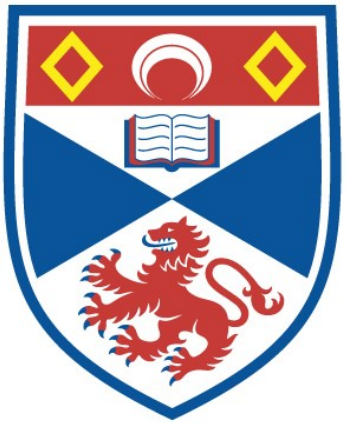
CODING OF OBJECT PARTS, VIEW, ORIENTATION AND SIZE IN THE TEMPORAL CORTEX OF THE MACAQUE

Elisabeth Wachsmuth

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



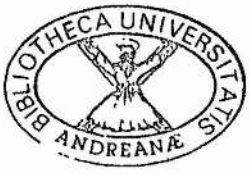
1996

CODING OF OBJECT PARTS, VIEW, ORIENTATION AND SIZE IN THE TEMPORAL CORTEX OF THE MACAQUE.

by:

Elisabeth Wachsmuth

June, 1995



A thesis submitted to the University of St. Andrews for the Degree of Doctor of Philosophy in the School of Psychology.

DECLARATION FOR THE DEGREE OF Ph.D.

I, Elisabeth Wachsmuth, hereby certify that this thesis, which is approximately 52'000 words in length, has been written by me, that is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Signed: Date: 16/6/95.

I was admitted as a research student in October 1992 and as a candidate for the degree of Doctor of Philosophy in October 1992; the higher study for which this is a record was carried out in the University if St. Andrews between 1992 and 1995.

Signed: Date: 16/6/95.

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulation appropriate to 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.

Signature of Supervisor: Date: 16/6/95

THESIS COPYRIGHT DECLARATION

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 the abstract will be published, and that a copy of the work my be made and supplied to any *bona fide* library or research worker.

Signed: Date:

ACKNOWLEDGEMENTS

I would like to thank my supervisor, David Perrett, for all the support he has given me during these three years, and for his helpful comments and contributions in producing this thesis. His unlimited enthusiasm (especially in the lab), encouragement and enormous knowledge of the different issues of visual perception have guided me throughout this three year period and prepared me to move on in my courier and life. Also, I would like to thank Mike Oram for all his help and patience with teaching and introducing me to the methods of electrophysiological recordings and statistical analysis of the data obtained. Thanks to Lesley Harrison and Nathan Emery for sharing constant battles with the equipment and helping to collect the data. I am also grateful to Hilda Dickie, Wendy Taylor, Jo Hutchison, Mary Latimer, all the workshop technical staff and Sean Earnshaw for helping with the monkeys, the equipment and the photography respectively.

A big thank you goes to Laura Allen and the White House inhabitants for sharing the experiences of working towards a PhD and to Allison Robertson and Dave Carey for their patience in proof reading my thesis. Furthermore, I am very grateful to my parents and family for their support and encouragement. Finally, the biggest and a very special thank you to Jonathan in supporting me in every respect throughout this period.

Abstract

The study examined the importance of (1) component parts, (2) view, (3) orientation and (4) size in the neural encoding of the sight of a complex object in the temporal cortex of the macaque. Studies focused on cells selectively responsive to the sight of the head/body but unresponsive to control stimuli. **(1)** Cells responsive to the static whole body were tested with two component parts of the body. 44% (29/66) of cells responded to the whole body and to *one* of the two body regions tested: 23 to the head; 6 to the body. 36% (24/66) responded independently to *both* regions of the body when tested in isolation. The remaining cells were selective for the *entire* body and unresponsive to component parts. Similar selectivity for component parts was observed amongst cells responsive to moving heads/bodies (18 cells tested). **(2)** 90% (66/73) of cells (selectively responsive to static or moving head/bodies) tested were sensitive to perspective view (viewer-centred). Comparable levels of view sensitivity were found for responses to the whole body and its parts. Contrary to some influential models of object recognition these results indicate view-specific processing for both the appearance of separate object components and for integration of information across components. **(3)** The majority of cells tested (18/25, 72%) were selectively responsive to a particular orientation in the picture plane of the static whole body stimulus. 7 cells generalised across all orientations (4 cell with pure generalisation; 3 cells with superimposed orientation tuning). Of all cells sensitive to orientation, the majority (15/21, 71%) were tuned to the upright image. **(4)** The majority of cells tested (81%, 13/16) were selective for a particular stimulus size. The remaining cells (3/16) showed generalisation across a 4 fold decrease in size from life-sized. Interestingly, all size sensitive cells were tuned to life-sized stimuli. These results do not support previous suggestions that cells responsive to the head and body are selective to the view but generalise across orientation and size. Here, extensive selectivity for size and orientation is reported. It is suggested that object part, view, orientation and size specific responses might be pooled to obtain generalising responses. Experience appears to affect neuronal coding in two ways: a) Cells become selective for multiple object components due to spatial and temporal association between parts; and b) more cells become tuned to views, orientations and image sizes commonly experienced.

LIST OF CONTENTS

Abstract 5

Chapter I 11

GENERAL INTRODUCTION 11
Problem of object recognition 11
What method is appropriate for the aim of this investigation? 12
Thesis outline 13

Chapter II 14

ANATOMICAL PATHWAYS FOR OBJECT RECOGNITION 14
The segregation of visual information in early visual cortex 15
The dorsal and ventral visual processing pathways 17
Ventral pathway 18
Receptive field sizes 19
Anterior STS 20
Summary of visual processing stages for object recognition 21

Chapter III 23

OVERVIEW OF FUNDAMENTAL OBJECT RECOGNITION MODELS 23
Psychological object recognition models 23
Template matching 24
Feature Analysis 25
Structural description 28
The role of components in object recognition 28
Hierarchical organisation: Bottom-up vs Top-down processing 35
Physiological evidence for the role of component parts in object recognition 37
View discrimination: object-centred vs viewer-centred 42
a) Object-centred representations 42
b) Viewer-centred representation 44
Single view representation 45
Multiple view representations 45
Physiological evidence: object-centred vs viewer-centred 45

Chapter IV 49

GENERAL EXPERIMENTAL METHODS: SINGLE UNIT RECORDING IN THE MACAQUE 49
Pre-surgical training and fixation task 49
Implant construction 51
Surgery 52
Recording techniques 53
Data collection 54

Data analysis 55
Recording sites 56
Perfusion and Histology 58

Chapter V 60

THE ROLE OF COMPONENT PARTS IN OBJECT RECOGNITION 60
INTRODUCTION 60
Cellular sensitivity to objects and their parts 61
METHODS 63
Static visual stimuli 63
Testing methods 64
Testing visual stimuli in motion 64
Single cell data analysis 65
Population data analysis 65
RESULTS 66
Coding of parts 67
a) Cells only responsive to the head 67
b) Cells only responsive to the body (torso and limbs) 68
Coding the entire body 68
Cells only responsive to whole body 68
Cells responsive to multiple body parts 68
Population estimates of time course responses 69
Cell population coding one component part (Head alone) 69
Cell population coding multiple parts 70
Cell population coding Whole Body alone 70
Overall population estimates of time course responses 71
Coding of parts in motion 71
Cells only responsive to one component part in motion 72
Coding the entire body in motion 72
Cells responsive to multiple body parts in motion 72
Cells only responsive to whole body in motion 72
Histological Localisation 73
DISCUSSION 73
Cell sensitivity to the body 73
Implications for models 74
a) Cells selective for one body part 74
b) Cells selective for multiple body parts 75
c) Cells selective for the whole body 76
Coding body parts in motion 76
Coding of isolated object parts 77
Learning the association between parts 77
Population estimates of time course responses 79

Chapter VI 81

VIEW SPECIFICITY AND COMPONENT PARTS 81
INTRODUCTION 81

Object-centred representation 81
Viewer-centred representation 82
Cellular sensitivity to object view 83
METHODS 84
Visual static stimuli 84
Visual stimuli in motion 84
Data analysis 85
View discrimination index 85
RESULTS 85
Cell responses to static heads/bodies 86
Cells with viewer-centred properties 86
Cells coding only one static body part 86
Coding the entire body 87
Cells responsive to multiple static body parts 87
Cells only responsive to the static whole body 87
Cells with object-centred properties 87
Cell responses to heads/bodies in motion 88
Motion sensitive cells with viewer-centred properties 88
Coding of single body parts in motion 88
Coding the entire body in motion 88
Cells responsive to multiple body parts in motion 88
Cells responsive to only the whole body in motion 89
Motion sensitive cells with object-centred properties 89
View discrimination indices 89
a) View discrimination: Static whole body vs static head alone 89
b) View discrimination: Static whole body vs body alone 90
c) View discrimination: Whole body in motion vs effective body part alone in motion 90
Population estimates of time course responses 90
Histological Localisation 91
DISCUSSION 91
Object-centred coding 91
Viewer-Centred coding 92
View discrimination occurs for both the whole body and its effective parts 92
Associations and learning configurations 93
Population estimates of time course responses 93

Chapter VII 94

EFFECT OF ROTATION ON OBJECT RECOGNITION 94
INTRODUCTION 94
Behavioural studies in Humans 94
Short term memory 94
Simultaneous matching of shapes 94
Successive matching of shapes 95
Long term memory 95
Identification and naming tasks 95
Practice effects 96

Models explaining rotation effects 97
Behavioural studies in other species: monkeys, apes and pigeons 99
Generalising across orientations 99
Inversion effects 100
Neuropsychological studies in humans 104
Ventral lesions: Visual agnosia 106
Dorsal lesions: Optic ataxia 106
Neuropsychological studies in monkeys 106
Ventral lesions 106
STS lesions 111
Single cell studies 112
Orientation specificity 112
Orientation generalisation 113
METHOD 114
Visual stimuli 115
Testing methods 115
Data analysis 116
RESULTS 117
Generalising across orientation 118
Generalising across all orientations with some tuning 118
Selectively responsive to a particular orientation 118
Distribution of orientation tuning 118
Orientation Discrimination Index 119
Population estimates of time course of responses 120
Histological reconstruction 120
DISCUSSION 121
The more anterior the more generalisation 122
Build to object-centred representations from viewer-centred representations 124
Familiarity and experience 124
Response latencies at the population and single cell level 126
Advantage of orientation specific coding 128
Orientation coding in dorsal and ventral stream 128

Chapter VIII 130

EFFECT OF SIZE CHANGE ON OBJECT RECOGNITION 130
INTRODUCTION 130
Behavioural studies in humans 130
Models explaining size transformations 131
Exceptions to the rule 132
Neuropsychological studies in humans 134
Neurological studies in monkeys 135
Single cell studies 137
Size specificity and generality 137
Effects of size change on memory for stimulus shape 140
Evoked potential studies in humans 141
METHODS 142
Testing methods 142

Visual stimuli 143
Data analysis 143
RESULTS 144
Size Discrimination Index 145
Population estimates of time course of responses 145
Histological reconstruction 146
DISCUSSION 146
What is size generalisation and specificity important for? 146
What exactly is generalisation and specificity? 147
Previous and empirical findings 147
Familiarity and Experience 150
Coding different image sizes 150
Population response pattern to different stimulus sizes 151

Chapter IX 152

GENERAL DISCUSSION 152
Cells which code information by inhibition 152
What kind of information do single cells code? 153
Little object-centred coding at more anterior parts of the processing pathway 155
An outline of a object processing scheme 155

References 158

Appendix 178

Chapter I

GENERAL INTRODUCTION

"We who investigate the workings of the human eye occupy a quite special place in the body of science, for our concern is both with the objective and subjective. The eye is the window in that wall which separates the body from the mind and we may look both out and in. Never in history has the view been more exciting than it is today when new techniques on every side invite us to explore gardens that for centuries have been locked to all but speculations."

(W. A. H. Rushton, Prentice Lecture, 1963)

Problem of object recognition

Visual object recognition is a fundamental part of our everyday activity. An object can be recognised despite variations in the object's orientation, distance and size, part occlusions, perspective view and lighting conditions. These different viewing conditions induce profound changes in the shape of the image on the retina. So how can the brain cope with this constantly changing retinal information, achieving perceptual *invariance* and *constancy*[1] ? Furthermore, the word 'recognition' implies that the viewed object has been encountered and registered before. This suggests that there must be some sort of representation for the encoding of that object held, by some means, by the neuronal mass of the brain. The neuronal activity responsible for the representation of an object may reflect long-term memory. So that when an object is viewed, the incoming (retinal image) information selectively triggers a particular neuron, or set of neurons (see later), within high levels of the visual processing pathway. The firing pattern and the output of this neuronal representation is relatively unique to the viewed object, allowing recognition to occur. How are these representations formed and what is their configuration? Once a representation of an object is formed, what (if any) transformational processes of the incoming information have to occur to match these representations?

[1] The term 'constancy' refers to the capacity to recognise the physical attributes of an object independent of viewing conditions. For example, size constancy refers to recognition of size-independent of viewing distance. This ability differs from the capacity to recognise an object's identity or class independent of its size, orientation, or view (e.g. a person standing upright and another person standing on his head can both be recognised as humans). This latter capacity is referred to here as 'invariance' or 'generalisation'.

The aim of this thesis is to add to the understanding of how the brain codes an object (or objects of the same class). The following questions will be addressed: What are the underlying mechanisms for the brain to achieve recognition that is *independent* of object part occlusion, retinal view, orientation and size? What is the role of object parts in representing the entire object? And how are *specific* retinal object views, orientations, and sizes coded?

What method is appropriate for the aim of this investigation?

Damage or disconnection to the temporal cortex of man or monkey results in defects in object processing. It has, therefore, been suggested that there are similarities between how the two species process and code visual information about objects. Hence, great emphasis has been placed on the study of the monkey brain in attempting to understand how complex perceptual processes such as object recognition occur in both human and non-human primates.

Different methods of studying the human and monkey brain all contribute to the understanding of visual information processing. Such methods include the study of brain damaged patients who often show similar behavioural effects as animals with brain lesions (where the lesion is in a homologous area to that of the human brain); behavioural studies in both neurological intact humans and monkeys; human brain mapping studies such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI); recording event related potentials (ERP) in both humans and monkeys; and single unit recording in vivo (applied here) carried out predominantly in monkeys (though limited recordings have also been performed in human subjects; Fukamachi et al., 1973; Kelly et al., 1987; Yamashiro et al., 1989). Different methods of investigation provide us with different information about the workings of the brain. The studies of lesioned subjects provide us with knowledge about the involvement of a particular brain area in carrying out behavioural tasks, which often leads to suggestions on the function of the disabled brain area. Lesion studies of the temporal cortex implicate the underlying areas involved in object recognition but do not reveal any detail of how information is processed. Other studies of the brain (PET, fMRI, ERPs etc.) reflect the overall neuronal activity (or oxygen consumption by brain areas) while a controlled task is carried out by the subject. This task can, in some cases be repeated (fMRI and ERP) and though temporal resolution of

these methods is high, spatial resolution is relatively low. The methods mentioned above would fail to indicate neuronal populations selectively coding different information, if the populations were anatomically intermixed. Furthermore, if a brain area is shown to be significantly more active than other areas during a PET scan, it is difficult to determine whether such neurons are activated because of the visual stimuli or the experimental task involved. Single cell studies, on the other hand, provide us with information at the neuronal level rather than measuring the neuronal response of a population. Hence, if two or more populations of neurons perform different functions in the same anatomical area, then selectivity can be observed at the single cell level. Thus, the individuality or similarity between the selectivity of different neurons can be established. The study reported here examined the mechanisms underlying object recognition. This was achieved by examining neuronal responses of single cells when the subject is viewing an object in a particular circumstance (transformation).

Thesis outline

To provide a context for interpreting physiological findings, the anatomical pathways involved in object recognition are outlined first (chapter II). In chapter III, general studies of object recognition relevant to *all* empirical studies (though with strong implications on the role of component parts and view coding) are described and discussed. Thereafter, general experimental methods used to investigate the selectivity of visual cells in the anterior part of the superior temporal sulcus (one area of the temporal cortex) involved in the coding of the visual appearance of human (or monkey) heads/bodies are described (chapter IV). Finally, a series of experimental studies are presented, each carried out in respect to the questions addressed above: Coding of object parts, view, orientation and size; chapters V, IV, VII and VIII, respectively. The findings will be analysed, described and discussed in detail. It is pointed out, that chapter VII on orientation and chapter VIII on size coding have more detailed and extensive introductions, referring to experiments and models which were not discussed in chapter III, though relevant to the empirical studies under examination. The final chapter (IX) will comment on the overall findings from empirical studies and will debate the function and role of object parts, view, orientation and size in constructing an all encompassing object description which is able to cope with different types of image transformations.

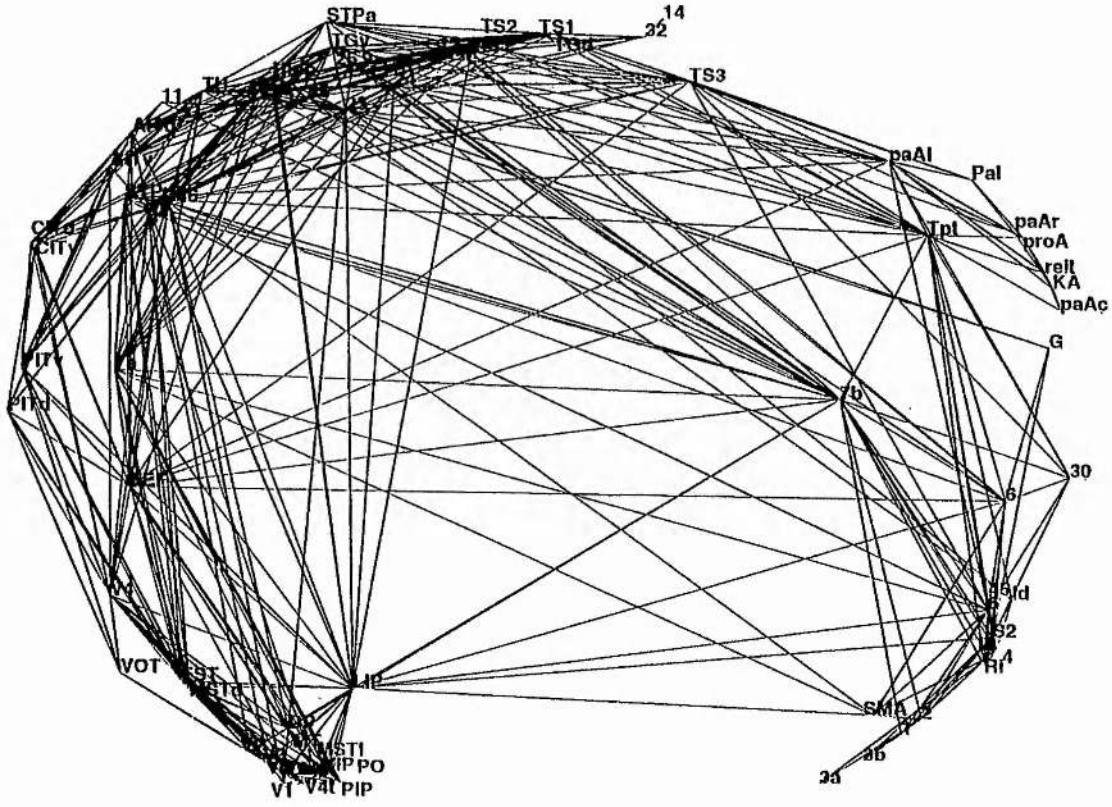
Chapter II

ANATOMICAL PATHWAYS FOR OBJECT RECOGNITION

This chapter will provide an brief overview of the anatomical structures and pathways involved in visual object recognition. It is to provide the reader with sufficient anatomical background to aid understanding and discussions on possible integration of information from different visual cortical areas discussed in subsequent chapters.

It has been suggested by Felleman and Van Essen (1991) that over half of the primate neocortex surface area is involved in visual information processing. There is, however, no single uniform area, rather, several separate areas which are distributed over the whole cortex (Pandya and Yeterian, 1985; Zeki and Shipp, 1988; Felleman and Van Essen, 1991; Young, 1992; 1993). Several parallel processing pathways which are highly specialised are therefore suggested. Thirty-two separate cortical areas for processing visual information have been identified by Fellman and van Essen in 1991, of which 25 areas deal solely with visual information, whilst the remaining 7 areas additionally receive input from other sensory areas hence are referred to as visual association areas. Felleman and Van Essen argue that there are at least 305 connections between these 32 areas, i.e. approximately 10 projections from and to each cortical area. These visual areas are organised and categorised into ten hierarchically different levels (Felleman and Van Essen, 1991). More recent studies on the topological organisation of the different cortical areas which apply a more objective analysis, argue the presence of two relatively independent hierarchically organised systems (the dorsal and ventral stream, see below) which reconverge in the principal sulcus (area 46) and the superior polysensory areas (Ungerleider and Mishkin, 1982; Young, 1992; 1993). Young (1992) included 73 different cortical areas in his study and suggested that there are 758 connections between them. 18% of these connections are found to be one-way projections. It is argued that this connectivity pattern is far from arbitrary, since it only represents 15% of the possible connections between these areas (Fig. 2.1). Not all cortical areas, therefore, are interconnected (Felleman and Van Essen, 1991; Young, 1992; 1993) and two main visual processing streams have been argued on

Figure 2.1. Connections of cortical areas: The topological organisation of the entire macaque cortical processing system. A total of 758 connections between the 73 areas are represented, of which 136 (18%) are one-way (reproduced from Young, 1993).



anatomical, neurophysiological and neuropsychological grounds (Newcombe and Russell, 1969; Ungerleider and Mishkin, 1982; Mishkin et al., 1983; Maunsell and Newsome, 1987; Young, 1992).

The segregation of visual information in early visual cortex

When an image is projected onto the retina at the back of the eye, the visual information is carried from the retina to the lateral geniculus nucleus (LGN); a complex structure consisting of six layers. Visual information of motion, form and colour are believed to be segregated by the early visual system into parallel anatomical pathways (extensively reviewed elsewhere; Livingstone and Hubel, 1984; 1987; Zeki, 1990a, b). In the LGN, cells of the upper four layers forming parts of the parvocellular (P) system, tend to be wavelength selective. Cells in the lower two layers, on the other hand, are part of the magnocellular (M) system and are sensitive to contrast, dynamic form (form in motion) and motion of the visual stimulus.

The striate cortex (also called V1 or the primary visual cortex, corresponding to the human cortical area 17) receives inputs from the retina via the LGN and optic radiations in a point to point manner. The M layers project mainly to layer 4Cα of V1. Cells in this cortical layer give rise to two new kinds of signal, direction and orientation selectivity of line stimuli (Hubel and Wiesel, 1962; 1968). The P layers project to layers 4Cβ of V1, where the cells also detect two new kinds of signals. Cells in these layers can be wavelength or orientation selective. Furthermore, the cortical colour system is obtained from the P input to the striate cortex and the cortical motion system from the M input. Information about stimulus orientation, however, results from both the M and P input.

The cortical form system is divided into two parallel pathways (Zeki, 1990a, b). One system (V1 interblob - V2 interstripe - V4) is concerned with static form, intimately linked to colour and hence derived from the P-system. The other system (V1 layer 4B - V2 thick stripe - V3) is concerned with dynamic form (independent of colour) derived from the M-system (Livingstone and Hubel, 1984; Shipp and Zeki, 1989; Zeki, 1989). In the visual processing pathway, it can be expected to find a close relationship between form and colour, since every form has a colour and every colour, enclosed in space, has a form.

The prestriate visual areas, which are specialised to process different kinds of information such as form, colour and motion, are divided into area V2 surrounding V1 (also called secondary visual cortex, corresponding to the human area 18), the V3 complex (corresponding to human area 19), the V4 complex, and MT (also called V5) and its satellite areas (Albright, 1984, 1992; Albright, Desimone, et al., 1984; Zeki, 1976; 1990a, b). However, these areas are not connected to each other in a single hierarchical chain. Sub-regions of the striate cortex (V1) send separate outputs to each of these visual areas.

As mentioned above, neuronal signals carrying specific information about the visual scene are already distinct before they reach the specialised areas of the prestriate visual cortex. Layers 2 and 3 of V1 contain regions commonly referred to as 'blobs' revealed by cytocrome oxidase staining. Approximately half of the cells found in the blobs are wavelength selective but not orientation selective, whereas cells found outside the blobs are orientation selective but not wavelength sensitive. This suggests that the pathways for form and colour are segregated within and may be distributed separately by V1. Such segregation of information occurs across the depth of the cortex in individual layers (Zeki, 1990a, b; 1991).

V1 projects information either directly to the different prestriate cortical divisions (V2, V3, V4, MT, PO and PIP), or via an indirect pathway passing through V2 (see e.g. Kuypers et al., 1965; Cragg, 1969; Van Essen and Zeki, 1978; Zeki, 1978a, b; Rockland and Pandya, 1981; Livingstone and Hubel, 1984; Shipp and Zeki, 1985; Shipp and Zeki, 1989; Zeki, 1989; 1990). Studies have shown that V2 is arranged in sets of alternative thick and thin stripes defined by staining characteristics, separated by very thin interstripes (Tootell et al., 1983; Livingstone and Hubel, 1987; Zeki, 1989). Cells found in V2 thin stripes receive input from cells located in the blobs of V1 and project to V4 (De Yoe and Van Essen, 1985; Shipp and Zeki, 1985; Zeki, 1989). These V4 cells are wavelength or colour selective. If the V4 cells are truley colour selective, they show colour constancy, where the cell response is selective to the colour of the image independent of the ambient light's wavelength composition (Zeki, 1990). Other V4 cells receive input from V2 interstripe cells. These V4 cells are selectively responsive to stimulus orientation but do not show any wavelength selectivity (Desimone and Schein, 1987). In summary, some V4 cells are selectively

responsive to the form and orientation of the stimulus independent of colour. Other cells are selectively responsive to a particular form in a particular colour (Zeki, 1976; Van Essen and Zeki, 1978; Zeki, 1978a, b; Desimone, Schein et al., 1985; Desimone and Schein, 1987; Kobatake and Tanaka, 1993; 1994). Nonetheless, some V4 cells, whether sensitive to stimulus orientation or not, show sensitivity to wavelength or colour.

Projections from V4 cells are directed mainly to the temporal lobe and the cortex of the ventral superior temporal sulcus (STS). V4 cells also project to the parietal cortex (see e.g. Felleman and Van Essen, 1991; 1992; Young and Yamane, 1992).

The dorsal and ventral visual processing pathways

Lesion studies of various areas in the brain have enlightened us a great deal about the anatomical arrangements of the visual processing pathways. Striate cortex damage results in a blind region in the visual field depending on which part of the striate cortex is damaged. Damage to anatomical areas which occur later in the visual pathway, for example the temporal or parietal lobe, can produce quite different results (Newcombe and Russell, 1969). Temporal cortex damage results in agnosia (Milner, 1958; Kimura, 1963; Landsdell, 1968; Benson et al., 1974; Meadows, 1974), posterior damage results in motion processing impairments (Zihl et al., 1983) and damage to the parietal cortex can cause visual spatial impairments (McFie et al., 1950; Ratcliff and Newcombe, 1973).

Even though the complex pathway through which visual information flows in humans is still unclear, it has been known for a long time that there are two major fibre bundles leaving the occipital cortex (Flechsig, 1896; 1920). In the monkey, these bundles project to two different areas in the brain: (a) The superior longitudinal fasciculus projects to the posterior parietal area and then to the frontal lobe: V1→ V2 / MT → VIP / (MST → FST / LIP) (Fig. 2.2). This pathway based on the M input (Maunsell, 1992), has been named the "where" (Mishkin, 1972, 1983; Pohl, 1973), or the dorsal pathway (De Yoe and Van Essen, 1988), since it is involved in spatial perception of where an object is in the environment. More recent studies have suggested that this pathway may be more appropriately called the "how" pathway,

since it seems to be involved in visumotor control (Goodale, 1993; Milner and Goodale, 1993). The dorsal pathway, however, does not only project to the parietal areas, but also projects from area MST/FST via the STS to the temporal cortex [posterior and anterior superior polysensory area (STPp and STPa)] (see Fig. 2.2). (b) The ventral pathway projects from V4 via the inferior longitudinal fasciculus to the temporal lobe [posterior, central and anterior IT (PIT, CIT, AIT) to STPa]. This second pathway, is suggested to be responsible for the recognition of objects and hence has been named the "what" pathway (see below).

The temporal lobe is divided by the superior temporal sulcus (STS) which runs along its length into the superior temporal gyrus (primary auditory cortex including area KA) (dorsally) and the inferior temporal gyrus (ventrally).

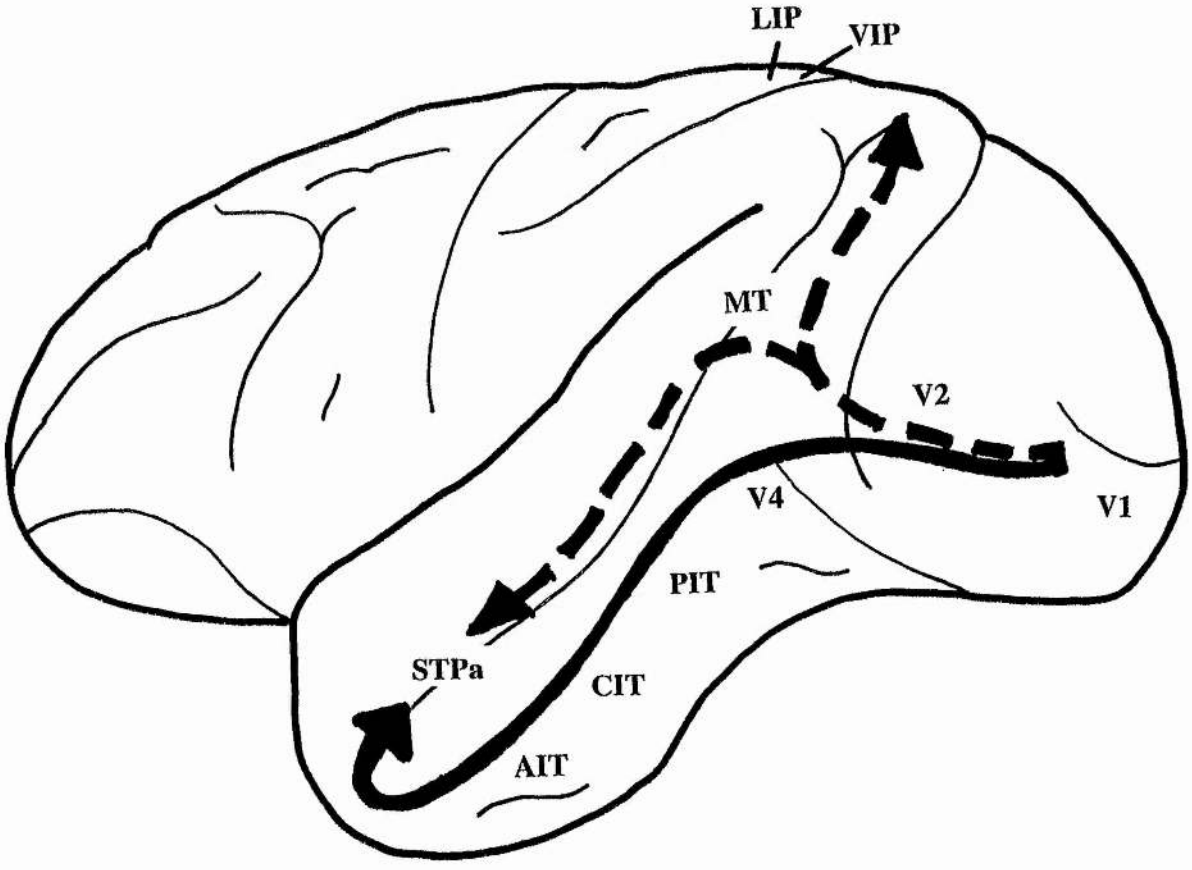
This thesis is mainly concerned with the ventral pathway of object recognition. However, it should be noted that the two pathways are not completely independent and anatomically distinct, rather there are many interconnections between these two main streams of visual information processing. Direct correspondence between the two subcortical areas and their pathways are suggested to occur (though see e.g. Livingstone and Hubel, 1987; Maunsell and Newsome, 1987).

Ventral pathway

Effects of bilateral removal of the inferior temporal cortex (IT) in the monkey have been reviewed by Wilson (1957); Cowey and Weiskrantz (1967); Gross (1973); Dean and Weiskrantz (1974); Dean (1976). These IT lesions resulted in severe impairment in visual discrimination, especially discriminating 2D- and 3D-shapes and colours.

The IT stretches from V4 in the prestriate visual cortex to the anterior temporal pole, lying ventrally to the superior temporal sulcus (STS). Different cytoarchitectonic areas which compose IT are PIT, CIT and AIT (also called TE1, TE2, TE3 and TEO; Seltzer and Pandya, 1978). Cells found in these areas are predominantly selectively responsive to visual stimuli (see e.g. Dean, 1976; Tanaka et al., 1991; 1992; 1993; Kobatake and Tanaka, 1994). The posterior part of the inferior temporal cortex, known as PIT or area TEO, is involved in the ability to discriminate between objects, whereas

Figure 2.2. A schematic representation of the left view of the macaque monkey brain illustrating the **two visual pathways**. The dorsal ("where") pathway represented by the dashed line and the ventral ("what") pathway indicated by the solid line.



the anterior part, area AIT or also called TE, would appear to be involved in object representation and visual memory (Iwai and Mishkin, 1968; Cowey and Gross, 1970; Iwai and Mishkin, 1990).

Von Bonin et al. (1942) suggested that the ventral cortical pathway is responsible for object recognition. Each prestriate area of this pathway sends projections across the splenium of the corpus callosum to the prestriate area of the other hemisphere. In addition, the tail of caudate, the claustrum, the amygdala and the hippocampal complex all receive input from the areas in the ventral pathway (Yeterian and van Hoesen, 1978). These areas, however, are believed to play a greater role in higher cognitive information processing rather than in pure object recognition processes (Brothers and Ring, 1993). Ungerleider and Mishkin (1982) argue that even though the ventral pathway (V1-V4-PIT-AIT) is essential for the analysis and coding of the physical dimensions of visual stimuli and the recognition and identification of visual objects, it is unlikely that this pathway is involved in even higher visual processing such as motivation and emotions. It was suggested that for object recognition the IT cortex might be the final stage and that the amygdala codes the emotional and social significance of the recognised object (Ungerleider and Mishkin, 1982).

Receptive field sizes

In 1972, Gross and his colleagues found that single cells in the inferior temporal cortex have large visual receptive fields (Gross et al., 1972), which, however, are activated by quite complex stimuli compared with cells found by Hubel and Wiesel (1968; 1970) in the striate and prestriate cortex. Cells in early visual areas (e.g. V1) have 'classic' receptive fields. That is, the cell will only respond if the optimal visual stimulus is presented in the correct location within the visual field. Such receptive fields are often very small. However, the further along the ventral pathway one maps out the receptive fields of cells, the larger the receptive fields become. That is, cells in V4 have on average a receptive field of 4 degree^2, cells in PIT of 16 degree^2, and cells in the anterior IT 150 degree^2. Since approximately 60% of the cells in the IT region have bilateral receptive fields, presenting the optimal stimulus of a cell nearly anywhere within the visual field, will result in the activation of the cell and therefore

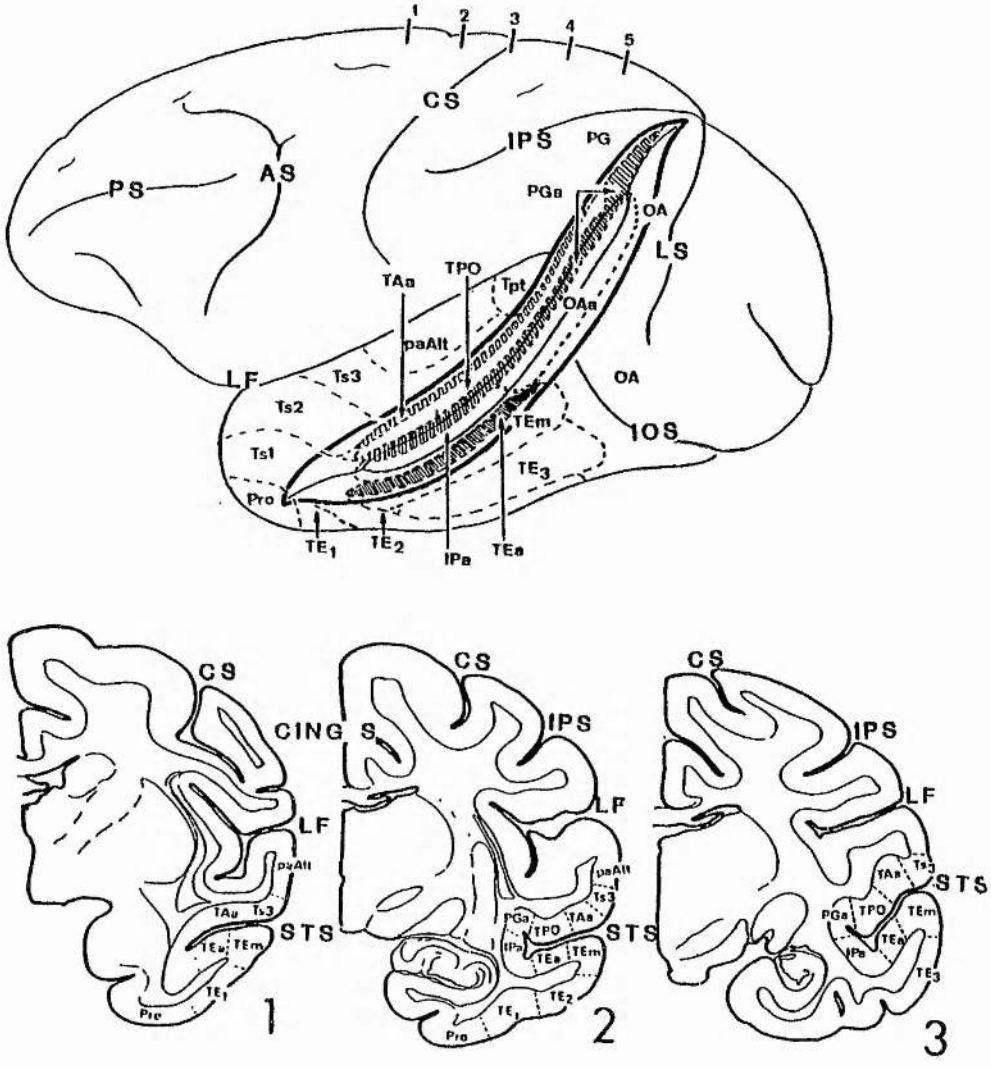
does not influence cell responses. The inferior temporal cortex, therefore, could provide the neuronal mechanism for stimulus equivalence across the whole retina (Gross, Rocha-Miranda et al., 1972; Gross et al., 1977).

Anterior STS

Visual information is then passed from the anterior IT to the anterior superior temporal polymodal area (STPa) in the superior temporal sulcus (STS). The STS runs from the lunate sulcus (LS) and area PG by the occipitoparietal border to area TG in the temporal pole (Seltzer and Pandya, 1978; 1989). Seltzer and Pandya (1989) suggested that the superior temporal gyrus can be subdivided into several areas: Ts1, Ts2, Ts3, paAlt, Tpt and KA (primary auditory area). In addition, the STS is divided into the upper bank, lower bank and the fundus. The upper bank of the STS is divided into three regions: area TAa, TPO (1-4) and PGa, whereas the lower bank is subdivided into areas IPa, TEa, TEm (present in anterior STS) and OAa (present in posterior STS) (Fig. 2.3). Cells found in the upper bank and fundus of the STS (area PGa and IPa) have been found to be polymodal (i.e. responsive to visual, auditory and somatosensory stimuli) and hence the area has been named the superior temporal polymodal (STP) area (Bruce et al., 1981). The STP is equivalent to T3 of Jones and Burton (1976) and overlaps with area TPO and PGa of Selzer and Pandya (1978). Cells selectively responsive to only visual stimuli (unimodal responses) have been found in areas TAa, TEa and OAa (see e.g. Gross, Rocha-Miranda et al., 1972).

Different projections arise from the anterior and posterior STP (STPa/STPp). Posterior areas, such as area PGa, project to unimodal parasensory association areas of the parietal lobule (area PGm and IPS) and the inferior temporal region (area TE) (Barnes and Pandya, 1992). Furthermore, it is suggested that the multimodal area TPO (especially from the middle region, TEO2, TEO3) found in the STS projects back to unimodal somatosensory areas in the parietal cortex (area PGm), to the middle inferior temporal parietal lobule (areas PFG, rostral PG), to auditory areas in the superior temporal gyrus (areas Ts3, paAlt, and Tpt), and to visual areas in the rostral portion of the superior temporal gyrus (area Ts1 and Ts2) and posterior parahippocampal gyrus (caudal area TF, and midportion area TL) (Seltzer and Pandya, 1991; Barnes and Pandya, 1992; 1994). Area IPa in the rostral depth of the STS projects to

Figure 2.3. The lateral surface of the cerebral hemisphere of the macaque monkey showing the architectonic parcellation of the superior temporal sulcus (STS) (enclosed in heavy dark lines), superior temporal gyrus, and inferotemporal region. Below, coronal sections of the hemisphere, taken at the levels indicated on the lateral surface, to show the architectonic areas of the STS (reproduced from Selzer and Pandya, 1989).



parahippocampal areas IPL and STG, to orbital areas 11 and 14 and to lateral areas 10, 12 and 46 of the frontal lobe. More anterior regions of the STP, such as area TEa and TEm, receive visual input from area TE1, TE2 and TE3 of the inferior temporal cortex and projects back to these areas, and in addition projects to areas in the ventral temporal lobe such as perirhinal cortex and areas TF and TL of the parahippocampal gyrus (Seltzer and Pandya, 1991; Barnes and Pandya, 1992). Furthermore, areas TEa and TEm also project to orbitofrontal areas 11 and 12 and lateral frontal areas 8 and 46 (Seltzer and Pandya, 1989).

In this thesis I will investigate the neuronal response of visual selective cells found in the anterior STS, predominantly from the upper and lower bank of the STPa.

Summary of visual processing stages for object recognition

It has been suggested (Perrett and Oram, 1993; 1994) that there are several (at least six) visual processing stages involved in object recognition: V1 is involved in extracting edges and contours from the visual scene. This information is passed on to V2 where cells are responsible for defining more general contours, which leads to the more precise definition of simple edge configuration in V4. By combining the selectivity to these edge configurations in PIT, simple shape selectivity is achieved. By merging this information on specific simple shape, more complex shape selectivity occurs in CIT. Selectivity increasing in shape complexity is accomplished by combining information obtained from earlier areas, resulting in high and complex shape selectivity in PIT of abstract forms and object-features. These forms and features may even be highly view-, orientation- and size-dependent. Finally, the STPa may be involved in constructing an object-description which is able to generalise across different visual attributes such as the orientation, size and view of the object. Such object-centred descriptions, however, could also be found in cortical areas subsequent to STPa, for example, the temporal pole.

Depending on the degree of recognition required for a task, not all stages of the visual processing pathway have to be included for recognition to occur. That is if recognition of, for example, a simple geometric shape of an object is sufficient (say recognising a big blob on top of the upright cylinder, for discriminating a tree from a house), the information processing pathway up to PIT be will be adequate. If, however, more detailed information about an object is required for recognition, such as the

particular features present and the orientation of those features, higher visual processing areas are essential.

Chapter III

OVERVIEW OF FUNDAMENTAL OBJECT RECOGNITION MODELS

This chapter will provide an overview of some models and theories of object recognition relevant to the empirical studies of this thesis. Historically, earlier models of object recognition put forward by Marr (1978) and Biederman (1987) have a profound impact on the object recognition literature and are therefore described and discussed here. More detailed information about experiments and models which are specifically relevant to chapters concerned with coding orientation and size information are described and discussed in the empirical chapters. Some overlap and repetition can therefore be expected which will be kept to a minimum.

Psychological object recognition models

Visual object recognition is a fundamental part of our everyday activity. The brain is able to compare sensory information with internal representations of objects apparently independent of the object's orientation, size and distance, part occlusions and lighting conditions. When viewing an object, a representation or description of that object stored in memory has to be accessible for recognition to occur. A good visual description of an object will ensure that important information about the object is easily accessible, whereas irrelevant information is hard to access if taken into account at all. Generally, it is important that an object has a unique description, for recognition of that object to occur. If a description employs several representations of an object (see later) then the description has to indicate which of the representations is valid in the viewed circumstance. The description also has to be quite broad (stable) to allow changes in viewing conditions (e.g. size, orientation, view, part occlusion), though at the same time, has to be sensitive to small changes of the object if these changes are crucial for e.g. interaction, social signals etc.. Furthermore, a description of an object has to be accessible with limited amounts of information available allowing the computational processing being economical. Such a *single* description is difficult to design, and therefore one might expect to find a 'space of representations' depending on the visual task involved (Plaut and Farah, 1990).

Many models of object recognition (e.g. Ullman, 1989) have been derived from early two-dimensional (2D) pattern recognition models. These early models can be divided into three groups: a) *Template analysis*, where a pattern representation consists of a series of 'templates' each of which fit some corresponding image/picture-like representation; b) *Feature analysis*, where the pattern representation consists of sets of unchanging features (each coded by feature detectors, e.g. 'demons'); and c) more recently suggested *structural descriptions*, which provide a language with which one is able to construct a theory of pattern and object recognition.

Template matching

Early work on pattern recognition examined the recognition of alpha-numeric patterns. Stimulus constancy for alpha-numeric patterns, however, is quite different from stimuli constancy for object recognition, since letters and numbers are two dimensional (2D), whereas objects are three dimensional (3D). There might be a change in the form, size and orientation of 2D letters or numbers as observed in 3D objects, but other problems unique to 3D object recognition, such as different view-points and shadows, cannot be taken into account with alpha-numeric stimuli. Nonetheless, it is still worthwhile to discuss the problem of recognising alpha-numeric patterns, since it gives us an opportunity to question certain theories for recognising 2D and 3D objects.

Template matching theories (Neisser, 1967; Pinker, 1984; Humphreys and Quinlan, 1987) would provide a simple way of explaining how we recognise alpha-numeric patterns. According to these theories, a template for each known character is stored in long term memory. An incoming pattern is simply matched against all existing templates and if there is sufficient correlation between the incoming pattern and a certain template (though no one has yet provided a satisfactory theory for how this correlation is measured), the pattern is recognised. Note that the whole incoming image is matched against various templates and that the image is not divided into component parts of the pattern/object viewed. To allow slight changes in the image's attributes (such as orientation or size) the incoming pattern is standardised by, for example, rotating the 2D pattern to the same orientation and adjusting the size to the same scale as that of the template. Palmer et al. (1981) further suggest that mental

rotation of input images could occur in 3D, so that images of 3D objects could be aligned with a template of a 'prototypical' view of that object. The 'template' consists of an image-like representation of the possible form of the relevant alpha-numeric character.

These standardisation procedures used on the incoming image (arising from the original descriptive template theories) are, however, questionable, since they may not be sufficient to reliably carry out recognition of the pattern/object. The resulting pattern of a hand-written character after size and orientation normalisation could still be substantially different from the corresponding template of the character, so that it could still be classified or matched against the wrong template. This could easily occur with letters which appear very similar, such as 'U' and 'V'. To achieve pattern recognition with a system of templates, it seems that there would have to be a whole series of templates for each pattern, requiring a very large storage capacity of the brain. In addition, if the incoming pattern is matched to all the templates present in the brain in a serial fashion, then it would make it difficult to understand the process of recognition, since the time required for success would be far too long. If, on the other hand, there was an alternative system of template matching utilising parallel processing and high tolerance of each template, object recognition models using templates might be plausible (see below).

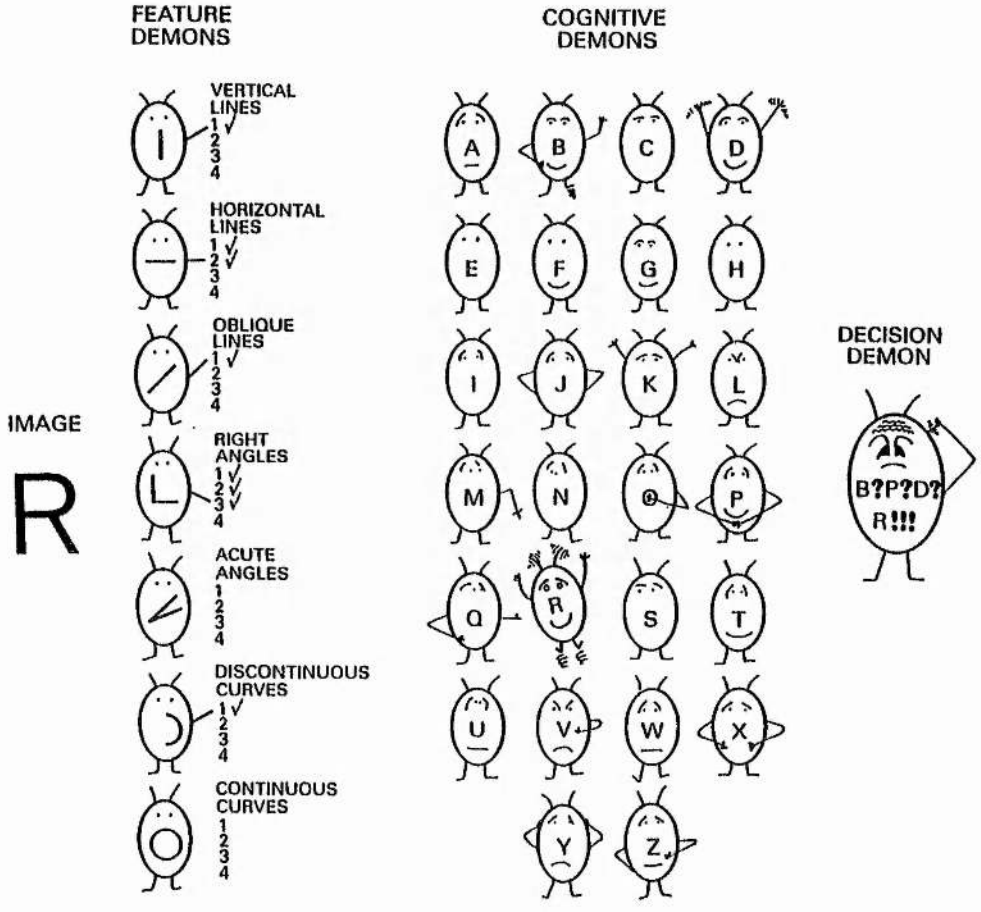
Feature Analysis

Feature analysis models (Selfridge, 1959; Sutherland, 1968) for pattern recognition were very popular amongst psychologists during the 1960's and early 70's. These models suggest that combinations of *features* of the incoming pattern are identified and used as the basis for subsequent processing. This is different from previously described template matching theories, since component parts of the object are involved in recognition rather than there being a template for the whole incoming image of an object. Neurophysiologists such as Hubel and Wiesel (1962) provided evidence of neurons that were detectors for 'edges', 'bars' and 'corners' in the cat's visual cortex. These neurons, suggested to be 'feature detectors', were used in support of feature analysis models, such as the 'Pandemonium system' described by Selfridge (1959). This model, originally designed for encoding Morse signals, was adapted by

Neisser (1967), and Lindsay & Norman (1972) to a system suitable for alpha-numeric pattern recognition. This scheme for processing visual patterns has a number of subcomponents called 'demons' - which could be analogous to neuronal feature detectors. The incoming pattern is first analysed by 'feature demons', which selectively respond to particular local configurations, such as right angles, vertical lines, curves etc. (see Fig. 3.1). The information from the feature demons is then combined and presented to the 'cognitive demons' which represent a particular letter (or object). For example, a 'T-demon' requires input from a vertical- and a horizontal-bar-feature detector and a third feature detector reporting the perpendicular interaction of the vertical bar dissecting the horizontal bar at its mid-point and at the top of the vertical bar. Other cognitive demons would also be activated, though to a lesser extent, by some of the same features, e.g. a demon for the letter 'J'. Finally, the 'decision demon' selects the right letter based on which cognitive demon "shouts" loudest. This "shouting" (the activity of the demon) could be compared with proportional neuronal firing from feature detectors such as those found by Hubel and Wiesel (1962; 1968). In this model, individual incoming pattern are represented as sets of critical features. The processing of the pattern is in a hierarchical fashion which becomes more and more abstract as information is processed.

Feature analysis models were used by Barlow (1972; 1985) and other physiologists to understand and explain the properties of simple cells found in V1 and cells responsive to more complex images (e.g. faces) present in higher visual areas of the monkey. It could be suggested that simple cells may play the role of feature demons and pass on their information to cognitive demons located further along the visual processing pathway. In this case, cognitive demons would respond to particular patterns/objects. That is, a particular object is represented in the brain by the firing of one or several 'gnostic' cells (Barlow, 1972). This notion also brought about terms like 'cardinal cells' (which were believed to be numerous), 'yellow Volkswagen detector' (Harris, 1980) and 'Grandmother cell' (Lettvin, 1969; see appendix in Barlow, 1994) - that is, a person is able to recognise their grandmother because somewhere in their brain are single cells selective for the image of the grandmother. A Grandmother cell would generalise across multiple images of the grandmother independent of lighting, distance, etc.. This specificity carries much biological importance for object

Figure 3.1. A Pandemonium system for classifying letters (see text for further explanation) (adapted from Selfridge, 1959).



recognition. Single unit recordings in the macaque have shown that there are cells responsive to the identity of one face (but not to others) and capable of generalising across viewing conditions (Perrett et al., 1984, Young, 1993). On the other hand, cells have been found which responded to the human face/body regardless of its identity. One could therefore suggest, that there are different systems in the brain responsible for high specificity, such as grandmother cells, and other systems with greater and varying tolerance for recognition.

The finding of highly visual selective cells to complex objects such as faces, hands, bodies and fractal patterns (see e.g. Gross, Rocha-Miranda et al., 1972; Perrett et al., 1982; 1984; Miyashita, 1988; Miyashita and Chang, 1988; 1991; Wachsmuth et al., 1993; 1994) found in IT and anterior STS would support sparse population coding, where single cells show great specificity for behaviourally relevant stimuli (Konorski, 1967; Barlow, 1972; 1985; Young and Yamane, 1993). If, however, the neuronal coding was very tight, as tight as a 1:1 cell to concept mapping, then one would expect to find that the system has a problem in deciding which cell represents the best 'fit' of the incoming image. Furthermore, if there was only one cell which was to respond when viewing e.g. your grandmother, then this neuronal response might not be detectable above spontaneous activity (noise) of other surrounding non-grandmother cells.

'Population' theories, on the other hand, anticipate that cells would have quite broadly tuned responses to a wider range of stimuli, since the underlying factors of perception and behaviour is the distributed pattern of activity in a neuronal population (Freeman, 1975; Georgopoulos et al., 1994). Baylis et al. (1985) point out that cells selectively responsive to faces are often broadly tuned in that they will respond to a wide variety of different faces. Hence, a population of cells might be very selective and tightly tuned to a particular class of pattern/stimuli, but tolerate variations within the stimulus class (e.g. identity). Rolls (1994) elaborates on this theory by suggesting that cells do coarse coding and have a more specific tuning for particular objects with a graded response region centred around the optimal response. This graded response region would allow for generalisation across changes in e.g. size, position, orientation etc.. Even though such a system might control a combinatorial explosion, very large numbers of neurons with a good learning capacity are required.

