline, as stimulus familiarity increases over trials (x-axis). The stimulus was repeated every 3 or 35 trials alternately.

While both these studies show changes in stimulus selectivity with repeated exposure to stimuli, it is not clear whether these effects have a short or a lasting effect on neuronal receptive fields. The changes in selectivity reported by Rolls et al. (1989) are very rapid and may actually be indicative of a form of short term memory, specific to face-selective neurons, which could allow the faces of new individuals to be quickly learnt.

13. At least some visual neurons in IT show experience-dependent plasticity. There may be multiple mechanisms, and time course of this plasticity remains unclear. The task-relevance of stimulus features can influence selectivity.

Finally, we shall consider a form of associative learning that has been extensively studied in IT neurons. In this section we have already seen differing degrees of view independence (figure 2.16) reported by Logothetis et al. (1995) that are suggestive of a hierarchical model of processing and plasticity in TE, with increasing levels of invariance perhaps formed by associating the outputs of the preceding level. Booth & Rolls (1998) suggest that the natural transformations of objects under normal viewing conditions may lead to view-invariant representations, as close temporal activation of specific features causes them to become associated together (see also Földiák, 1991, 1997)[8]. Association learning in IT has been specifically investigated in a number of studies in IT, and this form of learning is discussed in detail in the next section.

---

[8] Though this hypothesis might lead us to expect that invariance should develop by the progressive widening of view tuning curves, rather than via bimodal responses to views 180° apart (as seen in figure 2.5).

2.4.3. Association Learning

A series of studies have investigated the responses of inferotemporal neurons after pair association learning. The basic task is as follows: a set of pictures is sorted into pairs. In each trial, the monkey is presented with a cue stimulus on the monitor. This disappears, and after a short delay, a choice of two stimuli is presented, the paired-associate of the cue, and one from a different pair. The monkey is rewarded for correctly choosing the paired-associate of the cue stimulus. With sufficient trials, monkeys can learn to perform this task to a high level of accuracy.

Sakai & Miyashita (1991) recorded from neurons in anterior IT (recording sites included both TE and PRh) in monkeys that had learned the pair association task described above. During the cue period, neurons showed selective responses to particular pictures, with the majority of cells responding to at least two of the pictures. This is as would be expected, given the results of experiments described in the previous sections. More interesting, however, was the pattern of responses of those neurons that responded to two or more of pictures. In these cells, the maximal and second best responses tended to be evoked by pictures that were part of a pair. These neurons, showing selective cue responses to both pictures of the paired-associates, were termed "pair-coding" neurons, and their behaviour is illustrated in the top part of figure 2.19.

A second class of neuron was also found which, like the pair-coding neurons, had selective responses to particular pictures during the cue period. However, when the paired-associate of a picture was shown as a cue, there was little initial response in these cells. Instead, there was growing and sustained activity during the delay period, until the choice of stimuli was presented. These cells were termed "pair-recall"

neurons, where the paired-associate of a cue-optimal picture gave the highest activity in the delay period (figure 2.19 – bottom).

This study provides strong evidence that inferotemporal visual neurons can acquire their selectivity for patterns through associative learning, though the role of the different types of neurons is not clear. Visual association learning is known to depend on the integrity of limbic structures, since bilateral removal of the hippocampus and amygdala prevents monkeys from relearning the pair association task, as does disruption of the pathway from visual areas to the medial temporal lobe through rhinal cortex (Murray et al., 1993). Subsequent studies have provided more detail about the influence this backwards projection has on inferotemporal neurons.

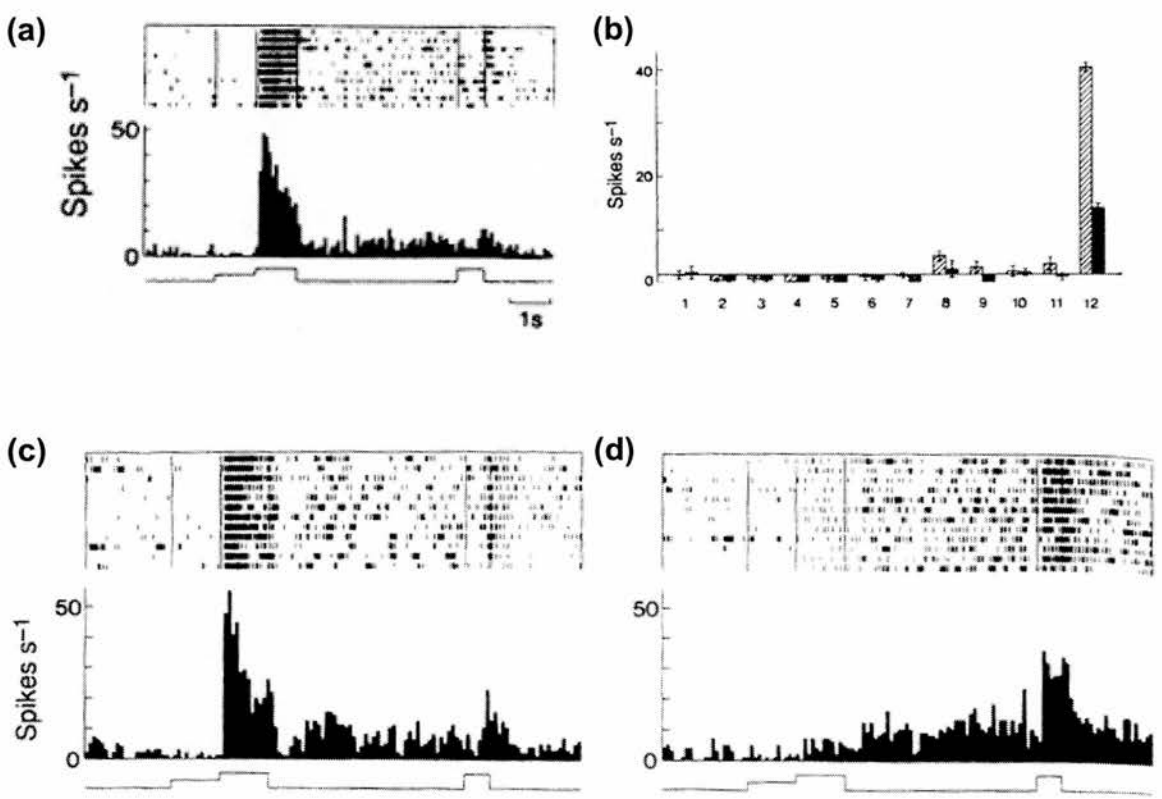


figure 2.19 – From Sakai & Miyashita (1991). *Top* – "Pair coding" neuron. *(a)* Rastergram and SDF showing the response to the stimulus that elicited the highest response during the cue period. *(b)* Mean discharge rates during the cue period for the same neuron, with the response to all 12 pairs of pictures shown (light and dark bars). The maximal and second best responses belong to paired stimuli. *Bottom* – "Pair recall" neuron. *(c)* Trials for the cue which elicited the highest response. (d) Trials for the paired-associate. There is little activity with the cue presentation but growing activity during the delay period.

Higuchi & Miyashita (1996) first transected the anterior commissure (AC), removing the input to anterior IT neurons from the other hemisphere, then trained monkeys in the pair association task with a first set of pictures (set A). After training was complete they measured the response of IT (area TE) neurons as a prelesion control. Then, inferotemporal neurons were disconnected from the backwards projection from the medial temporal lobe, with a unilateral lesion to entorhinal and perirhinal cortex. Monkeys then relearned the task with the original stimuli (set A), and a novel set of images (set B). Recordings were then made from the same area as in the prelesion control.

They found neurons selective to both the set A and the set B stimuli. However, while there was a tendency for neurons to show response correlation to paired-associates in the control recordings, this behaviour had been abolished by the lesion. Paired responses were seen neither for the set of images learned prior to the lesion (set A), nor with those learned after it (set B).

Thus the lesion appears not only to have destroyed the ability of IT neurons to form new associations, but also to have disrupted the associations already formed (i.e. set A). This neuronal retrograde amnesia implies that the pair-encoding properties of visual neurons in inferotemporal cortex are dependent on a backwards projection from rhinal cortex, rather than based on associations formed within IT cortex itself.

Secondly, because highly selective visual responses were seen to set B (even though training with this set occurred after the lesion), this backwards projection is therefore clearly unnecessary for the kind of experience-based plasticity described earlier. This is consistent with a view that learning in inferotemporal neurons can be shaped by

multiple parallel mechanisms, a point which will be discussed further at the end of this section.

Further evidence that TE neurons can be activated by a backwards memory signal from limbic areas has been provided by Naya et al. (2001), who compared the visual latency of responses in areas TE and PRh (area 36) while monkeys performed the pair association task. They found that the latencies of neurons in area 36 were significantly longer than those in area TE, suggesting that the perceptual signal reaches TE neurons first. However, when they looked at the time course of activity to the paired-associate of the cue-optimal stimulus, the reverse was found, with neurons in area 36 responding first, and those in area TE following much later (figure 2.20).

Finally, it is of interest to ask how rapidly visual stimuli become associated in IT. This question has been investigated recently (Messinger et al. 2001), and the results show that changes in selectivity can be seen over the course of an experiment.

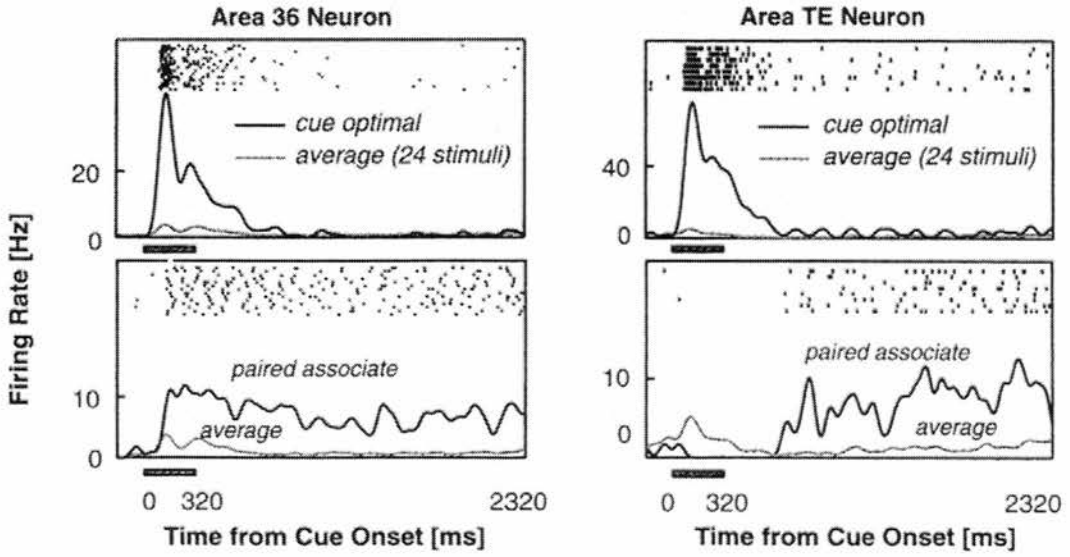


figure 2.20 – From Naya et al. (2001). Rastergrams and SDFs for neurons in area 36 (*left*) and TE (*right*). The response of each neuron to the stimulus that evoked the highest response during the cue period is pictured top. The response to the paired-associate of this cue optimal stimulus is pictured bottom. Although both neurons respond to cue and paired-associate images, activity relating to the paired-associate begins first in area 36.

Recording from both TE and PRh, they found that neurons initially selective to a single stimulus began to respond increasingly to the paired-associate as learning progressed. This is illustrated in figure 2.21.

14. Neurons in TE respond selectively to specific stimuli. They can also respond to other images that are repeatedly paired (i.e. closely temporally associated) with these stimuli. However, responses to paired stimuli occur with higher latency.

2.4.4. Summary

The studies presented in this section reveal that IT neurons do more than simply code for features immediately present in the visual world; their responses and selectivity is also shaped by a number of factors including experience of the immediate and longer-term visual past, working memory, and associations between separate elements of the world that have been appeared together frequently, or even in close temporal proximity.

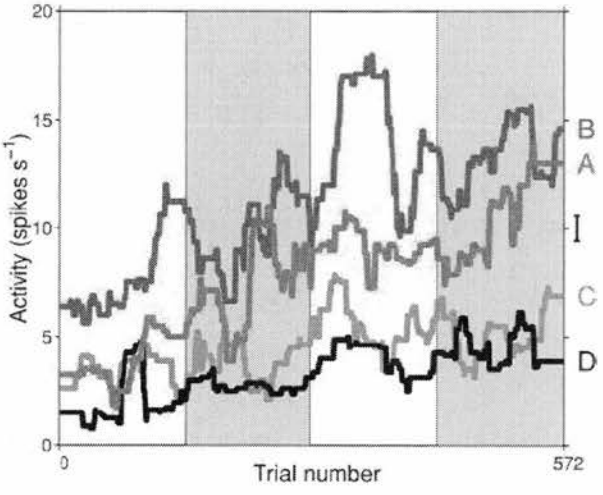


figure 2.21 – Stimulus associations developing in an IT neuron, from Messinger et al. (2001). Monkeys were trained in a pair association task with 2 sets of paired stimuli (A&B, C&D). The neuron pictured was initially selective for a single stimulus, but developed a pair-selective response to stimuli A&B as training with this set of images progressed.

About one-third of IT neurons have been shown to undergo short-term response suppression to repeatedly presented stimuli, and a similar (probably overlapping[9]) proportion also show a longer-term decline in response with increased stimulus familiarity. These effects could clearly play a role in the outcome of stimulus optimisation experiments, where a large number of variations on a stimulus might need to be shown in order to explore a region of stimulus-space. It would therefore be of great interest to know whether this decline in responsiveness is correlated with stimulus similarity, or if it requires the stimuli be completely identical.

The effects relating to association learning and working memory would be expected to cause little interference with optimisation experiments, though these studies do make it clear that optimisation would provide only a partial, and not complete, description of the properties of these cells.

Finally, we have also seen that TE neurons can undergo plasticity in their receptive fields, probably to allow the animal to represent novel objects or faces. Although the studies carried out have been fairly consistent in the proportion of cells reported to encode aspects of a new stimulus space (i.e. around 20%), it is also possible that all TE neurons exhibit plasticity in the adult animal, at least to some degree. If a sparse-distributed code (see Rolls et al., 1989) is used to represent the visual world in IT, both increases and (more often) decreases in response might be observed over the course of an optimisation experiment, as a new population code is established to

---

[9] Li et al. (1993) note that neurons showing a decline in responsiveness with increasing stimulus familiarity account for most of the match suppression described in their earlier study (Miller et al., 1991).

represent the feature space. However, the timescale over which we can expect these kind of changes to occur (and the importance of the behavioural relevance of the task) remains far from clear from the studies carried out to date.

2.5. Guidelines for Optimisation

This chapter has reviewed the known properties of cells in IT, and extracted a set of guidelines which should be taken into account in the design of an optimisation algorithm for use in IT cortex. To review the main points –

1. The responses of TE neurons are non-linear.

2. The response of a TE neuron is increases gradually, as a stimulus becomes closer to an optimal set of features.

3. TE neurons are selective for the orientation, and often colour, of a stimulus, but can show invariance for size, position and stimulus boundary definition. There may be optimal settings for these parameters which produce the greatest response.

4. If a cell responds to an object (e.g. a head) irrespective of view, multiple runs or multiple solutions would be required for optimisation to produce an adequate description of the cell.

5. Knowledge of population codes in IT may allow optimisation to be restricted to a more limited image subspace, improving the chances of success.

6. A technique relying on the manipulation of two-dimensional images is more likely to find success in TE and TEO. Investigations of STS may require a methodology based manipulation of three-dimensional objects.

7. A response can outlast stimulus presentation time by hundreds of milliseconds. Response persistence can be cut short by subsequent stimuli, whether or not these stimuli are themselves effective in driving the cell.

8. SOA, rather than stimulus duration, affects the response magnitude and duration.

9. Response duration under RSVP conditions equals SOA + 60ms.

10. The majority of stimulus information is carried by the spike count, rather than in temporal patterns in firing.

11. Response latency increases in STS and IT as stimulus contrast is reduced. This can result in latency shifts of hundreds of milliseconds.

12. Habituation can occur to visual stimuli. Working memory can potentiate responses to relevant stimuli. Both effects are weak.

13. Some visual neurons in IT show experience-dependent plasticity. The task-relevance of stimulus features can also influence selectivity.

14. Neurons in TE respond selectively to specific stimuli. They can also begin to respond to other images that are repeatedly paired with these stimuli. These responses to paired stimuli occur with higher latency.

3. Systematic Methods for Investigating Receptive Fields

The studies discussed in the previous chapter have revealed many interesting properties of neurons in STS and IT. However, the experimental methods used to study these cells rely on the assumptions of investigators, and make use of a limited and predetermined set of stimuli. Thus we can say that a given cell responds better to a picture of a face *compared to the other stimuli tested*, but we cannot say with any certainty what the actual trigger features of the cell are, or indeed be sure that there is not a different stimulus that would produce an even greater response. This chapter reviews some of the more systematic methods that have be used to study the receptive fields of sensory neurons, both earlier in the visual system, and elsewhere in the brain.

3.1. Stimulus Optimisation Studies

The space of possible sensory inputs is enormously large, which makes determining the trigger features of neurons a very difficult task. For example, consider a static image presented on a computer screen to a visual neuron. If the screen is 64 pixels wide by 64 pixels tall, and each pixel can be one of sixteen grey levels (ignoring, for now, the additional complexity of colour), there are $16^{4096} \cong 1.2 \times 10^{4932}$ possible different images.

How can we hope to find the optimal stimulus in such an enormous space? Stimulus optimisation studies have made use of the assumption that the response is some continuous function of the input – thus we can make an informed guess at what might be a better stimulus from the known responses to stimuli we have already presented.

3.1.1. The Neuronal Response Function

In the visual modality, we can imagine an image with n pixels (or n degrees of freedom) to be a point in an n-dimensional space, with every possible image occupying a unique location in this space. The neuronal response function maps this space of possible stimuli to a scalar response, generally the firing rate in a given interval.

If we consider an image with only two degrees of freedom (i.e. a two pixel image), we can picture the neuronal response function as a three-dimensional surface, where the height of the surface above the stimulus plane reflects the firing rate of the neuron to a particular two-pixel image (see figure 3.1).

As additional pixels are added to the image, new axes are added to the stimulus space, and the response function becomes a hypersurface in n+1 dimensions, where n is the number of free parameters in the image.

While we have assumed that each parameter of the stimulus represents the luminance of an individual pixel, this need not be the case. A variety of basis functions can be

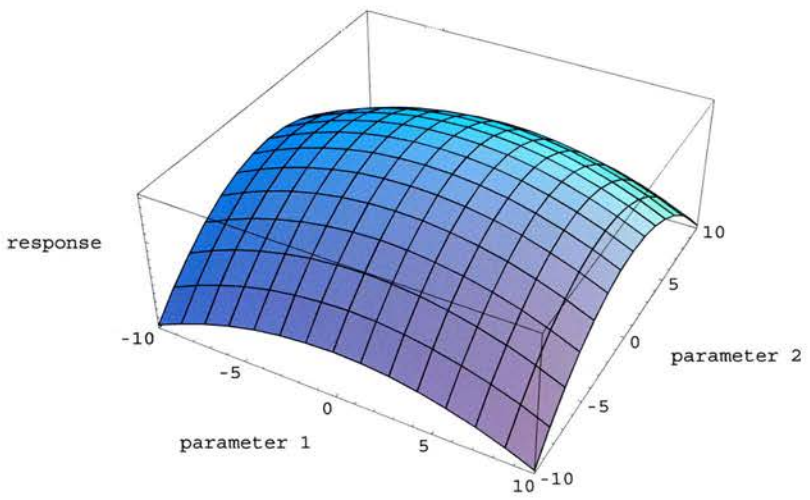


figure 3.1 – hypothetical neuronal response surface for a stimulus described by two parameters. Each point on the plane describes an individual stimulus, with the height of the surface above the plane reflecting the response of the neuron to a particular stimulus.

used to represent images, such as sines and cosines of varying frequency (Fourier transform) or wavelet basis functions. However, if the new basis is an orthonormal transform of the original, as in the examples above, then the general topography of the neuronal response surface will be maintained, since orthonormal transforms are equivalent to rotations and/ or translations of the axes in n-dimensions.

Assumptions regarding the neuronal response function –

(i) The existence of a neuronal response function, as described, implies that the response of the neuron is instantaneous (i.e. it is not affected by a memory of past stimuli). Results in the previous chapter indicate that this may not always be true for IT neurons, but this simplifying assumption is necessary for optimisation to be a possibility.

(ii) If we are to explore the surface in any systematic manner, it is also necessary that the surface is continuous along the chosen axes. While we have seen in the previous chapter that this is the case with axes chosen arbitrarily by the experimenter, this may not hold using the more traditional bases used to represent images, such as a pixel basis, or some orthonormal transform of this basis.

(iii) Finally, we must assume that there is not a large number of local maxima present on the surface which may prevent a search technique from finding the true maximum.

3.1.2. Introduction to Optimisation Methods

Optimisation algorithms are designed to find the maximum (or minimum) of a function which depends on one or more variables, generally with the assumption that,

for reasons of computational effort, we wish to evaluate the function as few times as possible. This is particularly true when we are trying to optimise the response of a neuron, since there is a limit to the length of time over which a stable recording from a cell can be obtained, and therefore a limit to the number of potential stimuli we can test.

Optimisation is a large field and a number of different algorithms exist (for a review, see Press et al., 1988). There is no perfect algorithm – the best choice depends on the nature of the function to be optimised, and on the information available. It is instructive, however, to see why one of the most widely applicable algorithms for optimisation, the method of steepest ascent[10], is not appropriate for use with visual responses.

The method of steepest ascent evaluates the gradient (a vector, in the multidimensional case) by measuring the change in response from the current location to a series of small perturbations directed along each axis. It can then proceed directly up the gradient to a new point where the response is better. Because the method takes a series of small steps towards an optimum, this acts as a filtering mechanism that ameliorates the effect of noisy gradient estimations.

However, obtaining the complete gradient vector at each location required by steepest ascent is impractical with a high-dimensional stimulus. For example, with the 64x64 image described earlier, we would need to test 4096 images in order to determine the full gradient vector at each location.

---

[10] The steepest descent form of this algorithm is discussed in detail in Widrow (1985).

Despite this, it is possible to obtain a noisy, or low-resolution, estimate of the gradient simply by testing the response to a single image. Although it is now no longer possible to proceed *directly* uphill along the steepest gradient, we can still move in a general uphill direction. The dynamics of this strategy have been investigated in detail using quadratic performance surfaces by Widrow (1985). This analysis shows that, convergence is faster if function evaluations are used to take a large number of steps based on coarse gradient information, rather than using function evaluations to obtain a detailed description of the gradient, and taking a lower number of steps. This is especially true where the dimensionality of the search is high, and this strategy is used by the ALOPEX algorithm, discussed below.

3.1.3. ALOPEX

The ALOPEX[11] technique (Harth & Tzanakou, 1974; Tzanakou et al., 1979) was the first attempt to apply optimisation methods to the characterisation of sensory receptive fields.

Relatively low resolution images (up to 32x32 pixels) were presented to cells in the frog visual tectum, and the images were updated based on response feedback from the cell. The algorithm, which implements hill climbing in multidimensions, is described in detail in box 3.1. Noise is added to an underlying image (initially of uniform luminance) then a response to that noisy image is obtained. The visual noise is then correlated with the change in response, and the underlying image updated accordingly. Essentially, ALOPEX makes a estimate of the gradient along a randomly chosen direction, then attempts to move up that gradient. The algorithm itself can be

[11] An acronym for ALgorithm Of Pattern EXtraction

compared to the method of linear random search (LRS), for which a proof of convergence, and rate estimations for a quadratic performance surface, can be found in Widrow (1985).

Computer simulations using the algorithm (Harth & Tzanakou, 1974) demonstrated that the method was capable of characterising not just simple linear receptive fields,

Box 3.1 – The ALOPEX Algorithm

The screen is divided into N×N elements (pixels) with intensitities $i_1 \ldots i_{N^2}$

The stimulus at iteration n can be viewed as a vector of pixel intensities –

$$I(n) = \begin{bmatrix} i_1(n) \\ \vdots \\ i_{N^2}(n) \end{bmatrix}$$

$$\text{with } i_j(n) = b_j(n) + r_j(n)$$

where $b_j(n)$ is the pixel bias, and $r_j(n)$ is a random variable drawn from a normal distribution with zero mean.

Each image, $I(n)$, produces a scalar response, $R(n)$, from the system.

Every iteration, for each pixel, biases are updated by cross-correlating the change in intensity with the change in the response –

$$\Delta b_j(n) = \mu \Delta R(n-1) \Delta i_j(n-1)$$

where $\Delta R(n-1) = R(n-1) - R(n-2)$, $\Delta i_j(n-1) = i_j(n-1) - i_j(n-2)$ and μ is a learning rate parameter.

Thus, the bias (or underlying intensity) of each pixel is raised if, in the preceding two iterations, the intensity of that element and the response changed in the same direction. Otherwise the bias is reduced.

but also the receptive fields of non-linear model complex cells, with multiple runs producing alternative optimal stimuli.

Optimisation runs on real visual cells produced images that showed a high degree of correlation with receptive fields characterised using scan techniques, as illustrated in figure 3.2 (the scan method is applicable in the case of linear tectal cells and is similar to reverse correlation, described in a later section). However, a major criticism of the ALOPEX experiments was that they failed to reveal anything new about the cells under investigation.

A problem with the technique lies in the limit to spatial resolution. As the number of pixels is increased, the spectrum of the input becomes more uniformly distributed over the Fourier plane, leading to a decrease in the power falling within any given part of the cell's receptive field. An approach to this problem is to use different basis functions to represent the image, and restrict the search to a subspace of the possible inputs. This approach is developed in the following chapter.

3.1.4. Simplex in Auditory Cortex

ALOPEX makes use of approximate gradient computations to search for peaks in the neuronal response surface. However, other optimisation algorithms exist which do not

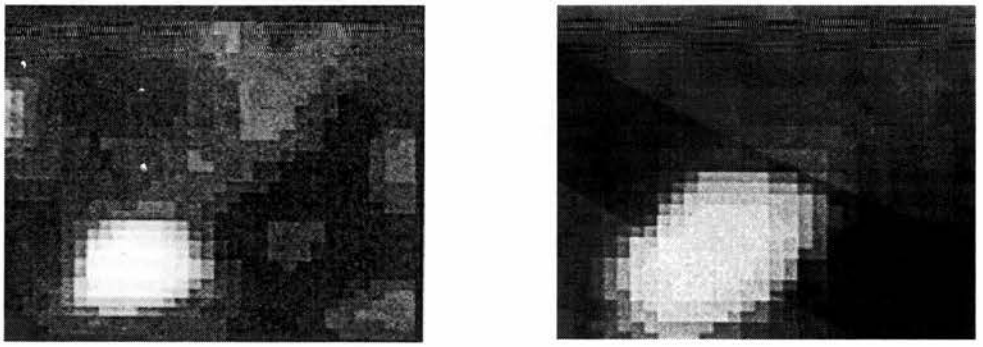


figure 3.2 – *(left)* An averaged result from 3 ALOPEX runs on a visual tectal cell. (right) The receptive field of the same cell as characterised by a scanning technique. Both images have been spatially smoothed. [from Tzanakou et al., 1979].

require explicit calculation of the gradient, and which can behave more robustly when the function values are noisy. One such algorithm is simplex (Press et al., 1988; see box 3.2), which was used by Nelken et al. (1994) to optimise high dimensional[12] stimuli in cat auditory cortex.

Primary auditory cortex has a tonotopic organisation, with a frequency gradient in one direction, with biaural interaction bands (where ipsilateral input suppresses or enhances the response to contralateral stimulation) orthogonal to it. Despite this, many cells are unresponsive to simple tones, responding strongly only to complex stimuli of various kinds.

The stimuli space searched by Nelken et al. (1994) consisted of single tones bursts (1-D stimuli), two tone bursts (2-D stimuli), four tone bursts (4-D stimuli) and nine tone bursts (characterised by centre and spacing parameters, and therefore 2-D). All the multiple tone bursts consisted of the tones presented simultaneously. The response was measured from population activity consisting of several cells, rather than a single unit.

The optimisation runs met with moderate success in that many of the stimuli found by the process were much more effective than the starting stimuli. Figure 3.3 (left) illustrates two search runs with different outcomes – the first leading to more effective stimuli, the second failing to improve from the starting point.

---

[12] though the stimuli in this study are very low dimensional compared to the visual stimuli used by ALOPEX

It is also interesting to note from the successful run illustrated that the method finds a local maximum relatively early in the process, with subsequent steps only serving to move the other points of the simplex closer to the maximum that has already been found.

Multiple optimisation runs were performed on many cells (figure 3.3, right) to assess the consistency of convergence points. The authors note that, although the runs produced apparently widely different results, there is some evidence that the optimised tones tended to cluster around certain frequencies in each cell, suggesting that combinations of specific tones may be the relevant features for these cells.

Unlike the ALOPEX studies, where receptive fields studied had already been characterised by simpler methods, Nelken et al. (1994) obtained novel results in auditory cortex. However, in the higher dimensionality of visual space, the additional

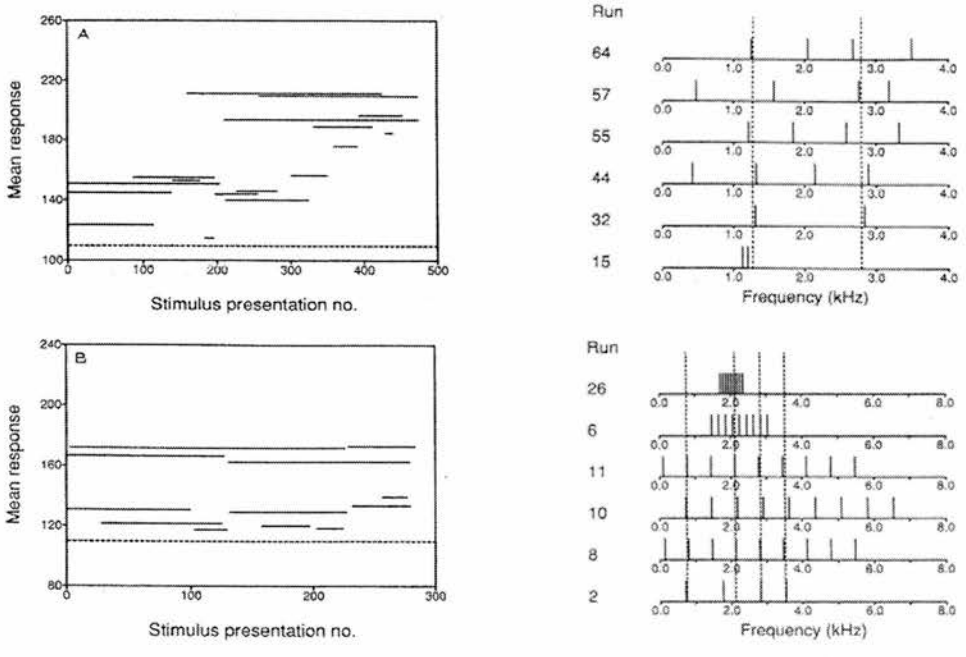


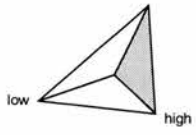
figure 3.3 – *(left)* Two example search runs. In the successful top run, the stimuli were nine-tone bursts (with two variable parameters). The responses to each of the three stimuli described by the vertices of the simplex can be seen at each time step. The lower unsuccessful run used two-tone bursts. *(right)* Multiple runs in two cells. Although there was considerable variation in the optimised stimuli, the dotted lines can be seen to indicate tones which were common to many results. [from Nelken et al., 1994].

information provided by a measure of the gradient (even if noisy and approximate) is likely to improve the likelihood of convergence.

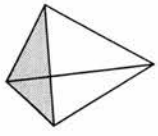
Box 3.2 – The Simplex Method

A simplex is a geometrical figure consisting of, in N dimensions, N+1 vertices, and all interconnecting lines, faces, etc. In three dimensions, a simplex is a tetrahedron. The function is evaluated at each vertex, with the algorithm attempting to choose new vertices that will take the simplex to the maximum (or minimum) of the function, while maintaining non-degeneracy.

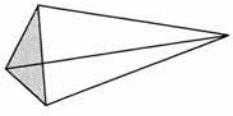
1 *Start of round.* The function is evaluated at each of the vertices of the simplex.



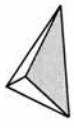
2 *Reflection.* The smallest vertex is reflected to the opposite side of the hyperplane defined by the other vertices, replacing the old vertex if better. If the new vertex is better than the previous best, proceed to step 3. If the new vertex is better than the second worst vertex, this is the end of the round (the direction of the next reflection will be different). Otherwise, proceed to step 4, since (at best) we have simply moved the worst vertex.



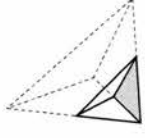
3 *Expansion.* A new vertex at twice the distance is generated, replacing the old vertex, if better. Since a better vertex was found by moving along the direction in step 2, an even better point may be found by moving further. This is the end of the round.



4 *Contraction.* A new vertex is generated at half the distance between the smallest vertex and the hyperplane defined by the other vertices, replacing the old vertex, if better. If the old vertex is replaced, this is the end of the round, since the simplex has been compressed towards the higher values of the function that lie close to the hyperplane. Otherwise, proceed to step 5.



5 *Multiple Contraction.* Each vertex, except the largest, is replaced by a new vertex at half the distance to the maximum vertex. This occurs when the smallest vertex could not be improved by any of the previous steps, useful near a local maximum.



3.2. Other Systematic Methods

3.2.1. Reverse Correlation

The reverse correlation technique was first developed by Jones & Palmer (1987) to quantify the 2-D spatial structure of simple cell receptive fields. The method involved presenting a rapid continuous sequence of bright and dark spots randomly positioned on a 16x16 grid (making a total of 512 distinct stimuli), with each spot displayed for a few tens of milliseconds (generally 50ms). Spikes were recorded throughout the presentation sequence, and whenever a spike occurred, the index of the stimulus currently on-screen was recorded.

When data acquisition was completed, a process termed "reverse correlation" was carried out to determine the set of stimuli that were present when spikes were emitted from the neuron. Two 2-D arrays were maintained, one for bright stimuli and one for dark stimuli. For each spike, the stimulus displayed was resolved into its x, y and contrast components, and +1 was added to the corresponding (x,y) position of the appropriate array. This process is illustrated in figure 3.4.

The 2-D spatial response profile of the cell is then obtained by subtracting the response to dark stimuli from the response to bright stimuli. This was necessary because simple cells have little or no spontaneous activity, so the inhibitory effect of bright stimuli must be estimated in this manner. An example of the bright and dark responses, and the spatial response profile given by their subtraction is shown in figure 3.4 (right).

Reverse correlation also allows the evolution of the receptive field profile across time to be studied, since it is possible to correlate spikes with the stimuli that precede them

by varying time intervals. However, in this study, the relatively long duration (50ms) of stimulus presentation makes this of limited value.

This form of reverse correlation can only be used to study cells that significantly change their probability of firing based on the presence or absence of single small spots of light. This clearly limits the value of the technique with cells later in the visual system. However, a technique has recently been described (Ringach et al., 1997) which makes use of different image basis functions, allowing them to be matched to the properties of the cell under study. This allows the technique to be applied to cells that do not respond to the presentation of single pixels, but do fire in response to e.g. sinusoidal gratings.

Reverse correlation can only be used to characterise fully cells that have a linear (or close to linear) spatial response, such as the simple cells in this study. Because single elements of the receptive field are stimulated in isolation, there is no scope for

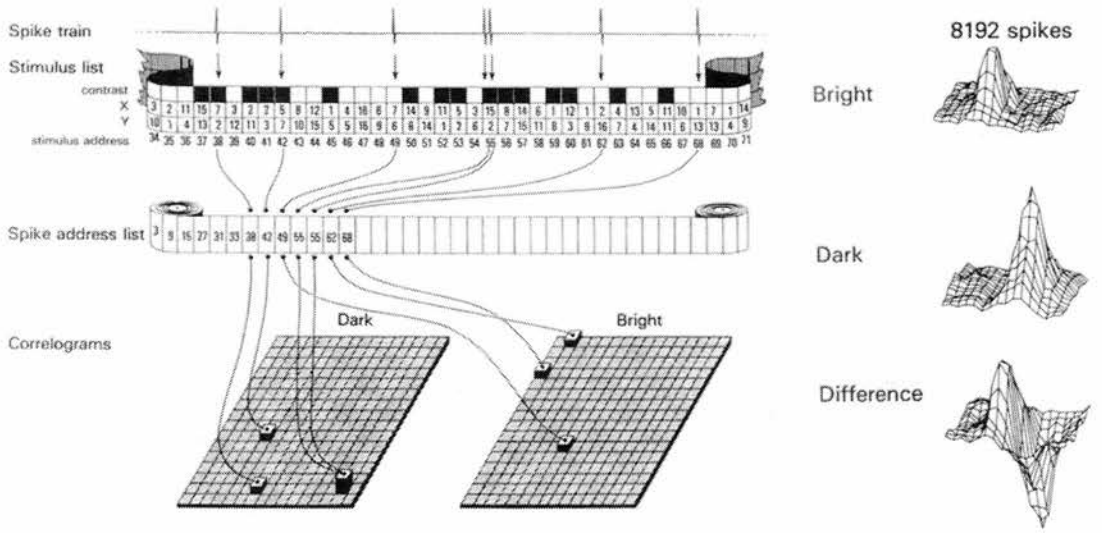


figure 3.4 – *(left)* The reverse correlation procedure. Stimuli are encoded in a randomised presentation list. When spikes occur, the stimulus address is recorded in a separate list. Two separate arrays (for bright and dark stimuli) correlate spikes with the locations of the stimuli that produced them. *(right)* The arrays after 8192 spikes had been recorded from a simple cell. The lower figure shows the gabor-like spatial response profile obtained by subtracting the dark from the bright array. [from Jones & Palmer, 1987].

evaluating the spatial interactions that may occur when multiple regions of the receptive field are stimulated simultaneously. The method can therefore be considered to be a subset of the more powerful Gaussian white noise approach, which is able to characterise these receptive field non-linearities, and is discussed in the following section.

3.2.2. Gaussian White Noise Analysis

White noise analysis is a general tool for characterising the input/output behaviour of linear and non-linear systems, which has been applied to the mapping of visual receptive fields (for reviews see Marmarelis & Marmarelis, 1978; Sakai, 1992; DeAngelis et al., 1995). The receptive field of a cell is covered with a square grid, and the luminance of individual pixels is modulated by independent Gaussian white noise processes. The output of the cell is then cross-correlated with the stimulus sequence to obtain the first-order cross-correlation, which completely characterises the linear response of the cell. This process is equivalent to the reverse correlation technique described above, and produces three dimensional entities describing the response across spatial location and time.

However, the power of the technique lies in its ability to characterise non-linear neurons, such as complex cells, where the response depends on interactions between stimuli at different positions (or times). Unlike the reverse correlation technique, where the luminance of each pixel is varied in isolation, white noise analysis varies the luminance of every pixel simultaneously. This allows higher order cross-correlations to be computed, such as the second order cross-correlation between the luminance of two individual pixels and the response.

Unlike optimisation methods, it is not necessary to assume that the response of the cell under study is instantaneous (i.e. memoryless) because correlations can be evaluated between the response and the presence of pixels with varying time delays. These high-dimensional correlations are hard to visualise and interpret but have proved useful for evaluating models of complex cells (e.g. Jacobson et al., 1993).

White noise analysis tests random locations in image space, rather than concentrating the search on images that are likely to produce high responses. This means that, while the technique is very powerful in fully characterising the first- and second-order correlation functions of cells earlier in the visual system, it would be of little value in a later area such as IT. This is because IT neurons have such complex trigger features that responses to an enormous number of stimuli would have to be collected in order to characterise these very high order correlations.

3.2.3. Modelling Techniques

An alternative technique for characterising the spatial response of a cell is to construct a model that attempts to reproduce the non-linear input/output relationships of the cell. This model can then be inverted to find the optimal stimulus for the model and, ideally, also for the cell.

This method was first applied by Lehky et al. (1992) to the study of complex cells in V1. First they presented a bank of four hundred stimuli to each cell, and measured the average response to individual stimuli over several presentations. The stimuli included simple patterns, such as sinusoidal gratings, gabor functions, and oriented bars, along with more complex patterns including textures and 3-D synthetic surfaces.

Each cell was then separately modelled with a 3-layer neural network (see figure 3.5),

using the backpropagation algorithm (Rumelhart et al., 1986) to train the network to produce the correct response for each of the input patterns. The input layer sampled images with a large number of on- and off-centre units, arranged into two spatially superimposed hexagonal arrays, designed to replicate the organisation of cells in the retinal and lateral geniculate nucleus (LGN). The network had a single output unit, which was trained against a normalised time-windowed cell response. The number of hidden units was varied in order to best fit the data, while preserving the ability of the network to generalise.

The networks were trained using a subset of the patterns presented to the real cells, with the remainder reserved to test the ability of a trained network to generalise its response to novel patterns. Networks with sixteen hidden units were found to be able to capture the input/output relationships well, with a correlation of 0.95 between network output and the response of the actual cell to the training patterns.

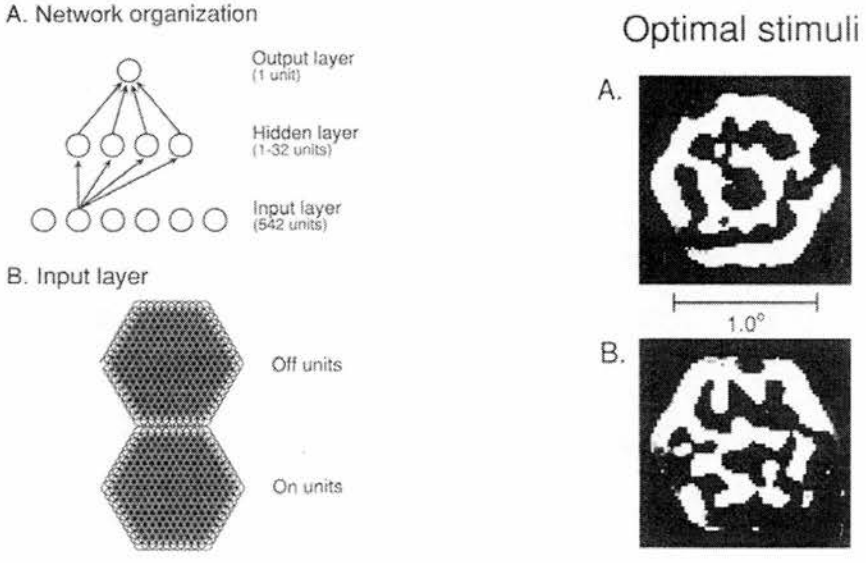


figure 3.5 – *(left)* Organisation of the 3-layer feedforward network used to model complex cell input/output relations. Each layer was fully connected to the next layer. The input layer consisted of 542 units, divided into parallel (and spatially superimposed) sets of on- and off-centre units, both with antagonistic centre/surround receptive fields. *(right)* Two examples of 'optimal' stimuli, obtained by inverting networks after training. [from Lehky et al., 1992].

The predictive power of the networks was found to be somewhat lower, with correlations in the range 0.4 – 0.94 between the network response to untrained stimuli and the recorded responses from cells. While the fit to training data could be improved by increasing the number of hidden units, the predictive power of the network was not similarly improved, suggesting that the increased degrees of freedom in the model with more than sixteen hidden units caused it to over-fit the data. The optimal stimulus for each network was then obtained by inverting the network (Linden & Kinderman, 1990) and typical examples of the optimal input patterns are shown on the right hand side of figure 3.5. Sadly, these optimal patterns were never tested on the real cells, since the training and inversion procedures took sufficiently long to make this impractical.

This is an interesting technique, but its usefulness in studying cells further along the visual stream is questionable, primarily because each step away from the retina adds many more degrees of freedom and any model would require ever larger numbers of input/output pairs in order to constrain it adequately. Replacing the antagonistic centre-surround input units with simple cell-like gabor filters has the potential to improve the power of technique, certainly in the study of the complex fields found in V1 (which are assumed to be a non-linear combination of a number of simple cell inputs), but this is unlikely to extend the reach of the technique to areas much further up the visual system.

3.3. Summary

The space of sensory stimuli is enormously large. Optimisation methods make certain simplifying assumptions about the neuronal response function in order to systematically search locations in the space where the response is increasing. It is

better to use stimulus presentations to take many small steps, based on a rough estimate of the gradient, rather than use them to gather more accurate gradient information at a single point. However, attempts at stimulus optimisation using real neurons have met with mixed success.

The SIMPLEX method (used in auditory cortex) does not use gradient information and is likely to be of limited use in visual optimisation experiments. This is because the first round requires N+1 stimulus presentations (where N is the dimensionality of the stimulus space – potentially a large number in visual experiments) before any optimisation takes place. Each subsequent round requires the response to a previous stimulus to be known before the next can be generated and presented. In contrast, the ALOPEX method makes use of gradient information, and is easily adapted to a paradigm where multiple stimuli are generated and presented at once, with optimisation rounds taking place in between. This may be useful because it allows a large amount of information to be obtained when an animal is fixating, with optimisation and (potentially lengthy) stimulus generation calculations taking place between fixation periods.

Gaussian white noise analysis generates stimuli distributed randomly in stimulus space, correlating them against the response. It makes fewer assumptions about the neuronal response function, and can characterise linear, non-linear and time-dependent aspects of the response. It has proved useful in characterising cells early in the visual system, such as simple and complex cells. Neural networks can also be used to model the response of non-linear cells, then inverted to produce an optimal input. However, since neither technique concentrates on areas of stimulus space where the response is high, both procedures would require enormous numbers of stimuli to fully

characterise cells in higher visual areas such as IT.

It is clear from the studies reviewed in this, and the preceding chapter, that a method suitable for systematically characterising receptive fields in IT should have the following properties –

(i) It should be able to characterise non-linear fields

(ii) It should explore areas of the stimulus space where the response is greatest

(iii) Images should be represented using basis functions that are related to the properties of preceding visual areas

An algorithm designed to possess these properties is presented in the following chapter.

4. Computerised Stimulus Optimisation for Visual Cortical Neurons

4.1. Introduction

An understanding of the stimulus selectivities of visual neurons is important, if we are to understand the representations and transformations carried out by the visual system. While the receptive field properties of neurons early in the visual system are fairly well understood (e.g. in the LGN and V1), this is certainly not the case for later visual areas, such as V4, STS and IT.

Effective stimuli for these neurons have traditionally been characterised manually, but these methods are restricted by a limited and predetermined set of stimuli, based on the assumptions of the investigators. In the previous chapter, we reviewed some of the more systematic methods for characterising sensory neurons, and introduced the idea of a neuronal response surface that may allow us to search for effective stimuli in a systematic manner.

This chapter presents a novel computational method for determining effective stimuli for visual neurons. The system is inspired by ALOPEX (Harth & Tzanakou, 1974; Tzanakou et al., 1979), similar to the LRS method of search described by Widrow (1985). This chapter presents the design of the system, which takes into account considerations based on some of the properties of cells in STS and IT discussed earlier. Results that have been obtained with artificial hardware and software simulated neurons are also presented in this chapter.