In summary, feature analysis models (or related models described above) are able to 'learn' to give more or less emphasis to a feature depending on the importance of this feature in discriminating between the different incoming patterns. However, more out-dated feature analysis models are insufficient for describing pattern and object recognition, since incoming patterns are described as sets of features which act like many different templates (Neisser, 1967). This would bring one back to the same problems as described for original restricted template matching models; i.e. too many templates would have to exist to cover every possible view, orientation, size etc. of an object ever seen. Several newer models of object recognition have attempted to solve this problem and are described above (see also e.g. Rolls, 1994). Nonetheless, the models described above are very weak in describing the spatial relationship between the different features/parts in relation to a reference frame of an object (see below).

Structural description

Structural description provides a symbolic 'language' with which one can construct a model of how pattern or object recognition is achieved. This language, however, is not linguistic, but rather consists of a set of abstract or propositional descriptions about a particular object configuration. Thus, an object or a pattern is described in terms of its components, which allows precise description of the structural arrangements of the object's components (Selfridge, 1959).

There are two different approaches to structural descriptions for object recognition: 1) Parts of objects are specifically related to a perceptual reference frame, such as the major axis of the object (Marr and Nishihara, 1978; Hinton, 1981; Marr, 1982; Humphreys, 1983; Lowe, 1985); or 2) Objects are recognised on the basis of their volumetric parts and the parts relative position to each other (Biederman, 1987): recognition by components (RBC).

The role of components in object recognition

Marr and Nishihara (1978) argued that mental representations of objects occur in a 3D, volumetric (as opposed to only the surface of an object) and object-centred form. In an object-centred representation the view-point is not relevant, since description of the object is based on an object-centred co-ordinate system where

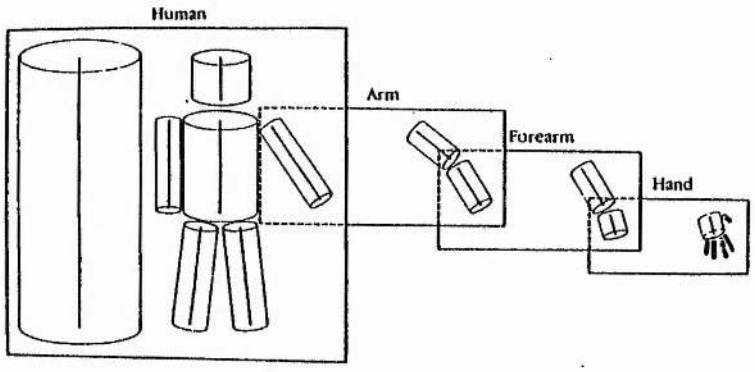
components of the object are described in relation to each other, or to the major axis of the object (see later). In their model, information about an object is processed through a series of four stages: The first stage (primal sketch), represents an object in terms of its edges, bars, and blobs which carry characteristics of orientation, width, length, contrast, and position. In the second stage, the 2½ D sketch, cues for depth, surface texture, and contour occlusion are added. A 3D structural description is achieved in the third stage, by describing the object in an object-centred fashion and classifying the object with the help of stored structural descriptions. Finally, in the fourth stage, the structural description is used for semantic interpretation. Marr and Nishihara (1978) suggest that for describing an object, one has first to set up a co-ordinate system which is determined by the overall shape of the object itself. Each 3D model description specifies the following: 1) the model axis: a single axis which defines the overall volumetric shape of the object; 2) the relative spatial arrangement, orientations and the length of the component axes; and 3) the volumetric primitives (generalised cones) which are associated with each axis. 4) Then, repeatedly, each component axis becomes a model axis for its sub-components, until the wanted level of detail is reached (see Fig. 3.2a). Alternatively, for recognising parts of the whole object the model allows direct access to the information at any subordinal level. For example, a limb of a human body can be described as an (model) axis of an component part relative to its major axis (of the whole body). This model axis of the arm becomes a major axis if components such as the upper arm, forearm and the hand are described in relation to their major axis (arm itself) (Marr and Nishihara, 1978).

To describe these volumetric primitives, Marr's model of object recognition uses 'generalised cones'. A generalised cone "refers to the surface created by moving a cross section along a given smooth axis. The cross section may vary smoothly in size, but it shape remains constant." (Marr, 1982). These generalised cones do not reveal any detail of the object's part such as surface pattern or texture, but have a cylinder like property. This simplifies the object structure substantially, resulting in a 3D 'stick-figure' (see Fig. 3.2b). In addition, this kind of representation does not use information obtained from surface patterns.

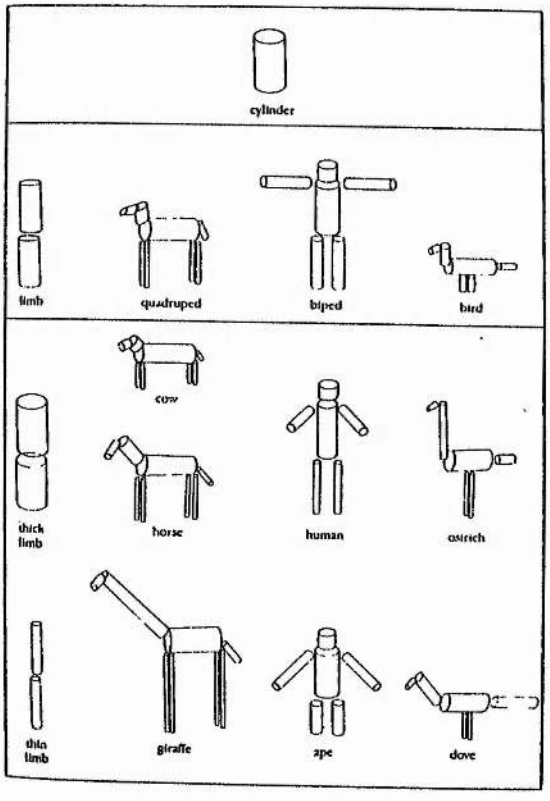
Marr realised that occluding contours (silhouettes) were very effective visual stimuli for object recognition. He argued that any silhouette could arise and be

Figure 3.2. Marr and Nishihara's (1978) model of object recognition. a) A hierarchy of 3D models. Each box shows the major axis for the figure of interest on the left, and its component axes to the right. b) A catalogue of 3D model descriptions at different levels of specificity (reproduced from Marr and Nishihara, 1978).

a)



b)

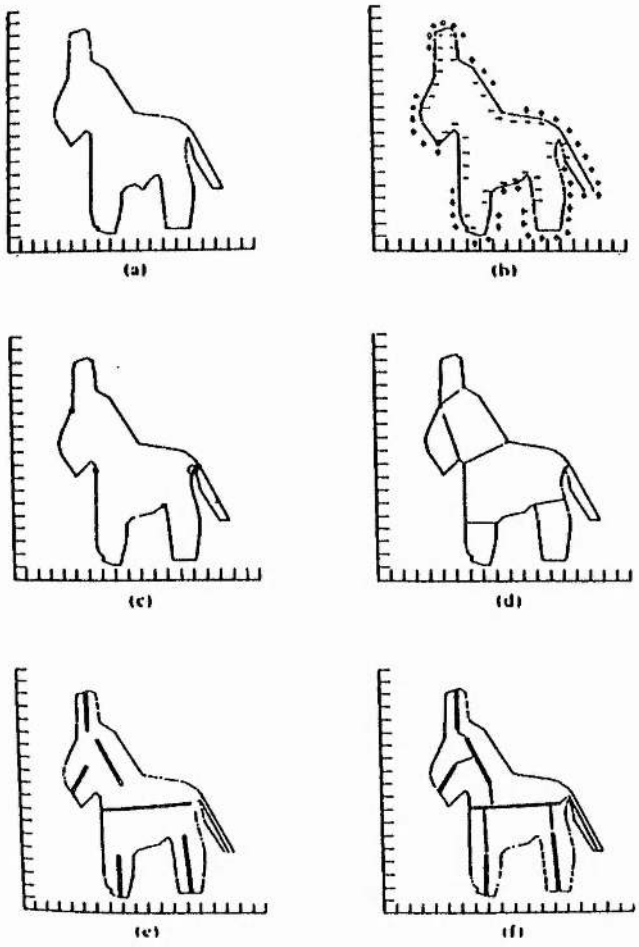


constructed by using a number of different 3D shapes. However, we are all capable of recognising a silhouette as the same 3D object. He therefore suggested that there must be some kind of constraints on the possible range of interpretations of the contour, which might be due to assumptions made by the visual system. Such assumptions were suggested to be that 1) each point of a silhouette represents a point on the surface of the 3D object; 2) two points close together in a silhouette will also be close together on the object; and 3) all the points on the contour line of a silhouette will lie in a single plane of the object viewed. These assumptions lead Marr to conclude that any silhouette of an object (and therefore the 3D object itself) is based on one or several generalised cones which are described in relation to the principal axis of the object (Marr, 1982).

The next question addressed was how these axes are derived from the object's occluding contours. Marr and Nishihara (1978) suggested that the complex contour is *segmented* into a number of distinct parts (generalised cones). They gave an example of a toy donkey silhouette and segmented its contour into distinct parts by finding the different concavities in the contour (Fig. 3.3). This was done by first labelling the different concave sections, which brings out the strong segmentation points. Then the outline was divided into a set of smaller segments using the strong segmentation points just defined. Next, these points are connected to other strong, neighbouring segmentation points of the contour in a straight line. Finally, the component axes are determined which are then linked up and related to each other.

Marr (1977) suggested that where there is a concavity in the contour of an object, there is a discontinuity indicating a change - the contour is not continuous. Nonetheless, he argued that from occluding contours alone one is unable to derive enough relevant information about the object for its recognition. This is because information about, for example, the parts of the object which are not defined by the overall contours of the object, such as internal parts (eyes, mouth, surface detail etc.), are lost (Marr, 1977). For example, the silhouette of a human body can reveal the torso, all four limbs and the head. If, however, any of these body components are pointing toward the viewer, they would not be visible in a silhouette. Furthermore, no information about internal parts of an object is given, so that a model of object recognition based on solely the silhouette or the 2½ D sketch (such as that of Seibert

Figure 3.3. a) How the major axes are defined from an outline of a toy donkey. b) Convex (+) and concave (-) sections are labelled. c) Strong segmentation points are found. d) The outline is divided into a set of smaller segments making use of the points found at c) and rules for connecting these to other points on the contour. e) The component axis is found for each segment. f) The axes are related to one another (this lines).(Reproduced from Marr and Nishihara, 1978.)



and Waxman, 1991) is implausible. Marr does not provide an answer to this problem in his theory.

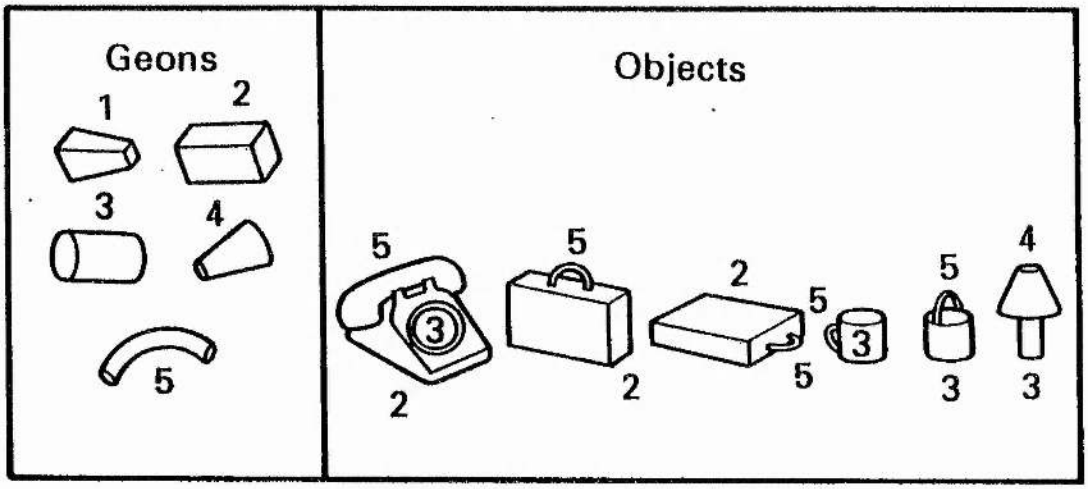
It was put forward by Hoffman and Richards (1984) that different parts of an object intersect at a contour of concave discontinuity of their tangent planes. In other words, whenever two surfaces interpenetrate, they will always meet in a concave discontinuity. So that where one part penetrates the other there is a concave cusp formed. Therefore, any concavity of an object will indicate the division between the contours of two intersecting parts of the object. They also pointed out that in the contours of a smooth object, the concavities can be found by looking at the sites of the object contour with the greatest negative curvature, i.e. the most inward pointing part of the silhouette of the object will indicate the presence of a concavity.

In contrast to Marr and Nishihara's (1978) theory, however, Hoffman and Richards (1984) suggest that the segregation does not depend on the parts being generalised cones, but rather is independent of the nature of the parts of the object.

Examples such as the 'face-vase' image support suggestions that the parts of an object can be predicted from the reversal of their figures. If the vase in this image is perceived, the concavities point inwards, defining the base, the stem and the bowl of the vase. However, when the faces are perceived, the concavities point into the face, defining the parts of the chin, nose, mouth, forehead etc.. Hence, depending on how the contour of the image is divided into parts, by looking at interpreting the concavities prior to recognition, different objects can be perceived (in this example either a vase or two faces each with its own distinct set of parts).

Biederman's theory (1987) of object recognition takes elements from the theories of Marr & Nishihara (1978), Hoffman & Richards (1984), as well as from the works of Lowe (1987). According to Biederman's theory, perceptual recognition of objects is a process by which the image is segmented into cone-components (called 'geons') which are derived from five properties of edges in a 2D image. Geons, no more than 36 in number, are like Marr's generalised cones but more diverse (Fig. 3.4). They all arise by translating a 2D outline along an axis which can either be straight or curved. The 2D outline can, during the transitional process expand, contract and bend. Because of these additional properties, a more flexible shape recognition is produced

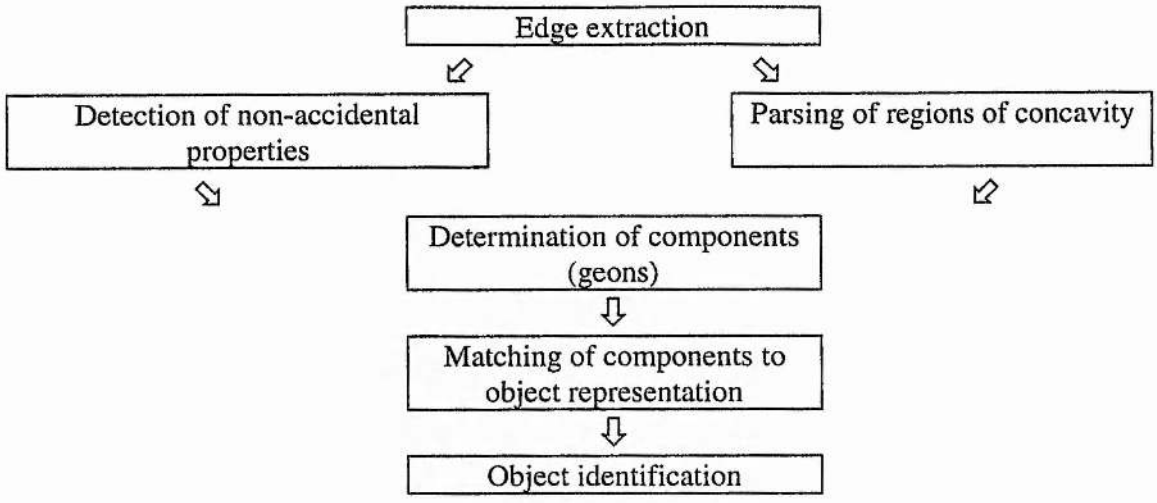
Figure 3.4. Left side: Some examples of Biederman's volumetric primitives called 'geons'; **Right side:** Examples of objects made up of a number of geons (suitcase, lamp, mug etc.). Note, that the spatial relationship between the geons is highly important (see text); (reproduced from Biederman, 1987).



compared to Marr's model, though Marr's model does offer more precise information about the spatial arrangements of the components.

Biederman's (1987) model of object recognition starts off with extracting edges from the image of a viewed scene of 3D objects. The edges enhance the presence of non-accidental properties, and hence indicate that these properties are reflected by the image from the real world. So that if, for example, a set of parallel lines are present in an image, it can be assumed that these lines are a property of the edges of an object in the real world and are parallel in 3D space. Biederman points out that because these non-accidental properties are constant over different view-points they can be used as additional information to determine the geons used for the object representation stored in memory. The other main information needed to determine the geons, is derived by segmenting the silhouette of the object image at regions of sharp concavity, resulting in a set of object parts (as done in Marr's model, 1977). These determined parts are then matched against the representations of the different geons. From the nature and arrangements of these geons, the structural model of the object can be defined (every known object has its own structural model based on its geons, the geon's relative sizes, orientations, place of attachment, etc.), resulting in the recognition of the object. It is very important to note the spatial arrangement between the geons of an object, since this builds the base of recognition of that object. Two different objects may have the same number and type of geons, but different spatial arrangements between them. Biederman (1987) uses terms like "on top of" or "attached at the side" to specify the spatial arrangement between the geons. Some of these descriptions (e.g. on top of), however, are based on a *viewer-centred* co-ordinate system, since they describe the position of the geon in relation to the viewer. Such view based terms should not be used in Biederman's *object-centred* description, since he explicitly points out that there is no top or bottom in a representation of an object.

A summary of the different stages involved in Biederman's model of object recognition is given below:



Biederman (1987) carried out some experimental work to support his theory of object recognition by components. Subjects were to name as fast as possible simplified line drawings of objects presented for 100 ms. The complexity of these objects ranged from complete objects composed of only two geons to complete objects composed of nine geons. Randomly interspersed between the presentation of complete images of objects, were images of different incomplete objects. For example, a 9-component object could have been presented with 1, 2, 3, 4, 5 or 6 components missing. Alternatively a 3-component object could have been presented with only two components present. Biederman found that subjects take longer to name and make more recognition errors if some parts of an object are missing irrespectively of how complex the complete version of the object was. However, for the more complex objects, recognition error was low (< 10%) if more than three geons of the object were presented (Biederman, 1987). Furthermore, the addition of colour, brightness and texture did not significantly improve recognition response times (Biederman and Ju, 1988).

In an different experiment, Biederman (1987; Biederman and Gerhardstein, 1992) measured subjects' response times and error rates in naming images which were degraded in several ways: 1) the information for the object's geons were present, i.e. the contours of an object were detected at regions where they can be replaced by coliniarity or smooth curvature (non-accidental properties), and 2) the information for the geons of the object were missing, i.e. the contours were detected at regions of

concavity so that coliniarity and smooth curvature of the segments bridges the concavity (Fig. 3.5). The results showed that recognition was slightly faster and more accurate (approximately by an error rate of 30%) if the information for the geons was present. Biederman (1987) concluded that, as in the previously described experiment, recognition of an object is carried out on the basis of the object's geons. However, the suggestion that recognition of an object is better if regions of sharp concavities are left intact in a fragmented image, than the recognition of an image where the object is fragmented in such a way that the concavities are removed, may be questionable. Biederman and Gerhardstein (1992) argued that in the first instance geons are more easily found for recognition purposes than when the concavities are removed from the image. However, this is debatable since priming and recognition of an image is best if enough information is available in the image to support perceptual closure. Perceptual closure is based on the Gestalt principle of filling in gaps in an contour to achieve the most meaningful forms of an image (Snodgrass and Feenan, 1990). Biederman's image of a fragmented object where regions of sharp concavities are intact could aid in Gestalt principle closure rather than an image of a fragmented object where the concavities are removed. Hence, this could influence the results found in these experiments.

In addition, Biederman and Cooper (1991b) studied object recognition with a priming technique, where a priming image was briefly presented before the target image of an object. They investigated whether or not a priming effect is due to 1) the object's features, such as e.g. edges and vertices; 2) the object's model, i.e. what the object is (name of object class); or 3) the component parts (geons) of the object. The subjects were presented with a priming set of familiar objects which had half of their contours deleted by either deleting every second feature from each part (for investigation point 1), or removing half of the parts of the object (for investigation point 3). The subjects were then asked to view a second set of images (the primed images) which could either be: a) the identical object image as the priming image; b) the complementary image with the missing contours of the priming object; or c) a different example of the priming object (such as a grand piano and a ordinary piano). The results showed that target images were recognised equally fast and accurate if the image had been primed with either the identical or the complementary image. If

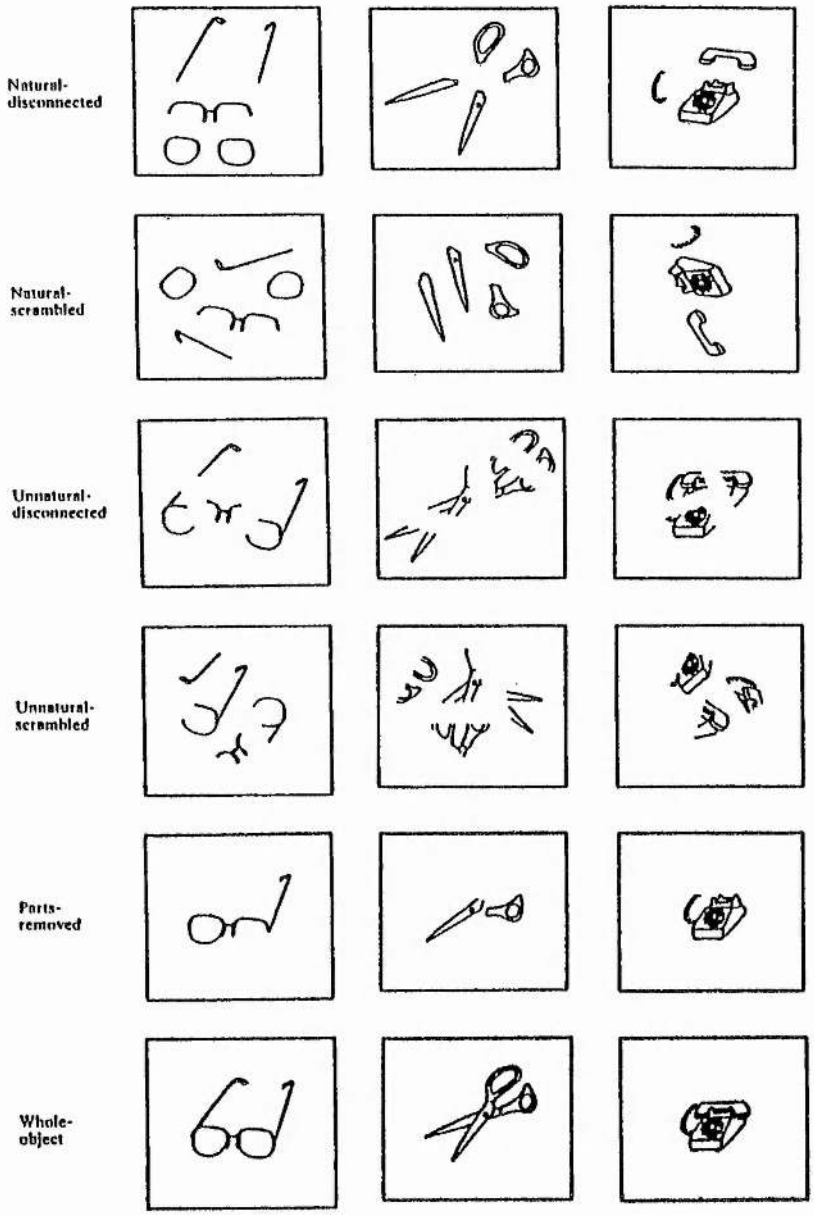
Figure 3.5. Examples of objects being parsed in two different ways: Left column: original intact version. Middle column: sites of concavities still intact: the contours have been deleted in regions where they can be replaced through colliniarity or smooth curvature; Right column: parsing at sites of concavities, resulting in colliniarity or smooth curvature of the segments bridging the concavities (reproduced from Biederman, 1987).

priming occurred with a different exemplar of the same object class as the target image, RTs and error rates were higher than when using the other priming conditions. Biederman concluded that visual priming was not based on image vertices and edges and that a visual image representation would have to be more global than that specifying the image features. This would leave two plausible types of the representations named above: either a representation based on the component parts of the object, or a representation which is a specific object model itself (e.g. a grand piano). However, when the component parts (geons) of an object were removed, RTs and error rates were much better if the target image was primed with the identical image, rather than with the complementary image. Hence, priming only occurred when the convex components (geons) were present in priming image (as well as in the primed image). It was concluded that "*all* the visual priming in object naming can be attributed to the explicit (actual) presentation of the components in the image" and the spatial arrangement between them (Biederman and Cooper, 1991b).

Hierarchical organisation: Bottom-up vs Top-down processing

The theories described up to this point are based on bottom-up processing, i.e. recognition of parts of an object lead to the recognition of the whole object. A controversial view of these theories has been promulgated by Baker-Cave and Kosslyn (1993) who argue that object features (or the whole object) are matched directly with the image stored in long-term memory without first being parsed. This would be in support of 'top-down' processing theories of object recognition. Testing involved exposing subjects to line drawings of objects where the image was segmented (parsed) in different ways (Fig. 3.6). Segmentation was made at a) 'natural' boundaries and b) arbitrary sites of the object. In addition, half of these resulting line drawing of the disconnected objects were scrambled. Furthermore, images of objects where parts of the object had been removed were tested. Subjects were presented with an image of an object for 1 sec (in some cases only 200 ms) and were asked to name the object as fast as possible. As can be expected, RTs and error rates for correct naming of the object were fastest and most accurate if the whole object was presented. There was, however, no difference between RTs and error rates for the different types of disconnected conditions. That is, the way an object is parsed does not influence recognition. If the

Figure 3.6. Examples of stimuli drawings used by Baker-Cave and Kosslyn (1993).



disconnected object parts (natural or unnatural parse) were scrambled, naming RT and error rates were higher than when the parts were simply disconnected from each other. It was concluded that one critical factor for recognition of an object was the spatial relationship between the parts of the object (scrambled or proper organisation). Furthermore, Baker-Cave and Kosslyn (1993) found no interaction between the effect of (i) the way an object was parsed and (ii) the arrangement on the page (scrambled or unscrambled). Baker-Cave and Kosslyn (1993) suggested on the basis of their results that in normal viewing conditions the overall shape of an object is encoded without the image being parsed into parts.

Unfortunately, Baker-Cave and Kosslyn (1993) did not look in greater detail at the effect of part deletion, which might influence and/or diminish the quality of object recognition. They only showed that a whole intact object can be recognised and named faster than an image of the same object with some of its parts removed. According to Marr's theory, part deletion should not greatly influence the recognition process, since recognition can start at different levels of the 'stick-like model', i.e. one can access the model at different levels of detail depending on the need for specific information on e.g. a particular part of the object. On the other hand, Biederman's theory of object recognition very much depends on the object' components, so that deletion of parts would have a greater impact on the recognition process.

Earlier theories based on bottom-up processing, where components of an object are identified and recognised prior to the recognition of the whole object, were put foreword by Lowe (1985), Tarr (1989), Ullman (1989) and Kosslyn (1990). It was suggested that object recognition is based on processing information from edges, or regions of colour or texture (Tarr and Pinker, 1989; Ullman, 1989; Kosslyn, Flynn et al., 1990). These features of the object are then matched against the representation in long-term memory. All these theories are based on a viewer-centred object recognition mechanism (see below), since the information used for recognition (edges, colour and texture) is very specific to certain views of the object rather than to the overall configuration of the object.

Physiological evidence for the role of component parts in object recognition

It has been proposed that the inferotemporal cortex (IT) plays a central role in visual pattern and object recognition (Gross, 1973; Ungerleider and Mishkin, 1982; Schwartz et al., 1983; Desimone et al., 1984; Tanaka and Fujita, 1991). Many physiological studies have been carried out in this region of the monkey brain, some of which will be discussed below.

One of the attempts to study shape selectivity of individual cells in the IT cortex examined how IT cells can extract information about the overall shape of an object by analysing the local boundaries of the image (Schwartz, Desimone et al., 1983). Schwartz et al. (1983) investigated whether any object can be described in terms of its boundaries curvatures (the contours). They argued that the visual system describe these boundaries in two steps. First, the boundary orientation function of the shape, which is the orientation of the shape's boundaries measured at regular intervals around the perimeter, is determined. This boundary orientation is then expanded in a Fourier series. These Fourier series, which have a particular frequency (number of lobes), amplitude (lobe indentation) and phase (orientation), were named 'Fourier descriptors' (FD). A FD can also be described as a shape which is systematically varying in boundary curvatures. According to Schwartz's theory, every shape is described by its own sets of FD.

Schwartz and his colleagues (1983) explored the selectivity of single cells in the IT cortex of the macaque monkey by presenting visual stimuli of different complexity. The cells were tested with slits of light, complex objects and FD's. The FD's were white on black background (or vice versa, see Gross 1972) presented in a semi-random fashion at the centre of gaze. The stimulus was oscillated 1° per second for a presentation duration of 2.5 seconds. It was noted that 54% of the cells tested responded to specific FD's (different cells being maximally activated by different FDs); 26% did not show any response to any FD tested; 13% responded to all FD's; and 7% showed a multimodal tuning curve, i.e. responded to several but not all FDs tested. These response patterns were found to be the same for most IT cells even if the size and position of the optimal stimulus was changed (though there might have been a change in response amplitude). The study suggested that IT cells may be involved in the analysis of global boundary orientation. However, this cannot be the only function

of cells in the IT cortex, since some cells have been found to be sensitive to a particular shape in a particular colour or texture, or sensitive to 3D-objects, faces or hands (Desimone, Albright et al., 1984; Tanaka and Fujita, 1991; 1993; Kobatake and Tanaka, 1994) - all which normally are much more complex stimuli than FD. Some of these cells, nonetheless, were in addition sensitive to FD's. The conclusion drawn, was that selectivity for boundary curvature is not a characteristic of all IT cells, nor is it the only response property of many of these cells.

Another systematic attempt to determine the mechanism for selectivity for a particular stimulus dimension, such as shape, was conducted by Desimone and colleagues (1984). Previous work concentrated on recordings of cells selectively responsive to stimuli which were either simple, as for example white and coloured bars, or complex such as patterns, hands and faces (see e.g. Gross, Rocha-Miranda et al., 1972; Perrett, Rolls et al., 1982). Desimone et al. (1984) tested cells in the IT cortex systematically with simple stimuli, such as bars with different length and width, 2D patterns, and with a wide range of 3D objects. Once the preferred object/pattern of a cell was identified, they tried to isolate the critical stimulus which was responsible for the response of the cell. They also studied cells in the superior temporal sulcus which were selectively responsive to face stimuli.

This investigation showed that most cells in the inferior temporal cortex responded to many different visual stimuli and therefore it was concluded that these cells cannot be narrowly tuned 'detectors' for a particular object. This is in contrast to theories of 'Grandmother cells' and 'Yellow Volkswagen detectors' discussed earlier.

Half of the IT cells which showed some selectivity seemed to be selective for shape, colour, texture or a combination the three (Desimone et al., 1984). Of these cells, some showed the same kind of sensitivity to length, width, or a coloured bar as the cells found in earlier stages of the visual pathway (striate and prestriate cortex). As pointed out in chapter II, the receptive field of IT cells are much larger (almost always include the centre of gaze and usually extend into both the contralateral and ipsilateral visual field), than the receptive fields of cells found in the striate and prestriate cortex.

Other cells found by Desimone (1984) were only sensitive to stimuli with some kind of texture, or complex 3D objects. These cells would often fail to respond to

simple stimuli such as a bars. It was concluded that the IT cortex must play an important role in higher shape recognition.

IT cells which respond selectively to faces or hands were first identified by Gross et al. (1969). In 1972, Gross and his colleagues showed that there are IT cells which best respond to photographs of faces. Later investigations (Desimone, Albright et al., 1984), however, showed that cells which are selective for a specific object are very rare throughout the IT cortex. Perrett et al. (1982) demonstrated a more frequent appearance of cells selectively responsive to faces in the anterior part of the superior temporal sulcus (STS) which partially overlaps with the IT cortex. Nonetheless, the cells appeared to be distributed non homogeneously - a particular clumped distribution for particular features with cell type patches of 1-4 mm^2 (Perrett, Smith et al., 1984; 1985; Harries and Perrett, 1991).

Perrett and colleagues (1982; 1983; 1984) tested face sensitive cells for their neuronal responses to the sight of faces with component parts, such as the eyes, the mouth or the hair, being occluded from sight. It was reported that most of these cells responded significantly better to the face with certain parts of it occluded than to various control objects and base line activity (S/A). However, most cells responded better to the whole face than to just isolated parts of the face. Different cells were found to be maximally activated by different individual component parts of the face. Furthermore, some cells only required one face part present whereas other cells required several parts present. For the remaining cells the response magnitude was highly reduced when any part of the face was occluded, i.e. the entire face was necessary to drive the cell.

The overall configuration of the component parts of a face also seem to play a highly important role in the neuronal firing of these face responsive cells (Perrett, Rolls et al., 1982; 1984; Desimone et al., 1984). A face has several different shapes, contours, colours and textures which all have to be arranged in a certain order to compose a face. Perrett et al. (1982; 1983) found that when this arrangement was upset (i.e. jumbled faces) then the cell response of most cells was greatly reduced.

Desimone (1984) suggested that, as in other visual areas, cells in IT are not specific for a particular object stimulus, but rather a complex stimulus will activate many different IT cells. He concluded that the neuronal representation of an object in

the IT cortex is the pattern of activity of a population of cells rather than one specific cell (see also Tanaka, 1993). However, face and hand selective cells are an exception to this system (Desimone et al., 1984). This might be due to the fact that, for primates, faces play an extremely important role in social communication and therefore, might have been in favour during selective evolutionary pressure to build a separate (but not necessarily independent) neuronal mechanism coding biologically important objects such as faces and hands. Nonetheless, faces may not be extraordinary exceptions from objects since face stimuli could activate many millions of cells selective for face features, which might be analogous to the representation of other complex objects.

More recent studies of object selectivity in different parts of the IT cortex have been carried out by Tanaka et al. (1991) and Kobatake and Tanaka (1994). They suggested that the IT cortex is divided into two areas: the posterior (PIT) and anterior inferior temporal cortex (AIT). PIT receives its information from V4 and in turn projects to a wide posterior-anterior range of AIT. V4 also projects to the posterior part of area AIT (see chapter II and e.g. Desimone et al., 1980; Baizer et al., 1991).

Tanaka and his colleagues (1991; Kobatake and Tanaka, 1994) tested single cells in different cortical areas of the ventral pathway (areas V2, V4, PIT and AIT). All the cells recorded from in the these areas were classified into six groups:

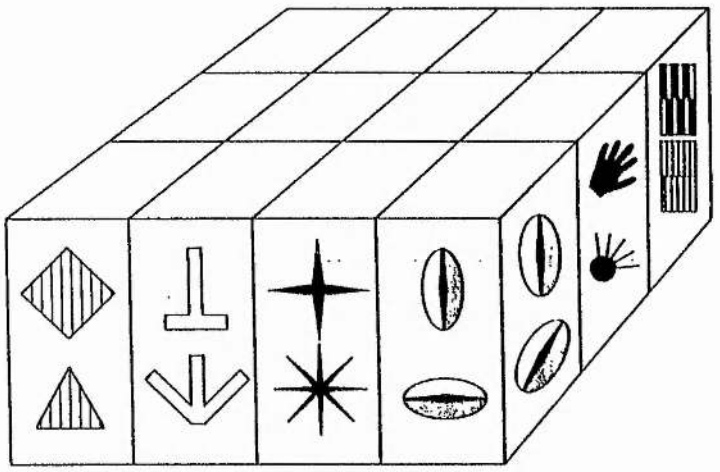
1) Primary cells: Maximally activated by slits or dots, where the exact shape of the stimulus is not crucial;
2) Texture cells: Maximally activated by textures such as stripe or dot patterns;
3) Elaborate cells: Maximally activated by pattern features which are more complex than only orientation of contours, colour blocks or simple textures;
4) Others: Cells which responded to a change in colour of a block or some specific movement;
5) Weak: Cells responding only very weakly to any stimuli tested; and
6) Unresponsive: Did not at all respond to any of the visual stimuli tested.

The distribution of these cells was as follows: Texture, weak and unresponsive cells were scattered throughout the whole IT cortex. V2 cells were mainly selectively responsive to texture but not shape (texture cells), whereas primary cells were mainly found in V4 and PIT. 38% of V4 cells required quite complex features for a cell to respond (elaborate cell), whereas the percentage of such cells selective for complex

features (e.g. bars or discs with adjusted size, orientation and/or colour) increased up to 49% in PIT. The highest relative proportion of elaborate cells, however, were found in the remaining anterior part of the IT cortex. Between 67% and 75% of AIT cells were highly selective in their response to complex features (such as a particular shape with a particular texture and/or colour, faces or hands) and did not respond to other complex features tested (Tanaka et al., 1991; Kobatake and Tanaka, 1994). There seemed to be an abrupt change in the distribution of the cells between the posterior and anterior part of the IT (Tanaka and Fujita, 1991). Hence, Tanaka concluded that the IT cortex is divided into posterior and anterior IT cortex, and cells in the anterior part require more complex visual stimuli than only orientation of a contour or a certain colour of the object. Therefore, object recognition and its coding advances significantly in the anterior IT, where an integration process of pattern information occurs based on the selectivity of earlier V4 and PIT cells (Kobatake and Tanaka, 1994).

Fujita and colleagues (1992; Tanaka, 1992) suggested that cells in the anterior IT cortex (AIT, TE) are arranged in columns (Fig. 3.7). This would be a convenient method to represent the image of a whole object and is supported by the finding that AIT cells lying closely together seem to have related and similar selectivities of stimuli. This was investigated by vertical and oblique penetrations through area AIT. First the critical feature of a cell in the middle of a penetration was determined. The critical feature was obtained when simplifying the original complex object stimuli by gradually removing features but still having the cell responding maximally. Consequently, a set of stimuli with different activation properties of the cell tested (optimal, suboptimal and ineffective) was produced. These were used to test the responsiveness of the cells along the same penetration as the first recorded cell. This study showed that cells in vertical penetrations had identical or similar stimuli selectivities. It was suggested (Fujita, Tanaka et al., 1992; Tanaka, 1992), that area AIT is built up of many columnar modules in which cells with similar and overlapping stimuli selectivities, cluster together. These cell clusters were on average 0.4 mm in width, which would results in an individual column to be approximately 0.16 mm^2 in square width. It was therefore suggested that there are about 800-1200 columns in the anterior IT of the monkey (Fujita, Tanaka et al., 1992).

Figure 3.7. A schematic diagram of the columnar organisation in area TE. The average size of columns across the cortical surface is 0.4 mm. Cells in one column have similar but slightly different selectivities as indicated by shapes (reproduced from Tanaka, 1992).



Tanaka et al. (1992) proposes that modules in area AIT are selectively activated when an object is viewed. This activity is specific to a subset of modules, which represents a particular partial feature of the object. However, the feature is not coded in the activity of one specific cell, but rather in the pattern activities of many cells in the same module. The whole object is then coded as the activity patterns of a particular subset of modules, whereas the individual features which make up the object, are coded in activities distributed over a population of cells in corresponding modules.

View discrimination: object-centred vs viewer-centred

A further question is addressed: *Is the representation which is produced object-centred, i.e. view-independent* (Marr and Nishihara, 1978; Marr, 1982; Biederman, 1987), *or view-centred, i.e. view-dependent* (Koenderink and van Doorn, 1976; Perrett, Smith et al., 1985; Tarr and Pinker, 1989; Ullman, 1989; Kosslyn, Flynn et al., 1990; Poggio and Edelman, 1990; Seibert and Waxman, 1992a, b; Baker-Cave and Kosslyn, 1993)?

a) Object-centred representations

An object-centred representation allows the object to be recognised when the object or the parts are seen from any view-point (hence is view-independent). A major advantage of this kind of representation is that only a single description of each object's spatial structure would have to be stored in long-term memory, to allow recognition of the object from any (even a previously inexperienced) view.

Marr and Nishihara (1978) argued that the mental description of an object carries information which is invariant to transformations such as rotation of the image (both in the picture plane e.g. upside-down, and around the model axis) or displacement of the object in distance, changing the image retinal size. As described above, Marr (1982) suggests a computational model of object recognition from the retinal image to the object representation stored in long-term memory. In this model, the goal of early visual processing is to establish a 2½ D sketch which represents a viewer-centred description of the object's surfaces. The 2½ D sketch provides information about the description of the surfaces, texture, shading and local motion present in the image with respect to the observer. The primal and the 2½ D sketch are

early representations of the object which need further processing to establish a representation with a co-ordinate system independent of the view observed.

Marr and Nishihara's (1978) model of object recognition is based on an object-centred co-ordinate system in which the position of every generalised cone is described ultimately with respect to the model axis, i.e. the principal axis of the major component. In this hierarchical system, each component axis becomes a major/model axis for its subcomponents. Therefore, only one representation is needed, since everything can be related back to this axis. This structural description of an object is totally view-independent and is based on the spatial arrangements between the object and its parts.

Having an object-centred representation, the question arises how, from different views of the same object, is the same object-centred description accessed? Marr (1982) gives a vague answer to this question, by suggesting that the major axis of an object (which is essential to the recognition of the object), is derived from the image itself. Hence, the same object-centred description will be produced irrespective of the view of the object.

According to Biederman's theory, the view-point is also unimportant for recognition since everything is related to a major axis and can be described by the spatial arrangement of the object's geons (the components) in relation to each other. Furthermore, not only the overall configuration of the object is represented in a object-centred manner, but also each individual geon has its own object-centred description and can be recognised independently. Therefore, like Marr's theory (although not as clearly described), object recognition in Biederman's model is view-independent. This system allows different objects to be made of the same kinds of geons, but with a different spatial arrangement between them, as for example a cup and a bucket. If the "handle-like" geon of a cup is removed and situated at the top of the cylinder one recognises the object as a bucket rather than the cup, despite the same kind and number of geons used to generate these objects. However, if certain characteristics of a geon are occluded, then the recognition task becomes increasingly difficult. Biederman argued that generalisation across different view-points is achieved because each geon and the different spatial arrangement of the geons are "qualitatively coded" and many geons can be used to recognise the same object. Nonetheless, if the key part of a geon

is self-occluded then the description must be view-sensitive. For instance, the pointy part of a cone is occluded from sight when viewing the geon from below and therefore not distinguishable from the bottom view of a cylinder. This argument holds true for the entire object description, in that if a key geon is occluded from sight and recognition is qualitatively diminished, then the description has to be view-sensitive since the object cannot be recognised with equal ease from any view-point.

Experimental evidence for an object-centred representation was provided by Gerhardstein and Biederman (1992). Their experiments were based on priming techniques where recognition times and accuracy of line drawings of object were recorded. Images were first primed by the same object irrespective of the view of the object, i.e. the rotation in depth of the priming object. They found that there is no difference in recognition abilities of an object's view, if the geon structural description of the priming and target images of the same object are equal (which would be no change in view) or highly similar, i.e. less than a 45° difference of rotation in depth. In other words, if the same geons are visible irrespective of different views of the object (only up to 45° difference tested), recognition occurs at the same speed and accuracy. In addition, Gerhardstein and Biederman (1993) tested and supported this finding with nonsense objects composed of a set of geons and constructed on the basis of his model of recognition by components. Nonetheless, one can still argue that with such a model as Biederman's, one still needs some view-dependent information for rotations greater than 45° in depth. Say the whole body is an object and the head was one of its geons. The back of a head does not have eyes, but the front does, hence, the presence of the eyes indicates the view of the head. Nonetheless, if both views (with a rotated difference of 180°) of the head can be recognised with equal speed as a body or give equal priming, then Biederman would be correct in suggesting an object-centred representation.

b) Viewer-centred representation

A representation of an object in a viewer-centred manner relies on view-point-specific information about smaller aggregates of the visual characteristics of the object. There are two main streams of viewer-centred theories: 1) Only one view of the object is stored (Koenderink and van Doorn, 1979; Palmer, Rosch et al., 1981; Jolicoeur,

1985); and 2) Multiple representations are stored in long term memory (Tarr and Pinker, 1989; Edelman and Bülthoff, 1990; Cutzu and Edelman, 1992).

Single view representation

Palmer et al. (1981) suggested that the view of the object which is stored in memory is a 'canonical' view, a view which maximises the most marked and noticeable information about the object. In most cases, Palmer suggested this view to be a ¾ view of the object, i.e. 45° rotated away in depth from the front view. Such a model suggests that there is only one view-dependent representation stored and that the incoming image is transformed in a certain way, such as using mental rotation, to match this stored canonical view. Note, that this is different from an object-centred description, since the stored representation is described by a co-ordinate system based on the viewer.

Multiple view representations

Tarr and Pinker (1989) suggested a model of object recognition where several views of the object are represented in memory. An incoming image is transformed (e.g. mentally rotated) to match the nearest stored view or the canonical view of that object.

Edelman and his colleagues (Edelman et al., 1990; 1991; 1992) also favoured a multiple-view model. They suggested that the stored image of the object is in a multi-representational, two dimensional, view-dependent form. However, they suggested that the transformation process of the incoming image is not by mental rotation, but rather the novel views of an image are interpolated between the stored views.

Multi-representational models suggest the presence of cellular units coding different specific views of an object, such as e.g. front view, back view and the profiles of a head (Perrett, Oram et al., 1991).

Physiological evidence: object-centred vs viewer-centred

Some physiological evidence is related to the issue of determining whether the representation of objects stored in memory is in an object-centred or viewer-centred fashion. Two major papers will be discussed: Hasselmo et al. (1989) argued for an

object-centred representation (but not solely, see below), whereas Perrett et al. (1991) argued that the representation is in a viewer-centred form in the STPa.

Hasselmo et al. (1989) suggested that there are three classes of neurons involved in the representation of objects. Cells coding in 1) an object-centred form; 2) an intermediate form; and 3) a viewer-centred form. They recorded from and tested a total of 37 cells in the superior temporal sulcus (STS) of the monkey with different faces, different identities and different views. Twenty-eight of these cells were tested with 2D video derived images. They tested 13 cells with two different faces and 8 views [6 in the horizontal plane (depth): front view (0°), 45°, 90°, back view (180°), 270°, 315°; and upright and upside-down], and 15 cells were tested with a more extensive range of 21 views.

Hasselmo et al. (1989) analysed the neuronal responses of these 37 face selective cells by using a two-way ANOVA where the group factor was the effect of the identity of the face, and the treatment factor was the angle of view. Eighteen cells responded selectively to faces discriminated between different faces. It was argued, that on the whole, this discrimination was independent of the view of the person's head, showing a constant difference in the neuronal response of the cell to the two faces at all views tested and an object-centred representation for identity was suggested. However, out of these 18 identity specific cells tested, 16 cells were found to be sensitive to identity *and* view. These cells should be classified as viewer-centred (and not object-centred as classified by Hasselmo et al, 1989.), since the cells responded to particular views. Therefore, one might be able to suggest that the coding of identity is in an object-centred manner, but not the coding of view.

The remaining 19 (19/37) face selective cells, responded to all faces equally well. 16/19 of these cells showed different responses to different viewing angles and therefore can also be classified as coding head information in a viewer-centred manner. The study by Hasselmo et al. (1989) further suggested, that most cells are broadly tuned across views, i.e. the neuronal response is relative invariant with respect to the view from which the face was seen. Because of this coexistence of object-centred and viewer-centred cells they suggested that the cortical area represents the processing stage at which three dimensional objects are computed from different views of the object.

Hasselmo et al. (1989) argued that because the front view of the head carries more information than the back view, the majority of the cells respond better to a view where some part of the front of the head was visible, rather than to views were the back of the head was visible. Note though, that there was no clustering of optimal views around the front view or the profile view. However, as Perrett et al. (1991) point out, the statistical analysis used by Hasselmo et al. would not truly represent the optimal view distribution of a cell population. For example, a bimodal cell maximally tuned to the two profiles would have been incorrectly classified as maximally tuned to the front view of the head, since in Hasselmo et al.'s analysis the response of the two profiles would cancel each other out and the reported preferred view would be the front view of the head.

Perrett et al. (1991) who investigated similar issues as Hasselmo's (1989) study, concluded differently. In their study most cells (110/119) showed a viewer-centred manner of coding with an unimodal tuning to one specific view. Four cells (4/119) showed an object-centred manner of coding by responding to all views equally well, and 5 cells (5/119) showed mixed properties, responding to all views better than base line activity (S/A) and control stimuli, but also discriminate between views. Perrett et. al. also tested 56 cells for sensitivity of identity. They found only six cells which showed such discrimination.

In addition, unlike Hasselmo et al. (1989) who argued that the front and profile view is coded by more cells than any other view, Perrett et al. (1991) suggested that all views (front, two profiles and back views) are equally well represented in memory. They found that the majority (52/73, 71%) of view sensitive cells were optimally tuned to visual angles 'on-axis', i.e. within 22.5° either side of one of the characteristic views which were especially the face view (0°), the two profiles (90° and 270°) and to a lesser extent the back view (180°) of the head. The remaining 21/73 cells were optimally tuned to visual angles of the head which were 'off-axis'. Of all the viewer-centred cells tested, most (54/73) cells showed relative narrow tuning for their optimal perspective view.

Perrett et al. (1991) found two classes of viewer-centred cells [cells responding to one or more (but not all) views better than S/A and control objects]. 99/110 cells displayed a unimodal pattern, having one view to which they responded significantly

better than to other views. 11/110 cells showed a bimodal distribution where two non-adjacent views (e.g. the two profiles of the head) were significantly better than other views.

Furthermore, Perrett et al. (1991) estimated the tuning for cells and showed that for most cells, 60° of rotation of the head in depth reduced the magnitude of neuronal response to half of that of the optimal view. Therefore, only four cell populations are needed to code all possible views of the head. Perrett et al. (1991) suggested that the differences between his study and Hasselmo's could be based on the formula applied by Hasselmo et al. (1989) to calculate statistical findings. Perrett and colleagues argued that this formula is insufficient and misleads the results since it is a formula designed for calculating cells with an unimodal neuronal distribution and not cells with a bimodal distribution (see above).

Perrett et al. (1991) classified only a small number of cells (4/119) as object-centred. The authors suggested that these object-centred cells could arise from getting input from different viewer-centred cells (Perrett, Smith et al., 1984; 1985; 1989). This pooling mechanism has been supported by findings that latencies for object-centred cells are longer that latencies for the viewer-centred cells (Perrett, Hietanen et al., 1992). It was, therefore, suggested that more cells coding in an object-centred manner might be found at a later stage of the processing pathway, subsequent to the STPa sampled.

Chapter IV

GENERAL EXPERIMENTAL METHODS: SINGLE UNIT RECORDING IN THE MACAQUE

"How much easier it would be to understand the brain if (single) neurons would light up when they talk to each other."

(Lord Sherrington)

Single cell responses were recorded from six awake macaque monkeys [*Macaca mulatta*, 2 females (J, E) wt. 4-8 kg and 4 males (B, H, D, S) wt. 5-8 kg] with an initial age of approximately three years. The majority of data presented in this thesis were obtained from subjects E and S. This chapter will describe the general experimental methods applied, with specific details of each experiment (such as stimuli used) presented subsequently in the relevant chapters.

Pre-surgical training and fixation task

The subject was first trained to climb voluntarily into a primate chair. The monkey was made familiar with being situated in the chair for an initial period of 30-60 minutes, which later was extended up to three hours, due to an increasing complexity of the training session and tasks to be learned by the monkey. This was achieved by restricting and closely controlling the amount of water and fruit available in the home cage (deprivation up to 12 hrs prior to training/recording session) and rewarding the monkey in the primate chair with liquids and fruits. One of the principal aims was to train the monkey on a colour discrimination task where a green LED light signalled reward and a red LED light instructed the monkey to withhold behavioural responses (see below).

To familiarise the monkey with the discrimination task, several basic applications were employed. The monkey was first trained to feed himself from a small plastic container held by the experimenter. It was not able to see the content of the box and was therefore forced to reach out and tap the box, resulting in the experimenter tilting it to enable the monkey to see the content and reach for the reward. Once the

monkey was able to feed himself, two coloured boxes were introduced. A red container with no food reward and a green container with a small piece of favourable food. The subject was to learn the association of the green coloured box with food reward. When the 90% correct criterion was reached, the coloured boxes were replaced with one white box having a LED (red or green light) attached to it. Now the monkey had to discriminate the colour of the LED before knowing whether the box would contained a food reward or not. Eventually, the floating LED light was distanced further and further away from the box and thereby from the monkey. Once the monkey was able to carry out the discrimination task with the LED placed at a distance of 4 m (the ultimate testing distance), the food reward was changed to liquid reward by introducing lick-tubes. The lick-tubes were attached in front of the monkey's mouth. As soon as the monkey touched the end of the lick-tubes with its tongue, the circuit between the chair and the lick-tubes closed, activating the solenoid driven pump system and resulting in liquid administration through the lick-tubes to the monkey. For this pump system to work, a very tight and secure contact between the monkey's tongue and the lick-tubes was required. Even with this occurring, the solenoid pumps were found to be relatively unreliable in delivering liquid. Since especially during the training phase, the delivery of fruit juice reward is crucial, a new Infusion pump system (Hamilton MicroLab 941, Dundee, Scotland) was introduced. This system was found to be highly reliable in delivering the required quantity of liquid at any given time. In addition, the amount of liquid delivered could be closely controlled. And hence, during the training period, larger quantities of liquid could be delivered decreasing the amount as the monkey became familiar with the task. At first, longer periods of the LED stimulus were presented in order for the subject to learn the task. Subsequently, however, the presentation time was progressively reduced down to one second during which discrimination had to occur for possible reward. The LED stimuli were first presented in an experimenter controlled fashion, having a greater number of green stimuli than red stimuli. Once the monkey correctly carried out the discrimination task, the presentation pattern was gradually changed to a computer controlled pseudo-random order. It was also possible to manually change the LED position on the screen to five different locations (+/- 10° above or below, or +/- 15° to the left or the right of the central LED position).

The final result of the training period was as follows: A short 500 ms signal tone, to indicate the start of a trial, obtained the monkey's attention. Once fixation occurred on the LED light presented on a white wall (at eye level) at a distance of four metres, the monkey's task was to discriminate the colour of the LED. Licking, while the LED light was green, resulted in fruit juice reward. The monkey was to withhold licking in order to avoid delivery of a weak saline solution while the LED was red. Towards the end of the training period the LED colour discrimination task was carried out at a high level of accuracy with fast responses (300-500 ms). This was achieved regardless of whether objects or images of objects were presented simultaneously with the LED light. At this stage, operation procedures for neurophysiological single unit experiments were prepared.

Implant construction

An implant frame was constructed from two stainless steel recording rings (internal diameter of 16 mm, outer diameter of 19 mm, 10 mm deep) fitting a David Kopf hydraulic micro-drive, and two plastic tubes (5 mm internal diameter) used for the restraining of the monkey's head during recording sessions. These tubes were made of either PTFE or Tuffnal, since experience showed that perspex tubes are prone to cracking and breaking. Tuffnal was found to be the strongest and most lasting type of substance especially with the use of ethanol for cleaning purposes. A proportional plan (1:1) of the implant was drawn onto graph paper which was then placed underneath a glass plate on which construction of the implant occurred. On the graph, the orthogonal axes represented the saggital and interaural planes of the monkey's head. Therefore, precise positioning of the recording rings was possible. The rings were centred on the glass plate with the stereotaxic co-ordinates between 12 and 15 mm to the left and the right of the implant mid-line and between 11 and 15 mm anterior to the interaural plane. The two plastic tubes (5 mm diameter) were positioned perpendicular to the mid-line (corresponding to the interaurial plane in the monkey) in front and behind the wells with a distance of approximately 65 mm between them. This layout of the two wells and two tubes was then fixed to hold in position by cementing them down to the glass plate with dental acrylate (Autenal Dental Products Ltd., Harrow, England). Small amounts of liquid dental acrylic was applied to the wells and the glass, making

sure that the acrylic was not placed higher than 2 mm on the walls of the rings allowing good grip for the micro-drive at a later stage. The plastic tubes were fixed in position in a similar way to the glass. By joining up the dental acrylic between all four parts in a cross-like fashion, the base of the implant was formed.

Once the dental acrylic hardened, small quantities of water were used to 'float-off' the implant from the glass. The implant was now readily prepared for the operation and therefore placed with the estimated anterior and posterior co-ordinates in a sterotaxic holder (David Kopf Instruments, California).

Surgery

Twenty-four hours prior to the surgical operation, the subject was taken off liquid and food. On the day of the operation, the monkey was given an intra-muscular injection of 1 ml of atrophine to reduce mucous secretion, 1 ml dupocilin to reduce bacterial infection and sedated with a weight dependent dose of ketamine (Vetalar, 10 mg/kg, Park Davis and Co., Gwent). Once sedated, the head was closely shaven and swabbed with alcohol and iodine. In addition, a drop of paraffin oil (Atropine) was placed in each eye of the monkey to protect and avoid the drying of the eyes. Finally, a barbituate anaesthetic (Sagatal, 60 mg/ml, May and Barker Ltd., Dagenham) was administered intravenously. The usage of a Butterfly vein set (JERUM, Surflo IV catheter; obtained through the district of St. Andrews vet Dr. Wilson) connected by a infusion set (Bioset) to a Hartmann's solution drip (compound sodium lactate, Baxter) was introduced to aid the administration of anaesthetic drugs during the operation. Once the cessation of the gabella reflex took place, the monkey's head was carefully positioned in the David Kopf stereotaxic frame. Under full sterile conditions the operation proceeded. To monitor and regulate the monkey's body state, a breathing counter and a rectal thermometer linked to a heating plate were set up. The depth of anaesthetic level were closely monitored by regularly checking the breathing rate (30-40 per min) and stretch reflexes.

By making an anterior-posterior incision, the skull was exposed and the skin pulled back. Connective tissue covering the area of interest of the skull (equivalent to the size of the implant) was reflected and the skull cleaned. The already fixed implant was now lowered down onto the skull using the David Kopf stereotactic instrument

set-up with the predetermined co-ordinates (which are to be such that the implant is 0° lateral/medial and 0° anterior/posterior). The relative size and position of the recording rings were marked on the skull and the implant was temporarily raised. Using an electric drill (constantly applying distilled water to prevent an increase in bone temperature), the bone underlying the two ring marks were removed, leaving the dura intact and thereby establishing a cranial defect. The implant was once again lowered down onto the exposed skull by the stereotaxic apparatus. To be able to fix the implant securely in position onto the skull, several small elongated holes were drilled into the skull for the insertion of 6-8 small (1.0 cm in length, with two cross-bars of approximately 1.0 cm in length and 0.3 cm wide) stainless steel H-shaped pieces which were subsequently locked into position. In addition, 0.5 cm long stainless steel screws fixed to the skull and stainless steel wire were added around the implant. Finally, a larger mass of liquid dental acrylic was once more applied to the implant frame, the screws, the wires and to the skull to securely fixate the implant.

A small amount of topical antibiotic (PEP, 3% powder) (Intervet Laboratories Ltd., England) was placed onto the exposed dura in the wells and around the implant. The well rings were then covered with plastic aerated caps.

The monkey was returned to its home cage where it regained consciousness after approximately two hours and was allowed to recover from surgery. Re-training sessions started after one to two weeks after the operation until pre-surgical performance in the colour discrimination task was achieved. Restraining of the head was introduced gradually by inserting two metal rods through the plastic tubes of the implant and fixing them to the primate chair.

Recording techniques

Experimental recording was carried out between three and five times a week for a period of up to two years. During this time, the head implant was cleaned regularly with either a very dilute antiseptic solution (Boot's Antiseptic Disinfectant) or Heprun Saline solution. After cleaning the dura, a small amount of PEP was applied to prevent bacterial infection before the head caps were replaced on to the recording wells.

Before each recording session, approximately 3.0 ml of topical anaesthetic (Xylocaine 40 mg/ml, ASTRA Pharmaceuticals Ltd.) was applied to the dura. A David Kopf micro-positioner was fixed to the recording well, allowing a 'tungsten in glass' microelectrode (outer diameter, 0.5 mm, Merrill and Ainsworth, 1972) to be inserted into the temporal cortex through a transdural guide tube with a outer diameter of 1 mm (which was inserted 3-5 mm through the dura). The electrode was first lowered by hand to the approximated depth of the target height, aiming for the top of the anterior part of the Superior Temporal Sulcus (STS) (areas TPO, PGa, TAa of Seltzer and Pandya, 1978) and then for exploration of these areas further lowered by a hydraulic micro-drive.

Data collection

The neuronal activity recorded by the electrode was preamplified (NeuroLog head stage NL 100) and further amplified (NeuroLog NL 104) for the analysis of the signal. The signal was subsequently filtered with a 50 Hz filter in conjunction of a with low-pass (800 Hz) and a high-pass (20 kHz) filter (NeuroLog NL125) and finally displayed on an oscilloscope (Telequipment DM63) and audiomonitor. A voltage window (Digitimer DM130) allowed the digitisation (conversion to a TTL pulse) of selected cell activities. Once the spike activity was converted to digital signals and it was stored in 5 ms time bins on a P.C. compatible computer using CED1401 (Cambridge Electronic Design). To determine whether fixation and eye movements influenced the response pattern of cells, eye movements were recorded with each trial using an infrared reflection system (ACS, modified to allow recording of both horizontal and vertical signals from one eye), and stored with 8 bit accuracy on the computer.

Once a cell was successively isolated by its spike wave form, the cell responses were tested with different stimuli. The LED light and the visual stimuli were presented from behind a shutter with a rise time of less than 15 ms. A large aperture (6.5 cm diameter) electromechanical shutter (Compur), or an alternative (20 cm square) liquid crystal shutter (Screen Print Technology Ltd.) with an inbuild infra-red camera to monitor eye movements was used [Compact medium-high resolution camera (JVC), RS Components; resolution: 500 horizontal by 582 vertical pixel; and a wide angle 16

mm TV-lense, Computar; sensitivity 0.5 lux (at F 1.4)]. The shutter opened for 1.0 sec after a 0.5 sec warning tone which allowed the monkey to prepare fixation of the LED position before the shutter opened. This enabled the monkey to lick several times for multiple juice rewards during the trial period. Visual stimuli (2D) were projected onto the white wall in front of the monkey on which the LED light was positioned, and 3D stimuli were presented in front or to either side of the LED. The test stimuli were interspersed with control stimuli and a no-stimulus condition [the LED alone, to measure spontaneous activity (S/A)].

Data analysis

Since most cells in the STPa respond with a latency of approximately 100 ms (+/- 30 ms; Oram and Perrett, 1992) the magnitude of cell activity on individual trials was assessed over the 1/4 second time period occurring 100-350 ms after stimulus onset. For some cells [with late response onset (>200 ms) or inhibitory responses] a 1/2 second time period (100-600 ms post-stimulus) was used to assess the cell's activity.

Cell responses to the head/body presented in different viewing conditions (see empirical chapters), controls and no stimulus conditions (spontaneous activity, S/A) were compared by using 1-way and 2-way ANOVAs and post-hoc tests [protected least significant difference (PLSD); Snedecor and Cochran, 1980]. N.B. All illustrations presented in this thesis display neuronal responses to all stimuli tested. If in a figure there is no histogram to a test condition, then this is due to absolutely no response activity during any of the five trials carried out. Furthermore, it is pointed out that in the figures of this thesis a schematic representation of a fire extinguisher was used to indicate responses to control objects.

Post stimulus time histograms (PSTH) were plotted for population estimates of time course response analysis. This was done by calculating and plotting the average of the normalised neuronal activities of each of the cells to different stimuli. To ensure that each cell contributes an equal amount to the population estimate, several computational steps were performed. First, the response magnitude of each cell to a particular (test) stimulus was normalised to the magnitude of the difference between the peak response (measured over 5 ms) of the optimal (best) stimulus and the average neuronal response to spontaneous activity (S/A). These individual normalised

responses were then averaged across all cells in 5 ms time bins. Then, these averaged normalised responses were once more normalised in the same manner as described above. This procedure resulted in S/A being set at 0% response and the peak of the most effective stimuli was set at 100% response. Stimulus presentation was set for all cells at point 0 ms. Some of the cell responses had been collected with slightly different time bins (4.8, 5.0 and 5.2 ms) and to be able to combine these responses, a spike destiny function was obtained by convolution of the raw spike train data with a half 50 ms Gaussian trailing backward in time. Furthermore, this half Gaussian was chosen to avoid the response curve to reflect neuronal activity that occurred before the time (spikes) analysed. A threshold point, to determine response onset, was chosen on the basis of pre-stimulus time period, i.e. the threshold was the upper limit of the 95% confidence interval of the pre-stimulus period (S/A) (Oram and Perrett, 1992).

Furthermore, to get a better indication of the cell population response, cumulative response curves (% of the maximum response) were plotted over time. The average population response rate was calculated after normalising each cell's activity (S/A=0, Max Response=100). The cumulative population response is defined as the running total of this average population response from 200 ms pre-stimulus onset. This process allowed the possibility to compare levels of average neuronal response at a given time in a situation where all previous responses above S/A from 200 ms pre-stimulus presentation are taken into account.

Recording sites

After each recording session, frontal and lateral X-ray photographs were taken using a Portable X-Ray Apparatus (Type MX-2) to localise the electrode (exposure times of 4-5 sec at 16KV). All the pictures were taken in the same manner and at the same distance, allowing direct comparison. One set of X-ray photographs were chosen for base comparison. On these, the mid-line of the skull and the interaural plane were marked. In addition, a parallel line was drawn 25 mm above the interaural plane. By superimposing the photographs, the position and angle of each electrode track was determined. This was done by measuring the distance between the electrode's tip and the electrode's intersection of the 25 mm line and the mid-line (measurements taken from the frontal photographs). In addition, the distance from the electrode to the

interaural plane (measurements taken from the lateral photographs) was also recorded from two points, the tip of the electrode and the intersection of the 25 mm line. Furthermore, the final depth of the electrode was noted by measuring the perpendicular distance between the electrode tip and the interaural plane (the depth recorded by the micro-drive was also noted). The 3D-position of each electrode track within the skull was then reconstructed on a PC compatible computer.

In some subjects (B, D, J, H), micro-lesions (10 microamp DC for 30 sec) made at the end of some electrode tracks, subsequently identified using standard histological techniques, allowed reconstruction of the electrode position within the brain. In addition, reference markers were made in all subjects by injection of horseradishperoxidase (HPR) and fluorescent dyes true blue and diamadino yellow at the end of some electrode tracks. In one subject (E), five (two in the right and three in the left hemisphere) dye injections (Indian ink, Winsor and Newton) were made along electrode tracks after perfusion. Particular tracks, where numerous and clear cell classification was achieved, were determined on the basis of data obtained during the 2 years of recording. The same co-ordinates and depth as those tracks were used for marking electrode tracks with black ink. Red ink was used for electrode tracks believed to terminate in the amygdala. A glass syringe (Scientific Glass Engineering PTY. Ltd.) with a fine plastic tube extension containing the coloured dye was used in place of an electrode. Once the acquired position and depth was obtained, frontal and lateral X-rays were taken. Then, very small amounts of dye were introduced as the syringe was slowly, and with great care, retracted.

The fixed brain was cut with a microtome and the sections were photographed onto slide film to allow projection (and therefore magnification) onto a blank wall. The grey matter boundaries were then traced onto paper from where it was transferred to the computer by again tracing the lines with the mouse. This technique allowed accurate reconstruction of the brain, since distortion of the tissue (often occurring during mounting of tissues) did not occur. The micro-lesions and the injected markers were then located by visually inspecting the brain tissue. Finally, micro-drive depth measurements of all investigated cells (each classified by their responses), electrode track co-ordinates obtained from the X-ray photographs and the co-ordinates from

markers and lesions were combined resulting in full reconstruction of the brain areas of interest with the cells investigated.

Perfusion and Histology

Once the last recording session was completed, the monkey was given a sedating dose of ketamine (1.0 ml Vetalar) followed by a lethal dose of barbiturate anaesthetic (2-3 ml intra-perineal). When the animal was in deep anaesthesia, the thorax was cut open to expose the heart. The heart was freed of the covering pericardium, a large bore cannula was inserted into the heart's left ventricle. To reduce the volume of perfusate required, the descending aorta was clamped off, allowing circulation of the perfusate being pumped by the heart to the upper torso only. A tube was inserted to the right atrium to allow (out) flow from the body.

To prepare the body for the fixative perfusion fluid an initial wash was carried out using 5 litres of 0.1M phosphate buffered saline +0.2% $NaNO_3$ (pre-fixative wash) at a temperature of 37°C to remove blood and avoid blood clotting. To support the heart muscle in its pumping role a mechanical centrifugal pump (C16-C, Charles Ansten Pumps Ldt.) was used. Perfusates (pre-fixative and fixative solutions) were situated approximately 1.3 m above the body to allow good and continuous flow of the solutions.

After buffer wash, 5 litres of phosphate buffered fixative composed of 4% paraformaldehyde and 0.5% glutealdehyde were pumped through the upper part of the monkey's body resulting in the stiffening of the neck and face. The head was now readily perfused, separated from the body and the intact skull exposed and put into phosphate buffered fixative.

For orientation and reconstruction purposes, the head was then placed in the stereotaxic frame and eight vertical stereotaxic injections of Indian ink were made at -5, +5, +15 and +25 mm relative to the interaural plane, 10 mm lateral from the mid-line. Five and 15 mm lateral from the mid-line and 10 and 20 mm respectively from the interaural plane, another eight horizontal injections were made. This was done for both hemispheres before the brain was removed from the skull. Craniotomy was performed and the brain set free.

The exposed brain was now infiltrated with a series of sucrose solutions at 10, 20 and 30% at 4°C or alternatively 2% dimethylsulphoxide and glycerol at concentrations of 5, 10 and 20% (Rosene et al., 1986) allowing the possibility of storage without freezing artefacts occurring. To be able to cut the brain with a Cryostat (Bright Instruments Company Ltd., Huntingdon, England), the brain had to be frozen by putting it into isopentane cooled by dry ice to -75° for 45 min. Subsequently, the brain was moved to the Cryostat where it was allowed to equilibrate for 1-2 hrs. During the process of brain sectioning, the microtome was kept at a temperature of -30°C, and sections of 25 or 50 microns thick were cut and selected at an interval 0.25 or 0.5 mm. The sections were free floating and mounted onto gelatine treated slides. They were then stained using Cresyl Violet for Nissl substance. In addition, every 0.25 mm the coronal brain surface was photographed with a colour slide film aiding in the reconstruction of the brain (see earlier).

Chapter V

THE ROLE OF COMPONENT PARTS IN OBJECT RECOGNITION

Two important issues have been raised in the context of object recognition. These concern (a) the role of the object's components and (b) the frame of reference used for specifying the spatial relationship between components of that object. Two types of frames of reference have been considered: *object-centred* descriptions relate an object's component parts to a framework based on the object itself; whereas *viewer-centred* descriptions relate an object's component parts to a framework based on the observer. This chapter presents a physiological study of the role of the component parts in the coding of one type of complex object, the body. Chapter VI then goes on discussing the coding of perspective view of the body and its component parts.

INTRODUCTION

Several theories of object recognition have suggested that the processing of an object's parts play an important role in the initial stages of recognition (Marr and Nishihara, 1978; Marr, 1982; Biederman, 1987). Recently, however, Baker-Cave and Kosslyn (1993) have suggested that even though coding of parts is important for recognition, parts are processed only *after* the processing of the configuration of the whole object (for a more detailed review of these models see chapter III).

To isolate the component parts, the image of the object may be segmented at regions of sharp concavity of the external boundary (Marr and Nishihara, 1978; Hoffman and Richards, 1984; Biederman, 1987). Each of the resulting image regions can then be treated as if it corresponded to a volumetric primitive (i.e. a 3D component). Theories differ as to the type of volumetric primitives thought to be used in object recognition. Marr suggested that objects could be built up from a set of generalised cones (Binford, 1971; Marr and Nishihara, 1978). By contrast, Biederman postulates a more extensive set of 36 types of cone-components called geons (Biederman, 1987).

Marr has suggested that the position, 3D orientation and size of each cone component of an object is described in relation to the object's principal axis (the

longest axis of the object; Marr and Nishihara, 1978). Biederman (1987) relates geon components to other geons rather than to the object's principle axis. In Biederman's model the position and orientation of each geon is specified relative to other geons in qualitative terms (e.g. geon 1 to the side of geon 2 and joined at the expanded end). Both these object-centred models suggest that the entire 3D object description can be accessed from the sight of key component parts. Thus, even when parts of an object are occluded from sight, recognition can occur on the basis of the remaining visible geons or cone components. From both Marr's and Biederman's theories, one might expect to find cellular units late in the visual pathway which respond selectively to one type of object and are activated by the sight of any major part of that object (a major part of an object can include several geons).

The processes of accessing the entire 3D description from the sight of one 3D part is similar to the 'completion' property of parallel distributed processing (PDP) networks (Kohoen et al., 1981). Networks trained on a 2D image and tested with any large part of the trained image will settle into a pattern of activity within the network, which is equivalent to activity pattern generated by the complete training image.

Cellular sensitivity to objects and their parts

As pointed out in chapter II, inferotemporal cortex (IT) is believed to play a central role in visual pattern and object recognition (Gross, 1973; Ungerleider and Mishkin, 1982). Most physiological studies of this area have been concerned with the coding of geometrical features (such as bars, circular areas or fourier descriptors; Gross, Rocha-Miranda et al., 1972; Schwartz, Desimone et al., 1983; Desimone, Albright et al., 1984; Tanaka and Fujita, 1991; Fujita, Tanaka et al., 1992; Komatsu and Ideura, 1993) which may occur in several different objects.

Studies in IT and neighbouring cortex within the superior temporal sulcus (STS) have also revealed populations of cells with greater visual selectivity that respond preferentially to particular complex biologically important stimuli such as hands and faces (Gross, Bender et al., 1969; 1972; Perrett, Rolls et al., 1982; Kendrick and Baldwin, 1987; 1989; Desimone et al., 1990). These cells offer an opportunity to determine the importance of an object's component parts in the processing of the entire

form of the object. Hence these cells can be used to investigate the neurobiological validity of psychological and computational models of object recognition.

Early studies focused on the importance of facial components for cell responses to the whole face. Different cells were found to be selective for different regions of the face (some tuned to the eyes, others to the mouth) (Perrett, Rolls et al., 1982). Though not systematically studied, it was also noted that most cells responded to several regions tested in isolation (Desimone, et al., 1984; Perrett, et al., 1982). Facial characteristics, particularly when seen from the front, are defined by differences in pigmentation and surface structure. The parts cannot be segmented from points of maximum concavity in the external contour (silhouette). Hence, models such as those of Marr and Biederman which embody segmentation at concave points in the external boundary may be less applicable to the processing of the internal structure of the face.

In the present study the investigation of the role of components in the processing of whole objects is extended by comparing the response of cells to the entire static body and to two major parts: the head with the body occluded from sight and the body with the head occluded from sight (preliminary results of have been reported previously; Perrett et al., 1993; Wachsmuth, Oram et al., 1993).

As pointed out in chapter II, cells with different types of selectivity have been found in the polysensory area of the anterior STS. This area includes cells selectively responsive to the whole body in motion (walking). These cells do not respond to a wide variety of control stimuli when moving in different directions, but respond to a whole body stimuli (often with a preferred view) walking in a particular direction (Oram et al., 1993; Oram and Perrett, 1994a, b, c). Furthermore, these form and motion sensitive cells do not respond to the same optimal stimulus presented stationary (static stimuli). Part of the present study investigated whether the coding of component parts for cells responsive to the static body is carried out in a similar manner for cells selectively responsive to the whole body in *motion*. Cells selectively responsive to the body in motion were tested for their response pattern to the whole body and component parts (head alone, body alone, legs alone) presented in motion (optimal direction of whole body movement). Since the aim of this investigation was to compare the static and the motion condition of these whole body sensitive cells, only walking movements were investigated. Other cells responsive to movements such as

rotation around body axis, up-down movements, etc. were not the subject of the current study.

METHODS

General single unit recording methods in the macaque monkey brain were applied which have been described earlier (see Chapter IV).

Static visual stimuli

Neuronal responses were recorded for both static real 3D objects and 2D objects (video disk images and slides). The visual stimuli tested were the a) whole human body (head and body), b) head alone, and c) body alone (defined here as torso, arms and legs; see for example Fig. 5.1). The posture in all cases was bipedal. The human body stimuli were photographed onto a 200 ASA ectocrome slide film, with the person standing against a light grey background. Slides with occluded body parts were produced by first taking the whole body slide and then occluding the unwanted parts with black tape.

Alternatively the stimuli were filmed with a video camera (JVC BY-110E), recorded on 3/4 inch U-Matic videotape, edited on a JVC editing suite (control unit RM-88U) and transferred on a laser video disc (RLV Mk II, Optical Disc Corp.). The video stimuli were then replayed with a video disc player (Philips VP406 LaserVision Disc Drive) and projected onto the display screen (using a Sony colour video projector VPH-1041QM).

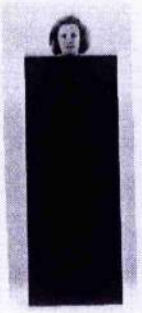
Where 2D stimuli did not activate cells, 3D stimuli were used. 3D head alone and body alone stimuli were created by occluding the unwanted parts with a large sheet of black cardboard or curtain material.

For most cells, before neuronal responses to the visibility of component parts were collected, multiple views of the entire body stimuli were tested to define the optimal view of the object (see chapter VI): front view (0°), left profile (90°), back (180°), and right profile (270°). In addition to these, four intermediate views (45°, 135°, 225° and 315°) were tested for some cells (see chapter VI). For each cell, a number of different control stimuli were tested (2D and 3D). These included complex 3D objects of different sizes, shapes and textures (lab coats, chairs, etc.), simple 2D

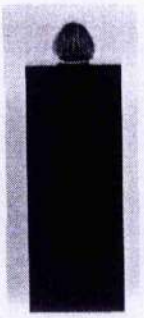
Figure 5.1. Examples of stimuli used for testing: Whole Body, Head alone, Body alone, Legs alone stimuli presented in different views.

Whole body view 315°

Whole Body view 180°



Head alone view 0°



Head alone view 180°

Body alone view 0°

Body alone view 180°

Legs alone view 0°

Legs alone view 180°

geometrical shapes (bars, spots and gratings) and simple 3D forms (cylinders, balls, boxes, etc.).

Testing methods

Every cell from which neuronal activity was recorded was first tested in an exploratory way by presenting a series of static and moving 3D objects (including bodies), and tactile and auditory stimuli. Cells found to be responsive to static views of the whole body were then further investigated for sensitivity to body parts. After identifying the optimal stimulus with a 3D whole body, further testing was carried out with 2D stimuli. Strimuli were presented in a computer-controlled pseudo-random order. The cell was then tested with at least five trials of each of the whole body, its component parts, and different control objects. Testing was performed, where possible, with stimuli of the (cell's) preferred view (see Chapter VI), control stimuli and no stimulus. The views used for testing components were the same as used for testing the whole body.

Testing visual stimuli in motion

Some cells were found to be selectively responsive to the sight of whole bodies in motion (walking), but not to the same visual image when static. These cells were tested in the same fashion as cells maximally responsive to static images (see above) thought the stimuli used were real 3D presentations of the experimenter walking, both forewards and backwards in different directions (towards and away, and moving from the left to the right, or the right to the left, of the monkey). The responses of these cells to the whole body when presented in its optimal view (see chapter VI) and moving in its optimal direction (see Oram and Perrett, 1994b, c) were significantly greater than responses to control stimuli moving in the same direction or base line activity (S/A). These cells were further tested with stimuli of the whole body and its component parts presented in the optimal view and moving in the optimal direction. That is, six stimulus conditions were used. Each condition was presented at least five times in a pseudo-random fashion: 1) Whole body in motion; 2) Head alone [body (torso and limbs, including legs) occluded] in motion; 3) Body alone (head occluded) in motion; 4) Legs alone (head, torso and arms occluded) in motion; 5) moving control objects

(see above) and 6) a no stimulus condition (S/A). The direction of movement and the view was the same for each stimulus.

All stimuli were presented in 3D with the appropriate part of the body occluded. This was done by holding a large card board in front of the component part and walking with the board in place in the optimal direction. The experimental set-up, testing method and data analysis was otherwise the same as carried out for cells tested with static stimuli.

Single cell data analysis

Cell responses to the whole body and components (head alone; body alone and legs only if in motion), controls and S/A were compared by using 1-way ANOVA and post-hoc tests [protected least significant difference (PLSD), Snedecor and Cochran, 1980]. All cells tested were classified on the basis of their neuronal response pattern to the different visual stimuli tested. They could either respond to none, one, or two body parts.

In some cases, post-stimulus time rastergrams were plotted to display neuronal activity during five single trials. These illustrations clearly indicate the pattern of response latencies of individual trials for a cell.

Population data analysis

An alternative way of classifying the same data into four cell populations than by looking at the individual cell's neuronal responses to body parts in comparison to S/A and control objects, was applied by normalising all cell responses to allow direct comparison between cells. This was done by presenting the different individual cell data in form of a scatter plot. For direct comparison between different cells, neuronal responses to different visual stimuli were normalised for each cell by applying the following formula:

NORMALISED RESPONSE TO THE HEAD = (H - S/A) / (BEST RESPONSE - S/A)
NORMALISED RESPONSE TO THE BODY = (B - S/A) / (BEST RESPONSE - S/A)

where H = average neuronal response to the head alone stimuli
B = average neuronal response to the body alone stimuli
S/A = average neuronal response to the no stimulus condition

BEST RESPONSE = average response to the (optimal) stimuli which resulted in the largest neuronal firing rate difference from S/A and control objects compared to other visual stimuli. Note that the best response can either be the average response to an effective component part or the whole body stimulus depending on the relative firing rates.

Furthermore, post-stimulus time histograms (PSTH) and cumulative response curves were plotted for population estimates of time course responses (see data analysis in general method chapter IV).

RESULTS

For five subjects[2], a total of 7287 cells were screened for neuronal responses. Of these, 23% (1691/7287) were found to be visually responsive, including cells responsive to moving, or static visual stimuli (see e.g. Bruce, Desimone et al., 1981; Oram, Perrett et al., 1993) and visual general cells (where no apparent specific visual stimuli drives the cell). Of the visually responsive cells in the cortex of the anterior STS a total of 77 cells which were found to be responsive to the static whole body (i.e. have significantly greater response to the whole body than to control objects and S/A) were tested for selectivity to static component parts of the body. In addition, 25 cells selectively responsive to the whole body in motion were tested with component parts of the body moving in the optimal direction. All the cells included in this study did *not* selectively respond to a wide variety of different control objects tested.

For 66/77 of the cells selectively responsive to static faces/bodies, responses were measured for two body parts [head alone; body alone (torso, arms and legs)] and the whole body (head and body). The remaining 11/77 cells were tested for the whole body and only one part, and were therefore not included in the analysis of coding component parts, but included in the view discrimination study (see chapter VI). Cells were categorised on the basis of their response to the two parts: they could either respond to none, one or two parts.

The optimal stimulus for a given cell was defined as the stimuli causing maximal change in the firing rate relative to baseline activity (S/A). For the majority of

[2] It is pointed out that I have only been involved in training and recording from two of these subjects (E and S). Since I designed the experimental tests applied, the majority of data described was obtained from these two subjects. Some data was obtained from previous monkeys, where the tests were carried out for identifying the cell responses, though this data was not analysed or described elsewhere.

cells (73/77), optimal stimuli produced excitatory responses. Four cells gave inhibitory responses to optimal stimuli. For clarity of explanation, 'greater' response is defined as greatest change from S/A (whether excitatory or inhibitory).

Coding of parts

29 cells (44%) responded to *only one of the two* component body parts when tested in isolation.

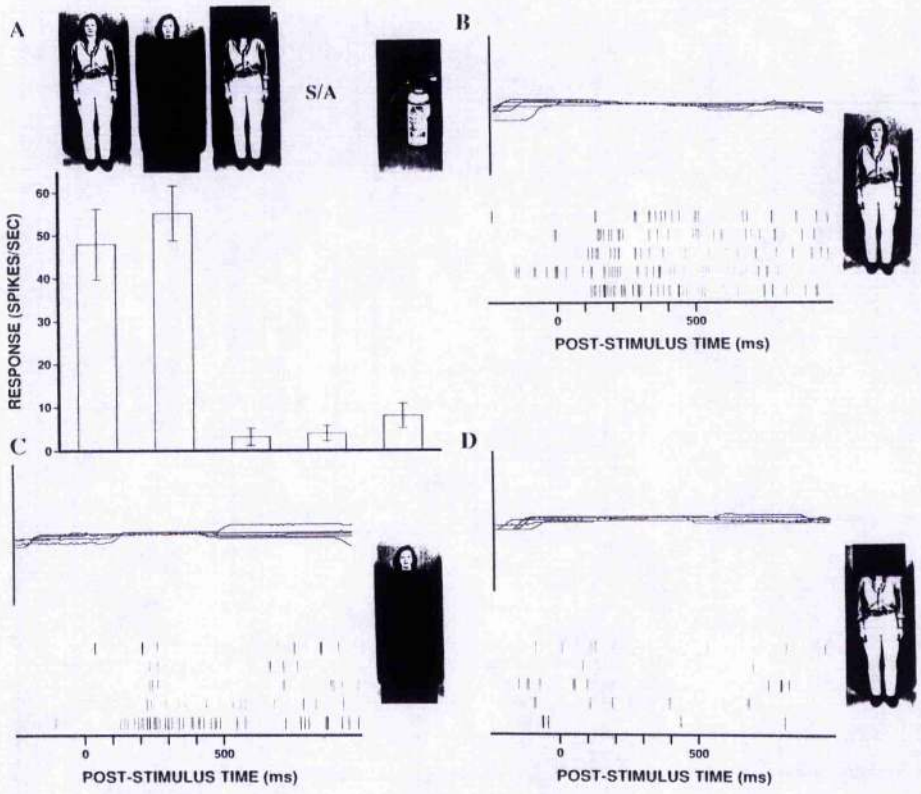
a) Cells only responsive to the head

For 23 cells (35% of the 66 cells tested) the response to the sight of the head alone (with the rest of the body occluded from sight) was significantly greater than that to controls and to S/A. Additionally for these cells, the body alone (with the head occluded from sight) did not produce a response that was significantly different from controls or S/A. This pattern of response is shown in Fig. 5.2a. For this cell the presence of the head was both necessary and sufficient to account for the response to the whole body.

Figs. 5.2b-d also display responses of the same cell recorded on the five individual trials with the whole body, head alone and body alone stimuli. Responses to the whole body (Fig. 5.2b) occurred at approximately 120 ms after stimulus onset, with an initial transient burst (lasting approximately 350 ms) following a response decline, though remaining substantially greater than the pre-stimulus activity typical for the cells found in STS (Oram and Perrett, 1992). Fig. 5.2c shows responses of similar latency and time course of activity during trials when the head alone was presented. When the body was presented in isolation (Fig. 5.2d), however, there was no change in cell activity in comparison to the pre-stimulus period. Thus, the body presented without the head was an ineffective stimulus for this particular cell.

In addition, Figs. 5.2b-d indicated that for all stimulus types the position of the eyes was held constant (with +/- 5 degrees) during the time of neuronal response analysis (100-350 ms). Thus, differences in cell responsiveness for the different types of visual stimuli (effective or ineffective) were not due to different patterns of fixation.

Figure 5.2. Neuronal responses of a cell only responsive to the head. (a) Histogram: Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) to the whole body, its components (view 0), spontaneous activity (S/A) and control stimuli are illustrated for one cell (E29_33.56). The responses to the whole body and the head only stimuli were not significantly different ($p>0.05$), whereas both these stimuli produced a significantly higher neuronal response than activity rates produced by the body alone stimulus, S/A, or control objects ($p<0.005$ each comparison). ANOVA: $F(7,39)=26.1$ $p<0.0005$. **(b-d) Rastergram** displays of responses of one cell (E29_33.56) on five trials for each of the stimulus conditions. Each trial (originally in pseudo-random order) is represented by a single row of ticks, each tick indicates one action potential. Pre-stimulus time is given at the figure base. Horizontal eye movements are shown for each trial in the upper sections. Scale = +/- 100 degrees for horizontal (eye movements were recorded over a range of positions +/-20 degrees from straight ahead). The responses to the whole body (left panel) and the head tested in isolation (left centre panel) were not significantly different ($p>0.05$), whereas both these stimuli produced a significantly higher neuronal response than activity rates produced by body alone (right centre panel) or control objects (right panel) ($p<0.005$ each comparison). ANOVA: $F(7,39)=26.1$ $p<0.0005$.



b) Cells only responsive to the body (torso and limbs)

9% (6/66) of the cells showed a response to the body tested in isolation, which was significantly different to the response to control objects and S/A (e.g. Fig. 5.3). For these cells the responses to head alone did not differ from S/A or the response to control objects.

The 29 cells described in this section were responsive to one of the two body parts tested in isolation. This classification could include cells which show hidden sensitivity to multiple body parts. Thus even though a cell responds only to one component when tested in isolation, the other component may influence the response to the whole object. For 16 of these cells, there was no significant difference between the response to the effective component part and response to the whole body. For these cells the response to one part was necessary and sufficient to account for the response to the whole body. Whereas the remaining 13 cells showed a significant difference between the responses to the effective component part and the whole body stimuli. In some cases (7/29 cells), the effective body part did not trigger an average response which was as large as the response to the entire body visible (Fig. 5.4). On the other hand, for 6/29 cells the response to the effective body part presented in isolation resulted in significantly greater response than the response to the entire body (Fig. 5.5).

Coding the entire body

Cells only responsive to whole body

20% (13/66) of the cells showed a response depending on the visibility of the whole body. These cells responded to *neither* component part when tested in isolation. For these cells the whole body stimulus produced a response significantly different from that to either of the body components (head alone and body alone), control objects and S/A. The responses to the components individually did not show any statistical difference from responses to controls or S/A (e.g. Fig. 5.6).

Cells responsive to multiple body parts

The remaining 36% (24/66) of the cells studied showed responses to *both* components of the body when tested in isolation. The responses to components and the

Figure 5.3. Neuronal responses of a cell only responsive to the body (head occluded). Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) to the whole body, its components (view 0), spontaneous activity (S/A) and control stimuli are illustrated for one cell (B77_21.54). The whole body and the body alone stimuli produced no significant difference in the cells activity rate ($p>0.05$), whereas both these stimuli produced significantly higher neuronal activity than activity rates produced by head alone stimulus, S/A, or control object stimuli ($p<0.005$ each comparison). ANOVA: $F(4,39)=11.4$ $p<0.0005$.

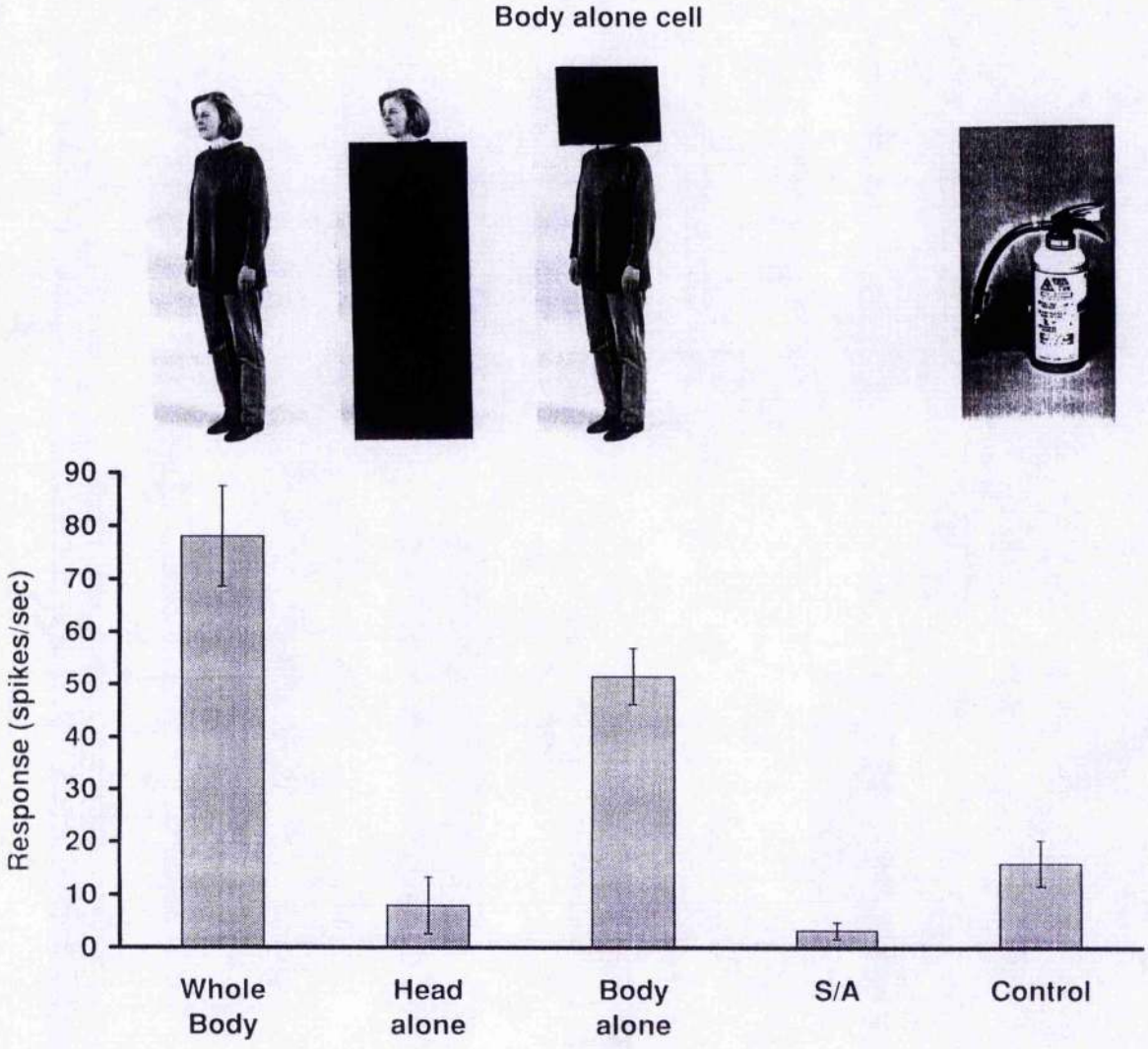


Figure 5.4. Neuronal responses of a cell only responsive to the head (body occluded). Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) to the whole body, its components (view 0), spontaneous activity (S/A) and control stimuli are illustrated for one cell (D29_29.82R). The responses to the whole body and the head only stimuli were not significantly different ($p>0.05$), whereas both these stimuli produced a significantly higher neuronal response than activity rates produced by the body alone stimulus, S/A, or control objects ($p<0.05$ each comparison). ANOVA: $F(4,23)=33.66$ $p<0.0005$.

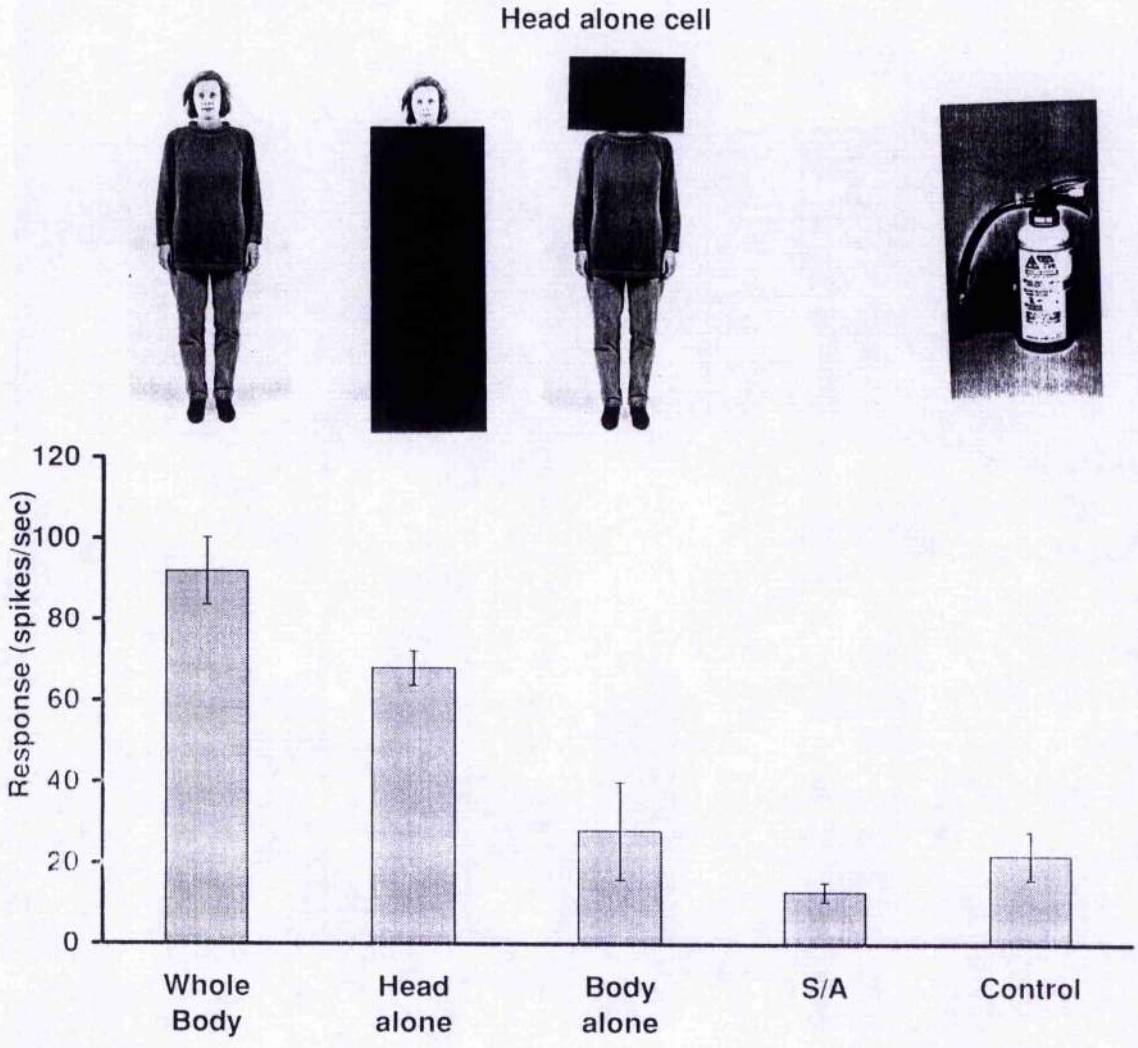


Figure 5.5. Neuronal responses of a cell only responsive to the head. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) to the whole body, its components (view 0), spontaneous activity (S/A) and control stimuli are illustrated for one cell (E82_37.00L). The responses to the whole body and the head only stimuli were significantly different ($p>0.05$) whereas both these stimuli produced a significantly higher neuronal response than activity rates produced by the body alone stimulus, S/A, or control objects ($p<0.05$ each comparison). ANOVA: $F(4,20)=21.2$ $p<0.0005$.

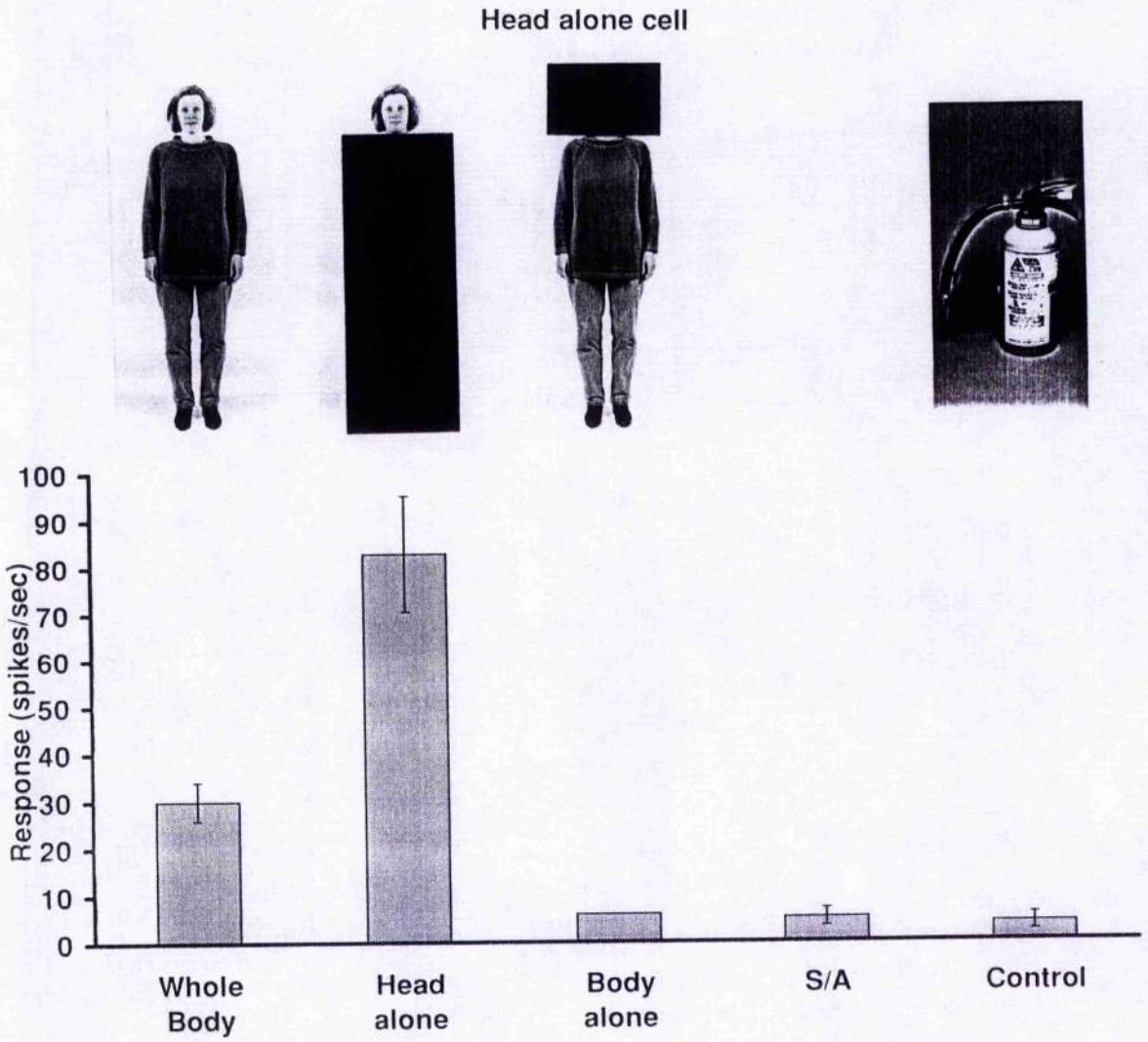
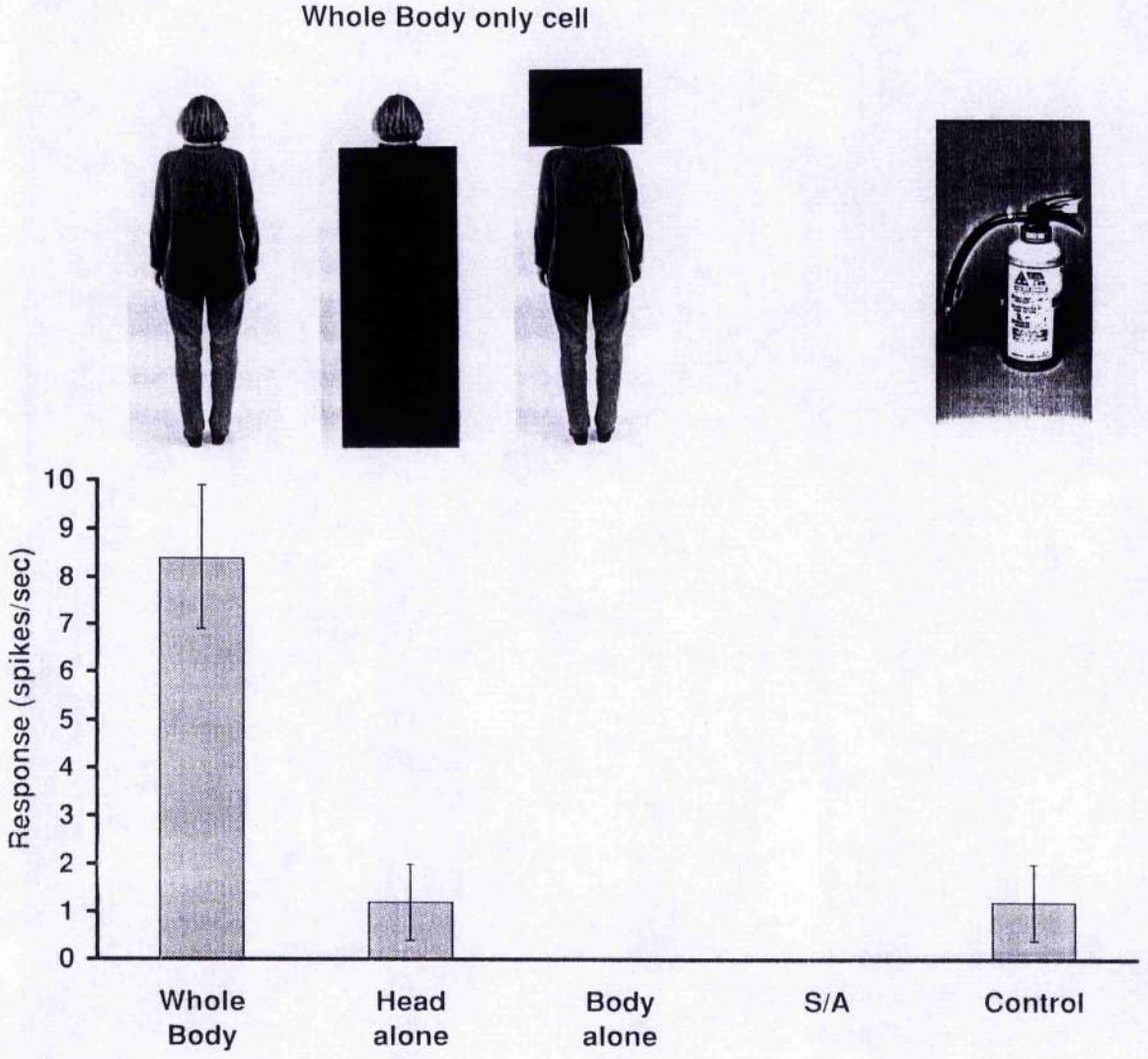


Figure 5.6. Neuronal responses of a cell only responsive to the whole body. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) to the whole body, head alone, body alone stimuli (view 180) and spontaneous activity (S/A) are illustrated for one cell (D107_35.41). The whole body stimuli gave a significantly higher response than the other stimuli tested ($p<0.0005$ each comparison). The head tested in isolation and the body tested in isolation did not show any significant difference from the S/A ($p>0.05$ each comparison). ANOVA: $F(3,23)=35.3$ $p<0.0005$.



whole body were significantly different from the responses to control objects and S/A. Typically the response to the whole body was greater than the response to the individual components (e.g. Fig. 5.7).

An alternative way of classifying the same single cell data into four cell populations was applied by presenting the individual cell data in form of a scatter plot, where the neuronal responses of individual cells were normalised for direct comparison (see Fig. 5.8). Clear clustering of cell types can be observed.

Population estimates of time course responses

To analyse the population's neuronal response to particular stimuli, post stimulus time histograms (PSTHs) and cumulative response curves were produced. For a given cell population, the neuronal responses to particular stimuli were plotted as a percentage of the optimal (maximal) response over time. At point 0 ms, stimulus presentation occurred.

Cell population coding one component part (Head alone)

An example of a cell population which is selectively responsive to only one component part, in this case the head presented in isolation, is given. 14 cells which have been tested for the entire body and two component parts were included in the population analysis (see Fig. 5.9). Neuronal population response to the effective stimuli (whole body and head presented in isolation in the cell's optimal view) occurs at a similar time after stimulus presentation. It is interesting to note that the average population response to the whole body stimulus peaks in the PSTH at a lower rate [at 88% of the maximal (head alone stimuli) response]. However, the response to the whole body is sustained at a higher rate for a longer time than the response to the effective body part. Nonetheless, up to 350 ms after stimulus presentation, the head alone cell population will respond equally well to the head presented in isolation as to the entire body stimulus. After this point, at a cumulative response greater than approximately 32% of the maximum response, the cell population will respond faster to the entire body present than only to the (effective) component part.

Figure 5.7. Neuronal responses of a cell responsive to multiple parts. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) to the whole body, its components (view 270), control stimuli and spontaneous activity (S/A) are illustrated for one cell (D30_27.72). The cell responded more to the whole body stimuli ($p<0.05$ each comparison) than to any other stimuli. The cell also responded significantly more to either of the body regions tested in isolation (head alone; body alone stimuli) than to S/A or control objects ($p<0.005$). ANOVA: $F(4,18)=14.1$ $p<0.0005$.