4.2. System Design

4.2.1. General Overview

This section introduces a system designed to search for optimal stimuli for visual neurons. The system has a closed-loop architecture (see figure 4.1) such that the generation of stimuli is not predetermined, but is instead guided by the responses of the cell under study. At any point during an experiment, the location in the search space is described by a set of parameters, which represent the current (or "base") image. Each stimulus is generated by adding noise to the parameters prior to their transformation into an image. When a response to this stimulus has been recorded from the cell, the parameters are updated by correlating the added noise with the resultant change in response. Thus, the characteristics of the stimuli generated evolve over time, as the system performs an approximate gradient ascent in stimulus space.

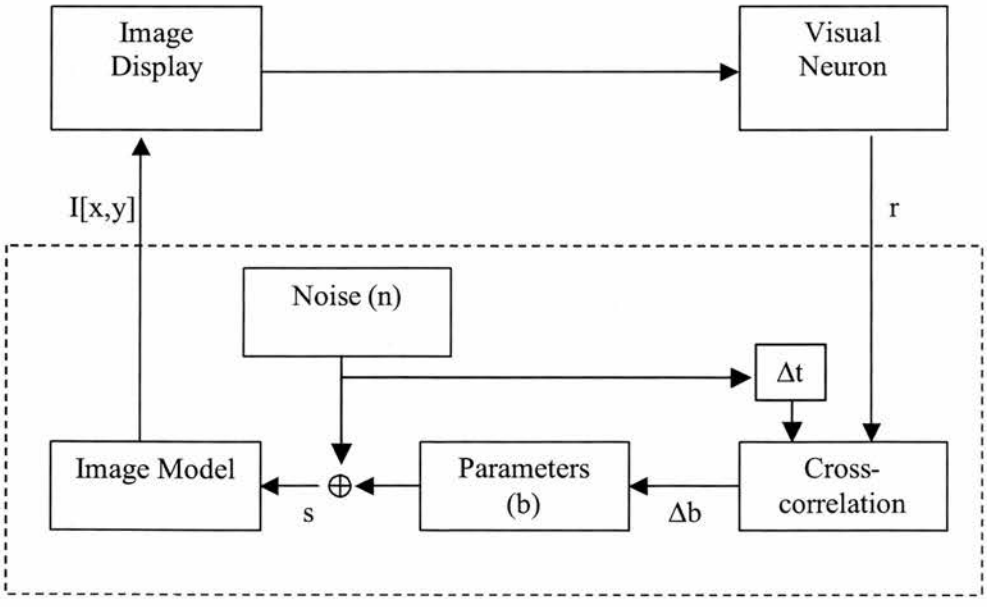


figure 4.1 – Closed-loop design of the stimulus optimisation system presented in this chapter. Symbols are defined in the text.

4.2.2. Correlation Algorithm

The search algorithm presented here optimises a set of parameters which represent an image. The actual representation of the image does not affect the operation of the search, and is therefore discussed later. However, for the purposes of this description, it may be useful to imagine that the parameters simply represent the intensities of each of the pixels that comprise the image.

If, after i iterations of the search, the current underlying (or "base") parameter values are denoted by the vector quantity $b^{(i)}$, then a stimulus $s^{(i)}$ is generated by perturbing the parameters by a quantity of noise, $n^{(i)}$ –

$$s^{(i)} = b^{(i)} + n^{(i)} \tag{1}$$

where each element of the noise vector $n^{(i)}$ is drawn from a Gaussian distribution with zero mean and standard deviation σ.

The stimulus parameters, $s^{(i)}$, are then transformed into an image, with the specifics of the transformation being dependant on the image model in use. This image is presented to the cell under study, resulting in a scalar response, $r^{(i)}$. The base parameter values are then updated by correlating the response with the added noise vector –

$$b^{(i+1)} = b^{(i)} + \alpha(r^{(i)} - \bar{r})n^{(i)} \tag{2}$$

where α is a small constant determining the learning rate, and $\bar{r}$ is the average response over the previous few iterations. After this correlation has been evaluated, and the base parameters updated, the stimulus for the next iteration is then generated

by adding a new set of noise to the new base parameters, as in (1).

The result of equations (1) and (2) can be summarised as follows – equation (1) first chooses a random direction in parameter space to test at each iteration, equation (2) then performs gradient ascent along the randomly chosen path, moving the parameters in the direction chosen if the response has increased, and in the opposite direction if the response falls.

This algorithm performs gradient ascent, but rather than fully evaluate the gradient at each point (which would require at least as many stimulus presentations as there were parameters), it instead takes many small steps, making a rough estimate of the gradient each time[13].

Note that the change in response on each iteration is evaluated with respect to a short term average over past iterations, $\bar{r}$. This average will tend towards the response to the base parameters (as $\alpha \to 0$ & $\sigma \to 0$), because each measured response is drawn from randomly distributed positions centred on this location.

4.2.3. Design Constraints

(a) Block Design

Updating the parameters after every stimulus presentation, as described above, is not readily achievable within the context of a single unit recording experiment. Firstly, the response latencies of visual cells are significant, with V1 responses tending to

---

[13] For a proof that this kind of strategy results in faster convergence on a quadratic performance surface, especially when the dimensionality of the search is high, see Widrow (1985), p. 162.

have latencies in the range 40-100ms (Vogels & Orban, 1991), and by the time visual information reaches later areas such as IT and STS, response latencies have risen to 60-220ms (Richmond et al., 1983; this thesis). Latencies of this order set an upper limit on the presentation rate at around 5-10Hz, significantly less than the 50Hz presentation rate at which we are able to record specific responses to stimuli presented in an RSVP[14] sequence (Keysers et al., 2002).

Computational issues also limit the presentation rate if parameter updates occur with every stimulus-response pair. Depending on the complexity of the image model, it may take a significant amount of time to compute an image from the stimulus parameters.

For these reasons, the design of the system was modified to instead generate and present short "blocks" of stimuli, each composed of many different images, which can be presented rapidly and continuously in an RSVP sequence. This style of presentation also fits well with the typical length of fixation periods that are obtained during recording sessions.

Each block of stimuli is created by adding multiple independent noise patterns to the base parameters. In the *i*th block of stimuli, the *j*th perturbation of the base parameter vector is now given by –

$$s^{(i)(j)} = b^{(i)} + n^{(i)(j)} \tag{3}$$

---

[14] In RSVP, the continuous recording of responses can simply be shifted to align each response to the stimulus that evoked it.

and after this block of stimuli has been presented, the base parameters are updated according to –

$$b^{(i+1)} = b^{(i)} + \alpha\left\langle (r^{(i)(j)} - \bar{r}^{(i)}) n^{(i)(j)} \right\rangle_j \tag{4}$$

where $r^{(i)(j)}$ is the response to the *j*th stimulus of block *i*, $\bar{r}^{(i)}$ is the average response during the block and $\langle\ \rangle_j$ denotes the expected value across the j perturbations.

As in (2), the parameter update performs gradient ascent on the response, as the correlation is proportional to the partial derivative for independent noise (see box 4.1). After the updates to the parameters have been computed, the next block of stimuli is then generated by adding a new set of noise to the new base parameters.

Box 4.1 – Proof that Correlation is Proportional to Gradient

The gradient of the performance surface, $f(\underline{x})$, is a column vector designated $\nabla(f)$ –

$$\nabla(f) = \begin{bmatrix} \dfrac{\partial f(\underline{x})}{\partial x_1} \\ \vdots \\ \dfrac{\partial f(\underline{x})}{\partial x_n} \end{bmatrix} \tag{1}$$

If we perturb the values of the input vector by an amount $\Delta\underline{x}$, then a linear approximation (which will be accurate when the step size is low) to the resultant change in the response, Δf , will be given by –

$$\Delta f \cong \Delta\underline{x}^T . \nabla(f) \tag{2}$$

The search algorithm evaluates the correlation –

$$\begin{aligned} \left\langle \Delta\underline{x}.\Delta f \right\rangle &= \left\langle \Delta\underline{x}.\Delta\underline{x}^T . \nabla(f) \right\rangle \quad \text{substituting, from (2)} \\ &= \left\langle \Delta\underline{x}.\Delta\underline{x}^T \right\rangle \nabla(f) \\ &= Cov[\Delta\underline{x}]\nabla(f) \end{aligned}$$

and

$$\nabla(f) = Cov[\Delta\underline{x}]^{-1} \left\langle \Delta\underline{x}.\Delta f \right\rangle \tag{3}$$

In this case, each element of $\Delta\underline{x}$ is independent, so –

$$\left\langle \Delta\underline{x}.\Delta f \right\rangle = \sigma^2 \mathbf{I} \nabla(f) \tag{4}$$

showing that the correlation is proportional to the gradient.

The block presentation algorithm spends more time gathering information about the gradient around the current location before moving to a new location. Stimuli can be more rapidly presented, and the computations involved in parameter updates and image generation can be carried out in between fixation periods.

(b) Multiple solutions

A further addition to the system allows multiple parameter optimisations to be carried out simultaneously. It is possible that the neuronal response surface may have multiple peaks, corresponding to different patterns, sharing few, if any, features in terms of retinal input. For example, in chapter 2 we saw how some neurons in IT and STS can produce invariant responses to particular objects, irrespective of which view of the object was presented.

The perturbation j of the kth base parameter vector after i blocks is now given by –

$$s^{(i)(j)(k)} = b^{(i)(k)} + n^{(i)(j)(k)} \tag{5}$$

and the correlation equation becomes –

$$b^{(i+1)(k)} = b^{(i)(k)} + \alpha\left\langle (r^{(i)(j)(k)} - \bar{r}^{(i)(k)}) n^{(i)(j)(k)} \right\rangle_j \tag{6}$$

Note that stimulus generation and correlations are carried out completely independently for each base parameter vector. This provides multiple independent solutions to the optimisation.

At each step of the optimisation, the block of stimuli to be presented will consist of multiple perturbations of multiple base images. Stimuli based on different parameter vectors are randomly interleaved in the continuous presentation sequence (see figure

4.2), which reduces the risk that a neuron under study will habituate to any particular image feature.

(c) Response Sampling in RSVP

The response to each stimulus presented as part of a block is obtained by cutting the spike train into slices of lengths equal to the stimulus presentation time, after taking the response latency of the neuron (which is assumed to be constant across stimuli) into account. This is shown in figure 4.3.

As discussed in chapter 2, responses to effective stimuli typically outlast the stimulus presentation time, by about 60ms in STS and IT (Keysers, 2000). This effect is also illustrated on figure 4.3, where spikes that relating to the effective stimulus can be

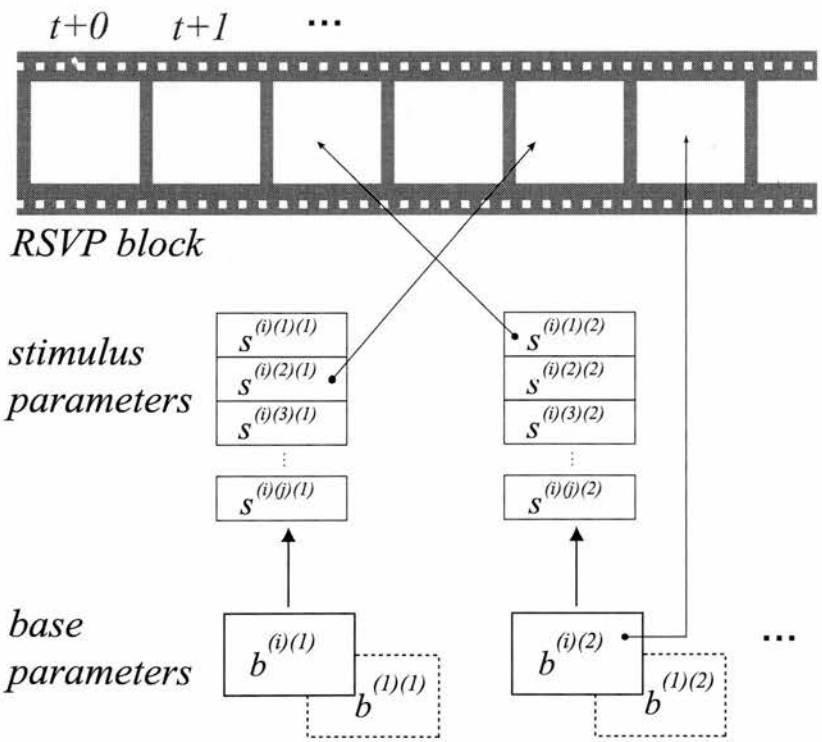


figure 4.2 – An RSVP block consisting of randomly interleaved stimuli, which are derived from one or more base parameter vectors. Stimuli may also be generated directly from the base parameters themselves, without the addition of noise. Copies are kept of the original base parameters, prior to any parameter updates (indicated by the dotted boxes). These may also be included in a block, as a baseline measure. Note that every stimulus is included once in the sequence, but only some of the arrows have been drawn, for clarity.

seen to fall into one (or more) of the windows belonging to previous and subsequent stimuli.

There are two simple solutions to this problem. Firstly, stimulus presentation can be slowed down, which will dilute the effect of spike "overspill" between stimulus windows. Because neurons in STS and IT tend to show response persistence until a new stimulus is displayed, it should make little difference whether this slowing takes the form of longer stimulus presentation, or the introduction of gaps between stimuli (see figure 2.9b).

A second solution is to present each stimulus multiple times within a block, and use an average spike count across the multiple presentations when calculating the correlation in (6). This will not only reduce the effect on noise on the response estimate, but will also dilute the influence of spike overspill between stimuli.

(d) Baseline Measures

Monitoring the progress of the optimisation process requires a baseline measure because cells may habituate to a particular image (see chapter 2) or exhibit a change in responsiveness over time. For this reason, a copy of the initial parameters is made,

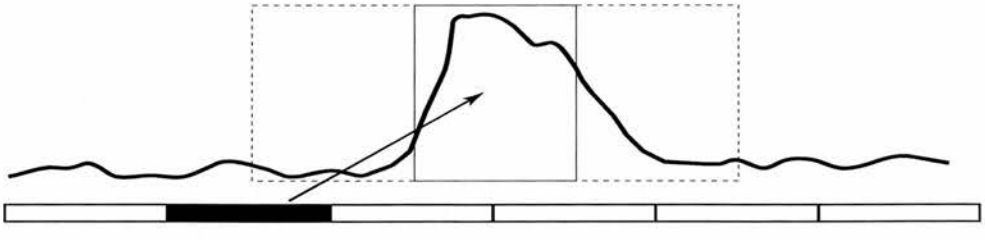


figure 4.3 – Response sampling in RSVP. The continuous recording sequence is sliced into windows, each of length equal to the stimulus presentation time. Taking cell latency into account, a windowed spike count is obtained for each stimulus. The lower filled rectangle indicates an effective stimulus producing the response above. Lower unfilled rectangles represent previous and subsequent stimuli. Spikes falling within the window indicated by the arrow would be attributed to the effective stimulus. Spikes falling with the earlier and later windows (shown by dotted rectangles) would be attributed to the previous and subsequent ineffective stimuli.

allowing an image created from these parameters to be inserted in every stimulus presentation block (see figure 4.2). An image created from the base parameters can also be shown in each block to evaluate the response to the stimulus generated from the parameters in the absence of noise.

(e) Starting Point of Search

The neuronal response surface may contain vast areas where the response (and therefore the gradient) is close to zero. Stimulus optimisation is most likely to succeed if the search can initialised in the region of a peak.

This can be achieved if the neuron is first tested with a standard test battery of images, and the image(s) producing the best response selected for optimisation. This requires the image model to support both a forward transform (from parameters to image) and a backward transform (from image to parameters).

4.2.4. Two-stage Image Model

A stimulus parameter vector, $s^{(i)(j)(k)}$, is converted into a pixel image by means of the two-stage image model outlined in figure 4.4. The image model determines how stimuli are generated from a set of parameters, and therefore constrains the type and range of images that can be generated. The first stage is dependent on the particular representation employed, and the details of each implementation are discussed in the sections that follow. The output of the first stage is always a virtual pixel based image, in YCbCr colour space. This is acted on by the second stage, which is common to all representations, and is therefore discussed first here.

YCbCr colour space (Bhaskaran & Konstantinides, 1997) is used in all stimulus representations because it separates out luminance from chromatic information. The

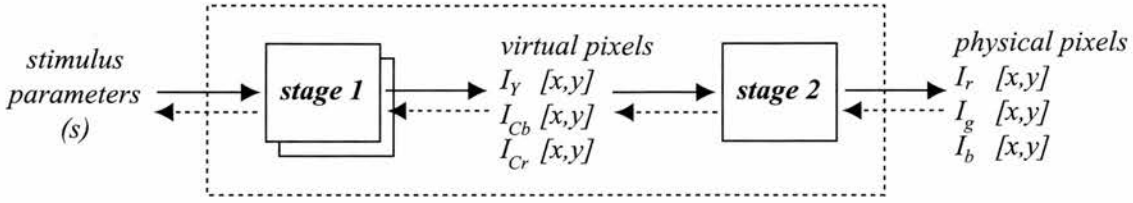


figure 4.4 – The two-stage image model. Stage 1 is dependent on the particular image representation in use, and transforms the stimulus parameter vector into a pixel image basis, with three matrices representing virtual pixels in YCbCr colour space. Stage 2 is invariant of image model and converts virtual pixels into physical RGB pixels, ready for display on a monitor. The forwards transform (solid arrows) is carried out for every stimulus generated. The backwards transform (dashed arrows) occurs once, prior to generation of the first block of stimuli, if the optimisation process is initialised with a particular starting image.

---

output of the first stage is a matrix, $I_Y[x,y]$, containing the luminance values of each virtual pixel and, if optimisation is being carried out in colour, two additional matrices $I_{Cb}[x,y]$ and $I_{Cr}[x,y]$ contain the blue-yellow and red-green chromatic information respectively. Each YCbCr triplet is converted into its equivalent RGB value by first shifting Y from the range [-0.5..0.5] into the range [0..1] then applying the following transformation –

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{7}$$

Thus (0, 0, 0) in YCbCr colour space describes a mid-luminance grey and is translated to (0.5, 0.5, 0.5) in RGB space. Increasingly positive values of Y generate pixels with luminance above the mid-level, increasingly negative values generate pixels with below mid-level luminance. Similarly, positive values of Cb generate increasingly blue colours, negative values generate increasingly yellow colours; positive values of Cr generate redder colours, negative values generate greens.

Finally, each virtual pixel can correspond to many physical pixels in the final image

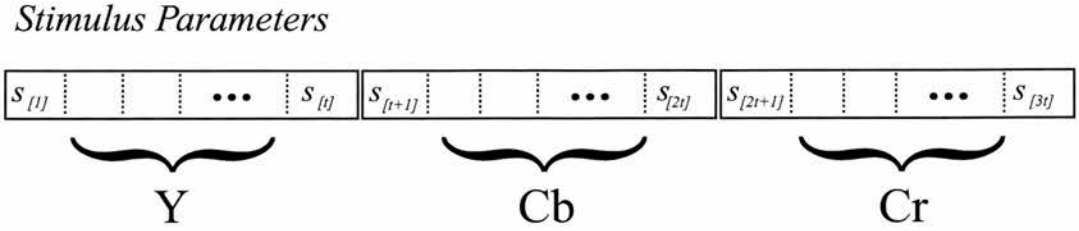


figure 4.5 – Organisation of pixel data in a stimulus parameter vector, for the pixel image model. Each parameter specifies attributes of a single pixel. If the optimisation is being carried out for luminance (Y) alone, only the first set of parameters is used. Otherwise, two additional sets of parameters specify the chromatic information (Cb & Cr) for each pixel. *t* is the total number of virtual pixels in the stimulus image.

---

prepared for display. Width and height magnification factors can be separately controlled to determine the size (and therefore the spatial resolution) of the stimulus.

4.3. Pixel Image Model

The pixel image model is the simplest of the image models implemented, with each parameter simply representing the luminance of a particular pixel. If the optimisation is being carried out in colour, two additional sets of parameters specify the chromatic components for each pixel. Therefore the first stage of the image model is effectively an identity transform, with only a simple repackaging of coefficients taking place.

If the resolution of the image to be optimised is *x* pixels wide by *y* pixels high, the total number of parameters will therefore be *xy* (for achromatic optimisation) or *3xy* (in colour). The packaging of information in the parameters is shown in figure 4.5.

The following sections describe results obtained with this pixel image model.

4.3.1. Simulation with Model Complex Cell

A complex cell model was constructed in order to simulate the dynamics of the optimisation algorithm with a non-linear response. In this case, the single presentation form of the algorithm (equations (1) and (2)) was used. A test of the block

presentation algorithm is described in the section that follows this one.

The complex cell was modelled as a summation of multiple simple cell inputs (figure 4.6a), where each simple cell consisted of a linear spatial filter, based on a gabor function, followed by a threshold non-linearity (Movshon, Thompson & Tolhurst, 1978a; Jones & Palmer, 1987). The complex cell was modelled as a sum of several of these simple cell inputs (Movshon, Thompson & Tolhurst, 1978b), with the spatial filter of each input differing in phase, sampled evenly across the range [0,2π].

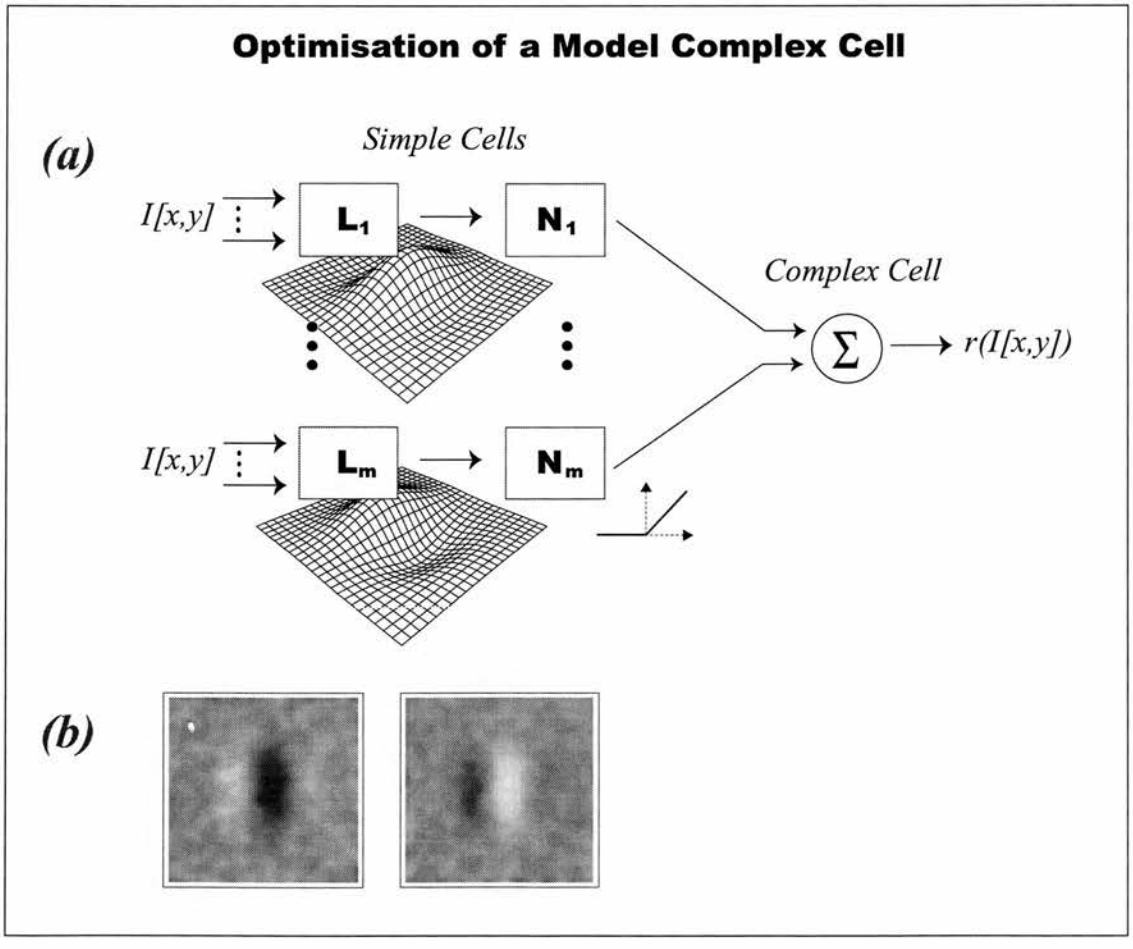


figure 4.6 – (a) A complex cell modelled as a summation of simple cell inputs. Each simple cell is modelled as a linear (L) and static non-linear (N) cascade – i.e. a spatial filter followed by a threshold. The complex cell sums parallel LN channels, whose linear spatial filters differ in phase. (b) The results of two optimisation runs carried out on the model complex cell (Gaussian filtered). Each run was initialised with identical parameters, but different random seeds. The optimised images both produce near-maximum output from the complex cell model, but the patterns differ in phase, with the bright area of one stimulus occupying the same spatial location as a dark region in the second. [α =0.002, σ =0.1, i=2000]

The results of two optimisation simulations with the model complex cell are shown in figure 4.6b. In both cases the parameters were initialised at zero, to produce a uniform mid-luminance grey image, but with a different random seed in each simulation. The optimised images both produce near-maximal output from the complex cell model, but the patterns differ in phase, with light regions in one image corresponding to dark regions in the other.

This simulation illustrates how the algorithm can find alternative maxima in the neuronal response surface, where the system under study is a non-linear function of the input (i.e. the regions where light excites and inhibits the response are not fixed).

4.3.2. Coarse to Fine Optimisation

The optimisation method presented allows the spatial resolution of the image to be changed mid optimisation, thus allowing us to rapidly optimise a coarse stimulus (described by a low number of parameters), yet change to a higher resolution image when performance reaches a plateau.

This is illustrated by the simulation shown in figure 4.7. In this case, an image is optimised by carrying out gradient descent[15] on an quadratic error function, which was calculated as the mean squared error of pixel differences between the current image and a 64 x 64 pixel target image.

---

[15] On a real neuron, we are trying to make our way uphill on the performance surface. However, with an error measure, we wish to do the opposite. The algorithm will perform gradient descent (rather than ascent) if the sign of the learning rate constant (α) is reversed in equation (2).

Initially, there are only $1/16^{th}$ as many parameter as pixels in the image, with each parameter controlling a 4 x 4 pixel block. After 1500 iterations, performance is no longer improving, and the remaining error being is due to residual detail below the resolution of the optimisation. The number of parameters is then increased, such that each pixel is controlled by a single parameter, and performance starts to increase again, with additional fine detail in the target image being discovered.

In this example, converge is faster with a coarse stimulus representation because the target image is naturalistic and has local correlations between pixel values (any point in the image tends to be like its immediate neighbours). Therefore a coarse

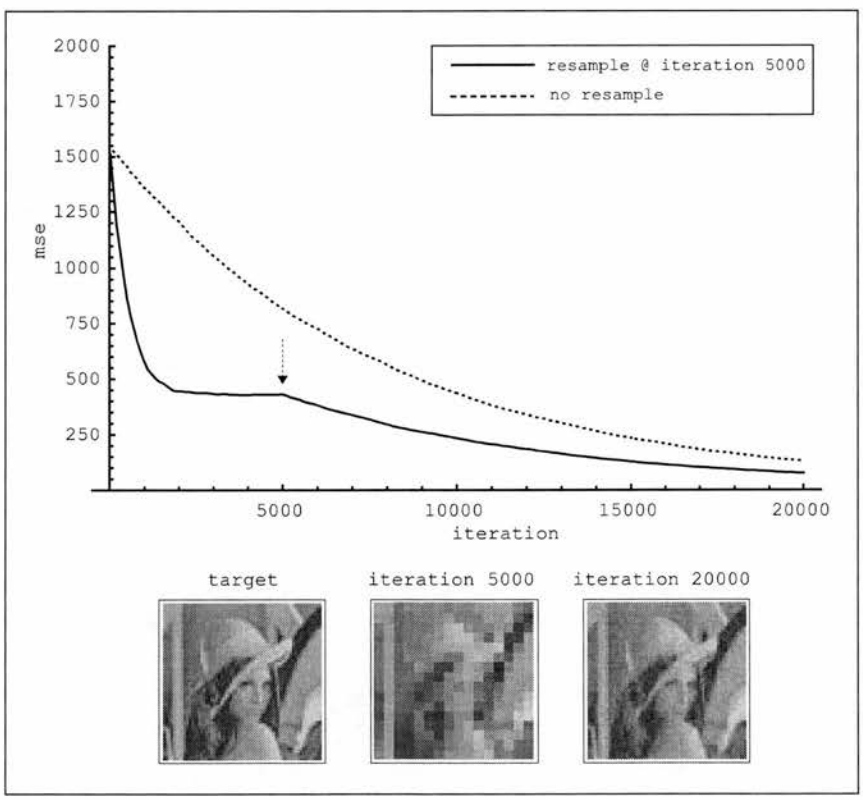


figure 4.7 – Optimisation of an image by error descent. The quadratic error function measured the difference between the search image and a 64 x 64 pixel target image (*lena*, an image used widely in image processing). The error was the mean squared pixel difference (where each pixel value ranged from 0-255). Top trace – Each parameter represented the luminance of one pixel (total of 4096 parameters). Bottom trace – initially, each parameter represented the luminance of a 4 x 4 pixel block (total of 256 parameters). At iteration 5000 (arrow), the number of parameters was increased (using a backwards transform on the image) such that each parameter controlled a single pixel (4096 parameters, as in top trace). This simulation used the single presentation form of the optimisation algorithm (equations (1) & (2)). In all cases the values of the convergence parameters were the same. [α=0.001, σ=0.05, average over=10].

representation of the image can account for a large proportion of the error measure.

To what extent is this technique likely to be useful with a real visual neuron? If the optimal stimulus for a cell is a mix of both low- and high-spatial resolution information (e.g. a face) *and* a significant response can be obtained when the low-spatial resolution information alone is presented (i.e. the cell responds to a coarse representation of the stimulus) then the coarse-to-fine optimisation technique should prove experimentally useful. There is at least some evidence that IT cells do indeed respond significantly to coarse versions of their preferred stimulus (e.g. Rolls et. al., 1995 showed this with faces) and I shall return to this question later in the thesis.

4.3.3. Test with Artificial Neuron

In order to verify the working of the block design algorithm within the context of the complete physiology system, a hardware artificial neuron was built which could be tested in place of a real visual cell. The artificial neuron consisted of a square wave generator connected to two photoresistors which were attached to the monitor used for stimulus presentation. The photoresistors were connected as a "push-pull" pair, with one increasing the frequency of the square wave as the light intensity increased, and the other decreasing the frequency as light intensity increased.

The results of an optimisation run on this artificial neuron are shown in figure 4.8. After one parameter update, a bright spot has appeared under the photoresistor which increased the output frequency, with a corresponding dark spot evident under the inhibitory one. With subsequent presentations the spots become more clearly defined, and, after five to six parameter updates, the artificial neuron reaches its maximal output rate.

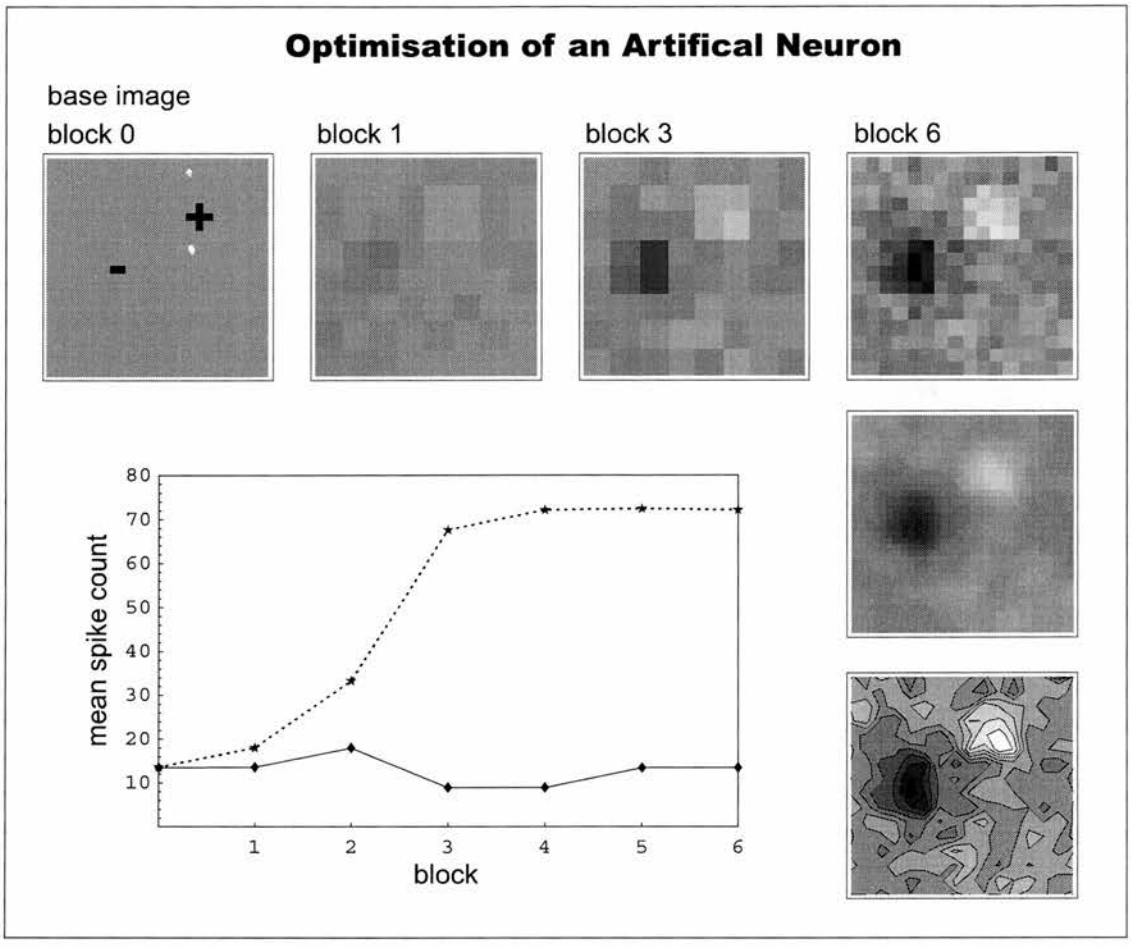


figure 4.8 – Optimisation run with an artificial neuron, consisting of two photoresistors attached to the display screen and modulating the frequency of a square wave produced by a function generator. The approximate location of each photoresistor is shown in the top left of the figure. Light falling on the photoresistor indicated by a + increased the output frequency, whereas light falling on the photoresistor indicated by a – decreased the output frequency. The top row of images shows the evolution of the base parameters across presentation blocks. Middle and lower right hand side images show Gaussian smoothed and contours versions of the final image. The graph shows the response to the base parameters (upper curve) and the unoptimised start image (lower curve) across presentation blocks. Note that the parameters were resampled after block 3 to increase the spatial resolution (see text). [α =0.005, σ =0.1, 100 stimuli per block]

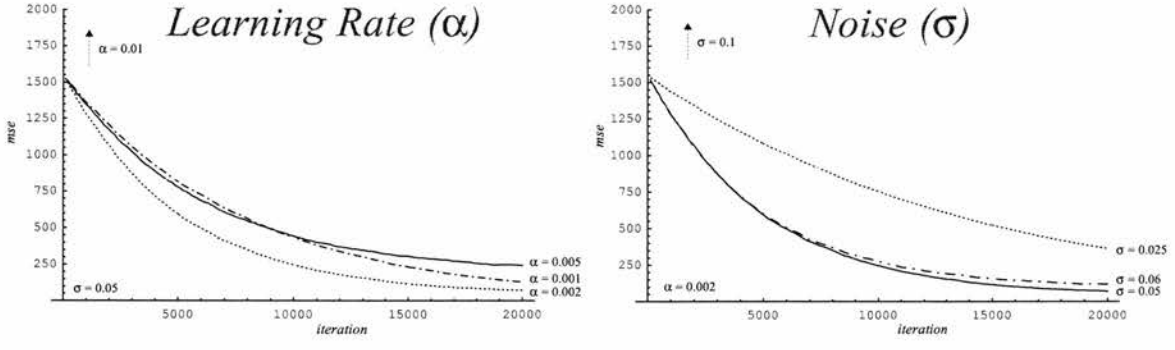
This example also illustrates how the resolution of the image can be changed from coarse to fine during an optimisation run. After the third parameter update, the base image was resampled to initialise a new parameter set with twice the spatial resolution. As described earlier, this allows a coarse image to be rapidly optimised, then the spatial resolution increased to provide a finer picture of the receptive field.

Box 4.1 – Choice of Convergence Parameters

The convergence parameters, α & σ, must be chosen with respect to the system under study.

- The value of α should be chosen such that the value of $\alpha(r^{(i)} - \bar{r})$ (see equation 2) will be << 1, otherwise the optimisation will be unstable. However, if it is too low, little learning will take place. The dynamic range of the response will determine a reasonable range for this parameter.
- The value of σ is dependent on the nature of the image model employed, and determines the magnitude of the noise, $n^{(i)}$, added to the image parameter vector. This should be chosen to be large enough such that the noisy images produce measurably different responses from the response to the base image, or no learning can take place. If too large, σ may produce an unreliable estimate of the gradient (depending on the nature of the performance surface).

Given the constraints above, the graphs below illustrate the effect on convergence over a range of values for these parameters (using the single step form of the algorithm).



Pixel-based optimisation (gradient descent on a quadratic error function – for additional details on method see figure 4.7), using a 64x64 pixel target image. Left – Varying the value of the learning rate (α). When the learning rate is too high ($\alpha = 0.01$), convergence is not achieved. Lower values ($\alpha<0.01$) result in stable convergence. In this stable range, higher values of α result in faster convergence initially, but stabilise with a higher mean squared error. Lower values convergence less rapidly, but stabilise closer to the optimal values ($\alpha = 0.001$ is still converging at the end of the trace shown, and will eventually stabilise at an error level below that of higher values of α). Right – Varying the value of the noise standard deviation (σ). As with the learning rate, when the value of σ is too high, convergence is not achieved. Within the stable range, lower values result in slower convergence initially, but stabilise with lower error levels.

4.4. Spatial-Frequency Image Model

For any n-pixel image, one requires an n-dimensional space to represent the set of all possible images, with every possible image occupying a unique location in this space. In the pixel based image model described above, we can imagine that each parameter describes the distance along a specific coordinate axis, where each axis represents the intensity of a particular pixel. The spatial frequency model, described here, is based on a transformation of the coordinate axes, where each axis now represents the contribution of a specific spatial frequency to the image. This transformation is achieved by a Fourier transform[16] (an orthonormal transform, equivalent to a rotation and/or reflection of the coordinate axes).

Specifics of the implementation of this image model (along with a scaling step that is discussed later) are shown in figure 4.9.

4.4.1. The Fourier Transform and Stimulus Generation

Each stimulus generated by the optimisation algorithm is picked from a random location in parameter space, with the probability density described by an n-dimensional Gaussian centred on the current values of the parameters.

It is worth noting that this probability density will have the same shape in image space, irrespective of whether the parameters represent a pixel basis, a Fourier basis, or any other orthonormal transform of the axes. This is because orthonormal transforms are isometric (i.e. form preserving). Thus the Fourier transform, by itself,

---

[16] See Press et al., (1988) for details of the specific implementation used here.

will have no effect on either the type of stimuli generated, or the performance of the optimisation.

However, the spatial-frequency based model does allows us to develop two different strategies to constrain the search –

(i) We can simply reduce the number of parameters and thus restrict the stimulus search to a lower dimensional subspace. This allows *a priori* knowledge about the tuning of cells to be used to constrain the search[17]. This idea is investigated further in the following two chapters, where we test the hypotheses that, at high stimulus presentation rates (where the visual system has little time to process one stimulus being presented with

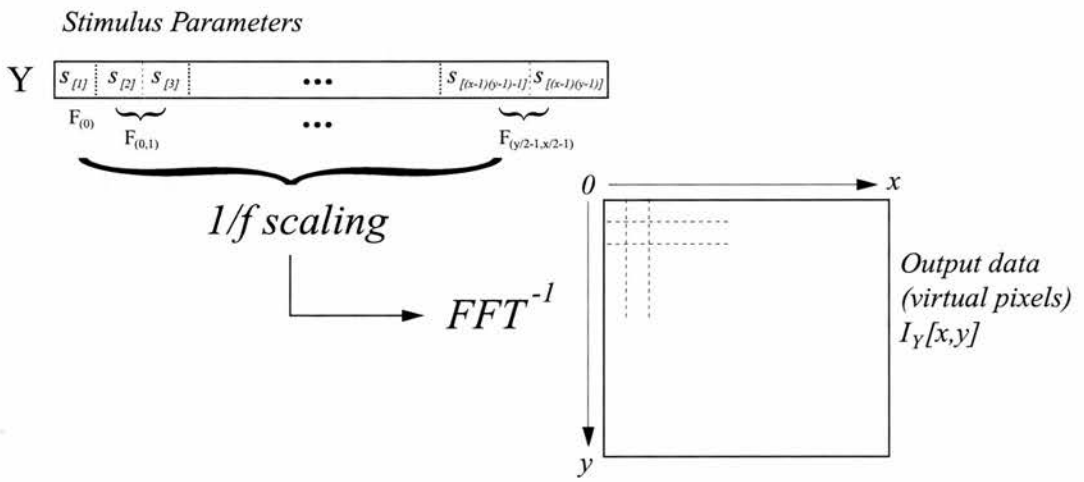


figure 4.9 – Stimulus parameterisation with the 1/f frequency image model. Stimulus parameters represent Fourier coefficients. For an image with $x \times y$ pixels, there are $(x-1)\times(y-1)$ parameters. The first parameter represents F(0), the DC component of the image. Subsequent parameters are grouped into pairs, representing real and imaginary parts of each spatial frequency. Nyquist frequencies are set to zero (and are not represented by parameters) resulting in an decrease in dimensionality. Each parameter pair is scaled by a factor of 1/f (where f is the frequency represented by the pair) prior to the inverse Fourier transform. Only the Y channel is shown. An identical process takes place for the Cb and Cr channels when optimisation is carried out in colour. [For a complete description of how individual spatial frequencies are encoded by the parameters see Appendix 4 and Press et al., 1988].

[17] The same idea is discussed with respect to the reverse-correlation technique by Ringach et al. (1997).

the next one), the response of neurons in STS and IT may be determined only by low-spatial frequency and achromatic aspects of a stimulus.

(ii) Alternatively, we can try to incorporate knowledge about the statistics of natural images into the generative image model. It is likely that neurons in the visual system of an organism are tuned to features of its natural environment. Analysis of these natural images has revealed consistent statistical properties, which can be used to constrain the search. This idea is considered next.

It is obvious that both of these approaches reduce the generality of the optimisation technique. However, if the constraints placed on the search volume are well chosen, and matched to the properties of the cells under study (and cells providing input to the cells under study), this should not be detrimental to the accuracy of the results.

4.4.2. 1/f Amplitude Spectra of Natural Images

The visual environment of an animal is not random, but highly structured. The pattern of light reflected by a scene is determined by specific physical laws, and reflects the properties and configuration of the independent objects within it. This structure limits the range of images the organism will encounter, such that natural images only occupy a small fraction of the total image space. Therefore we can say that there is redundancy in the retinal input.

Field (1987) has described one consistent statistical feature of natural images in terms of their amplitude (or power) spectra. In contrast with white noise images, which have a flat amplitude spectra, natural images show the greatest amplitudes at low

frequencies, with decreasing amplitude as frequency increases. The amplitude falloff with frequency is approximately *1/f*, a relationship that holds for a wide range of natural images. If natural images have stationary statistics (over the whole population, the statistics at one location are no different from any other location) then all the redundancy reflected in correlations between pixels will be captured by the amplitude spectrum of the data.

We can modify the spatial frequency image model to produce noise patterns that have a 1/f spectrum (and therefore the same kinds of correlations between pixels as natural images) by introducing a scaling step (shown in figure 4.9) between the parameters and the Fourier transform. This alters the dynamic range of each parameter, such that parameters representing higher frequencies have a reduced effect on the amplitude of the spatial frequency that they represent. This yields correlated noise patterns like the images shown in figure 4.10.

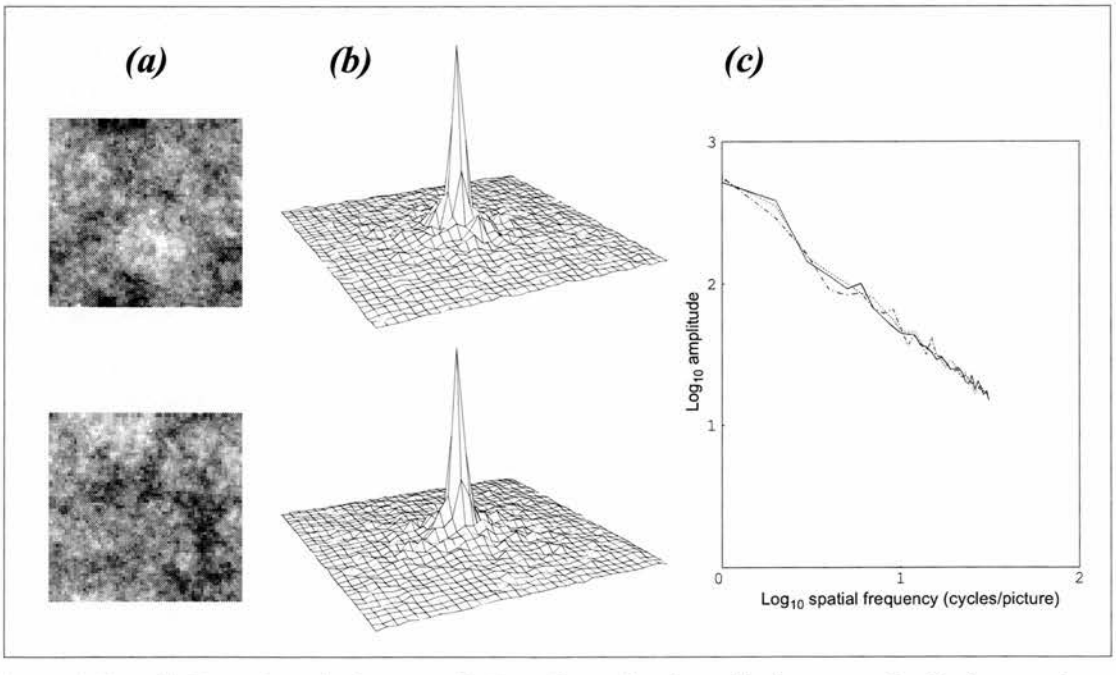


figure 4.10 – *(a)* Example noise images. *(b)* Two-dimensional amplitude spectra for the images shown to the left. 0 spatial frequency is shown in the centre of the plot. Amplitude decreases rapidly with increasing spatial frequency at all orientations. For clarity, each point on the grid is a 2x2 average region of the spectrum. *(c)* Amplitude spectra for these images (plus a third, not pictured). With log-log coordinates, the spectra fall off by a factor of 1/f (indicated by a slope of –1). [After Field, 1987].