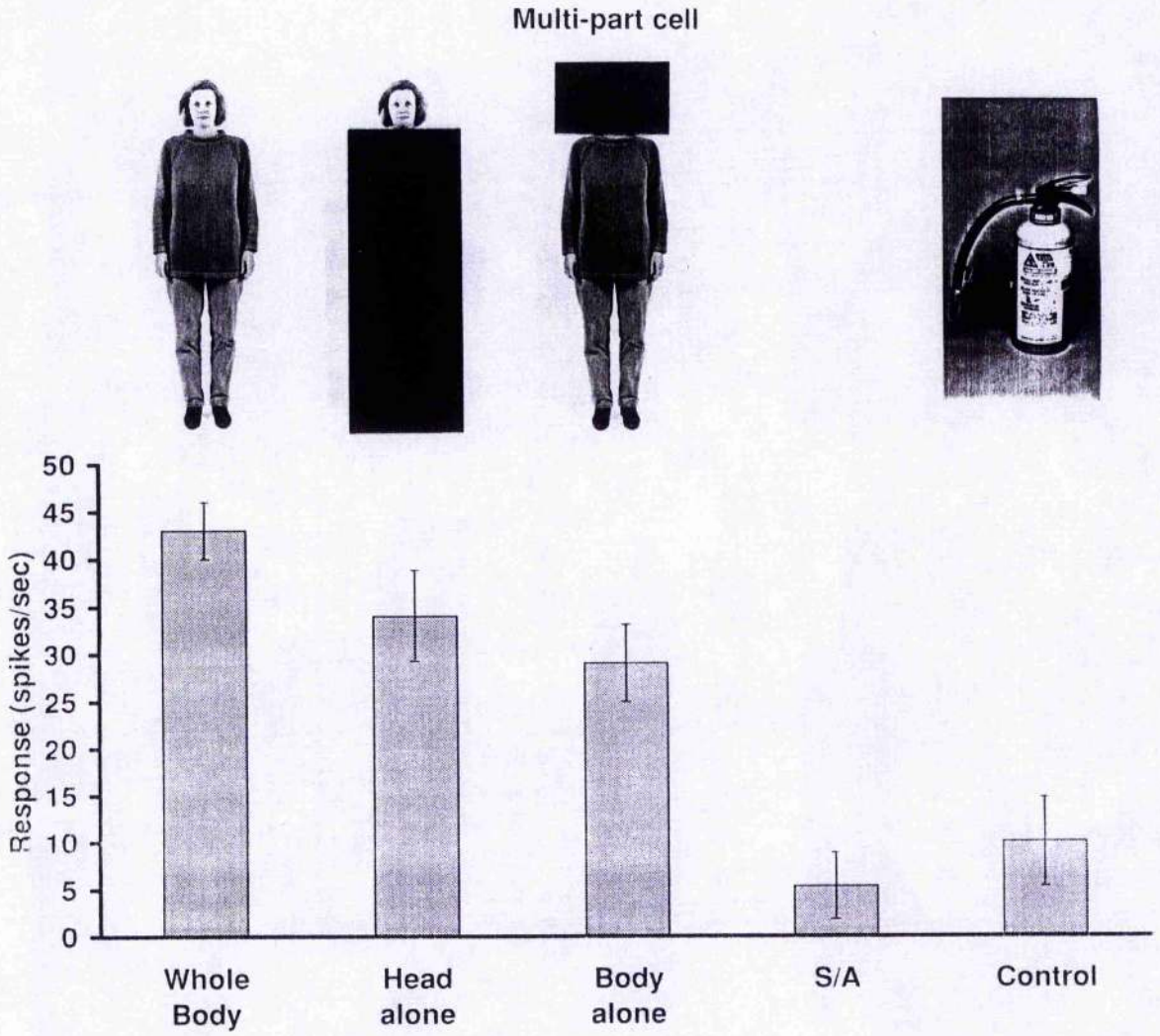


Figure 5.8. Population classification: Scatter plot, indicating four cell populations (n=66 cells). Normalised cell responses to Body alone stimuli and Head alone stimuli are plotted for each cell (see text). Whole only cells = filled triangles; Head alone cells = open circles; Body alone cells = open diamonds; Multi-part cells = filled squares (as classified using one-way ANOVA analysis of responses).

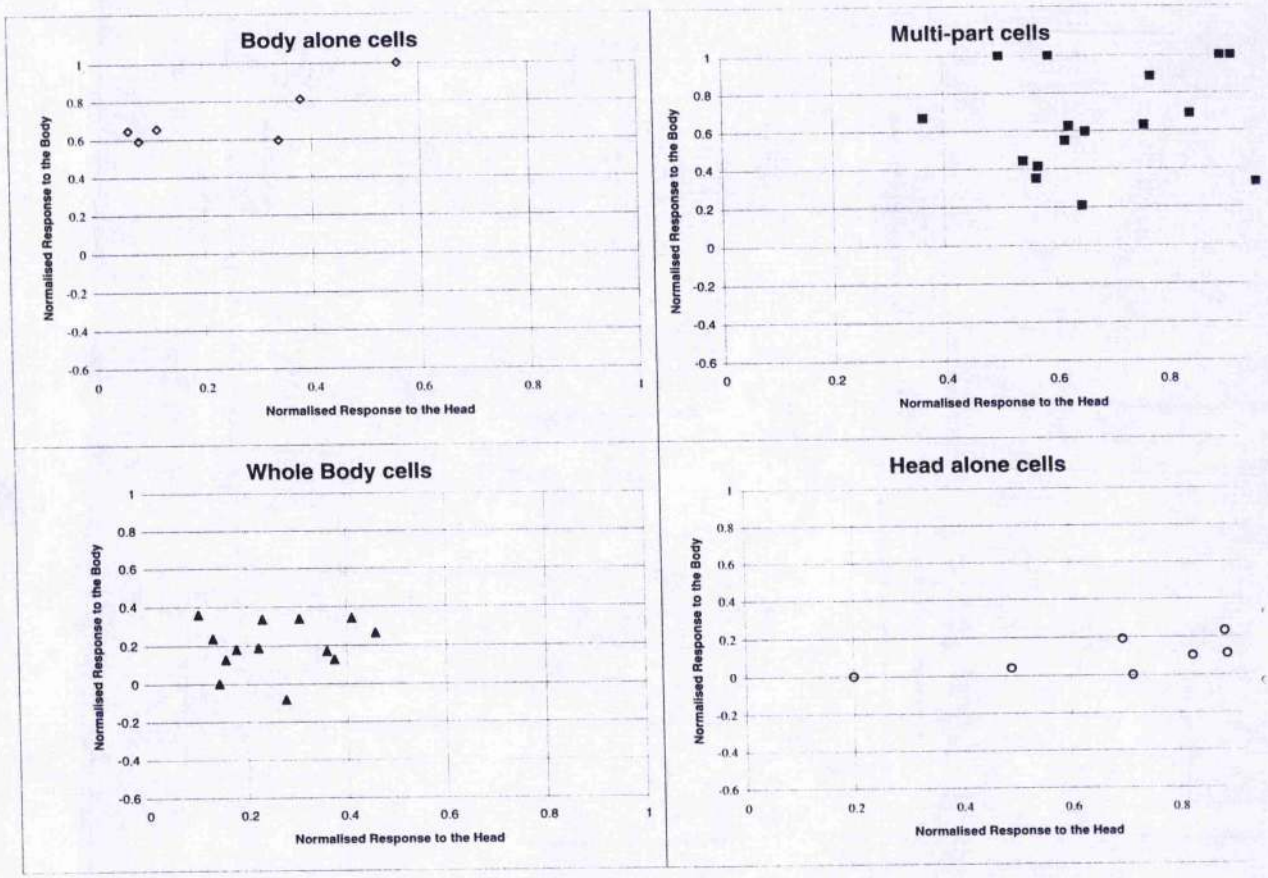
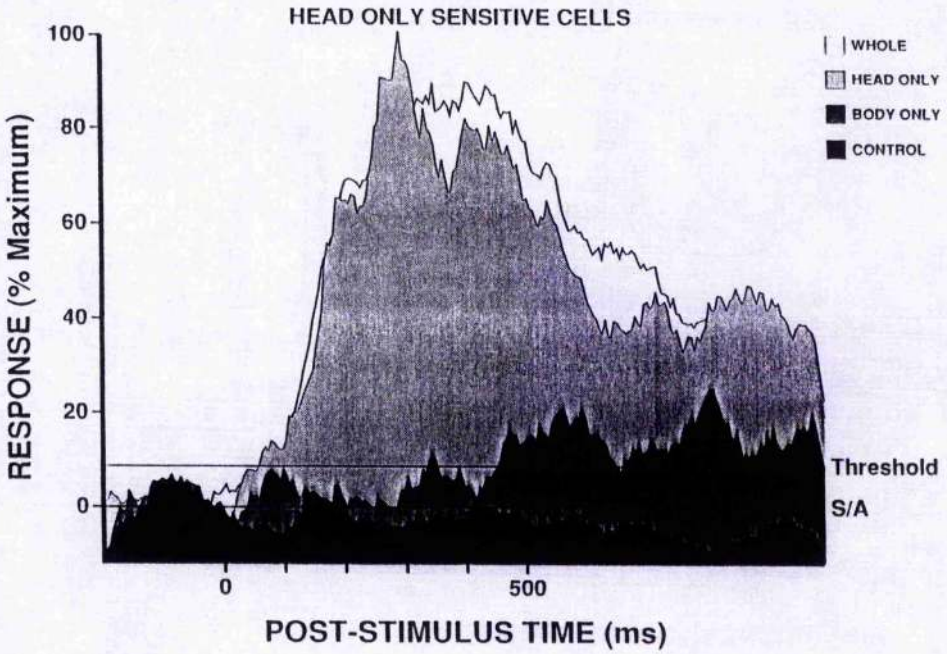
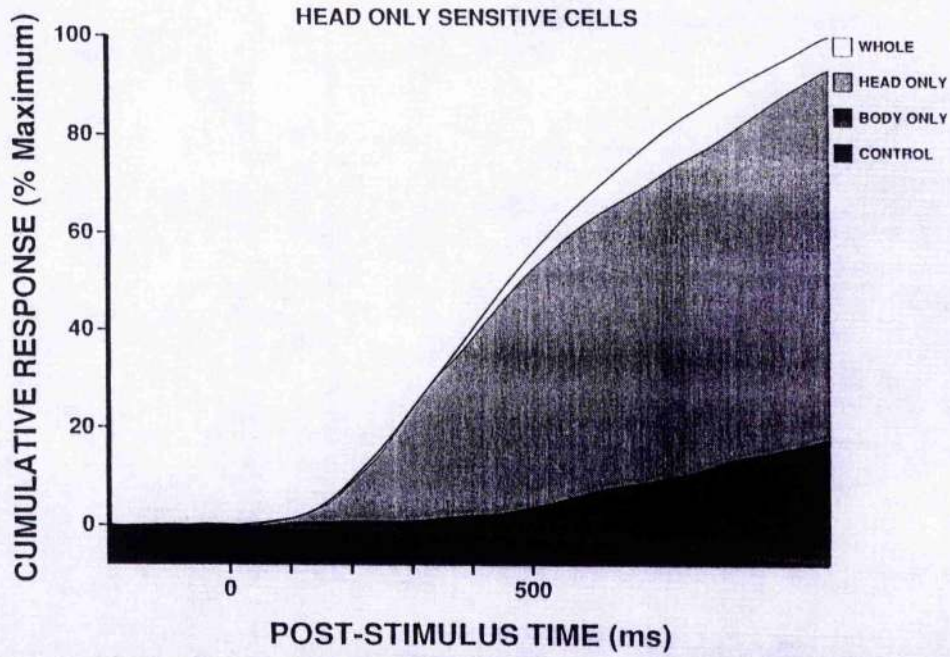


Figure 5.9. Cell population response to the Whole Body stimulus, component part (Head alone, Body alone) stimuli and control objects. a) Combined responses of 14 cells **selectively responsive to the static head presented in isolation** to different stimuli: The clear area displays the population response to the whole body stimulus, medium grey area to the head alone stimulus, dark grey area to the body alone stimulus and the black area population response to various control objects. Firing rate is expressed as a percentage of the peak response to the best (in this case the head presented in isolation) stimulus. Stimulus presentation occurs at point 0 ms. b) Cumulative Response curve of head alone population estimate. The colour coding for different stimuli responses is identical to (a). Firing rate is expressed as a percentage of the maximum response to the best (in this case the head alone) stimulus. Stimulus presentation occurs at point 0 ms. The average population response rate was calculated after normalising each cell's activity (S/A=0, Max Response=100). The cumulative population response is defined as the running total of this average population response from 200 ms pre-stimulus onset.





Furthermore, it is interesting to note the very slight increase of neuronal population activity to the ineffective (body presented in isolation) stimulus after approximately 450 ms post-stimulus presentation. The decrease in cumulative response to control objects is due to the neuronal activity to controls being lower than the activity to S/A.

Cell population coding multiple parts

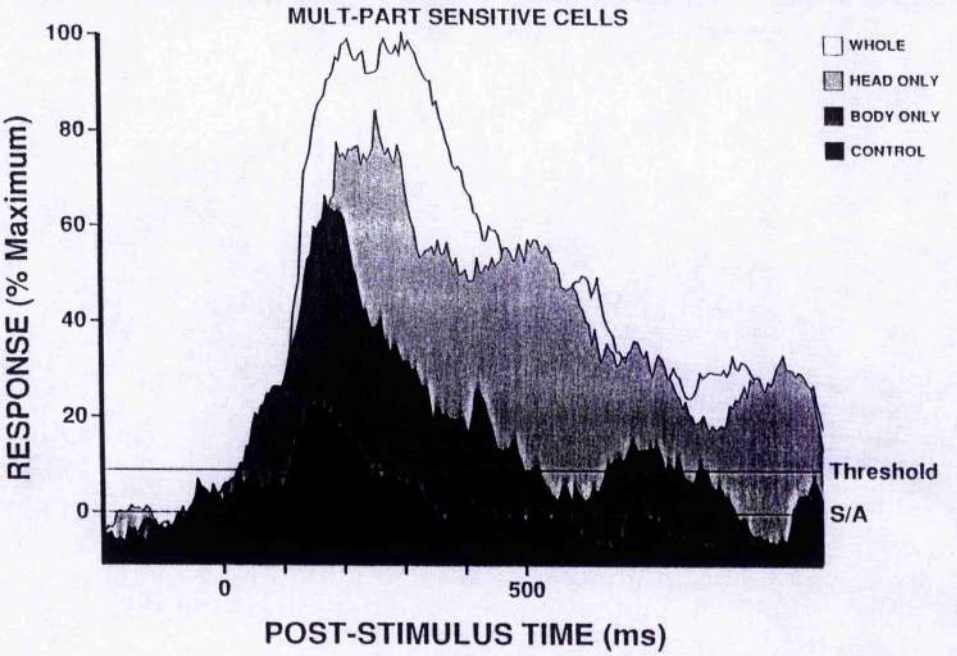
12 cells which were selectively responsive to the whole body and the body parts stimuli were included in this population analysis. A PSTH with all combined responses to different stimuli was drawn (Fig. 5.10a). The multi-part cell population starts to respond to different whole body or body part stimuli with a latency of approximately 90 ms after stimulus presentation (N. B. the very early response onset to the body alone stimuli is caused by a few cells with exceptionally early response onset, causing the average latency to be very early). From the graph one can clearly see that the population response to the whole body has the fastest rise time and peaks the highest. The population responses to the head alone presented in isolation peaks second highest at approximately 85% of the maximum response. The population response to the body component presented in isolation, on the other hand, only rises up to 65% of the maximum response and decreases much faster.

The cumulative response curve of all cells selectively responsive to the whole body and the body parts stimuli indicates that the population response to the whole body is always faster and stronger than responses to component parts presented in isolation (see Fig. 5.10b). Furthermore, the population response to the head presented in isolation is at most times greater than the population response to the body alone without the head.

Cell population coding Whole Body alone

A PSTH of the combined neuronal responses of 6 cells selectively responsive to the whole body but not to either component parts presented in isolation (whole body alone cells) is illustrated in Fig. 5.11. Response on-set occurs at approximately 90 ms after stimulus presentation. It is interesting to note that as a population, small neuronal responses to component parts occur at 190 ms for the head alone stimulus or 220 ms

Figure 5.10. Cell population response to the whole body stimulus, component parts (Head alone, Body alone) stimuli and control objects. a) Combined responses of 12 cells **selectively responsive to the static whole body and component parts (multi-part cells)** to different stimuli: The clear area displays the population response to the whole body stimulus, medium grey area to the head alone stimulus, dark grey area to the body alone stimulus and the black area population response to various control objects. Firing rate is expressed as a percentage of the peak response to the best (in this case the whole body) stimulus. Stimulus presentation occurs at point 0 ms. b) Cumulative Response curve of multi-part population estimate. The colour coding for different stimuli responses is identical to (a). Firing rate is expressed as a percentage of the maximum response to the best (in this case the whole body) stimulus. Stimulus presentation occurs at point 0 ms. The average population response rate was calculated after normalising each cell's activity (S/A=0, Max Response=100). The cumulative population response is defined as the running total of this average population response from 200 ms pre-stimulus onset.



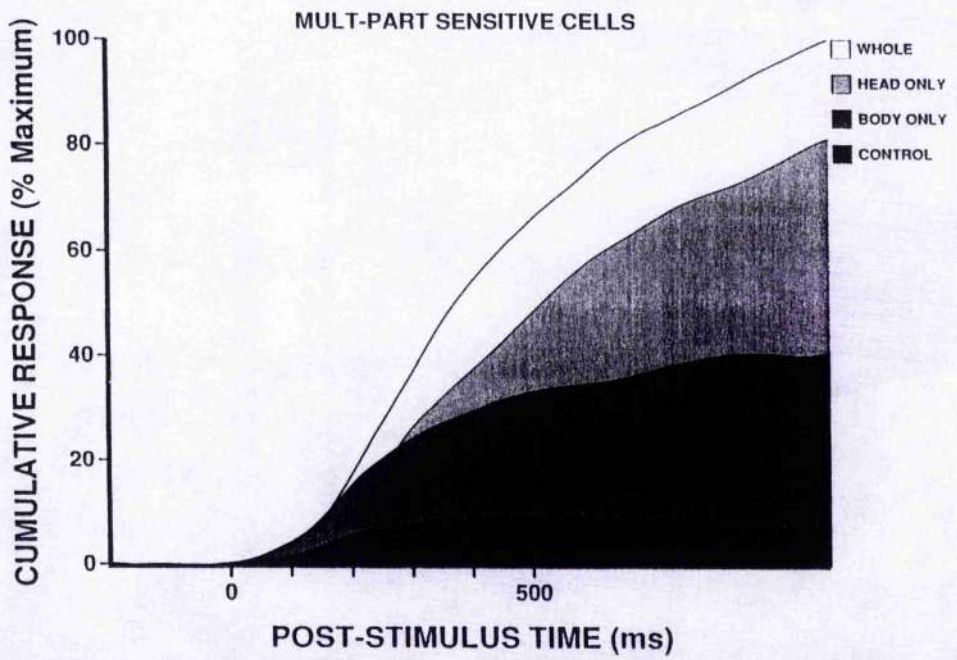
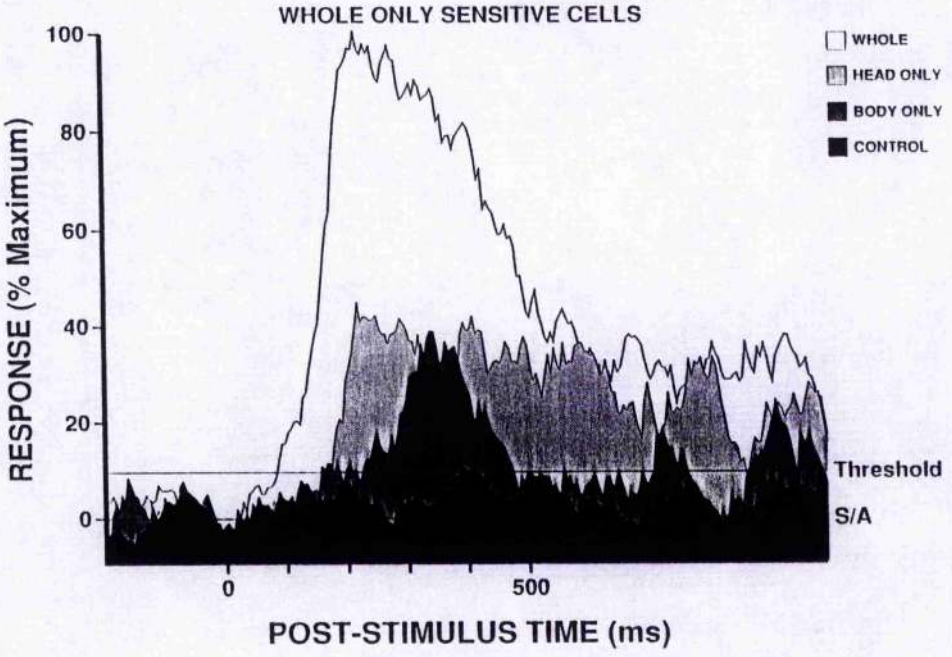
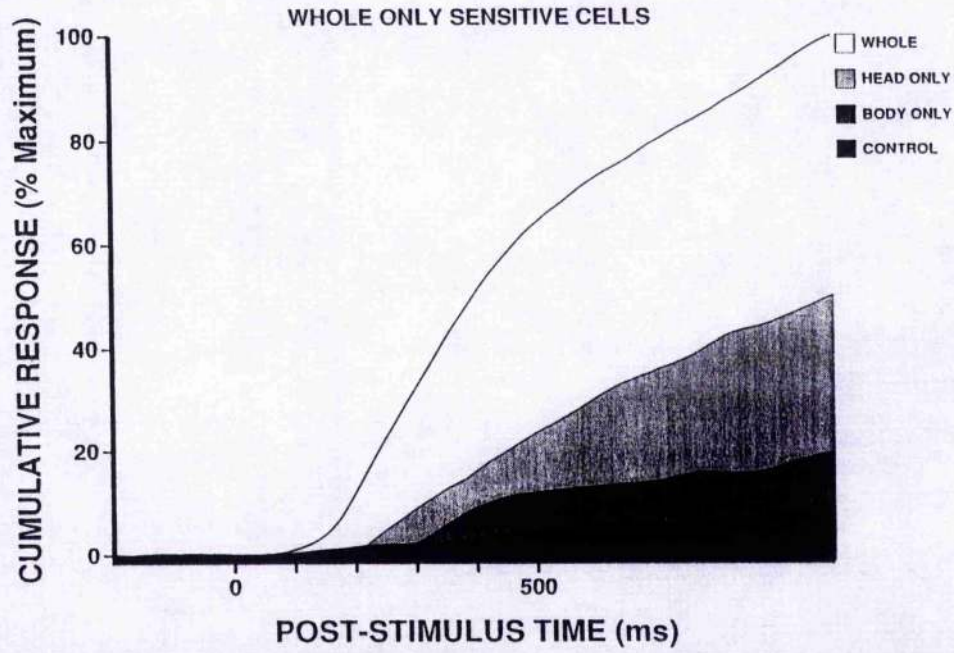


Figure 5.11. Cell population response to the whole body stimulus, component part (head alone, body alone stimuli) and control objects. a) Combined responses of 6 cells **selectively responsive to only the static whole body** to different stimuli: The clear area displays the population response to the whole body stimulus, medium grey area to the head alone stimulus, dark grey area to the body alone stimulus and the black area population response to various control objects. Firing rate is expressed as a percentage of the peak response to the best (in this case the whole body) stimulus. Stimulus presentation occurs at point 0 ms. b) Cumulative Response curve of whole body population estimate. The colour coding for different stimuli responses is identical to (a). Firing rate is expressed as a percentage of the maximum response to the best (in this case the whole body) stimulus. Stimulus presentation occurs at point 0 ms. The average population response rate was calculated after normalising each cell's activity (S/A=0, Max Response=100). The cumulative population response is defined as the running total of this average population response from 200 ms pre-stimulus onset.





for the body alone stimulus after stimulus presentation. These small responses might not have been picked up during single cell analysis since individual data was examined over a 250 ms time period, from 100 ms to 350 ms after stimulus presentation. Indeed, neuronal activity to the visibility of parts of the whole body might not have been strong enough to elicit a response in the individual cell, but might well have accumulated to a small response when looking at the whole body only cell population. However, the cumulative response curve clearly indicates the main point to be made, that the whole body cell population responds much faster at any cumulative response level to the entire object present (see Fig. 5.11b). In addition, at any time the response to the whole body stimulus is greater then the cumulative response to component parts.

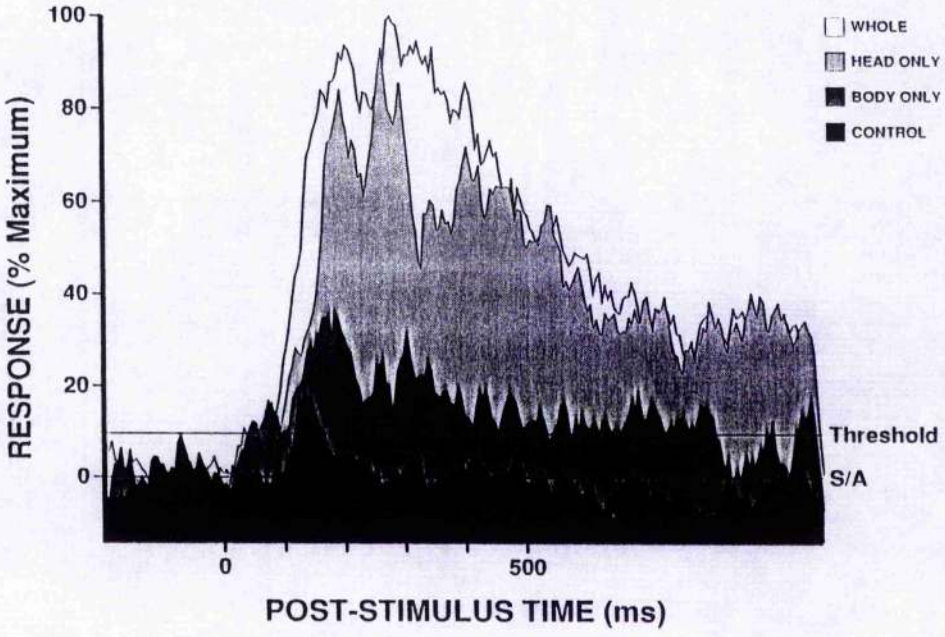
Overall population estimates of time course responses

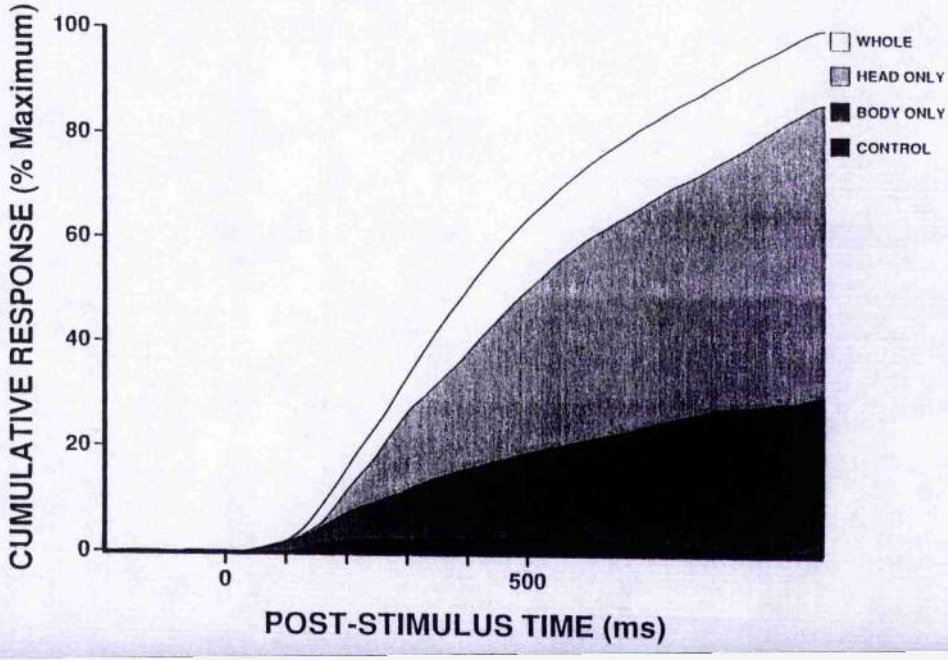
To investigate what the overall neuronal response pattern is of the STS when viewing the whole body and/or its component parts, all cell types described above were included in a population estimate. To illustrate this, 35 cells responsive to the entire body or its component parts were included in a PSTH (see Fig. 5.12a) and cumulative response curve (see Fig. 5.12b) for a population estimate. The whole body stimulus triggers the greatest response at any response level and time. Nonetheless, as a population, the neuronal response to the head presented in isolation is very large. Looking at the cumulative response curve for all cell types (Fig. 5.12b), responses to the whole body occur first and at any time result in the greatest activation. It is pointed out, that the great activity to the whole body stimulus is partly due to the whole body only cells included in this investigation.

Coding of parts in motion

25 cells, selectively responsive to the whole body stimulus in motion, have been tested for their sensitivity to component parts. 18/25 cells were tested with six stimulus conditions (5 trials per stimulus presented in a pseudo-random order): 1) Whole body moving in optimal direction; 2) Head alone moving in optimal direction; 3) Body alone moving in optimal direction; 4) Legs alone moving in optimal direction; 5) Control objects moving in optimal direction and 6) a no stimulus condition (S/A). The remaining 7 cells were tested for neuronal responses to the whole body and only

Figure 5.12. Cell population response to the whole body stimulus, component part (head alone, body alone) stimuli and control objects. a) Combined responses of 35 cells **selectively responsive to the static whole body and/or component part presented in isolation** (i.e. all cells tested) to different stimuli: The clear area displays the population response to the whole body stimulus, medium grey area to the head alone stimulus, dark grey area to the body alone stimulus and the black area population response to various control objects. Firing rate is expressed as a percentage of the peak response to the best (in this case the whole body) stimulus. Stimulus presentation occurs at point 0 ms. b) Cumulative response curve of a population estimate including all cell types. The colour coding for different stimuli responses is identical to (a). Firing rate is expressed as a percentage of the maximum response to the best (in this case the whole body) stimulus. Stimulus presentation occurs at point 0 ms. The average population response rate was calculated after normalising each cell's activity (S/A=0, Max Response=100). The cumulative population response is defined as the running total of this average population response from 200 ms pre-stimulus onset.





one component part presented in isolation and hence these cells were only included in the analysis of view discrimination (see chapter VI). Therefore, 18/25 of the cells were classified in a similar way as was done with static stimuli.

Cells only responsive to one component part in motion

17% (3/18) of the cells were found to be selectively responsive to one moving component part presented in isolation. That is, the response to the sight of the one component part presented in isolation when moving was significantly greater than the responses to moving control objects and to S/A. In addition, for these cells, other component parts presented in isolation and in moving in the optimal direction did not produce a response that was significantly different from moving controls or S/A. An example of such a pattern of response is shown in Fig. 5.13. For this cell the presence of the head in motion was both necessary and sufficient to account for the response to the whole body. All (3) cells reported here were found to be selectively responsive to the head presented in isolation. No cells were found to be selectively responsive to only the body (with the head occluded from sight) or only the legs presented in isolation moving in the optimal direction. It is, however, pointed out that number of cells selectively responsive to only one component part is very small.

Coding the entire body in motion

Cells responsive to multiple body parts in motion

61% (11/18) of the cells studied showed responses to all (tested) components of the body when presented in isolation and presented in motion. The responses to components and the whole body were significantly different from the responses to moving control objects and S/A. Typically the response to the moving whole body was greater than the response to the individual components in motion (e.g. Fig. 5.14).

Cells only responsive to whole body in motion

The remaining 22% (4/18) of the cells showed a response depending on the visibility of the moving whole body in the optimal orientation. These cells responded to none of the component part (head alone, body alone, legs alone in motion) when tested in isolation. For these cells the moving whole body stimulus produced a

Figure 5.13. Neuronal responses of a cell only responsive to the head in motion. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) to the whole body in motion, its components in motion (view 0 move 90), spontaneous activity (S/A) and moving control stimuli are illustrated for one cell (E102_39.31L). The responses to the whole body and the head alone stimuli in motion were not significantly different ($p>0.05$), whereas both these stimuli produced a significantly higher neuronal response than activity rates produced by the body alone stimulus, S/A, or moving control objects ($p<0.05$ each comparison). ANOVA: $F(5,29)=14.527$ $p<0.0005$.

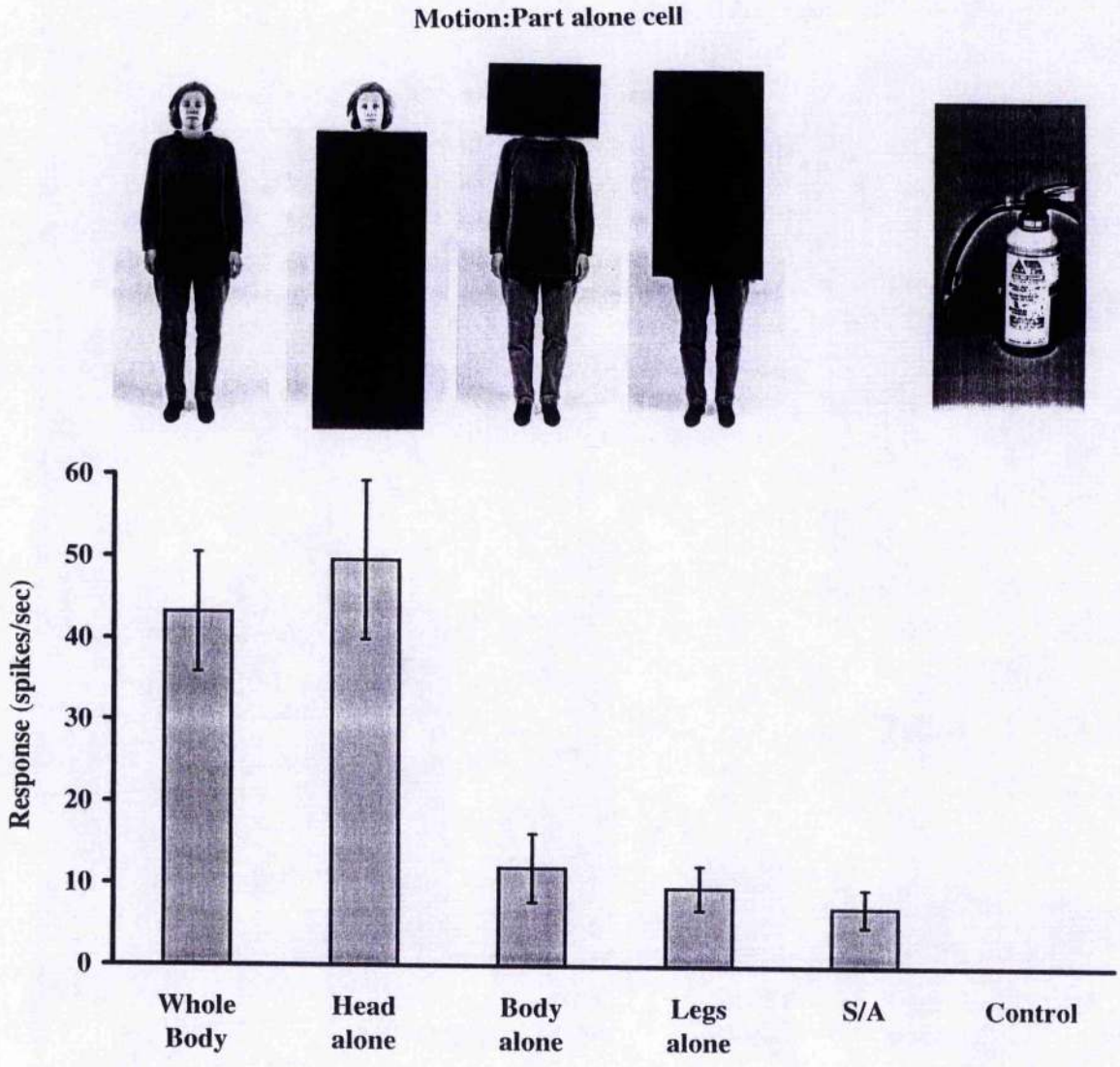
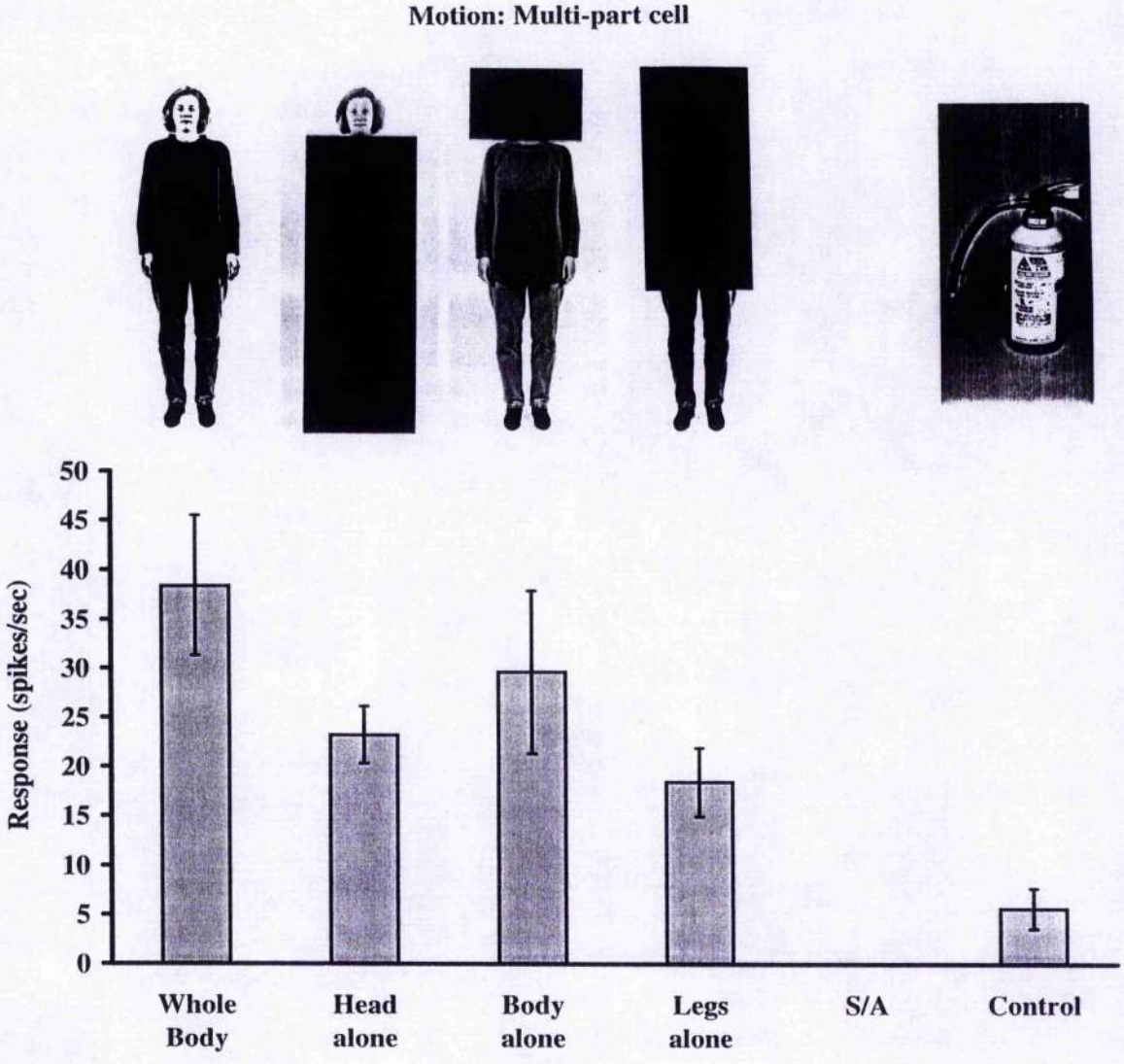


Figure 5.14. Neuronal responses of a cell responsive to multiple parts in motion. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) to the whole body in motion, components in motion (view 0 moving 0), moving control stimuli and spontaneous activity (S/A) are illustrated for one cell (E88_35.41R). The cell responded more to the whole body stimuli in motion ($p<0.05$ each comparison) than to any other stimuli. The cell also responded significantly more to either of the body regions tested in isolation (head alone, body alone and legs alone moving stimuli) than to S/A or moving control objects ($p<0.005$). ANOVA: $F(5,29)=8.7$ $p<0.0005$.



response significantly different from that to other moving body components, moving control objects and S/A. The responses to the moving components individually did not show any statistical difference from responses to moving controls or S/A (e.g. Fig. 5.15).

Histological Localisation

Reconstruction of cell position in 5 monkeys revealed that cells responsive to the head and body were located in the upper bank (and to a certain extent in the lower bank) of the anterior STS. Fig. 5.16 illustrates the position of the different cell types recorded in one monkey (J) and Fig. 5.17 gives an example of the STPa for another monkey (E) at 15 mm anterior to the interaural plane (for more extensive cortical maps of the STS tested see Appendix 1 & 2). The different cell types were intermixed within the cortical areas sampled. This is true for both cells selectively responsive to the optimal stimulus in motion (Fig. 5.17b) or presented stationary (Fig. 5.17a). Nonetheless, in subject E, there seems to be a hint of a static head alone cell cluster at a quite anterior site (between 15-16 mm anterior). Furthermore, it is pointed out that even though in subject E a wide range of the STS was sampled [from 6 mm to 17 mm anterior of the interaural plane (in X-ray mm), see Appendix 2 for maps of the STS indicating recording tracks], most cells were found to be located at a quite anterior site. There was, therefore, no indication that the more anterior in the visual processing pathway one samples, the more generalising cell responses become, i.e. find a greater number of cells to be selectively responsive to both the whole body and the component parts.

DISCUSSION

Cell sensitivity to the body

Previous studies of form processing of biological stimuli in the temporal cortex have focused on cell responses to the sight of the face and other views of the head (Perrett, Rolls et al., 1982; Perrett et al., 1984; 1985; Kendrick and Baldwin, 1987; 1988; Hasselmo, Rolls et al., 1989; 1991; 1992). In the present study, it was not surprising to find cells responding selectively to the sight of the head alone, since the face carries much social information.

Figure 5.15. Neuronal responses of a cell only responsive to the whole body in motion. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) to the whole body in motion, head alone in motion, body alone in motion, legs alone in motion stimuli (view 0 moving 0), moving control stimuli and spontaneous activity (S/A) are illustrated for one cell (E18_31.79R). The whole body moving towards 0 stimuli gave a significantly higher response than the other stimuli tested ($p<0.0005$ each comparison). The head tested in isolation and the body tested in isolation did not show any significant difference from the S/A ($p>0.05$ each comparison). ANOVA: $F(5,29)=42.53$ $p<0.0005$.

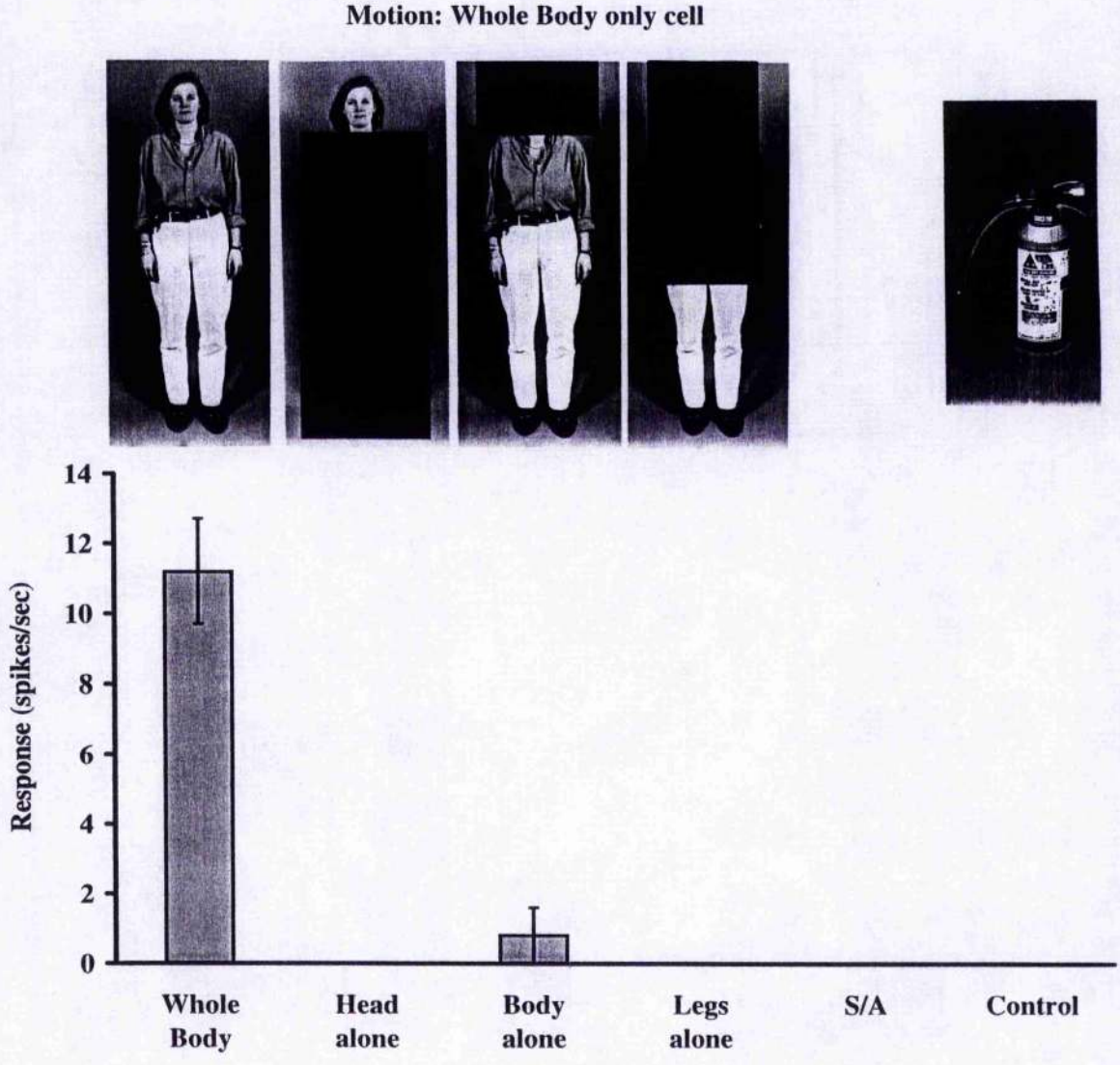


Figure 5.16. Histological reconstruction: Upper Left: Schematic drawing of the right hemisphere of the macaque brain. Shaded region indicates the location area STPa. **Upper Right:** Frontal section of the brain of one monkey (J) 9.5 mm anterior to the interaural plane. The superior temporal sulcus of the left and right hemispheres indicated by the shaded areas. **Lower:** Series of frontal sections every 1 mm (15.5 to 6.5 mm anterior to the interaural plane) showing the location of cells responsive to the head alone (open circles), body alone (filled circles), whole body only (open triangles) and responsive to all parts (filled triangles). The thick lines indicate the brain surface, the thin lines show the boundary between white and grey matter. Cells were located in both the upper bank and fundus of the superior temporal sulcus.

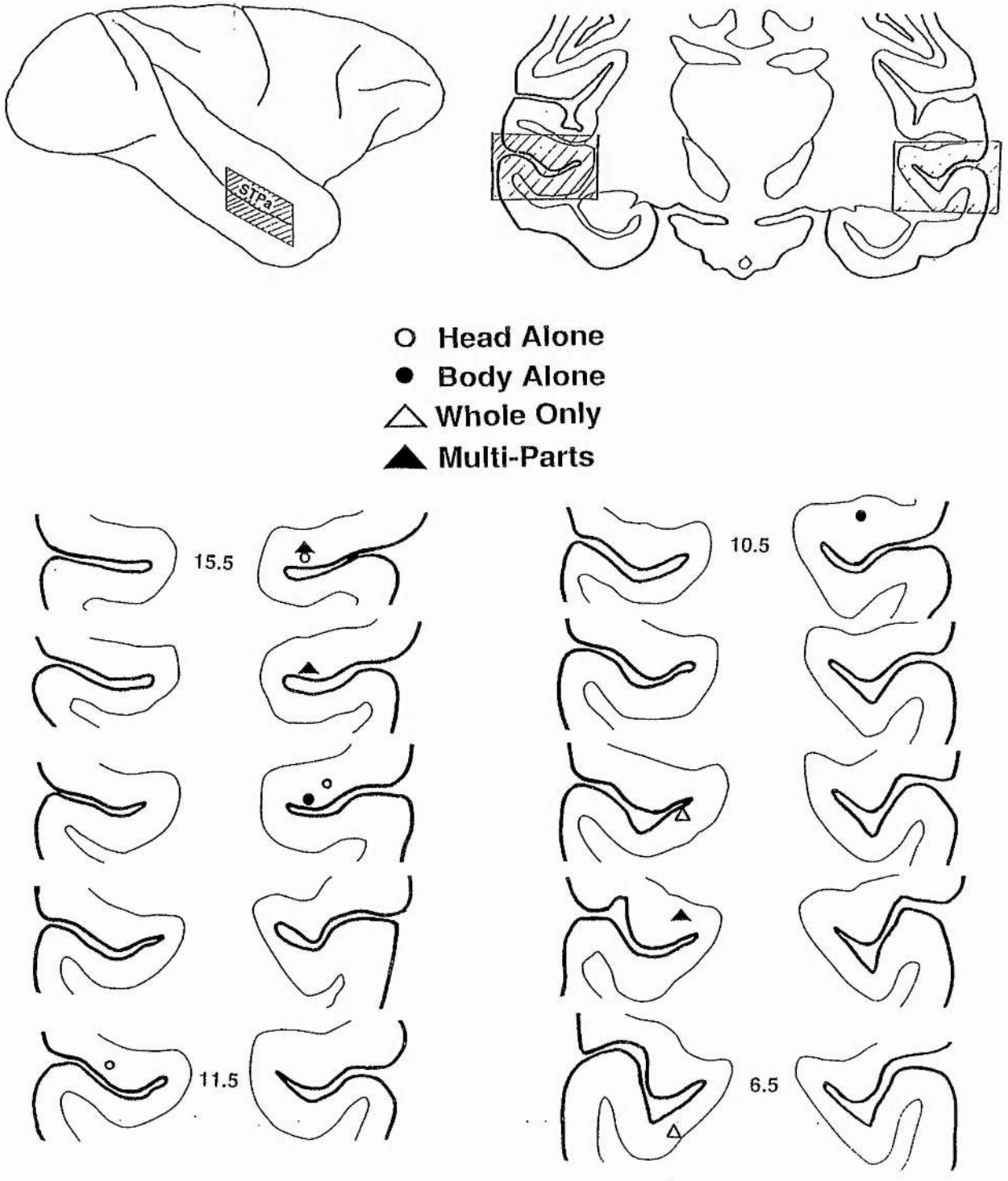
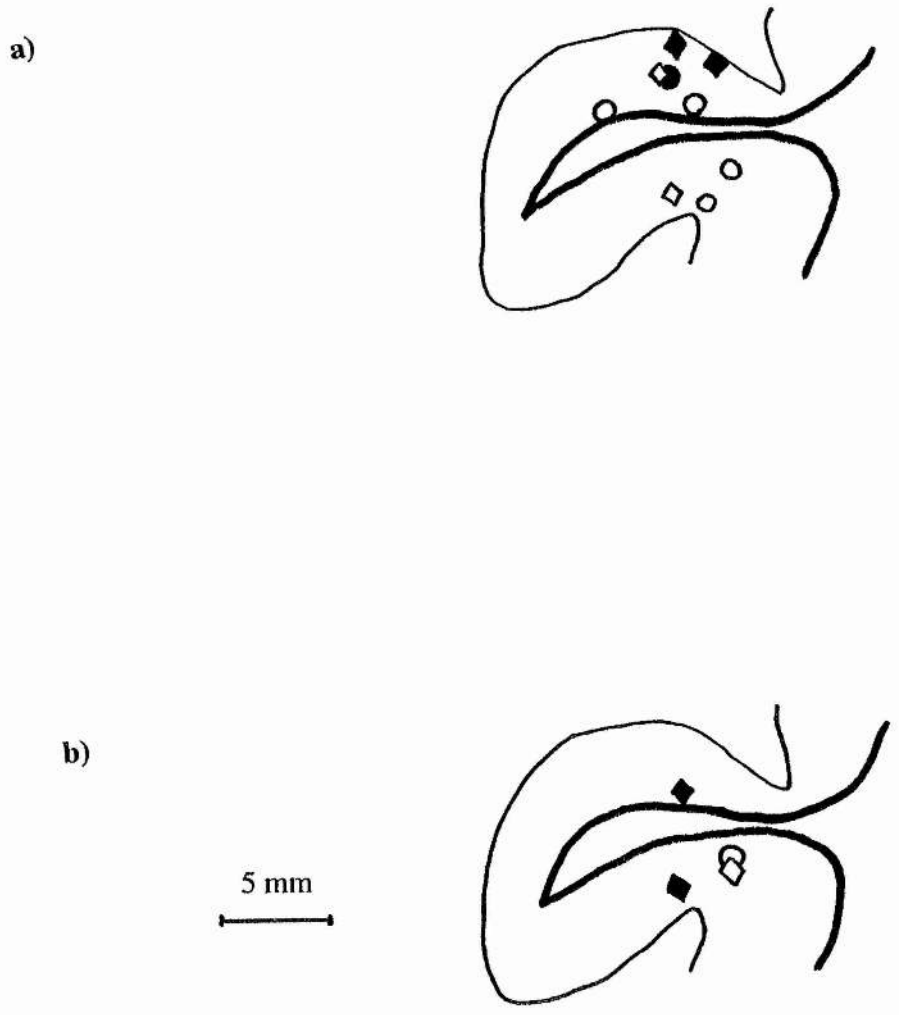


Figure 5.17. Histological reconstruction. A frontal section of subject E taken at 15 mm anterior to the interaural plane showing the location of: **(a)** cells responsive to the **static** head alone (open circles), body alone (filled circles), whole body only (open triangles) and responsive to multi-parts (filled triangles); (b) cells responsive to the **moving** head alone (open circles), moving body alone (filled circles), moving whole body only (open triangles) and responsive to moving multi-parts (filled triangles). The thick lines indicate the brain surface, the thin lines show the boundary between white and grey matter. Cells were located in both the upper bank and fundus of the superior temporal sulcus (see Appendix 1 and 2).



The surprising result of the current study was the extent to which information arising from parts of the entire body, other than the head, influences cell responses. The responses of the majority of cells (65%, 43/66) carried information about regions of the body other than the head. There were three types of cell for which body information was found to be important. For the 6 cells responsive to the body alone, information about the head presented in isolation was insufficient to drive responses. For 24 cells independent responses could be measured to both the head and to the rest of the body. Finally for 13 cells, spatially continued information from the head and the body was critical before the cell responded.

The high proportion of cells found in the present study sensitive to body information suggests that other studies overlooked the importance of the body. It is unlikely that 4 new classes of cells have been uncovered, rather additional information processing abilities of cells that have already been described have been revealed. In previous studies cells have been described as 'face responsive' or even dubbed 'face cells'. The present analysis suggests such labelling may be inappropriate, as it can underestimate or bias interpretation of cell selectivity and functions. The function of cells may require an integration of information from multiple body parts including the eyes, face, hands, body posture etc. (Perrett, Hietanen et al., 1992).

Implications for models

a) Cells selective for one body part

44% (29/66) of the cells studied responded to only one of the two body parts tested. Hence the component parts of an object appear to be coded separately within higher visual association cortex. One might suppose that such findings fit models suggesting an initial encoding of objects in terms of their component 3D volumetric parts (Marr and Nishihara, 1978; Biederman, 1987). Detailed consideration (below) indicates that the current findings do not fit these models.

The neurophysiological recordings indicate response sensitivity to body components that are more complex than simple 3D volumetric shapes. First, the cells are unresponsive to a wide variety of simple and complex control stimuli. If a cell was selective for a particular geon (e.g. a cylinder shape) then one would expect responses both to the shape of a human body which would include several cylindrical components and to a great variety of control objects (e.g. a mug, a broom etc.) which

also have cylindrical components. Second, most cells showed sensitivity to body view (see chapter VI). Indeed, some cells discriminate between mirror symmetrical views (e.g. left profile but not right profile) which contain identical geometric shapes. If a cell response to one body part is to be explained by the presence of a particular 3D volumetric shape (geon or cylinder), then the cell should be equally responsive to all views where this shape remains visible. Even though the psychological models of object recognition proposed by Marr and Nishihara (1978) and Biederman (1987) suggest that object components are encoded independently during visual processing, physiological data reported here show that the complexity of the component parts, which are separately analysed within the STS cortex, is much greater than suggested by those models (i.e. generalised cones or geons).