It is important to note, however, that Field (1994) has also argued that much of the redundancy in natural scenes is not captured by correlations between pixels, and suggests that pyramid codes (e.g. based on Gabor functions), rather than pure frequency based representations are necessary to model the statistical regularities in natural images. The Fourier model presented in this section should therefore simply be considered as a first tentative step towards modelling natural image statistics, and we shall return to the subject of wavelet coding in the discussion.

4.4.3. Test of the 1/f Image Model with Artificial Neuron

The 1/f image model has also been tested with the artificial neuron described earlier. The results of one test are shown in figure 4.11. Convergence to the optimal stimulus is faster than with the pixel model as can be seen be a comparison of this figure and figure 4.7. This would be expected, because the photoresistors were sensitive to light falling over a fairly wide spatial extent. With the 1/f image model, there would be a high signal to noise ratio for the parameters controlling the amplitude of the low spatial frequencies. In contrast, there would be relatively low signal to noise ratios in the pixel model, since each pixel would have a comparatively small influence on the response.

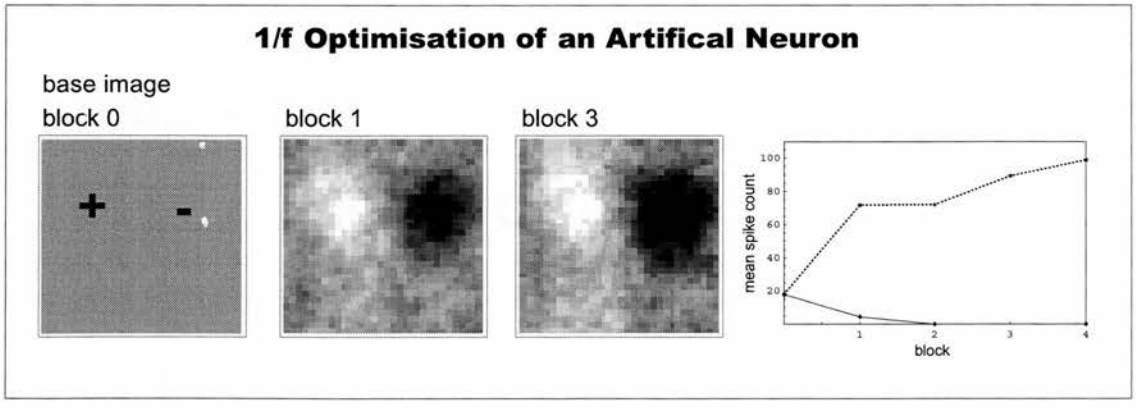


figure 4.11 – Optimisation run on the artificial neuron with the 1/f frequency image model. Convergence is more rapid, and the images clearer than those obtained with the pixel model described earlier. + and – indicate approximate photoresistor positions (see figure 4.7 for full details). [α =0.005, σ =0.03, 100 stimuli per block]

4.4.4. Experimental Results

Experimental tests using the optimisation system on real IT visual neurons proved inconclusive. The system was tested with 11 cells, but it was rarely possible to achieve a stable recording over a long period. Before running an optimisation test, visual neurons were identified as such by first testing their responses using a screening set of images (similar or identical to the set shown in Appendix 1).

During optimisation, the maximum number of iterations achieved was 11, but more generally only 5 or 6 iterations were obtained. The results are shown in Table 4.1 and Figure 4.12. There is little evidence of increasing response to the images across iterations.

The table shows the parameters used for each test, the images (if any) used to initialise each set of parameters, and notes that were recorded during the test. In general, the presentation time was 56ms per stimulus (4 frames) with a 167 ms gap between stimuli (12 frames).

Stimulus presentation blocks always included noiseless images based on the current parameters, as well as the original base images (after parameterisation), as a baseline measure. However, the images after parameterisation (to 32 x 32 virtual pixels) produced a severely reduced response compared to that obtained with the original images testing during the screening phase, and in general, the baseline measures can be seen to be extremely variable.

Table 4.1 – Optimisation Tests in IT Cells

Date	*Cell*	*Test*	*Notes*
11/4/00	250/7	1	Pixel optimisation. Terminated after 2 iterations, monkey not working.
26/4/00	254/8	1[a,b]	1/f colour optimisation with 3 parameter sets initialised from screening images (boxc_, M1_00c, ColorMc_) α=0.005, σ=0.02, 8 iterations. 32x32 vitual pixels. 198 images per iteration (66 per parameter set).
26/4/00	254/8	2[a,b]	1/f colour optimisation with 3 parameter sets (default initial images). Parameters as above, 5 iterations. Computer crash terminated test.
27/4/00	255/6	1[a,b]	1/f colour optimisation with 3 parameter sets (default). α=0.02, σ=0.02, 8 iterations. 32x32 vitual pixels. 198 images per iteration (66 per parameter set).
23/5/00	258/4	1	1/f optimisation – cell lost.
2/6/00	266/15	1[a,b]	1/f achromatic optimisation with 2 parameter sets (default). α=0.01, σ=0.03, 11 iterations. 32x32 vitual pixels. 200 images per iteration (100 per parameter set).
19/6/00	271/9	1[a,b]	1/f colour optimisation from 3 grey images. α=0.005, σ=0.03, 5 iterations. 32x32 vitual pixels. 300 images per iteration (100 per parameter set).
29/8/00	285/2	1[a]	1/f colour optimisation. Single parameter set. α=0.01, σ=0.1, 5 iterations. 32x32 vitual pixels. 100 images per iteration. Lost cell? retest with screening set showed different selectivity.
21/2/01	302?/5	1[a]	1/f achromatic optimisation with 2 parameter sets (default). α=0.01, σ=0.1, 5 iterations. 32x32 vitual pixels. 200 images per iteration (100 per parameter set). Cell probably lost during test.
15/3/01	307/16	1	1/f colour optimisation with 3 parameter sets (Handc_ , 2 default). α=0.05, σ=0.1, 0 iterations. 32x32 vitual pixels. "Aborted".
4/4/01	311/3	1[a]	1/f colour optimisation with 3 parameter sets (Alienc_, 2 default). α=0.05, σ=0.1, 5 iterations. 32x32 vitual pixels. 180 images per iteration (60 per parameter set). Lost cell, monkey not working.

[a] Complete results for this test are shown in a separate figure.

[b] The 1/f image model used in this test was slightly different to that described in the thesis. Values for σ are therefore not comparable to the later results.

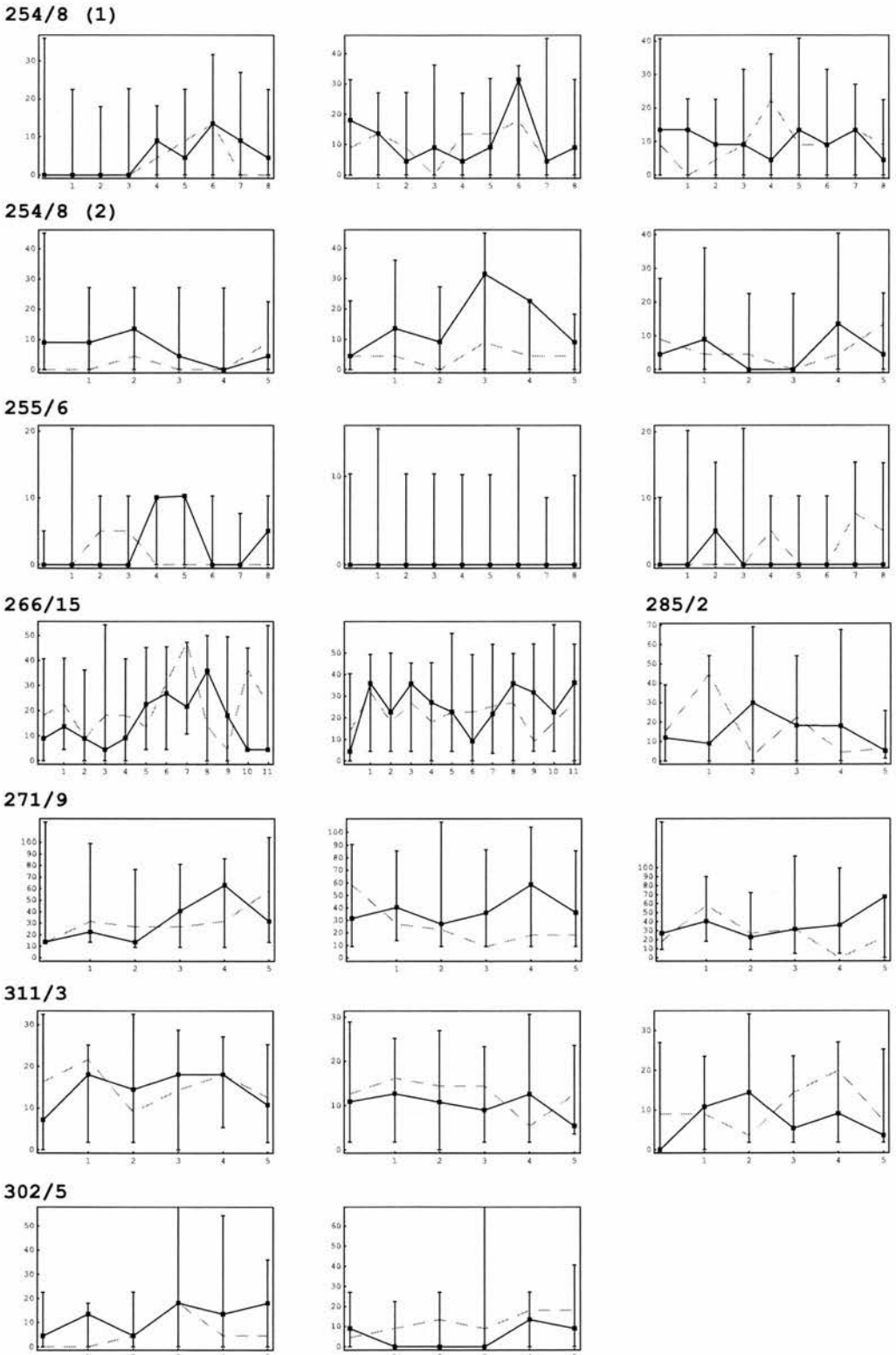


figure 4.12 – Optimisation runs on mutliple IT cells. Separate graphs (arranged from left to right) are shown for a cell where multiple parameter sets were simultaneously optimised. Parameter sets were initialised as described in Table 4.1, with the leftmost graph corresponding to the first initial image listed (and so on). A single copy of the original base image was included in the block of images presented every iteration to act as baseline measure (for further details – see text and figure 4.2). The response to this image is shown by the grey dashed line. The solid lines (and marks) indicate the response to the base image, with error bars showing the maximum and minimum responses produced by the search images. X-axis labels indicates iteration number, Y-axis labels indicate spike count.

4.5. Discussion

The stimulus optimisation technique is shown to reveal multiple receptive fields for non-linear artificial neurons. This indicates that it should be more applicable than the reverse correlation technique for non-linear neurons such as complex cells (see Földiák, 2001). It may also be suitable for studying stimulus selectivity in higher areas, such as V4 or IT. However, this is likely to require a stable recording from a single cell over long periods of stimulus presentation, which was never achieved during the course of the work described in this thesis. Particularly worrisome was the lack of consistency in the measured responses to the original images over the course of optimisation testing on real cells, which were intended to act as a baseline measure. If there is a very high level of noise present in the responses attributed to each stimulus, the process of optimisation is likely to be very slow.

However, the main problem with stimulus search is likely to be one of dimensionality. The speed of convergence is dependent on the number of parameters used to represent the stimulus. It is possible to reduce the dimensionality of the problem if we are prepared to search only a subspace of image space. This may be applicable in some cases, and in the next chapter we consider whether a search of the specific subspace corresponding to low spatial frequency achromatic images might prove a practical approach in STS and IT. Parallel work which seeks to understand receptive field properties in terms of natural image statistics may also prove useful to generating candidate stimuli that are more likely to contain the trigger features of a visual cell.

4.5.1. Natural Image Statistics

Several authors have argued convincingly that visual processing is best understood in

terms of transforming or reducing the redundancy present in the natural environment, in order to produce a representation that can represent natural images efficiently. Barlow (1961) first suggested that the goal of sensory processing was to represent the information in the natural environment with minimal redundancy. Atick & Redlich (1990, 1992) apply the principle to retinal ganglion cells, showing their receptive fields can be understood in terms of a strategy for decorrelating their outputs in response to the 1/f amplitude spectrum of natural images. Field (1987, 1994) discusses how a set of self-similar gabor wavelet filters are effective at transforming the high-order redundancy present in natural images into first order redundancy in the response distribution of the coefficients, where each input is represented by the minimum number of active units (i.e. a sparse code). Olshausen & Field (1996) show that a network designed to discover sparse representations produces receptive field structures similar to those determined for simple cells in V1, if presented with a large number of natural image samples.

This line of study has been effective in providing explanations for the structure of early visual receptive fields, and as the techniques become more powerful, will also enable the prediction of receptive field structures later in the visual system.

4.5.2. Relationship to Stimulus Optimisation

Field (1994) suggests that a first approximation to natural scenes may be achieved by producing a random sum of orientated wavelet basis functions, where the probability of any particular coefficient taking on a high value is very low (in essence, inverting the visual code proposed for V1). Later work (see Schwartz & Simoncelli, 2001) suggests that natural images contain consistent relationships between the values of

wavelet coefficients across neighbouring positions, orientations and scales.

Simoncelli's work shows structure beyond correlation – the magnitude of one filter response is not predictive of the magnitude of a second (in fact the correlation is zero) but is instead predictive of the *variance* of the second filter, with the strength of the dependency varying depending the on the specific pair of filters chosen.

As further regularities are discovered, these could be incorporated into a generative stimulus model, to produce potential stimuli that contain the kinds of features present in natural images, and therefore likely to be trigger features of visual neurons.

However, although natural images contain redundancy, wavelet representations do not allow for simple reduction in the number of parameters used to represent images. In fact, the basis may be overcomplete (i.e. more basis functions that pixels). As Olshausan & Field (1996) point out, there is no reason to suppose that an image with N pixels is composed from N independent causes. Rather, the regularities present in natural images are described by the distribution of (i.e. sparse), and relationship between, the wavelet coefficients.

It is worth noting that this will be non-trivial to incorporate in the image model and search proposed in this chapter, because it is based on statistically independent parameters. It is possible that other types of search, such as genetic algorithms, may prove a more effective base for implementing the kinds of constraints described above.

5. Coarse to Fine Processing in Neurons of Temporal Visual Cortex – An Introduction to Chapters 6 & 7

5.1. Introduction

The visual system is often described as a hierarchy of processing modules, each specialised to a particular attribute of the visual scene. However, more recent work (Sugase et al., 1999) has suggested that there may also be a *temporal* hierarchy of information processing in the visual system, with coarse information about a stimulus becoming available before the fine detail. Recording from neurons in inferior temporal cortex (IT), they found that the early part of a cell's response related only to the global stimulus category (e.g. face or non-face) with the more detailed information about the stimulus (e.g. facial expression) available only in the later (> 50 ms) component of the response.

The idea that processing in the visual system proceeds in a coarse to fine manner has been proposed by a number of authors (Carpenter & Grossberg, 1987; Schyns & Oliva, 1994; Ullman, 1995; Nowak & Bullier, 1997). Coarse to fine processing in terms of spatial frequency is proposed by Parker et al. (1992) to explain differential reaction times to sinusoidal gratings. Delorme et al. (2000) found that, in a rapid categorisation task, colour makes little difference to speed or accuracy, leading to a suggestion that the first wave of visual information is essentially low spatial frequency and achromatic, with colour and the high frequencies following later.

One potential physiological basis for these theories may lie in the differential processing speeds of the magnocellular (M) and parvocellular (P) pathways from the retina to area V1 of the cortex. The distinction between M & P pathways originates in the primate retina (for reviews see Milner & Goodale, 1985; Merigan & Maunsell,

1993) with two distinct classes of retinal ganglion cells. One class (the Pα cells) have large cell bodies and dendritic radiations, a transient response to visual stimulation and, because all three cone types make excitatory synapses on these cells, they are spectrally broad-band. In contrast, the second class of cells (Pβ) have medium to small sized cell bodies, small dendritic radiations, produce sustained responses and generally receive excitatory synapses from only one or two types of cones, thus conferring spectral sensitivity. The outputs from these two classes of retinal cells form two anatomically distinct pathways to visual cortex. The M pathway begins with Pα cells which project to the magnocellular layers of the lateral geniculate nucleus (LGN) and thence to layer 4Cα of cortical area V1. In contrast, Pβ cells project to the parvocellular layers of the LGN and thence to layer 4Cβ of V1.

The two systems can be said to transmit different regions of the "window of visibility" (Watson & Ahumada, 1985) in that the P system appears to provide greater spatial resolution, colour selectivity and responds to slowly changing stimuli. In contrast, the M system is effectively colour-blind, but is much more sensitive to rapidly changing stimuli (Merigan & Maunsell, 1993). It was once thought (Livingstone & Hubel, 1988) that the ventral visual stream was dominated by input from the P system, but more recent work (Ferrera, et al., 1994) has shown that, in V4, neurons may be driven by either system and there is an almost equal contribution by the two inputs across the whole population of V4 cells.

Consistent differences have been observed between the *latency* of the magnocellular and parvocellular pathways. In the LGN, the earliest magnocellular responses precede parvocellular responses by 10 ms (Marrocco, 1976; Maunsell et al., 1999) and, in V1,

the latencies in the parvo-recipient layer 4Cβ are about 20 ms longer compared to 4Cα (Nowak et al., 1995).

If an achromatic low-frequency signal about a stimulus reaches V1 prior to the colour high-frequency information, does the visual system compensate for this latency difference in the later stages of processing, or might this be a plausible explanation for the latency difference between global and fine information found by Sugase et al. (1999)?

The answer to this question has profound implications for the rapid serial visual presentation technique (RSVP), particularly at the high presentation rates we intend to employ for stimulus optimisation experiments. If colour and high frequency information is not reaching IT until 50 ms after the low frequencies, this would set a upper bound for presentation rate in an RSVP paradigm. Were we to go faster than that rate, we would either be measuring a component of the response relating only to the achromatic gross features of the stimulus or, even worse, a response to components of the current stimulus mixed up with the colours and fine detail from some previous stimulus.

The latter part of this thesis is concerned with exploring these ideas. The literature concerning colour and frequency is reviewed in the remainder of this chapter, particular those studies which make reference to issues of latency. The two experimental chapters that follow present neurophysiological experiments investigating the latency of colour (Chapter 6) and frequency (Chapter 7) information in a population of cells in IT.

It is worth explicitly pointing out that the experiments in the following chapters make

use of a limited and restricted set of stimuli, marking a departure from the methodologies discussed up to this point in the thesis. However, in the experiments that follow it is not necessary for us to know what the *optimal* stimulus for a cell is – as long as we can find an *effective* stimulus amongst the limited range of stimuli, then it will be possible to investigate the timing and characteristics of colour- and frequency-specific aspects of the response. This information should be valuable in shaping the design of future optimisation methodologies.

5.2. The Role of Colour

5.2.1. Evidence from Physiology

The extent to which visual neurons in IT cortex are sensitive to colour is not clear from the literature. In the macaque, Tanaka et al. (1991) found that colour was relevant for only about 10% of their sample and that proportion fell to about 7% for those cells whose receptive fields were considered to be elaborate (responding to a face or particular combination of shape and texture). An earlier study (Gross et al., 1972) had detected colour sensitivity in some cells in IT, but the extent of its importance was not accurately quantified.

A study specifically looking at colour sensitivity in anterior IT was carried out by Komatsu et al. (1992). The colour of simple geometric shapes was varied systematically and colour was found to influence the response of the vast majority of cells tested (~90%). Unfortunately, colour was only studied with those cells that could be driven by simple shapes, and cells responding to more complex stimuli (such as faces and natural scenes) were not tested.

A second study (Komatsu & Ideura, 1993) varied shape (simple geometric figures)

and texture in addition to colour. Again, a high degree of colour sensitivity was found, with 69% percent of neurons selective for colour (similar to the proportion found to be shape selective). There was no evidence of any interaction between shape and colour selectivity (i.e. the colour preference of a cell did not depend on the shape).

Perrett et al. (1982) recorded from face selective cells in rhesus superior temporal sulcus (STS) and noted that colour did not seem to play an important role, with only one out of 18 cells showing a reduction in response when the faces were viewed through a colour filter. However, there has been little systematic study of the effect of colour in neurons with complex stimulus selectivity (or indeed the latency of colour information), and we must look beyond physiology for further evidence of its role.

5.2.2. Lesion and Imaging Studies

In addition to the role that IT cortex plays in object recognition, there have been a number of studies examining the effect of lesions on colour processing. Heywood et al. (1995) found almost complete impairment of hue discrimination after ablation of the inferior temporal lobe in macaque monkeys, yet luminance discrimination was relatively spared. Similarly, Dean (1979) found no retention of a colour discrimination task after IT ablation, though after retraining, hue discrimination thresholds were found to be unaltered with respect to their preoperative levels. It is possible, however, that slight departures from isoluminance in this study may have allowed the monkeys to discriminate on this basis when relearning the task. Horel (1994) used cold-suppression of the dorsal aspect of IT to examine colour and form discrimination in trained macaques, finding disruption of colour discrimination even though form discrimination was spared.

Imaging studies also indicate that IT has a role in the processing of colour. Takechi et al. (1997) used PET to look at the cortical areas involved in colour, luminance and positional discrimination (again in the macaque). Using simple square stimuli, they found significant activation in the posterior part of IT cortex in the (colour – brightness) and (colour – position) subtraction pairs.

fMRI imaging has been used in the human (Zeki & Marini, 1998) to look at the areas involved in colour processing with more natural stimuli. Their stimuli included common objects and landscapes that were presented in full colour, achromatic or false colour conditions. The (colour – achromatic) subtraction resulted in an area of activation extending anteriorly *beyond* V4 in the temporal lobe into areas the authors suggest may be analogous to monkey IT. Interestingly, this activation is not found in the (false – achromatic) subtraction, suggesting that the role of this inferior temporal area anterior to V4 is specific to the role colour plays in object recognition, rather than simply processing colour in a more abstract capacity.

Together, the above studies suggest that colour processing is an important property of inferior temporal cortex in both the monkey and human, though the extent to which this is specific to object processing is unclear (particularly in human).

5.2.3. Psychophysics

Psychophysical studies provide the majority of evidence that colour information is delayed with respect to form in the visual system, though there has been a great amount of debate in the literature about the precise role it plays in object recognition.

Several studies have indicated that colour is unimportant for object recognition, with subjects responding just as quickly and accurately to black and white line drawings or

photographs when compared with colour photographs (Ostergaard & Davidoff, 1985; Biederman & Ju, 1988; Davidoff & Ostergaard, 1988). However, there been a greater number of studies which do not support this position, finding that colour and detail consistently enhance the speed of reaction or accuracy (Price & Humphreys, 1989; Wurm et al., 1993; Humprey et al., 1994; Lee & Perrett, 1997).

There are major methodological differences between all the studies and a potential explanation for the conflicting data is given by Price & Humphreys (1989), who suggest that no advantage for colour is found when the subjects are required to make *very fast decisions* or the stimuli are masked. This contrasts with the situation when subjects can react *in their own time* where a consistent advantage for colour is found. Perhaps subjects reacting quickly have access only to coarse representation, which does not include colour?

Delorme et al. (2000) required both monkey and human subjects to make a rapid categorisation (food or non-food / animal or non-animal) of briefly presented (32 ms) stimuli, which were either colour or achromatic photographs. Colour was found to make little difference in terms of either accuracy or reaction time in the majority of subjects. Interestingly, they noted that there was a strong correlation between the accuracy impairment for achromatic images and reaction time, with the slowest reacting subjects showing the highest impairment with achromatic images and the faster subjects categorising equally well in both conditions. Like Price & Humphreys (1989), they suggest that colour is not an important cue when reactions must be made quickly, but that it can be used as a relevant feature when a subject takes longer to respond.

Chapter 6 will therefore set out to answer two questions – firstly to what extent are cells in IT sensitive to colour, particularly with complex or naturalistic objects and, secondly, does colour information arrive quickly enough to contribute to a response in a rapid presentation paradigm?

5.3. The Role of Spatial Frequency

5.3.1. Evidence from Physiology & Lesion Studies

Rolls et al. (1985) varied the spatial frequency composition of facial stimuli while recording from neurons in STS of the macaque. They found a wide variety of frequency tuning curves, but noted that the majority of neurons tested would respond well both to low-pass filtered faces, containing only these frequencies below 8 cycles per faces, and to high-pass images, containing only frequencies above 8 cycles per face. This fits well with the results of Fiorentini et al. (1984) who found that spatial frequencies above or below 8 cycles/face are adequate to recognise different individuals' faces. Rolls et al. (1985) did not attempt to establish whether there was any differential latency between the cells' responses to low- and high-pass filtered images.

However, evoked potential studies in the human do provide evidence for spatial frequency dependent changes in latency (Parker & Salzen, 1977a, 1977b; Parker & Salzen, 1982). The time to peak of a visual evoked response triggered by the onset, contrast reversal, or offset of a sinusoidal grating was measured, and found to increase with increasing stimulus spatial frequency. The magnitude of this increase was generally found to be around 20ms/octave, and persisted even when the stimuli are matched for apparent contrast.

If this increase in latency is mediated by differential speeds of the M and P pathways, then is it possible to establish a relationship between stimulus spatial frequency and processing latency by taking into account the spectral sensitivity tuning curves of the two different channels? Unfortunately, this is not straightforward. At a given eccentricity, neurons in the two pathways have almost identical spatial resolution (Crook et al., 1988) and there is a great deal of overlap in the spatial frequency tuning of the two pathways (Shapley & Lennie, 1985). However, lesion studies in the LGN of the macaque (Merigan & Maunsell, 1993) have provided a somewhat clearer picture. M pathway lesions (Mergian et al., 1991a) allow the characteristics of the P system to be studied in isolation, showing that its spatial frequency sensitivity peaks at around 1 cycle per degree. In contrast, P pathway lesions (Merigan et al., 1991b) show the M system has a much lower peak spatial frequency sensitivity at around 0.1 cycles per degree. While there is considerable overlap between the spatial frequency response of the two systems, spatial frequency response over 10 cycles per degree is entirely dependent on the P system. A possible explanation for these findings, despite the similar spatial resolution of individual cells mentioned earlier, probably reflects the higher sampling density of the more numerous P pathway cells (Merigan et al., 1991b).

5.3.2. Evidence from Psychophysics

The processing delays for high spatial frequencies implied by evoked potential studies appear to have clear psychophysical correlates. Reaction times (RTs) to sinusoidal gratings (Breitmeyer 1975; Parker 1980; Parker & Salzen, 1982) have been found to increase monotonically with the spatial frequency of the grating, with estimated delays ranging from 20 – 32 ms per octave increase in spatial frequency. However, at

least one study (Barr, 1983) found no significant difference between RTs between high and low frequency gratings.

A matching task paradigm (Parker & Dutch, 1987) attempted to more accurately quantify the delay by removing the variability present in earlier reaction time tasks. Subjects were instead required to adjust a visual probe such that its onset was perceived as being simultaneous with that of a grating. In this case, a much smaller delay of around 7 ms per octave was found. Parker & Dutch (1987) suggest that a latency lag of this order of magnitude could have been swamped by motor variability in earlier studies, and this may account for the failure of Barr (1983) to find any significant difference in latency.

However, it is clear from the studies above that any mapping of this pattern of spatial frequency processing delays onto the parvocellular and magnocellular systems requires delays to be found *within* as well as *between* the two systems (Parker, 1980). For instance, the gratings used by Parker & Dutch (1987) lie entirely within the range (as determined by lesion studies) of both the P & M pathways, which would allow either pathway to mediate the responses. Despite this, there is a steady increase in both reaction and perception time as spatial frequency increases.

This would seem to support a view of the visual system where there is a *gradual* progression in the availability of spatial frequency information, from coarse to fine. If this is the case, then we might ask how recognition systems integrate the different waves of spatial frequency information. Two distinct possibilities have been suggested, which are termed the temporally anisotropic and temporally isotropic models (Parker et al., 1992). In the temporally anisotropic model, the integration

process is dependent on the availability of information progressing from coarse to fine, with the earlier and coarser information used to prime and contextualise later waves of increasingly fine detail. In contrast, the temporally isotropic model assumes that there is no inherent bias in the integration process, information is simply integrated as and when it becomes available.

A number of studies have attempted to establish which model is correct, by presenting sequences of spatial frequency filtered images in rapid succession, and asking subjects to rate the perceived quality of the image sequence as a whole (Parker et al., 1992; Parker et al., 1997; McSorley & Findlay, 1999). In these studies, subjects rated more highly image sequences where the order of presentation swept from low to high frequencies compared to those sequences where the order of presentation ran from high to low. This was interpreted as strong evidence that spatial frequency integration is temporally anisotropic.

These studies would seem to strongly support the view that spatial frequencies are processed and integrated in a coarse to fine manner, and that the differential time courses of global and fine information found in neurons in IT may be a reflection of this integration process, with the global information acting as a "header" to set the context in target areas for the fine detail information that is about to follow (Sugase et al., 1999). It is worth pointing out that global information, as defined in the study of Sugase et al. (1999), may simply equate to the low spatial frequencies, since these would be sufficient to make the "global" categorisation between human face, monkey face or non-face images, but not the "fine" categorisation between different facial expressions, which would require the higher spatial frequencies.

If this is the case, then it is interesting to ask how the visual system reacts when images are presented in very rapid succession. The evidence presented above would seem to suggest that neurons in higher visual areas, such is IT and STS, would only have time to respond to the information that was first to arrive, corresponding to the coarse stimulus information carried by the low spatial frequencies. Inherent in this view, however, is an assumption that the late arriving fine detail information about one stimulus is suppressed by coarse information from the following stimulus (see figure 5.1a).

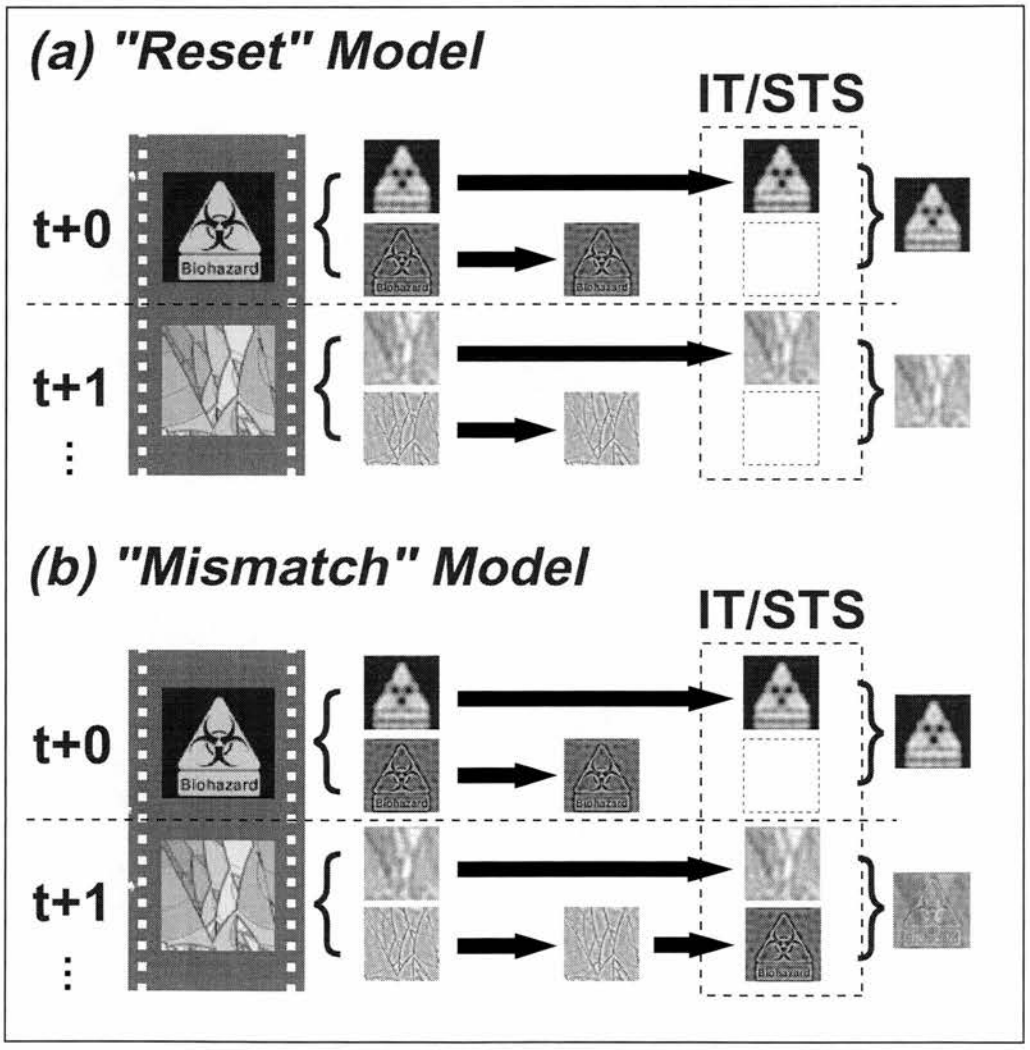


figure 5.1 – At high presentation stimulus rates, the relatively slow processing rate of high spatial frequencies might either result in (a) "Reset" - only the low frequencies reach higher visual areas, since change of stimulus generates a signal which resets the visual processing system. (b) "Mismatch" - low and high spatial frequencies of different stimuli are inappropriately combined in upstream integration areas.

If the visual system lacks a "reset" signal of this type, then different processing speeds of each spatial frequency channel would allow the potential for a "mismatch" between different aspects of stimuli, particularly at high processing speeds, when the information is combined upstream in areas such as STS and IT (see figure 5.1b).

The aim of the experiment described in Chapter 7 is to first establish whether low and high spatial frequency information arrives with differing latencies in higher visual neurons, after which the above models will be considered further.

6. The Role of Colour in the Response of Neurons in Temporal Visual Cortex to Complex Stimuli

6.1. Experimental Hypothesis

The work summarised in Chapter 5 suggests that the colour information is used by the visual neurons in temporal cortex, but the extent of its importance in the processing of complex visual stimuli is unclear. The evidence also suggests that the colour signal may be delayed with respect to the luminance signal. This leads us to the following experimental hypothesis:

As stimulus presentation rate increases, any response due to colour information in the stimulus will be progressively diminished because the colour signal will not arrive in time to contribute to the response, resulting in an equal response to achromatic and colour stimuli at the highest presentation rate.

6.2. Methods

6.2.1. Recording Techniques

The subject (male *Macaca mulatta*, age 6 years) was seated in a primate chair and head restrained. Neural signals were recorded using standard methods (Oram & Perrett, 1992). Neurons were localised to the upper and lower banks of the superior temporal sulcus and inferotemporal cortex (see figure 6.2, Appendix 3 for details). The subject's eye position was monitored (accuracy ±1°; IView, SMI, Germany). A 486 PC and Cambridge Electronics CED 1401 interface recorded eye position, spike arrival times and measured stimulus onset times.

6.2.2. Stimulus Presentation

Stimuli (256 x 256 pixels) were presented centrally on a Sony GDM-20D11 monitor

(72Hz refresh rate, image size 10°x 10°) which was attached to an Indigo2 Silicon Graphics workstation. Stimuli were presented against a black background. Onset and duration of the stimuli were measured using light sensitive diodes on the monitor screen. If the measured stimulus duration differed from the intended duration, the data for that stimulus sequence was discarded. Sequence presentation commenced when the subject's gaze remained within a fixation window ± 5° of the monitor centre for >500 ms and terminated if the subject's gaze moved outside the fixation window. Fixation was rewarded with fruit juice delivery. Activity relating to the first and last image of each sequence was discarded.

6.2.3. Visual Stimuli

The stimulus set consisted of 38 colour images (256 x 256 pixels) including photographs of human and monkey heads, animals, everyday objects and abstract figures. Monkey (two individuals) and human (one individual) head images formed complete rotational sequences around the head with 45° spacing, such that 0° was the front (or facial) view of the head and 180° was the rear. The full set of stimuli can be seen in Appendix 1, with those stimuli considered to be abstract (and treated separately in the results) marked as such.

False-colour and achromatic versions (figure 6.1a) of each image were prepared as outlined below. First, images were transformed to YCbCr colour space, which has separate luminance (corresponding to the CIE **Y** primary) and chromaticity components (Bhaskaran & Konstantinides, 1997).

forward transform	back transform
$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$	$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.0 & 0.0 & 1.4021 \\ 1.0 & -0.3441 & 0.7142 \\ 1.0 & 1.7718 & 0.0 \end{bmatrix} \begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix}$

Achromatic images were generated simply by setting the chromatic components (Cb & Cr) to zero for each pixel, followed by a transform back to RGB. This process always produces a valid RGB triplet.

False-colour images were prepared by reflecting the colour of each pixel in turn about the origin of the chromatic (CbCr) plane, keeping Y constant. The process often generates invalid RGB triplets which correspond to colours that cannot be produced on a standard monitor (e.g. very bright pure blue). When this occurred, points were moved back towards the origin of the chromatic plane until displayable colours were obtained (effectively reducing the saturation of the colour). The false-colour algorithm (and its implementation in the C programming language) is contained in an appendix.

A digital photometer (Tektronix, Model J6523-2) was used to test the success of these transformations in maintaining both overall image luminance and contrast edges within images. Overall image luminance was measured by placing a perspex diffusion plate between the computer screen and photometer. Contrast borders were tested by individually measuring a series of colour patches (diameter 1°) before and after image transformation. The process was judged to be satisfactory for the purposes of the experiment, with most measured luminances falling within ± 10% of the pre-transformation levels (figure 6.1, b & c).

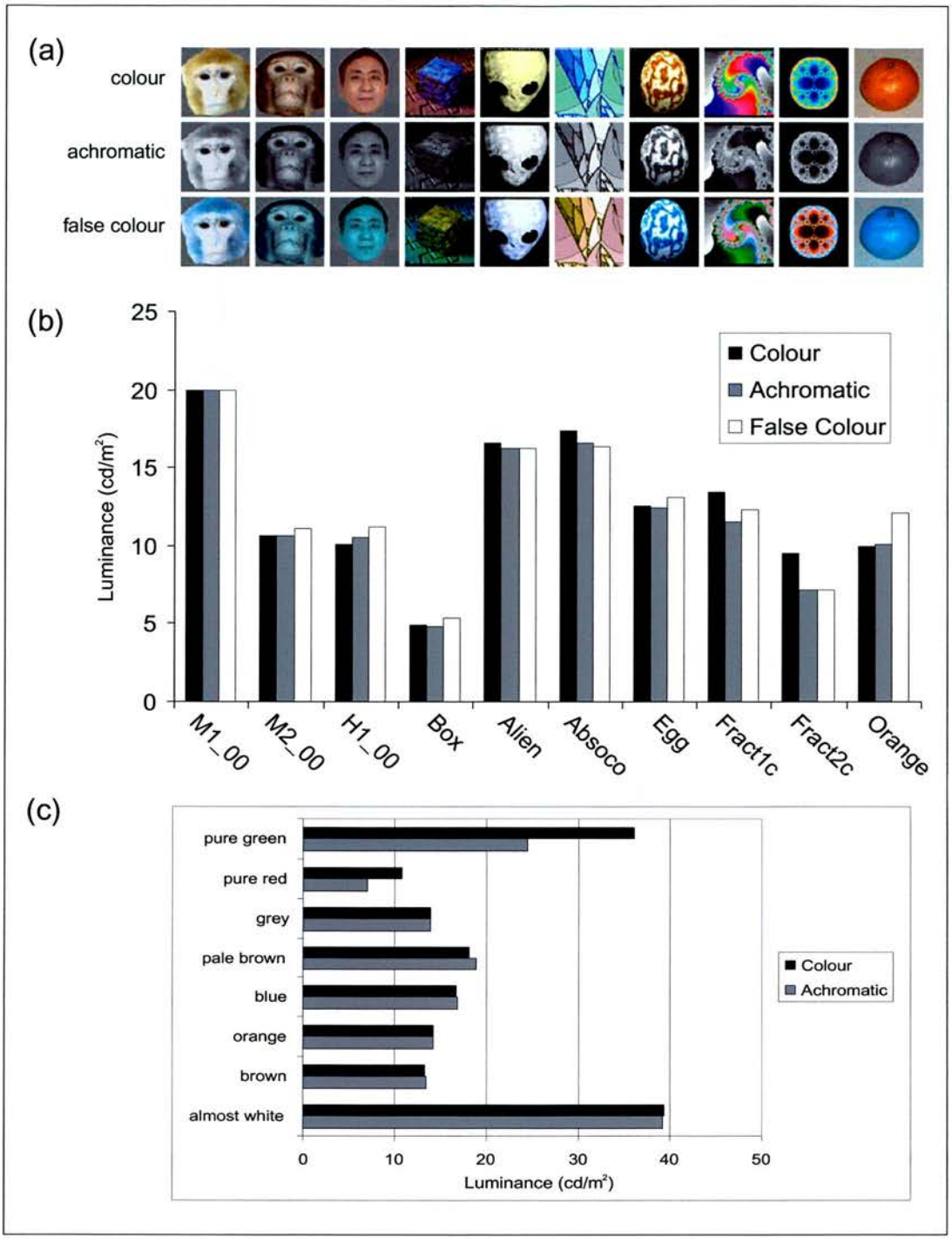


figure 6.1 – *(a)* Colour, achromatic and false-colour versions for a selection of the images used. *(b)* Average image luminance for each of these stimuli and their colour transformed counterparts. Image Luminance was measured using a Perspex diffusion plate placed in front of the monitor. *(c)* Luminance measurements for small patches (circle, diameter 1°) within the images, with a description of the approximate shade.

6.2.4. Procedure

(a) Screening Phase

On detection of a neuron, we first conducted a screening test to establish whether any of the stimuli were sufficient to evoke a visual response.

In Experiment 1, each cell was tested using only the 38 colour images. Stimuli were presented in a random sequence with a stimulus presentation time of 111 ms (8 frames) and no gap between stimuli. Each stimulus was repeated several times during the screening phase (median 30, range 11 – 58). 32 cells were found to have a preference for particular stimuli and were further tested. Preference was judged by eye from the set of peri-stimulus time histograms (PSTHs) computed online.

Experiment 2 differed only in the number of stimuli used for the screening phase, each cell being tested with the 38 achromatic images in addition to the original 38 colour images. This was intended to remove any potential bias towards colour sensitive cells in the sample. Due to the increased number of stimuli present, there was a lower number of stimulus repetitions during this screening phase (median 20, range 9 – 34). 18 cells showed stimulus preference and were tested further.

(b) Colour Response Test

A total of 50 cells were found to have a preference for at least one of the stimuli during the screening phase. For each cell, we selected its "best" and "worst" images (those producing highest and lowest responses respectively) from the screening set, along with 3 other images that had produced intermediate responses. This was intended to reduce the contamination produced by adjacent stimuli since it has been shown that the neural response to a stimulus typically outlasts stimulus presentation

time by about 60 ms (Keysers et al., 2001).

A cell was then tested using a set of stimuli for each of the images chosen above. Each set included the colour image, achromatic and false-colour transformed versions, as well as an additional set of 4 frequency-filtered colour and achromatic versions of the image that will be discussed in full in the next chapter. In total, the complete test set was comprised of 35 stimuli per cell.

For the first 11 cells in Experiment 1, no false-colour images were present in the test set. These cells were tested using only 6 versions of each image, with a total test set of 30 stimuli per cell.

The stimuli were presented in 4 different stimulus duration conditions (table 6.1), interleaved in blocks of a single presentation rate. Within a block, stimuli were presented in pseudo-random order, with the constraints that no 2 identical stimuli were ever presented in immediate succession and each stimulus was presented an equal number of times. Each block contained a number of repetitions of each stimulus, such that the length of each block was the same, and thus the total presentation time was identical for each of the four conditions (i.e. there were more repetitions of the shorter duration conditions). The presentation rate blocks were also

Table 6.1 – Experimental Conditions

Condition	*Stimulus presentation time*			*Time taken for 1 complete cycle (s)*	*Stimulus repeats per block*
	Stim. (ms)	*Gap (ms)*	*Total (ms)*		
S56G167	55.6	166.7	222.2	7.8	3
S56G0	55.6	-	55.6	1.9	12
S28G0	27.8	-	27.8	1.0	24
S14G0	13.9	-	13.9	0.5	48

randomly ordered. The median number of stimulus repeats in the slowest condition (S56G167) was 17 (range 6 – 42) and the median number of stimulus repeats in the fastest condition (S14G0) was 209 (range 119 – 333).

Condition S14G0 only took place when a sufficiently strong response was obtained and was often carried out post-hoc in isolation, after the data from the other conditions had been examined. Only 7 cells were tested at this rate.

6.2.5. Response Analysis

A cell's response to a particular stimulus in the sequence was calculated by aligning segments in the continuous recording which contained an occurrence of the stimulus. Each segment lasted from 250 ms before stimulus onset to 550 ms after stimulus offset. The peri-stimulus time histogram (PSTH) was generated by summing across all the aligned segments, and represented the response triggered by that particular stimulus against a background of activity evoked by all the surrounding stimuli.

"Best" and "worst" stimuli for a cell at the end of the *screening phase* were simply judged by eye from the set of PSTHs calculated on-line and were selected (along with 3 intermediate images) for the colour response test.

After the *colour response test*, the sets of stimuli were ranked from "best" to "worst" separately for each condition, based on the cell's response to the colour stimuli. This ranking and the cell latency was calculated as follows –

First, the responses were summed across trials (bin-size = 1 ms) and smoothed (gaussian, σ = 20 ms). A control period was defined as the 200 ms preceding stimulus onset. The latency of response onset was measured as the first 1ms time bin at which

the firing rate exceeded the mean + 2.58 σ (i.e. $p < .005$) of activity during the control period, for at least 15 consecutive bins (i.e. 15 ms). Where this criterion was not met, a fixed latency of 100 ms was assumed for that stimulus. If a latency could not be detected for any of the stimuli, the cell was excluded from the analysis of that condition.

Next, the response to each stimulus was measured in a time window starting with the latency measured above and lasting for the length of the stimulus + half the gaussian width of the smoothing filter (10 ms). Stimulus sets were ranked according to this windowed response, and cell latency was defined as being the onset time of the maximum response.

6.2.6. Population Analysis

A spike density function (SDF) was calculated from the raw spike counts for every cell and stimulus by smoothing with a gaussian (σ = 5 ms). A single normalising factor for each cell was calculated as the maximum value of the SDF for the colour version of the best stimulus. Colour, achromatic and false-colour responses to the best stimulus were weighted by this factor, so that every cell would have an equal contribution in the population response, with the colour response acting as baseline.

Finally, population curves were calculated for each condition as the average SDF for colour, achromatic, and false-colour versions of the best stimulus.

For the latency aligned population curves, an additional step took place prior to averaging, with each SDF shifted in time such that time 0 reflected the detected cell latency, as measured above.

6.3. Results

6.3.1. Stimulus Preference

Stimuli were categorised as being abstract if no natural colouring existed for that image (see Appendix 1). Those stimuli which had natural colouring were broken down into faces (which included all head views) and non-faces. The category of the preferred (or "best") stimulus for each cell is shown in table 6.2. Where a cell's preferred stimulus was not consistent across presentation rate conditions, or if no response was obtained for any condition (excluding S14G0 since only 7 cells were tested at this rate), the cell was categorised as having unknown preference.

A list of stimulus preferences for each individual cell is presented in Appendix 2, and approximate electrode tracks and cell locations are reconstructed in Appendix 3. These are also summarised in figure 6.2.

We first consider the effect of colour on the whole population of cells tested (experiments 1 & 2 combined). Although it seems unlikely that screening with colour stimuli alone (experiment 1) introduced a sampling bias in favour of chromatically tuned cells (since luminance information is still present in these colour images), this point is considered later in this section.

Table 6.2 – Preferred Stimulus Category

Preferred Stimulus Category	*Number of cells*	*Percentage*
abstract	6	12
non-abstract (face) (non-face)	26 (22) (4)	52 (44) (8)
unknown	18	36
total	50	100

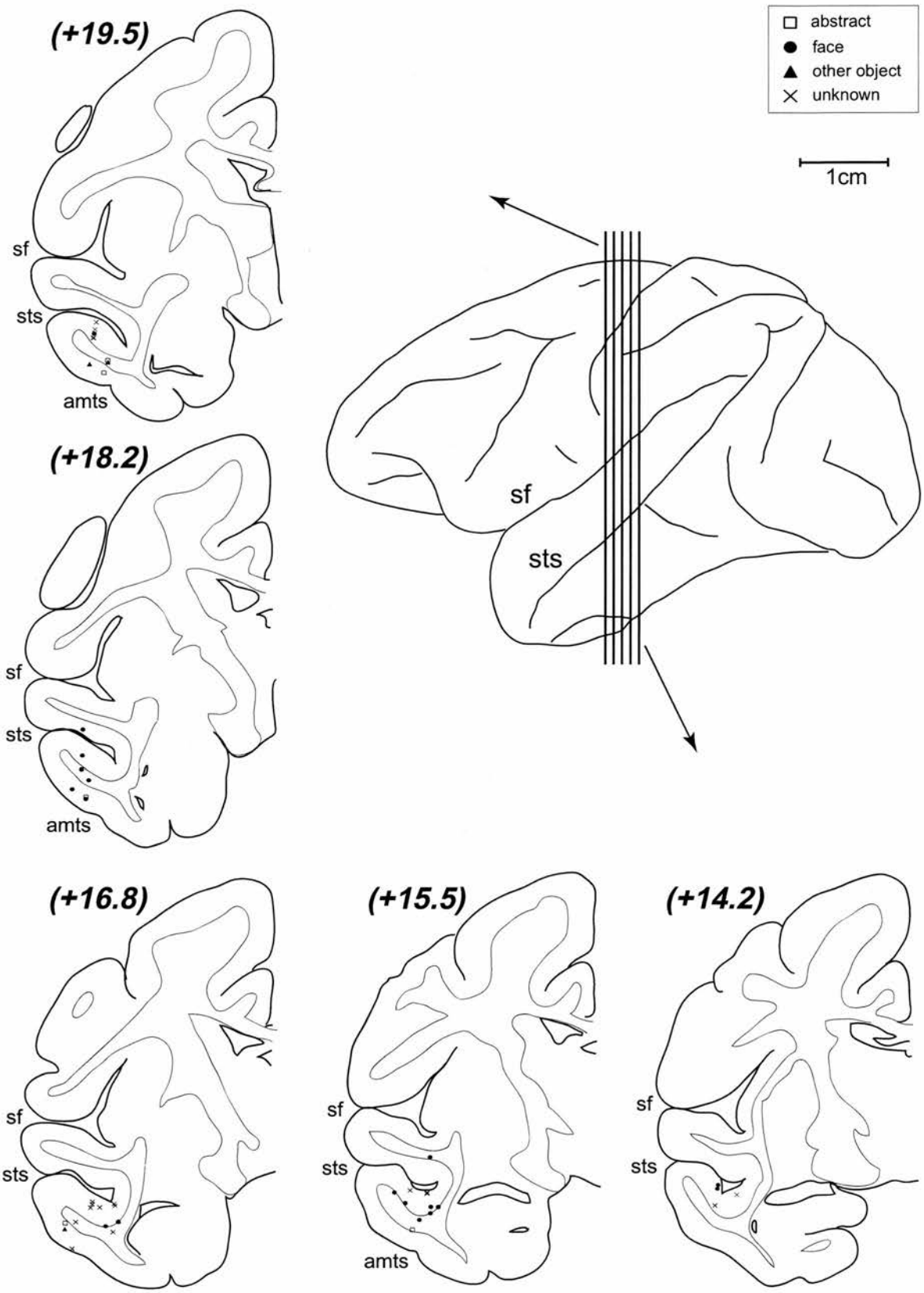


figure 6.2 – Coronal sections showing the location and stimulus preference of cells tested in this, and the following, chapter. *sts* – superior temporal sulcus, *amts* – anterior middle temporal sulcus, *sf* – sylvian fissure. Bracketed figures indicate approximate stereotaxic position on the anterior – posterior axis.

6.4. Whole population analysis

6.4.1. As a population, cells in STS/IT are strongly colour-tuned

Neurons in STS/IT exhibit a high degree of chromatic tuning in addition to their shape tuning, as illustrated by the response of a single cell (figure 6.3) and the population response to the best stimulus (figure 6.4). Achromatic versions of the best stimulus produce, at a population level, much weaker responses than the original colour images. The reduction in response is even greater when stimuli are falsely coloured, suggesting that these cells not only have a preference for certain colour profiles, but this tuning extends to inhibition of the shape response, when an incorrect colour profile is present. This initial qualitative description is backed by a formal statistical treatment of the colour and shape responses in the following sections.

The population includes cells with a wide range of response latencies (median 91, range 58 – 141 ms) and we can see the extent of chromatic tuning even more clearly if this variation in response latency is removed by aligning the data from each cell on the cell latency (figure 6.4b).

The stimulus aligned population histogram shows what looks like a clear latency difference in the population response, with colour images producing earlier responses than achromatic or false colour images. However, when the responses are aligned on cell latency, this effect disappears, implying that the earliest responding cells are particularly colour sensitive, with little response to the achromatic and false colour versions of their preferred stimulus. This relationship between cell latency and colour sensitivity is examined in a later section of the results, where it will be considered whether it is reasonable to treat the cells in this study as a single population.

If we are correct in our hypothesis that the colour signal will be delayed with respect to the luminance signal, then given the effects noted above, we should see two separate effects in the population response –

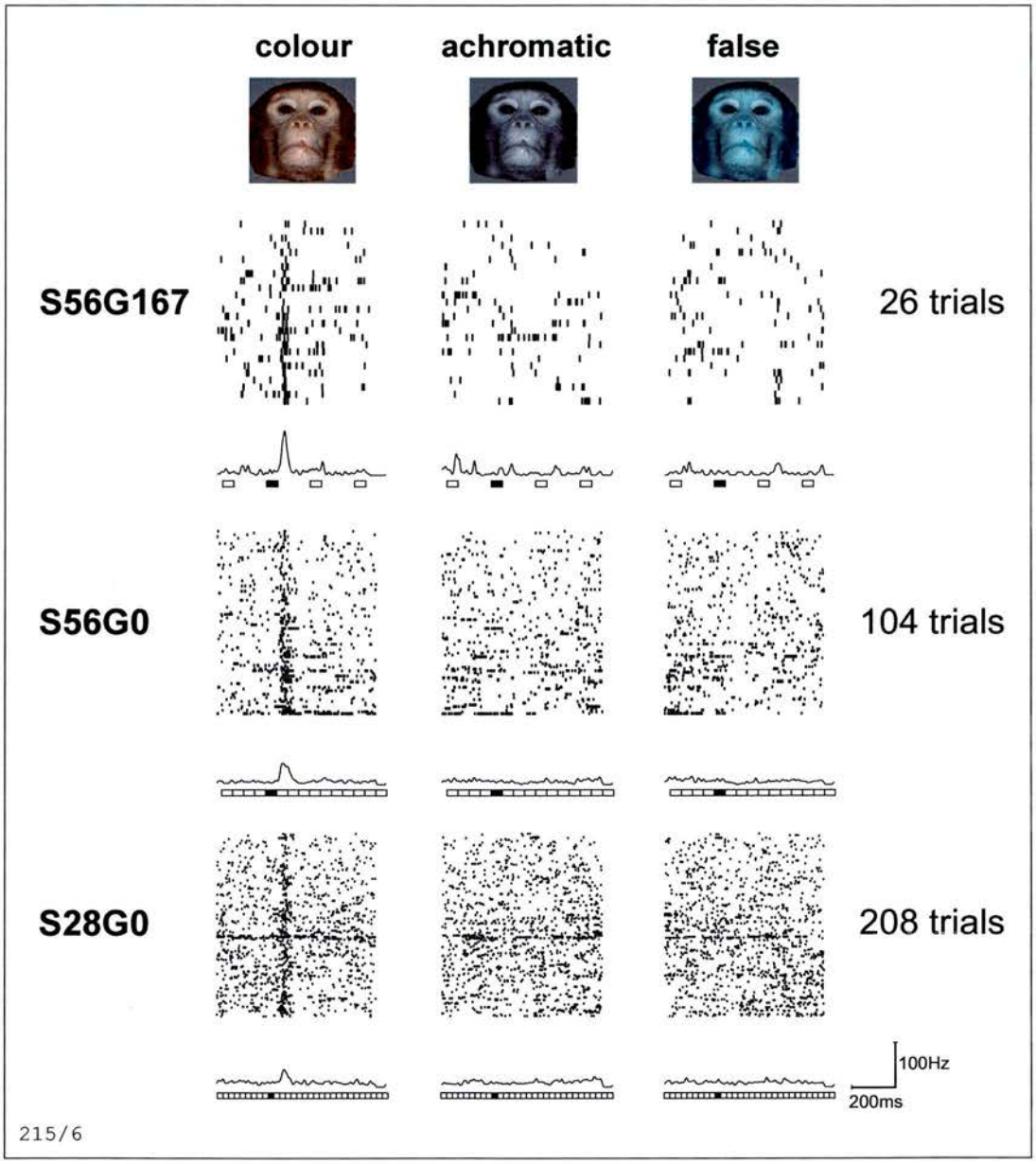


figure 6.3 – Response of a single cell to colour, achromatic and false colour versions of its best stimulus in different presentation rate conditions (56 ms + 167 ms gap top, 56 ms middle, 28 ms bottom). This particular cell was not tested at the fastest rate of 14 ms. SDFs were created by summing the data across trials and smoothing with a gaussian (σ = 10 ms). Stimulus presentation is shown by a filled rectangle below the SDF and hollow rectangles show previous or subsequent stimulus presentations. The latency of this cell (66ms) was at the lower end of the cell population tested. Scale is indicated in the bottom right corner.

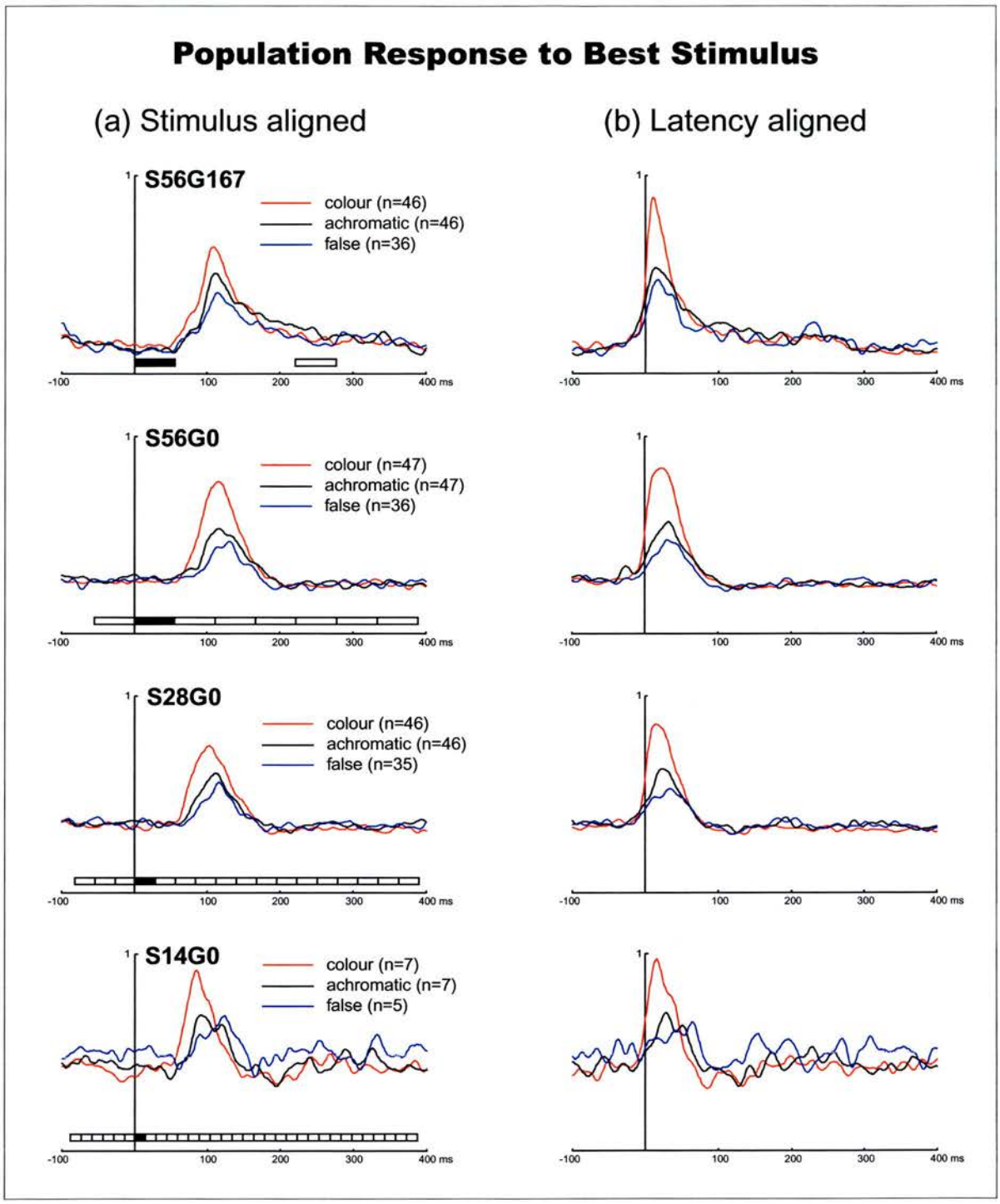


figure 6.4 – *(a)* Stimulus-aligned and *(b)* Latency-aligned normalised average population responses to colour, achromatic and false-colour versions of the best stimulus for each cell. Slowest presentation rate at the top, fastest at the bottom. Filled rectangle indicates stimulus presentation, outlined rectangles show earlier or subsequent stimulus presentations. Each set of responses for a cell was aligned to the detected latency of the colour response, as described in methods. n's indicate the number of cells contributing to each SDF.

(i) Colour, achromatic and false-colour curves should be initially similar, then diverge, reflecting the delayed contribution of colour-specific information to the response

(ii) Any difference between the curves should be progressively reduced as presentation rate increases, since the appropriate colour information will not arrive in time to contribute to the response

In fact, neither of these effects is seen. Preference for the colour version of the stimulus appears almost immediately at the start of the population response (figure 6.4b) and this preference is consistent across the different presentation rate conditions, and is clearly apparent even at the fastest presentation rate of 14 ms.

6.4.2. Inhibitory responses to stimuli

The worst stimulus for a cell (i.e. that stimulus producing the lowest activity during response assessment) often resulted in inhibition, with the response falling below baseline (figure 6.5). This inhibition was also dependent on stimulus colour, with achromatic and false colour versions of a stimulus resulting in less inhibition than the original colour image. This effect is consistently seen across the different presentation rate conditions.

Due to the technique we are using, the baseline is the average response to all stimuli presented, rather than a measure of the true spontaneous firing rate of the cell. The periods before and after the presentation of a particular stimulus will therefore contain a fraction of the spikes relating to *all other* stimuli. We would therefore expect the worst stimulus to produce a drop in firing compared to the baseline simply because, during the presentation of that stimulus, we are not measuring that fraction of the

spikes relating to effective stimuli.

This is, however, not sufficient to account for the pattern seen – firstly, because there is a consistent pattern in the responses to worst stimulus and secondly, because the ordering (of colour, achromatic and false colour responses) seen with the best stimulus is perfectly reversed with the worst stimulus, as would be expected in the case of a colour-tuned inhibitory response.

6.4.3. Time course of colour and shape tuning – population statistical analysis

The time course of shape and colour tuning was accurately established using a sliding window statistical test to measure the probability of discrimination between stimuli as a function of time.

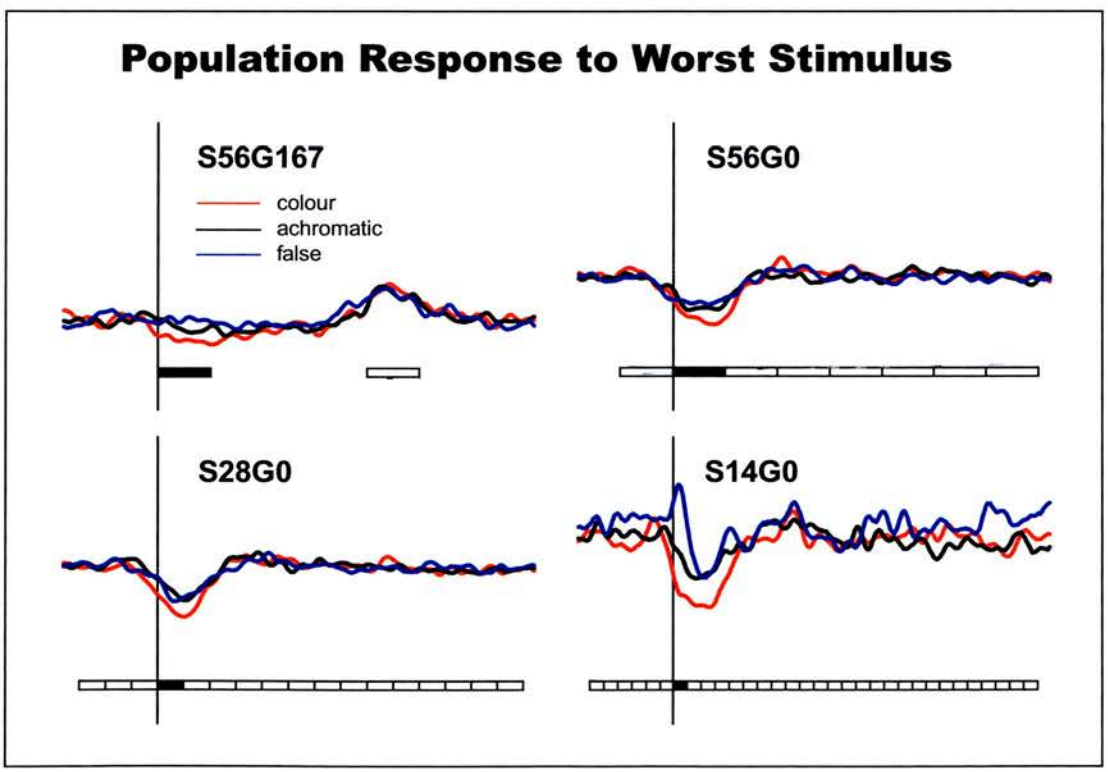


figure 6.5 – Latency-aligned normalised average population responses to colour, achromatic and false-colour versions of the worst stimulus for each cell. Each set of responses for a cell was aligned to the detected latency of the colour response to the best stimulus, as described in methods. Filled rectangles indicate stimulus presentation, outlined rectangles show previous or subsequent stimulus presentations. Note that the responses have been aligned to cell latency, such that response onset coincides in time with the onset of the stimulus evoking the response.

For each cell, and each presentation rate, normalised and latency-aligned SDFs were calculated as previously described. Shape tuning was then determined by comparing the SDFs for the *achromatic* stimuli with related one-way ANOVA (5 levels, corresponding to the five different stimuli), performed separately for each 1 ms time bin. Colour tuning was measured in a similar manner, with a one-way ANOVA (3 levels) comparing the responses to colour, achromatic and false-colour versions of the best stimulus for each cell. The analysis was restricted to the subset of cells for which we had collected false-colour responses. There was an insufficient number of cells tested at the 14 ms presentation rate to perform the analysis for this condition.

The results of this analysis can been see in figure 6.6. Shape and colour discrimination have an almost equal onset time, with shape leading colour by not more than 5 ms at the uncorrected $p = .01$ level (lower dashed line) and no difference when considered the same criterion level with Bonferrori correction (upper dashed line). The Bonferroni level was calculated by dividing the criterion level ($p = .01$) by the number of time bins over which the analysis was performed (500) and is, in fact, over-corrected since there is a high degree of correlation between consecutive time bins due to firstly the response properties of the cells themselves and secondly due to the smoothing procedure used to create the SDFs.

If we consider the overall pattern of colour and shape discrimination, the data appears to suggest that, in fact, the colour signal dominates the earlier part of the response, with optimal colour discrimination peaking before that of shape by 10 – 20 ms. The latter part of the response is dominated by shape discrimination, which outlasts colour discrimination by around 20 ms in the zero gap conditions.

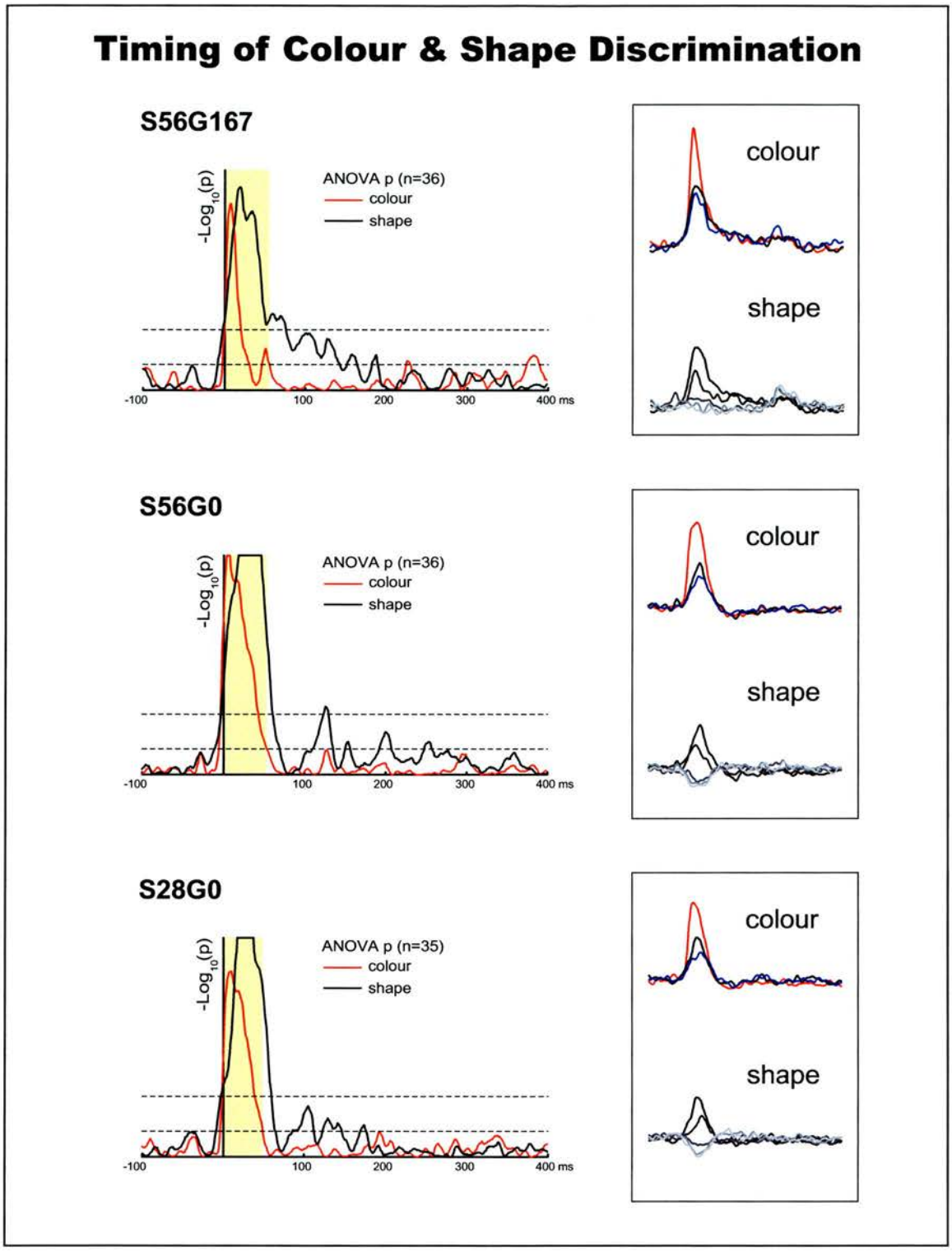


figure 6.6 – Time course of colour and shape discrimination, measured using sliding window ANOVAs. SDFs were calculated for each stimulus, normalised and latency aligned (see method). Shape tuning was determined by a related one-way ANOVA (5 levels) and compared responses to the achromatic stimuli (insets, bottom – darkest line = best stimulus, lightest line = worst). Colour tuning was measured in a similar manner, with a one-way ANOVA (3 levels) comparing responses to colour, achromatic and false-colour versions of the best stimulus for each cell (insets, top – colour = red, achromatic = black, false = blue). This analysis was performed separately for each 1msec time bin. Probability is plotted on a negative log scale, and curtailed at $-\text{Log}_{10}(p)=17$. Lower dotted line shows significance at the p = .01 level (uncorrected). Upper dotted line shows the same level Bonferroni corrected. Yellow shaded area indicates duration of colour discrimination at the uncorrected level and is used to define the window in the cell by cell analysis. [Software used to perform the analysis: Mathematica 4, Wolfram Research].

This pattern is particularly evident in slowest presentation condition, where a 168 ms gap is present between each stimulus. During the entire gap period, the response continues to carry information on the shape of the previous stimulus and this is extinguished only when the next stimulus is presented. However, this "visual memory" for stimulus shape appears to be colour blind, since colour discrimination is as short lived as in the conditions where no gap exists.

6.4.4. Colour tuning of individual cells

The colour responses of individual cells were assessed by comparing windowed spike counts for each individual trial for the best colour, achromatic and false-colour (when presented) stimuli. The window was determined from the population analysis above, and was calculated separately for each condition as follows. The window began with the first 1 ms time bin when colour discrimination exceeded the $p = 0.01$ (uncorrected) level. It ended when population colour discrimination fell back below this criterion level, with the constraint that the window was at least as long as the stimulus presentation time.

Colour sensitivity was measured with a one-way ANOVA with 2 or 3 levels corresponding to colour, achromatic and false colour responses (false colour responses were only measured in a subset of the cells).

The variance of spike counts has previously been found to be approximately proportional to the mean response (Tolhurst et al, 1981; Dean, 1981), therefore it was first necessary to perform a square-root transform on the raw spike counts prior to the statistical tests.

Table 6.3 shows the proportion of cells where a significant effect of colour was found for each condition. 70% of the cells in our population showed a significant effect of

Table 6.3 – Colour Tuning as a Function of Presentation Rate

Condition	*Stimulus presentation time (ms)*	*Gap time (ms)*	*n*	*Mean Response onset (ms)*	*Window duration (ms)*	*Percentage of neurons discriminating colour* [a]
S56G167	55.6	166.7	46	94	51	41
S56G0	55.6	-	47	94	59	68
S28	27.8	-	46	83	58	65
any	-	-	50	-	-	70 [b]

[a] Percent of neurons with a significant ANOVA testing the effect of stimulus colour on neuron response ($p < .05$)
[b] Percent of neurons with a significant effect of colour in at least one condition ($p < .05$, Bonferrori corrected)

colour in at least one of the conditions ($p < .05$, corrected). This breaks down into 72% of the cells screened using just colour stimuli (experiment 1) and 67% of the cells screened using colour and achromatic stimuli (experiment 2). This difference was not significant ($\chi^2 = 0.70$, d.f. = 2, n.s.) and it therefore seems unlikely that there was any substantial sampling bias towards colour sensitive cells when screening took place with colour stimuli alone (experiment 1).

6.4.5. Colour sensitivity index

In order to assess whether there was any relationship between colour sensitivity and cell latency, a colour sensitivity index was calculated for each cell –

$$colour\ sensitivity\ index = \frac{\bar{c} - \bar{a}}{\bar{c} + \bar{a}}$$

where $\bar{c}$ was the mean spike count for the best colour stimulus and $\bar{a}$ was the mean spike count for the achromatic version of the best stimulus. The window was

calculated separately for each condition as described in the previous section.

The colour sensitivity index ranges from +1 (preference for colour stimuli) to -1 (preference for achromatic stimuli) with a value of 0 indicating that a cell responded identically to the colour and achromatic stimuli. Extreme values were unlikely to be obtained, however, because spike counts were not corrected by subtraction of the background firing rate from stimulus response (this resulted in a measure that was too sensitive to noise).

Figure 6.7(a) shows colour sensitivity plotted against latency for each cell. Latency and colour sensitivity has been averaged across the 3 conditions (where possible) to produce a single figure for each cell. There is a negative correlation between latency and colour sensitivity ($r = -0.394$, $n = 50$, $p < .01$ two-tailed), with the most colour sensitive cells tending to respond earliest at around 70 – 90 ms.

The figure also indicates cells where there was no significant effect of colour on response – as would be expected these cells have colour sensitivity indices close to 0 but there is no tendency for these cells to have similar latencies.

6.4.6. The relationship between latency, colour sensitivity and stimulus type

The same data is presented in figure 6.7(b) with cells labelled with the category of their preferred stimulus. Cells classed as unknown (where preference changed across conditions, or if no response was obtained for any condition) are not shown. While face responsive cells are present across the full range of latencies, they tended to have shorter latencies than non-face cells (face cells, mean latency = 85 ms; non-face cells, mean latency = 102 ms) . This difference was found to be significant ($t = 2.91$, $df = 30$, $p < .01$ two-tailed).

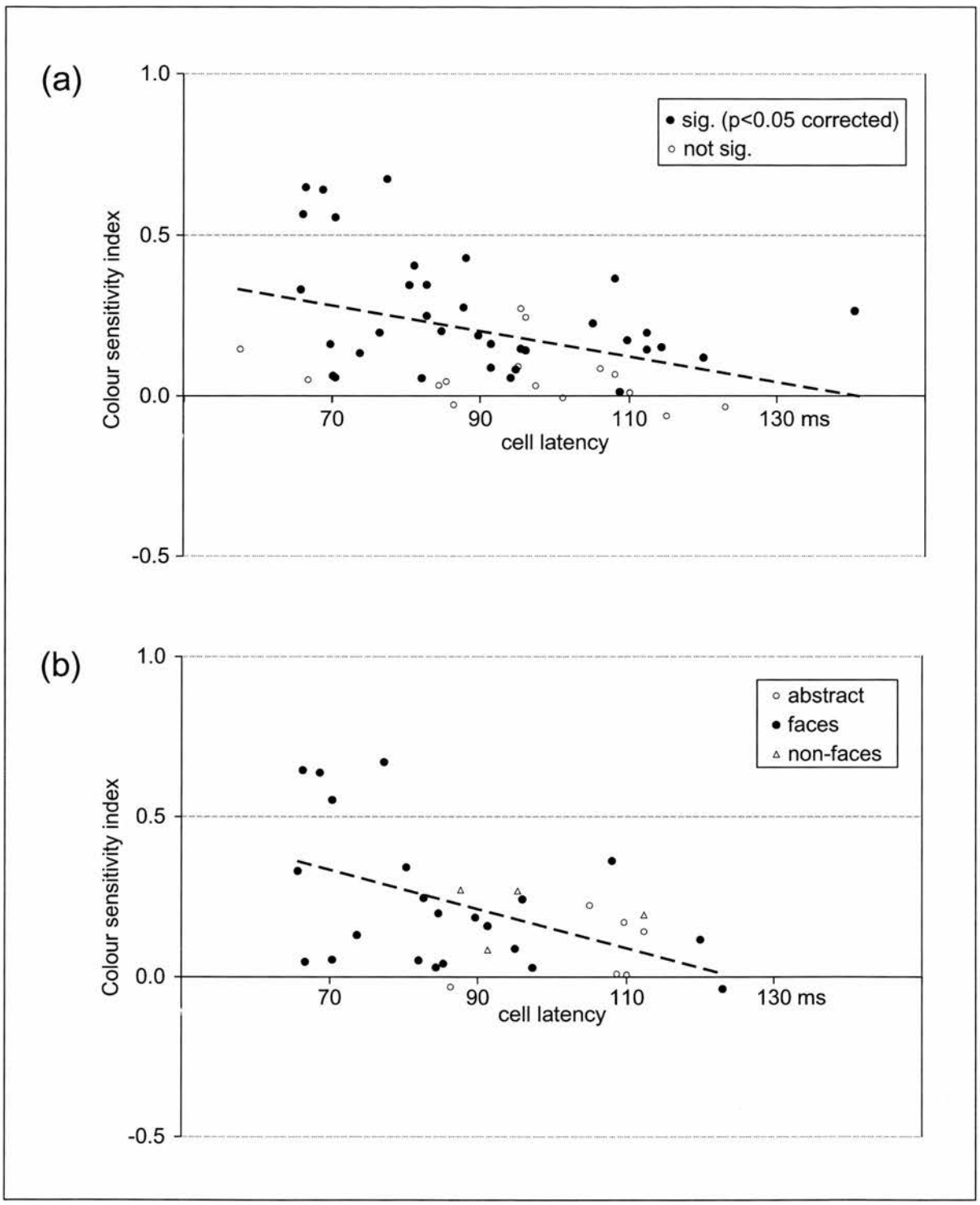


figure 6.7 – Colour sensitivity and cell latency. Colour sensitivity was defined as the difference between colour and achromatic responses over the sum of colour and achromatic responses. Increasingly positive values represent increasing preference for colour, increasingly negative values represent a preference for the achromatic stimuli. A value of around zero indicates similar responses to colour and achromatic stimuli. Colour sensitivity and latency figures were averaged across all conditions. (a) All cells. Filled circles show those cells which had a significant effect of colour ($p<0.05$ corrected) in at least one condition. Hollow circles indicate cells where no significant effect was found. The dashed line shows the least-squares regression for all cells. (b) The category of a cell's preferred stimulus is indicated by the labelled symbols. Those cells which were categorised as unknown (having inconsistent stimulus preference) are not shown. The dashed line shows the least-squares regression for the face-selective sub-population.

Because face cells tend to have shorter latencies, and there is a negative correlation between latency and colour sensitivity, this is reflected in the mean colour sensitivities for the two groups (face cells, mean colour sensitivity = 0.23; non-face cells, mean colour sensitivity = 0.13). However this difference was not significant. ($t = -1.34$, $df = 30$, $p = 0.191$, n.s.)

While the data suggests that there is some tendency for cells responding to abstract stimuli to have a lower degree of colour sensitivity, the sample is too small to draw any firm conclusions.

6.4.7. A Single Population?

As implied by the population average histograms (figure 6.4), the results of the preceding section show that there is a relationship between colour sensitivity and cell latency (figure 6.7), with the earliest cells tending to be more colour-sensitive than later cells. In order to verify that the results of the statistical analysis were robust across the population as a whole (and not biased by a small subset of early latency and particularly colour-sensitive cells), the population was divided into 3 groups –

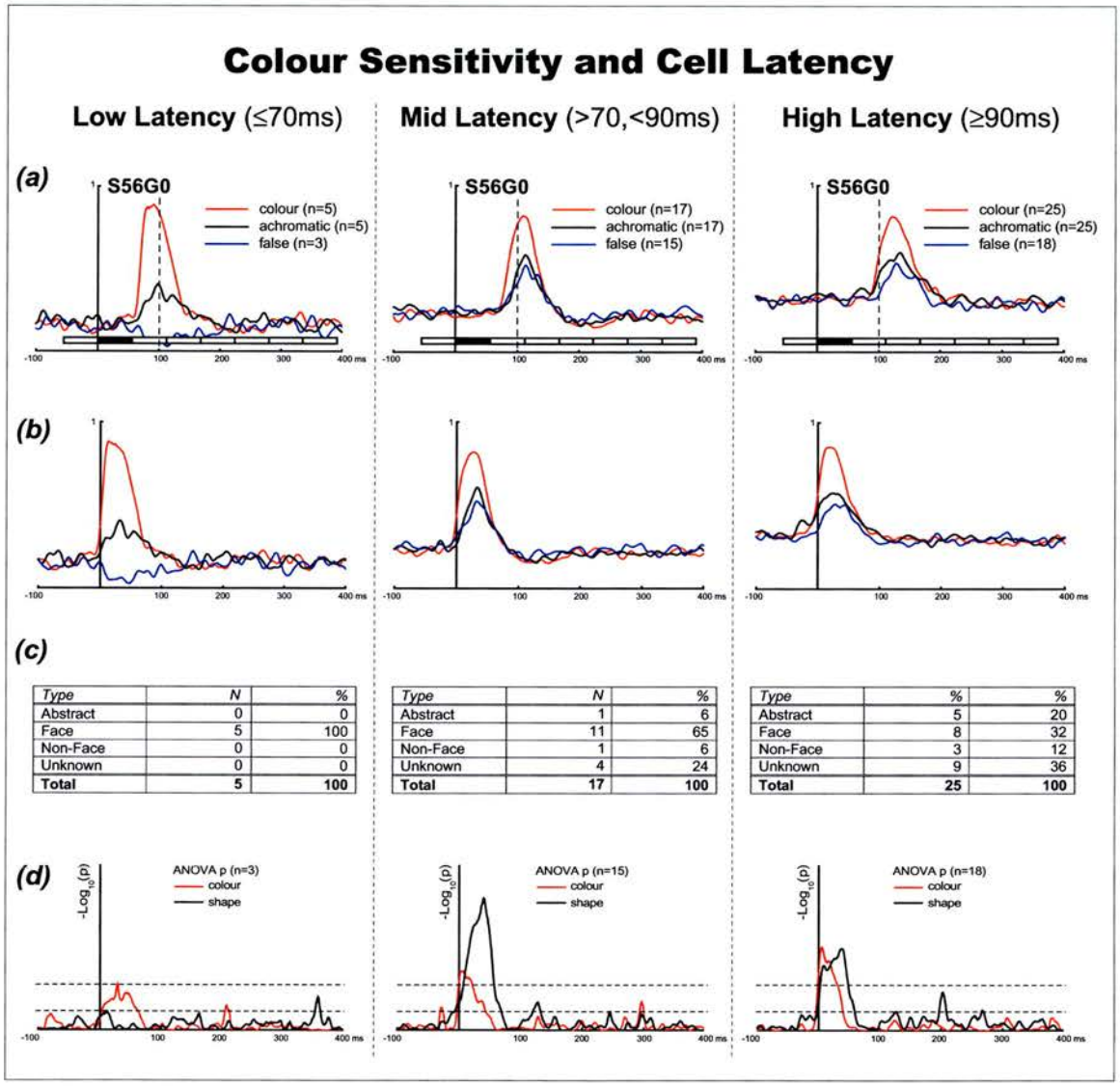
(i) The earliest cells (Cell latency ≤ 70ms)

(ii) Middle latency cells (latency between 70 and 90ms)

(iii) Late cells (latency ≥ 90ms).

This division ensured approximately equal numbers of cells in (ii) & (iii), with (i) included to verify that early colour effect seen in figure 6.4 was indeed due to a small number of low-latency cells that produced little to no achromatic and false colour responses.

The analysis was carried out for the S56G0 condition only, and the results are presented in figure 6.8. Firstly, it can be seen that the population results obtained earlier are representative across latency. Colour images always produce the greatest response, irrespective of latency sub-group.



Type	N	%
Abstract	0	0
Face	5	100
Non-Face	0	0
Unknown	0	0
Total	**5**	**100**

Type	N	%
Abstract	1	6
Face	11	65
Non-Face	1	6
Unknown	4	24
Total	**17**	**100**

Type	%	%
Abstract	5	20
Face	8	32
Non-Face	3	12
Unknown	9	36
Total	**25**	**100**

figure 6.8 – The population sub-divided into 3 groups, based on cell latency. Only responses in the S56G0 condition are shown. *(a)* Stimulus-aligned and *(b)* Latency-aligned group average responses to the colour, achromatic and false colour versions of the best stimulus for each cell. The vertical dashed line shows 100ms, making the latency shift apparent between groups. *(c)* The preferred stimulus of each cell belonging to the latency groups, established as described at the beginning of the results section (i.e. across all presentation rate conditions). *(d)* The time course of colour and shape discrimination for each latency group. As before, this analysis was restricted to the subset of cells for which false colour data had been obtained. For a complete description, please refer to the legend of figure 6.6.

Furthermore, discrimination between stimuli on the basis of colour does not lag behind discrimination on the basis of shapes in either the mid- or high- latency groups (with the n's being too small in the early group to achieve significance).

It is noteworthy that the lowest latency group consists entirely of faces, produced little achromatic response, and also show a marked inhibition with false colour versions of the preferred face, an effect that is not seen in the other two latency groups.

6.4.8. The effect of colour on stimulus selectivity

Despite an overall reduction in response, many cells continued to show the same selectivity for achromatic and false colour stimuli. This can be seen in figure 6.9, which shows the population responses to colour, achromatic and false colour versions of each stimulus, with the stimuli ranked from left to right based on the response to

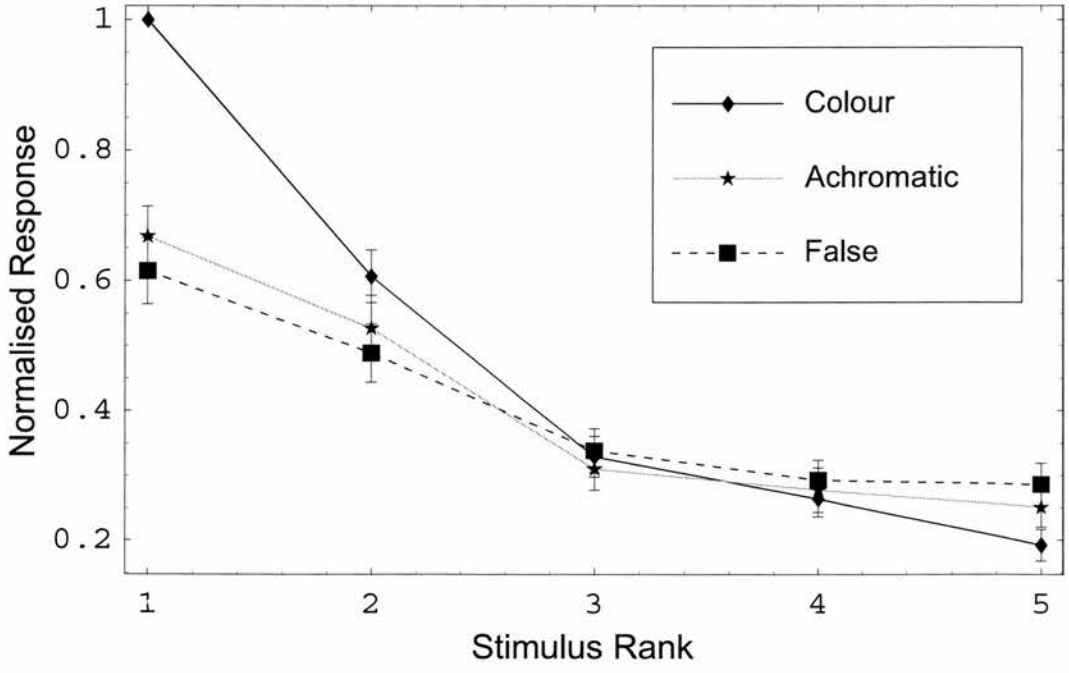


figure 6.9 – Population responses (n=36) for colour, achromatic and false colour versions of the 5 different stimuli tested with each cell. Stimuli were ranked from best (1) to worst (5) based on the responses to the *colour* version of each stimulus (see methods). Prior to the population averaging, responses were normalised to the best colour response for each cell, such that every cell has an equal contribution to the population response. The data shown is for the S56G0 presentation condition, and is restricted to the subset of cells for which false colour responses were obtained. Error bars indicate the standard error of the mean.

the *colour* version of each stimulus. Thus, any change in stimulus selectivity in the achromatic or false colour conditions should be reflected in a deviation from the overall downwards slope from left to right, an effect which is not seen.

There is considerable overlap between the error bars for rankings 4 and 5. This would be expected since, for each cell, stimuli were selected from the screening set on the basis that the *best*, *worst* and three intermediate stimuli were included. For many cells tested, however, there were only one or two stimuli that evoked a response, and as such, stimuli ranked 3 and 4 were typically ineffective at driving cells. As described earlier, the worst stimulus for a cell (rank 5) often caused inhibition, and it is therefore notable the ordering is reversed at this point (implying as mentioned earlier, that the correct colouring was often necessary for the inhibition to occur).

However, the population average masks considerable variation between cells. For the 36 cells examined individually in figure 6.10, less than 50% show unchanged selectivity for achromatic and false colour images, with the majority either showing a change in selectivity, or no selectivity to the achromatic or false colour stimuli.