The results are also incompatible with viewer-centred models of recognition which suggest that only the global shape (whole object) of a particular object view are represented (without independent part representation; Koenderink and van Doorn, 1976; 1979; Edelman and Bülthoff, 1990; Poggio and Edelman, 1990; Seibert and Waxman, 1991).

b) Cells selective for multiple body parts

One population of cells (36%, 24/66) studied was responsive to the whole body and to multiple parts of the body when tested in isolation. This population of cells is in accordance with the models of Biederman (1987) and Marr and Nishihara (1978). These models predict one global description of an object stored in memory which is independently accessible from the sight of any major object component. However, the view tuning of these cells suggests that even these cells do not support these models (see chapter VI).

The categorisation of cells applied here may have underestimated multi-component coding. Sensitivity to information from both the head and the body alone may have been present even for some of the cells categorised as responsive to only one part. First, for some of these cells (13/29) responses to the entire body were different to the response to the most effective part tested in isolation (e.g. Fig. 5.4), indicating that the 'non-effective' part could influence response when viewed together with the effective part. Second, for four cells view tuning was present for the entire body but was not evident for the effective part tested in isolation (see chapter VI).

c) Cells selective for the whole body

The models of Marr and Nishihara (1978), Biederman (1987) and Lowe (1987) do not predict the population of cells (20%, 13/66) which were only responsive to the entire body. For these cells there was no activity provoked by isolated parts. Thus, the cell activity provides a description of the appearance of the entire body but this description is not accessible from the isolated parts. The cell responses provide information about the overall appearance of the object but they do not provide information about independent parts of the object.

The responses of such cells may fit suggestions of Baker-Cave and Kosslyn (1993). These authors propose a model of visual processing in which the overall appearance of the entire object is processed before the parts. If the cells selective for the entire body represent a description of the overall configuration that is processed first, then one would not expect them to be activated by independent parts. The overall configuration of an object could perhaps be revealed from an analysis of coarse image attributes (e.g. low spatial frequencies). This course of analysis could be independent of the detailed form of individual object parts. It is noted, however, that the majority of cells selective for the whole body (as all other cell classes) showed viewer-centred properties in coding (see chapter VI). View discrimination would not be predicted from a very coarse analysis of the body form. Thus, the findings reported here do not fit with the proposal of Baker-Cave and Kosslyn (1993).

Coding body parts in motion

The results of the empirical study reported here indicate that cells selectively responsive to the walking body stimulus code information about component parts of the body (in motion) in a similar fashion as cells selectively responsive to static stimuli of the body. That is, some cells (17%) respond to only to one component part when presented in isolation (and in motion), whereas other cells (22%) require the visible presence of all body parts in their correct configuration, i.e. the whole body walking. The majority of cells (61%), however, will respond to any one component part presented in motion and/or the whole body walking in the optimal direction. These form and motion sensitive cells will not respond to the optimal stimulus presented

motionless (see Oram and Perrett, 1994a, b). It is suggested, however, that for cells selectively responsive to *static* body stimuli, the response pattern to different stimuli (head alone, body alone, whole body) presented in motion would not change the response pattern of the cell (thought the overall response amplitude might be slightly decreased since the stimulus is optimal if presented stationary rather than in motion). Little empirical data comparing neuronal responses to moving and static stimuli is available at this point.

Coding of isolated object parts

One question arising from the current study is why parts of the body are coded independently of each other? Under natural viewing conditions individual parts of a body or indeed any object are often not fully visible. Such situations arise when a body is viewed from behind an intervening object or when one part of the body occludes the sight of another part. It would be impossible to recognise objects in such circumstances if the cortex only contained cells selective for the intact or entire object (such as that depicted in Fig. 5.6). The separate coding of object parts allows recognition under conditions of partial object occlusion.

If each of the major (articulating) parts of an object are coded separately from a small number of views, then recognition of the object is possible despite changes in the configuration of these parts. An alternative processing scheme could rely on cells selective for the entire object's configuration with different cells selective for different configurations. While this scheme is possible, it is also inefficient and would require large numbers of 'templates' or cells, each selective for one of the huge range of possible configurations. Thus, an advantage of separate coding for distinct body parts is that such coding is compact and accommodates the great variety in configurations of those body parts. It is noted that single cell coding for specific configurations may exist for some of the more meaningful body postures (e.g. crouching, bipedal, quadrapedal, see Perrett et al., 1982).

Learning the association between parts

One issue that is raised by the current study concerns the mechanism by which the nervous system integrates information about different parts of the same object. Several unsupervised learning mechanisms have been proposed whereby output units

in artificial or real neural networks 'learn' the pairing of independent input patterns when these inputs are associated over time (e.g. Rumelhart and Zipser, 1985; Foldiak, 1990; 1991). After learning the association, the output units of the network are able to respond to any of the input paired patterns.

Twenty-four of the cells (or 35, including walking cells) that are described here were selectively responsive to any of the component parts of the body tested in isolation. The body parts for which the cells' responses were conjointly sensitive are visually distinct, but nonetheless these two parts are physically related in the entire object. This physical (spatial) association means that the head and the body (in the same view) are frequently encountered together. It is speculated, therefore, that the sensitivity of cells responsive to multiple parts is established through a learned association of simultaneously presented inputs from cells selective to only one body part. Studies of Miyashita and colleagues (Miyashita, 1988; Miyashita and Chang, 1988; 1990) have demonstrated that single cells in the anterior and ventral temporal cortex do register the temporal association of abstract patterns which are presented sequentially with an interval of 15 seconds. A recent report by Sobotka and Ringo (1993), however, failed to find evidence that an association between complex arbitrary patterns (presented simultaneously in pairs) occurs in IT cells. They trained monkeys on a discrimination task to associate two paired abstract pattern and found that IT cells were not any differently activated by simultaneously associated patterns, than when two non associated patterns were viewed. There are, however, several reasons why such a result might have been found. Firstly, there was always a blank gap between the two associated patterns, allowing the subject to perceive the paired images as two independent entities (unlike a whole body composed of body parts). Furthermore, the paired images were randomised in their spatial location in respect to each other, i.e. sometimes pattern A would be presented at the top and pattern B at the bottom, whereas during the next presentation the order is reversed. Hence the spatial relationship between the patterns varied discouraging the perception of the pattern as one entity.

The scheme described above implicitly assumes that cells responsive to individual body parts should be activated by the visual input before cells responsive to multiple parts.

Population estimates of time course responses

Having discussed cell responses at the single cell level, it is now important to ask the question: what happens to the overall neuronal response of a brain area studied when viewing particular stimuli? STPa cells selectively responding to only one component part of the body also respond to the whole body stimulus (since the effective component part is present in this image). It is interesting to note that the overall neuronal response of head selective cells to the whole object is initially the same as the response to the head presented in isolation (see Fig. 5.9). Only after approximately 530 ms after stimulus presentation (once neuronal response to the whole body and the head alone stimuli reach a level of 32% of the maximum response), the head alone cell population discriminates between the two stimuli. That is, the response to the whole object is greater than the response to the head presented in isolation. Therefore, if there were only cells in the brain selectively responsive to the head presented in isolation and none of the other cell types describes above, then one would expect that at the behavioural level, reaction times (RTs) for discrimination between the head alone and the whole body presented should not differ if carried out by the cells within the first 350 ms.

Furthermore, whole body only cells, cells which only respond if both body parts (head and body) are present in their correct spatial relationship, will eventually also respond (though only slightly) to either component part presented in isolation (see Fig. 5.11). This activity pattern could be explained if one argues that the main aim of these cells is to signal the presence of a whole body in the visual field. Once this has been established, further information about the head and eventually the body part can be taken into account for, for example, social communication (facial expression, direction of attention, etc.).

Looking at the overall neuronal activity, including all cell type populations selectively responsive to the visibility of the head/body, then the response to the whole body stimulus is at all times and at any response level greater than the response to either component part presented in isolation. This can explain behavioural recognition

studies which suggest that whole objects are recognised faster than isolated parts of the object (Baker-Cave and Kosslyn, 1993). However, this response pattern does not necessarily suggest top-down processing (where the whole is recognised before component parts), but can be explained by the magnitude of population cell response to the whole object (body) in comparison to the population response magnitude to component part stimuli. That is, response onset latency may not differ between the different stimuli, but because there is a greater response to the whole body stimulus, any level is reached by the response to the whole body first.

Chapter VI

VIEW SPECIFICITY AND COMPONENT PARTS

As described earlier (chapter III) there have been two types of reference frames suggested for object recognition. An object-centred representation is a description which relates an object's component part to a framework based on the object itself. Such a representation would allow very effective object identification, but not carry any additional information about the object (e.g. what direction the object is pointing; or if the object is a person, in what direction does he/she look). A viewer-centred representation, on the other hand, is a description which relates an object's component parts (or the entire object) to a framework based on the observer. Such a framework would require several representations, each coding a specific view of the same object (e.g. front view, back view etc.). This chapter describes and discusses an empirical investigation of the role of component parts in coding an object's view.

INTRODUCTION

Object-centred representation

Both Marr and Biederman's accounts of object recognition can be considered object-centred (Marr and Nishihara, 1978; Biederman, 1987). In these models only one descriptive representation of the object is stored in long term memory (see also Lowe, 1987; Porrill et al., 1988). This description should be accessible from all view-points, provided the principle axis is fully visible (Marr and Nishihara, 1978).

As mentioned previously, under Biederman's account, object recognition can be based on a very small number of geons (see chapters III and V). Thus, as long as a sufficient number of geons are visible, recognition should not be affected by view. Under both Marr and Nishihara's and Biederman's object-centred scheme, perspective view should not affect the cellular mechanisms involved in the higher stages of object processing (beyond the limitations noted). One would expect, from both theories, to find cellular units late in the visual pathway which code an object in a way which is accessible from *all* views (i.e. cells that respond to all views of an object).

Psychological studies show that for brain damaged and normal subjects, ease of recognition is influenced by the visibility of an object's salient features and cues to the 3D view/orientation of the object (Warrington and Taylor, 1973; Humphreys, 1984; Warrington and James, 1986; Humphrey, 1989).

Viewer-centred representation

Representation of an object in a viewer-centred manner relies on descriptions of the object relative to the viewer. Such description includes a collection of the 2D visual characteristics of the object that are visible from a specific view-point. The number of characteristics present in any one viewer-centred description is therefore smaller than that of an object-centred description. Two main types of viewer-centred representation have been considered, in which a) only one view of the object is stored, or b) multiple views are stored in long term memory (Palmer, Rosch et al., 1981; Jolicoeur, 1985; Tarr and Pinker, 1989; Ullman, 1989; Edelman and Bülthoff, 1990; McMullen and Farah, 1991; Cutzu and Edelman, 1992; Verfaillie, 1992). Note that under scheme (a) where a single view is stored, the view of the object's components would be specified with respect to the viewer. Hence, this single description is not object-centred.

Palmer et al. (1981) has defined the 'canonical view' as the single view which reveals the maximal information about an object's salient features. For most objects (8/12 objects of those studied by Palmer et al., 1981), the canonical view lies between the 'front' and 'side' views, with the object's principal axis oriented 45° to the observer's line of sight. In addition, for heads the 45° (or 3/4) view is the most readily recognised in naming and matching tasks (Bruce et al., 1987; Perrett, Benson et al., 1994). Models envisaging storage of a single viewer-centered description (such as the canonical view) rely on a transformation of the incoming image to match the stored description. The transformation may involve processes akin to mental rotation (Shepard and Cooper, 1982). If an object is represented by a single canonical viewer-centered description, then one might expect to find greater numbers of cells selectively tuned to this view of an object rather than to other views of the same object.

Most viewer-centered models of object recognition suggest that *several* views of an object are represented in memory (Koenderink and van Doorn, 1979; Tarr and

Pinker, 1989; Ullman, 1989; Edelman and Bülthoff, 1990; Poggio and Edelman, 1990; Seibert and Waxman, 1991). The theories, however, differ as to the number and nature of the views stored. In these models the incoming image is either transformed to match to the nearest stored view, or alternatively the image is identified by interpolation from a minimum of three surrounding stored views. Multi-representational viewer-centred models suggest the presence of cells coding different specific views of the same object.

Cellular sensitivity to object view

Physiological studies have investigated the view sensitivity of cells responsive to faces in the temporal cortex (Desimone, Albright et al., 1984; Perrett, Smith et al., 1984; 1985; Kendrick and Baldwin, 1987; Hasselmo and Rolls et al., 1989a; 1989b; 1991; 1992). All studies reveal that the majority of cells are sensitive to change in perspective view. That is, most cells code in a viewer-centred fashion. Different cells are tuned to different optimal views (some to the face, some to the back of the head, etc.). It is interesting to note that in humans, evoked potential studies also indicate viewer-centred processing of the face in that some different views triggered differently shaped VEPs (Boetzel and Grüsser, 1989; Jeffreys, 1989; Jeffreys et al., 1992; Jeffreys and Turkmachi, 1992).

A small population of cells in the macaque STS has been found to respond equally to all views of the head (Perrett, Smith et al., 1984; 1985; Hasselmo, Rolls et al., 1989a; 1989b; 1991; 1993). The insensitivity of these cells to changes in view is in accordance with the definition of object-centred coding. Object-centred cells could arise from a combination of the outputs of different viewer-centred cells (Perrett and Oram, 1993).

The aim of the current study was to compare view tuning for the whole body with view tuning for component body parts. This aspect of the study could provide potential insight in the role of view-selective processes in the integration of information about separate object parts.

Furthermore, STPa cells which are selectively responsive to the whole body in a particular view and walking in a particular direction (but not activated by different body views or direction of motion) have been reported (Perrett and Oram, 1993; Oram and Perrett, 1994b, c, see also chapter V). Such cells were tested for their neuronal

response pattern when only isolated component parts of the whole body were presented in different views moving in the optimal direction. It was therefore possible to investigate whether view discrimination for component parts moving in their optimal direction is carried out in a similar manner as when the optimal stimulus is stationary.

METHODS

General testing and single cell recording methods, as described in chapter IV, were applied.

Visual static stimuli

2D and 3D visual static whole body and component parts (head presented in isolation, and body without the head presented in isolation) stimuli were used to test neuronal response patterns of cells selectively responsive to the head/body. These stimuli have been described in the method section of chapter V. For most cells multiple views of the stimuli were tested: front view (0°), left profile (90°), back of head (180°), and right profile (270°). In addition to these, four intermediate views (45°, 135°, 225° and 315°) were tested for some cells. Therefore, testing was performed with stimuli of the (cell's) preferred view, the opposite view (180° rotation was usually the least effective view), control stimuli and no stimulus. The views used for testing components were the same as used for testing the whole body. For each cell, a number of different control stimuli were tested (2D and 3D) (see chapter IV).

Visual stimuli in motion

Furthermore, cells selectively responsive to walking bodies were investigated for their neuronal response when moving body component parts (head alone, body alone, legs alone; see method section of chapter V) and whole body stimuli were presented in different views. The direction of motion (walking) of the test stimuli was always the most optimal direction for that cell, though the view of the head/body varied (0°, 45°, 90°, 135°, 180°, 225°, 270°). Generally, once the optimal direction of motion and the optimal view of the whole body was established, the cell was tested for

its response to whole body and body parts stimuli presented in the optimal view and the same stimuli rotated in depth by 180° (e.g. 0° and 180°; 45° and 225°).

Data analysis

The influence of view on cell responses to the whole and component parts was analysed off-line with 2-way ANOVAs with view and body part tested as main factors.

View discrimination index

A view discrimination index was computed for each viewer-centred cell from the view producing the greatest change in activity from S/A (best angle) and smallest change in activity (worst angle) to the whole body stimulus. The cell's activity was then measured to the component stimulus in the same two views. This was carried out for both static and motion cells.

The formula used to calculate these indices was the same as used previously (Oram and Perrett, 1992):

$$I_V = \frac{[(\text{response to best angle-S/A})-(\text{response to worst angle-S/A})]}{(\text{response to best angle-S/A})}$$

RESULTS

View sensitivity was investigated for a total of 73 cells which responded significantly different to the visual whole body stimulus than to a wide range of static and moving control objects and spontaneous activity (S/A). 53 of these cells were selectively responsive to the head/body presented stationary (static stimuli), whereas the remaining 20 cells were selectively responsive to the head/body in motion. For 53/73 cells (39 cells responding to static and 14 cells responding to moving head/body stimuli), view sensitivity was tested for the whole and at least two parts (head alone, body alone, legs alone), whereas for 20/73 cells view sensitivity was measured for the whole and only one part (either presented in static form or in motion depending on the preference of the cell). These latter cells are not included in the classification of cells,

but can be included in the analysis of how effective cells are in discriminating different views. In this section, cells selectively responsive to *static* faces/bodies and their responses to whole bodies and component parts of the body presented in different views are described first. The classification which is established thereby is then used to describe neuronal responses of cells selectively responsive to the whole body and sometimes its component parts when *in motion* (walking).

Cell responses to static heads/bodies

Cells with viewer-centred properties

47/53 (89%) of all the cells selectively responsive to the *static* head/body were found to show viewer-centred properties in their coding of the whole body stimulus. That is, neuronal responses to one view of the body (the best/optimal view) was significantly different (greater) than responses to other non-optimal (worst) views of the body. As should be evident from most the figures used in this thesis, different cells were found to be tuned to different views (some responded maximally to the front view of the body, others to the back, others to the left profile, and others to the right profile etc.). Below I will describe the findings, using the classification of cell types established in chapter V.

Cells coding only one static body part

29/53 cells were selectively responsive to only one (effective) component part. These cells responded to the other component part at rates not significantly different to S/A or control objects (23 head alone and 6 body alone cells were tested). 97% (28/29) of the cells studied displayed sensitivity to view. These cells responded significantly different to a minimum of two views of the whole body. Furthermore, the majority of these cells (90%, 26/29) with viewer-centred properties for the whole body also showed view-selectivity for the effective component part tested with the same views as the whole body stimuli (see e.g. Fig. 6.1). The remaining 3/29 cells (2 body alone and one head alone cell) discriminated between different views when the whole body was tested, but did not discriminate between views for the effective body part tested in isolation (see e.g. Fig. 6.2). Since one cell discriminated between different views of the head presented in isolation, this phenomena cannot be argued to be due to little

Figure 6.1. Neuronal responses of a **'head alone' cell with viewer-centred** properties to the whole body and the head alone. Histogram of response (spikes/sec) to different stimuli. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) of one cell (E99_40.16L) tested to view 0° and view 180° of the whole body and its components. Two-way ANOVA revealed a significant main effect of view [$F(1,24)=31.3$ $p<0.0005$]; body part tested [ANOVA: $F(2,24)=9.3$ $p=0.001$]; and interaction between these factors [$F(2,24)=5.2$ $p<0.05$]. Protected least significant difference tests (PLSD), post-hoc tests, indicated significant response discrimination between the different views of the entire body ($p<0.0005$) and of the effective body part tested in isolation, head only ($p<0.0005$).

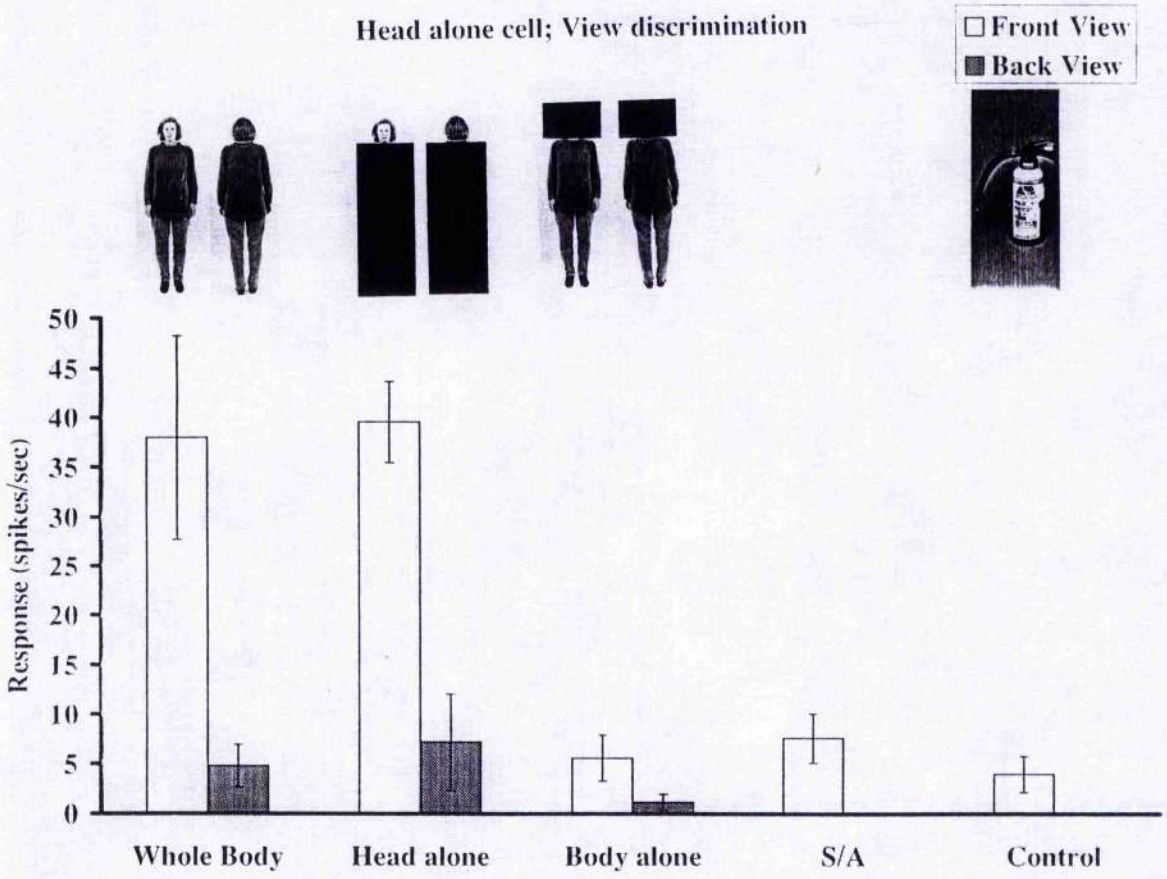
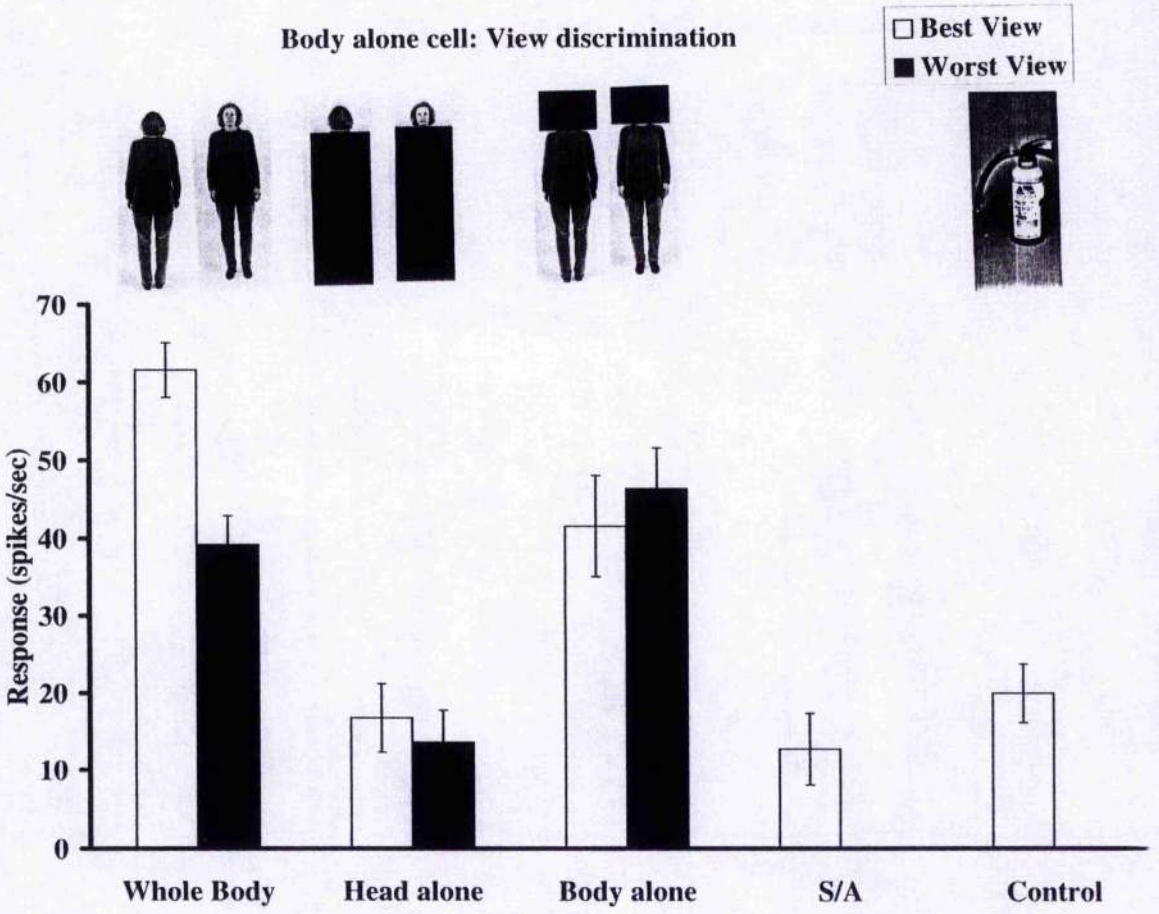


Figure 6.2. Neuronal responses of a **'body alone' cell with viewer-centred** properties to the whole body but not to the body alone. Histogram of response (spikes/sec) to different stimuli. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) of one cell (E92_38.20L) tested to view 180° and view 0° degrees of the whole body and its components. Two-way ANOVA revealed a significant main effect of view [$F(1,22)=4.5$ $p<0.05$]; body part tested [ANOVA: $F(2,22)=8.06$ $p<0.005$]; however, there was no interaction between these factors [$F(2,22)=0.12$ $p>0.5$]. Protected least significant difference tests (PLSD), post-hoc tests, indicated significant response discrimination between the different views of the entire body ($p<0.05$) but not for the different views tested of the effective body part (body alone) ($p=0.5$).



difference in the visual appearance of the stimuli tested when presented in different views.

Coding the entire body

Cells responsive to multiple static body parts

11/53 cells which were selectively responsive to the whole body and multiple component parts tested were tested for their sensitivity to different views of the whole body and the same views of the component parts. 7/11 cells were found to be coding the whole body and its component parts in a viewer-centred manner (see e.g. Fig. 6.3).

The view discrimination for the whole and parts was found to be compatible for all cells. That is, no cells responsive to multiple static body parts were found to only carry out view discrimination for the whole body but not its component parts. If a cell exhibited view discrimination between two views of an isolated body part, then the cell also showed the same direction of view discrimination for the whole body. Furthermore, view sensitivity was observed to be similar for both component parts.

Cells only responsive to the static whole body

13 of all the cells (13/53) tested were selectively responsive to the whole body only and did not respond to the component parts of the body at rates significantly different than to S/A and control objects. The majority of these cells (12/13, 92%) were found to be coding information about the whole body in a viewer-centred manner (see e.g. Fig. 6.4).

Cells with object-centred properties

11% (6/53) of all cells tested for their selectivity to different views of the head/body showed no preferred view for either the whole body or its components (see e.g. Fig. 6.5). It should be noted that some of these cells were classified as object-centred on the based of only two test views (best and worst view).

One cell (1/29) of all cells selectively responsive to only one component part of the body was found to be coding information about the whole body and the effective body part in an object-centred manner.

Figure 6.3. Neuronal responses of a cell with **viewer-centred properties to the body and its components**. Histogram of response (spikes/sec) to different stimuli. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) of one cell (B06_17.88R) tested to view 270 degrees and view 90 degrees of the whole body and its components. Two-way ANOVA revealed a significant main effect of view [$F(1,27)=11.4$ $p<0.005$]; but not for the body part tested [$F(2,27)=0.3$ $p>0.5$]; nor was there an interaction between these factors [$F(2,27)=0.7$ $p>0.5$]. Protected least significant difference tests (PLSD), post-hoc tests, indicated significant response discrimination between the different views of the entire body ($p<0.005$) and of the body parts tested in isolation, head only ($p<0.05$) and body only ($p<0.05$).

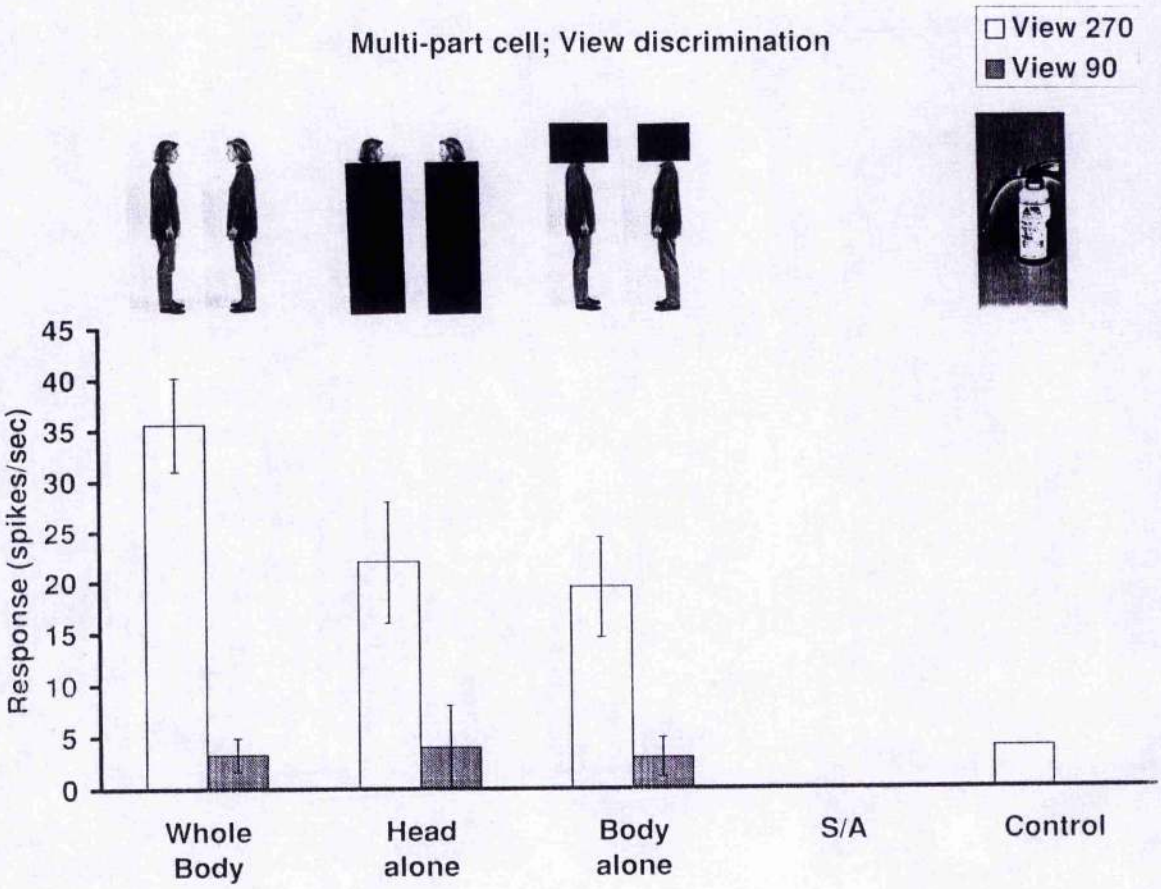


Figure 6.4. Neuronal responses of a **'whole body only' cell with viewer-centred** properties to the whole body. Histogram of response (spikes/sec) to different stimuli. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) of one cell (E08_26.50R) tested to view 0 degrees and view 180 degrees of the whole body and its components. Two-way ANOVA revealed a significant main effect of view [F(1,16)=10.46 p=0.0005]; body part tested [F(2,16)=7.3 p=0.006)]; and interaction between these factors [F(2,16)=11.2 p=0.001)]. Protected least significant difference tests (PLSD), post-hoc tests, indicated significant response discrimination between the different views of the entire body (the only effective stimuli, $p<0.0005$).

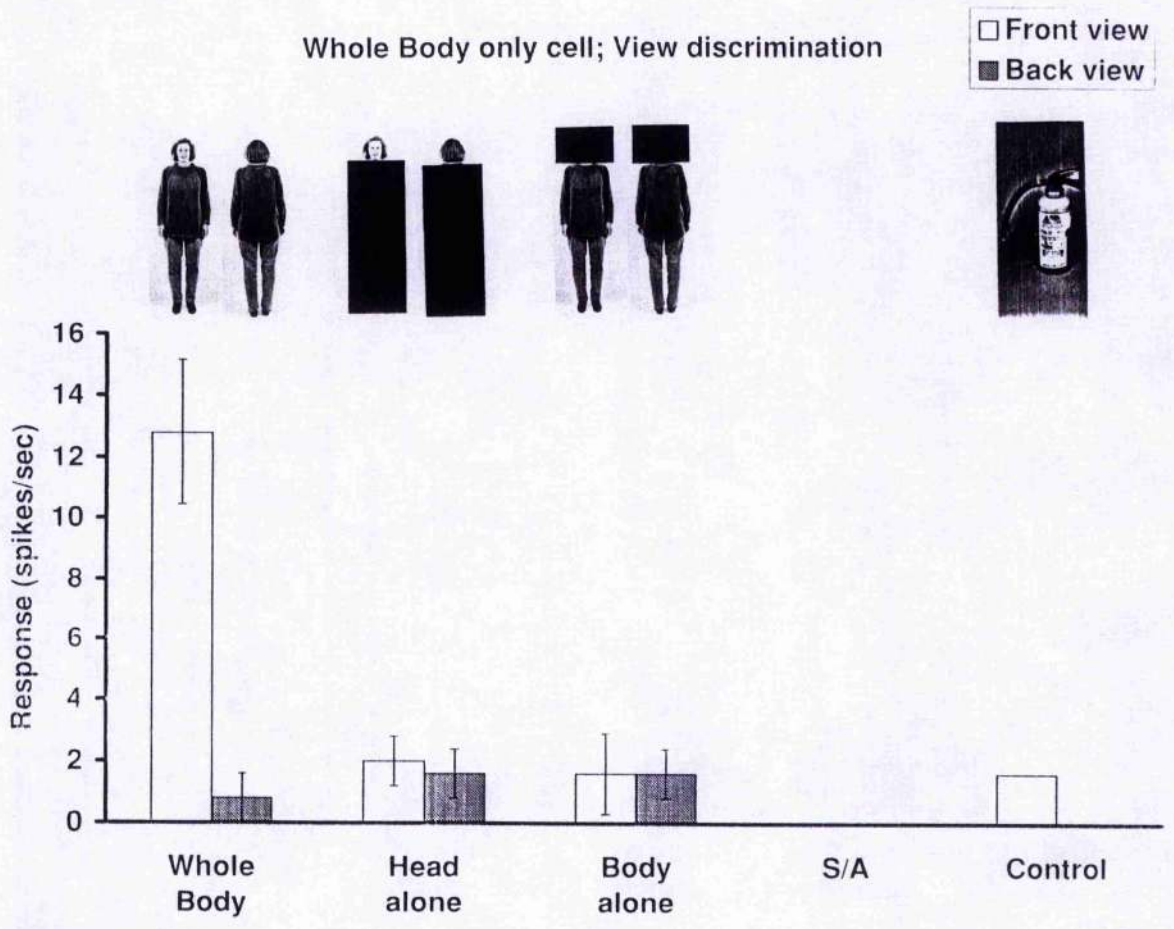
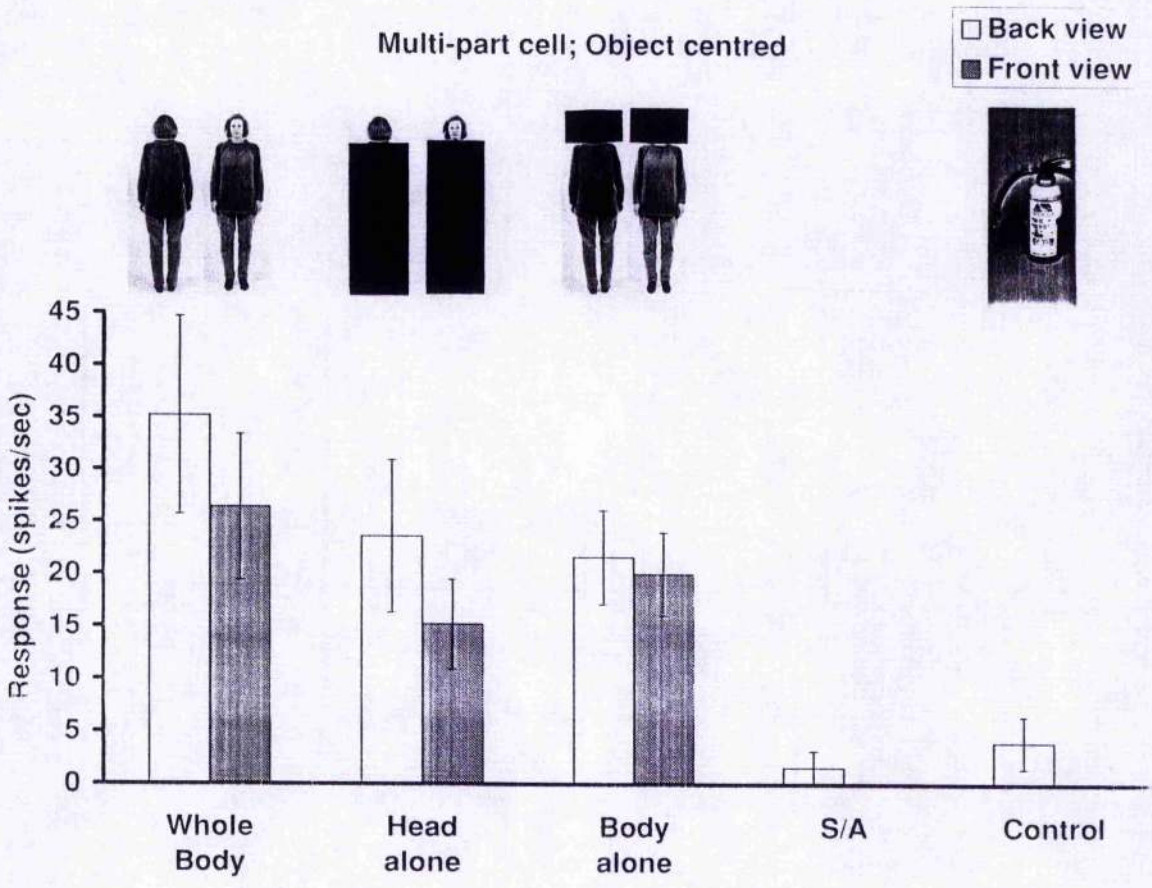


Figure 6.5. Neuronal responses of a cell with **object-centred properties to the body and its components**. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) of one cell (E79_35.14L) to the back (view 180) and front (view 0) views of the whole body and its components are shown. 2-way ANOVA showed no significant main effect of the body part tested [$F(2,24)=1.9$ $p=0.2$] and no effect of view [$F(1,24)=1.4$ $p>0.5$] and no interaction between view and part tested [$F(2,24)=0.2$ $p>0.5$]. All responses to the whole body and parts stimuli in either the front view or the back view were significantly greater than control stimuli and S/A [ANOVA for the views: $F(4,20)=6$ $p<0.005$; ANOVA for the back views: $F(4,20)=4.1$ $p<0.05$].



4/11 of the cells responsive to multiple static body parts showed object-centred properties when tested with different views of the stimuli (Fig. 6.5). All the cells coded information about effective body parts in the same object-centred manner as they coded information about the whole body.

One cell out of 13 static whole body only cells did not discriminate between different views of the entire body tested. This cell responded equally well to the entire body presented in different views, but did not respond any differently to different views of the component parts than to S/A and control objects.

Cell responses to heads/bodies in motion

Motion sensitive cells with viewer-centred properties

A total of 20 cells were tested for their responses to different views of the whole body and its component parts when moving (walking) in the optimal direction. 95% (19/20) of these cells showed viewer-centred properties of coding information about the head/body. As before, the previously established classification of cells (see chapter V) is used to organise the findings.

Coding of single body parts in motion

Seven cells were selectively responsive to the whole body and only one component part when moving in the optimal direction. All the cells (100%) coded information about the whole body in a viewer-centred manner (see e.g. Fig. 6.6). Of these viewer-centred cells, only one cell discriminated between different views to the whole body stimulus but not to the same views of the effective body part.

Coding the entire body in motion

Cells responsive to multiple body parts in motion

8/9 cells selectively responsive to multiple body parts tested, coded information about the entire body in a viewer-centred manner. The majority (6/8) of these cells generalised the view-sensitive coding from the whole body to the component parts tested in the same views (see e.g. Fig. 6.7). Two cells, however, only coded information about the whole body in a viewer-centred manner and not information about isolated body parts presented in the same views moving in the same direction.

Figure 6.6. Neuronal responses of a **'head alone in motion' cell with viewer-centred** properties to the whole body and the head alone moving in the optimal direction (move towards 225°). Histogram of response (spikes/sec) to different stimuli. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) of one cell (E94_38.18L) tested to view 180° and view 0° of the whole body and its components. Two-way ANOVA revealed a significant main effect of view [$F(1,32)=10.6$ $p<0.005$]; body part tested [ANOVA: $F(3,32)=10.6$ $p<0.0005$]; and interaction between these factors [$F(3,32)=6.76$ $p=0.001$]. Protected least significant difference tests (PLSD), post-hoc tests, indicated significant response discrimination between the different views of the entire body ($p<0.0005$) and of the effective body part tested in isolation, head only moving towards 225° ($p<0.0005$).

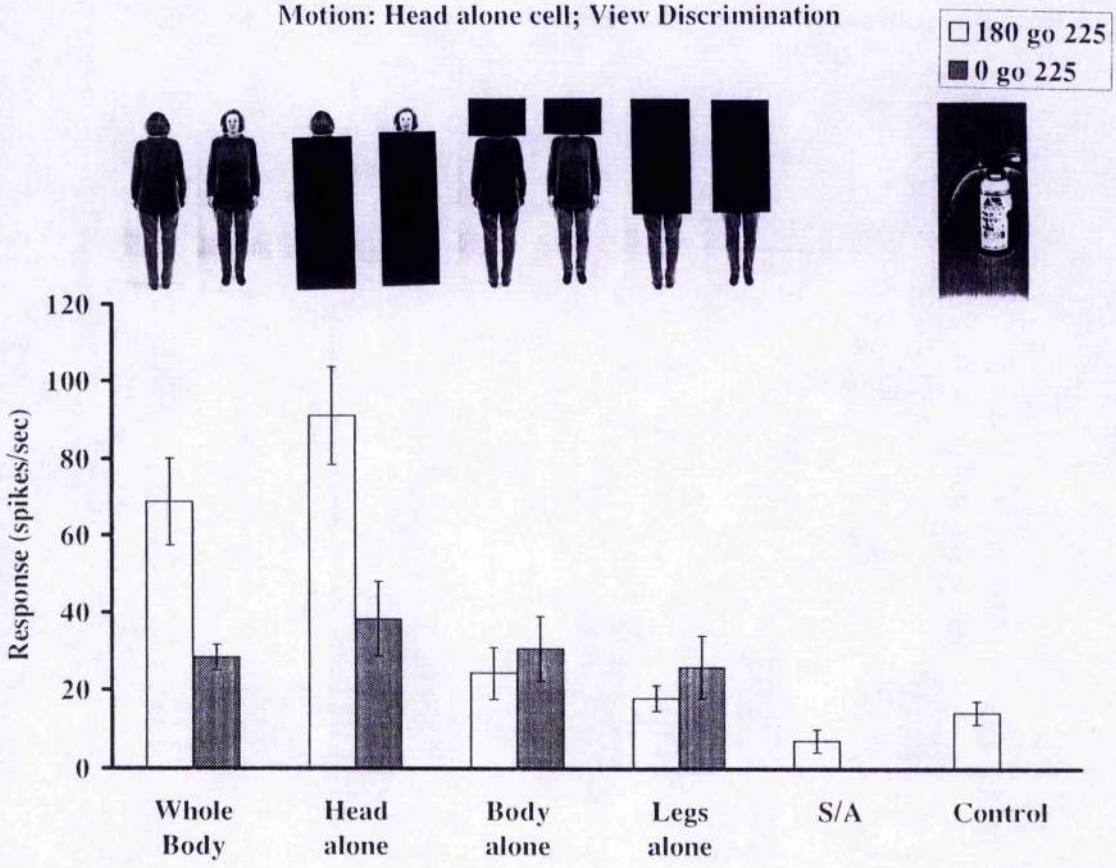
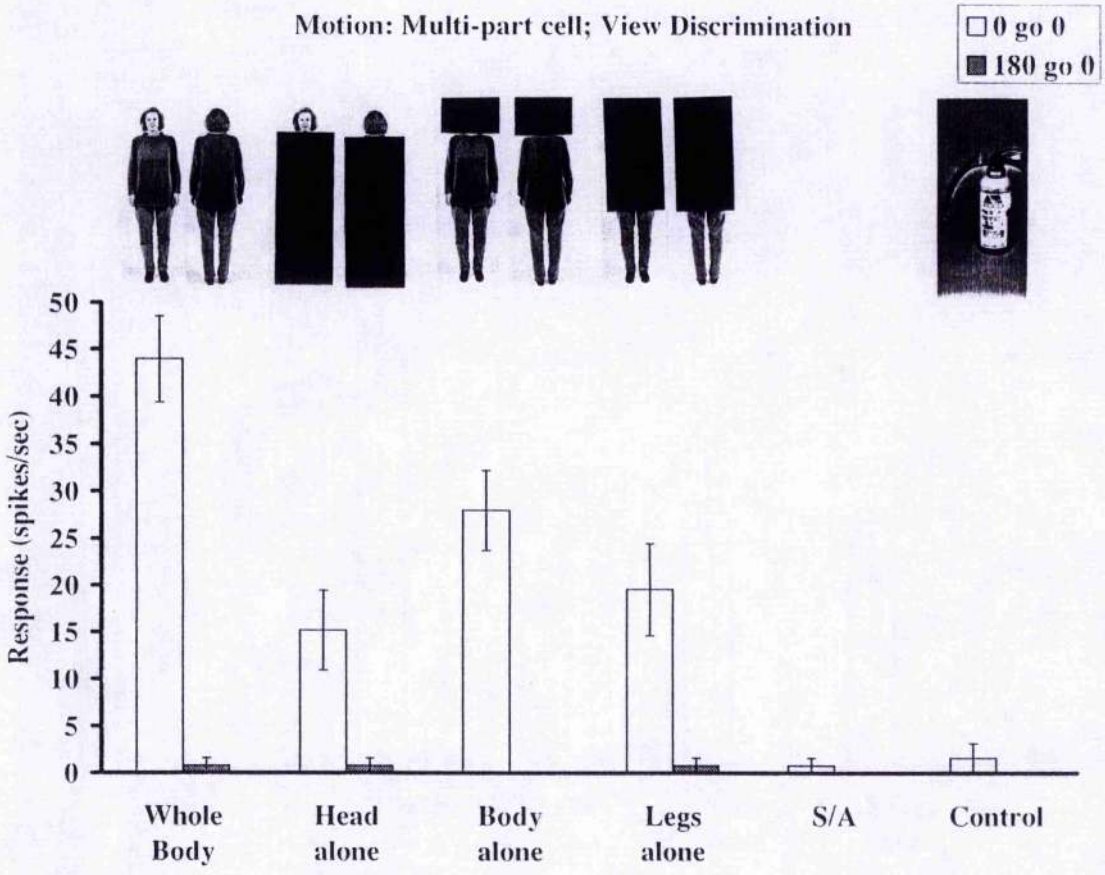


Figure 6.7. Neuronal responses of a cell with **viewer-centred properties to the body and its components in motion**. Histogram of response (spikes/sec) to different stimuli. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) of one cell (E88_35.34L) tested to view 180° and view 0° of the whole body and its components when moving 0° (i.e. towards the monkey). Two-way ANOVA revealed a significant main effect of view [$F(1,32)=130.0$ $p<0.0005$]; and for the face/body stimuli tested [$F(3,32)=7.8$ $p<0.0005$]. However, there was no significant difference between responses to component parts of the body presented in the optimal view and moving in the optimal direction (protected LSD $p>0.05$). Furthermore, there was an interaction between these factors [$F(3,32)=7.88$ $p<0.0005$]. Protected least significant difference tests (PLSD), post-hoc tests, indicated significant response discrimination between the different views of the entire body ($p<0.0005$) and of the body parts tested in isolation, head only ($p<0.0005$), body only ($p<0.0005$) and legs alone ($p<0.0005$).



This was not due to smaller responses to the component parts resulting in a non-significant difference, but rather information about the view of component parts are suggested to be coded differently than the views of the whole body stimulus.

Cells responsive to only the whole body in motion

Finally, all 4 cells selectively responsive only to the whole body moving showed viewer-centred properties of coding the effective whole body stimulus. That is, the cells responded to one view of the whole body but not to other views of the same stimulus (see e.g. Fig. 6.8).

Motion sensitive cells with object-centred properties

Only one cell (1/20) of all the cells tested showed object-centred properties. This cell generalised across different views for the whole body and the component parts stimuli (see Fig. 6.9).

View discrimination indices

a) View discrimination: Static whole body vs static head alone

The responses of 34 viewer-centred cells were used in a population analysis to compare the efficiency of view discrimination for the static whole body and the head presented alone (Fig. 6.10a). Cells were only included in this analysis if they responded to the static whole body *and* to the head when tested in isolation.

The distribution of I_v values is shown in Fig. 6.10a for the whole body and for the head alone. For an index value of 1.0 response to the worst view was the same as S/A. Index values > 1.0, arise when the cell response to the best view was greater than S/A, and the response to the worst view was less than S/A. I_v can have a negative value if a neuronal response is numerically greater for the 'worst' view than the 'best' view. This can only occur when the index is computed for the component body parts, since the best and worst views were defined on the basis of responses to the whole body.

For cells responsive to the head, the distribution of I_v for the entire body was not significantly different from the distribution of I_v when the head was tested alone (matched pairs t=-0.77, df=33, p=0.45).

Figure 6.8. Neuronal responses of a **'whole only in motion' cell with viewer-centred** properties to the whole body. Histogram of response (spikes/sec) to different stimuli. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) of one cell (E18_31.79R) tested to view 0° and view 180° of the whole body and its components moving towards 0° (i.e. towards the monkey). Two-way ANOVA revealed a significant main effect of view [$F(1,32)=12.5$ $p=0.001$]; body part tested [$F(3,32)=32.3$ $p<0.0005$)]; and interaction between these factors [$F(3,32)=16.17$ $p<0.0005$)]. Protected least significant difference tests (PLSD), post-hoc tests, indicated significant response discrimination between the different views of the entire body (the only effective stimuli, $p<0.0005$).

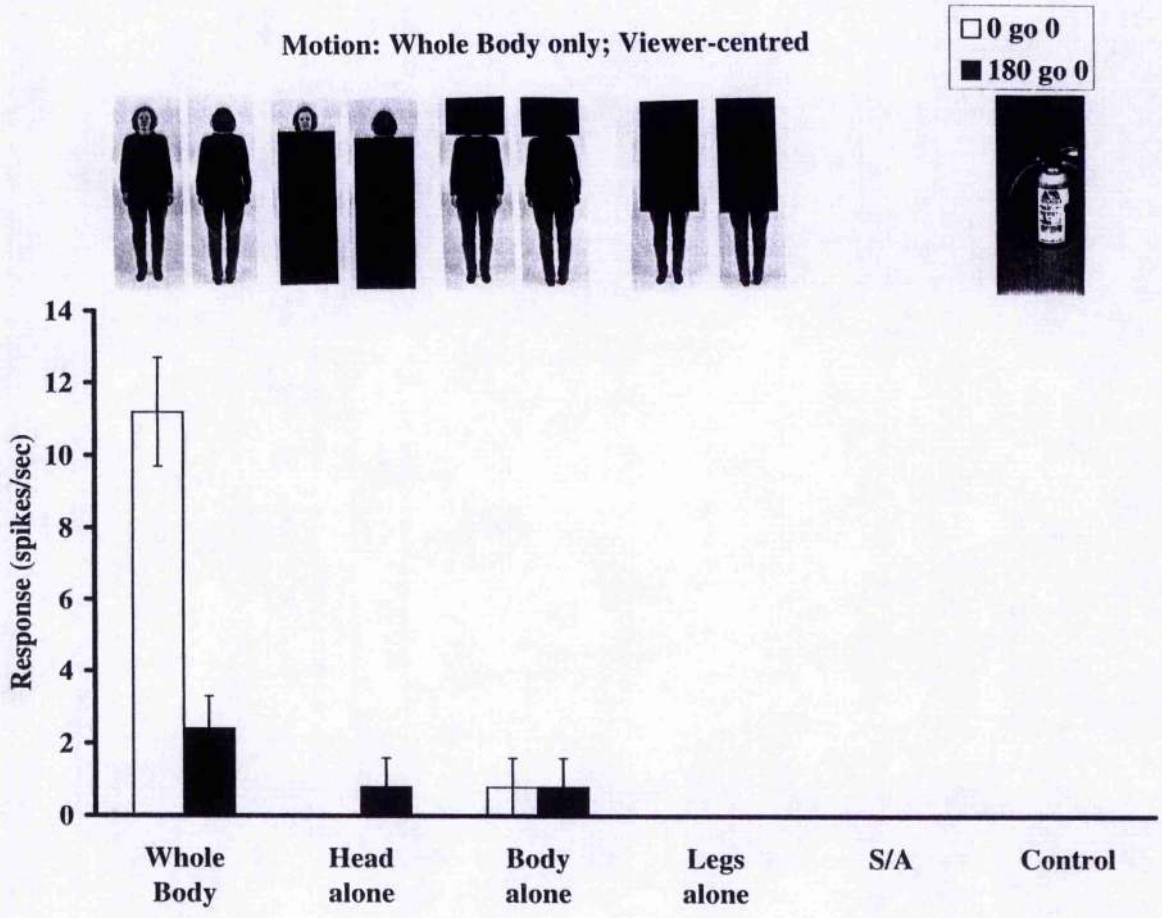


Figure 6.9. Neuronal responses of a cell with **object-centred properties to the body and its components in motion**. Upper: photographic representation of stimuli used for testing. Lower: mean responses (+/- 1SE) of one cell (E96_39.75L) to the right profile (view 270°) and left profile (view 90°) of the whole body and its components moving towards 270° are shown. 2-way ANOVA showed no significant main effect of the body part tested [F(3,30)=0.94 p=0.4] and no effect of view [F(1,30)=0.13 p>0.5] and no interaction between view and part tested [F(3,30)=0.12 p>0.5].