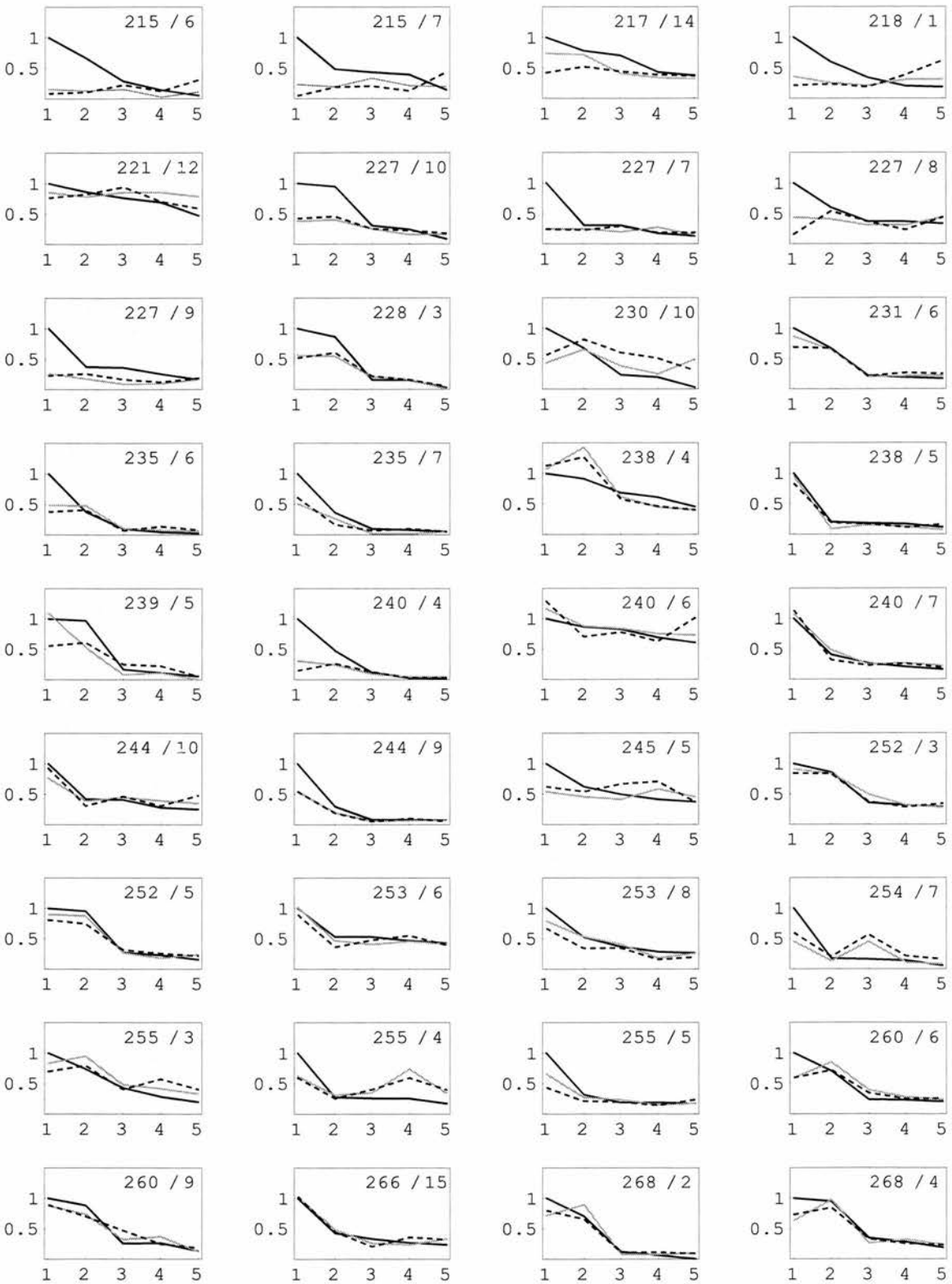


figure 6.10 – Individual stimulus selectivity tuning curves for each of the 36 cells included in the population average shown in figure 6.8. The solid black lines, grey lines and dashed lines indicate colour, achromatic and false colour responses respectively. The axes are not labelled for reasons of clarity, but are identical to those in the population figure. Cell number is indicated in the upper right corner.

6.5. Face-selective sub-population analysis

6.5.1. Face-selective sub-population also shows colour tuning

Almost half our cells (n=22) consistently responded best to an image of a face or head view. The effect of colour on this sub-population of face-selective cells (figure 6.11) is consistent with the effect of colour on the population overall, with the original (or naturally coloured images) producing far greater responses that achromatic or falsely coloured images.

6.5.2. Colour tuning of individual face-selective cells

The colour tuning of individual face-selective cells was measured using the same technique as previously described in the overall population analysis (table 6.4). 68% of these cells showed a significant effect of colour in at least one of the conditions ($p < .05$, corrected).

6.5.3. Colour sensitivity and latency

The negative correlation between latency and colour sensitivity found for the population as a whole is largely due to the contribution of a number of face-selective cells (figure 6.7b). If the analysis is restricted to the face-selective sub-population, a significant correlation is obtained ($r = -0.438$, $n = 22$, $p < .05$ two-tailed), with those face-cells responding earliest tending to be the most strongly colour tuned.

Table 6.4 – Face-Selective Cells' Colour Tuning as a Function of Presentation Rate

Condition	*Stimulus presentation time (ms)*	*Gap time (ms)*	*n*	*Mean Response onset (ms)*	*Window duration (ms)*	*%age of neurons discriminating colour* [a]
S56G167	55.6	166.7	22	91	51	36
S56G0	55.6	-	22	87	59	72
S28	27.8	-	22	78	58	72
any	-	-	22	-	-	68 [b]

[a] Percent of neurons with a significant ANOVA testing the effect of stimulus colour on neuron response ($p < .05$)
[b] Percent of neurons with a significant effect of colour in at least one condition ($p < .05$, Bonferrori corrected)

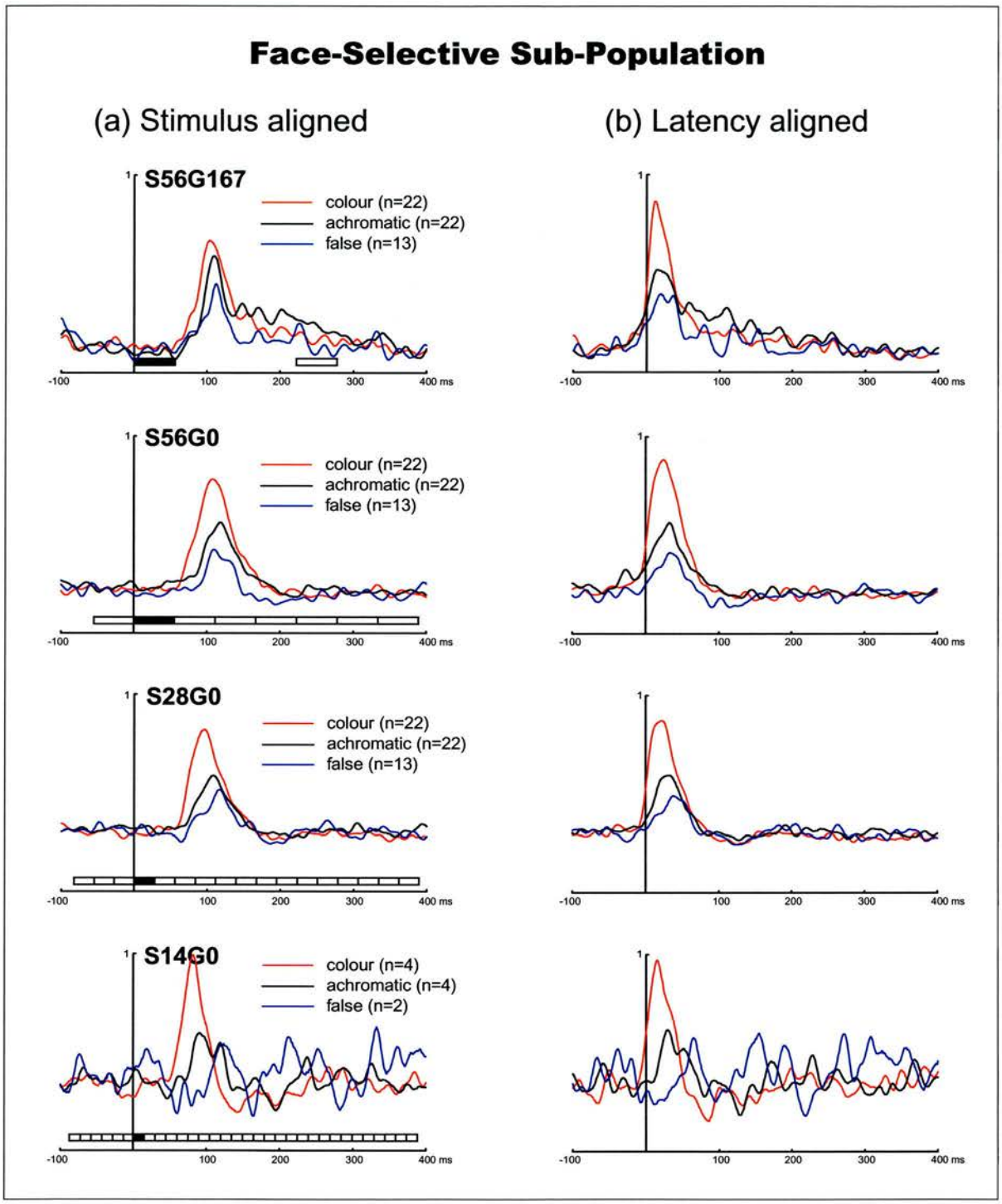


figure 6.11 – *(a)* Stimulus-aligned and *(b)* Latency-aligned normalised average responses to the colour, achromatic and false-colour versions of the best stimulus. This analysis is restricted to the sub-population of cells whose preferred stimulus was a face. Slowest presentation rate at the top, fastest at the bottom. Filled rectangle indicates stimulus presentation, outlined rectangles show earlier or subsequent stimulus presentations. Each set of responses for a cell was aligned to the detected latency of the colour response, as described in methods.

6.5.4. Colour aids in the discrimination of face orientation

A particularly dramatic example of a colour sensitive face-tuned cell (colour sensitivity index = 0.56) can be seen in figure 6.12, with data was obtained during a screening phase in experiment 2. The cell responds best to the front view of a face, with the response decreasing sharply as the face turns away. In contrast, there is almost no response to the achromatic head views. Also shown is the response to a frequency filtered (low-pass) version of the face, which produces little response from

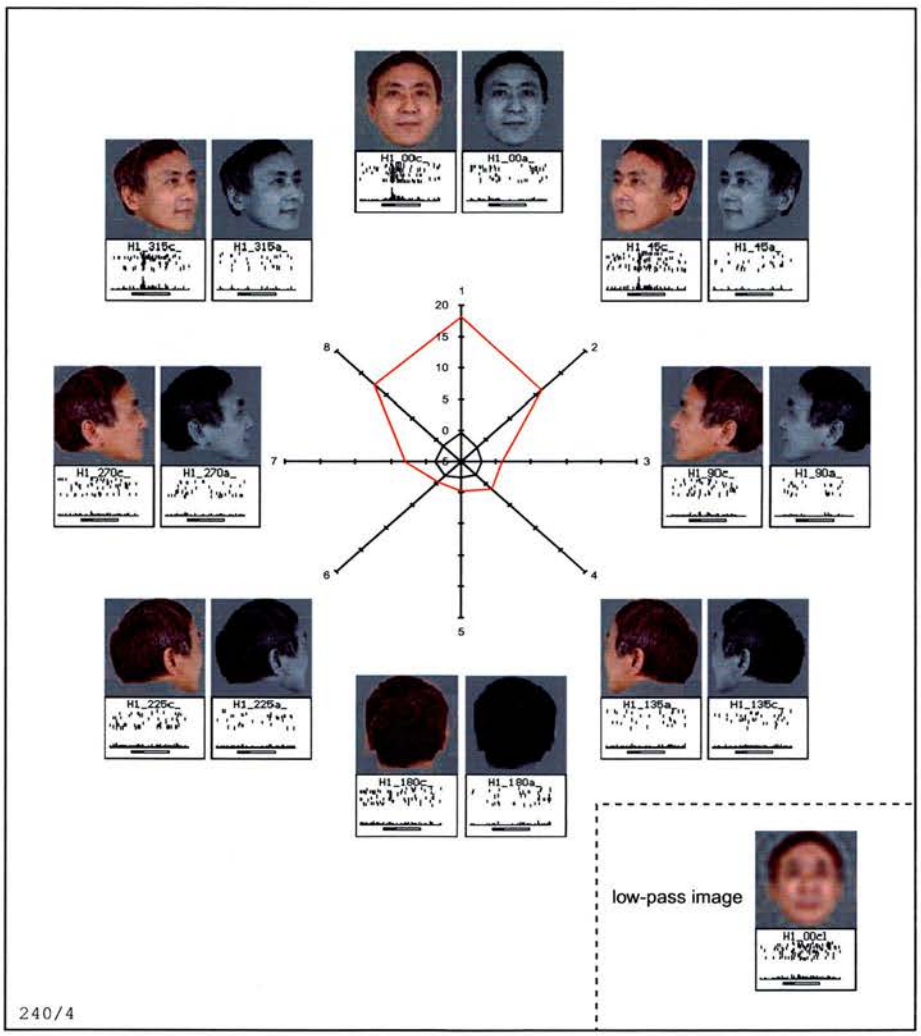


figure 6.12 – An example colour-tuned face cell. The polar plot in the centre of the figure shows the average spike count in a fixed window (beginning at the detected cell latency and with width equal to the stimulus length + ½ the gaussian width for latency detection). The red line and black lines show colour and achromatic responses respectively. The stimuli themselves and corresponding PSTHs are shown around the edges of the plot. Inset, bottom right – the response to a frequency filtered low-pass version of the 0° colour face view, showing that the cell is not simply a "pink-blob" detector.

the cell. This is strong evidence that the cell is truly a colour sensitive face-tuned cell, and not simply a "pink-blob" detector.

6.6. Discussion

6.6.1. Timing of colour information

This experiment provides no evidence to support the idea that colour information is delayed with respect to luminance in inferior temporal cortex. As a population, cells show strong colour tuning and discriminate on the basis of stimulus colour just as early as they discriminate on the basis of shape. Colour tuning is evident across all the presentation rates tested and even at the fast presentation rate, where stimuli are presented in rapid succession for only 14 ms with no gap, there is a clear preference for colour over achromatic stimuli. Thus, colour discrimination did not fail even at the highest presentation rates.

How then do we interpret the result of Delorme et al. (2000), suggesting that rapid reactions are made on the basis of a first wave of achromatic coarse visual information? There is certainly no evidence in the present study to support the idea that the first wave of information is achromatic. There is, however, evidence of a differential effect of colour and shape in "visual memory", apparent in the cells' response during the gap between stimuli (figure 6.6). Delorme et al. (2000) presented stimuli briefly against a black background, then the stimulus disappeared while the subject responded. The results of the present study suggest that, during this gap period, neurons in temporal cortex continue to represent the shape of the last stimulus, but not its colour. It might be suggested, therefore, that the subjects in the study of Delorme et al. (2000) are not actually responding on the basis of a first wave of visual

information, but instead to a colour-blind memory of the stimulus that is signalled by these cells.

However, where a stimulus is immediately followed by a mask, we would expect a quite different pattern of results. Any visual input following the stimulus presentation would abolish the visual memory (compare figure 6.6 top with middle and bottom) and the subject would instead be forced to react on the basis of the response *during* the stimulus presentation. In this case, our results would lead us to expect a clear advantage for colour over achromatic images. This pattern of results is evident in e.g. Lee & Perrett (1997) who briefly presented images of famous faces sandwiched between masking images, and found that colour images were recognised more accurately than achromatic images, even though the stimulus presentation time was comparable to the study of Delorme et al. (2000).

The presence of colour discrimination in the earliest part of our populations' response strongly suggests that either the higher areas of the ventral pathway are exclusively fed by the P pathway, or (more likely) that any latency difference between M & P pathways, as found in V1 (Nowak et al., 1995) has been corrected for by the time visual information reaches higher levels of processing. Indeed, Maunsell et al. (1999) suggested that any latency difference found in V1 between the pathways may be illusory, simply because there are ten times more parvocellular neurons than magnocellular neurons, allowing for a far greater degree of convergence on postsynaptic cells, and perhaps providing P-recipient neurons with a sufficient level of excitatory input to cross threshold more quickly than their M-recipient counterparts, despite latency differences present in the individual inputs themselves.

6.6.2. Proportion of colour tuned cells

The proportion of cells showing a significant level of colour tuning is around 70%. This is remarkably similar to the figure obtained by Komatsu et al. (1992) using simple colour patches, and would suggest that there is a high degree of colour tuning in IT, regardless of whether the shape selectivity of a cell is simple (i.e. capable of being driven by simple geometric shapes), as in their study, or complex (e.g. selective for face and head view), as in the present study. It is not clear why Tanaka et al. (1991) found a much lower proportion of colour-tuned cells.

The range of latencies of those cells categorised as face-selective was 66 – 123 ms, with the sample skewed towards the earlier part of this range (mean 85 ms). The colour insensitivity of face cells noted by Perrett et al. (1982) was from a population of cells with somewhat longer latencies than this (range 80 – 180 ms, mean not provided but approximately 125 ms). This might suggest that the presence of separate colour-sensitive and colour-insensitive populations of face cells in temporal cortex, with those producing colour invariant responses having longer latency. There is also evidence to support this view in our results (see also figure 6.7b), with a significant negative correlation between latency and colour sensitivity both for the face-selective subset and the population as a whole.

In a later study of face-selective cells in temporal cortex, Perrett et al. (1992) note that cells which are view-independent (i.e. respond equally to different views of the head) tend to have longer latencies (by about 10 ms) than those that are view-selective. They suggest this may be evidence of a hierarchical processing scheme, where the outputs from cells responsive to particular views of an object synapse on a single neuron upstream to produce the view-invariant response. We might speculate that the

negative correlation we find between colour sensitivity and latency is evidence of an analogous system of processing for object colour, where several cells, each tuned for a different possible colour of an object, make synapses on a single upstream neuron to produce a colour-invariant response.

6.6.3. The nature of colour tuning

Many of the stimuli used in this experiment had natural colours, i.e. colours that are generally associated with that particular pattern (e.g. a face), and the response to these stimuli was often greatly suppressed when the images were falsely coloured. It seems possible that inferotemporal cortex contains cells that become tuned by visual experience to the specific conjunctions of colour and shape that represent commonly occurring objects in the visual world. For instance, although many cells in temporal cortex respond to images of faces and bodies of different orientations and sizes, the majority are found to be tuned to real life sized and upright orientations (Perrett et al., 1998; Ashbridge et al., 2000). This might explain why naturally coloured scenes activated a region of the human brain corresponding to inferior temporal cortex in the macaque (Zeki & Marini, 1998) but falsely coloured scenes did not. However, since the population of cells we tested was selected on the basis that they responded to at least one image from our screening set (and this set largely consisted of natural images) it is impossible to tell to what extent the cells in IT code for the shapes and colours of naturally occurring objects.

Finally, this experiment has also shown that colour plays a central role in the selectivity of IT neurons. While some cells showed the same stimulus selectivity when achromatic or false colour images were presented, the majority did not. Thus, the idea that colour plays a secondary role to object form in IT (i.e. useful to provide

extra detail, but not necessary) gains little support from the results of this study.

6.6.4. Implications for stimulus optimisation experiments

One of the motivations behind this experiment was to find out whether, at the high rates of presentation required by stimulus optimisation, colour information would arrive too late to contribute to the measured response. This possibly was suggested by a range of studies, which I discussed in Chapter 5. If this were the case, it would be possible to present achromatic-only images, and reduce the dimensionality of the search space by two-thirds. However, the results of this experiment suggest otherwise. Colour is a necessary attribute for effective stimuli in a large proportion of IT cells. Achromatic images produce greatly reduced responses, and this reduction is evident from the response onset. In IT at least, it seems that effective optimisation will require the use of colour stimuli.

6.6.5. Concluding comments

It is clear from the present study that any latency differences present in V1 (and earlier in the visual system) between colour information, in the P pathway, and achromatic information, in the M pathway, are either illusory or have been corrected for *before* the information about a stimulus reaches the higher visual areas such as IT.

This study suggests a strong role for colour in object recognition, and provides no evidence to support the idea of a first wave of purely achromatic information in the visual system.

7. The Role of Spatial Frequency in the Response of Neurons in Temporal Visual Cortex to Complex Stimuli

7.1. Experimental Hypotheses

(i) At any presentation rate, discrimination between patterns based on high spatial frequencies will occur at increased latency, and with a later peak, than discrimination of patterns based on low spatial frequency.

(ii) Additionally, the "reset" model (introduced in Chapter 5) would predict that, as stimulus presentation rate increases, any response due to high frequency information in the stimulus will be progressively diminished because the high frequency signal will not arrive in time to contribute to the response. Thus, the response to high-pass stimuli will tend towards baseline, with low-pass stimuli generating an equal response to the corresponding unfiltered images at the highest presentation rate. Neither of these effects would be predicted by the "mismatch" model – the high frequency aspect of response would still occur, simply with higher latency.

7.2. Methods

Since this experiment was carried out in parallel with the colour experiment described in the previous chapter, only the visual stimuli and those aspects of the analysis that differed are described in detail here.

7.2.1. Visual Stimuli

Frequency filtered versions of the stimulus set (Appendix 1) were prepared using a Fast Fourier Transform algorithm (Press et al., 1988). Each image was transformed into frequency space and multiplied with circularly symmetric high- and low-pass filters. The high and low-pass filters were exactly complementary, with the cut-off

frequency at 8 cycles per image. Thus the low-pass filter preserved all frequencies less than or equal to 8 Hz, and the high-pass filter all the remaining frequencies above 8 Hz. The cut-off frequency was chosen with reference to the study of Rolls et al. (1984) who found that almost all of the face selective cells they tested produced at least half the maximal response to images filtered either above or below this frequency. The DC component was treated separately, and was maintained in both high- and low-pass images, in order to preserve the overall luminance. No adjustment was made to the contrast of the resultant images, since this may have distorted the neuronal response.

In addition to the colour images, frequency filtering was also carried out on the achromatic images prepared in the previous chapter. Thus, for each of the stimuli, there were 6 different versions (an example is shown in figure 7.1).

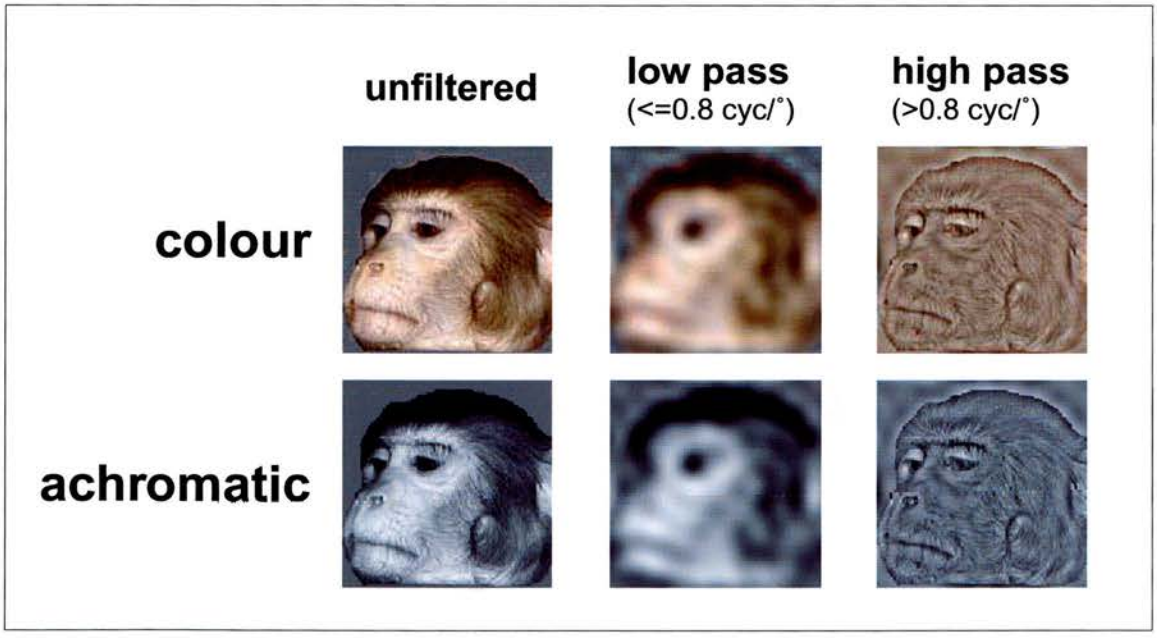


figure 7.1 – The six different versions of one of the stimuli used in the experiment. The top row shows the original image, and corresponding low- and high-pass versions of the image. Frequency filtering was carried out as described in the methods section. The bottom row shows corresponding achromatic versions of the stimuli, with the colour filtering carried out as described in the previous chapter. The DC components of the image were preserved in both the low- and high-pass filtered conditions to maintain overall image luminance.

7.2.2. Procedure

(a) Colour Stimuli

Cell preference and latency was established as described in the previous chapter, based on the cell's response to the colour unfiltered stimuli.

The normalising factor for each cell was calculated as the maximum value of the SDF (spike density function) for the unfiltered colour version of the best stimulus. High- and low-pass responses to the best stimulus were weighted by this factor, so that every cell would have an equal contribution in the population response, with the response to the unfiltered colour stimulus acting as baseline.

(b) Achromatic Stimuli

Achromatic versions of the stimuli were analysed separately using exactly the same procedure as above, but with ordering and latency detection based on the cell's response to the achromatic unfiltered stimuli. Because achromatic stimuli produced greatly reduced responses from the majority of cells tested, there were approximately 20% fewer cells in each of the achromatic conditions, with the excluded cells failing to produce a response with measurable latency to any of the unfiltered achromatic stimuli.

In this case, the normalising factor for each cell was calculated as the maximum value of the SDF for the unfiltered achromatic version of the best stimulus. High- and low-pass responses to the best stimulus were weighted by this factor, so that every cell would have an equal contribution in the population response, with the unfiltered achromatic response acting as baseline.

7.3. Results

7.3.1. Stimulus Preference

The stimulus preference for each cell tested was established in the previous chapter, but is reprinted below for reference.

Table 7.1 – Preferred Stimulus Category

Preferred Stimulus Category	*Number of cells*	*Percentage*
abstract	6	12
non-abstract	26	52
(face)	(22)	(44)
(non-face)	(4)	(8)
unknown	18	36
total	50	100

Each cell was tested with both colour and achromatic frequency filtered images (see figure 7.1). Because colour images consistently produce larger responses in the majority of cells tested (see previous chapter) we shall first examine the effect of frequency filtering with colour images. In the second part of this section we will examine the smaller responses evoked by the achromatic versions of the stimuli.

7.4. Colour Stimuli

7.4.1. Qualitative description of frequency tuning

(a) Single Cell

Figure 7.2 shows the response of a single cell to the frequency filtered versions of its best stimulus in the four different presentation rate conditions. Also shown in this figure are the responses to unfiltered, high- and low-pass stimuli superimposed and aligned in time.

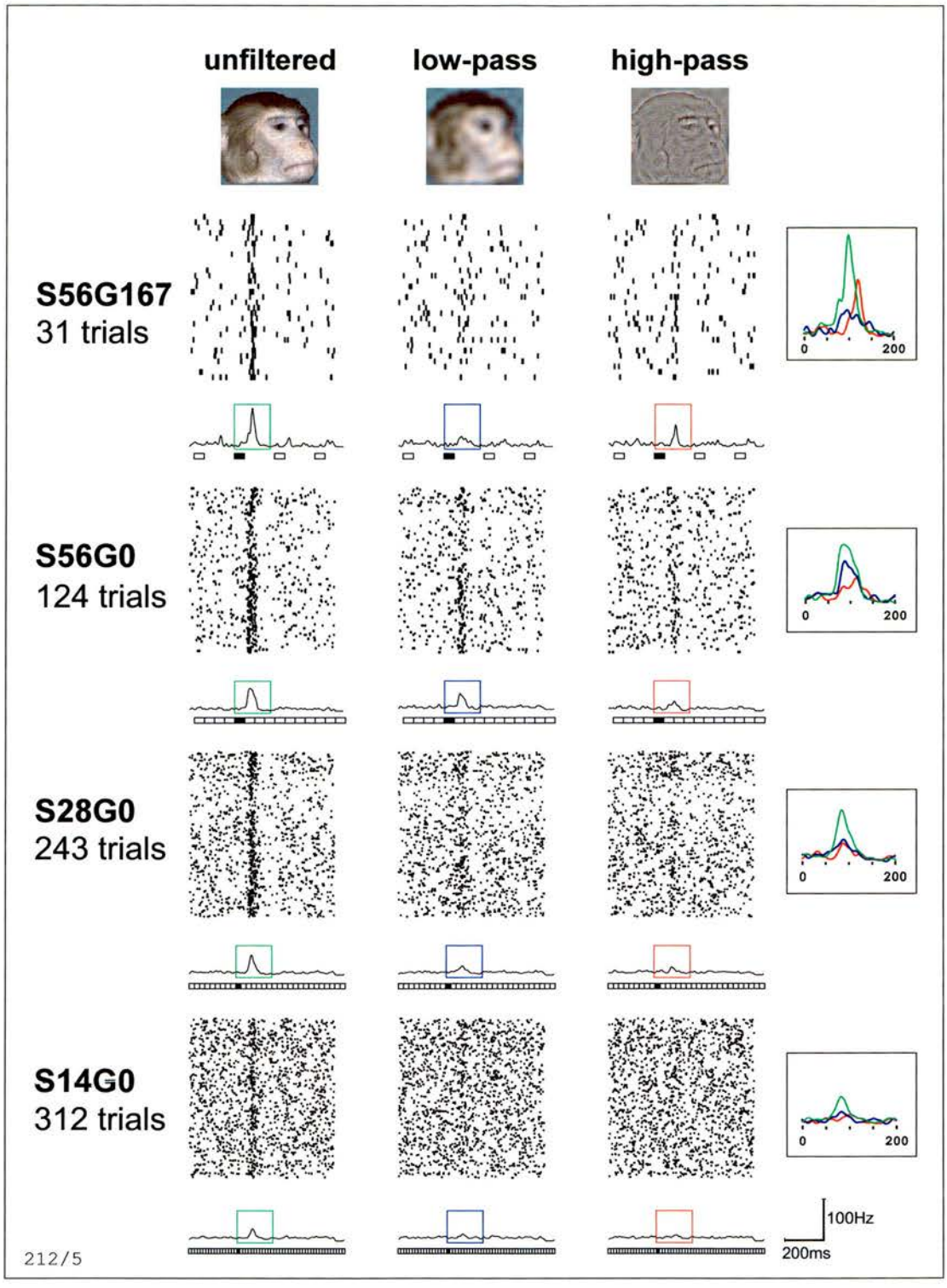


figure 7.2 – Response of a single cell to the unfiltered, low- and high-pass versions of its best colour stimulus in different presentation rate conditions. (top row: 56 ms + 167 ms gap, 2nd row – 56 ms, 3rd row – 28 ms, bottom row – 14ms). SDFs were created by summing the data across trials and smoothing with a gaussian (σ = 10 ms). Stimulus presentation is shown by a filled rectangle below the SDF and hollow rectangles show previous or subsequent stimulus presentations. The latency of this cell was 66 ms. Scale is indicated in the bottom right corner. Insets (right) show regions of the response to unfiltered (green), low-pass (blue) and high-pass (red) images magnified and time aligned, with stimulus onset at 0 ms.

Firstly, it can be seen that this cell can be driven by *either* the low- *or* the high spatial frequency components of an image, even though the two stimuli share no spatial frequencies in common. This property was noted in an earlier study by Rolls et al. (1984). However, the responses to both of the filtered stimuli are substantially less than to the unfiltered original, an effect that is repeated across the presentation rate conditions.

If our hypothesis that high spatial frequency information about a stimulus is delayed with respect to the low spatial frequencies, then we should expect to see the following effects in this cell, and the population response as a whole –

(i) Unfiltered and low-pass responses should have an equal onset in time, leading the high-pass response.

(ii) The unfiltered and low-pass responses should be initially identical, then diverge to reflect the delayed contribution of the high spatial frequency stimulus information present in the unfiltered image

Furthermore, the "reset" model would additionally predict that –

(iii) Any difference between unfiltered and low-pass curves should be progressively reduced as presentation rate increases, since high frequency information will not arrive in time to contribute to the response

(iv) The high-pass curve should tend towards the baseline as presentation rate increases, for the same reason as in (iii)

Although some of these effects can be seen in the response of this single cell, the picture is not consistent. There does appear to be a latency difference between the

high- and low-pass responses in some of the presentation rate conditions (particularly S56G167), but little evidence that unfiltered and low-pass responses are equal during the early part of the response.

With reference to the "reset" model in particular, we can also see that, even though the high-pass response diminishes as the presentation rate is increased, this is matched by a similar reduction in the low-pass response, with little indication that the low-pass and unfiltered responses become equivalent at the fastest rate.

(b) Population Results

Results from the population as a whole are shown in figure 7.3, and it is clear that the cells are being driven by information carried across the whole range of spatial frequencies, with a sharp reduction in response when either the high or the low frequency components of the image are removed.

There is, however, little evidence for any of the latency effects we are looking for in the population response. In the fastest condition (S14G0), the initial 15-20 ms of response is apparently dominated by a low frequency response, with the high frequencies coming in later, but this does not appear to be the case in the other presentation rate conditions. Condition S14G0 was only carried out with those cells giving a sufficiently strong response, and as such, the number of cells is relatively low (n=7) in this condition. Condition S56G0 shows what might appear to be an earlier response to the low-frequencies, but this effect disappears when the responses are latency aligned.

There is also no evidence that the high frequency curve tends towards baseline as presentation rate increases, or that the difference between low-pass and unfiltered

curves is reduced as speed increases (with, again, the exception of the fastest condition, S14G0).

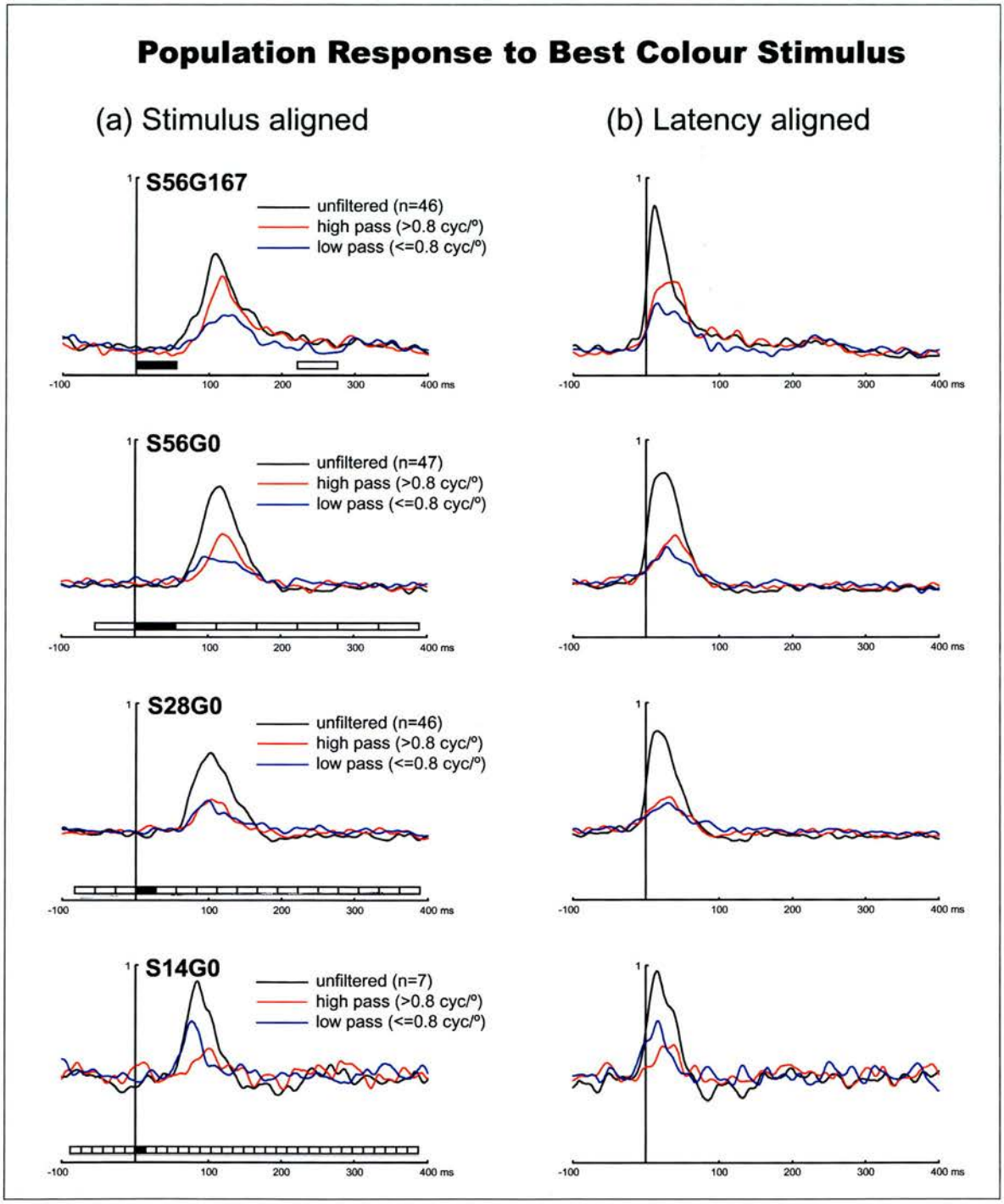


figure 7.3 – *(a)* Stimulus-aligned and *(b)* Latency-aligned normalised average population responses to unfiltered, high- and low-pass versions of the best colour stimulus for each cell. Slowest presentation rate at the top, fastest at the bottom. Filled rectangle indicates stimulus presentation, outlined rectangles show earlier or subsequent stimulus presentations. Each set of responses for a cell was aligned to the detected latency of the unfiltered response, as described in methods.

7.4.2. Time course of frequency tuning – population statistical analysis

The time course of tuning for high- and low spatial frequencies was assessed quantitatively using a sliding window statistical test to measure the probability of discrimination between stimuli as a function of time.

For each cell, and each presentation rate (with the exception of S14G0, where there was an insufficient number of cells to perform the analysis), normalised and latency-aligned SDFs were calculated as previously described. The time course of overall shape discrimination was then determined by comparing the SDFs for the unfiltered colour stimuli with a related one-way ANOVA (5 levels, corresponding to the five different stimuli), performed separately for each 1 ms time bin. High- and low-frequency tuning was established using a similar procedure with two further 5 level ANOVAs over the high- and low-pass images respectively, and thus describing the time course of stimulus discrimination by the population when only these restricted frequency subsets were available.

The results of this analysis can be seen in figure 7.4. and is tabulated in table 7.2. Despite the much stronger level of discrimination based on the (generally higher) responses to the high-pass images, onset and offset times of discrimination for low- and high- pass stimuli are approximately equal, with nothing to suggest that pattern discrimination based on low-frequencies precedes that of high frequencies.

Table 7.2 – Onset[1] of pattern discrimination at the p=0.01 (corrected) level

Condition	*Unfiltered (ms)*	*High-pass (ms)*	*Low-pass (ms)*
S56G167	-6	4	6
S56G0	-7	2	1
S28G0	-9	-4	3

[1] onset times are relative to detected cell latency

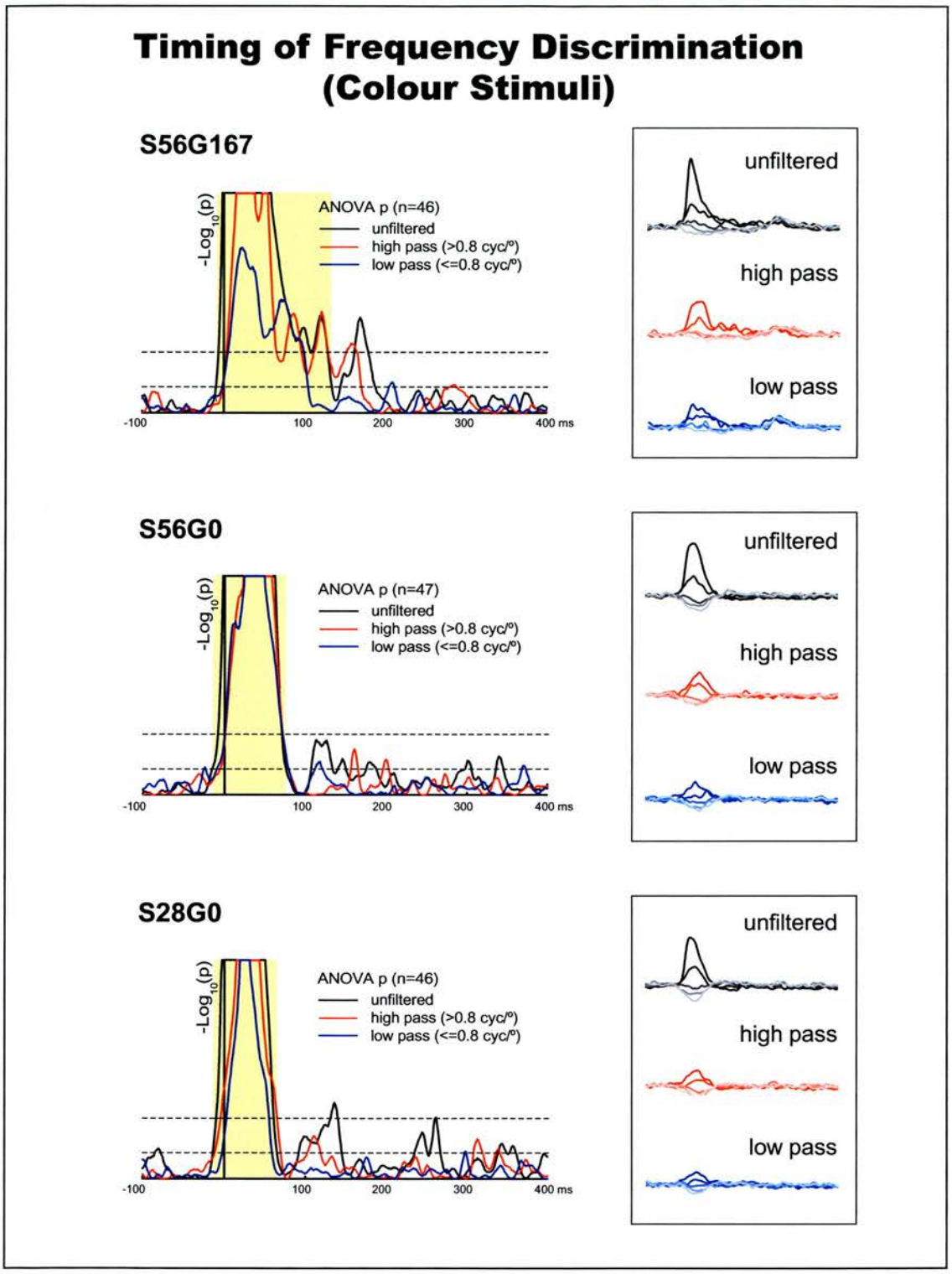


figure 7.4 – Time course of frequency discrimination, measured using sliding window ANOVAs. SDFs were calculated for each stimulus, normalised and latency aligned (see method). Overall shape discrimination was determined by a related one-way ANOVA (5 levels) and compared responses to the unfiltered colour stimuli (insets, top – darkest line = best stimulus, lightest line = worst). Tuning for high- and low spatial frequencies was measured in a similar way, with ANOVAs across the high- and low-pass colour stimuli respectively (insets, middle and bottom). This analysis was performed separately for each 1msec time bin. Probability is plotted on a negative log scale, and curtailed at $-\text{Log}_{10}(p)=17$, which is the limit of the numerical accuracy of the software used to perform the analysis (Mathematica 4, Wolfram Research). Lower dotted line shows significance at the $p = .01$ level (uncorrected). Upper dotted line shows the same level Bonferroni corrected. Yellow shaded area indicates duration of overall shape discrimination at the uncorrected level and is used to define the window in the cell by cell analysis.

7.4.3. Frequency sensitivity index

The statistical analysis above finds no difference in response onset time between high- and low-frequency stimuli *within the context of each cell's response.* However, because the responses have been previously aligned on detected cell latency (based on the unfiltered response), this type of analysis would mask potential latency differences *between* cells that may be differentially tuned to the high- and low frequencies. If there was, for example, a shorter-latency sub-population of cells that responded only to the low frequencies, this would explain why there appeared to be a earlier response to low- compared to high-frequencies in the stimulus-aligned population response, but this effect disappeared when the data was latency-aligned (see figure 7.3 (a) S56G0, S14G0).

Therefore, we now assess whether there was any relationship between frequency sensitivity and cell latency, using a frequency sensitivity index calculated individually for each cell –

$$frequency\ sensitivity\ index = \frac{\bar{h} - \bar{l}}{\bar{h} + \bar{l}}$$

where $\bar{h}$ was the mean spike count for the high-pass version of the best colour unfiltered stimulus and $\bar{l}$ was the mean spike count for low-pass version. The window was determined from the population analysis above, and was calculated separately for each condition as follows. The window began with the first 1 ms time bin when unfiltered shape discrimination exceeded the p = 0.01 (uncorrected) level. It ended when shape discrimination fell back below this criterion level, with the constraint that the window was at least as long as the stimulus presentation time.

The frequency sensitivity index ranges from +1 (preference for high-pass stimuli) to -1 (preference for low-pass stimuli) with a value of 0 indicating that a cell responded identically to the high- and low-pass stimuli. Extreme values were unlikely to be obtained, however, because spike counts were not corrected by subtraction of the background firing rate from stimulus response (this resulted in a measure that was too sensitive to noise).

Figure 7.5 shows frequency sensitivity plotted against latency for each cell. Both latency and frequency sensitivity have been averaged across the 3 stimulus presentation rate conditions S56G167, S55G0 and S28G0 (where possible) to produce

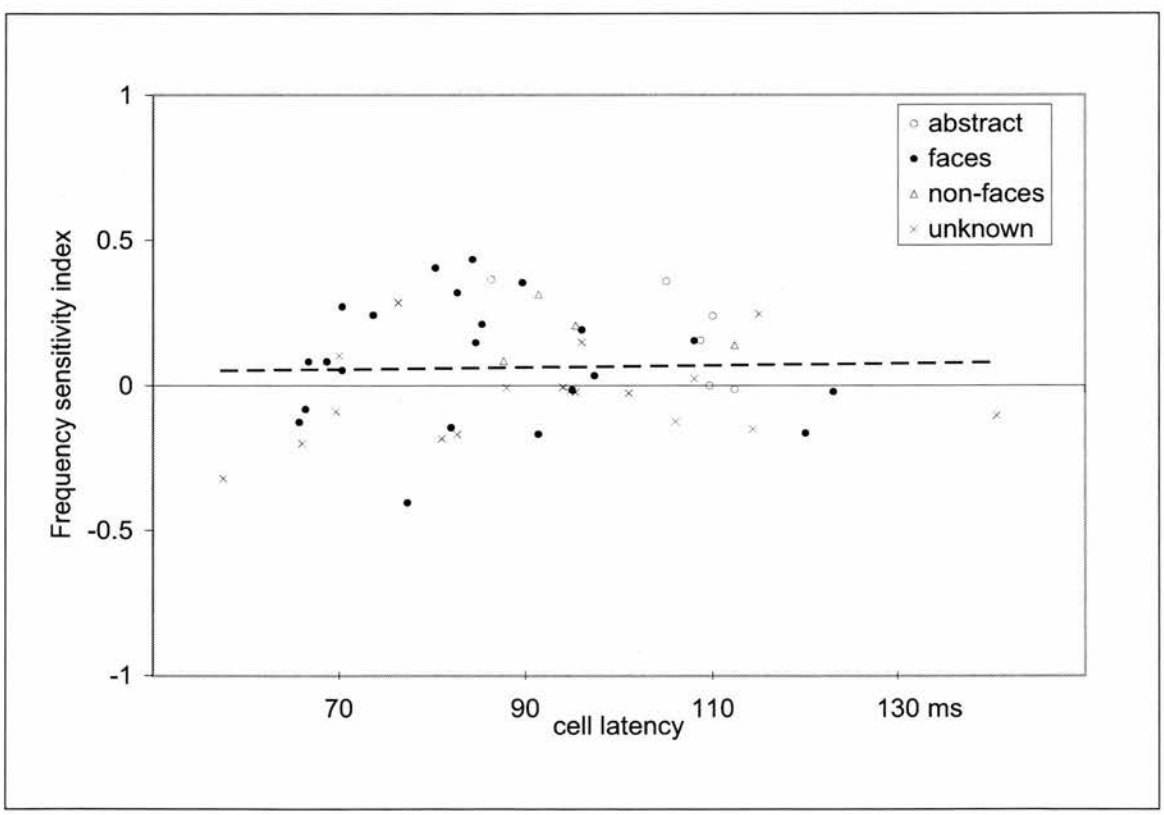


figure 7.5 – Frequency sensitivity and cell latency. Frequency sensitivity was defined as the difference between high- and low-pass responses over the sum of high- and low pass responses. Increasingly positive values represent preference for high-pass stimuli, increasingly negative values represent a preference for the low-pass stimuli. A value of zero indicates equal responses to high- and low-pass stimuli. Frequency sensitivity and latency figures were averaged across all conditions. The category of a cell's preferred stimulus is indicated by the labelled symbols. The dashed line shows the least-squares regression.

a single latency and frequency sensitivity figure for each cell. Although the earliest responding cells in our population do appear to have a preference for the low-frequency stimuli, there is no significant correlation between these two properties ($r = .044$, $n = 50$, n.s.).

7.5. Achromatic Stimuli

7.5.1. Qualitative description of frequency tuning

Achromatic stimuli consistently produced smaller responses than their colour counterparts, as would be expected given the results of the previous chapter. This is particularly evident in the population responses to achromatic stimuli (figure 7.6) where the low frequency response is almost completely abolished by the absence of colour.

Since there is no consistent response to low-pass achromatic stimuli (except in condition S56G167), this makes it rather difficult to compare the time courses of responses to the different versions of the stimuli. In the condition where a low-pass response is obtained, however, the onset and offset times of responses to the unfiltered, high- and low-pass images appear to be similar, despite their difference in magnitude.

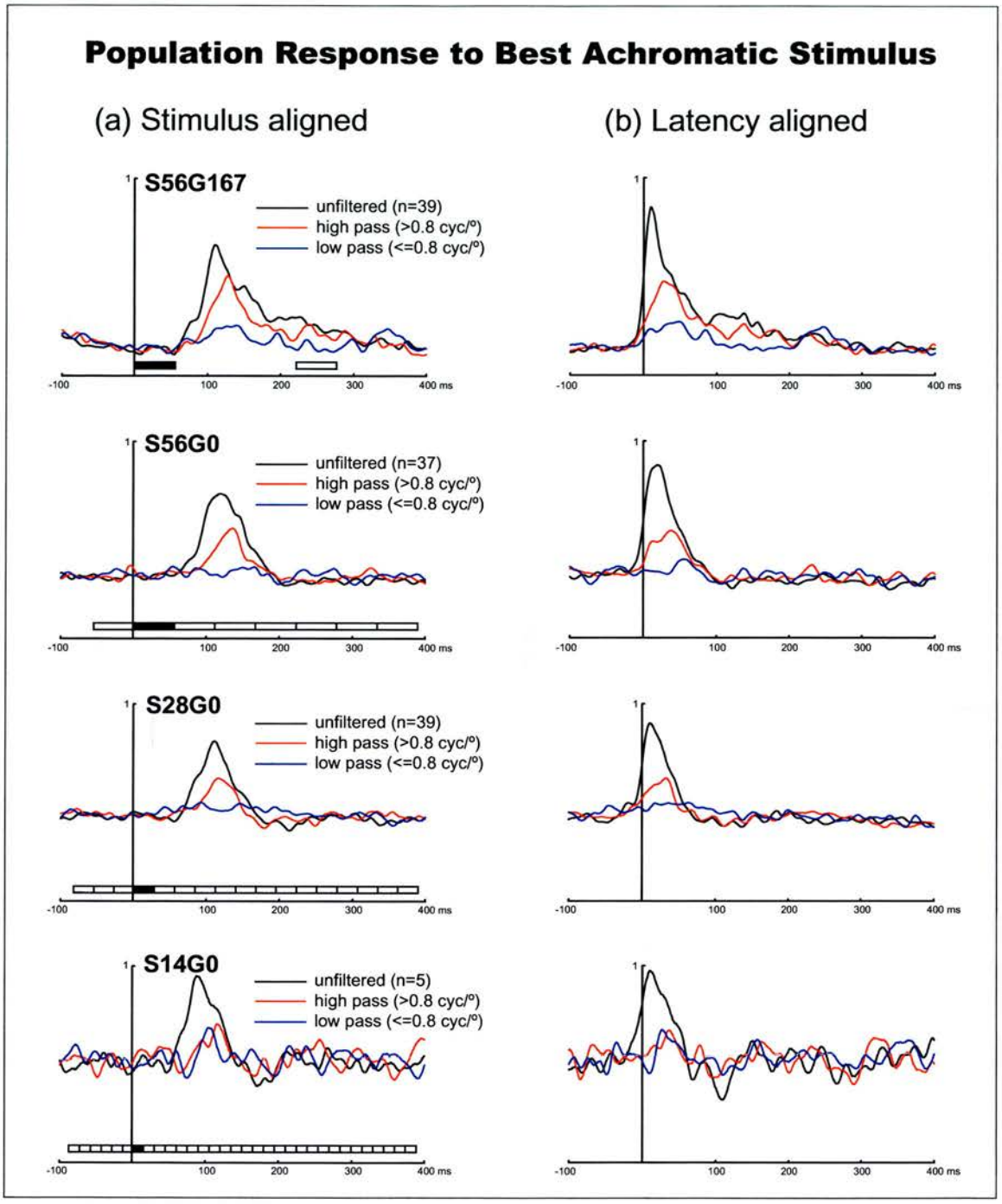


figure 7.6 – *(a)* Stimulus-aligned and *(b)* Latency-aligned normalised average population responses to unfiltered, high- and low-pass versions of the best achromatic stimulus for each cell. Slowest presentation rate at the top, fastest at the bottom. Filled rectangle indicates stimulus presentation, outlined rectangles show earlier or subsequent stimulus presentations. Each set of responses for a cell was aligned to the detected latency of the unfiltered response, as described in methods.

7.5.2. Time course of frequency tuning – population statistical analysis

A statistical analysis of the discrimination of achromatic stimuli was carried using the method previously described (figure 7.7). Despite the very small responses to the low-pass achromatic stimuli, the population was found to be discriminating amongst these stimulus versions to a statistically significant level.

It is however, difficult to draw any solid conclusions from this analysis. Although the population appears to be discriminating on the basis of high frequency information approximately 10-20 ms before low frequency discrimination reaches a comparable threshold, the difference in magnitude makes this conclusion open to question. Overall the time courses of high- and low-frequency discrimination appear similar, with both curves appearing to peak at similar times. There is certainly no evidence to support the hypothesis that high spatial frequency information has a longer latency than low frequency information.

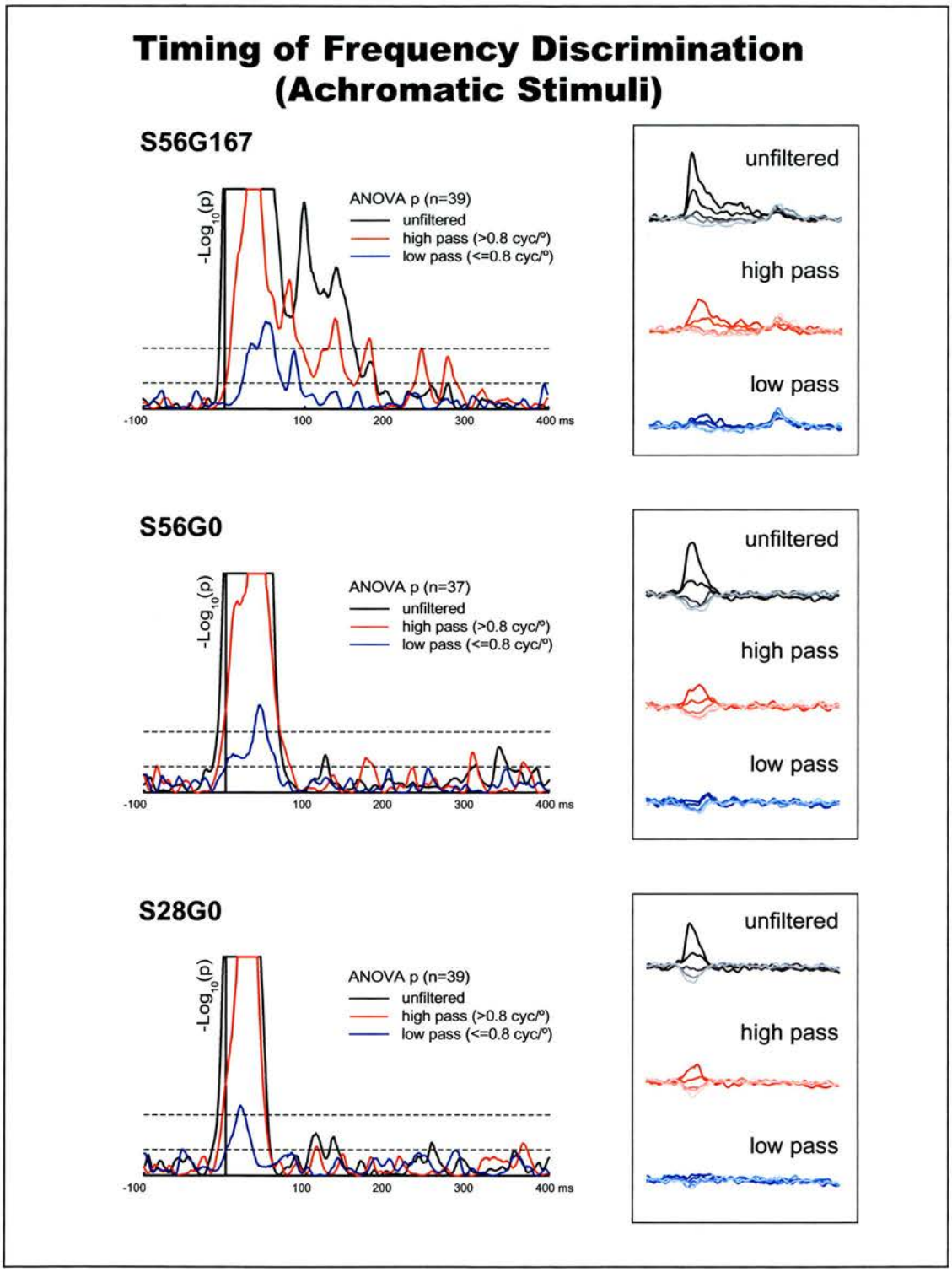


figure 7.7 – Time course of frequency discrimination, measured using sliding window ANOVAs. SDFs were calculated for each stimulus, normalised and latency aligned (see method). Overall shape discrimination was determined by a related one-way ANOVA (5 levels) and compared responses to the unfiltered achromatic stimuli (insets, top – darkest line = best stimulus, lightest line = worst). High and low frequency tuning were measured in a similar way, with ANOVAs across the high- and low-pass achromatic stimuli respectively (insets, middle and bottom). This analysis was performed separately for each 1msec time bin. Probability is plotted on a negative log scale, and curtailed at $-\text{Log}_{10}(p)=17$. Lower dotted line shows significance at the $p = .01$ level (uncorrected). Upper dotted line shows the same level Bonferroni corrected. [Software used to perform the analysis: Mathematica 4, Wolfram Research].

7.6. Discussion

7.6.1. Timing of frequency information

Although somewhat less clear than the previous study, this results of this experiment fail to support the idea that high spatial frequencies have a delayed arrival time with respect to low spatial frequencies in IT/STS. This is demonstrated by the following observations –

(i) unfiltered and low-pass responses differ from the onset of response

(ii) unfiltered and low-pass responses remain different even at the fastest presentation rate (condition S14G0), a difference that can only be due to the absence of the high spatial frequencies

(iii) high-pass responses are present even when stimuli are presented in a rapid continuous sequence, with each stimulus present for only 14 ms (conditon S14G0)

(iv) rising phase of stimulus discrimination shows an equal time course, irrespective of whether discrimination is on the basis of low-pass or high-pass stimuli

(v) spatial frequency preference is not correlated with latency across different cells

It is therefore quite clear that neurons in the area studied have access to all spatial frequencies of a stimulus, from the very beginning of their response, and that neither of the models postulated in the introduction is required to explain our findings.

This result supports the use of the RSVP technique for stimulus optimisation experiments. Evidence for either the "reset" and "mismatch" models would have reduced the experimental value of the RSVP technique. If the high spatial frequencies were lost at high presentation rates, it would be impossible to fully characterise the response properties of cells in the area under investigation. Alternatively, if the high spatial frequencies were simply delayed, this too would be problematic, since the window for capturing the response to a stimulus can only be as long as the stimulus itself, allowing us to measure the cell's response to either the low- or high-frequency aspect of the stimulus, but never both.

7.6.2. Theoretical implications for coarse to fine processing

It could be argued that the cutoff point chosen to construct the low- and high-pass images used in this study did not ensure a clear division between the magnocellular and parvocellular pathways. While this is certainly the case (witness the greatly increased response from the low-pass colour images compared to the achromatic versions – showing clear involvement of P pathway inputs), the considerable overlap in sensitivities between the two systems would make it practically impossible to create stimuli that fulfilled this criteria. Furthermore, the psychophysical and physiological studies discussed in the introduction would also strongly imply that there are spatial frequency related delays within (as well as between) the systems, and that the stimuli used in the present study should have been sufficient to expose any latency delays that were present.

In the context of the results of this study, how do we interpret the findings of Sugase et al. (1999), showing that global information about a stimulus is present in inferotemporal neurons before the fine detail? It certainly seems that there cannot be a

simple correspondence between the 'global' and 'fine' information of their study and the low- and high-spatial frequencies of the present study. In many ways this is not surprising, since the kind of properties exhibited by neurons in STS/IT (such as view tuning to facial stimuli) imply that the inputs that drive these cells already possess considerable complexity, and are more likely to relate to meaningful aspects of complex real world objects, rather than specific spatial frequency channels.

If there are latency differences between different spatial frequency channels earlier in the visual system, the results presented here imply that this information must be integrated in a way that compensates for the differential arrival time of information, and this integration has must have already taken place by the time stimulus information reaches higher visual areas involved in recognition, such as the area studied here. This view is compatible with the idea that integration proceeds according to a temporally anisotropic model, but suggests that this may be largely irrelevant for recognition processes, which receive a signal only after the integration process is complete.

An alternative explanation might be that latency differences seen earlier in the visual system could be cancelled out by anatomical features, such as a convergence of more numerous higher-spatial frequency P pathway synapses on higher visual neurons. These could provide sufficient EPSPs on a IT/STS neuron to cause it to fire, despite having higher latency than a smaller number of low frequency inputs. This would be consistent with the higher sampling density of P pathways cells mentioned in the introduction (Merigan et al., 1991b).

7.6.3. Concluding comments

In this and the preceding study we have tested the theory that visual information is processed in distinct waves, with the first wave consisting of low spatial frequency achromatic information, and coloured fine detail information following later.

Recording from neurons in the higher visual areas involved in recognition, we find no evidence for this model. These cells respond selectively to colour and high frequencies even at the fastest presentation rates tested, and there is no evidence for a differential time course of colour or frequency discrimination within the responses themselves. We thus reject the model of visual recognition proposed most recently by Delorme et al. (2000).

We can therefore be confident that the RSVP technique used here and elsewhere (Keysers et al., 2001) is sufficient to explore the *complete* stimulus selectivity of higher visual neurons, with respect to colour and spatial frequency, rather than just a restricted subset of these properties.

8. Summary of Main Findings and Conclusions

8.1. Optimisation

The first part of this thesis described a novel computerised approach to the problem of characterising visual receptive fields. The method employs a correlation based search algorithm and an image model which can be used to tune the search to the properties of the cells under study.

Optimisation simulations using both pixel and frequency based image models showed that the algorithm can be highly effective at finding optimal stimuli for artificial and simulated non-linear neurons. Attempts to find optimal stimuli for visual neurons in IT and STS were less successful. However, the method may prove more applicable than the reverse correlation technique (Jones & Palmer, 1987; Ringach et al., 1996) or ALOPEX (Harth & Tzanakou, 1974; Tzanakou et al., 1979), firstly because it is able to provide multiple solutions for non-linear receptive fields and, secondly, because it allows the search to be restricted to particular stimulus subspaces, which could be chosen with reference to known properties of the particular area under study.

Also introduced was the idea of coarse-to-fine optimisation, where we initially optimise a low resolution image (with a low dimensionality), and increase the resolution when effective stimuli are found. In Chapter 7 it was found that low-pass images can be effective stimuli for at least some cells in STS and IT. This approach may go some way towards helping with the problem of searching high dimensional spaces that is discussed below.

8.2. Problems Encountered with Optimisation

Optimisation in a neurophysiological context presented considerable technical difficulties. Obtaining a stable recording for a long period of time proved problematic, and typical optimisation attempts only completed 4 iterations before the cell was lost. Stable recordings can be achieved with floating electrodes over longer periods (days – months) and this is therefore an area which can be readily improved upon.

An unexpected problem was that RSVP sequences of the noise images produced by the algorithm seem inherently less rewarding to the animal than RSVP sequences of meaningful or well structured images. It was typically found that the monkey would be eager to view the sequence of meaningful screening images that were initially used to characterise the cell, but lost interest rapidly when presented with the set of images generated by the optimisation software.

Depending on the complexity of the image model, the generation of search images for each iteration of the optimisation can also take a significant time, which introduces long delays between stimulus presentations. This is partly a side-effect of the system design used, with separate computers carrying out the image generation and response recording. With a faster computer carrying out both tasks, it would be possible to carry out the correlations and image generation in closer to real time. A further improvement to the design might be to run multiple optimisations in parallel, with updates and new image generation taking place for one set of parameter values while a stimulus relating to a different parameter set was being displayed. This would compensate for the inherent latency between stimulus display and the response becoming available.

The primary problem however, is the high dimensionality of the search space. Even when the optimisation process is started with an image that is capable of (at least partly) driving a cell, the search method described in this thesis may take steps downwards, producing images that lead to worsened responses from the neuron rather than better ones. Why is this the case? Although there may be a path leading upwards towards the optimal stimulus, the vast majority of possible steps at any one point may actually lead downwards. The situation can be compared to that of a blindfolded man trying to find his way up a narrow and twisting mountain path, with precipitous drops on either side of the way.

Even with a well behaved and regular performance surface (e.g. the quadratic response function used in some of the simulations in this thesis), high dimensionality can make convergence troublesome. Consider the two-dimensional optimisation pictured below, in which we are trying to find the optimal value for 2 parameters, p0 & p1. The dotted circular lines show the contours of a quadratic response function.

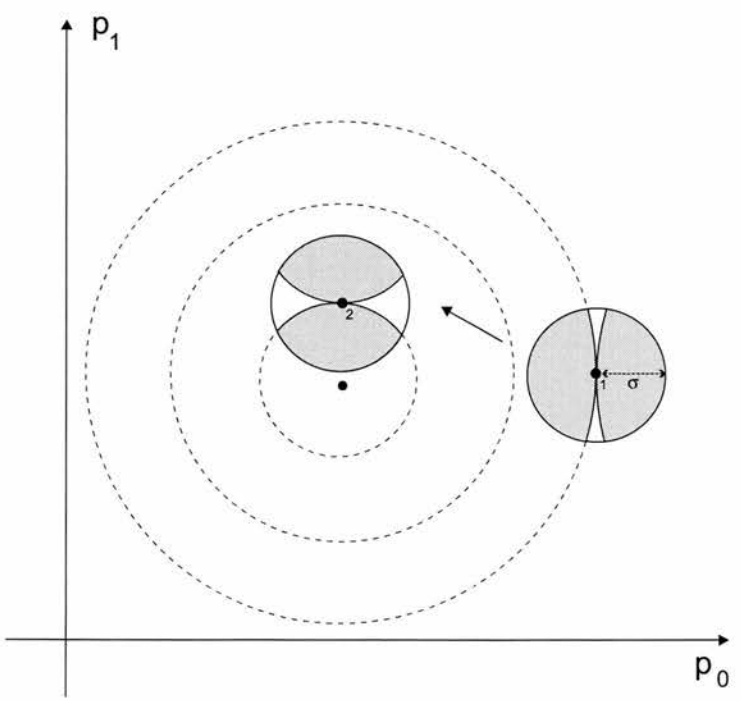


figure 8.1 – Optimisation on a quadratic performance surface. See text for details.

At point 1, the optimisation algorithm will generate random search points with a two dimensional gaussian distribution around the current parameter values. The solid circle around point 1 indicates one standard deviation of this distribution. Consider the following three possibilities -

(i) If the randomly chosen point lands in the left shaded area within this circle, then the new image will result in a better response, and the optimisation will move the parameters towards these new values, up the performance surface.

(ii) Conversely, if the randomly adjusted parameter values produce a point that lies in the right hand shaded area, a worse response is obtained, and the optimisation will proceed directly away from this point, again resulting in a step up the performance surface.

(iii) However, if the point chosen lies within the two unshaded segments of the circle, then not only is a worse response obtained, but when we try to move in the opposite direction, we still end up with a worse response. In this case, the optimisation process will take a step downhill.

When we are far away from the optimal parameter values, these unshaded areas will be small (as at point 1), since the contour lines of the performance function will tend to bisect the distribution of noisy images.

However, nearer the optimal value (e.g. point 2), the curvature of the contour lines is higher, and the unshaded areas occupy a larger proportion of the space in which our

samples fall. Thus, the closer we are to a peak, the higher the probability that optimisation steps will fail to improve the response[18].

As the dimensionality of the search increases, then so does the size of the unshaded area relative to the shaded area. At high degrees of dimensionality it becomes vanishingly unlikely that a randomly chosen point will allow us to proceed uphill.

Furthermore, this example has used a relatively forgiving performance surface, which is unlikely to be a good model for a neural performance surface. If instead the contour lines are thin and elongated, then it is easy to see that it will be much harder to find a point that lies uphill, and the potential for making downwards steps is much increased.

8.3. Colour and Spatial Frequency Response in IT

In the second part of this thesis, I explored with idea of restricting the search to specific stimulus subspaces, with a view of reducing the dimensionality of the optimisation problem. The colour and spatial frequency response of neurons in STS and IT was investigated, inspired by theories of coarse to fine processing in the visual system, and most recently a study by Delorme et al. (2000), which had suggested that the first wave of visual information about a stimulus may be coarse and achromatic. Under RSVP conditions, where stimuli are presented rapidly and continuously, we hypothesised that the response due to colour and high-spatial frequency aspects of a

---

[18] The figure also shows how the search will get closer to finding optimal parameter values as σ is reduced. This effect is illustrated in box 4.1.

stimulus would be progressively diminished as the presentation rate increased, resulting in equivalent responses to both the original stimulus and its low-frequency achromatic counterpart.

The population of neurons studied in IT and STS showed a marked reduction in response to the achromatic and low-spatial frequency stimuli, an effect that persisted, even at the highest presentation rate of 14ms per stimulus. Analysis of the onset of discrimination revealed differential responses to colour and achromatic stimuli during the earliest part of the response, failing to support the hypothesis that the first wave of information about a stimulus was achromatic. Similarly, discrimination between low-spatial frequency stimuli did not precede that of high-spatial frequency stimuli, but had an equal onset time, failing to support the hypothesis that the first wave of visual information consisted of low spatial frequencies only.

Sugase et al. (1999) found that the early component of the response in IT (the first 50ms) signalled only coarse aspects of a stimulus, with fine detail following later. This was attributed by Delorme et al. (2000) to a slower processing speed within the visual system for high spatial frequency and colour information. This thesis has shown that this cannot be the case, since colour and high-spatial frequency information about a stimulus is present from the response onset. Therefore, alternative explanations must be sought to explain Sugase's findings. One possibility is that the emergence of fine detail discrimination over time might reflect the effect of lateral inhibitory connections between several neurons that are initially activated by the stimulus, an effect suggested by many PDP models (see Oram & Perrett, 1992).

We also found that 70% of cells tested in STS/IT were strongly colour tuned, producing differential (and mainly reduced) responses to the luminance matched achromatic and false colour versions of their preferred colour stimulus. Furthermore, at least half of the individual cells examined also showed a change in the ranking of stimulus preference when the colouring of stimuli was manipulated, suggesting a strong role for colour in the stimulus selectivity of cells responsive to complex and naturalistic objects. It has been suggested that object recognition is based on an abstract edge-based representation of objects (Biederman & Ju, 1988) and cells have been described in IT that show invariant responses to objects irrespective of how their boundaries are defined (Vogels & Orban, 1996). The results presented here do not directly support this view, though it seems possible that there is a hierarchy of stimulus representations within IT, perhaps with the outputs of neurons selective for basic properties of objects, such as colour and texture, being combined into higher-level object representations that can be invariant of these properties. Consistent with this idea, this thesis also found a significant negative correlation between colour sensitivity and response latency, such that colour sensitivity was lower in those cells with longer latency, suggestive of the progressive emergence of invariant properties.

In general, the finding of this thesis are broadly positive in terms of using RSVP to rapidly explore the stimulus selectivity of cells in STS and IT. However, the work of Sugase et al. (1999) does suggests that there may be broad activation across a group of cells to a particular stimulus (e.g. a face), followed by competition over the next few tens of milliseconds which leads to a representation encoded by a small number of cells best tuned to the features of the stimulus. If this is the case, RSVP in itself may not be sufficient to characterise in full the stimulus selectivity of these neurons.

8.4. Ideas for Future work

An obvious path for future work to take would be to roughly characterise IT cells by testing them with a large bank of images using RSVP. When the approximate preference of the cell is established (e.g. a preference for geometrical shapes, or faces), an appropriate image model could be selected for optimisation. These image models could each have a very restricted domain, with perhaps just tens of parameters describing e.g. eye size and spacing, face colour, and so on (in the case of faces). Obviously, this lacks much of the generality we hoped for at the start of this work, but is likely to prove a much more tractable problem.

Rather than trying to search for optimal stimuli in IT, an alternative approach might be to progressively characterise the visual system, starting in V1, where optimisation techniques such as the one presented here are likely to be successful. Such an approach would be able to combine the known properties of a particular stage of the visual system with along with theoretical considerations (such as a drive towards sparse representations), thus placing more constraints on the range of stimuli to be tested in order characterise neurons in a subsequent stage.

The RSVP technique also presents an number of unknowns that would be useful to investigate in order to progress with the optimisation methodology. For example, it would be very interesting to know if any habituation occurs for repeatedly presented stimulus over the sort of timescales relevant for optimisation (i.e. tens of minutes), and furthermore to know how similar images need to be to produce habituation. e.g. do cells habituate to faces in general, or just a particular view of a specific face. Trying to find the optimal stimulus is a hard enough problem without a moving target.

8.5. Concluding Comments

The work described in this thesis represents a first attempt to apply optimisation techniques to the complex receptive fields found in STS and IT, and will hopefully inspire future work in this direction. With hindsight, the work was perhaps too ambitious and general in its approach, and future attempts may find it profitable to constrain the generality of the image model while the effectiveness of different search algorithms are evaluated on the relatively unknown responses surfaces of these neurons.

9. Appendix 1 – Stimuli used in Experiments

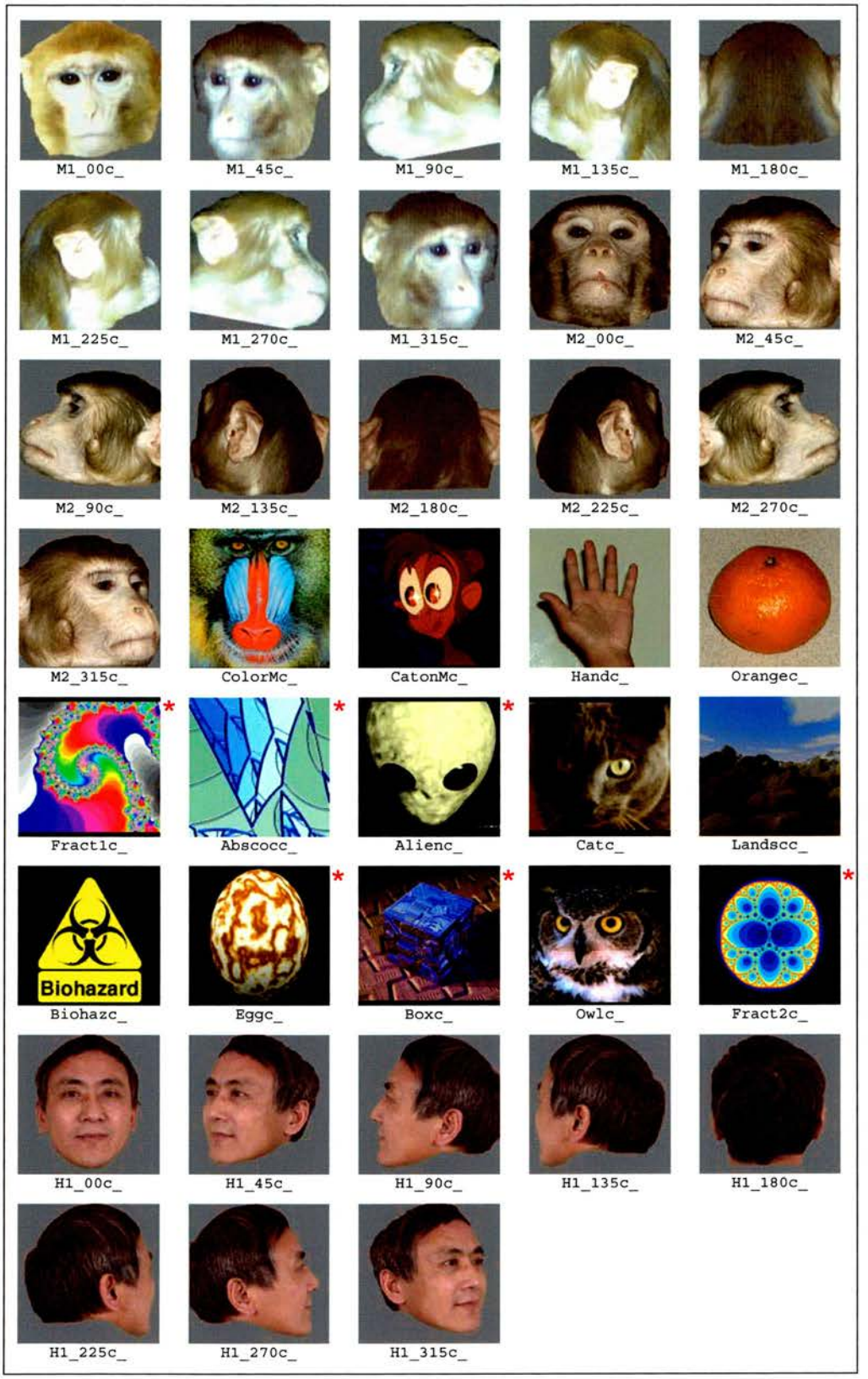


figure A1.1 – The full set of 38 colour images used in the experiment. For each image there were corresponding achromatic and false-colour versions (not shown). Images classified as abstract for the purposes of cell classification are indicated with a red asterisk to their upper right.

10. Appendix 2 – Table of Cell Properties

10.1. Cells in Colour & Frequency Experiments

Pen	Cell	S14G0			S28G0				S56G0				S56G167				
		best	worst	latency	best	worst	latency	csi	fsi	best	worst	latency	csi	fsi	best	worst	latency	csi	fsi
207L	9				H1_135c_	Fract2c_		112	-0.02	0.00	H1_135c_	Eggc_		144	0.00	0.00	H1_135c_	Biohazc_		113	-0.09	-0.06
208L	4				M2_130c_	Handc_		88	0.11	0.06	M2_180c_	Eggc_		95	0.14	0.10	M2_180c_	Handc_		105	0.48	0.41
209L	10				H1_00c_	Handc_		81	0.26	0.22	H1_00c_	Boxc_		86	0.28	0.31	H1_00c_	Boxc_		102	0.02	0.54
210L	3				M2_315c_	Eggc_		84	0.17	-0.09	M2_315c_	Eggc_		93	0.17	-0.15	M2_315c_	Eggc_		97	0.14	-0.27
211L	4				M2_315c_	Orangec_		66	0.04	0.04	M2_315c_	Orangec_		69	-0.02	0.01	M2_315c_	H1_135c_		65	0.12	0.20
212L	5	H1_315c_	Fract2c_	64	H1_315c_	Fract2c_		62	0.31	-0.14	H1_315c_	Fract2c_		65	0.37	-0.20	H1_315c_	Fract2c_		70	0.32	-0.03
212L	6	H1_00c_	Abscocc_	65	H1_00c_	Abscocc_		68	0.14	0.12	H1_00c_	Landscc_		75	0.15	0.24	H1_00c_	Abscocc_		78	0.11	0.36
212L	9				Boxc_	Calc_		100	0.00	0.01	Boxc_	H1_45c_		116	0.13	0.04						
214L	6				Handc_	ColorMc_		92	0.15	-0.05	Handc_	Eggc_		91	0.16	-0.08	Handc_	Fract2c_		103	0.50	0.75
214L	7				M2_00c_	Orangec_		82	0.19	-0.13	M2_00c_	Boxc_		123	0.24	-0.24	M2_00c_	Boxc_		155	-0.07	-0.13
214L	9				M2_00c_	Boxc_		80	0.12	-0.13	M2_00c_	Eggc_		110	-0.03	-0.01	M2_00c_	Calc_		102	0.00	0.25
215L	6				H1_00c_	Abscocc_		65	0.55	-0.14	H1_00c_	Abscocc_		66	0.70	-0.26	H1_00c_	Abscocc_		68	0.69	0.15
215L	7	H1_315c_	Fract2c_	56	H1_315c_	Fract2c_		62	0.49	-0.01	H1_315c_	Fract2c_		65	0.65	-0.06	H1_315c_	Fract2c_		79	0.78	0.32
217L	14										H1_00c_	Biohazc_		138	0.10	-0.09	H1_00c_	M1_00c_		143	0.43	-0.12
218L	1				Calc_	Abscocc_		112	0.16	-0.02	Calc_	Abscocc_		105	0.50	0.11	Calc_	Abscocc_		107	0.43	0.37
221L	12										Calc_	Eggc_		130	0.11	0.11	Eggc_	M1_180c_		58	-0.01	-0.11
227L	10	M1_00c_	Owlc_	83	M1_00c_	Owlc_		78	0.37	-0.17	M1_00c_	Owlc_		89	0.45	-0.25	M2_00c_	Fract1c_		81	0.22	-0.09
227L	7				Fract2c_	Alienc_		57	0.48	-0.11	M1_00c_	M2_00c_		83	0.59	-0.41	Eggc_	M2_00c_		58	0.63	-0.08
227L	8				Calc_	H1_00c_		67	0.24	0.03	Landscc_	Eggc_		96	0.40	-0.15	Landscc_	H1_00c_		80	0.58	-0.42
227L	9				M1_00c_	Owlc_		71	0.52	-0.20	M1_00c_	Owlc_		81	0.60	-0.55	M1_00c_	Owlc_		80	0.91	-0.46
228L	3				M2_90c_	M2_00c_		92	0.02	-0.02	M2_225c_	M2_00c_		98	0.26	0.40	M2_90c_	M2_00c_		98	0.13	0.06
230L	10				M2_00c_	Eggc_		115	0.23	-0.17	M2_00c_	Biohazc_		118	0.41	0.01	Alienc_	Eggc_		110	-0.19	-0.29
231L	6				H1_135c_	ColorMc_		71	0.17	-0.06	H1_135c_	Eggc_		88	0.09	-0.16	H1_135c_	Eggc_		87	-0.11	-0.21
233L	4				Orangec_	Calc_		60	0.07	-0.20							Fract2c_	Calc_		55	0.22	-0.44
235L	6				H1_270c_	Abscocc_		80	0.23	-0.18	H1_270c_	Boxc_		97	0.33	0.06	H1_270c_	Boxc_		77	0.04	0.57
235L	7				H1_90c_	Eggc_		72	0.16	0.07	H1_90c_	Abscocc_		72	0.29	0.16	H1_90c_	Abscocc_		97	0.58	0.99
238L	4				Abscocc_	H1_00c_		81	0.08	0.22	Abscocc_	Orangec_		98	-0.03	0.33	Abscocc_	Orangec_		151	-0.03	0.17
238L	5				Handc_	Eggc_		84	0.15	0.25	Handc_	Eggc_		95	0.09	0.31	Handc_	Eggc_		95	0.02	0.38
239L	5				H1_00c_	ColorMc_		73	0.07	0.16	M2_00c_	ColorMc_		78	-0.01	0.25	H1_00c_	Handc_		78	0.53	0.44
240L	4	H1_00c_	Fract2c_	69	H1_00c_	Fract2c_		65	0.49	-0.08	H1_00c_	Fract2c_		67	0.55	-0.01	H1_00c_	Alienc_		79	0.63	0.25
240L	6				Abscocc_	Eggc_		115	-0.05	0.07	Abscocc_	Eggc_		112	-0.04	0.16	Abscocc_	Handc_		99	0.12	0.24
240L	7				H1_180c_	M1_180c_		79	0.05	0.35	H1_180c_	Boxc_		87	0.00	0.50	H1_180c_	Boxc_		87	0.04	0.45
244L	10				M1_00c_	Eggc_		83	0.04	0.06	M1_00c_	Eggc_		94	0.16	-0.02	M1_00c_	Calc_		108	0.07	-0.08
244L	9				M2_180c_	Handc_		79	0.27	0.14	M2_180c_	Boxc_		76	0.31	0.48	M2_180c_	Boxc_		93	0.17	0.34
245L	3				M1_315c_	Calc_		106	0.08	-0.12												
245L	5										Biohazc_	Eggc_		88	0.43	-0.01						
251L	1																M2_00c_	Boxc_		115	-0.07	0.25
252L	3				M2_315c_	Abscocc_		82	-0.02	0.09	M2_315c_	Abscocc_		82	0.03	0.13	M2_315c_	Alienc_		92	0.12	0.41
252L	5				M2_315c_	Biohazc_		67	0.10	0.12	M2_315c_	Biohazc_		77	0.04	0.29	M2_315c_	Biohazc_		67	0.02	0.41
253L	6				Alienc_	Orangec_		102	0.01	-0.04	Alienc_	M1_270c_		100	-0.03	-0.01						
253L	8				Handc_	Eggc_		103	0.24	0.11	Handc_	Owlc_		105	0.14	0.06	Handc_	Eggc_		129	0.21	0.25
254L	7				Boxc_	M1_270c_		95	0.21	0.20	Boxc_	M1_00c_		103	0.39	0.35	Boxc_	M1_00c_		117	0.07	0.52
255L	3				Alienc_	M2_00c_		108	0.18	0.08	Alienc_	Biohazc_		119	0.14	0.06	Alienc_	Landscc_		110	0.10	-0.18
255L	4				Alienc_	Eggc_		117	0.08	0.00	Alienc_	Eggc_		102	0.23	-0.05	Alienc_	Eggc_		110	0.20	0.04
255L	5				Orangec_	Eggc_		86	0.38	0.13	Orangec_	Boxc_		89	0.28	0.16	Orangec_	Owlc_		88	0.17	-0.03
260L	6				M2_00c_	Boxc_		69	0.05	-0.05	M1_00c_	Boxc_		80	0.28	-0.14	M2_00c_	Abscocc_		60	0.15	-0.09
280L	9				M2_00c_	Orangec_		60	0.09	0.03	M1_45c_	Eggc_		81	0.06	0.38	M2_00c_	Eggc_		69	0.03	-0.11
266L	15	Abscocc_	Eggc_	74	Abscocc_	Eggc_		84	-0.04	0.22	Abscocc_	Eggc_		79	0.01	0.49	Abscocc_	Eggc_		96	-0.07	0.39
268L	2				M2_90c_	Eggc_		85	-0.02	0.06	M2_90c_	Biohazc_		98	0.15	-0.27	M2_225c_	Biohazc_		101	0.10	0.16
268L	4	M2_270c_	Eggc_	84	M2_270c_	Boxc_		82	0.13	-0.01	H1_270c_	Eggc_		96	0.24	-0.16	H1_270c_	Eggc_		108	0.07	0.10

Notes -
(1) Top shaded area shows experiments where no false colour images were shown
(2) Bottom shaded area shows experiments where screening phase contained both colour and achromatic images

10.2. Cells in Contrast Experiment

Pen	Cell
302	4
304	10
307	10
307	15
307	16
311	3
311	6
312	9
313	13
314	2
314	6
314	8
316	7
317	3
321	8
321	10

11. Appendix 3 – Reconstruction of Electrode Tracks and Cell Positions

11.1. X-Rays

11.1.1. Measurements

Lateral and frontal X-rays were taken at the end of every recording session and show the position of the electrode relative to bone landmarks. The trajectory and position of the electrode tip were calculated from the X-ray relative to a co-ordinate system superimposed on the visible landmarks. For frontal X-rays, a horizontal axis was drawn between the two auditory canals, and a second, vertical axis, perpendicular to the first, at the midpoint of the skull. The position (x, y) of the electrode tip was then measured, along with the position of a second point on the path of the electrode (20mm away from the tip), in order to allow the trajectory to be established. For lateral X-rays, a horizontal axis was drawn between the external auditory meatus and the orbital ridge, and a vertical axis, perpendicular to the first, also drawn through the auditory meatus. As with the frontal X-rays, measurements (z, y) of the tip of the electrode and a point along its path were then taken.

Because auditory canals were not always visible on frontal X-rays, only the medio-lateral measurement (x) was retained, and was combined with the superior-inferior (y) and anterior-posterior (z) measurements from the lateral X-rays to provide a complete 3-dimensional position.

11.2. Histology

11.2.1. Final Recording and Microlesions

At the end of the final recording session, electrolytic microlesions were made, that marked approximate anterior and posterior boundaries of the recording region. Three

lesions were made in total, consisting of a single large posterior lesion (current of 40μA for 30 seconds) and two smaller anterior lesions at differing depths (shallow – 20μA, 20s.; deep – 30μA, 30s). As in all previous recording sessions, frontal and lateral X-rays were taken.

11.2.2. Perfusion

The monkey was given an injection of ketamine to sedate it and, after 10-15 minutes, a lethal dose of barbiturate (sagatal). After a few minutes the effect of the barbiturate was verified by the absence of the gabella reflex (closure of the eyelids following contact with cornea).

The electrode and stereotaxic apparatus were kept in place throughout perfusion. The monkey was removed from the chair, and the thorax cut open to expose the heart. The pericardium was removed and a large bore cannula inserted into the left ventricle. The descending aorta was clamped so that the upper torso and head only were perfused. An incision was made in the right auricle to allow outflow of fluid from the circulation.

Solution was passed through the large bore canula and into the circulation using a mechanical centrifugal pump (C16-C, Charles Ansten Pumps Ltd.). A pre-fixative wash of phosphate buffered saline and 0.2% sodium nitrate (for vasodilation) was passed through the monkey to remove blood from the system. Approximately 5 litres of solution was required to flush out all the blood. The perfusing liquid was then changed to a phosphate buffered fixative solution (4% paraformaldehyde). After approximately 5 litres of fixative was passed through the monkey, the muscles of the head and neck went rigid and perfusion was complete. The cannula was removed

from the heart, and the head severed from the body.

11.2.3. Histology

The head was first placed in a stereotaxic frame with ear bars and orbital ridge grips, and skin and muscle were removed. The skull, including the implant and stainless steel wells, were then removed using bone cutters and a drill. The brain was removed from the skull and the frontal lobes cut away. The brain was then sunk in successively higher concentrations of sucrose solution (10, 15, 20, 30%) over a period of one month. Before sectioning, the brain was immersed in a bath of isopentane, cooled to below minus 40°C using dry ice (CO_2). After 25 minutes, the brain was removed from the isopentane and placed in a cryostat (Bright Instruments Co., Huntingdon, UK) at minus 15°C, with the cerebellum forming a base for the sectioning. The brain was left for two hours to equilibrate in temperature.

The brain was sliced into sections of 25μm thickness. A digital photograph was taken every 250μm, and the section retained and placed in a set of bays filled with 0.1M phosphate buffer and 0.9% NaCl. The retained sections were transferred to dishes containing water and guided onto glass microscope slides. Once dry, the sections were Nissl stained (for cell bodies) and coverslipped. Drawing were made of one in every four sections (i.e. with 1mm spacing) from the grey and white matter boundaries as visible from the digital photographs.

11.3. Co-registration of X-rays and Histology

In order to determine the position of cells relative to the structures of the brain, it is necessary to establish a mapping between X-ray co-ordinates and photographs of the

actual tissue sections.

First, the location of the large posterior lesion was identified on one of the Nissl stained slides (figure A3.2). This provided not only a common point in the two co-ordinate systems, but also a scale factor for the medio-lateral (x) axis, since a distance to the midline could be measured in both cases. In the absence of definitive locations for the remaining two anterior lesions, uniform tissue shrinkage was assumed, allowing this scale factor to be applied to the remaining two axes. Each cell could now be located to a known point on a known section.

Confidence in the accuracy of this mapping was greatly increased when the calculated positions of the two anterior lesions were found to fall on two areas of tissue that had been provisionally, but not conclusively, identified as the locations of these lesions.

11.4. Electrode Tracks and Cell Positions

The figures that follow show a sequence of coronal sections (traced from the digital photographs) with the individual electrode tracks and cell positions superimposed on the nearest section. The spacing between sections is 1mm, with the most anterior section first and the most posterior section last. In stereotaxic co-ordinates, these slices are located from approximately + 19.5 mm to + 8.8 mm on the A-P axis.

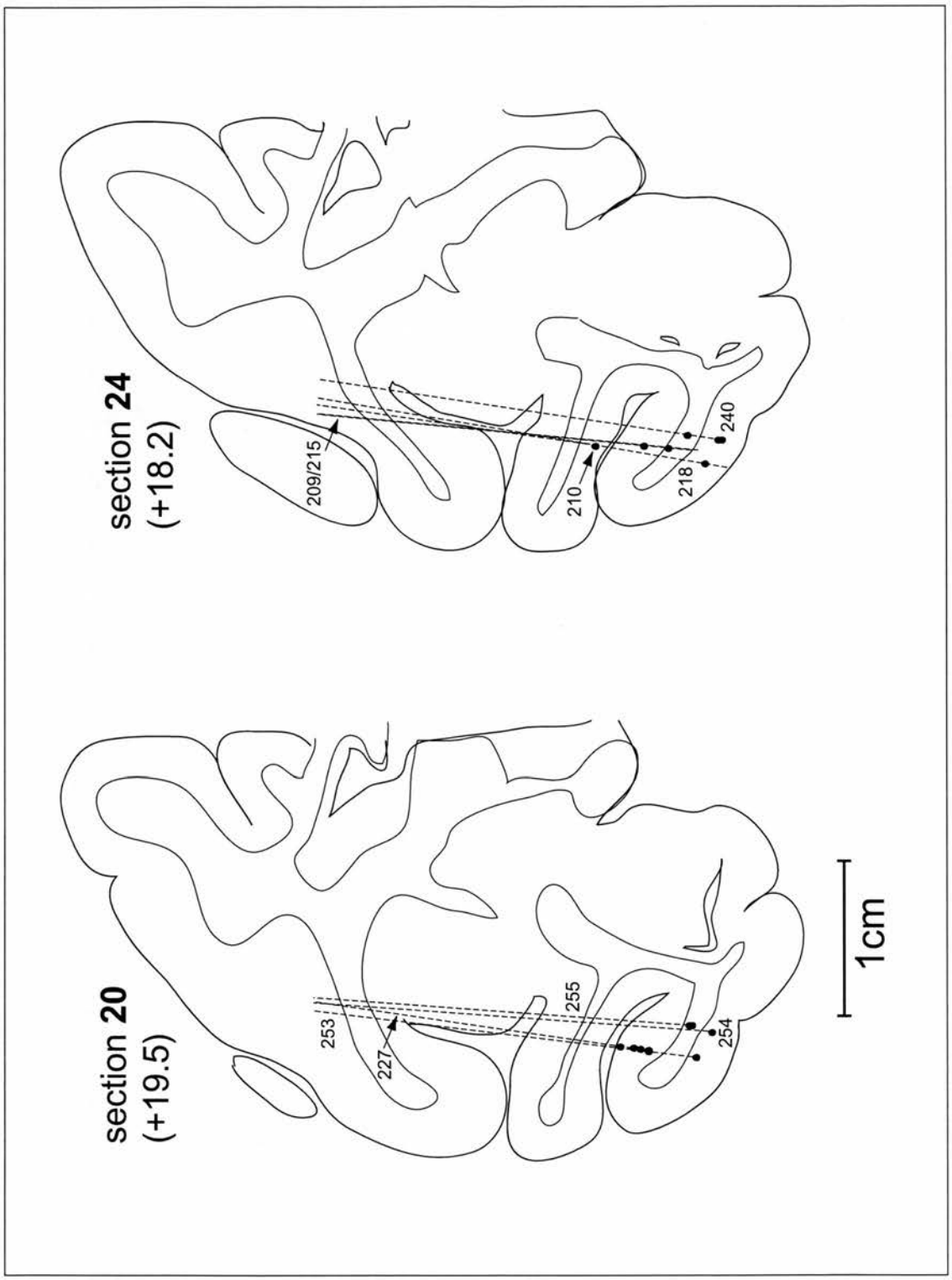


figure A3.1 – Coronal sections of the left hemisphere showing electrode tracks (dashed lines) and penetration numbers, cell positions (filled circles) and lesion positions (outlined circles).