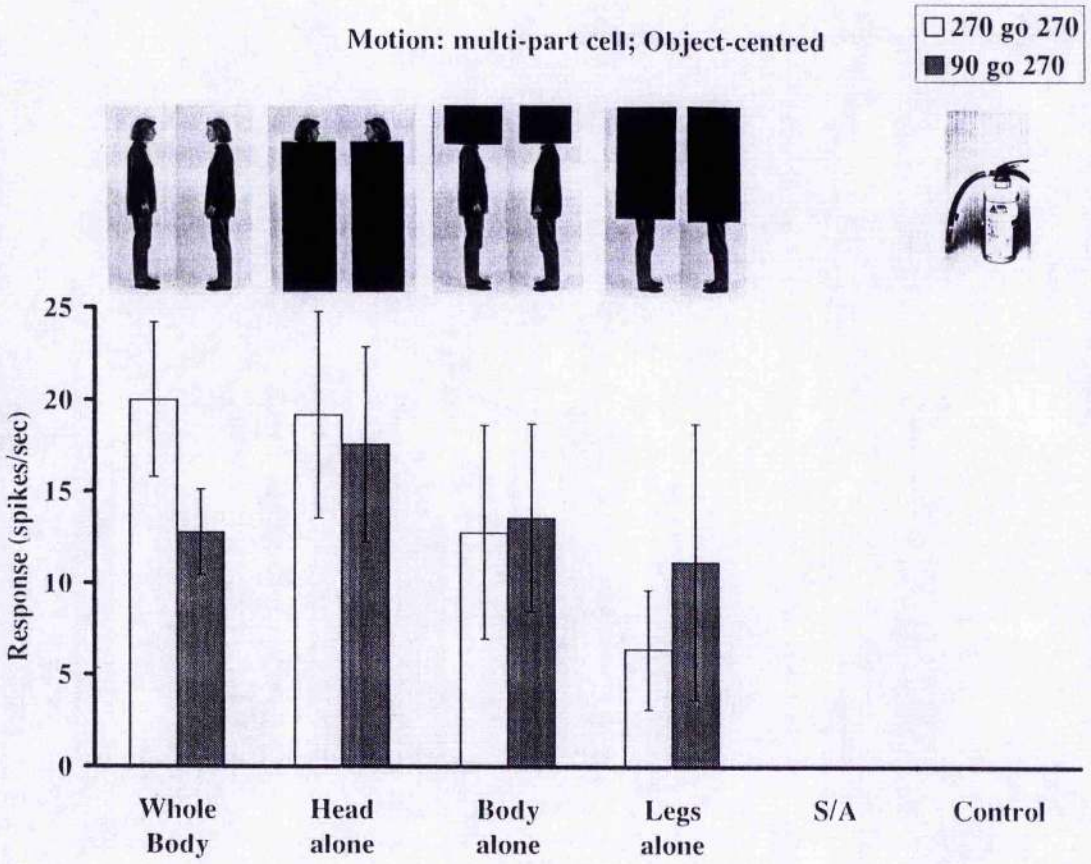
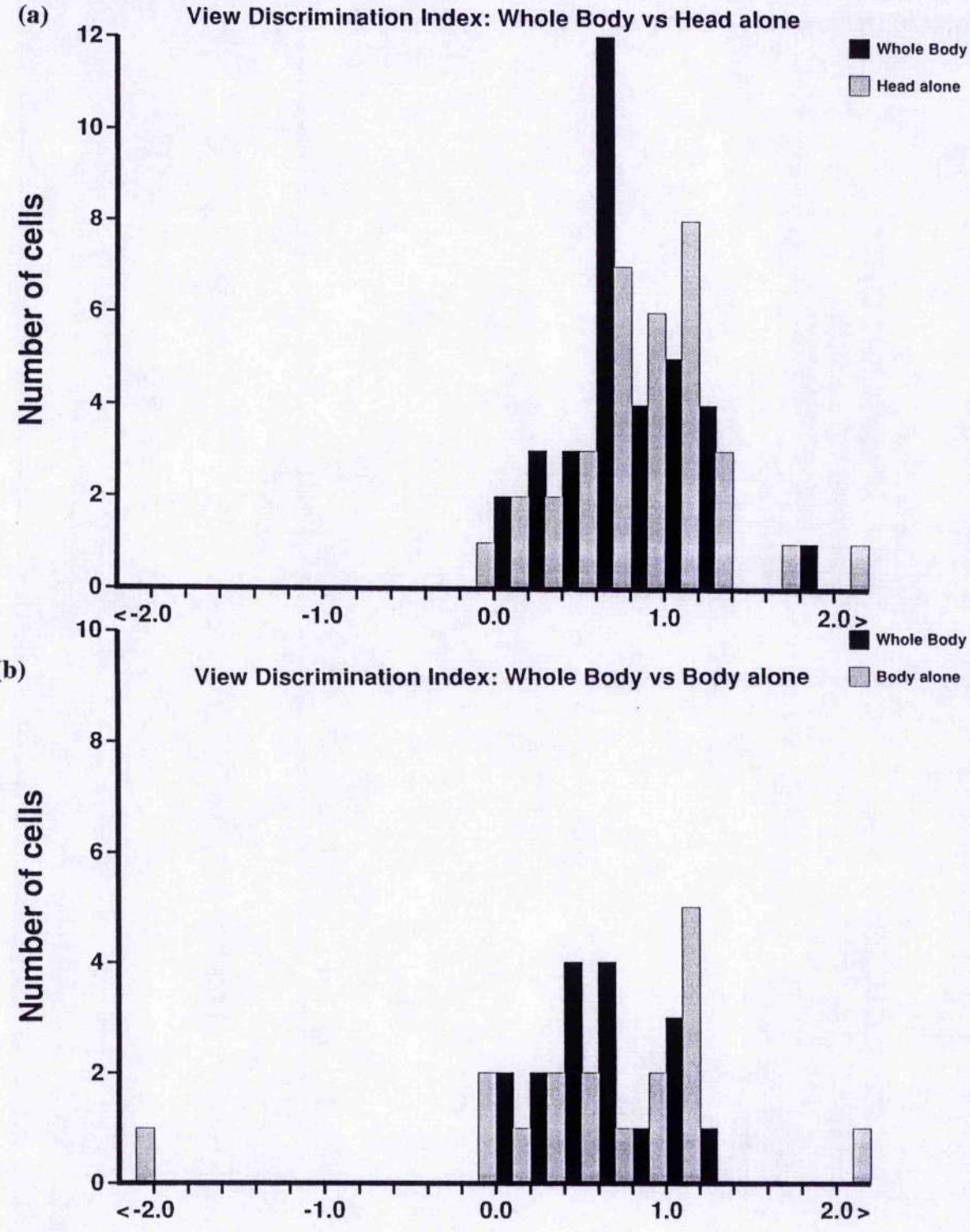


Figure 6.10. View discrimination indices. *a) Whole Body vs Head alone:* An index of view discrimination was computed (see text) for 34 viewer-centred cells responsive to the whole body and the head alone. The black bars display the ability of the cell population to discriminate between views of the whole body stimuli. The light grey bars display the ability of the same cells to discriminate between views of the head alone. The distributions of index values were not significantly different (t=-0.77, df=33, p=0.45). *b) Whole Body vs Body alone.* A similar comparison was made for 17 viewer-centred cells responding to the whole body and the body alone. The black bars display the population's ability to discriminate between views of the whole body stimuli. The grey bars display the population's ability to discriminate between views of the body only stimuli. The distributions did not significantly differ (t=-0.61, df=16, p>0.5).



b) View discrimination: Static whole body vs body alone

The efficiency of view discrimination for the static whole body and the body alone was computed in a similar way for 17 cells (Fig. 6.10b). These cells were responsive to both the whole body *and* to the body without the head visible. The distribution of I_v values obtained with the whole body visible was not significantly different to that obtained from the body alone (t=-0.61, df=16, p>0.5).

This analysis was repeated for 11 cells responsive to multiple parts and again it was shown that the distribution of I_v for the entire body did not significantly differ from the distribution of I_v for the head alone (matched pairs t=-1.39, df=10, p=0.2) nor the body alone (matched pairs t=-0.53, df=10, p>0.5).

Thus, as a population of cells there was no significant difference of quality for view discrimination between viewing the entire body or its isolated parts.

c) View discrimination: Whole body in motion vs effective body part alone in motion

16 viewer-centred cells selectively responsive to the whole body and at least one component part moving in the optimal direction were included in this view discrimination analysis. The data was analysed in the same manner as the data obtained from cells responding to static stimuli. A t-test showed that the distribution for Iv for the whole body in motion did not differ from the distribution of Iv for the effective component part in motion (t=-1.87, df=15, p>0.05). It is suggested that information about component parts of an object/body is processed in a similar manner whether the image is stationary or in motion.

Population estimates of time course responses

29 (viewer-centred and object-centred) cells selectively responsive to the *static* head/body tested for their response for at least two different views were included in a population estimate analysis (see Fig. 6.11). A PSTH was drawn indicating that as a population of cells the overall neuronal activity to the best view is always greater than the overall activity to the worst view. However, the main point of this illustration is to show that there is no difference in response onset (latency) to the whole body best view stimulus compared to the whole body worst view stimulus. Both stimuli types

Figure 6.11. Population estimates of time course responses. Cell population response to different whole body views and control objects. a) Combined responses of 29 cells to different Whole Body views: best and worst view (see text). The clear area displays the population response to the best view of the whole body stimulus, dark grey area to the worst view of the whole body stimulus and the black area population response to various control objects. Firing rate is expressed as a percentage of the peak response to the best view of the stimulus. Stimulus presentation occurs at point 0 ms.

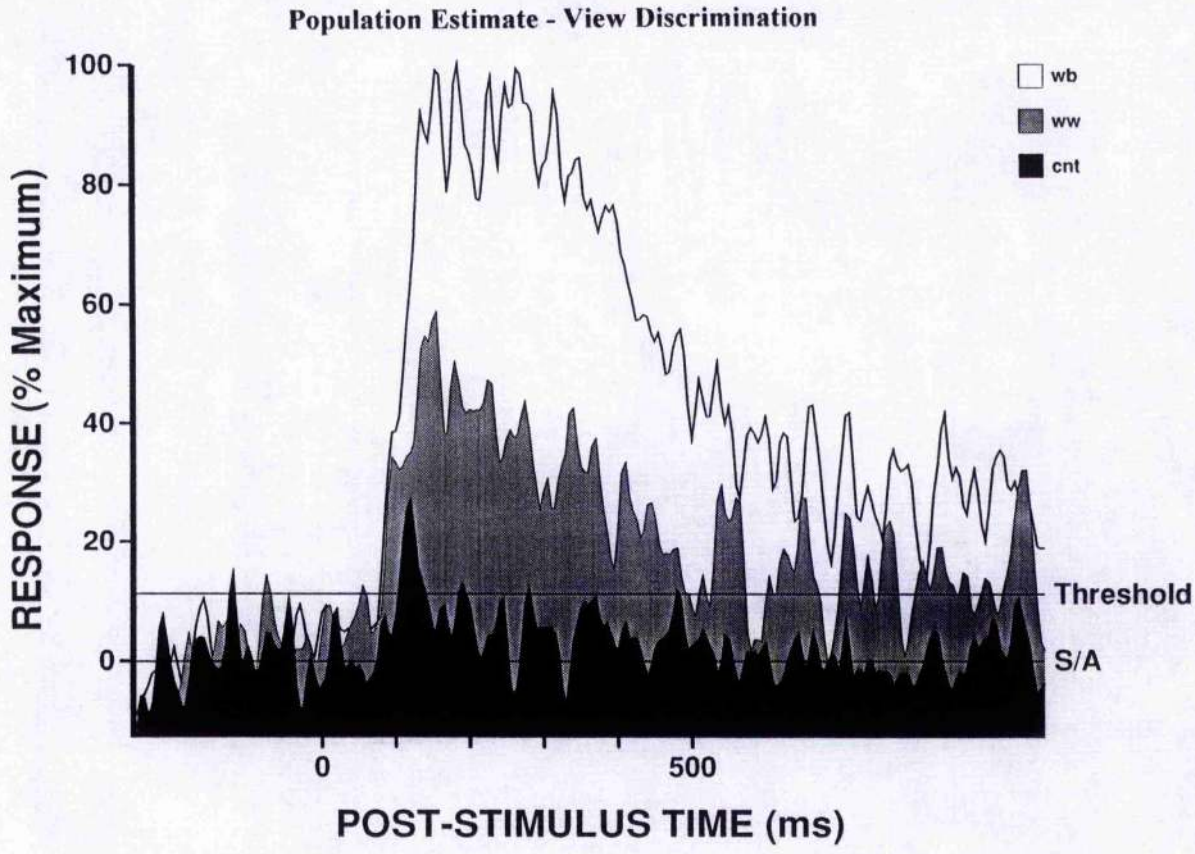
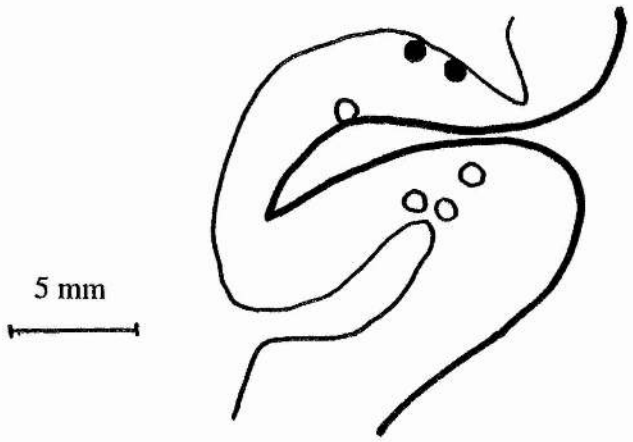


Figure 6.12. Histological reconstruction. A frontal section of subject E taken at 15 mm anterior to the interaural plane showing the location of cells coding head/body information in a viewer-centred manner (open circles) and cells coding head/body information in an object-centred manner (filled circles). The thick lines indicate the brain surface, the thin lines show the boundary between white and grey matter. Cells were located in both the upper bank and fundus of the superior temporal sulcus (see Appendix 3).



evoke a response which reaches threshold (95% confidence level, see chapter IV) at approximately 90 ms after stimulus presentation.

Histological Localisation

The reconstruction of cell position as described in chapter IV and V indicated that the cells tested were located in the upper and lower bank of the anterior STS. Viewer-centred cells were present in both banks and throughout the length of the STS sampled (see Appendix 3 or Fig. 6.13 for reconstruction of one monkey E; a total of 23 cells classified). It is interesting to note that the small number of cells coding information about the static head/body in an object-centred manner were mainly found at quite anterior locations.

DISCUSSION

Before the empirical data on view discrimination reported above is discussed, it is pointed out that cell responses to whole bodies and component parts in motion perform view discrimination in a very similar way as cells selectively responsive to the static head/body. Therefore, view discrimination for the entire cell population, i.e. cells selectively responsive for static and moving heads/bodies, are discussed.

Object-centred coding

The analysis of view sensitivity revealed that representation of the body and its parts within area STPa occurs mostly in a viewer-centred fashion (discussed below). Only 10% (7/73) of the cells studied were found to respond to all views of the body that were tested and can therefore be considered as coding in an object-centred manner. An object-centred cell may receive multiple inputs from cells with viewer-centred properties, accounting for the responses of object-centred cells. It is possible that such a pooling mechanism resulting in object-centred coding occurs in areas subsequent to the STS. One would therefore expect to more frequently encounter cells with object-centred properties in areas which receive input from the STS or other areas with cells which have predominantly viewer-centred properties. Such areas might not be exclusively visual areas, such as the amygdala and the hippocampal complex.

Viewer-Centred coding

The remaining 90% (66/73) of the STPa cells studied were selective for view. This result is in marked contrast to models which employ object-centred representations since these representations should be accessible from all views (e.g. Marr and Nishihara, 1978; Biederman, 1987; Lowe, 1987; Porrill, Pollard et al., 1988; Hasselmo, Rolls et al., 1989; 1994; Logothetis et al., 1994; 1995).

It emerged that different cells were selective for different views of the whole body. This parallels the observation that different cells in the STPa are tuned to a range of head views (Desimone, Albright et al., 1984; Perrett, Smith et al., 1985; Hasselmo, Rolls et al., 1989b; Perrett, Oram et al., 1991). Thus, no single canonical view of the body can account for or represent the entire range of view tuning observed (Oram and Perrett, 1994a). The results are therefore consistent with models suggesting multiple viewer-centred representations stored in memory (Koenderink and van Doorn, 1979; Tarr and Pinker, 1989; Ullman, 1989; Edelman and Bülthoff, 1990; Poggio and Edelman, 1990; Seibert and Waxman, 1991; Cutzu and Edelman, 1992). As noted in chapter V, the results are consistent with a view sensitive representation of parts, rather than a view sensitive representation of the global form which is common to most viewer-centred models.

View discrimination occurs for both the whole body and its effective parts

The present study did not set out to define the distribution of view tuning amongst cells responsive to the body. Nevertheless, view tuning for the whole body appears to have similar properties to view tuning for the component parts. For a minority of cells (6/73 cells, 8%) view tuning was apparent for the whole body but was not present for the effective component part(s) which provoked responses when tested in isolation. As a population of cells, analysis showed view discrimination (as measured by the index) exhibited to the whole body which was equivalent to the view discrimination manifest to the head alone or to the body alone. That is, there was no improvement in quality of view discrimination for the entire body, compared to a situation in which only one body part was visible. The view preference exhibited by cell responses to the whole body generalised to the responses to the effective component part (e.g. Fig. 6.1 and 6.3).

Associations and learning configurations

It was apparent in the present study that cells tuned for multiple parts exhibited compatible view tuning for these parts. If such a cell responded more to the front than the back view of the head, then the cell was also more responsive to the front than the back view of the rest of the body. This compatible view tuning supports the speculation that sensitivity to multiple body parts arises through experience of the association between the parts. For example the face is seen in association with the front view of the body. If cells learn the association of parts, then it follows that the cells will tend to be selective for the same view of these parts (assuming that the parts are normally to some degree co-aligned).

Population estimates of time course responses

The overall response pattern of the entire cell population which was selectively responsive to the static head/body was investigated. From the PSTH it is evident that cells respond equally fast to the whole body in its preferred view than to the whole body in its non-preferred view. If mental rotation (Shepard and Metzler, 1971) occurred at a neuronal level, i.e. a rotation process was applied to the incoming image or the representation stored, then one would expect to find that the response onset to the best view would be earlier than the response onset to the worst view (the view to be transformed/rotated). This time difference in latency could be argued to be due to the mental rotation process. However, no such time difference was observed and therefore, the data presented here does not support models suggesting mental rotation.

Chapter VII

EFFECT OF ROTATION ON OBJECT RECOGNITION

INTRODUCTION

Rotating an object in the picture plane does not influence the relative positions of the object's parts visible from one perspective view. The overall direction of the parts, however, relative to the observer, and relative to the gravitational and scene-based axes does change. For example, if the whole body is viewed upright and from the left profile, then the nose points to the left of the viewer. If, however, this view of the body is inverted, even though the body remains in left profile view, the nose now points towards the right of the viewer. If a representation of an object present in higher visual areas can be activated independent of the view and orientation relative to the observer, then such a representation will be referred to as *object-centred.* On the other hand, if the representation of an object is preferentially activated by a specific orientation, then the representation will be referred to as *viewer-centred.*

Behavioural studies in Humans

Short term memory

Simultaneous matching of shapes

Shepard and Metzler (1971) presented subjects with two shapes simultaneously. Subjects were asked to determine whether the shapes were the same or were mirror images of each other. All subjects reported that the method used, was to imagine a three-dimensional rotation of one of the objects and thereby matching it to second image. It was found that reaction times (RTs) to perform such a task increased as the angle between the first and second image increased. This increase of reaction time was taken as an estimate of the time to 'mentally rotate' one image in such a way that it could be directly matched against the other image. Hence, the larger the angle of rotation between the two images, the longer the reaction time needed for mental rotation to be completed.

Shepard and Metzler (1971) therefore suggested that mental rotation occurred at a speed of 60° per second. It should be noted, however, that inverted images of most

familiar objects do not require 3 seconds for recognition. Thus, mental rotation phenomena (at this speed) appear too slow for general object recognition. Moreover, Corballis and Nagourney (1978) argued that if mental rotations did occur, one would have to expect that if the orientation of the viewed object is changed relative to the orientation of the stored representation, there would be a linear function relating stimulus orientation to recognition times. RT/orientation functions are, however, often non-monotonic, with 180° rotations (inverted images) being processed faster than 120° rotations (Jolicoeur, 1985; McMullen and Jolicoeur, 1990; 1992).

Successive matching of shapes

Tarr and Pinker (1989) asked subjects to learn a series of letter-like characters. The subjects were then shown one character presented in a different orientation from the trained character and had to recognise the image as the same or as a mirror reflection of the trained image. Initially, the greater the angle of rotation between the learned image and the test image, the longer the RTs, however, a practice effect did arise (see below). They also found that recognition of characters that were a mirror reflection from the trained images, was performed with equal efficiency independent of the orientation of the stimulus. That is, the time taken to match an upright letter-like object to its upright mirror image was the same as the time taken to match an inverted letter-like object to its inverted mirror image. It should be noted that mirror images preserve the orientation of many of the object's component features, thus matching mirror images and matching across a rotation of >45° may involve fundamentally different processes.

Long term memory

Identification and naming tasks

Jolicoeur (1985) measured recognition times of line and colour drawings of common objects presented in different orientations. RTs increased as the angle of rotation increased from the 'upright' orientation (i.e. the orientation in which the object was most commonly experienced). Jolicoeur (1985) therefore suggested that the representation for most objects is viewer-centred and that the upright orientation is 'canonical'. Similarly Palmer et al. (1981) suggested that an object can have a

canonical perspective view which maximises the visibility of the object's most salient features.

This suggestion that most objects have a canonical orientation is supported by neuropsychological studies. Turnbull et al. (in press) describe a patient who shows difficulties in establishing the canonical orientation of objects, though shows little impairment in recognising objects. This is independent of the patient's ability to perceive orientation, since she is able to correctly carry out orientation matching tasks even matching the orientation of two objects.

Practice effects

Jolicoeur (1985) also found that after practice recognising all test orientations, subjects became less affected by the stimulus orientation and that the speed of recognition of all non-upright orientations was similar. Familiarity with an object in multiple orientations, therefore, diminishes both the effect of rotation on object recognition and the canonical nature of the upright orientation. This might be interpreted as the speed of mental rotation becoming faster with practice. Modification of the speed of mental rotation, however, did not generalise to untrained objects.

McMullen and Jolicoeur (1990; 1992) presented images of line-drawn objects and asked subjects to either name the object or to locate a dot which was placed near the top or the bottom of the object. A linear increase in RTs for both tasks was found when the image was rotated away from the upright orientation (for 0° to 120° rotation). The effect of orientation on object naming decreased as images were viewed more frequently. One can speculate that experience in recognising one object (at different orientations) could result in the establishment of a greater number of orientation specific (viewer-centred) templates for that object, and consequently the angular distance required for the matching process would be decreased, reducing all RTs. Nonetheless, such practice effects were less prominent when the subject had to indicate the top and the bottom of rotated objects (dot location task). In this task, the linear effect of stimulus orientation on top-bottom discrimination remained despite practice. McMullen and Jolicoeur (1992) therefore argued that an object-centred representation cannot be used to carry out the top-bottom discrimination task[3]. Instead, knowledge

[3] Most object-centred representational schemes use the 'major axis' of an object as an internal spatial reference for describing the disposition of object parts (Marr and Nishihara, 1978; Biederman, 1987).

about the location of object features (or the orientation of the object's principle axis) relative to the viewer is critical for determining the top and bottom of objects.

To account for these effects one needs to differentiate recognition of whether an object is upright with respect to self or gravity and recognition of an object without regard to its orientation. These two tasks are differentiated by the studies of McMullen and Jolicoeur (1992).

Models explaining rotation effects

Rock (1973) suggested that recognition of disorientated objects occurs by first identifying the orientation of the object (principal axis). In this theory the object is described relative to the environmental and gravitational upright (see also Cooper and Shepard, 1973; Jolicoeur, 1985). An object becomes increasingly difficult to recognise as it is rotated away from an upright position. This is because it is increasingly harder to assign the specific orientation to the object. Rock argues that the effect on recognition ability caused by changes of orientation in the picture plane is greater than changes of orientation in the depth plane (i.e. change in perspective view). This is because change of view occurs independent of the direction of the object relative to the environmental and gravitational upright.

To account for recognition of the object rather than the object's orientation several authors (e.g. Jolicoeur, 1985; Tarr and Pinker, 1989; Ullman, 1989; 1992) suggest that the retinal image produced by an object is matched against a viewer-centred representation stored in memory applying an alignment process. Different theories suggest different numbers of stored (view and or/orientation specific) representations. Palmer et al. (1981) and Jolicoeur (1985) stress that only one single (canonical) view and orientation of the object is needed for the representation of an object, whereas others (Koenderink and van Doorn, 1979; Perrett, Smith et al., 1985; Tarr and Pinker, 1989; Ullman, 1989; Edelman and Bülthoff, 1990; Cutzu and Edelman, 1992; 1994; Logothetis, Vetter et al., 1994) suggest that there are several views represented and stored in memory. Most theorists would agree that stored representations are based on views and orientations with which the observer is familiar. Some models which include multiple representations of different views do

not require multiple representations of different image orientations and sizes (Seibert and Waxman, 1991; 1992a ,b).

Jolicoeur (1992) suggests that the orientation of an object's representation is specified using a retinal co-ordinate system (an object is upright when it is aligned with the naso-temporal division of the retina). If the object being viewed is not in its upright orientation, then the visual system either applies mental rotation to the image, or processes the image on the basis of certain critical features of the object. These two methods can also work in parallel. Mental rotation in the picture plane transforms the image using the shortest 2D rotational path possible to achieve a match with the canonical upright representation. Tarr and Pinker (1989) suggest that several representations are stored in memory (each representation corresponding to one specific orientation commonly experienced). An input image is matched to the canonical orientation or the closest matching representation possible (through mental rotation).

One major problem with transformation models is that they assume some kind of recognition before the correct transformation process can be carried out. As Corballis (1988) points out, the visual system cannot rotate something to its canonical upright position without knowing what the object is, since without knowing the object one does not know what its canonical orientation is. To overcome this problem, Tarr and Pinker (1989), Jolicoeur (1992), Ullman (1989) acknowledge that specific details/characteristics or key features of objects are identified prior to matching with stored representations (using mental rotation or other transformation processes). The nature of these characteristics or features and how they are determined remains a weakness in these models.

Unlike the 2D approaches of the above models, Lowe's (1987) model of object recognition is based on matching the input images to stored 3D representations. This matching process is founded on the spatial organisation of the object's edges and corners etc. As with the model of Marr and Nishihara (1978) there is no reason that the object's orientation in the picture plane should affect the efficiency of matching. However, processing in both object-centred models can be more difficult when the object's main features or the object's principle axis is obscured due to an unusual perspective view (e.g. foreshortened axis).

Interpolation models (Poggio and Edelman, 1990; Edelman and Weinshall, 1991; Cutzu and Edelman, 1992; Intrator and Gold, 1992) use multiple 2D descriptions of an object from different perspective views. These 2D views are stored as clusters of views in a 'representational space'. When sufficient views are stored, the appearance of any view falling in-between those stored, can be specified by interpolation between (or a linear combination of) those views that lie adjacent to the input view. This is perhaps most easily imagined in the orientation domain. For example, the appearance of an object rotated 45° from upright could be interpolated from two stored descriptions: one where the object is represented in its upright position and a second description where the object has been rotated 90°. The effectiveness of matching is then judged on how well the input image fits a stored view or the interpolation between stored views. Interpolation models do not need to perform any processes akin to mental rotation (linear transformations or normalisation) in order to match an image to stored representation. Increased RTs for unusual orientations or views occur because these views lie between stored views and require more complex or time consuming interpolation.

With a sufficient number of 2D descriptions contained in views the same amount of information about an object can be specified as that by a single 3D description (Poggio and Edelman, 1990). Models utilising multiple views also have the flexibility to expand the number of stored views/orientations representations. That is, if an unusual orientation of an object becomes experienced due to viewing practice a new representation of that view can be added to those already stored. This leaves the problem of how one builds or links a collection of 2D representations together to represent the same object. Though such links could depend on continuity of experience (Foldiak, 1991; Perrett, Oram et al., 1991; Seibert and Waxman, 1991; Logothetis, Vetter et al., 1994; Oram and Perrett, 1994a).

Behavioural studies in other species: monkeys, apes and pigeons

Generalising across orientations

At the behavioural level it is apparent that image orientation affects visual processing in monkeys. Monkeys can discriminate a change in orientation in the picture plane of a simple grating pattern of 2-7° depending on the task (Vogels and

Orban, 1994). Holmes and Gross (1984a, b) show that normal monkeys can differentiate *objects* rotated through an angle of 30°.

Of more relevance here is the capacity of animals to generalise across orientations. Monkeys can learn to discriminate the difference between pairs of faces and having done so show generalisation across a change in stimulus orientation during discrimination (Rosenfeld and van Hoesen, 1979). Thus rotation in the picture plane does not prevent face recognition in normal monkeys (Ockleford et al., 1977; Dittrich, 1990), although it may influence the way in which faces are processed.

Inversion effects

Several studies (Bruce, 1982; Overman and Doty, 1982) have claimed, that monkeys process information about faces in a different way from humans. Bruce (1982) claimed that in contrast to humans (see later; Yin, 1969; 1970), monkeys show no inversion effect for recognising and remembering faces. That is, monkeys recognise upright and inverted faces with equal ease.

Bruce explored a variety of tasks, where pairs of stimuli were presented simultaneously in an upright or inverted orientation. After training discrimination of a given stimulus pair to a criterion of 90% correct, new images of the same stimuli were presented with the expression, size or lighting of the original image changed (transfer stimuli). No significant difference in the learning phase was observed when the face images were presented upright or upside down. In addition, monkeys were able to recognise the transformed face stimuli and inverting the image did not affect performance in transformation tasks. Bruce (1982) concluded that monkeys do not show an inversion effect, and that this lack might be attributed to the "macaque's precocious development, smaller cortex, and lack of hemispheric specialisation".

It has also been pointed out (Bruce, 1982) that the representation of faces may occur predominantly in cortical areas present in humans, that are not present in the macaque monkey (though Bruce did not specify which areas), accounting for the differences of the presence of an inversion effect. The human cortical mantle is several times greater in size than the cortical mantel of the monkey. However, if evolutionary development of the human brain was responsible for the visual inversion effect, why lose the ability to generalise across orientations? Surely, a primordial face recognition

system developed in ancestral primate species would be more effective if the system can recognise faces equally well and fast when presented in any orientation, than a system that recognises one orientation better than another. Moreover, neuronal mechanisms for processing faces appear similar in macaques and humans in terms of orientation sensitivity.

Bruce's (1982) speculation was based on several earlier findings. Young humans (children under 10 year old) do not exhibit an inversion effect when faces are viewed at different orientations, as has been observed in human adults. That is, children can remember faces equally well irrespective of the orientation of the image. The ability to establish representations (for face recognition) that are orientation- and shape-specific appears to develop only around the age of ten (Carey and Diamond, 1977). In addition, this orientation specific representational system appears to be lateralised to the right hemisphere. This was suggested since the right-hemisphere advantage (left visual field presentation superior to right) for faces is first noted in human children who are approximately 10 years old (Carey and Diamond, 1977). If the right hemisphere face processing mechanisms fails, other mechanisms may be able to carry out the task of face recognition (Yin, 1969; 1970; De Renzi et al., 1994; Rapcsak et al., 1994). Such mechanisms might be similar to those applied by young children, e.g. using a 'piece-meal' strategy which would not be sensitive to inversion (Carey and Diamond, 1977).

More recent studies indicate that both the inversion effect and a hemispheric lateralisation of face processing occurs in children of younger ages (Flin and Dziurawiec, 1989; de Schonen, 1992). Such evidence calls into question Bruce's (1982) interpretation of face processing mechanisms.

Experimental and clinical studies have suggested that the human brain is designed to process information about the upright face and that the right hemisphere is the main site for this information processing. Studies in monkeys, however, suggest that face processing does not show any hemispheric asymmetry in other primates (Overman and Doty, 1982, though see Hamilton and Vermeire, 1988), but rather takes place in both hemispheres.

However, monkeys might process information about upright and inverted faces in the same way. These claims were assessed by Perrett et al. (1988) who used

behavioural and physiological methods to study processing of face configuration at different orientations.

In one task, Perrett et al. (1988) measured the behavioural response (response times) of monkeys to discriminate between faces (presented in different orientations) and non-faces (images of objects and scenes). Images were presented successively in a GO-NOGO paradigm where the monkey was trained to lick for fruit juice reward when a face was presented and to withhold licking when a non-face stimulus was presented. The results showed no effect of orientation of face images on the behavioural RTs. The monkeys, however, may have been able to solve the task using a 'piecemeal feature detection strategy' and the inversion effect may depend critically in utilising configurational clues. In a second task therefore, monkeys were trained to discriminate normal from jumbled face configurations stimuli, presented in upright, horizontal and inverted orientations. Once the monkeys reached a criterion level of 90% correct (on upright normal vs jumble face discrimination) the RTs to all stimulus orientations were recorded. It was found that RTs increased as the image was rotated away from the upright position. A similar experimental paradigm revealed comparable effects of stimulus orientation on the ability to discriminate normal and jumbled faces for human subjects. Thus, humans and monkeys both show an 'inversion effect' when processing configuration of stimuli that have been experienced predominantly in one (upright) orientation.

In a further experiment (Perrett, Mistlin et al., 1988) monkeys were retrained on further face configuration judgements, but this time with a new set of face and jumbled face stimuli all presented in an inverted orientation. Having pre-trained on inverted orientations for normal and jumbled faces, the target face in a horizontal orientation (i.e. rotated 90° from the training orientation) was still processed with a delay (increased RTs)[4]. Interestingly, any interpretation based on mental rotation of stimuli to the gravitational upright orientation would predict that the effects of this pre-training would speed up decisions about horizontal faces. An interpretation that is favoured here is that extensive experience with a complex object which occurs predominantly in one orientation, leads to specific neural 'templates' for that object. Some of these templates may be selective for particular features (parts) of the object, whereas other

[4] Equivalent effects have been found for human subjects trained to recognise unfamiliar faces in specific non-upright orientations (Tarr, 1995; personal communication).

templates may be sensitive to the combination and configuration of parts. Both types of template are likely to be maximally sensitive to the orientation most frequently encountered.

The face inversion effect or any decrement in performance for objects presented in unusual orientations, may simply reflect the lack of experience at particular orientations which in turn will cause a lack of relevant orientation specific templates. Of course, some face classification tasks can be solved by reference to individual distinctive features or object parts (Rapcsak, Polster et al., 1994). In these tasks, orientation (normal or unusual) of the image is irrelevant. (N.B. Parts of objects can be complex items made up of several features; some templates coding parts are presumably also sensitive to configuration.) The sensitivity to the orientation (and configuration) may underlie the Thatcher Illusion (Thompson, 1980). In this illusion an inverted mouth embedded in an otherwise upright face appears 'normal' when the entire image is viewed rotated by 180° in the picture plane, presumably because the features can now be recognised in their gravitationaly normal orientation.

Phelps and Roberts (1994) compared the memory of humans, squirrel monkeys and pigeons for upright and inverted faces. Monkeys and pigeons were presented with a sample image (upright or inverted) and immediately afterward with a pair of images of which one was the same as the sample image. Both test stimuli were presented in the same orientation as the sample stimuli. Humans, on the other hand, had to first view lists of 24 (upright or inverted) sample images before being tested on pairs of images of which one matched a sample image. Again, the sample and test stimuli were presented in the same orientation. Using a collection of human faces as stimuli, both monkey and human subjects were less accurate at recognising inverted stimuli compared to upright stimuli.

In addition, Phelps and Roberts (1994) tested humans, squirrel monkeys and pigeons on recognition of other classes of primate faces (human, great ape and a variety of new and old world monkey species). Colour images were presented either in their upright or inverted orientation. Squirrel monkey and human subjects were significantly more accurate in recognising upright human and ape faces compared to the inverted versions. Thus, both species found it harder to recognise upside-down faces (a face inversion effect). For monkey faces and outdoor scenes there was no

effect of stimulus orientation on recognition accuracy. Pigeons, on the other hand, showed no effect of orientation on the recognition of any class of stimuli. It is interesting to note a practice effect where monkeys improved at matching the inverted human and ape face images until accuracy for recognising inverted images was at the same standard as for upright images. Thus practice removed the advantage of upright orientation.

The behavioural evidence from this study and from that of Perrett et al. (1988) indicates that both monkeys and humans show a face inversion effect and possess similar orientation sensitive systems for recognising faces. The effects of stimulus orientation are particularly apparent in both species where, for recognition of the stimulus, processing of the configuration is required (Perrett et al., 1988). Pigeons by contrast, appear to use recognition strategies that are less sensitive to stimulus orientation.

Neuropsychological studies in humans

It has been suggested that orientation-independent object recognition is accomplished by mental rotation (Shepard and Metzler, 1971; Jolicoeur, 1985). Farah and Hammond (1988) on the other hand argue that orientation-invariant object recognition does not rely on mental rotation processes. They tested a patient with a stroke in the territory of the large right middle cerebral artery on his ability to recognise rotated objects and carry out mental rotation tasks. It was found that even though the patient was able to recognise objects presented in various orientations he was unable to carry out mental rotation tasks.

Another patient described by Turnbull et al. (in press) shows that recognition of objects can occur without knowledge of the orientation of the object with respect to gravity. This patient seems unable to indicate the canonical orientation of an object. Turnbull et al. (in press) suggest (following Milner and Goodale, 1993) that there may be two pathways in normal object recognition, the ventral-pathway which displays object-centred properties and the dorsal pathway which shows viewer-centred properties. Equally, such observations could indicate orientation-dependent viewer-centred descriptors present in the ventral system with access to view-independent

descriptors and semantic associations, but with no access to parietal systems necessary for finding orientations with respect to egocentric or gravitational axes.

Human studies have shown that when a face is inverted, subjects have much greater difficulties in recognising and memorising the face than when the face is presented in its upright position. Yin (1970) studied the recognition of upright and inverted unfamiliar faces in patients with focal brain damage. Each subject first viewed a set of 40 singly presented upright faces, then was tested on a series of 24 upright image pairs, having to indicate which of the two images he/she had seen previously. The experiment, was repeated with new set of faces presented upside-down. Subjects with posterior right hemisphere lesions performed at a lower rate than normal subjects or patients with other unilateral brain injury (3 left anterior, 7 left posterior and 3 right anterior) when face images were in the upright orientation. When the images were inverted, all subjects made more errors than with upright images. Surprisingly, patients with posterior right hemisphere lesions did as well as the normal control subjects, whereas patients with other unilateral damage showed greater problems recognising inverted images compared to control subjects.

It is interesting to note that the dissociation did not occur for a second type of object (houses). Yin (1969; 1970) concluded that the difficulties in recognising and memorising inverted faces were two fold: "a general factor of familiarity with mono-orientated objects" (objects normally viewed in the upright orientation) and "a special factor related to faces" (Yin, 1969). In addition, Yin suggested that the right hemisphere mainly deals with the processing of upright faces, since brain damage of the right hemisphere only influenced the processing of upright but not inverted faces (Yin 1970).

Yin (1969; 1970), Carey and Diamond (1977) and DeRenzi (1994) have suggested that in humans an upright face is processed in terms of its feature configuration, whereas an inverted face is processed in a piecemeal manner, with each feature analysed separately. One can speculate that the inversion effects (increased reaction time or decreased accuracy of processing) are dependent on configuration coding and that in humans the right posterior cortex is specialised for face processing using configurational cues (Perrett et al., 1988).

Ventral lesions: Visual agnosia

A patient (DF) who suffered damage to posterior and mainly ventral areas (though she has spared striate areas) and damage to the parasaggital occipitoparietal region was studied (Milner et al., 1991). DF suffers from visual form agnosia and showed a profound impairment in recognising the orientation and size (see chapter VIII) of stimuli. She could not match the orientation of two stimuli presented simultaneously or accurately apply the verbal labels 'horizontal' or 'vertical'. Unexpectedly, DF was able to carry out a 'posting' task in which a hand held card had to be fitted into a rectangular slot of different orientations. She was unable, however, to indicate the orientation of the slot without performing the motor movement (Milner et al., 1991). Thus, orientation appears to be processed in DF's dorsal stream and capable of guiding visuo-motor co-ordination when dealing with objects.

Dorsal lesions: Optic ataxia

Patients with more dorsal lesions (i.e. to parietal cortex) show a converse pattern of deficits. These patients show no visual form agnosia and can verbally label the orientation of objects but are impaired in carrying out visuo-motor tasks. A patient (VK) who suffered bilateral damage to the parieto-occipital lobe (Jakobson et al., 1991) was studied. VK's performances were significantly different from control subjects when carrying out a prehension task. In contrast to DF, VK was unable to adjust her grasp appropriately to the size and the orientation of the object to be picked up. Patients with dorsal stream damage are able to recognise complex line drawings and recognise the correct orientation of objects, but cannot make the correct prehension movement towards the object (Jeannerod, 1981; Perenin and Vighetto, 1988). These studies indicate that the mechanism which underlies conscious perception of an object's qualities such as size and orientation, is quite distinct from the mechanism underlying visual guidance of prehension.

Neuropsychological studies in monkeys

Ventral lesions

Lesions in monkeys to the occipito-temporal areas but not the dorsal parietal areas produce a profound impairment in the ability to learn and discriminate objects

(for reviews see Dean, 1976; Ungerleider and Mishkin, 1982). These findings, however, do not specify whether the ventral or dorsal visual pathways are involved in the capacity to generalise across different viewing transformations, or are just involved in object learning. Several studies have addressed this question of the involvement of temporal cortex in object constancy more directly (Humphrey and Weiskrantz, 1969; Ungerleider and Brody, 1977; Holmes and Gross, 1984a, b; Weiskrantz and Saunders, 1984; Walsh et al., 1992).

Weiskrantz and Saunders (1984) trained unoperated control monkeys and monkeys with varying lesions to parietal cortex, foveal prestriate cortex [combined V4 and posterior inferotemporal cortex (PIT)], anterior inferotemporal cortex (AIT), and posterior superior temporal sulcus (pSTS), on a number of visually presented 3D objects each seen under constant conditions. They then transformed the appearance of objects either in size, 3D orientation (a compound change of perspective view and orientation with respect to gravity), or lighting (shadows). The monkeys were then tested on their ability to generalise recognition across the transformed views of the training objects. Testing continued afterwards with transformed views trained as new targets.

Normal unoperated monkeys were to some degree impaired on the initial trials of generalisation testing (transformation of orientation produced 36% errors). This performance may be better than chance because in most problems particular features, such as object colour, could be used to aid generalisation (across orientation and other transformations). With continued testing, normal subjects performance with orientation transformed objects improved to an average of 28% errors.

Monkeys with foveal prestriate and AIT lesions were found to have great difficulty in the initial learning of untransformed objects. Training continued for these monkeys on the initial problems to criterion (90/100 correct). In one experiment both foveal prestriate and AIT lesion groups showed inferior performance relative to parietal cortex lesions on generalisation across orientation. This generalisation deficit was not only present on initial transfer trials, but was also maintained across the subsequent 10 trials. In a separate experiment, generalisation across orientation was not markedly impaired for monkeys with prestriate and AIT lesions relative to normal controls during the first trials of generalisation (AIT 55%, Prestriate 33%, Normal 36%

errors). Nonetheless, generalisation performance for both ventral lesions (AIT and prestriate) was significantly impaired relative to normal controls after extended testing. N.B. during this extended testing all monkeys may have been treating the novel transformed versions of the target as independently rewarded patterns. One cannot be sure that the monkeys were aware that the transformed versions of one object were different instances of the same object: at no stage did the monkeys witness a continuous transformation between alternative viewing conditions.

To summarise, normal monkeys do not show perfect ability to transfer learning across orientation transformation. Monkeys with prestriate or anterior inferotemporal lesions have problems learning about objects generally. Both these groups are impaired on generalisation across orientation once they have learned a standard orientation problem. It should be noted that the impairment is relatively mild compared to normal animals, the two ventral system lesions produce greater impairment in orientation generalisation than parietal or posterior STS lesions. The mildness of the impairment in the study of Weiskrantz and Saunders may be due to the tasks being soluble using transformation independent cues such as colour or particular shape features (pointed, or jaggy).

Since visual information reaches IT cortex by travelling through the prestriate regions (V4 and PIT), foveal prestriate damage will automatically deprive the IT cortex of visual information. Hence, it should be expected that foveal prestriate and AIT damage both produce deficits in learning about objects. One might also expect equal deficits in generalisation performance. It is argued here that one needs experience with multiple instances of the particular features or shapes, and the transformations between these to build representations that generalise across change in size or orientation. In the experiments of Weiskrantz and Saunders (1984), subjects did not experience continuous transformations of individual objects. The subjects could perhaps rely on preoperative experience of orientation and size transformations of particular features and shapes. If orientation and size generalised representations of such features are located in AIT cortex, then they would be available for normal monkeys and monkeys with pSTS or parietal lesions, but inaccessible for subjects with foveal prestriate lesions and directly damaged AIT.

Weiskrantz and Saunders (1984) interpret their results as indicating the posterior part of the IT cortex being involved in defining viewer-centred information about features of objects and the anterior parts of IT cortex containing generalising (object-centred) representations. The authors argued that the results were consistent with representations used for object recognition becoming less view specific (and more prototypic, object-centred or generalising) as one moves anterior in the temporal cortex. The gradient of this progression is, however, not defined by the lesion study of Weiskrantz and Saunders (1984); viewer- and object-centred representations could both exist within anterior temporal regions.

Holmes and Gross (1984b) examined generalisation across orientation for the recognition of 3D letter shapes after IT lesions in the macaque monkey. The monkeys were trained to discriminate simultaneously presented pairs of stimuli (J vs π, and P vs T). Training was continued to 90/100 correct and then subjects were *overtrained* to the same criterion with partial reinforcement. Both stimuli in a pair were transformed on probe trials in their orientation (original orientation 0°, transformed orientations 90°, 180° and 270°), or their real size (described below). It was found that animals with IT lesions took longer to learn the original discriminations, but generalised across orientation transformations in a way directly comparable to the unoperated controls. These results appear difficult to account for if one speculates that orientation general descriptions of objects and features are only present in AIT. Without AIT, subjects should fail to generalise over orientation because they would be deprived of the only neural apparatus which is tolerant to pattern orientation.

There may be, however, particular simple features present in the stimulus pairs that can be analysed by early visual cortex in an orientation-independent manner. For the P vs T stimuli in the study of Holmes and Gross (1984b), the detection of the presence of a small closed circular feature could be sufficient to allow discrimination at any orientation. Hypothetical feature detectors would not have to be maximally activated by the **D** component of the P stimulus to enable discrimination independent of orientation. For the J vs π stimulus pair, a multi-sided feature $\square$ that tolerates one side missing (Π) would be activated by the π but not the J stimuli in the 4 test orientations. This explanation is not implausible given the nature of cell selectivity recently described in V2 (e.g. Fig. 7.12b, Kobatake and Tanaka 1994).

In another study, Holmes and Gross (1984a) tested monkeys with IT lesions on a two-choice visual discrimination task, using either different 2D patterns or 3D objects, or identical patterns of objects but presented in different orientations. Monkeys with IT lesions were impaired in learning to discriminate two different patterns or objects. The IT lesion group were, however, unimpaired in learning to discriminate two identical stimuli with one transformed in its orientation by a large amount (90° or 180° for 2D patterns, 60° or 120° for 3D objects). When the two stimuli differed by only a small angle of rotation (2D patterns: 30- 60°; 3D objects: 30°), monkeys with IT lesions were severely impaired. The authors (Holmes and Gross, 1984a) suggest that control subjects might be more proficient at using environmental cues when dealing with more complex objects (i.e. the angle an object makes with respect to gravity or its position with respect to landmarks). By contrast, animals with IT lesions might rely on coarse feature specific cues (e.g. big blob at the top for rewarded object). When a stimulus is rotated only a small amount (less than 60°) the IT lesioned animals would lose their strategy in discriminating the objects because key features would still occur in a similar location (e.g. the stimuli are indistinguishable because in both cases the big blob is in the upper part of the visual image). Interestingly, animals with IT lesions performed equivalently to control subjects in discriminating vertical and horizontal lines (0° vs 90°), but were impaired discriminating oblique line stimuli separated by the same angle (e.g. 45° vs 135°). Holmes and Gross suggest that the lines are processed in a different way than patterns and objects. That is, early prestriate and striate orientation sensitive mechanisms could be employed for the discrimination of lines. Holmes and Gross (1984a) conclude that IT cortex is involved in the discrimination of different patterns but is not necessary for the discrimination of two versions of the same pattern transformed in orientation by a large angle of rotation (>60°).

Walsh and Butler (in press) concluded along similar lines, suggesting that there may be two separate mechanisms for the perception of rotated objects. Walsh et al. (1992) found that monkeys with V4 lesions were unable to relearn discriminations between different shapes or between two versions of the same shape rotated through any angle between 30°-180° (despite being able to perform these tasks preoperatively). One can conclude that for discrimination tasks, involving two versions of the same

pattern rotated by more than 60°, there are usually visual differences in the stimulus pair which are evident on the basis of processing in V4. For more subtle changes in stimulus orientation (<60°), a more sophisticated pattern of processing needs to be executed in IT with orientation sensitive analysis of feature configuration.

Lesions of the parietal cortex can prevent monkeys from discriminating between two stimuli that differ only in orientation (Eacott et al., 1993). Parietal lesions do not, however, prevent discrimination between different shapes. Walsh and Butler (in press) argue that depending on the degree of rotation between the two test images (same shapes), different mechanisms might be utilised for perceptual discrimination. If two patterns (of the same shape) differ in orientation by more than 60°, then the pathway from the prestriate cortex to the parietal cortex is sufficient to support visual discrimination between stimuli. This pathway is sufficient to define the location of simple but salient features (e.g. there is a blob at the top of one of the patterns). If the patterns differ by some smaller angle of rotation, then the complete ventral pathway involving prestriate cortex and IT is necessary for discrimination (e.g. blob will be approximately at the top of both patterns).

STS lesions

Lesion studies of IT cortex rarely differentiate between anterior STS cortex and IT cortex. The cortex within the STS was commonly regarded as an extension of IT cortex (Dean, 1976), yet it may have different roles in recognition. Eacott et al. (1993) tested monkeys for their ability to learn to discriminate two simultaneously presented images of ASCII characters, where either the images differed in their shape (same orientation), or in their orientation of presentation (same shape rotated 90° or 180°). Monkeys with STS lesions (including anterior and posterior STS) were impaired in both the shape and orientation discrimination tasks. It is interesting to note that monkeys with IT lesions were only impaired in shape but not orientation discrimination.

In conclusion, neuropsychological lesion studies in both humans and monkey subjects suggest that PIT may support feature analysis in an orientation specific manner. If pattern discrimination and generalisation can be solved by feature analysis,

subjects with AIT lesions (leaving PIT feature analysis intact) will be able to solve the tasks. The STS appears to function similarly to the IT in recognition processes, though may have an additional role in the analysis of orientation.

Single cell studies

Orientation specificity

The majority of cells found in area V4 exhibit tuning for simple attributes of the image, including spatial frequency or stimulus width, length and colour. Some cells in V4 do show selectivity to slightly more complex features (e.g. preference for a circular but not square shape, or the angle between 2 contours; Kobatake and Tanaka, 1994). The majority of V4 cells sensitive to simple or more complex features exhibit sensitivity to orientation (Desimone and Schein, 1987; Haenny and Schiller, 1988; Kobatake and Tanaka, 1994). Kobatake and Tanaka (1994) show an example of a V4 cell which is selective to the orientation of the optimal stimulus (a white bar presented at 315°) but generalises across size change [two sizes were tested, short (20%) and long (100%), see chapter VIII]. The main anatomical input into inferotemporal (IT) cortex is through V4 and it is therefore perhaps not surprising that cells in IT are also orientation sensitive (see Perrett and Oram, 1993; Oram and Perrett, 1994a; Vogels and Orban, 1994).

For some features which do not have an axis of elongation such as a dark spot, a circular or square shape, and radially symmetrical patterns (e.g. high frequency Fourier Descriptors; Schwartz, Desimone et al., 1983; Gross, 1992) cells with preference for these features will automatically show tolerance of 90° or 180° rotation. In this sense, such feature detectors are orientation-invariant.

Tanaka et al. (1991) found that cells in the AIT which respond selectively to complex objects, such as faces and features, appear to be orientation selective. Ten AIT cells responsive to faces were tested for four orientations (0°, 90°, 180° and 270°). For these cells, the rotation of the optimal stimulus by 90° reduced the neuronal response by more than half. An additional 11 cells were reported which also exhibit orientation sensitivity, but were only tested qualitatively. The optimal orientation of the face stimulus varied from cell to cell, though the upright orientation was most commonly encountered (Tanaka et al., 1991).

Orientation generalisation

Physiological studies (Perrett et al., 1982, 1984, 1985; Hasselmo et al., 1989b) have shown that cells in the IT and STS cortex of macaque monkeys are selectively responsive to faces. In an initial study by Perrett et al. (1982) of selective cells in the STS, generalisation across object orientations appeared to be universal. All the face responsive cells studied (26) did not differentiate in response magnitude between upright, horizontal or inverted faces (Perrett et al., 1982). A closer examination of responses indicate that cell latencies can be affected by orientation even when response magnitude of the cells was similar for upright and inverted faces. Neural response latencies for 62% (16/26) of the cells were affected by the orientation of the image presented (Perrett et al., 1988). When the orientation of the image was changed from the upright orientation, the response latency of 10/26 cells were not affected. 15/26 cells responded with longer latencies to inverted or horizontally presented faces. Perrett at al. (1988) described this increased response latency as additional 'processing time' which may parallel increased RTs to identify faces or face configurations.

Hasselmo et al. (1989b) tested a small group of cells which responded to a particular movement (ventral flexion of the head; 'head nodding'). These cells continued to respond to the head movements even when the image was rotated in the picture plane, i.e. upside-down. This process altered the retinal or viewer-centred co-ordinate system, but not the object-centred co-ordinate system, since for the latter the position of the relative parts of the object are solely related to the object itself (i.e. the forehead moves towards the chest) and is therefore not dependent on the view. Their findings suggest that this small population of cells reflected the latter, the object-centred description of the movements. However, their results are limited in information about generalisation of rotation in the picture plane as they apply to stimulus motion.

In summary, physiological data shows that early visual areas (V1, V2, V4) exhibit orientation-specific coding of features of objects. Cells from these areas feed into more anterior visual processing areas (PIT, AIT), where cells are selectively responsive to progressively more complex features and object such as faces. From studies carried out so far, it appears that all IT cells exhibit orientation specificity

whether the cells are selective for elementary features or more complex objects. Studies in the STS indicate the presence of cells which are selective for complex objects and which respond irrespective of the stimulus orientation. Interestingly, such cells exhibiting orientation-invariant responses to faces require extra time for the processing of unusual orientations of the stimuli.

It is pointed out, however, that for single cell studies described above only very few orientations were tested. Indeed, in the majority of instances only the neuronal responses to the upright and the inverted stimulus was investigated. The empirical study reported here was concerned with the extent to which cells in the STPa show orientation invariance (generalise across different orientations), or orientation specificity in their responses to heads, bodies or whole bodies presented in multiple orientations, often up to eight different orientations in the picture plane. This allows us to address the question of how orientation in the picture plane is processed by the visual system and whether this is done in a similar fashion as object view (see chapter VI and Perrett et al., 1991). In addition, neuronal response timing factors to different orientations of the stimulus for a cell population will be described and discussed.