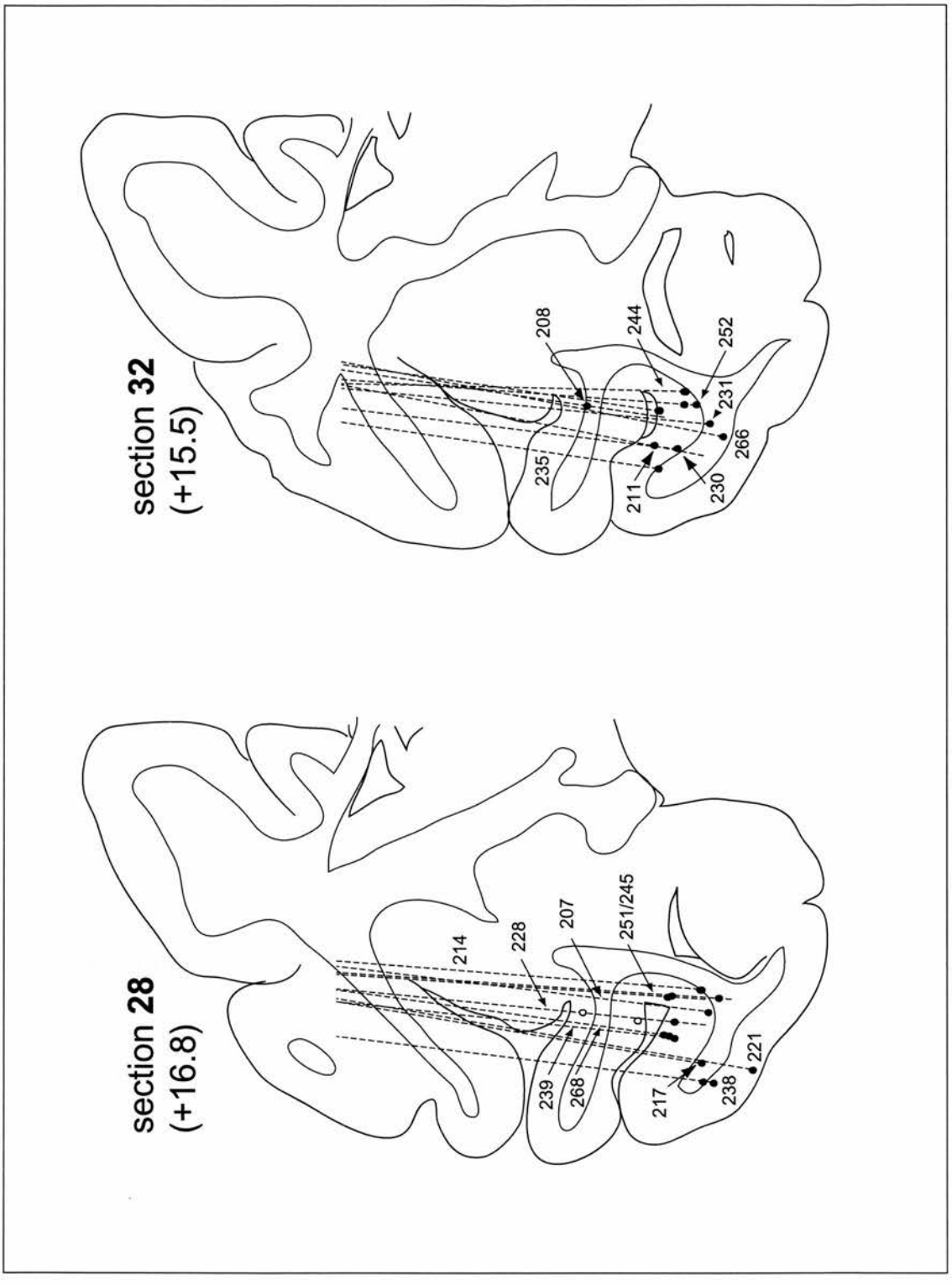


figure A3.1 (contd) – Coronal sections of the left hemisphere showing electrode tracks (dashed lines) and penetration numbers, cell positions (filled circles) and lesion positions (outlined circles).

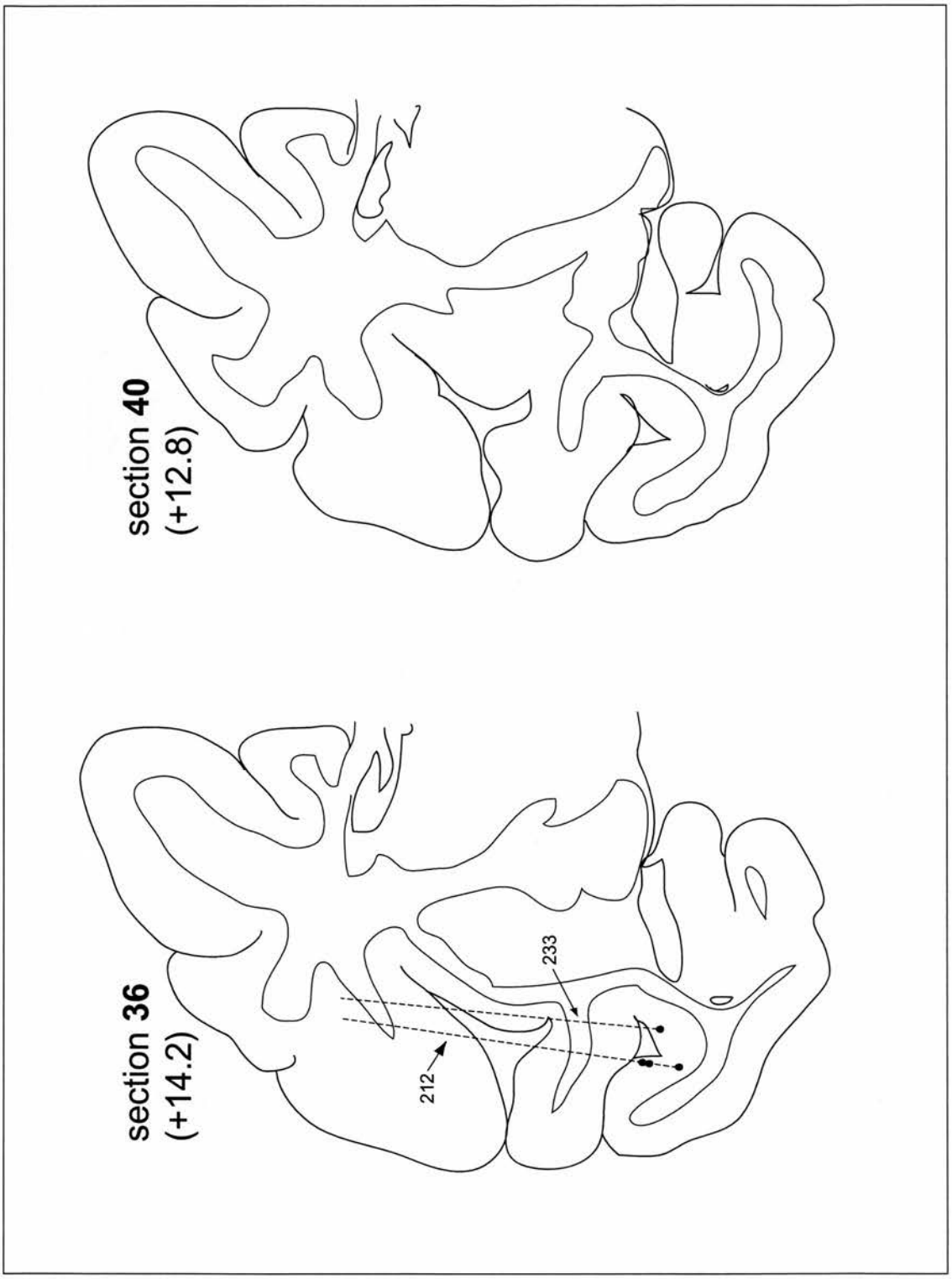


figure A3.1 (contd) – Coronal sections of the left hemisphere showing electrode tracks (dashed lines) and penetration numbers, cell positions (filled circles) and lesion positions (outlined circles).

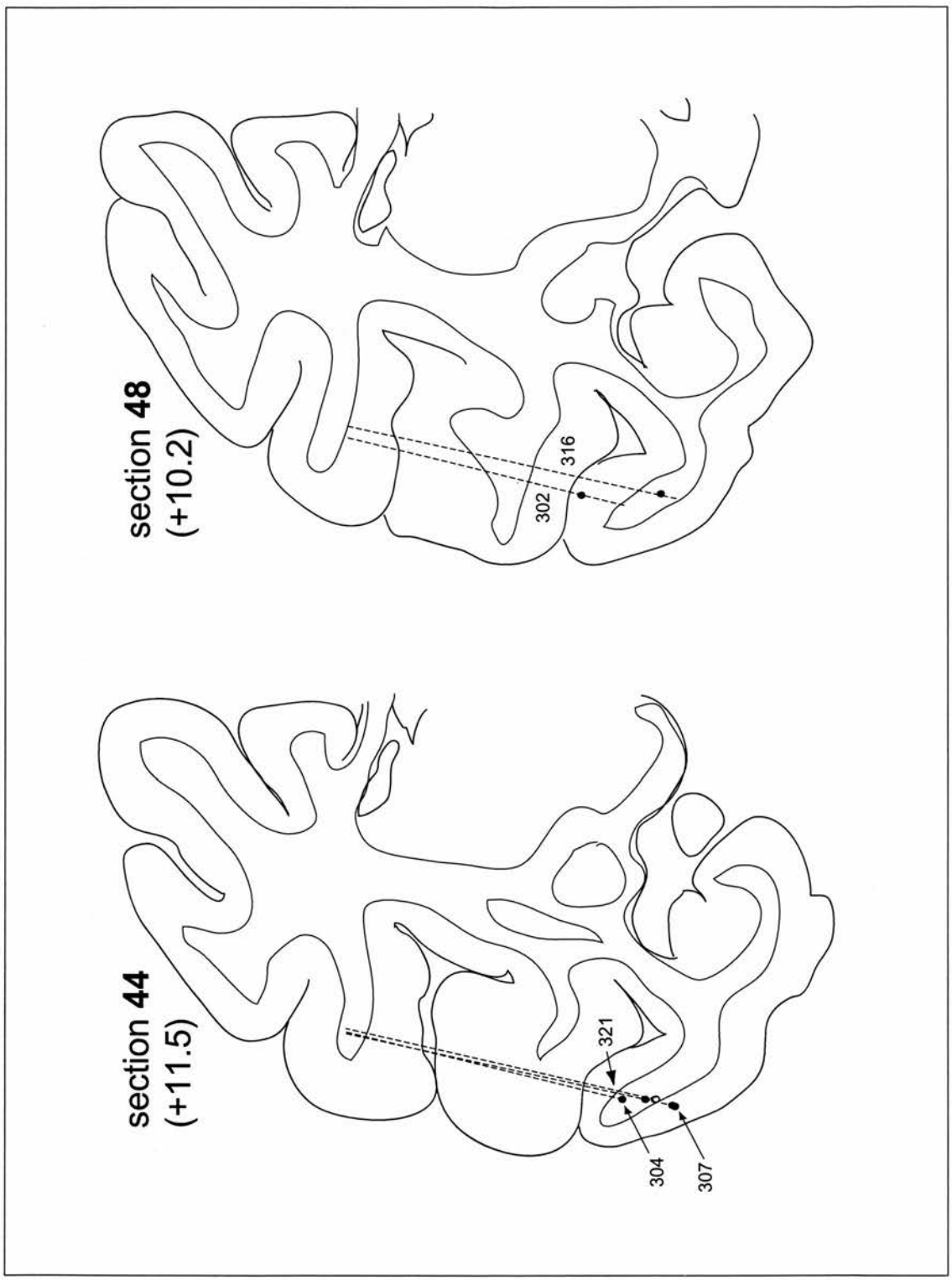


figure A3.1 (contd) – Coronal sections of the left hemisphere showing electrode tracks (dashed lines) and penetration numbers, cell positions (filled circles) and lesion positions (outlined circles).

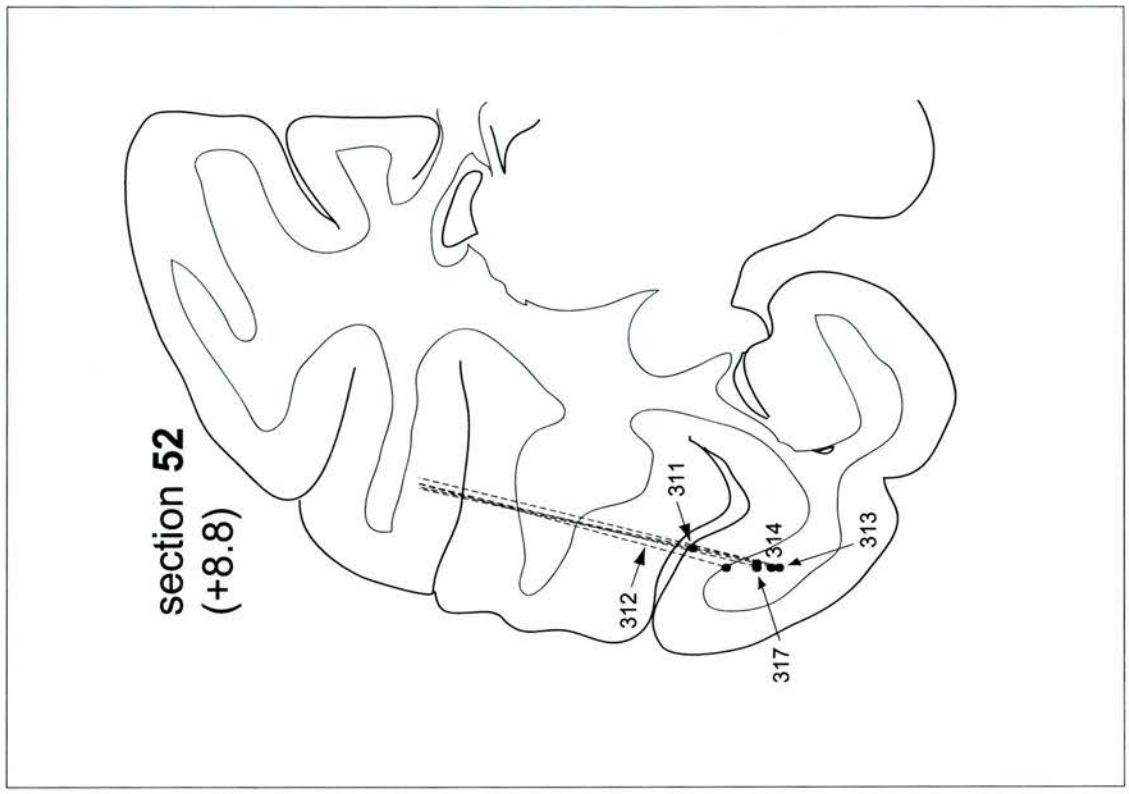


figure A3.1 (contd) – Coronal sections of the left hemisphere showing electrode tracks (dashed lines) and penetration numbers, cell positions (filled circles) and lesion positions (outlined circles).

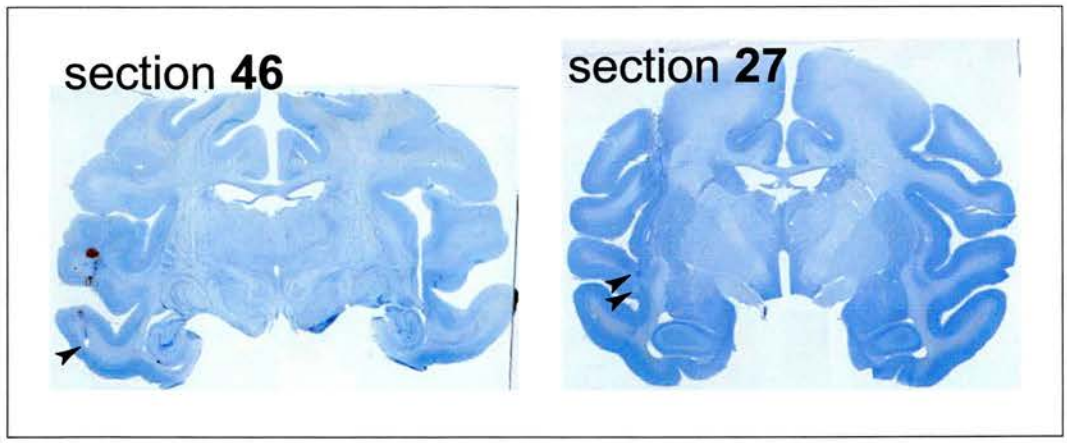


figure A3.2 – Nissl stained sections of the lesion sites. *Left* – the posterior lesion is indicated by an arrow. *Right* – the arrows indicate the probable locations of the smaller anterior lesions. These sites lie almost exactly in the region predicted by the co-registration and plotted on section 28 in the previous figure.

12. Appendix 4 – Optimisation C Code

This appendix contains a listing (in the C language) and commentary on the computer code written to implement the optimisation algorithm and interface it with the physiology system. It is intended to provide further detail, if required, to the description in Chapter 4.

Note that, for clarity and ease of navigation, each new function declaration is printed in bold type. Comments have been added in a different typeface, and follow the region of code that the comment relates to. The ↓ symbol indicates a single line of code which has been split between multiple lines.

12.1. optimise.c

```
/* Stimulus Optimisation
 * Robin Edwards 3/2/2000
 * Version 4.21a
 * optimise.c
 */

// 3.0  - [24/ 2/2000] first version. Tested OK with artificial neuron
// 3.01 - [28/ 2/2000] list passed to optimise() changed from type int to float
// 3.02 - [ 1/ 3/2000] now possible to change parameters during optimisation
// 4.0  - [27/ 3/2000] major rewrite to sort out the image model framework
// 4.01 - [29/ 3/2000] 1/f model complete. Combined fft.c & nrutil.c
// 4.02 - [08/ 5/2000] minor bug with float vs int type
// 4.03 - [12/ 6/2000] changed variance to st.deviation
// 4.04 - [13/ 6/2000] added decay term
// 4.1  - [19/ 6/2000] **changed image models**
// 4.2  - [10/ 8/2000] added arbitray basis function image model (Gabors implemented)
// 4.21 - [28/ 8/2000] initialise images with coefficient files

// Questions to resolve with Peter:
// Decay (& u) once per image or per iteration?

// CC optimise.c main.c fft.c gabor.c -lm -lgl -lX11 -w
// CC Tser06.c optimise.c fft.c gabor.c -lgl -lX11 -laudio -lm -w -s -o neuron9

/* Parameters to the optimisation (from struct para in Myvso04.h)
 * blocks_x       - x dimension of the grid
 * blocks_y       - y dimension of the grid
 * block_width    - pixel width of element
 * block_height   - pixel height of element
 * colour         - colour(YCbCr) or luminance based optimisation?
 * learning_rate  - how much we move every step
 * noise_stdev    - how noisy the images are
 * decay          - decay term
 * max_images     - how many noisy images to make
 * /

/* DEFINITIONS ***************************************************************************/
#define kMODEL_IDENTITY       0
#define kMODEL_CORRELATED     1
#define kMODEL_BASIS          2

#define kNOISY_ONLY           0
#define kNOISY_BASE_ORIGINAL  1

/* OPTIONS *******************************************************************************/
const int image_model = kMODEL_BASIS;
```

```
const int search_image_set = kNOISY_BASE_ORIGINAL;

/* INCLUDES ************************************************************************************/
#include "fft.h"
#include "gabor.h"
#include "Myvso04.h"

/* LOCAL STRUCTURES ****************************************************************************/
// image parameter set structure
struct img_pset {
     int blocks_x, blocks_y;
     int block_width, block_height;
     int colour;

     int n;          // number of parameters
     float *vals;    // pointer to values
};
```

This data structure stores a set of parameters encoding the image being optimised. The actual representation of the image (and also the number of parameters) depends on the image model in use. There may be multiple instances of this structure if more than one image is being optimised simultaneously.

```
// stimulus structure
struct stim {
     int orig;       // 1 = original parameters, 0 = current
     int pset;       // parameter set #
     long seed;      // random seed (0 = none)
};
```

This structure is created for every stimulus that is displayed. It records which parameter set an image is derived from, and the seed for added noise.

```
typedef unsigned char    byte;

/* LOCAL VARIABLES *****************************************************************************/
int num_psets;                          // number of parameter sets
struct img_pset *psets = NULL;          // pointer to parameter sets
struct img_pset *orig_psets = NULL;     // pointer to original parameter sets
struct stim *stims = NULL;              // stimuli
int iteration;                          // current iteration
char filename[32];                      // results filename
int pre_t, dis_t, aft_t;
char tmp[200];                          // error message buffer
static int iset = 0;                    // moved outside gasdev() so can be reset

basis_fn basis[BASIS_SIZE];

/* LOCAL COPIES
   These parameters can be changed from the PC user iterface between trials
   so we must keep copies here of the previous values */
float old_noise_stdev;
int old_colour;
int o_blocks_x, o_blocks_y, o_block_width, o_block_height;

extern struct stimulus_image *image_data;
extern struct parameters *para;
extern int serpt;

/* INTERFACE FUNCTION PROTOTYPES ***************************************************************/
int optimise(float *list, int flag);

/* LOCAL FUNCTION PROTOTYPES *******************************************************************/
int optimise_begin(float *image_list);
int update_parameters(float *response_list);
void optimise_end(void);

int write_data(float *response_list);
int make_img_set();
int default_pset(struct img_pset *pset);
int copy_pset(struct img_pset *src, struct img_pset *dst);
int resample_parameters();

// forward transforms
int pset_to_image(struct img_pset *pset, long seed, struct stimulus_image *img);
int identity_transform_f(struct img_pset *pset, float *nvals, float **ch_Y, float **ch_Cb, float↓
 **ch_Cr);
int basis_transform_f(struct img_pset *pset, float *nvals, float **ch_Y, float **ch_Cb, float↓
 **ch_Cr);
void scaled_basis_transform(float *src, float **ch, int width, int height);
int freq_transform_f(struct img_pset *pset, float *nvals, float **ch_Y, float **ch_Cb, float↓
 **ch_Cr);
void scaled_ift(float *src, float **ch, int width, int height);
int matrix_to_image(float **ch_Y, float **ch_Cb, float **ch_Cr, struct stimulus_image *img, int↓
 nx, int ny, int bw, int bh);
```

```
// backward transforms
int image_to_pset(struct stimulus_image *img, struct img_pset *pset);
int identity_transform_b(float **ch_Y, float **ch_Cb, float **ch_Cr, struct img_pset *pset);
int freq_transform_b(float **ch_Y, float **ch_Cb, float **ch_Cr, struct img_pset *pset);
void scaled_fft(float **ch, float *dst, int width, int height);
int load_coefficients(struct stimulus_image *img, struct img_pset *pset);
int image_to_matrix(struct stimulus_image *img, float **ch_Y, float **ch_Cb, float **ch_Cr, int
 nx, int ny, int bw, int bh);

void RGB_to_YCbCr(float r, float g, float b, float *Y, float *Cb, float *Cr);
void YCbCr_to_RGB(float Y, float Cb, float Cr, float *r, float *g, float *b);
void get_pixel_rgb(struct stimulus_image *img, int x, int y, byte *r, byte *g, byte *b);
void put_pixel_rgb(struct stimulus_image *img, int x, int y, byte r, byte g, byte b);
float gasdev(void);

/* INTERFACE FUNCTIONS ************************************************************************/

/* int optimise(float *list, int flag)
 * Interface to the rest of the program.
 * flag = 0: begin optimisation (*list = list of images)
 * flag = 1: update parameters  (*list = list of responses)
 * flag = 2: end optimisation   (*list = NULL)
 * returns 1 for success, 0 for failure
 */
int optimise(float *list, int flag)
{
     switch (flag)
     {
          case 0:
               if (!optimise_begin(list))
                    return 0;
               break;
          case 1:
               if (!update_parameters(list))
                    return 0;
               break;
          case 2:
               optimise_end();
               return 1;
     }

     old_noise_stdev = para->noise_stdev;
     old_colour = para->colour;
     o_blocks_x = para->blocks_x;
     o_blocks_y = para->blocks_y;
     o_block_width = para->block_width;
     o_block_height = para->block_height;

     sprintf(tmp,"%d total\n", para->total_images);
     write(serpt,tmp,strlen(tmp));
     return 1;
}
```

This is the entry point from the physiology system. When called initially, it can be passed a list of one or more images to be optimised. In subsequent calls, it is passed a list of windowed spike counts that describe the response to each of the stimuli.

```
/* LOCAL FUNCTIONS ****************************************************************************/

/* int optimise_begin(float *img_list)
 * Parameterises the images passed to the function in image_list
 * img_list[0]    : number of images
 * img_list[1..n] : indices of each image
 * if passed a negative number of images, create that many default parameter sets)
 * (minor hack = can't send -1 images, so we must subtract 1)
 */

int optimise_begin(float *img_list)
{
     int i;
     if (!img_list) return 0; // null pointer check

     srand48(time(0));        // random seed
     iteration = -1;

     // take some defaults from image 0
     pre_t = (image_data+0)->pre_t;
     dis_t = (image_data+0)->dis_t;
     aft_t = (image_data+0)->aft_t;

     // initialise basic functions
     if (image_model == kMODEL_BASIS)
     {
          if (para->blocks_x == para->blocks_y)
          {
               sprintf(tmp,"2\ncreating basis fns...\n");
```

```
            write(serpt,tmp,strlen(tmp));
            create_basis(para->blocks_x);
        }
        else
        {
            sprintf(tmp,"1\nimage dimensions must match!\n");
            write(serpt,tmp,strlen(tmp));
            return 0;
        }
    }

    num_psets = (int)img_list[0];
    if (num_psets < 0) num_psets = abs(num_psets) - 1;

    // allocate memory for parameter sets
    psets = (struct img_pset *)malloc(sizeof(img_pset) * num_psets);
    orig_psets = (struct img_pset *)malloc(sizeof(img_pset) * num_psets);

    if (!psets || !orig_psets) return 0;     // malloc error

    for (i=0; i<num_psets; i++)
    {
        (psets+i)->vals = NULL;
        (orig_psets+i)->vals = NULL;
    }

    // now create the parameter sets
    if ((int)img_list[0] < 0)
        // no images passed, create default parameters
        for (i=0; i<num_psets; i++)
        {
            sprintf(tmp,"2\ndefault pset %03d\n", i+1);
                write(serpt,tmp,strlen(tmp));

            if (!default_pset(psets+i))
            {
                sprintf(tmp,"1\nERROR: default pset %03d\n", i+1);
                write(serpt,tmp,strlen(tmp));
                return 0;
            }
        }
    else
        // otherwise encode each image
        for (i=0; i<num_psets; i++)
        {
            sprintf(tmp,"2\ncoding img %03d\n", i+1);
                write(serpt,tmp,strlen(tmp));

            if (!image_to_pset(image_data+((int)img_list[i+1]), psets+i))
            {
                sprintf(tmp,"1\nERROR: coding img %03d\n", i+1);
                write(serpt,tmp,strlen(tmp));
                return 0;
            }
        }
```

At the beginning of the optimisation process, the last set of images displayed by the physiology system is still held in memory. One or more of these images can be used as a starting point for the optimisation. Each image selected must be encoded into a set of parameters.

```
    // backup the parameter sets
    for (i=0; i<num_psets; i++)
        if (!copy_pset(psets+i, orig_psets+i))
        {
            sprintf(tmp,"1\nERROR: copying pset %03d\n", i+1);
            write(serpt,tmp,strlen(tmp));
            return 0;
        }
```

A backup is made of each of the parameter sets that have just been calculated. This allows an unoptimised image to be included in every iteration, as a baseline measure.

```
    if (!make_img_set())
```

This function creates the first set of stimuli by adding noise to the parameters set(s).

```
    {
        sprintf(tmp,"1\nERROR: making img set\n");
        write(serpt,tmp,strlen(tmp));
        return 0;
    }

    sprintf(tmp,"2\nOkay.\n");
    write(serpt,tmp,strlen(tmp));

    return 1;
}
```

```
/* int update_parameters(float *response_list)
 * Updates the parameters based on the responses contained in response_list
 * response_list[0] = number of responses
 * response_list[n] = response n
 */
int update_parameters(float *response_list)
```

A set of stimuli (images derived from noisy versions of each parameter set) has now been presented, and we have been passed a list containing the response to each stimulus.

```
{
     int i,j,k,n;
     float mean;
     float c;        // constant for each image
     float *dst, *orig;
     float decay;

     if (!response_list) return 0;

     // check we have the right number of spike counts
     if ((int)response_list[0] != para->total_images)
     {
          sprintf(tmp,"1\nbad # of responses [r=%03d, i=%03d]\n",(int)response_list[0],para↓
 ->total_images);
          write(serpt,tmp,strlen(tmp));
          return 0;
     }

     if (!write_data(response_list))
     {
          sprintf(tmp,"1\nERROR: writing results to disk\n");
          write(serpt,tmp,strlen(tmp));
          return 0;
     }

     for (i=0; i<num_psets; i++)
     {
          sprintf(tmp,"2\nUpdating param %03d\n",i+1);
          write(serpt,tmp,strlen(tmp));

          // calculate the mean response for each parameter set
          // (ignoring any noiseless images)
          mean = 0.0; n = 0;
          for (j = 0; j < para->total_images; j++)
               if (stims[j].pset == i && stims[j].seed) {
                    mean += response_list[j+1];
                    n++;
               }
          mean /= n;

          // add each noisy image to the parameter set
          for (j = 0; j < para->total_images; j++)
          {
               if (stims[j].pset == i && stims[j].seed) {

                    srand48(stims[j].seed); iset=0;         // seed the random number stream
                    c = response_list[j+1] - mean;
                    c *= para->learning_rate * old_noise_stdev;
                    dst = (psets+i)->vals;

                    // update parameters
                    for (k=0; k < (psets+i)->n; k++)
                         (*dst++) += c * gasdev();
               }
          }
```

The section above implements the correlation algorithm.

```
          // decay the parameters
          dst = (psets+i)->vals;
          orig = (orig_psets+i)->vals;

          decay = 1.0 - (para->decay);
          for (k=0; k < (psets+i)->n; k++)
               (*dst++) *= decay;
     }

     // do we need to resample the parameters?
     if ( (para->colour != old_colour) ||
          (para->blocks_x != o_blocks_x) ||
          (para->blocks_y != o_blocks_y) ||
          (para->block_width != o_block_width) ||
          (para->block_height != o_block_height) )
     {
          sprintf(tmp,"2\nresampling psets\n");
```

```
            write(serpt,tmp,strlen(tmp));

            if (!resample_parameters())
            {
                sprintf(tmp,"1\nERROR: resampling\n");
                write(serpt,tmp,strlen(tmp));
                return 0;
            }
        }
```

The size of virtual pixels can be changed mid optimisation to increase or decrease the resolution. This region calls the resampling process, if required.

```
        if (!make_img_set())
```

Finally, we generate the next set of stimuli.

```
        {
            sprintf(tmp,"1\nERROR: making new img set\n");
            write(serpt,tmp,strlen(tmp));
            return 0;
        }

        sprintf(tmp,"2\nOkay.\n");
        write(serpt,tmp,strlen(tmp));


        return 1;
}

/* void optimise_end(void)
 * Clean up, and write final values of parameters
 */
void optimise_end(void)
{
        int i;

        write_data(NULL);

        // free the memory used by the parameter sets
        if (psets)
        {
            for (i=0; i<num_psets; i++)
                if ((psets+i)->vals)
                    free((psets+i)->vals);
            free(psets);
        }
        if (orig_psets)
        {
            for (i=0; i<num_psets; i++)
                if ((orig_psets+i)->vals)
                    free((orig_psets+i)->vals);
            free(orig_psets);
        }

        // free the memory used by stimulus descriptions
        if (stims) free(stims);
}

/* int resample_parameters()
   We've updated the parameters, but the user has changed one (or more) of the settings.
   So... we need to recreate the parameters, keeping what we have 'learned' so far.
   The copied original parameters are also resampled, so we have a baseline to decay to.

   This code does not work with the arbirary basis functions since they are not orthonormal
*/
int resample_parameters()
{
        int i;

        if (image_model == kMODEL_BASIS)
            return 1;

        // free the current images
        for (i=0; i<para->total_images; i++)
            free((image_data+i)->pixel);
        free(image_data);

        // allocate memory for a new image for every parameter set
        image_data=(struct stimulus_image *)malloc(sizeof(struct stimulus_image) * num_psets);
        if (!image_data) return 0;

        // create a new image for every parameter set (no noise, obviously)
        for (i=0; i<num_psets; i++)
            if (!pset_to_image(psets+i, 0, image_data+i))
                return 0;
```

```
     para->total_images = num_psets;

     // free the current parameter sets
     for (i=0; i<num_psets; i++)
          if ((psets+i)->vals)
               free((psets+i)->vals);

     // create new parameter sets based on the images
     for (i=0; i<num_psets; i++)
          if (!image_to_pset(image_data+i, psets+i))
               return 0;

     return 1;
}

/* int write_data(float *response_list)
 * Write the parameters and spike counts for each derived image to disk
 * (if response_list is NULL only parameters are written)
 * Called by update_parameters() & optimise_end()
 */
int write_data(float *response_list)
{
     int i,j, ips;
     FILE *fp;
     char fname[32], suffix[32];

     if (++iteration==0)
     {
          time_t t = time(NULL);
          struct tm *tp = localtime(&t);
          strftime(filename, 32, "results_%H%M_%d%m%y_", tp);
     }

     for (i=0; i<num_psets; i++)
     {
          strcpy(fname, filename); // base filename
          sprintf(suffix,"%03d.txt",i+1);     // append parameter number & extension
          strcat(fname, suffix);         // full filename
          fp = fopen(fname,"a");
          if (!fp)
               return 0;

          // write parameter set
          fprintf(fp,"ITERATION #%03d ************************\n",iteration);
          fprintf(fp,"PARAMETER SET #%03d (%02dx%02dx%1d) ***********\n",
               i+1, (psets+i)->blocks_x, (psets+i)->blocks_y, (psets+i)->colour?3:1);
          for (j=0; j<(psets+i)->n; j++)
               fprintf(fp,"%5.3f ",(psets+i)->vals[j]);
          fprintf(fp,"\n");
          if (response_list)
          {
               fprintf(fp,"VARIABLES ******************************\n");
               fprintf(fp,"Learning Rate  = %f\n",para->learning_rate);
               fprintf(fp,"Noise Std.Dev  = %f\n",old_noise_stdev);
               fprintf(fp,"CELL RESPONSE **************************\n");
               fprintf(fp,"  Image    Seed     Spks\n");
               ips = para->total_images / num_psets;

               for (j = 0; j < para->total_images; j++)
                    if (stims[j].pset == i)
                         fprintf(fp,"   #%03d [%010d] %5.1f\n", j, stims[j].seed, ↓
 response_list[j+1]);
          }

          fprintf(fp,"\n");
          fclose(fp);
     }
     return 1;
}

/* int make_img_set()
 * Creates a set of images derived from the parameter sets
 */
int make_img_set()
```

This function creates a set of stimuli by adding noise to the parameter values. Each noisy set of parameters is then transformed into an image. Different functions implement the actual transformation, depending on the image model in use. Images based on noiseless and original versions of the parameters can also be included in the stimulus set, in order to monitor the progress of the procedure.

```
{
     int i,j;
     int pset_images;     // number of images per pset
     struct img_pset *ppset;
```

```
    // free memory occupied by original/last set of images
    for (i=0; i<para->total_images; i++)
    {
        if ((image_data+i)->load)
            { free((image_data+i)->pixel); (image_data+i)->load = 0; }
        if ((image_data+i)->filename)
            free((image_data+i)->filename);
    }
    free(image_data);

    // free memory occupied by last set of stimulus descriptions
    if (stims) free(stims);

    // There must be at least 3 images per parameter set, since we need 2 for the
    // original and noiseless images
    pset_images = para->max_images / num_psets;
    if (pset_images < 3) return 0;

    para->total_images = pset_images * num_psets;

    // allocate image memory
    image_data=(struct stimulus_image *)malloc(sizeof(struct stimulus_image) * para↓
->total_images);
    if (!image_data) return 0;

    // allocate memory for stimulus descriptions
    stims=(struct stim *)malloc(sizeof(struct stim) * para->total_images);
    if (!stims) return 0;

    // generate the stimulus descriptions
    for (i=0; i<para->total_images; i++)
    {
        j = i / pset_images;
        stims[i].orig = 0;
        stims[i].pset = j;
        stims[i].seed = lrand48() + 1;

        if (search_image_set == kNOISY_BASE_ORIGINAL)
        {
            // each first image is the original
            if (i % pset_images == 0) {stims[i].orig = 1; stims[i].seed = 0;}

            // each second is the noiseless
            if (i % pset_images == 1) stims[i].seed = 0;
        }
    }

    // now create each stimulus image
    for (i=0; i<para->total_images; i++)
    {
        sprintf(tmp,"2\nnoisy image %03d\n",i+1);
        write(serpt,tmp,strlen(tmp));

        if (stims[i].orig)
            ppset = orig_psets + stims[i].pset;
        else
            ppset = psets + stims[i].pset;

        if (!pset_to_image(ppset, stims[i].seed, image_data+i))
            return 0;
```

The forward transform is called to convert the parameters plus noise into an image.

```
    }
    return 1;
}


/* int copy_pset(struct img_pset *src, struct img_pset *dst)
 * Copies parameter set src to dst. Allocates memory for the copied values
 */
int copy_pset(struct img_pset *src, struct img_pset *dst)
{
    int i;

    dst->colour = src->colour;
    dst->blocks_x = src->blocks_x;
    dst->blocks_y = src->blocks_y;
    dst->block_width = src->block_width;
    dst->block_height = src->block_height;
    dst->n = src->n;

    dst->vals = (float *)malloc(sizeof(float) * dst->n);
    if (!dst->vals) return 0;

    for (i=0; i<dst->n; i++)
```

```
            dst->vals[i] = src->vals[i];

      return 1;
}


/*****************************************************************************************
 * default_pset() - Initialises the parameters.
 * The parameters are initialised such that the current image model will produce a mean
 * grey image when the parameters are fed through.
 *****************************************************************************************/
int default_pset(struct img_pset *pset)
{
      int i;     // counter
      int num;

      pset->colour = para->colour;
      pset->blocks_x = para->blocks_x;
      pset->blocks_y = para->blocks_y;
      pset->block_width = para->block_width;
      pset->block_height = para->block_height;

      // determine how many values we need
      switch (image_model)
      {
            case kMODEL_IDENTITY:
                  num = (pset->blocks_x) * (pset->blocks_y);
                  break;

            case kMODEL_CORRELATED:
                  num = (pset->blocks_x - 1) * (pset->blocks_y - 1);
                  break;

            case kMODEL_BASIS:
                  num = BASIS_SIZE;
                  break;
      }

      // allocate the memory
      if (pset->colour)
            pset->n = 3 * num;
      else
            pset->n = num;

      pset->vals = (float *)malloc(sizeof(float) * pset->n);
      if (!pset->vals) return 0;

      float *dst = pset->vals;

      // initialise the values to zero
      for (i=0; i < pset->n; i++)
            *dst++ = 0.0;

      return 1;
}
```

If the user has not selected an image to optimise, we instead optimise one (or more) mean luminance images. This function sets initialises the parameters to describe a mean luminance image.

```
/*****************************************************************************************
 * pset_to_image() - Implements the two-step image model.
 * The parameters and noise are added, and the values fed into the chosen image model,
 * resulting in an image ready for display.
 *****************************************************************************************/
int pset_to_image(struct img_pset *pset, long seed, struct stimulus_image *img)
{
      int i, nx, ny;
      float **Y = NULL;
      float **Cb = NULL;
      float **Cr = NULL;

      float *nvals;

      nx = pset->blocks_x;
      ny = pset->blocks_y;

      // add the noise and parameters
      srand48(seed); iset=0;
      nvals = (float *)malloc(sizeof(float) * pset->n);
      for (i=0; i < pset->n; i++)
            if (seed)
                  nvals[i] = pset->vals[i] + (gasdev() * para->noise_stdev);
            else
                  nvals[i] = pset->vals[i];
```

```
    // create matrices to hold the intermediate values
    Y = matrix(1,ny,1,nx);
    if (pset->colour)
    {
        Cb = matrix(1,ny,1,nx);
        Cr = matrix(1,ny,1,nx);
    }

    // pass the values through the first stage of the image model
    switch (image_model)
    {
        case kMODEL_IDENTITY:
            identity_transform_f(pset, nvals, Y, Cb, Cr);
            break;
        case kMODEL_CORRELATED:
            freq_transform_f(pset, nvals, Y, Cb, Cr);
            break;
        case kMODEL_BASIS:
            basis_transform_f(pset, nvals, Y, Cb, Cr);
            break;
    }
```

First, the parameters are transformed to a pixel basis, with a YCbCr colour representation.

```
    // now pass them through the second stage
    if (!matrix_to_image(Y, Cb, Cr, img, nx, ny, pset->block_width, pset->block_height))
        return 0;
```

Then the virtual pixels are converted to real RGB pixels ready for display.

```
    free_matrix(Y,1,ny,1,nx);
    if (pset->colour)
    {
        free_matrix(Cb,1,ny,1,nx);
        free_matrix(Cr,1,ny,1,nx);
    }

    free(nvals);

    return 1;
}
/*****************************************************************************************
 * x_transform_f() These functions implement the first step of the forward image model.
 * The noisy parameters are fed into an image model, producing matrices for the luminance
 * channel and (optionally) two colour channels.
 *****************************************************************************************/
int identity_transform_f(struct img_pset *pset, float *nvals, float **ch_Y, float **ch_Cb, float
 **ch_Cr)
{
    int bx, by;
    float *src = nvals;

    for (by = 0; by < (pset->blocks_y); by++)
        for (bx = 0; bx < (pset->blocks_x); bx++)
        {
            ch_Y[by+1][bx+1] = *src++;

            if (pset->colour)
            {
                ch_Cb[by+1][bx+1] = *src++;
                ch_Cr[by+1][bx+1] = *src++;
            }

        }
    return 1;
}
```

The parameters already represent the image in a pixel basis, so there is little to do here, simply copy the (linear) parameter set into a matrix of the appropriate size. This results in a YCbCr virtual pixel based representation.

```
int basis_transform_f(struct img_pset *pset, float *nvals, float **ch_Y, float **ch_Cb, float
 **ch_Cr)
{
    int num = BASIS_SIZE;

    scaled_basis_transform(nvals, ch_Y, pset->blocks_x, pset->blocks_y);
    if (pset->colour)
    {
        scaled_basis_transform(nvals+num, ch_Cb, pset->blocks_x, pset->blocks_y);
        scaled_basis_transform(nvals+num+num, ch_Cr, pset->blocks_x, pset->blocks_y);
    }
    return 1;
}

void scaled_basis_transform(float *src, float **ch, int width, int height)
{
```

```
    int x,y;
    int i,j,k;
    float w,s;
    float *ptr;

    for (y = 0; y < height; y++)
        for (x = 0; x < width; x++)
            ch[y+1][x+1] = 0.0;

    for (i=0, j=0, k=8, s=1.0; i<BASIS_SIZE; i++, j++)
    {
        ptr = basis[i].f;
        w = s * (*src++);
        if (j==k) {k*=4; s/=0.5; j=0;} // previously 2 - incorrect

        for (y=basis[i].y0; y<=basis[i].y1; y++)
            for (x=basis[i].x0; x<=basis[i].x1; x++)
                ch[height-y][x+1] += w * (*ptr++);
    }

}
```

Here, the parameters represent coefficients of a set of Gabor functions. The coefficients are scaled such that the image noise has a 1/f frequency spectrum. The end result is a YCbCr virtual pixel based image.

```
int freq_transform_f(struct img_pset *pset, float *nvals, float **ch_Y, float **ch_Cb, float
 **ch_Cr)
{
    int num = (pset->blocks_x - 1) * (pset->blocks_y - 1);

    scaled_ift(nvals, ch_Y, pset->blocks_x, pset->blocks_y);
    if (pset->colour)
    {
        scaled_ift(nvals+num, ch_Cb, pset->blocks_x, pset->blocks_y);
        scaled_ift(nvals+num+num, ch_Cr, pset->blocks_x, pset->blocks_y);
    }

    return 1;
}

void scaled_ift(float *src, float **ch, int width, int height)
{
    int x,y;
    float ***data, **speq;
    float fd;

    data = f3tensor(1,1,1,height,1,width);
    speq = matrix(1,1,1,2*height);

    // make sure all the coefficients are zero
    for (y=1; y<=height; y++)
    {
        for (x=1; x<=width; x++)
            data[1][y][x]=0.0;
        speq[1][y*2]=0.0;
        speq[1][y*2-1]=0.0;
    }

    // scale the coefficents by 1/f
    for (y=1; y<=height/2; y++)
    {
        for (x=1; x<=width/2; x++)
        {
            if (x==1 && y==1)    // f00
            {
                data[1][1][1] = (*src++);
                data[1][1][2] = 0.0;
                continue;
            }

            fd = 1.0 / sqrt( (x-1)*(x-1)+(y-1)*(y-1) );

            data[1][y][x*2]   = fd * (*src++);
            data[1][y][x*2-1] = fd * (*src++);

            if (y!=1)
            {
                if (x==1) // set the -f0's to be *f0
                {
                    data[1][height-y+2][1] =  data[1][y][1];
                    data[1][height-y+2][2] = -data[1][y][2];
                }
                else      // independent -f's
                {
                    data[1][height-y+2][x*2-1] = fd * (*src++);
                    data[1][height-y+2][x*2]   = fd * (*src++);
```

```
                    }
                }
            }
        }
        // inverse transform
        rlft3(data,speq,1,height,width,-1);

        // copy to channel
        for (y=1; y<=height; y++)
            for (x=1; x<=width; x++)
                ch[y][x] = data[1][y][x];

        free_matrix(speq,1,1,1,2*height);
        free_f3tensor(data,1,1,1,height,1,width);
}
```

In this case, parameters represent real and imaginary parts of the component frequencies of the image. Again, coefficients are first scaled to produce noise with a 1/f frequency spectrum. The end result is a YCbCr virtual pixel based image.

```
/*****************************************************************************************
 * matrix_to_image() - Implements the second step of the forward image model.
 * a matrix specifying the luminance channel is (optionally) combined with two colour
 * channels and an RGB image of the specified size is created, ready for display.
 *****************************************************************************************/
int matrix_to_image(float **ch_Y, float **ch_Cb, float **ch_Cr, struct stimulus_image *img, int↓
 nx, int ny, int bw, int bh)
```

This function simply converts a YCbCr virtual pixel representation into an RGB real pixels, ready for display.

```
{
    int bx,by;// block counters
    int x,y;  // pixel counters

    int tlx, tly, brx, bry;

    float r,g,b, Y, Cb = 0.0, Cr = 0.0;
    byte pix_r, pix_g, pix_b;

    img->width = nx * bw;
    img->height = ny * bh;
    img->pixel = (unsigned long *)malloc(sizeof(unsigned long) * img->width * img->height);
    if (!img->pixel) return 0;

    // the following need to be set for the external code
    img->load = 1; img->filename = NULL;
    img->cx = para->center_x; img->cy = para->center_y;
    img->pre_t = pre_t; img->dis_t = dis_t; img->aft_t = aft_t;

    for (by=0; by < ny; by++)
        for (bx=0; bx < nx; bx++)
        {
            Y = ch_Y[by+1][bx+1];
            if (ch_Cb) Cb = ch_Cb[by+1][bx+1];
            if (ch_Cr) Cr = ch_Cr[by+1][bx+1];

            // convert to R,G,B
            YCbCr_to_RGB(Y,Cb,Cr,&r,&g,&b);

            // bounds check
            if (r>1.0) r=1.0; if (g>1.0) g=1.0; if (b>1.0) b=1.0;
            if (r<0.0) r=0.0; if (g<0.0) g=0.0; if (b<0.0) b=0.0;

            pix_r = (byte)(r*255.0);
            pix_g = (byte)(g*255.0);
            pix_b = (byte)(b*255.0);

            tlx = bx * bw;
            tly = by * bh;
            brx = tlx + bw;
            bry = tly + bh;

            // write the pixels
            for (y=tly; y<bry; y++)
                for (x=tlx; x<brx; x++)
                    put_pixel_rgb(img, x, y, pix_r, pix_g, pix_b);

        }
    return 1;

}

/*****************************************************************************************
 * image_to_pset() - Implements the two-step inverse image model.
 *****************************************************************************************/
int image_to_pset(struct stimulus_image *img, struct img_pset *pset)
```

This, and the following functions are used at the start of the optimisation process to transform the pixel based image(s) that has been selected for optimisation into parameters (the form of encoding depends on the choice of image model).

```
{
     int nx, ny;
     float **Y = NULL;
     float **Cb = NULL;
     float **Cr = NULL;

     if (!img->pixel || !img->load) return 0;

     pset->colour = para->colour;
     pset->blocks_x = para->blocks_x;
     pset->blocks_y = para->blocks_y;
     pset->block_width = para->block_width;
     pset->block_height = para->block_height;

     nx = pset->blocks_x;
     ny = pset->blocks_y;

     // create matrices to hold the intermediate values
     Y = matrix(1,ny,1,nx);
     if (pset->colour)
     {
          Cb = matrix(1,ny,1,nx);
          Cr = matrix(1,ny,1,nx);
     }

     // pass the image through the first stage of the inverse model
     image_to_matrix(img, Y, Cb, Cr, nx, ny, pset->block_width, pset->block_height);

     // pass the values through the second stage of the inverse model
     switch (image_model)
     {
          case kMODEL_IDENTITY:
               identity_transform_b(Y, Cb, Cr, pset);
               break;
          case kMODEL_CORRELATED:
               freq_transform_b(Y, Cb, Cr, pset);
               break;
          case kMODEL_BASIS:
               // just load the pre-stored coeffs, since
               // it would take too long to calculate them
               load_coefficients(img, pset);
               break;
     }


     free_matrix(Y,1,ny,1,nx);
     if (pset->colour)
     {
          free_matrix(Cb,1,ny,1,nx);
          free_matrix(Cr,1,ny,1,nx);
     }
     return 1;
}
```

`int load_coefficients(struct stimulus_image *img, struct img_pset *pset)`

Because computing the Gabor representation of an image is very time consuming, we simply load a set of coefficients that have been pre-computed.

```
{
     int i;
     char filename[256];
     FILE *fp;

     // open the coefficient file
     strncpy(filename, img->filename, strlen(img->filename)-4);
     filename[strlen(img->filename)-4]='\0';
     strcat(filename,".txt\0");

     fp = fopen(filename,"r");
     if (!fp)
     {
          printf("coefficient filename = %s\n",filename);
          sprintf(tmp,"1\nNo coefficient file!\n");
          write(serpt,tmp,strlen(tmp));
          return 0;
     }

     // allocate memory for the parameters
     pset->n = BASIS_SIZE;
     if (pset->colour) pset->n *= 3;
```

```
    pset->vals = (float *)malloc(sizeof(float) * pset->n);
    if (!pset->vals) return 0;

    // then load them
    float *dst = pset->vals;
    for (i=0; i<pset->n; i++)
        fscanf(fp,"%f\n",dst++);
    fclose(fp);

    // now scale the parameters
    int j,k;
    float s;

    dst = pset->vals;
    for (i=0, j=0, k=8, s=1.0; i<BASIS_SIZE; i++, j++)
    {
        *(dst) *= s;
        if (pset->colour)
        {
            *(dst+BASIS_SIZE) *= s;
            *(dst+BASIS_SIZE*2) *= s;
        }
        dst++;
        if (j==k) {k*=4; s*=0.5; j=0;} // previously 2 - incorrect
    }
}

/*****************************************************************************************
 * x_transform_b() These functions implement the second step of the inverse image model.
 *****************************************************************************************/
int identity_transform_b(float **ch_Y, float **ch_Cb, float **ch_Cr, struct img_pset *pset)
{
    int bx, by;

    // work out the number of parameters
    pset->n = (pset->blocks_x) * (pset->blocks_y);
    if (pset->colour) pset->n *= 3;

    // allocate memory for them
    pset->vals = (float *)malloc(sizeof(float) * pset->n);
    if (!pset->vals) return 0;

    float *dst = pset->vals;
    for (by = 0; by < (pset->blocks_y); by++)
        for (bx = 0; bx < (pset->blocks_x); bx++)
        {
            *dst++ = ch_Y[by+1][bx+1];
            if (pset->colour)
            {
                *dst++ = ch_Cb[by+1][bx+1];
                *dst++ = ch_Cr[by+1][bx+1];
            }
        }

    return 1;
}
```

The first step already produced YCbCr virtual pixels. Just need to copy them into the parameter structure.

```
int freq_transform_b(float **ch_Y, float **ch_Cb, float **ch_Cr, struct img_pset *pset)
{
    int num;

    // work out the number of parameters
    num = pset->n = (pset->blocks_x - 1) * (pset->blocks_y - 1);
    if (pset->colour) pset->n *= 3;

    // allocate memory for them
    pset->vals = (float *)malloc(sizeof(float) * pset->n);
    if (!pset->vals) return 0;

    float *dst = pset->vals;

    scaled_fft(ch_Y, dst, pset->blocks_x, pset->blocks_y);
    if (pset->colour)
    {
        scaled_fft(ch_Cb, dst+num, pset->blocks_x, pset->blocks_y);
        scaled_fft(ch_Cr, dst+num+num, pset->blocks_x, pset->blocks_y);
    }
    return 1;
}

void scaled_fft(float **ch, float *dst, int width, int height)
{
    int x,y;
    float ***data, **speq;
```

```
    float fd;

    data = f3tensor(1,1,1,height,1,width);
    speq = matrix(1,1,1,2*height);

    // copy channel to matrix
    for (y=1; y<=height; y++)
        for (x=1; x<=width; x++)
            data[1][y][x] = ch[y][x];

    // forward transform
    rlft3(data,speq,1,height,width,1);

    // scale the coefficents by f
    // (& further scale by width*height/2 since I don't do this in inverse transform)
    float scale = width*height/2.0;

    for (y=1; y<=height/2; y++)
    {
        for (x=1; x<=width/2; x++)
        {
            if (x==1 && y==1)    // f00
            {
                (*dst++) = data[1][1][1] / scale;
                continue;
            }

            fd = sqrt( (x-1)*(x-1)+(y-1)*(y-1) ) / scale;

            (*dst++) = fd * data[1][y][x*2];
            (*dst++) = fd * data[1][y][x*2-1];

            if (y!=1)
            {
                if (x==1) // set the -f0's to be *f0
                {
                    data[1][height-y+2][1] =  data[1][y][1];
                    data[1][height-y+2][2] = -data[1][y][2];
                }
                else      // independent -f's
                {
                    (*dst++) = fd * data[1][height-y+2][x*2-1];
                    (*dst++) = fd * data[1][height-y+2][x*2];
                }
            }
        }
    }

    free_matrix(speq,1,1,1,2*height);
    free_f3tensor(data,1,1,1,height,1,width);
}
```

When transforming the parameters into an image, the coefficients are scaled such that the image noise has a 1/f spectrum. This must be compensated for during the initial encoding step above, such that the underlying initial image will remain the same.

```
/*****************************************************************************************
 * image_to_matrix() - Implements the first step of the inverse image model.
 * an RGB image in memory is resampled at the resolution specified and converted into
 * seperate luminance (and optional) colour channels.
 *****************************************************************************************/
int image_to_matrix(struct stimulus_image *img, float **ch_Y, float **ch_Cb, float **ch_Cr, int↓
 nx, int ny, int bw, int bh)
{
    int bx,by;// block counters
    int x,y;  // pixel counters

    int zero_x, zero_y, tlx, tly, brx, bry;

    float r,g,b, Y,Cb,Cr;
    long tot_r,tot_g,tot_b;
    byte pix_r, pix_g, pix_b, def_r, def_g, def_b;

    if (!img->pixel || !img->load) return 0;

    // establish a mapping between the original and new coordinate systems
    zero_x = (img->width - nx*bw) / 2;
    zero_y = (img->height - ny*bh) / 2;

    // get the default background colour from (0,0)
    get_pixel_rgb(img, 0, 0, &def_r, &def_g, &def_b);

    for (by=0; by < ny; by++)
        for (bx=0; bx < nx; bx++)
        {
            // calculate the average R,G,B for
```

```
                // this block in the original image

                tlx = zero_x + (bx * bw);
                tly = zero_y + (by * bh);
                brx = tlx + bw;
                bry = tly + bh;

                tot_r=0.0; tot_g=0.0; tot_b=0.0;
                for (y=tly; y<bry; y++)
                     for (x=tlx; x<brx; x++)
                     {
                          if (x >= 0 && y >= 0 && x < img->width && y < img->height)
                               get_pixel_rgb(img, x, y, &pix_r, &pix_g, &pix_b);
                          else
                               {pix_r = def_r; pix_g = def_g; pix_b = def_b;}

                          tot_r += pix_r;
                          tot_g += pix_g;
                          tot_b += pix_b;
                     }

                // get R,G,B in [0,1]
                r = (float)tot_r / (bw * bh * 255);
                g = (float)tot_g / (bw * bh * 255);
                b = (float)tot_b / (bw * bh * 255);

                // convert to Y,Cb,Cr
                RGB_to_YCbCr(r,g,b,&Y,&Cb,&Cr);

                // and store
                ch_Y[by+1][bx+1] = Y;
                if (ch_Cb) ch_Cb[by+1][bx+1] = Cb;
                if (ch_Cr) ch_Cr[by+1][bx+1] = Cr;
          }
     return 1;
}
```

This code resamples the selected image according to the resolution settings chosen by the user. The result is a matrix based YCbCr virtual pixel representation, ready for transformation using the selected image model.

```
/*****************************************************************************************
 * helper functions
 *****************************************************************************************/

void get_pixel_rgb(struct stimulus_image *img, int x, int y, byte *r, byte *g, byte *b)
{
     long index = x + (y * img->width);
     unsigned long tmp = *(img->pixel + index);

     // colour layout in pixmap is xBGR
     *r = (tmp&0xFF);
     *g = (tmp>>8)&0xFF;
     *b = (tmp>>16)&0xFF;
}

void put_pixel_rgb(struct stimulus_image *img, int x, int y, byte r, byte g, byte b)
{
     long index = x + (y * img->width);
     unsigned long tmp = ((long)r) | ((long)g<<8) | ((long)b<<16);

     *(img->pixel + index) = tmp;
}

void RGB_to_YCbCr(float r, float g, float b, float *Y, float *Cb, float *Cr)
{
     *Y  = ( 0.299 * r) + ( 0.587 * g) + ( 0.114 * b);
     *Cb = (-0.169 * r) + (-0.331 * g) + ( 0.500 * b);
     *Cr = ( 0.500 * r) + (-0.419 * g) + (-0.081 * b);

     *Y -= 0.5;
}

void YCbCr_to_RGB(float Y, float Cb, float Cr, float *r, float *g, float *b)
{
     Y += 0.5;

     *r = ( 1.0 * Y) + ( 0.0000 * Cb) + ( 1.4021 * Cr);
     *g = ( 1.0 * Y) + (-0.3441 * Cb) + (-0.7142 * Cr);
     *b = ( 1.0 * Y) + ( 1.7718 * Cb) + ( 0.0000 * Cr);
}

float gasdev(void)
{
     /* returns a normally distributed deviate, with zero mean
        and unit variance. From Numerical Recipes in C 2nd ed. */
```

```
    //static int iset=0;
    static float gset;
    float fac,rsq,v1,v2;

    if (iset==0) {
        do {
            v1=2.0*drand48()-1.0;
            v2=2.0*drand48()-1.0;
            rsq=v1*v1+v2*v2;
        } while (rsq >= 1.0 || rsq == 0.0);
        fac=sqrt(-2.0*log(rsq)/rsq);
        gset=v1*fac;
        iset=1;
        return v2*fac;
    } else {
        iset=0;
        return gset;
    }
}
```

12.2. gabor.h

```
// basis function structure
typedef struct {
    int x0,y0; // top left
    int x1,y1; // bottom right
    float *f;
} basis_fn;

#define BASIS_SIZE      2729

void create_basis(int size);
```

12.3. gabor.c

```
#include "gabor.h"
#include <stdio.h>
#include <stdlib.h>
#include <math.h>

#define PI3.14159265

extern basis_fn basis[BASIS_SIZE];

float gabor(float x, float y,
    float x0, float y0, float w, float f, float o, float p, float size)
{
    float k = 2 * 3.14159265 / size;
    x = x - x0;
    y = y - y0;

    return exp(-(x*x+y*y)/(2*w*w)) * cos(k*f*(x*cos(o)+y*sin(o))+p);
}

// create a five level gabor wavelet pyramid basis for an image
// of dimensions (size x size) pixels
void create_basis(int size)
{
    int i = 1;

    int x, y, sx, fc, f, o, p;
    float cx, cy, w;
    float *ptr;

    f = 1;
    sx = size;

    // create the DC component

    basis[0].x0 = 0; basis[0].y0 = 0;
    basis[0].x1 = size-1; basis[0].y1 = size-1;
```

```
    basis[0].f = (float *)malloc(sizeof(float) * size * size);
    ptr = basis[0].f;
    for (y=0; y<=size-1; y++)
        for (x=0; x<=size-1; x++)
            *ptr++ = 1;

    // now the gabors

    for (fc = 1; fc <= 5; fc++)
    {
        w = sx / sqrt(2.0*PI);
        for (o=0; o<=3; o++)
            for (cx = (sx-1.0)/2.0; cx<size ; cx+=sx )
                for (cy = (sx-1.0)/2.0; cy<size ; cy+=sx )
                    for (p=0; p<=1; p++)
                    {
                        basis[i].x0 = (cx-w*3.0)<0 ? 0:(cx-w*3.0);
                        basis[i].y0 = (cy-w*3.0)<0 ? 0:(cy-w*3.0);
                        basis[i].x1 = (cx+w*3.0)>(size-1) ? size-1:(cx+w*3.0);
                        basis[i].y1 = (cy+w*3.0)>(size-1) ? size-1:(cy+w*3.0);
                        basis[i].f = (float *)malloc(sizeof(float) *
                            (1+basis[i].x1-basis[i].x0) * (1+basis[i].y1-basis[i].y0));
                        ptr = basis[i].f;
                        for (y=basis[i].y0; y<=basis[i].y1; y++)
                            for (x=basis[i].x0; x<=basis[i].x1; x++)
                                *ptr++ = gabor( x,y,cx,cy,w,f,o*(PI/4.0),p*(PI/2.0),size );

                        i++;
                    }
        f*=2.0;
        sx/=2.0;
    }
}
```

12.4. fft.h

```
// FFT routine and other utilities from Numerical Recipes in C, 2nd Ed

float **matrix(long nrl, long nrh, long ncl, long nch);
float ***f3tensor(long nrl, long nrh, long ncl, long nch, long ndl, long ndh);
void free_matrix(float **m, long nrl, long nrh, long ncl, long nch);
void free_f3tensor(float ***t, long nrl, long nrh, long ncl, long nch, long ndl, long ndh);

void rlft3(float ***data, float **speq, unsigned long nn1, unsigned long nn2, unsigned long nn3,
 int isign);
void fourn(float data[], unsigned long nn[], int ndim, int isign);
```

12.5. fft.c

```
// FFT routine and other utilities from Numerical Recipes in C, 2nd Ed
#include <math.h>
#include <stdio.h>
#include <stdlib.h>

#define NR_END 1
#define FREE_ARG char*
#define SWAP(a,b) tempr=(a);(a)=(b);(b)=tempr

void nrerror(char error_text[])
{
    fprintf(stderr,"Numerical Recipes run-time error...\n");
    fprintf(stderr,"%s\n",error_text);
    fprintf(stderr,"...now exiting to system...\n");
    exit(1);
}

float **matrix(long nrl, long nrh, long ncl, long nch)
/* allocate a float matrix with subscript range m[nrl..nrh][ncl..nch] */
{
    long i, nrow=nrh-nrl+1,ncol=nch-ncl+1;
    float **m;

    /* allocate pointers to rows */
    m=(float **) malloc((size_t)((nrow+NR_END)*sizeof(float*)));
    if (!m) nrerror("allocation failure 1 in matrix()");
```

```
	m += NR_END;
	m -= nrl;

	/* allocate rows and set pointers to them */
	m[nrl]=(float *) malloc((size_t)((nrow*ncol+NR_END)*sizeof(float)));
	if (!m[nrl]) nrerror("allocation failure 2 in matrix()");
	m[nrl] += NR_END;
	m[nrl] -= ncl;

	for(i=nrl+1;i<=nrh;i++) m[i]=m[i-1]+ncol;

	/* return pointer to array of pointers to rows */
	return m;
}

float ***f3tensor(long nrl, long nrh, long ncl, long nch, long ndl, long ndh)
/* allocate a float 3tensor with range t[nrl..nrh][ncl..nch][ndl..ndh] */
{
	long i,j,nrow=nrh-nrl+1,ncol=nch-ncl+1,ndep=ndh-ndl+1;
	float ***t;

	/* allocate pointers to pointers to rows */
	t=(float ***) malloc((size_t)((nrow+NR_END)*sizeof(float**)));
	if (!t) nrerror("allocation failure 1 in f3tensor()");
	t += NR_END;
	t -= nrl;

	/* allocate pointers to rows and set pointers to them */
	t[nrl]=(float **) malloc((size_t)((nrow*ncol+NR_END)*sizeof(float*)));
	if (!t[nrl]) nrerror("allocation failure 2 in f3tensor()");
	t[nrl] += NR_END;
	t[nrl] -= ncl;

	/* allocate rows and set pointers to them */
	t[nrl][ncl]=(float *) malloc((size_t)((nrow*ncol*ndep+NR_END)*sizeof(float)));
	if (!t[nrl][ncl]) nrerror("allocation failure 3 in f3tensor()");
	t[nrl][ncl] += NR_END;
	t[nrl][ncl] -= ndl;

	for(j=ncl+1;j<=nch;j++) t[nrl][j]=t[nrl][j-1]+ndep;
	for(i=nrl+1;i<=nrh;i++) {
		t[i]=t[i-1]+ncol;
		t[i][ncl]=t[i-1][ncl]+ncol*ndep;
		for(j=ncl+1;j<=nch;j++) t[i][j]=t[i][j-1]+ndep;
	}

	/* return pointer to array of pointers to rows */
	return t;
}

void free_matrix(float **m, long nrl, long nrh, long ncl, long nch)
/* free a float matrix allocated by matrix() */
{
	free((FREE_ARG) (m[nrl]+ncl-NR_END));
	free((FREE_ARG) (m+nrl-NR_END));
}

void free_f3tensor(float ***t, long nrl, long nrh, long ncl, long nch, long ndl, long ndh)
/* free a float f3tensor allocated by f3tensor() */
{
	free((FREE_ARG) (t[nrl][ncl]+ndl-NR_END));
	free((FREE_ARG) (t[nrl]+ncl-NR_END));
	free((FREE_ARG) (t+nrl-NR_END));
}

// FFT routines follow...

void rlft3(float ***data, float **speq, unsigned long nn1, unsigned long nn2, unsigned long nn3,↓
 int isign)
{
	void fourn(float data[], unsigned long nn[], int ndim, int isign);
	void nrerror(char error_text[]);
	unsigned long i1,i2,i3,j1,j2,j3,nn[4],ii3;
	double theta,wi,wpi,wpr,wr,wtemp;
	float c1,c2,h1r,h1i,h2r,h2i;

	if (1+&data[nn1][nn2][nn3]-&data[1][1][1] != nn1*nn2*nn3)
		nrerror("rlft3: problem with dimensions or contiguity of data array\n");
	c1=0.5;
	c2 = -0.5*isign;
	theta=isign*(6.28318530717959/nn3);
	wtemp=sin(0.5*theta);
	wpr = -2.0*wtemp*wtemp;
	wpi=sin(theta);
	nn[1]=nn1;
```

```
    nn[2]=nn2;
    nn[3]=nn3 >> 1;
    if (isign == 1) {
            fourn(&data[1][1][1]-1,nn,3,isign);
        for (i1=1;i1<=nn1;i1++)
            for (i2=1,j2=0;i2<=nn2;i2++) {
                speq[i1][++j2]=data[i1][i2][1];
                speq[i1][++j2]=data[i1][i2][2];
            }
    }
    for (i1=1;i1<=nn1;i1++) {
        j1=(i1 != 1 ? nn1-i1+2 : 1);
        wr=1.0;
        wi=0.0;
        for (ii3=1,i3=1;i3<=(nn3>>2)+1;i3++,ii3+=2) {
            for (i2=1;i2<=nn2;i2++) {
                if (i3 == 1) {
                    j2=(i2 != 1 ? ((nn2-i2)<<1)+3 : 1);
                    h1r=c1*(data[i1][i2][1]+speq[j1][j2]);
                    h1i=c1*(data[i1][i2][2]-speq[j1][j2+1]);
                    h2i=c2*(data[i1][i2][1]-speq[j1][j2]);
                    h2r= -c2*(data[i1][i2][2]+speq[j1][j2+1]);
                    data[i1][i2][1]=h1r+h2r;
                    data[i1][i2][2]=h1i+h2i;
                    speq[j1][j2]=h1r-h2r;
                    speq[j1][j2+1]=h2i-h1i;
                } else {
                    j2=(i2 != 1 ? nn2-i2+2 : 1);
                    j3=nn3+3-(i3<<1);
                    h1r=c1*(data[i1][i2][ii3]+data[j1][j2][j3]);
                    h1i=c1*(data[i1][i2][ii3+1]-data[j1][j2][j3+1]);
                    h2i=c2*(data[i1][i2][ii3]-data[j1][j2][j3]);
                    h2r= -c2*(data[i1][i2][ii3+1]+data[j1][j2][j3+1]);
                    data[i1][i2][ii3]=h1r+wr*h2r-wi*h2i;
                    data[i1][i2][ii3+1]=h1i+wr*h2i+wi*h2r;
                    data[j1][j2][j3]=h1r-wr*h2r+wi*h2i;
                    data[j1][j2][j3+1]= -h1i+wr*h2i+wi*h2r;
                }
            }
            wr=(wtemp=wr)*wpr-wi*wpi+wr;
            wi=wi*wpr+wtemp*wpi+wi;
        }
    }
    if (isign == -1)
        fourn(&data[1][1][1]-1,nn,3,isign);
}

void fourn(float data[], unsigned long nn[], int ndim, int isign)
{
    int idim;
    unsigned long i1,i2,i3,i2rev,i3rev,ip1,ip2,ip3,ifp1,ifp2;
    unsigned long ibit,k1,k2,n,nprev,nrem,ntot;
    float tempi,tempr;
    double theta,wi,wpi,wpr,wr,wtemp;

    for (ntot=1,idim=1;idim<=ndim;idim++)
        ntot *= nn[idim];
    nprev=1;
    for (idim=ndim;idim>=1;idim--) {
        n=nn[idim];
        nrem=ntot/(n*nprev);
        ip1=nprev << 1;
        ip2=ip1*n;
        ip3=ip2*nrem;
        i2rev=1;
        for (i2=1;i2<=ip2;i2+=ip1) {
            if (i2 < i2rev) {
                for (i1=i2;i1<=i2+ip1-2;i1+=2) {
                    for (i3=i1;i3<=ip3;i3+=ip2) {
                        i3rev=i2rev+i3-i2;
                        SWAP(data[i3],data[i3rev]);
                        SWAP(data[i3+1],data[i3rev+1]);
                    }
                }
            }
            ibit=ip2 >> 1;
            while (ibit >= ip1 && i2rev > ibit) {
                i2rev -= ibit;
                ibit >>= 1;
            }
            i2rev += ibit;
        }
        ifp1=ip1;
        while (ifp1 < ip2) {
            ifp2=ifp1 << 1;
```

```
                theta=isign*6.28318530717959/(ifp2/ip1);
                wtemp=sin(0.5*theta);
                wpr = -2.0*wtemp*wtemp;
                wpi=sin(theta);
                wr=1.0;
                wi=0.0;
                for (i3=1;i3<=ifp1;i3+=ip1) {
                     for (i1=i3;i1<=i3+ip1-2;i1+=2) {
                          for (i2=i1;i2<=ip3;i2+=ifp2) {
                               k1=i2;
                               k2=k1+ifp1;
                               tempr=(float)wr*data[k2]-(float)wi*data[k2+1];
                               tempi=(float)wr*data[k2+1]+(float)wi*data[k2];
                               data[k2]=data[k1]-tempr;
                               data[k2+1]=data[k1+1]-tempi;
                               data[k1] += tempr;
                               data[k1+1] += tempi;
                          }
                     }
                     wr=(wtemp=wr)*wpr-wi*wpi+wr;
                     wi=wi*wpr+wtemp*wpi+wi;
                }
                ifp1=ifp2;
           }
           nprev *= n;
      }
}
```

13. References

Ashbridge, E., Perrett, D. I., Oram, M.W. & Jellema, T. (2000). Effect of image orientation and size on object recognition: responses of single units in the macaque monkey temporal cortex. *Cognitive Neuropsychology* 17: 13-34.

Atick, J. J. & Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Computation* 4: 196-210.

Atick, J. J. & Redlich, A. N. (1992). What does the retina know about natural scenes?. *Neural Computation* 4: 449-572.

Barlow, H. B. (1961). The coding of sensory messages. *Current Problems in Animal Behaviour*. Cambridge: Cambridge University Press.

Baylis, G. C. & Rolls, E. T. (1987). Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. *Experimental Brain Research* 65: 614-622.

Baylis, G. C., Rolls, E. T. & Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex of the superior temporal sulcus of the monkey. *Brain Research* 342: 91-102.

Baylis, G. C., Rolls, E. T. & Leonard, C. M. (1987). Functional subdivisions of the temporal lobe neocortex. *Journal of Neuroscience* 7: 330-342.

Bhaskaran, V. & Konstantinides, K. (1997). *Image & video compression standards: Algorithms and architectures (2nd Ed)*. Kluwer Academic Publishers.

Biederman, I. & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology* 20: 38-64.

Booth, M. C. A & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex* 8: 510-523.

Boussaoud, D., Desimone, R. & Ungerleider, L.G. (1991). Visual topography of area TEO in the macaque. *J. Comp. Neurol.* 306: 554-575.

Breitmeyer, B. G. (1975). Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision Research* 15: 1411-1412.

Brodmann, K. (1905). Beitrage zur histologischen lokalisation der grosshirnrinde. dritte mitteilung: die rindenfelder der niederen affen. *J. Psychol. Neurol.* 4: 177-226.

Buckley, M. J., Gaffan, D. & Murray, E. A. (1997). Functional double dissociation between two inferior temporal cortical areas: perirhinal cortex versus middle temporal gyrus. *Journal of Neurophysiology* 77: 587-598.

Carpenter, G. A. & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphic and Image Processing* 37: 54-115.

Chun, M. M. & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology, Human Perception and Performance* 21: 109-127.

Crook, J. M., Lange-Malecki, B., Lee, B. & Valberg, A. (1988). Visual resolution of

macaque retinal ganglion cells. *Journal of Physiology* 396: 205-224.

Davidoff, J. B. & Ostergaard, A. L. (1988). The role of colour in categorical judgements. *The Quarterly Journal of Experimental Psychology* 40A: 533-544.

Dean, A. F. (1981). The variability of discharge of simple cells in the cat striate cortex. *Experimental Brain Research* 44: 437-440.

Dean, P. (1976). Effects of inferotemporal lesions on the behaviour of monkeys. *Psychological Bulletin* 83: 41-71.

Dean, P. (1979). Visual cortex ablation and thresholds for successively presented stimuli in rhesus monkeys: II. Hue. *Experimental Brain Research* 35: 69-83.

DeAngelis, G. C., Ohzawa, I. & Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neurosciences* 18: 451-458.

Delorme, A., Richard, G. & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Research* 40: 2187-2200.

Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience* 4: 2051-2062.

Ferrera, V. P., Nealey, T. A. & Maunsell, J. H. R. (1994). Responses in macaque visual area V4 following inactivation of the parvocellular and magnocellular LGN

pathways. *Journal of Neuroscience* 14: 2080-2088.

Field, D. J. (1994). What is the goal of sensory coding. *Neural Computation* 6: 559-601.

Field, D.J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America. A* 4: 2379-2394.

Fiorentini, A., Maffei, L. & Sandini, G. (1984). Role of high spatial frequencies in face perception. *Perception* 12: 195-201.

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation* 3: 194-200.

Földiák, P. (1997). Learning constancies for object perception. In W. Walsh, J. J. Kullikowski (Eds). *Visual Constancies: Why Things Look As They Do*. Cambridge University Press.

Földiák, P. (2001). Stimulus optimisation in primary visual cortex. *Neurocomputing* 38-40: 1217-1222.

Fujita, I., Tanaka, K., Ito, M. & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360: 343-346.

Gross, C. G. (1973). Visual functions of inferotemporal cortex. In R. Jung (Ed). *Handbook of Sensory Physiology* (Vol 7-3B, pp. 451-482). Berlin: Springer-Verlag.

Gross, C. G., Bender, D. B. & Rocha-Miranda, C. E. (1969). Visual receptive fields of

neurons in inferotemporal cortex of the monkey. *Science* 166: 1303-1306.

Gross, C. G., Desimone, R., Albright, T. D. & Schwartz, E. L. (1985). Inferior temporal cortex and pattern recognition. In C. Chagas, R. Gattass & C. G. Gross (Eds). *Pattern Recognition Mechanisms.* (pp. 179-201). Berlin: Springer-Verlag.

Gross, C. G., Rocha-Miranda, C. E. & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology* 35: 96-111.

Harth, E. & Tzanakou, E. (1974). ALOPEX: A stochastic method for determining visual receptive fields. *Vision Research* 14: 1475-1482.

Hasselmo, M. E., Rolls, E. T. & Baylis, G. C. (1989a). The role of expression and identity in the face-selective response of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research* 32: 203-218.

Hasselmo, M. E., Rolls, E. T., Baylis, G. C. & Nalwa, V. (1989b). Object-centred encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. *Experimental Brain Research* 75: 417-429.

Heywood, C. A., Gaffan, D. & Cowey, A. (1995). Cerebral achromatopsia in monkeys. *European Journal of Neuroscience* 7: 1064-1073.

Higuchi, S-I. & Miyashita, Y. (1996). Formation of mnemonic neuronal responses to visual paired-associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *Proceedings of the National Academy of Sciences USA* 93: 739-

743.

Horel, J. A. (1994). Retrieval of color and form during suppression of temporal cortex with cold. *Behavioural Brain Research* 65: 165-172.

Hubel, D. H., Wiesel, T. N. & LeVay, S. (1977). Plasticity of ocular dominance columns in monkey striate cortex. *Philosophical Transactions of the Royal Society of London: Series B* 278: 377-409.

Humprey, G. K., Goodale, M. A., Jakobson, L. S. & Servos, P. (1994). The role of surface information in object recognition: studies of a visual form agnosic and normal subjects. *Perception* 23: 1457-1481.

Iwai, E. & Mishkin, M. (1969). Further evidence on the locus of the visual area in the temporal lobe of the monkey. *Experimental Neurology* 25: 585-594.

Jacobson, L. D., Gaska, J. P., Chen, H-W., Pollen, D. A. (1993). Structural testing of multi-input linear-nonlinear cascade models for cells in macaque striate cortex. *Vision Research* 33: 609-626.

Jeeves, M. A., Milner, A. D., Perrett, D. I. & Smith, P. A. J. (1983). Visual cells responsive to direction of movement and stimulus form in the anterior superior temporal sulcus of the macaque monkey. *Journal of Physiology* 341: 80P.

Jones, J. P. & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology* 58: 1187-1211.

Keysers, C. (2000). *The Speed of Sight: Neural Correlates of Perceptual Reports in*

RSVP. Ph.D. Thesis. University of St.Andrews.

Keysers, C., Xiao, D.-K., Földiák, P. & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience* 13: 90-101.

Kobatake, E. & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology* 71: 856-867.

Kobatake, E., Wang, G. & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology* 80: 324-330.

Komatsu, H. & Ideura, Y. (1993). Relationships between colour, shape, and pattern selectivities of neurons in the inferior temporal cortex of the monkey. *Journal of Neurophysiology* 70: 677-694.

Komatsu, H., Ideura, Y., Kaji, S. & Yamane, S. (1992). Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. *Journal of Neuroscience* 12: 408-424.

Kovacs, G., Vogels, R. & Orban, G. (1995). Cortical correlate of pattern backward masking. *Proceedings of the National Academy of Sciences USA* 92: 5587-5591.

Lee, K. J. & Perrett, D. I. (1997). Presentation-time measures of the effects of manipulations in colour space on discrimination of famous faces. *Perception* 26: 733-

752.

Lehky, S. R., Sejnowski, T. J. & Desimone, R. (1992). Predicting responses of nonlinear neurons in monkey striate cortex to complex patterns. *Journal of Neuroscience* 12: 3568-3581.

Li, L., Miller, E. K. & Desimone, R. (1993). The representation of stimulus familiarity in the anterior inferior temporal cortex. *Journal of Neurophysiology* 69: 1918-1929.

Linden, A. & Kinderman, J. (1990). Inversion of multilayer nets. In M. Caudill (Ed). *Proceedings of the International Joint Conference on Neural Networks, Washington, DC* (pp. 425-430). Hillsdale, NJ: Erlbaum.

Livingstone, M. S. & Hubel, D. H. (1988). Segregation of form, color, movement and depth: anatomy, physiology, and perception. *Science* 240: 740-749.

Logothetis, N. K., Pauls, J. & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology* 5: 552-563.

Marmarelis, P. Z. & Marmarelis, V. Z. (1978). *Analysis of Physiological Systems*. Plenum.

Marr, D. (1982). *Vision*. New York. W. H. Freeman.

Marrocco, R. T. (1976). Sustained and transient cells in monkey lateral geniculate nucleus: Conduction velocities and response properties. *Journal of Neurophysiology* 39: 340-353.

Maunsell, J. H. R., Ghose, G. M., Assad, J. A., McAdams, C. J., Boudreau, C. E. & Noerager, B. D. (1999). Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. *Visual Neuroscience* 16: 1-14.

McSorley, E. & Findlay, J. M. (1999). An examination of a temporal anisotropy in the visual integration of spatial frequencies. *Perception* 28: 1031-1050.

Merigan, W. H. & Maunsell, J. H. R. (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience* 16: 369-402.

Merigan, W. H., Byrne, C. & Maunsell, J. H. R. (1991a). Does primate motion perception depend on the magnocellular pathway? *Journal of Neuroscience* 11: 3422-3429.

Merigan, W. H., Katz, L. M. & Maunsell, J. H. R. (1991b). The effects of parvocellular lateral geniculate lesions on the acuity and contrast sensitivity of macaque monkeys. *Journal of Neuroscience* 11: 994-1101.

Messinger, A., Squire, L. R., Zola, S. M. & Albright, T. D. (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proceedings of the National Academy of Sciences USA* 98: 12239-12244.

Miller, E. K. & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science* 263: 520-522.

Miller, E. K., Li, L. & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science* 254: 1377-1379.

Milner, A. D. & Goodale, M. A. (1985). *The Visual Brain in Action*. Oxford.

Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transactions of the Royal Society of London: Series B* 298: 85-95.

Movshon, J. A., Thompson, I.D. & Tolhurst, D. J. (1978a). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology* 283: 53-77.

Movshon, J. A., Thompson, I.D. & Tolhurst, D. J. (1978b). Receptive field organization of complex cells in the cat's striate cortex. *Journal of Physiology* 283: 79-99.

Murray, E. A., Gaffan, D. & Mishkin, M. (1993). Neural Substrates of visual stimulus stimulus association in rhesus-monkeys. *Journal of Neuroscience* 13: 4549-4561.

Naya, Y., Yoshida, M. & Miyashita, Y. (2001). Backward spreading of memory-retrieval signal in the primate temporal cortex. *Science* 291: 661-664.

Nelken, I., Prut, Y., Vaadia, E. & Abeles, M. (1994). In search of the best stimulus: An optimization procedure for finding efficient stimuli in the cat auditory cortex. *Hearing Research* 72: 237-253.

Nowak, L. G. & Bullier, J. (1997). The timing of information transfer in the visual system. In J. Kaas, K. Rockland & A. Peters, *Extrastriate Cortex in Primates* (pp. 205-241). New York: Plenum.

Nowak, L. G., Munk, M. H. J., Girard, P. & Bullier, J. (1995). Visual latencies in

areas V1 and V2 of the macaque monkey. *Visual Neuroscience* 12: 371-384.

Olshausen, B. A. & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381: 607-609.

Optican, L. M. & Richmond, B. J. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. information theoretic analysis. *Journal of Neurophysiology* 57: 162-178.

Oram, M.W. & Perrett, D.I. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology* 68: 70-84.

Ostergaard, A. L. & Davidoff, J. B. (1985). Some effects of colour on naming and recognition of objects. *Journal of Experimental Psychology: Learning, Memory & Cognition* 11: 579-587.

Parker, D. M. & Dutch, S. (1987). Perceptual latency and spatial frequency. *Vision Research* 27: 1279-1283.

Parker, D. M. & Salzen, E. A. (1977a). Latency changes in the human visual evoked response to sinusoidal gratings. *Vision Research* 17: 1201-1204.

Parker, D. M. & Salzen, E. A. (1977b). The spatial selectivity of early and late waves within the human visual evoked response. *Perception* 6: 85-95.

Parker, D. M. & Salzen, E. A. (1982). Evoked potentials and reaction times to the offset and contrast reversal of sinusoidal gratings. *Vision Research* 22: 205-207.

Parker, D. M. (1980). Simple reaction times to the onset, offset and contrast reversal of sinusoidal gratings stimuli. *Perceptual Psychophysics* 28: 365-368.

Parker, D. M., Lishman, J. R. & Hughes, J. (1997). Evidence for the view that temporospatial integration in vision is temporally anisotropic. *Perception* 26: 1169-1180.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception* 21: 147-160.

Perrett, D. I., Hietanen, J. K., Oram, M. W. & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London: Series B* 335: 23-30.

Perrett, D. I., Oram, M. W. & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition* 67: 111-145.

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J. & Thomas, S. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Experimental Brain Research* 86: 159-173.

Perrett, D. I., Rolls, E. T. & Caan, W. (1982). Visual neurons responsive to faces in monkey temporal cortex. *Experimental Brain Research* 47: 329-342.

Perrett, D. I., Smith, P. A. J., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., Broennimann, R., Milner, A. D. & Jeeves, M. A. (1985). Visual analysis of body

movements by neurons in the temporal cortex of the macaque monkey: preliminary report. *Behavioural Brain Research* 16: 153-170.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D. & Jeeves, M. A. (1984). Neurons responsive to faces in the temporal cortex: Studies of functional organisation, sensitivity to identity and relation to perception. *Human Neurobiology* 3: 197-208.

Potter, M. C. & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology* 81: 10-15.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1988). *Numerical Recipes in C: The Art of Scientific Computing (2nd ed.)*. Cambridge University Press.

Price, C. J. & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. *The Quarterly Journal of Experimental Psychology* 41A: 797-828.

Pulfich, C. (1922). Die stereoskopie im dienste der isochromen und heterochromen photometrie. *Naturwissenschaften* 10: 533-564, 569-574, 596-601, 714-722, 735-743, 751-761.

Richmond, B. J. & Optican, L. M. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. quantification of response waveform. *Journal of Neurophysiology* 57: 147-161.

Richmond, B. J., Optican, L. M., Podell, M. & Spitzer, H. (1987). Temporal encoding

of two-dimensional patterns by single units in primate inferior temporal cortex. I. response characteristics. *Journal of Neurophysiology* 57: 132-146.

Richmond, B. J., Wurtz, R. H. & Sato, T. (1983). Visual response of inferior temporal neurons in awake rhesus monkey. *Journal of Neurophysiology* 50: 1415-1432.

Ringach, D. L., Sapiro, G. & Shapley, R. (1987). A subspace reverse-correlation technique for the study of visual neurons. *Vision Research* 37: 2455-2464.

Rolls, E. T. & Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Experimental Brain Research* 65: 38-48.

Rolls, E. T. & Tovée, M. J. (1994). Processing speed in the cerebral cortex, and the neurophysiology of backward pattern masking. *Proceedings of the Royal Society of London: Series B* 257: 9-15.

Rolls, E. T. (1991). Neural organisation of higher visual functions. *Current Opinion in Neurobiology* 1: 274-278.

Rolls, E. T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philosophical Transactions of the Royal Society of London: Series B* 335: 11-21.

Rolls, E. T., Baylis, G. C. & Leonard, C. M. (1985). Role of low and high spatial frequencies in the face-selective responses of neurons in the superior temporal sulcus of the monkey. *Vision Research* 25: 1021-1035.

Rolls, E. T., Baylis, G. C., Hasselmo, M. E. & Nalwa, V. (1989). The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. *Experimental Brain Research* 76: 153-164.

Rolls, E. T., Tovée, M. J. & Panzeri, S. (1999). The neurophysiology of backward visual masking: information analysis. *Journal of Cognitive Neuroscience* 11: 300-311.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds). *Parallel Distributed Processing* (Vol 1, pp. 318-362). Cambridge, MA: MIT Press.

Sakai, H. M. (1992). White-noise analysis in neurophysiology. *Physiological Reviews* 72: 491-505.

Sakai, K. & Miyashita, Y. (1991). Neural organization for the long-term memory of paired-associates. *Nature* 354: 152-155.

Schwartz, O. & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience* 4: 819-825.

Schyns, P.G. & Oliva, A. (1994). From blobs to boundary edges: evidence for time and scale dependent scene recognition. *Psychological Science* 5: 195-200.

Seltzer, B. & Pandya, D. N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Research* 149: 1-24.

Shapley, R. & Lennie, P. (1985). Spatial frequency analysis in the visual system.

Annual Review of Neuroscience 8: 547-583.

Sigala, N. & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415: 318-320.

Sugase, Y., Yamane, S., Ueno, S. & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400: 869-873.

Takechi, H., Onoe, H., Shizuno, H., Yoshikawa, E., Sadato, N., Tsukada, H. & Watanabe, Y. (1997). Mapping of cortical areas involved in colour vision in non-human primates. *Neuroscience Letters* 230: 17-20.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience* 19: 109-139.

Tanaka, K., Saito, H.-A., Fukada, Y. & Moriya, M. (1991). Coding visual images of objects in inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology* 66: 170-189.

Tolhurst, D. J., Movshon, J. A. & Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research* 41: 414-419.

Tovée, M. J. & Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. *Visual Cognition* 2: 35-58.

Tovée, M. J., Rolls, E. T., Treves, A. & Bellis, R. P. (1993). Information encoding

and response of single neurons in the primate temporal visual cortex. *Journal of Neurophysiology* 70: 640-654.

Tzanakou, E., Michalak, R. & Harth, E. (1979). The Alopex Process: Visual receptive fields by response feedback. *Biological Cybernetics* 35: 1-14.

Ullman, S. (1995). Sequence seeking and counter streams – A computational model for bi-directional information flow in the visual-cortex. *Cerebral Cortex* 5: 1-11.

Ungerleider, L. G. & Mishkin, M. (1982). Two cortical visual systems. In *Analysis of Visual Behavior* (ed. D. J. Ingle, M. A. Goodale & R. J. W. Mansfield). pp549-586. Cambridge, MA: MIT Press.

Vogels, R. & Orban, G. (1996). Coding of stimulus invariances by inferior temporal neurons. *Progress in Brain Research* 112: 195-211.

Vogels, R. & Orban, G. A. (1991). Quantitative study of striate single unit responses in monkey performing an orientation task. *Experimental Brain Research* 84: 1-11.

von Bonin, G. & Bailey, P. (1947). *The Neocortex of Maccaca Mulatta.* Urbana, IL: University of Illinois.

Watson, A. B. & Ahumada, A. J. (1985). A model of human visual motion sensing. *Journal of the Optical Society of America. A* 2: 322-342.

Widrow, B. (1985). *Adaptive Signal Processing*. Prentice Hall.

Wurm, L. H., Legge, G. E., Isenberg, L. M. & Luebker, A. (1993). Color improves object recognition in normal and low vision. Journal of Experimental Psychology:

Human Perception and Performance 19: 899-911.

Zeki, S. & Marini, L. (1998). Three cortical stages of colour processing in the human brain. *Brain* 121: 1669-1685.