METHOD

General single unit methods (see chapter IV) were applied to investigate the response pattern of cells in the STPa of the macaque monkey. All cells tested showed a significantly different response to the whole body than to both control stimuli and spontaneous activity (S/A). For each cell the optimal perspective view was first defined for the upright body (see chapter VI), and subsequently the cell was presented with that view of the whole body in different orientations. Note that the entire body stimulus was used for testing purposes. This is because cells selectively responsive to only one component part usually (see chapter V) responded to the entire body significantly better than to S/A and control objects. In addition, it is assumed that neuronal response selectivity to a particular orientation is held constant across different body parts (and the entire body) as seen with view. That is, if e.g. a 'head alone' cell is selectively responsive to a particular orientation of the head presented in isolation, then it is expected to find that the entire body presented in the same orientation as the head alone will elicit a cell response of the same magnitude (or larger, see chapter V) than

the response to the head presented in isolation. This makes sense, if one follows the argument from chapter V, that component parts are associated with each other. It is unlikely, therefore, to find e.g. a figure where the head is upright in respect to gravity and the body component is inverted.

Visual stimuli

The visual stimuli were all photographed onto 200 ASA ectocrome print film, with the subject (bipedal whole human body) standing against a light grey background. Eight pictures of the whole body presented in eight different perspective views (rotated in depth, see chapter VI) were taken. For each view, eight orientations in the picture plane were presented (resulting in 64 different stimuli, see e.g. Fig. 7.1). This was done by re-photographing the upright image rotated by 45° steps anti-clockwise onto a 200 ASA Ectocrome slide film, resulting in the following orientations: 0° (upright), 45°, 90° (horizontal), 135°, 180° (inverted), 225°, 270° (horizontal) and 315°. The images (at a size of 1.73 m) were then projected against a white wall at a viewing distance of 4 m resulting in the image size being 24.4°. The neuronal response of a cell was normally tested for eight (in some cases only four) different orientations.

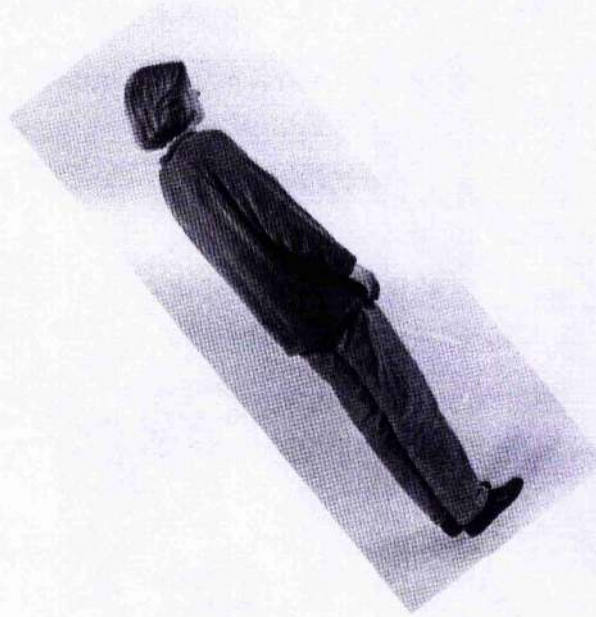
Besides the specific body stimuli tested, a range of control stimuli and a 'no stimulus condition' (S/A) were used. The control stimuli included slide images of complex 2D and 3D objects of different sizes, shapes, textures and orientations (e.g. broom, lab coats, chairs, pictures of different animals, etc.), simple 2D geometrical shapes (bars, spots and gratings) and simple 3D forms (balls, cylinders, boxes, etc.).

Testing methods

Every cell from which neuronal activity was recorded was first tested in an exploratory way by presenting a series of static and moving 3D objects (including bodies), tactile and auditory stimuli. Cells found to be selectively responsive to static views of the whole body were then further investigated for sensitivity to orientation. After identifying the optimal stimulus view with a 3D or 2D whole body stimulus (see chapter VI), further testing was carried out with at least five trials of each of the 2D slide stimuli described above. Stimulus order was presented in a computer-controlled pseudo-random order.

Figure 7.1. Examples of stimuli used for testing: Whole body stimuli presented in different orientations of the picture plane and different (depth) views.

View: 225°
Orientation: 0°



View: 225°
Orientation: 45°

View: 270°
Orientation: 180°

View: 135°
Orientation: 90°

For some cells which were selective in their response to a particular orientation of the stimulus, a receptive field test was carried out to ensure that the response (or more the lack of response) was due to the orientation of the stimulus, rather than the position of the optimal part in the receptive field. This was often only carried out as an observational test (little statistical data available). Therefore, if the cell was previously classified as maximally responsive to only one body part (head or body presented in isolation), the effective body part was randomly presented in three different locations of the visual field: at the top (0.9 m above central LED, angular displacement of 13°), the centre (on top of the central LED) and the bottom (0.9 m below the central LED). During these investigations the monkey was to fixate on the central LED to avoid eye-movements.

Data analysis

Cell responses to different test conditions were compared by using 1-way analysis of variance (ANOVA) and post-hoc tests (protected least significant difference (PLSD), Snedecor and Cochran, 1980) with a significance level of $p<0.025$ (see chapter IV).

In addition, a multiple linear regression analysis was carried out on cells tested with eight different orientations. This was done to estimate the best relationship between response and second order cardioid function of orientation of the stimulus (Perrett et al., 1991). This regression analysis calculates the values of the coefficients $\beta_{1\text{-}5}$ of the equation defined below for producing the highest correlation between cell responses and the angle of orientation.

$$R = \beta_1 + \beta_2 \cos(\theta) + \beta_3 \sin(\theta) + \beta_4 \cos(2\theta) + \beta_5 \sin(2\theta)$$

where R is the response, $\beta_{1\text{-}5}$ are coefficients and θ is the angle of the body orientation (adapted from Perrett et al., 1991).

A Chi-square test was then carried out on cell responses which produced a significant result with the regression analysis. This was done to compare the observed response rates with the expected response rates predicted by the equation defined above (significant relation, $p<0.05$). If the relation between the observed response values and the expected response values was significant, the equation defined above was used to define the optimal angle of orientation (θmax), the maximum response at

this orientation (Rmax), the sharpness of tuning (i.e. the average angle of rotation required to reduce the response of half of that if Rmax) and the angle of magnitude of any other peak in the orientation tuning (Perrett et al., 1991).

Furthermore, an orientation discrimination index was computed for 24 cells to investigate the cells ability to discriminate between the upright and inverted orientation of a whole body stimulus. All cells tested for the upright and the inverted body stimuli were included in this analysis, regardless of the cell classification (i.e. whether generalising to tuned to particular orientation). This formula used was adapted from the view discrimination index formula used in chapter VI:

Orientation Discrimination Index (I_O) = [(response to upright orientation- S/A) - (response to inverted orientation - S/A)] / (response to upright orientation - S/A).

In addition, post-stimulus time histograms (PSTH) and cumulative response curves were plotted for population estimates of time course responses (see general method chapter IV).

RESULTS

Of the visually responsive STPa cells of two subjects a total of 25 cells which were found to be responsive to the whole body (i.e. have significantly greater response to the whole human body than to control objects and S/A) were tested for selectivity to different orientations in the picture plane of the whole body. All the cells included in this study did not respond to a wide variety of different control objects tested. The majority of cells (18/25) were tested for the whole body in eight different orientations, whereas the remaining 7 cells were tested with less than eight orientations. Cells were then categorised on the basis of their response to the different orientations: 1) generalising across all orientations; 2) generalising across all orientations with orientation tuning superimposed on the generalised response; and 3) selectively responsive to a particular orientation.

Generalising across orientation

Seven cells (of 25 tested) responded to all orientations at a rate significantly above S/A and control stimuli. Four of these cells responded statistically equivalently to all orientations (see Fig. 7.2).

Generalising across all orientations with some tuning

Three cells showed orientation tuning superimposed on the generalised response (see Fig. 7.3).

Selectively responsive to a particular orientation

The remaining 18 cells responded to some, but not all orientations (see e.g. Fig. 7.4). To ensure that the cell was not only sensitive to the optimal stimulus in a particular region of the visual field, a visual field test was carried out. If the cell did not respond significantly different to the stimulus presented in a different location (see Fig. 7.4b), then the response pattern to different orientations of the stimulus was suggested to be due to the change in orientation rather than the presence of the stimulus in a particular location. In this case, the cell was then included in the analysis of orientation sensitive cells. Different cells were maximally responsive to different orientations.

Distribution of orientation tuning

The majority of cells (21/25, 82%) were selectively responsive for orientation in the picture plane. While most of the tuned cells (15/21) were selective for orientations close to upright (see e.g. Fig. 7.4), some (6/21) cells, however, were selective for non-upright stimulus orientations such as inverted images (e.g. Fig. 7.5).

All the cells displaying orientation selectivity displayed an approximately monotonic tuning function (e.g. Fig. 7.6). That is, the cell responded best to one orientation and showed a gradual decrease of response amplitude as the stimulus was rotated away from its optimal orientation.

The tightness of the tuning varied across the cells. Rotation of the stimulus away from optimal by 45° to 90° would reduce most cell responses by half.

To find the shape of orientation tuning in the picture plane a regression analysis for each cell was used to estimate the maximal response (Rmax) and the optimal angle

Figure 7.2. The mean responses (+/- 1SE) to the front (0°) view of the whole body and spontaneous activity (S/A) are illustrated for one cell (E83_38.31L). The cell response showed no significant difference between different orientations tested (**generalisation**), but responded to all orientations significantly different than to spontaneous activity and controls ($p<0.05$ PLSD each comparison). ANOVA: $F(9,40)=1.89$ $p<0.05$.

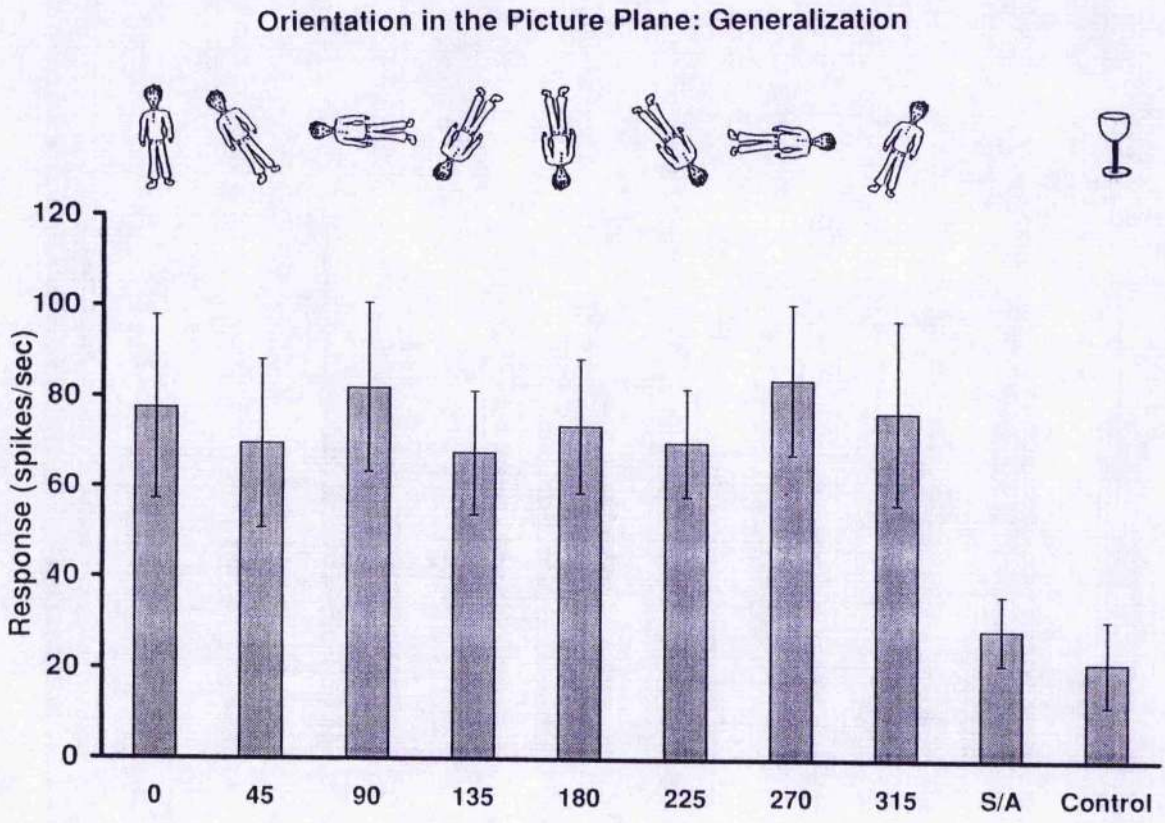


Figure 7.3. The mean responses (+/- 1SE) to the back (180°) view of the whole body and spontaneous activity (S/A) are illustrated for one cell (E64_33.06L). The cell responded best to the whole body stimuli when upright ($p<0.05$) (**tuned**), but also responded to other orientations better than to spontaneous activity (S/A) ($p<0.05$) (**generalising**). ANOVA: F(9,39)=3.9 p=0.001.

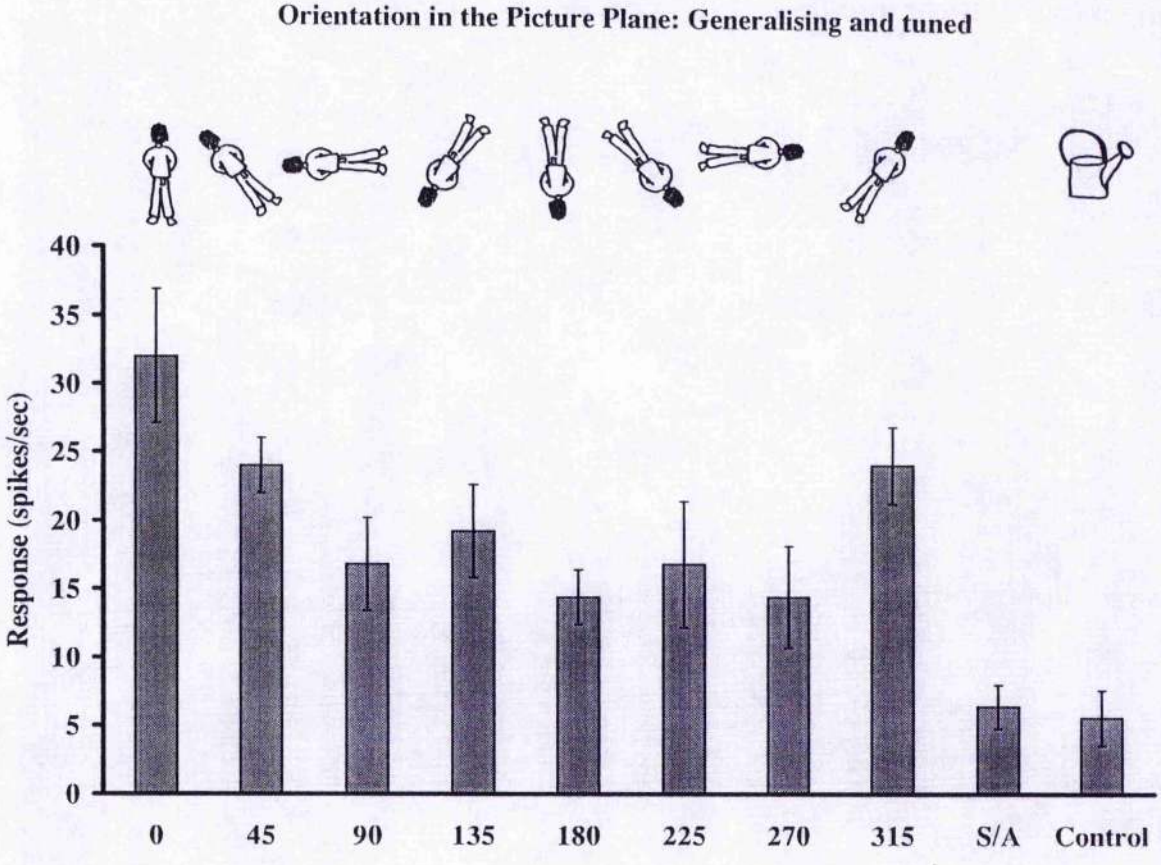
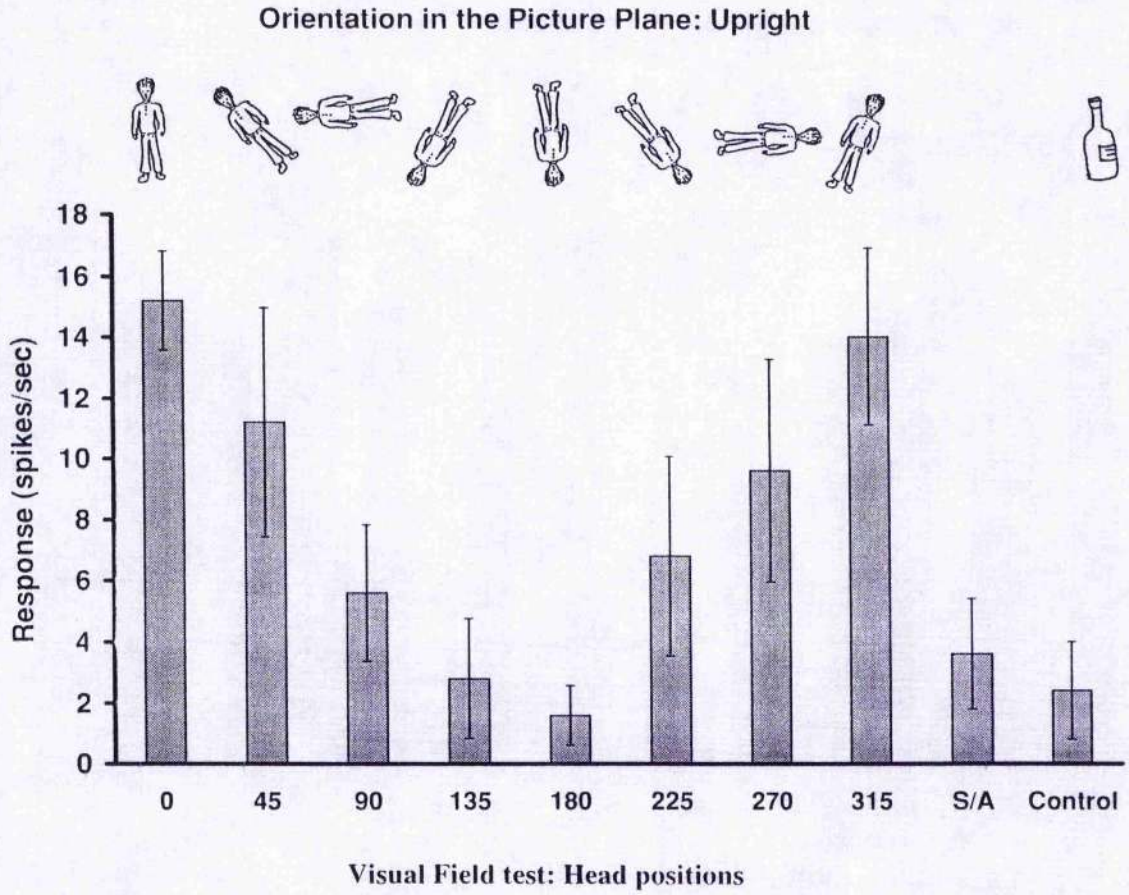


Figure 7.4. (a) The mean responses (+/- 1SE) to the front (0°) view of the whole body and spontaneous activity (S/A) are illustrated for one cell (E99_40.16L). The cell responded best to the whole body stimuli at **upright** position, but also responded to some other orientations better than to spontaneous activity (S/A) ($p<0.05$) (**tuned**). ANOVA: $F(9,40)=3.8$ $p<0.005$. (b) No significant difference in response at different positions (angular displacement of 13° up and down from the central fixation point) were found (ANOVA: $F(4,19)=6.5$ $p<0.005$). All the stimuli tested gave a significantly different response than the one to control stimuli and S/A ($p<0.05$).



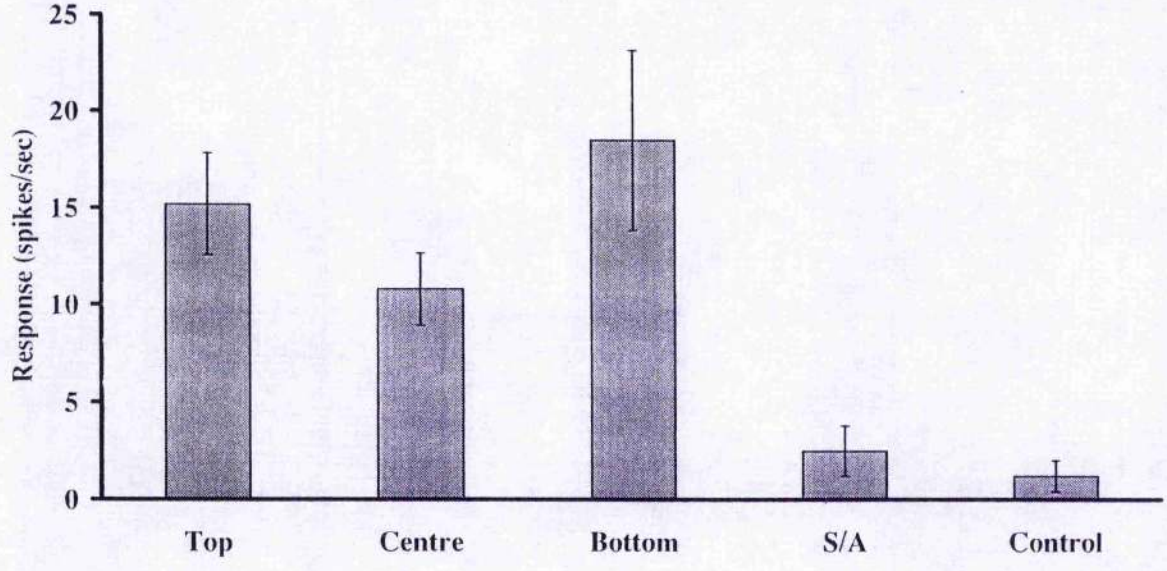


Figure 7.5. The mean responses (+/- 1SE) to the front (0°) view of the whole body and spontaneous activity (S/A) are illustrated for one cell (E61_31.82L). The cell responded best to the whole body stimuli when **inverted**, but also responded to some other orientations better than to spontaneous activity ($p<0.05$) (tuned). ANOVA: $F(9,40)=3.097$ $p=0.006$.

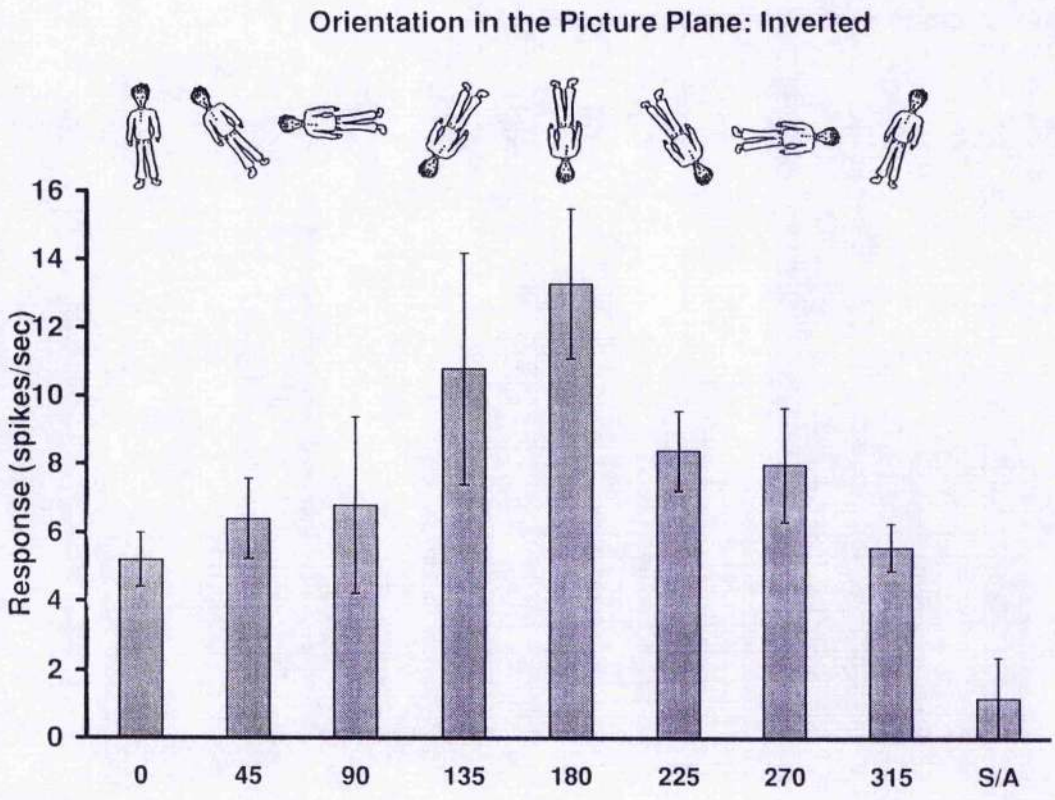
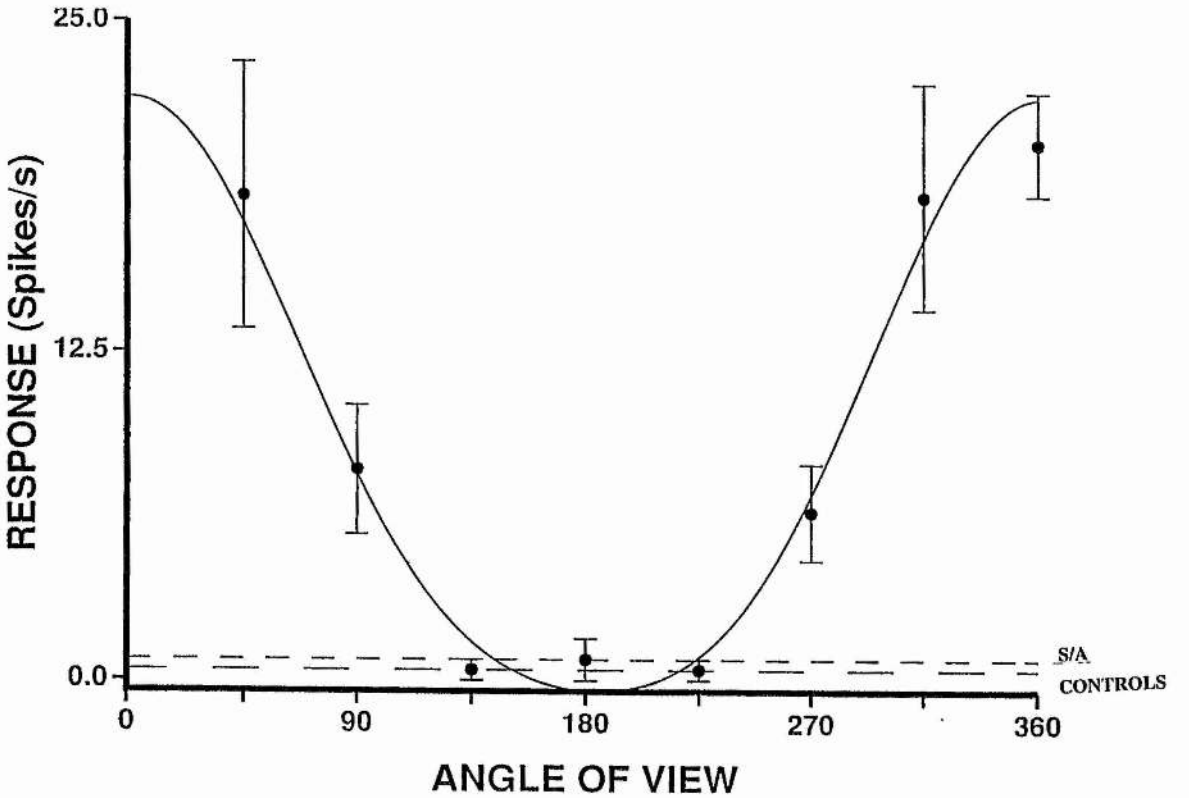


Figure 7.6. Responses of a **viewer-centred cell** (E86_38.43R) **with unimodal tuning** to orientation in the picture plane. The mean responses (± 1SE) recorded for eight different whole body stimulus orientations are illustrated (R^2 = 0.684; F(4,39) = 18.939, p<0.0005). The dashed lines indicate the mean response to control stimuli and S/A. Responses close to the upright (optimal) orientation (45° and 315°) were not significantly different than 0° (Protected LSD tests p>0.05) but were significantly greater than to controls and S/A (p<0.0005 each comparison). ANOVA: F(9,49)=12.38, p<0.0005.



of orientation (θmax) at which Rmax occurred (Fig. 7.7). This figure includes all tuning curves for 14 cells maximally sensitive to one orientation in the picture plane (see methods above) (Perrett et al., 1991). To combine different individual tuning curves, a normalisation was applied. This was done by setting the mean base line (S/A) firing rate of each cell to 0 and Rmax to 1.0. The regression analysis was then repeated resulting in 'normalised' individual tuning curves. The peak normalised responses (1.0 for each cell) were aligned at point 0 (optimal orientation for each cell) on the X-axis. If a normalised response curves falls below the point 0 (i.e. response < average S/A), then the cell response is inhibited to that orientation of the stimulus.

Having established tuning curves for each cell tested with eight different orientations of the stimulus, an average tuning curve for these cells was produced. This is done by averaging the coefficients obtained from the individual regression analysis for each cell (normalised responses were used) (Fig. 7.8).

Finally, having shown that different cells have different optimal orientations in which the stimulus is viewed, a polar plot was drawn to indicate the distribution of these optimal orientations across the cell population. All view-sensitive cells (14 cells) were used in this illustration, which included cells tested for 4 or 8 orientations. The polar plot indicates that the majority of cells sensitive to orientation are tuned to the upright orientation ± 22.5° of the stimulus (see Fig. 7.9).

Orientation Discrimination Index

The orientation discrimination index indicates the ability of the cells to discriminate between the upright and the inverted orientation of the body stimulus (see Fig. 7.10). This index shows that most cells are able to discriminate between upright and inverted body orientations (i.e. most cells have an orientation index value (I_O) which is greater than say 0.2 or smaller than -0.2). If I_O equals 0, then the cell will respond equally well to the upright and the inverted image. If the I_O equals 1.0, then the response to the inverted body is the same as S/A and the response to upright images is significantly greater than S/A. If $I_O > 1.0$, then the cell response to the upright orientation is greater than S/A and the response to the inverted orientation is less than S/A. At a negative I_O, the neuronal cell response to the inverted body is greater than the

Figure 7.7. Normalised tuning curves for cells selectively responsive to view (including cells which respond to all body orientations > S/A, but show some tuning superimponsed) tested for eight orientations in the picture plane. The tuning curves for 14 cells are estimated from the best fit cardioid function relating resonse to an orientation. S/A = 0 and maximal response = 1.0. Angle of orientation is expressed as an angle of rotation from optimal orientation (θmax).

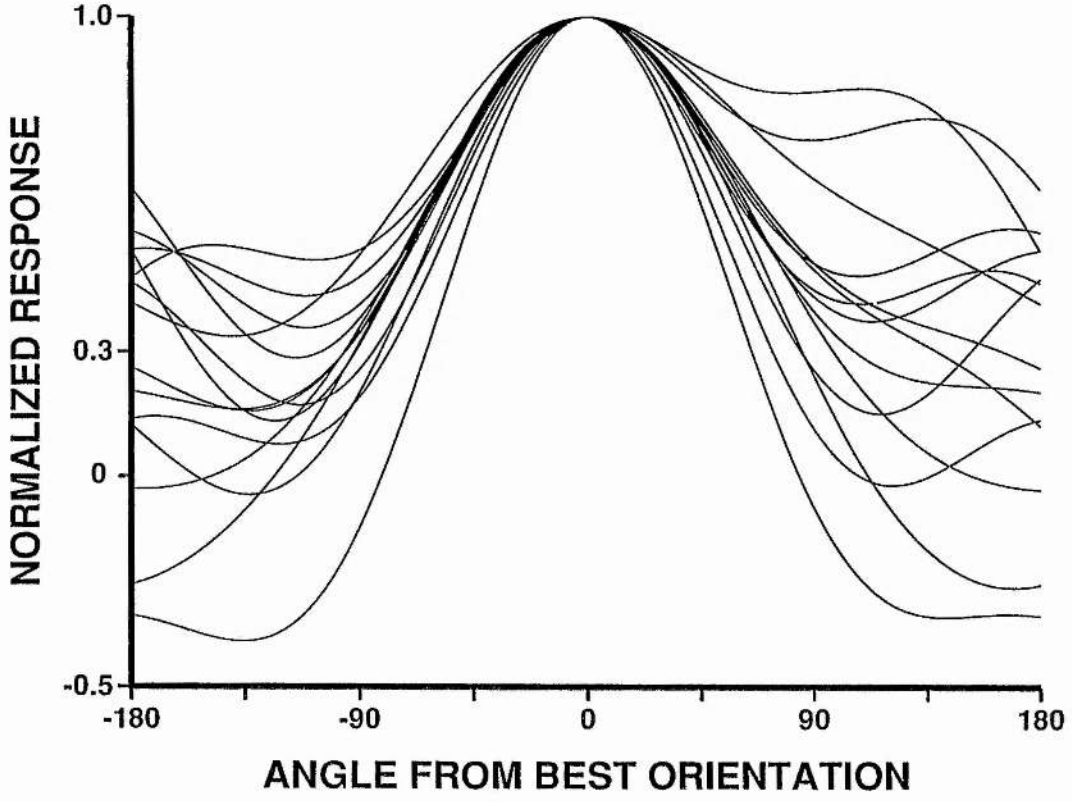


Figure 7.8. An **average tuning curve** from all 14 view sensitive cells (described in Fig. 7.7) which have been tested for eight different orientations in the picture plane was computed. S/A = 0 and maximal response = 1.0.

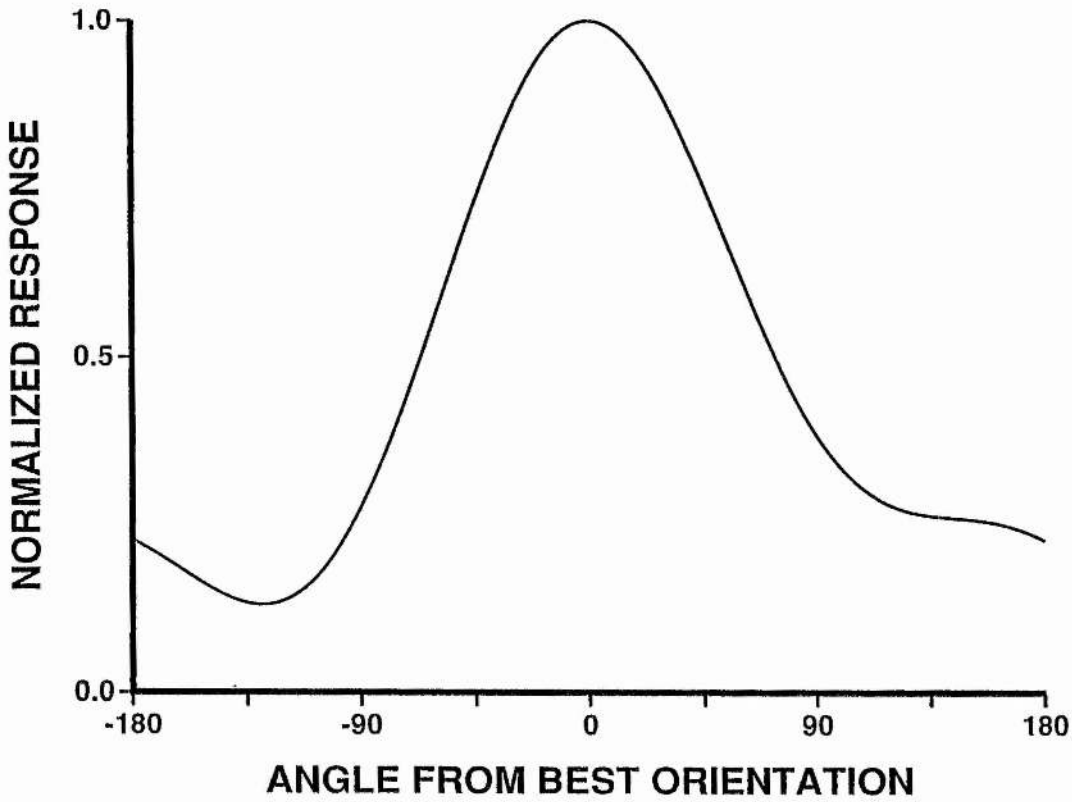
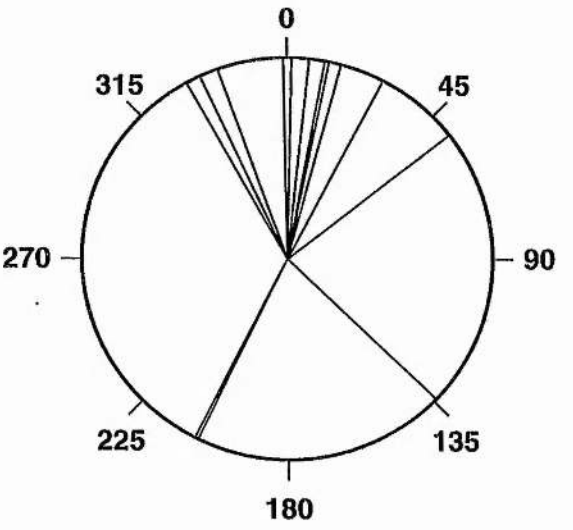
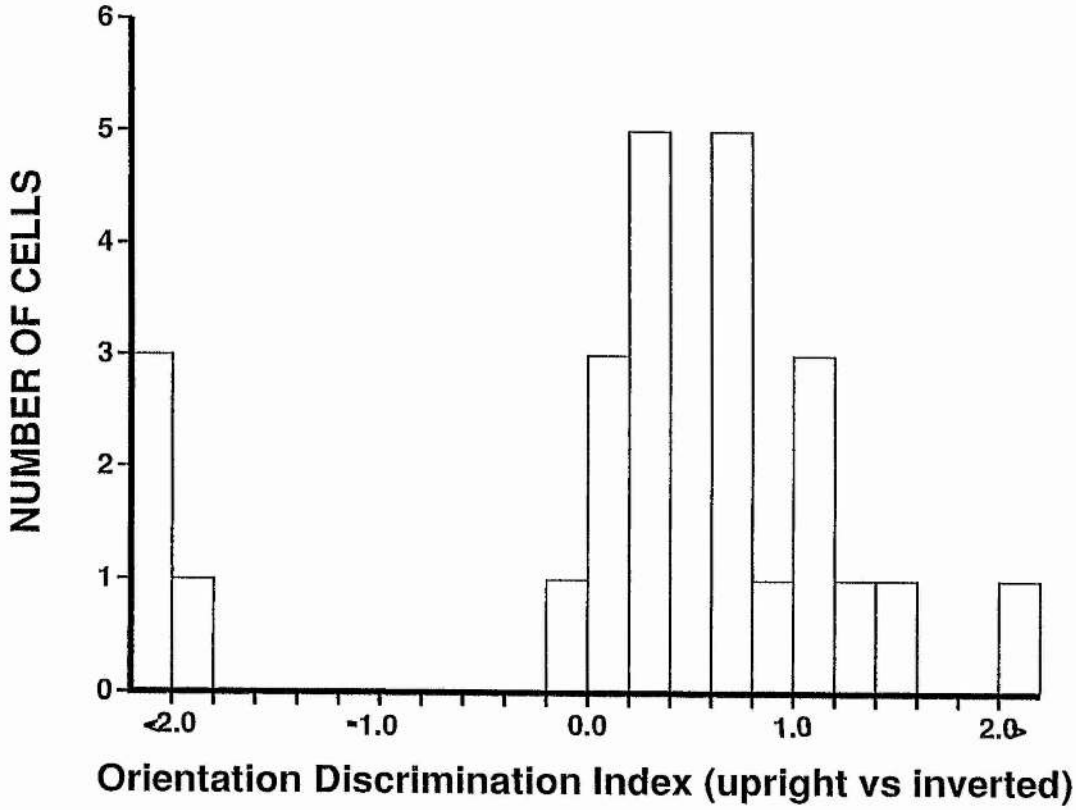


Figure 7.9. Polar plot: **Distribution of optimal response orientations** (tuning) across the population of cells sensitive to view (n=14). Each line presents the optimal orientation obtained through the regression analysis of an individual cell (see text).



Cell Population Estimate for optimal response orientation

Figure 7.10. Orientation Discrimination Index. An index of orientation discrimination was computed (see text) for 24 cells tested for responses to the upright and inverted body orientation.



cell response to the upright image (i.e. cell is tuned to a non-upright orientation and hence is able to discriminate between the two orientations).

Population estimates of time course of responses

To obtain population estimates of time course of responses of cells tested with different orientations in the picture plane (all cell types included), post stimulus time histograms (PSTH) were plotted for a population of 10 cells (Fig. 7.11a). At point 0 ms stimulus presentation occurred and thus, the average response on-set appears to be for all effective stimuli at approximately 60 ms. As a population, the response to the upright image was the greatest. After response on-set, average neuronal response increased for approximately the first 320 ms and then decreased down to threshold level after approximately 600 ms following response on-set (540 ms after stimulus presentation).

A cumulative response curve of the same data was plotted to obtain a better indication of the cell population response (all cell types included) (Fig. 7.11b). This graph allows one to compare levels of average neuronal response at any given time. Therefore, at for example, 400 ms after stimulus presentation, the cells selectively responsive to orientation will as a population respond most when a body/head stimulus is presented in its upright orientation. Also, if a threshold is chosen (to trigger, for example, behavioural response times) at say 40% of the maximum response, then the threshold is reached fastest when an image is presented in its upright orientation relative to the observer or gravitational axis (approximately 230 ms after response on-set). This response is faster than the population response to rotated images (e.g. horizontally viewed whole bodies, threshold reached only after 500 ms following response on-set).

Histological reconstruction

The histological reconstruction of 3 monkey brains showed that all cells tested for whole body stimuli presented in different orientations were located in the upper and lower bank of the STS (see Fig. 7.12 and Appendix 4 for one subject). There was no observed clustering of cell types. A chi-square test showed that cells tuned to the upright image are not more frequently found at more anterior sites than at more posterior sites of the STPa (chi^2=1.3, df=1, p=0.25, where group A were cells in

Figure 7.11. Cell population response to different stimulus orientations and control objects. a) Combined responses of 10 cells to different stimuli orientations: upright (0°), horizontal (90° and 270°) and inverted (180°). The clear area displays the population response to the upright body stimulus orientation, medium grey area to the horizontal body stimulus orientation, dark grey area to the inverted body stimulus orientation and the black area population response to various control objects. Firing rate is expressed as a percentage of the peak response to the best (in this case the upright) stimulus. Stimulus presentation occurs at point 0 ms. b) Cumulative Response curve of orientation population estimate (10 cells). The colour coding for different stimuli responses is identical to (a). Stimulus presentation occurs at point 0 ms. The average population response rate was calculated after normalising each cell's activity (S/A=0, Max Response=100). The cumulative population response is defined as the running total of this average population response from 200 ms pre-stimulus onset.

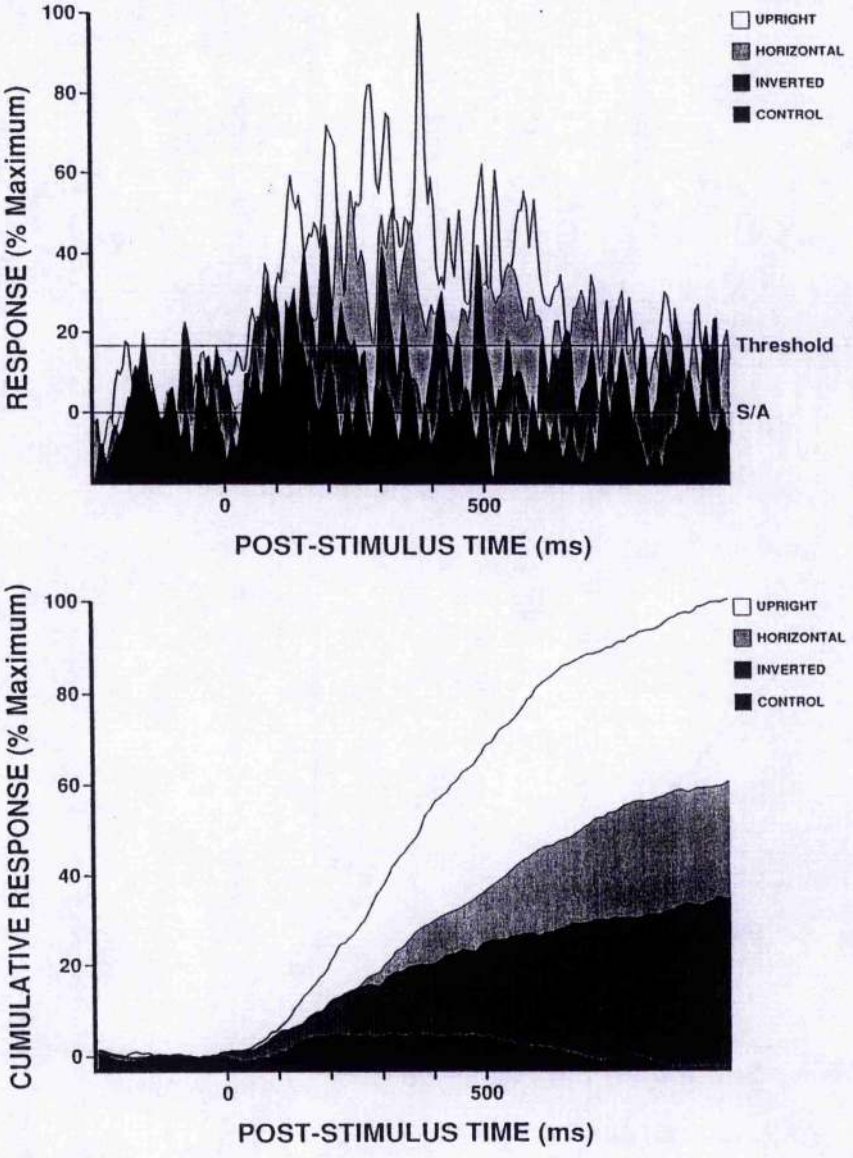
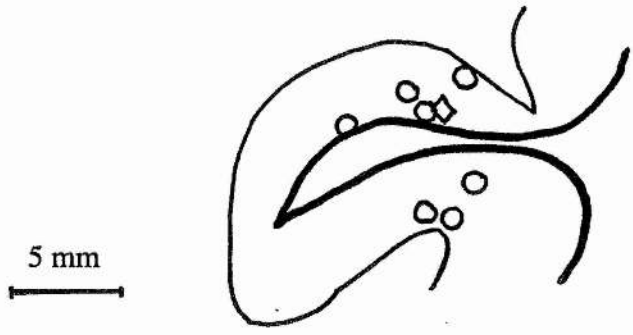


Figure 7.12. Histological reconstruction. A frontal section of subject E taken at 15 mm anterior to the interaural plane showing the location of cells tuned to upright orientation of the head/body (open circles) and cells generalising across all orientations of the head/body tested (open triangles). The thick lines indicate the brain surface, the thin lines show the boundary between white and grey matter. Cells were located in both the upper bank and fundus of the superior temporal sulcus (see Appendix 4).



sections <15 mm anterior, and group B were cells found in sections 15-17 mm anterior). Cells coding upright heads/bodies were therefore intermixed with cells coding non-upright heads/bodies and cells which generalised across different orientations in the picture plane.

DISCUSSION

Given the length of the introductory review of this chapter, it is appropriate to first recap the main findings:

Most human and animal studies indicate that the ventral stream (V4, PIT, AIT and STPa) is involved in the processing of information about an object's appearance including the orientation of the object (see e.g. Dean, 1976; 1991; 1992; 1993; Tanaka, Fujita et al., 1993; Kobatake and Tanaka, 1994). The IT cortex plays an important role in the formation of mental representations of objects for long term memory (Miyashita, 1990; Miyashita et al., 1993). Ideally, for recognition purposes, a single object-centred representation which generalises across viewing conditions would be most efficient. This would allow an object to be recognised irrespective of its perceived view, orientation, size, or even part visibility (i.e. part-occlusions). For behavioural interaction with an object, information about the object's orientation relative to the viewer has to be available to allow and guide interaction. For such visuo-motor control, viewer-centred representations are essential, although information about the objects' identity is irrelevant. It now appears that such information about 3D orientation (and size) is coded by cells in parietal cortex (area LIP, Sakata personal communication, 1995; Taira et al., 1990; Sakata and Taira, 1994).

Interestingly, different methods of investigation emphasise the importance of different types of representation in the temporal cortex. The study of brain lesioned subjects (both humans and monkeys) indicate the involvement of the temporal cortex in generalising across changes in orientation (see e.g. Wilson, 1957; Cowey and Weiskrantz, 1967; Gross, 1973; Dean and Weiskrantz, 1974; Dean, 1976), whereas lesions in the parietal cortex appear to disrupt identification of an object's orientation (Pohl, 1973; Ockleford, Milner et al., 1977; Ungerleider and Brody, 1977; Jeannerod et al., 1994). From such lesion work one might assume that viewer-centred information about objects would be found only in the dorsal stream of processing. Single unit

recording studies, on the other hand, reveal that the vast majority of cells in the temporal cortex are selective for object features or parts (see chapter V), view (see chapter VI), orientation and/or size (see chapter VIII). These studies contrast with the lesion studies and indicate the predominance of object representations that are highly specific to viewing conditions (i.e. viewer-centred representations). In the same physiological studies, however, a minority of temporal cortex cells are found to generalise across particular viewing conditions such as orientation.

To summarise neurophysiological studies, cells in early visual areas (V1, V2, V4) exhibit orientation-specific coding of elementary features of objects (Henry et al., 1974; Trotter et al., 1989; Kobatake and Tanaka, 1993). These cells feed into more anterior visual processing areas (PIT, AIT), where cells are selectively responsive to progressively more complex features and object such as faces but still exhibit orientation specific responses (Tanaka et al., 1991; 1992; 1993; Kobatake and Tanaka, 1994). IT cortex in turn projects to cells present in the STPa (Seltzer and Pandya, 1978) which have been shown to be selective for complex objects and which have previously been reported to respond irrespective of the stimulus orientation (Perrett, Rolls et al., 1982; Perrett et al., 1985; 1988). Interestingly, such cells exhibiting orientation-invariant responses to faces appear to require extra time for the processing of unusual orientations of the stimuli (Perrett et al., 1985; 1984; 1988). The empirical work reported here confirms the presence of cells with orientation-invariant responses but also indicates that most coding for complex objects in the STPa is orientation selective as was found in IT shape selective cells. That is, 82% (21/25) of the cells that were selective for the sight of the body tested here also showed some sort of selectivity to a particular orientation of the body. Furthermore, the Orientation Discrimination Index computed for all the cells tested in this study indicates that the cells which do discriminate between different orientation of the whole body stimuli are relatively good at such discrimination.

The more anterior the more generalisation

Indeed, it could be argued from the above observations, that the further along the ventral stream one goes, the greater the percentage of cells which generalise across viewing conditions in which an object is seen. However, single cell testing of the

temporal pole and the amygdala investigating whether more cells code (highly selective) visual information of objects in an object-centred fashion is still to be carried out. Nonetheless, from single cell studies in visual areas of the ventral stream up to the STPa, it could be suggested that the selectivity between objects rises while at the same time the sensitivity to the particular viewing conditions declines (in this sense the cells become less viewer-centred and more object-centred) (Mikami et al., 1992). As one progresses down the ventral stream, cells become also less sensitive to position (Desimone and Gross, 1979; Tanaka and Fujita, 1991; Kobatake and Tanaka, 1994; Ito et al., 1995) and show a greater tendency to generalise across object parts (Wachsmuth, Oram et al., 1993; 1994), see chapter V). The cells also become increasingly selective. Cells in the posterior temporal cortex may show size tolerance but are generally selective for simple features such as a circular shape. In the anterior temporal cortex, cells may again show size tolerance but are now more selective between patterns: particular cells will only respond to only circular patterns that are faces (Rolls and Baylis, 1986), see chapter VI). In addition, cells in visual areas of the ventral pathway up to and including AIT (V1, V2, V4, PIT) do not show the ability to generalise across change in orientation of the optimal stimulus (Desimone and Schein, 1987; Haenny and Schiller, 1988; Kobatake and Tanaka, 1994). Only cells in STPa start to show generalisation across orientation.

It is pointed out that in the present investigation cells tested for orientation selectivity to the whole body were found to be located at a very anterior site of the STS (see Appendix). At first glance it does appear as if more anterior sampling sites represent more cells tuned to the upright image, though statistically this is not the case (see result section). Furthermore, the notion that more generalised coding occurs at more anterior sites within the ventral visual processing pathway is not strongly supported by the anatomical findings of this study (see subject E). The recording of cells occurred at very anterior sites which were close to the temporal pole, making it unlikely that there is more generalised coding occurring at even more anterior sites of the STS. This does not rule out the existence of other areas with mainly generalised representations, though these areas might not be predominantly visual areas.

Building object-centred representations from viewer-centred representations

Models of object recognition based on physiological evidence suggest the construction of view-independent representations by pooling the outputs from appropriate view-specific representations (e.g. Perrett and Oram, 1993; Logothetis, Vetter et al., 1994). Such pooling operations can be based on simple learned association between different views of the same object seen in temporal succession as the object is rotated or the viewer moves around the object. While these pooling operations were first conceived to account for generalisation across perspective view, the same operations could account for generalisation in many domains (Foldiak, 1993). In this same manner the responses of neurones selective for the head or body capable of generalising across orientation, could be formed by pooling outputs of orientation sensitive cells, i.e. orientation specific representations being pooled to establish orientation general representations. Pooling across descriptions valid for specific lighting conditions (Hietanen et al., 1992) and across descriptions for independent articulating parts of the same objects (Wachsmuth, Oram et al., 1994), see chapter V) may be achieved by analogous processes.

Familiarity and experience

The study reported here has shown that different cell populations in the anterior STPa are selective for different object orientations of the same complex object (face/body). However, the majority (71%, 15/21) of cells which were selectively responsive to particular orientation of the face/body were found to be tuned to the upright orientation of the object. This neuronal response pattern is similar to that found in IT (Tanaka and Fujita, 1991). It is therefore suggested that experience with one (the upright) orientation of an object may enhance the neuronal representation for that object in the specific viewing conditions. This may explain the predominance of cells tuned to upright orientation.

It is interesting to note that the average orientation tuning curve is highly similar to the average view tuning curve obtained and described by Perrett et al. (1991) as well as the average tuning curve obtained by Oram et al. (1993) from direction of motion sensitive cells. Distribution of test direction was tightly grouped around cartesian axes up/down, left/right described by Oram et al. (1993). For the view

analysis, view distribution was relatively continuous but with more cells responding to views within 22° of a characteristic view of the head: front (0°) and profiles views (90°, 270°) (Perrett et al., 1991). In that study few cells were noted to be responding to the back. A recent re-evaluation of view tuning (Oram, Perrett and Wachsmuth, unpublished) indicates a preference of cells coding the front view (face). This was also noted by Hasselmo et al. (1989) though their testing was biased towards front views. The distribution of optimal orientations, however, was quite different from that found for view. In the empirical study reported here, a polar distribution of optimal orientation in the picture plane (Fig. 7.9) indicated that the majority of cells responded optimally to an orientation of the body which was close to upright. This pattern seems to be more similar to the direction tuning distribution than to the view tuning distribution, though there are no (or few) cells preferentially responding to horizontal orientations.

Most cells tested and recorded from in the empirical studies reported here showed selectivity to both human and monkey stimuli presented in their upright orientation (general observation). Furthermore, it is interesting to point out that even though monkeys may be more familiar with seeing other monkeys upside down than humans seeing other humans upside down, this inverted orientation is relatively weakly represented in the neuronal cell population of the monkey STPa. However, monkeys do spend the majority of their time with their heads in an upright position and therefore, their perception of and experience with other heads and bodies will be based (like in humans) on the gravitational upright. In addition, monkeys are not very familiar with seeing human bodies in a non-upright (gravitational) orientation. This frequent experience with the gravitational upright might account for the familiarity effect (greater number of cells coding the upright orientation of the body). Nonetheless, the cell population reported here might be questionable in its true representation of the average sample of the STPa cell population, since during exploratory testing upright 2D or 3D bodies were most commonly used (see general methods chapter IV). Therefore, cells which were tightly tuned to inverted or other orientations were harder to recognise and hence might have been missed for testing purposes.

Response latencies at the population and single cell level

The population estimates of time course responses computed are a measure of the *overall* neuronal response activity to different stimuli of a particular brain area investigated. In this study, the neuronal population response of STPa cells to whole body stimuli presented in different orientations are analysed. The PSTH and the cumulative response curve contain the same data but are displayed differently (see general methods chapter IV). The advantage of the cumulative response curves over population PSTHs is that one can clearly see the total population response at any given time. For instance, at 150 ms after stimulus presentation, a much clearer separation between responses to different stimuli are seen in the cumulative response cure than in the PSTH. This is because in the PSTH, the population response activity is illustrated for only that time (e.g. 150 ms) whereas in the cumulative response curve the activity prior to that time (e.g. from stimulus presentation up to 150 ms) is also taken into account.

If the brain was first to determine the identity of an image, i.e. whether it is a face/body or not, and only then determine the orientation of the stimulus (e.g. top-down processing), then it could be expected to find little difference in response latency and magnitude at early stages of discrimination with increasing differences at a later time. However, the overall neuronal response of the STS cells to the whole stimulus presented in different orientation does respond with greater magnitude to the upright image than to non-upright images (see Fig. 7.11b) and, therefore, discrimination can take place at an early stage, conflicting with top-down processing models.

The cumulative normalised response to a particular stimulus obtained from the averaged normalised PSTH of all cells (see chapter IV) are used to show that the neuronal population activity to the upright whole body stimulus is at any response level or time greater than the cumulative population response to the non-upright body orientations. Therefore, the overall cell responses are faster to upright stimuli than to non-upright versions of the same stimulus. This neurophysiological finding would suggest that recognition of upright bodies is always faster than recognition of non-upright bodies (see below). This is not because there is a latency difference in response onset, but it is caused by the differences in response magnitude to the different stimuli.

The cumulative population response curve indicates an initial faster and greater response to inverted images than to horizontally presented bodies (see Fig. 7.11). This could parallel the finding that inverted (180°) images are recognised faster than images rotated between 90° and 120° from upright (Jolicoeur, 1985; McMullen and Jolicoeur, 1990; 1992). The neuronal population included in this study is, however, relatively small and hence the evidence for this argument is preliminary.

Behavioural and psychophysical data from humans and monkeys suggests that upright images of complex objects (such as faces and bodies) are recognised faster than inverted or other non-upright orientations of objects (Yin, 1969; Shepard and Metzler, 1971; Jolicoeur, 1985; Perrett, Mistlin et al., 1988; McMullen and Jolicoeur, 1992; Phelps and Roberts, 1994). This inversion effect is reflected in physiological findings. Perrett et al. (1988) showed that some (58%) STPa cells selectively responsive to faces had a longer latency to faces rotated 90°-180° from upright than to upright faces. Unlike the present population study, level analysis was carried out on individual cells. For each cell, Perrett et al. (1988) set a threshold level at 95% confidence level based on pre-stimulus activity. Pre-stimulus activity was often very variable, resulting in a high threshold level. Because a threshold was used to investigate response pattern to different stimuli one does not know whether onset latency is equal for upright and non-upright stimuli. The relative response onset might have been equal, resulting in the same latencies for upright and non-upright stimuli. However, if the relative response magnitudes differed (stronger response to upright than to non-upright) then by the time the threshold level is crossed, a latency difference will appear showing a later latency to non-upright than to upright stimuli. This could explain the latency effect found by Perrett et al. (1988).

Looking at the overall neuronal activity of the brain while viewing rotated faces, different latencies have been observed to upright, horizontal and inverted images. Human scalp 'vertex-positive peak' (VPP) recordings to face stimuli presented in different orientations showed a linear increase in response time as the image was rotated away from the upright orientation up to 90° (Jeffreys, 1989; 1993). However, VPP latencies did not further increase when 90°-180° rotated stimuli were presented. The latency to these stimuli did not only stay constant as the angle of rotation increased

from upright (0°), but actually slightly decreased, resulting in faster latency to inverted than to rotated faces by say 120°. This overall neuronal activity result supports the argument stated above, that single cell neuronal population activity may underlie behavioural inversion effect responses (i.e. upright images are recognised fastest, and inverted images are recognised faster than faces rotated by e.g. 120° (Jolicoeur, 1985; McMullen and Jolicoeur, 1990; 1992).

Advantage of orientation specific coding

One question which arises is what the function of orientation specific information in the ventral stream is? First, the answer is that computing such information is an essential stage in the processing leading to the identification of an object, even if later stages of the identification process discard orientation information through generalisation. Second, information about orientation is important in its own right. This is particularly evident for information about the face and body. The anterior STS appears to be a site which integrates information relevant to the interpretation of social signals, such as where other individuals are directing their attention and actions (Perrett et al., 1989; 1992; 1993; 1994; Walsh and Perrett, 1994). To recognise where an individual is attending requires orientation specific and view specific information about the head posture. The left profile view of the head in its upright orientation may indicate that an individual is attending to the viewer's left, but the same profile view rotated in the picture plane by 45° to 135° from upright (in an anti-clockwise direction) indicates the individual's attention is directed towards the ground (Perrett et al., 1992). More generally, to interpret the significance of an individual's posture and actions one needs to specify the orientation of the face, body limbs and hands.

Orientation coding in dorsal and ventral stream

In the scheme of processing suggested here, feature selectivity is first evident in the ventral stream in an orientation specific manner. Gradually, as one progresses along the ventral stream, generalisation across orientation is dependant on pooling of appropriate outputs from earlier stages within this system. Processing of pattern information leading to object identification can thus proceed autonomously within the ventral pathway and is independent of input from the dorsal system.

As an alternative scheme of processing one could propose that orientation information about an object is specified in the dorsal stream to drive visuo-motor co-ordination (Goodale et al., 1991; Biederman and Cooper, 1992b) and that sensitivity to orientation in the ventral system depends on inputs from the dorsal stream. Certainly, such a scheme is plausible on anatomical grounds within the anterior STS which receives input from parietal cortex and IT cortex (see e.g. Seltzer and Pandya, 1978; 1984; Pandya and Yeterian, 1985; 1989; Young and Yamane, 1992; 1993). But, as was described, orientation specificity is prevalent throughout the ventral stream including the posterior and anterior IT cortex and these areas do receive input from the parietal cortex (Baizer, Ungerleider et al., 1991). It may be more appropriate to consider orientation information about objects being computed separately in the dorsal and parietal systems and utilised for different visual tasks. Realising what an object is and knowing whether it is upright (Turnbull, Laws et al., in press), or how to interact with it (Goodale et al., 1991) appear to be dissociable abilities. No doubt there is considerable interplay between the dorsal and ventral streams of processing. Indeed, interactions between individuals depends on an awareness of head and body postures and orientation. Such information may well be specified by the ventral systems including the STS before the information is passed (through STS outputs to parietal cortex, Harries and Perrett, 1991) to dorsal systems to guide appropriate behavioural output and interactions.

Chapter VIII

EFFECT OF SIZE CHANGE ON OBJECT RECOGNITION

INTRODUCTION

Perceiving and recognising an object independent of a change in distance or a change in physical size, does not seem a difficult task, but is, as will be discussed below. This perceptual *invariance* or *constancy* occurs even though the retinal image size of the object varies. The term 'size constancy' refers to the capacity to recognise the physical (real) size of an object independent of viewing distance. For example, when viewing an adult human at a distance of 0.5 m or 200 m, one can recognise that the person's height is the same, approximately 1.5-2.0 m, despite the 400 fold change in (retinal) image size. This ability differs from the capacity to recognise an object's identity or class independent of its size (for example a small person and a large person can both be recognised as humans; similarly a toy car and a real car can both be identified as cars). This latter capacity is here referred to as 'size invariance' or 'size generalisation'. This chapter will discuss how the brain codes an object (or objects of the same class) at different retinal sizes and how the brain achieves recognition which is independent of retinal size.

Behavioural studies in humans

Changing stimulus size has been found to impair recognition in a wide variety of paradigms. Jolicoeur (1987) exposed subjects to 20 shapes, where each shape was presented for 12 seconds during the learning phase of a recognition memory task. In subsequent testing in which subjects classified the familiarity of shapes (as old or new), Jolicoeur found that recognition latencies increased when stimulus size changed between learning and test phases. The effect of size change on recognition memory was found for meaningless shapes and for drawings of natural objects. Moreover, the greater the size change[5] (100% -25% versus 100%-50%) the greater the recognition impairment.

[5] To compare the effects of size change across different studies, we define the size of the original training or largest stimulus used in experiments as 100%, smaller sizes are expressed as a percentage the this maximum. We estimate stimulus size as a linear dimension (e.g. diameter, visual angle) rather than as an

Using stimuli of different sizes (100%, 75%, 50%, 40%), Jolicoeur and Besner (1987) found that the time for subjects to recognise the similarity of two simultaneously presented meaningless shapes increased when stimuli were transformed in size. Similar effects have been reported by other authors (Budesen and Larsen, 1975; Besner, 1983). Successive stimulus presentation matching paradigms produce similar results with meaningless shapes (Larsen and Bundesen, 1978), or real objects (Ellis et al., 1989; Dickerson, 1991).

Models explaining size transformations

The size transformation results described above are explainable in the context of alignment models (e.g. Ullman, 1989) which suggest that representations of objects in memory are specified in terms of size. When an object is viewed at a size that differs from the size of the stored representation an alignment testing process is initiated, whereby feature points on the stored model are transformed in size (view and orientation) so that they align with the corresponding points on the incoming image. If an alignment is successful and multiple sets of feature points match, then the object under view is effectively recognised. Note here, that alignment processes have been envisaged which match the input image to a single stored all encompassing (object-centred) description (Lowe, 1987; Porrill et al., 1988; Ullman, 1989) or a single (viewer-centred) description of the object as seen in a canonical (typical) view and orientation (Palmer, Rosch et al., 1981). The idea of a single canonical representation stored in memory, would explain the longer recognition times for stimuli presented in an unusual size (i.e. different from the stored representation of the object), since transformation processes, which bring stored representation and input image into alignment, presumably require time. These transformations have been referred to as 'mental adjustment of size' or 'size normalisation' (Shepard and Metzler, 1971; Budesen and Larsen, 1975; Besner, 1983; Larsen, 1985) and are thus analogous to 'mental rotation' described earlier. Therefore, similar problems arises, such as the

area. Thus, if a square stimulus with the sides 10° (area 100 square degree) is changed to a square stimulus with sides of 7.5° (area 56.25 square degrees), we refer to the stimulus sizes as 100% and 75% respectively.

question of which is the correct size to scale to and the direction of scaling (i.e. go smaller or larger) in a complex scene.

Exceptions to the rule

Most experiments show size scaling. Biederman and Cooper (1992b), on the other hand, suggested a size-invariant system of object recognition. They used a priming technique to investigate the role of size in object classification. They briefly presented subjects with two sets of 32 pictures. Objects in the second set were depicted as 1) the same exemplar, same size; 2) same exemplar, different size; 3) different exemplar from the same object class (e.g. grand piano and ordinary piano having the same name but different visual shapes), same size; or 4) different exemplar, different size. Stimuli in the two sets could be either 6.2° (100%) or 3.5° (56%) in diameter. It was found that naming responses for second set items were faster (i.e. 'primed'). The magnitude of priming was not significantly influenced by size change, hence Biederman and Cooper (1992b) concluded that object classification was invariant across size. This lies in contradiction to the size effects seen by others.

An object or shape identification system which generalises across metric attributes (e.g. size, orientation) could operate independently from other visual systems which specify size and orientation of objects but generalise across identity. Several authors, including Marr and Nishihara (1978) Biederman and Cooper (1992b), note the computational advantage of an object recognition system which ignores metric variations such as size, orientation and view. Such a system would not require separate representations of an object at different sizes, orientations or positions. Hummel and Biederman (1992) also note that even if different object sizes were represented separately, the brain would need to employ a further size-invariant system to determine whether differently sized images represented the same object.

Other priming experiments showed that size changes have an affect on episodic memory [memory for specific details (and context) in which an object was experienced]. Subjects were asked to judge whether the target image was the same shape as the first (priming) image, ignoring the size of the images. In these experiments, a change in stimulus size affected the same-different shape judgement carried out by the subjects (Cooper and Biederman, 1991).

In summary, earlier studies investigating effects of changes in stimulus size on recognition suggest the involvement of object image alignment and/or transformation of the incoming image to match the representation of that object stored in memory. These studies suggest that stored representations of objects include a specification of size. Biederman and Cooper's (1992b) priming experiments, on the other hand, suggest some generalisation across sizes.

Repetition priming for tasks involving classification of familiarity of famous people (from face images) is view or pose specific and does not generalise from face to body (Bruce and Young, 1986). Repetition priming has been interpreted as acting on visual codes (e.g. Face Recognition Units; Bruce and Young, 1986) and thus suggests that underlying visual representations of faces are view-specific and body part specific. These representations may also be size and orientation specific though the effect of size and orientation changes on repetition priming for faces have yet to be measured.

The priming tasks used by Biederman and colleagues to study object recognition are different in method from those used to study faces; Biederman and Cooper's (1992b) task involved two naming responses. It is not clear that the tasks tap the level of visual representation of objects, instead they could involve semantic or naming codes which are size and view-independent. There is evidence that size change has less affect on recognition tasks which tap higher level codes. Ellis et al. (1989) found that a small change in size (52% increase in *area*, i.e. 7.1% increase in *linear dimension*) did not affect successive matching for different exemplars in the same semantic class, but did affect the matching of different views of the same object.

From these findings and the findings described above, Biederman and Cooper (1992b) concluded that the ventral pathway (the 'what' pathway, leading to IT) stores size-invariant representations of objects, whereas the dorsal pathway (the 'where' pathway) stores size-dependent representations of objects. In support of their data they suggest that only the representations stored in the ventral system mediate object naming tasks, whereas the old/new judgement task is mediated by *both* the ventral and the dorsal representations of objects. The study of neurological patients sheds light on this suggestion.

Neuropsychological studies in humans

Neurological studies suggest that the ventral pathway is involved in processing object size. Cohen et al. (1994) studied and described two hemimicropsia patients with lesions in area 18 and 19 (posterior parts of the ventral pathway). These patients perceived an apparent reduction of the object's size when presented in one hemifield. Cohen et al. (1994) suggest that object size might be processed separately from other object characteristics such as colour or movement. These observations could suggest that parts of the ventral stream of cortical processing are involved with the perceptual appreciation of size. These areas may encode size specific information about visual features. That would certainly fit physiological observations of size sensitive feature coding in the early visual cortical areas of the monkey (V2 and V4; Desimone and Schein, 1987; Kobatake and Tanaka, 1994).

The visual coding of size information in the ventral stream of processing does not preclude additional analysis of size within the dorsal stream. Patient DF (Milner et al., 1991) who suffered damage to areas 18 and 19, and the parasaggital occipitoparietal region was able to appreciate size of objects when reaching to pick them up. DF adjusted the space between her fingers appropriately to the size of object during the reaching movement and before any contact with the item occurred. Yet, she showed impairment in matching different sized shapes of simultaneously presented stimuli (100%-50%; Goodale, Milner et al., 1991) and failed to accurately indicate the width of objects by the space between her thumb and index finger without reaching to grasp the objects.

It has been suggested that the ventral pathway from the occipital and temporal cortex is responsible for classifying shapes and recognising objects, whereas the dorsal pathway from the occipital to parietal cortex is responsible for visual control of actions. The dorsal stream must therefore be involved in the analysis of metric attributes of objects such as their size, orientation, view and position (Jeannerod, 1981; Ungerleider and Mishkin, 1982; Goodale and Milner, 1992; Jeannerod, Decety et al., 1994). To be able to recognise and classify an object, the orientation, size, view and position of that object are irrelevant. By contrast, to interact physically with an object, its orientation, size and position are critical for guiding hand and limb movements, whereas the object's identity is less relevant. Another way of characterising the

division, is the idea that the dorsal pathway is responsible for the real-time motor control, and the ventral pathway is involved in representations of objects stored in memory.

DF's abilities can be interpreted as indicating 1) intact visual pathways between visual cortex and the dorsal parietal systems involved in visual control of motor actions and 2) disrupted pathways to the ventral temporal areas which are involved in object recognition (Goodale and Milner, 1992). Thus, DF can use metric attributes of objects (i.e. stimulus orientation and size) to guide motor action using the dorsal cortical pathways, but she is impaired in recognising the orientation, size and shape of objects since these judgements rely on damaged ventral areas.

The dorsal stream, therefore, appears to be involved in processing size and orientation information. A conclusion supported by the observation of size and orientation sensitive neurons in posterior parietal areas (Sakata and Taira, 1994). Note that this does not exclude a role of the ventral system in processing the same stimulus dimensions for different purposes. Indeed, recognition of complex objects presumably depends on an initial analysis of the orientation and size of pattern elements in both the dorsal and the ventral streams.

In the intact brain it is probable that shape information analysed in the ventral system gains access to parietal cortex to guide motion. It may also be the case that the parietal cortex performs some analysis of shape (apart from size and orientation) independent of inputs from the ventral stream. If so, then patients like DF may be able to demonstrate motor movements guided by further aspects of shape (in addition to size and orientation).

Neurological studies in monkeys

Weiskrantz and Saunders (1984) trained monkeys with varying lesions (to parietal cortex, prestriate cortex, PIT, AIT, and posterior STS) on a number of visually presented 3D objects each seen under constant conditions. Subsequently the subjects were tested on transformed versions (orientation and size) of the original objects. Two transformations of size were used, one smaller and one larger than the original object. It should be noted that size transformation were not constant across different test objects: Increases in object size could range from 116% to 171% of the original object

(100%), and decreases in object size ranged from 50% to 67%. In one experiment, parietal lesions were found to produce less impairment in size generalisation compared to foveal prestriate, or AIT lesions. The impairment being evident on initial transfer trials and after repeated testing.

In a second experiment, normal monkeys made 28% errors on initial size transformation trials. After repeated testing with the new size transformed target objects, performance improved to 12% errors (Weiskrantz and Saunders, 1984). Successful generalisation may have been partly accomplished because of the relatively small size change during generalisation and because of information from visual features (e.g. colour) that are size-invariant. In the second experiment, AIT and prestriate lesions produced relatively mild impairments (compared with control subjects' performance) which were not visible on initial transfer trials (prestriate 24% and AIT 33% errors), but were evident with repeated testing (prestriate 18% and AIT 21% errors).

The interpretation of the effects of lesion on generalisation across size is similar to the interpretation of orientation invariance (for fuller discussion see chapter VII). V4 and PIT cortex may be involved in specifying size specific (viewer-centred) information about object features and AIT (or later areas in the ventral stream) may contribute to size-independent coding of these features (object-centred representations). The location of the lesions is therefore highly important in determining which ventral pathway areas are involved in generalisation processing.

In the study of Holmes and Gross (1984b) (also described in chapter VII), monkeys were highly trained on a simultaneously presented discrimination task (stimuli: J vs π, and P vs T). Both stimuli were transformed on probe trials in their real size (original size 100%, transformed sizes 50% and 200%). IT lesioned animals took longer to learn the original discriminations but generalised across size (and orientation) transformations in a way directly comparable to unoperated controls.

If one is to maintain the postulation that AIT contains much of the neural apparatus for generalising over viewing conditions then it is necessary to allow for at least some size generalisation in feature processing prior to AIT, otherwise it would not be possible to account for the results of Holmes and Gross (1984b). One alternative explanation of maintained size generalisation after IT lesions in the Holmes and Gross

study is that in the Wisconsin General Test Apparatus which is typically used to train and test subjects, the distance from the subject to the stimuli can vary. When the stimuli are presented, the monkey may be at the back of its transport cage before making the behavioural choice. It is therefore likely that the subject experiences the target over a range of different retinal sizes (the range of image sizes could be as large as the range during transformation testing). The monkey could learn that all the different retinal image sizes are associated with reward and thereby achieve generalisation, particularly with the protracted testing that IT lesioned monkeys require to reach criterion.

Single cell studies

Size specificity and generality

Studies of early visual areas such as the striate cortex (Kulikowski and Bishop, 1981; De Valois et al., 1982) have shown that cells are tuned to the size of individual elements or to the spatial frequency of repetitive patterns. Cell responses to sine wave stimuli have a bandwidth tolerance of 1.4 octaves [changing image size by a factor of 2.8 (e.g. from 100% to 36%) reduces cell responses to half (one octave is approximately one double]. Single cell studies in V4 have shown that V4 cells are selectively responsive to simple or complex features. The study of Kobatake and Tanaka (1994) compared large and small bars of the same width but differing in length 10-2° (100%-20%). Some cells were insensitive to change in length over this range. However size specificity for more complex shapes has not been extensively tested in V4 to date. Size selectivity is likely to be higher in V4 than in later areas PIT/AIT simply because receptive fields are much smaller in V4.

In subsequent stages of the ventral visual pathway (IT), cell response selectivity to complex stimuli generalise across different spatial frequencies (Rolls et al., 1985; 1987). Cells selectively responsive to the face responded well to both high- and low-pass spatial frequency filtered images of faces where the difference between low-pass and high-pass cut-offs was on average 3.2 octaves. This shows a convergence of visual information about the same type of object (face) from different spatial scales on to individual cells. Note, though, that spatial scale is not the same as size.

Shape sensitive neurons in IT and STS cortex respond relatively independently of stimulus position within the visual field. The cells have extremely large receptive fields (median size about 25° x 25°, or larger in more anterior regions; (Gross, Rocha-Miranda et al., 1972; Desimone and Gross, 1979; Schwartz, Desimone et al., 1983; Fujita, Tanaka et al., 1992; Gross, 1992; Tanaka, 1992).

Lueschow et al. (1994) investigated IT cell sensitivity to different image sizes during performance of a delayed matching-to-sample task in macaque monkeys. On each trial the subjects were presented with a series of picture stimuli and were trained to respond when the first stimuli of the picture series reappeared. The monkey was trained to ignore changes of size between presentations of the target picture. Two different stimuli sizes were used: the smaller typically 2° (50%) and the larger 4° (100%) wide (i.e. a two fold change; one octave). Lueschow et al. (Lueschow, Miller et al., 1994) found that some of the visually responsive and stimulus selective cells (57%, 21/37) showed no difference in their response to changes of stimulus size, i.e. these cells were invariant to size changes. Other cells (43%, 16/37), however, did show preference to stimulus size. Of these cells, 44% (7/16) responded to *all* objects at either the larger *or* the smaller size, whereas 56% (9/16) were selective for a *particular* stimulus at a *particular* optimal size. It should be noted that cells responding independently of stimulus shape might be considered visually unselective and, therefore, unlikely to be sensitive to size. Nonetheless, some cells were found to be responsive to any large shape (unselective for shape, but selective for size). Lueschow et al. (1994) suggest that object size is coded in the same way as other object features (orientation, colour, shape etc.). Information about object's features (size, shape and orientation), and information about the object's identity (shape selectivity with generalisation over change in size and orientation) is coded by different cell populations in the IT cortex. Depending on the type of information about an object needed for task solution, different populations of cells would be relevant.

A more recent study of shape selectivity in AIT confirms the existence of cell types that generalise across size change and other cell types more specifically tuned to a particular image sizes (Ito, Tamura et al., 1995). For each cell the optimal stimulus shape was first defined and then this shape was presented at different sizes with a size change of four octaves (6.25%, 12.5%, 25%, 50%, 100%, where at 100% the stimulus

size was at a distance of 57 cm with an visual angle of 31° x 29°). The width of tuning to size varied between cells. Out of a total of 28 cells tested, 43% (12 cells) were size selective with size changes of more than two octaves from optimal size (e.g. from 100% to 25%) eliminating responses (this tolerance is close to the tolerance level for generalisation in the empirical study reported below, see discussion). Of these 12 cells, 5 cells responded best to the largest size tested (>25°). For the remaining cells (16/28, 57%), size change was less influential. 21% of the cells tolerated a size change of at least four octaves (e.g. from 100% to 6.25%).

Ito et al. (1995) noted that neither size of the optimal stimulus nor the bandwidth of size tuning correlated with the posterior-anterior position or depth (cortical lamina) of the recording sites within the IT cortex. Given the relatively small number of cells sampled, the generality of this last observation is limited. The study of Ito et al. (1995) indicates that AIT processes information about an object in both a size-dependent and size-independent way.

To summarise, single cell studies of the IT cortex of the macaque monkey have shown that the majority of cells are broadly tuned to size. 79-82% of cells respond maximally to a particular size (with half of these cells exhibiting narrow-band size tuning and half exhibiting broad band tuning) and 18-21% of the cells generalise across all sizes tested (e.g. 100% to 6.25%) (Lueschow et al., 1994; Ito et al., 1995). Obviously, the degree of generalisation depends on the degree of size change tested.

The studies described so far have investigated size tolerance of cells sensitive to simple shape attributes, but it appears that selectivity for more complex shapes can be accompanied both by size specificity or size tolerance. Initial qualitative studies of 14 cells selectively responsive to faces in the STS (Perrett, Rolls et al., 1982) indicated size tolerance with cells responding to real faces both at a distance of 0.2 meters (subtending 70°, 100%) or 2.0 meters (10%). Rolls and Baylis (1986) studied cells selective for faces in IT and STS cortex for their responses to different stimuli sizes [100% (full size, 12° of visual angle at a viewing distance of 1m), 50%, 25%, and 12.5%, and 200% where the central part of the image was doubled in size]. Some cells (6/33) showed a consistent neuronal response over a wide (not specified) range of sizes, i.e. show size invariance. Other cells (13/33) showed broad tuning (i.e. the cells would respond to a wide range of different sizes but showed a response decrease with

very large or very small image sizes). 14/33 were tuned to a specific size (3 cells selectively responded to small images whereas 11 cells were selective for large stimuli).

In summary, within the temporal cortex (area IT and STS) the majority of cells appear to be size selective, whether the cells are tuned for particular simple features or for the multiple features present in a complex object. A minority (20%) of feature sensitive cells have been found in IT cortex which generalise over size change (from 100% to 6.25%). Furthermore, in the IT and STS cortex some 20% of cells, selective for complex objects (faces and whole bodies), generalise over large size changes (100% to 12%).

Effects of size change on memory for stimulus shape

Two types of short term memory effects (enhancing and suppressive) are evident in the responses of IT cortex cells (Miller and Desimone, 1991; Miller et al., 1993). The enhancement mechanism is evident when the subject has to actively hold a target image in mind and compare it to subsequently presented images (Miller and Desimone, 1994). In this case, a cell's activity is increased when the target image is presented for the second time. A suppressive effect is evident during straight forward stimulus repetition. Here, the cell's activity is decreased if the same image is presented more than once in quick succession. This cell behaviour occurs even when an image is repeated independent of whether the image is a target image or not. Thus, in the sequence ABBA, where stimulus A is the target, cell responses to B will be reduced on the second presentation.

It has been suggested that this type of suppressive mechanism may mediate automatic detection of image repetition (or registration of familiarity, (Rolls et al., 1982), whereas the enhancement mechanism may be involved in more cognitive 'working memory'. It is interesting to ask how change in image size affects these memory related effects. Lueschow et al. (1994) found that most cells showing suppressive effects with stimulus repetition and cells showing enhancement effects during matching to sample tasks exhibited size invariance in these effects. That is, a size change between first and second stimulus presentations did not seem to affect the magnitude of response suppression or enhancement. Mechanisms underlying both

short-term memory effects must, therefore, exhibit tolerance over size change. Note that these suppressive and enhancement effects can show size generality even though the tuning to single stimuli may show size selectivity. For example, a cell responds to A at 10° but not to A at 5°. Nonetheless, when changing the object size from 5° to 10° the cell may show suppressive effects.

Repetition priming tasks, where reaction times to classify pictures are reduced when the stimuli are repeated (Biederman and Cooper, 1991b), might be a psychological reflection of the short-term memory mechanisms apparent in the temporal cortex (Lueschow et al., 1994).

Evoked potential studies in humans

Physiological studies in humans have so far been restricted to visually evoked potentials. Jeffreys et al. (1992) recorded 'vertex-positive peaks' (VPPs exhibiting latencies of approximately 150 ms) from the scalp of human subjects to images of faces. Jeffreys et al. (1992; also see Boetzel and Grüsser, 1989) suggested that the VPP may reflect the activity of face selective cell populations in the temporal cortex (lingual and fusiform gyri Allison et al., 1994). Jeffreys et al. (1992), investigated the effect of changing the image size on the VPP. No significant differences in the shape and the maximum amplitude of the VPP were apparent when changing the stimulus size with fixation at the centre of the image [face length: 8° (100%), 4° (50%), 2° (25%), and 1° (12.5%) at a viewing distance of 64 cm]. With increasingly eccentric fixation points, however, the VPP was reduced in amplitude and delayed. The rate of reduction of the VPP with eccentricity, depended on the starting size of the stimulus, which is perhaps not too surprising as small faces are not resolvable in the periphery.

'Backward Masking' experiments also showed that stimulus size did not influence the VPP's evoked by faces (Jeffreys, personal communication). The brief presentation of a target face stimulus (S) was rapidly followed by a masking image (M) which was spatially superimposed on S. With targets presented for approximately 20-30 ms and M for 200 ms with a stimulus on-set asynchrony of 50 ms, size change (100%-14%) between S and M did not affect the shape, amplitude or duration and on-set of the VPP.

The empirical study reported here is concerned with the extent to which neuronal responses of cells in the temporal lobe, within the anterior superior temporal polysensory cortex (STPa) show an ability to generalise across different object sizes (size invariance), or show size specificity in their responses to heads, bodies or whole bodies (mainly whole bodies tested, see below) presented in different sizes. Previous studies suggest mainly size specific coding prior to STS and greater size generalisation in STS. It would be interesting to note whether size is coded by the visual system in a similar fashion as object orientation in the picture plane (see chapter VII). In addition, neuronal response timing factors to different sizes of the stimulus for a cell population will be described and discussed.

METHODS

Testing methods

General single cell recording methods and techniques have been described in detail in chapter IV. Each cell was first tested in an exploratory way using visual, auditory and somatosensory stimuli. If a cell was classified as selectively responsive to visual stimuli of the whole body, more specific testing was carried out (see chapter IV). Neuronal responses were recorded for both real 3D objects and 2D objects (video disk images and slides).

All cells tested showed a significantly different response to the whole human body than to control stimuli and spontaneous activity (S/A). For each cell the optimal perspective view was first defined for the upright body, and subsequently the cell was presented with that view of the whole body in different sizes. Note that the entire body stimulus was used for testing purposes. This is because cells selectively responsive to only one component part usually responded to the entire body significantly better than to S/A and control objects (see chapter V). In addition, it is assumed that size selectivity is constant across different body parts (and the entire body) as seen with view. That is, if for example a cell is selectively responsive to a component part (e.g. the head) presented in a particular size, then it is expected to find that the cell is also (maximally) activated by the whole body stimulus presented in the same size as the component part stimulus. This makes sense, if one follows the argument from chapter V, that component parts are associated with each other and therefore it is unlikely to

find e.g. a figure where the head is extremely small and the rest of the body is very large, or a very large head on a small body.

Visual stimuli

Cells were tested with photographic images (slides) of human bodies in varying sizes. The posture of the human body was in all cases bipedal. The human body stimuli were photographed onto a 200 ASA ectocrome slide film, with the person standing against a light grey background. Size stimuli were chosen to be ranging from 100% (1.73 m, head to toe height, subtending 24.4°), 75% (1.3 m, 18.5°), 50% (0.87 m, 12.3°) to 25% (0.43 m, 6.2°) at the viewing distance of 4m (see Fig. 8.1). Note that since all cells which have been included in this study were unselective for identity of the whole body stimulus (data not reported here), the most commonly encountered whole human body size by the monkey was ranging between 1.3m (75%) and 1.73 m (100%) (the size of other monkeys, lab technicians and other (Scottish) people being in contact with the monkeys).

In addition to the specific visual stimuli tested, each cell was tested with a number of different control stimuli (2D and 3D). These included complex 3D objects of different sizes, shapes and textures (lab coats, chairs, etc.), simple 2D geometrical shapes (bars, spots and gratings) and simple 3D forms (cylinders, balls, boxes, etc.).

Data analysis

An assessment period of a 1/4 or 1/2 second, occurring 100-350 ms after stimulus on-set, was used for data analysis.

Cell responses to the whole body presented in four different sizes (100%, 75%, 50% and 25%), controls and S/A were compared using 1-way analysis of variance (ANOVA) and post-hoc tests (protected least significant difference (PLSD), Snedecor and Cochran, 1980) with a significance level of $p<0.025$.

Individual single cell data were presented in the form of histograms indicating average neuronal responses (spikes per second) to different visual stimuli (whole body presented in different sizes, controls and no stimulus condition (S/A)).

A size discrimination index was computed for 16 cells tested for the largest (100%) and the smallest (25%) whole body stimulus size. An index formula similar to

Figure 8.1. Examples of stimuli used for testing: Whole body at size 100%, 75%, 50% and 25% presented in different views.

View: 0°
Size: 100%

View: 0°
Size: 75%

View: 0°
Size: 50%

View: 0°
Size: 25%

View: 315°
Size: 100%

View: 315°
Size: 75%

View: 315°
Size: 50%

View: 315°
Size: 25%

the one used to compute the orientation discrimination index (see chapter VII) was applied.

Size Discrimination Index (I_s) = [(response to 100% - S/A) - (response to 25% -S/A)] / (response to 100% - S/A).

Furthermore, a post-stimulus time histograms (PSTH) and cumulative response curves were plotted for population estimates of time course responses (see general method chapter IV).

RESULTS

Of all the cells tested in the STPa cortex which were visually selective for the whole body stimuli (see chapter V), a total of 16 cells were tested for size sensitivity with different image sizes of the whole body. Two categories of responses were observed.

a) Generalising across image sizes

Three cells showed size generalisation, where they responded to all images sizes at a rate greater than to S/A and control object. Of these 3 cells, 1 (see Fig. 8.2) showed no difference in response amplitude to different image sizes. This cell shows perfect generalisation across different sizes of the optimal stimulus.

Two cells, however, displayed some size tuning superimposed on a generalised response (e.g. Fig. 8.3).

b) Tuned to optimal image size

The remaining 13 cells (81%) were responsive to one (or more) but not all image sizes tested and hence showed size specificity (e.g. Fig. 8.4).

Thus, the majority of cells 15/16 showed sensitivity to size (including cells which displayed some size tuning superimposed on a generalised response, Fig. 8.3). Interestingly, all tuned cells were selective for the whole body at largest (100%, e.g. Fig. 8.4) or second largest (75% e.g. Fig. 8.5, cell B) projection sizes. These two sizes would correspond to images of real humans encountered at the projection distance. As for orientation the response tuning to image size varied across cells some showed

Figure 8.2. The mean responses (+/- 1SE) to the whole body (view 0°) and spontaneous activity (S/A) are illustrated for one cell (E61_31.82L). The cell showed no significant difference between different sizes tested (**generalisation**), but all sizes gave a significant different response than to spontaneous activity (S/A) ($p<0.05$). ANOVA: $F(4,20)=9.6$ $p<0.05$.

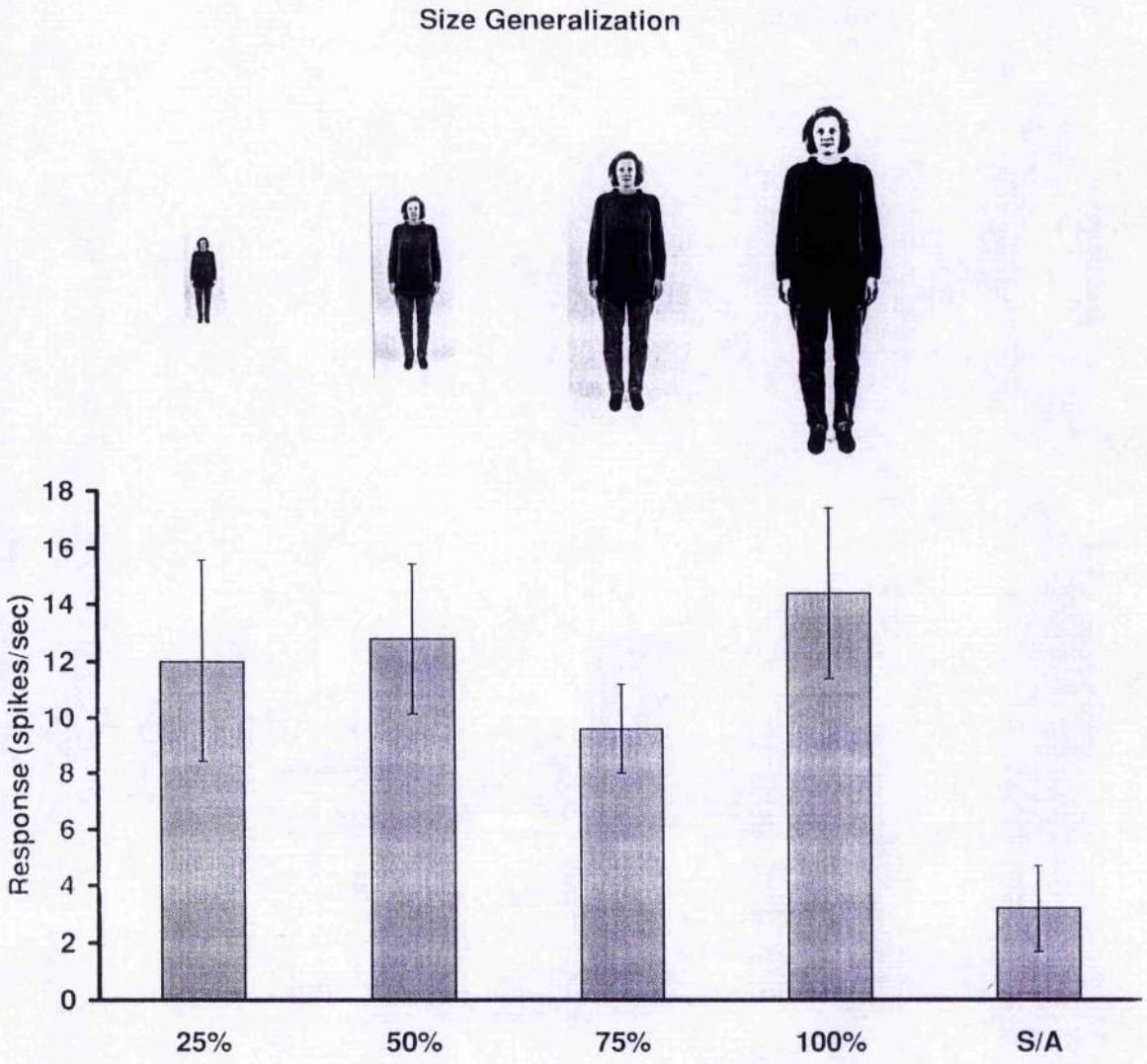


Figure 8.3. The mean responses (+/- 1SE) to the whole body (view 0°) and spontaneous activity (S/A) are illustrated for one cell (E99_40.16L). The cell showed some tuning to the whole body stimuli when tested at maximal (100%) size ($p<0.05$). The cell also responded to all other sizes better than to spontaneous activity (S/A) (**broad tuning**) ($p<0.05$). ANOVA: $F(5,24)=9.6$ $p<0.0005$.

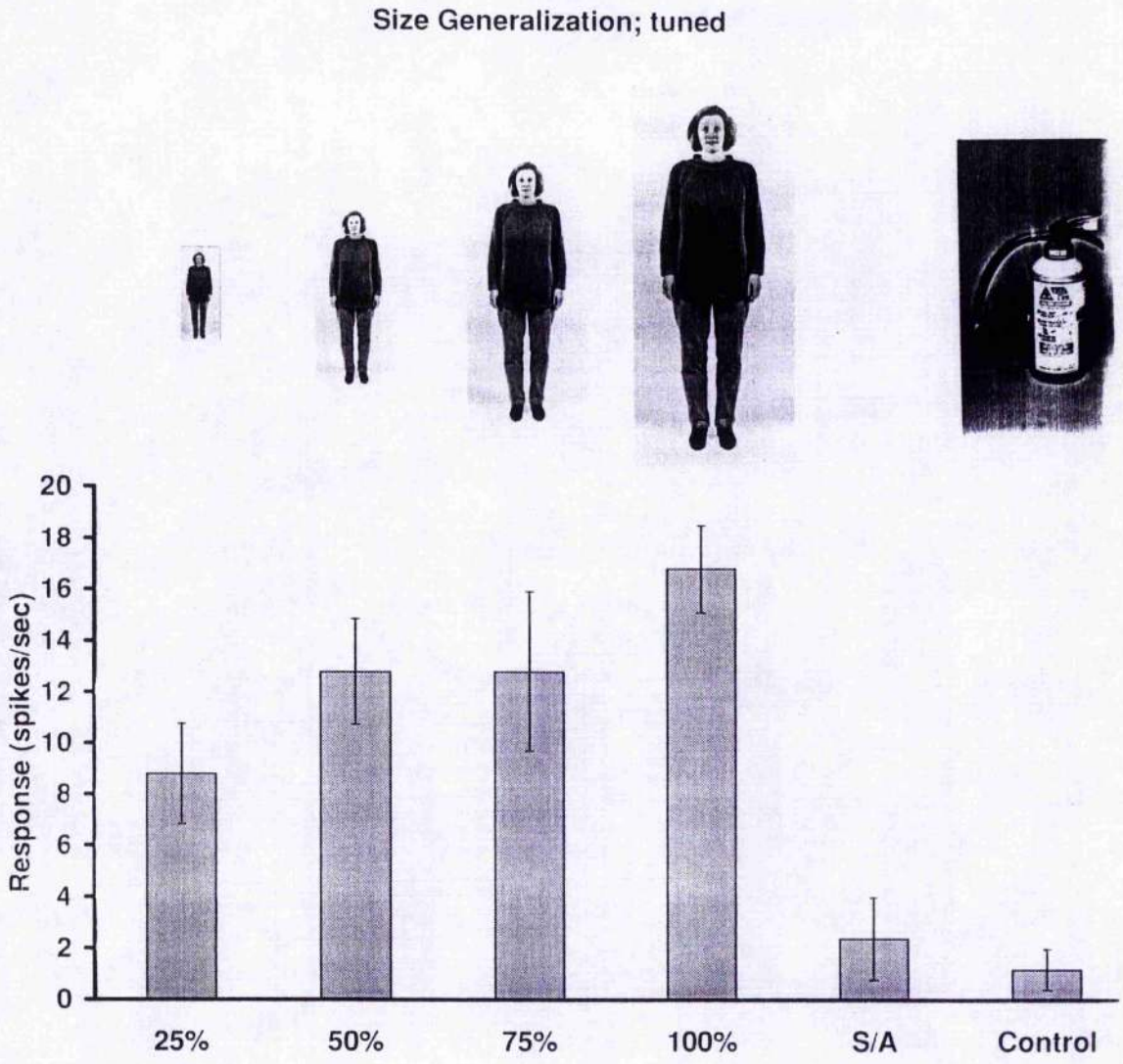


Figure 8.4. The mean responses (+/- 1SE) to the whole body (view 0°) and spontaneous activity (S/A) are illustrated for one cell (E83_37.00L). The cell responded only significantly different from S/A and controls to the whole body stimuli at 100% size ($p<0.001$). No stimulus in any other size gave a significantly different response from S/A and control stimuli (**narrow tuning**) ($p>0.1$). ANOVA: $F(5,23)=4.6$ $p<0.005$.

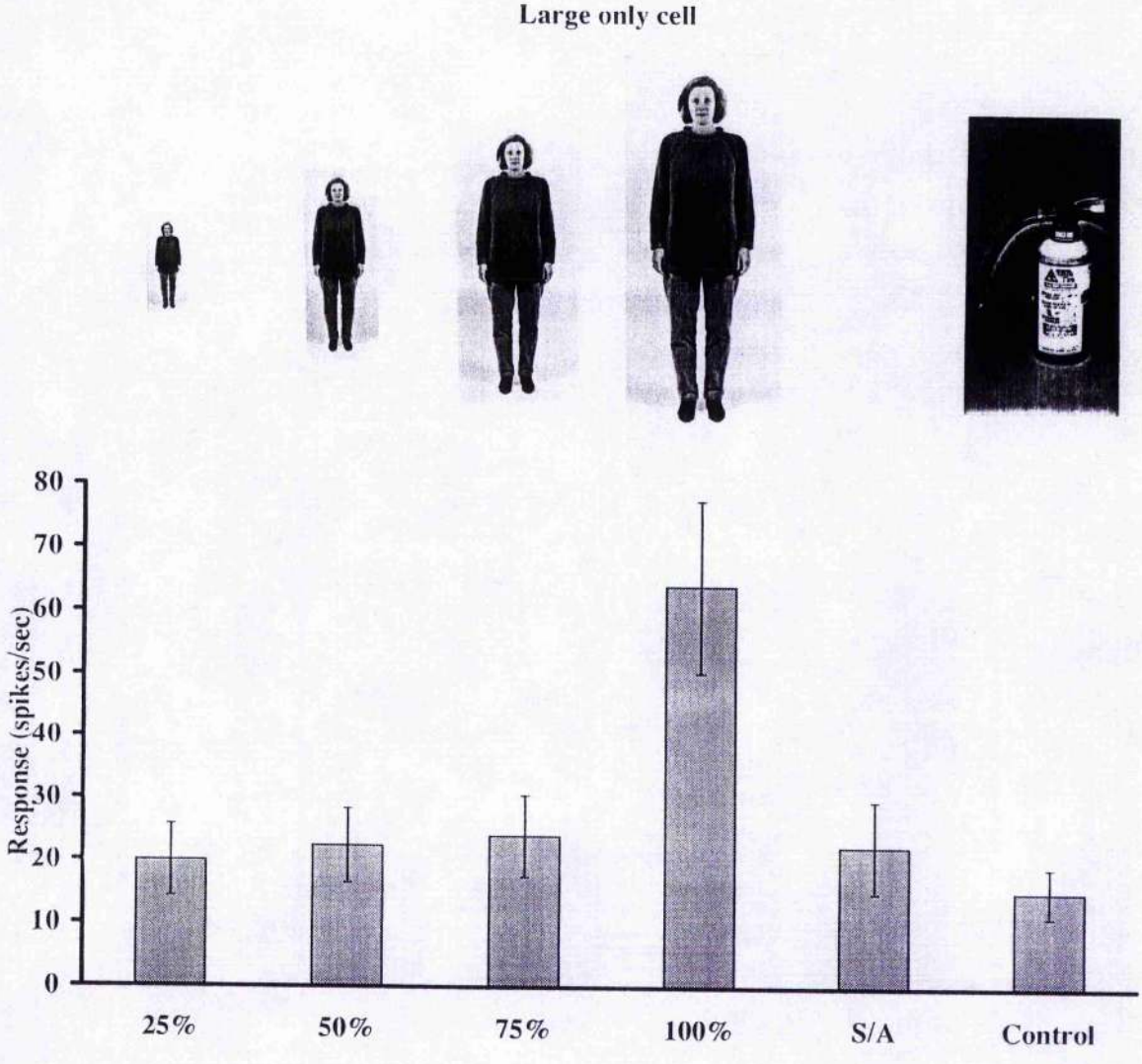
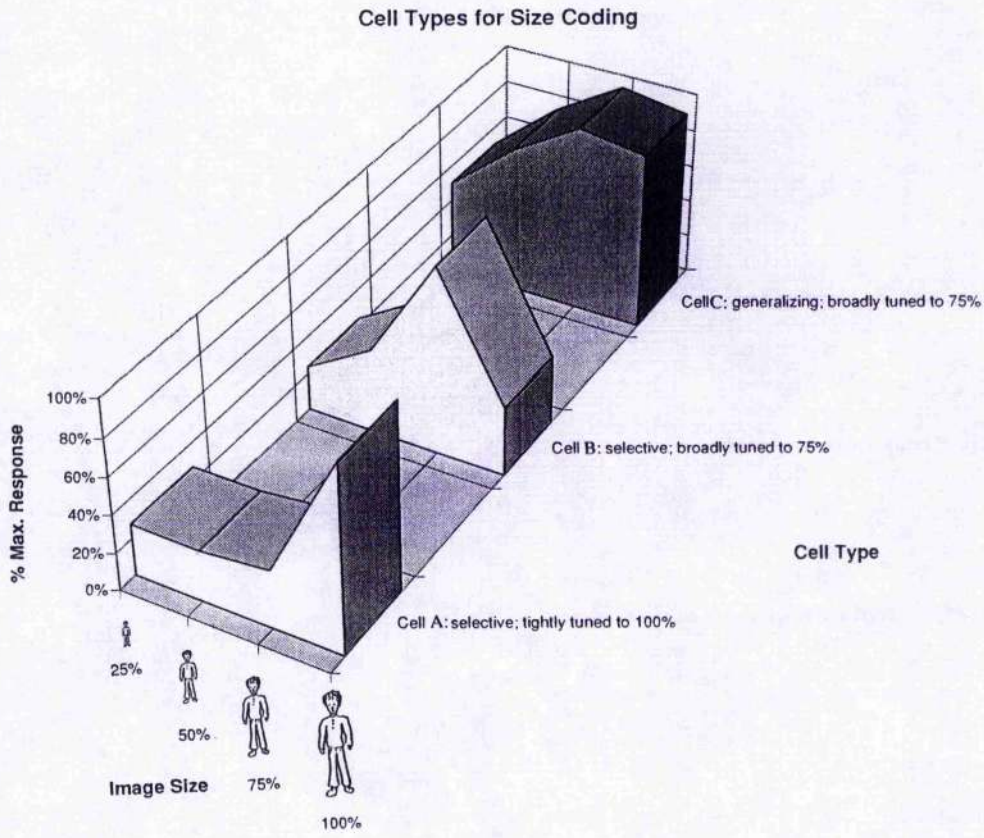


Figure 8.5. The responses of 3 cells tested with four different image sizes are illustrated. Size defined as a % of maximum size (head to toe 1.73m, visual angle 24.4°). Responses to different stimuli are expressed as a % of the maximum of cell activity (where spontaneous activity = 0%). Cell A was maximally activated by the largest stimulus size. Cells B and C were maximally activated by the by second largest image size (75%). Tuning curves are progressively broader for cells A, B and C.



narrow tuning and response to only one size of the body at a rate significantly different than to control objects and S/A (see Fig. 8.5, cell A); others showed broader tuning with the responses declining gradually as the image size progressively changed from the optimal size (e.g. Fig. 8.5, cell B and C).

Size Discrimination Index

A total of 16 body selective STPa cells which had been tested for very large and very small body stimuli were included in this analysis, irrespective whether the cells responses were tuned to a particular size or not (see Fig. 8.6). The size discrimination index shows that individual cells are quite good at discriminating between the largest (100%) and the smallest (25%) stimulus size, that is most cells have an size index value (I_S) greater than 0. If I_O = 1.0, then the response to the smallest body stimulus = S/A. If I_O > 1.0, then the cell response to the largest (100%) size > S/A and the response to the smallest body size < S/A. If the I_O value is close to 0, the cell generalises across sizes. At an negative I_O, the neuronal cell response to a small body stimulus is > than the response to a large stimulus. This situation is not observed here, since no cells tuned to small images have been reported. Finally, as a population of cells, good in discriminating occurs since most cells are selective to a particular size (see above).

Population estimates of time course of responses

Post stimulus time histograms (PSTH) were plotted for a population estimates of 12 cells tested with different sizes (all cell types included) (Fig. 8.7a). At point 0 ms stimulus presentation occurred. Both stimuli representing real life-sized images at the projected distance (100% and 75%) triggered greatest neuronal responses. Population response to the largest stimulus tested (100%, life-size) shows a slightly faster response on-set than smaller stimuli sizes tested. It is interesting to note that the neuronal response to visual control stimuli is slightly inhibited after stimulus presentation.

A cumulative response curve of the same data was plotted to obtain a better indication of the cell population response (all cell types included) (Fig. 8.7b). This graph allows one to compare levels of average neuronal response at any given time.

Figure 8.6. Size Discrimination Index. An index of size discrimination was computed (see text) for 16 cells tested for responses to the largest (100%) and the smallest (25%) body size.

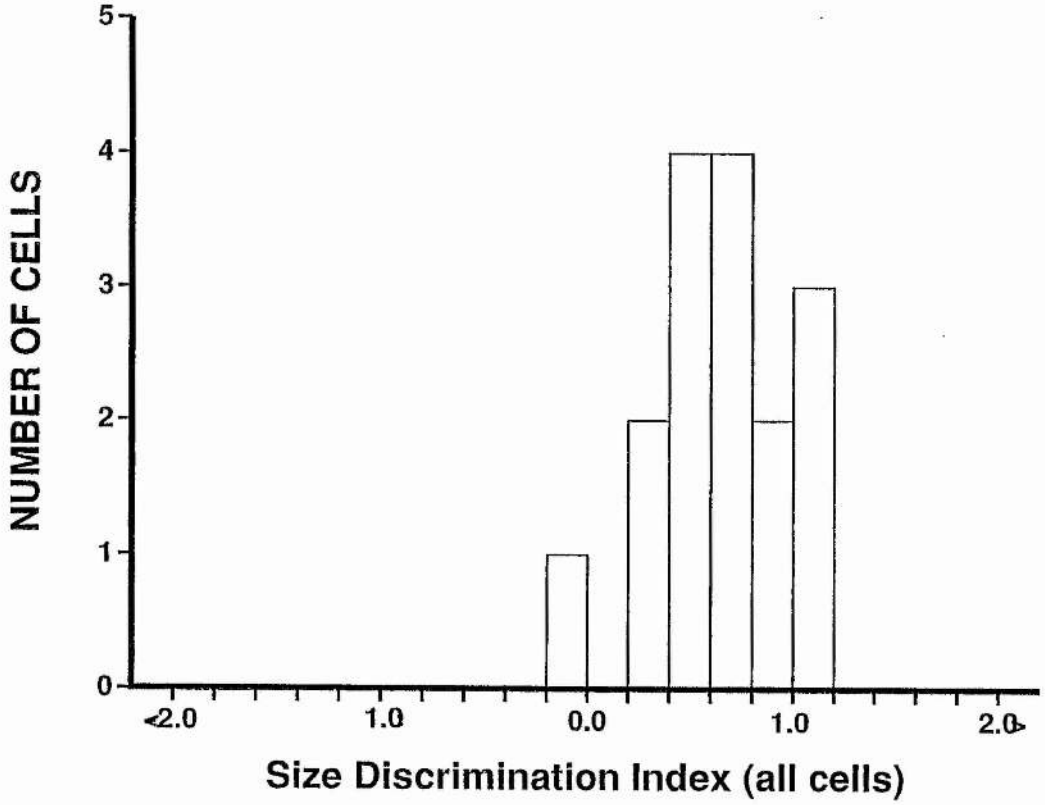
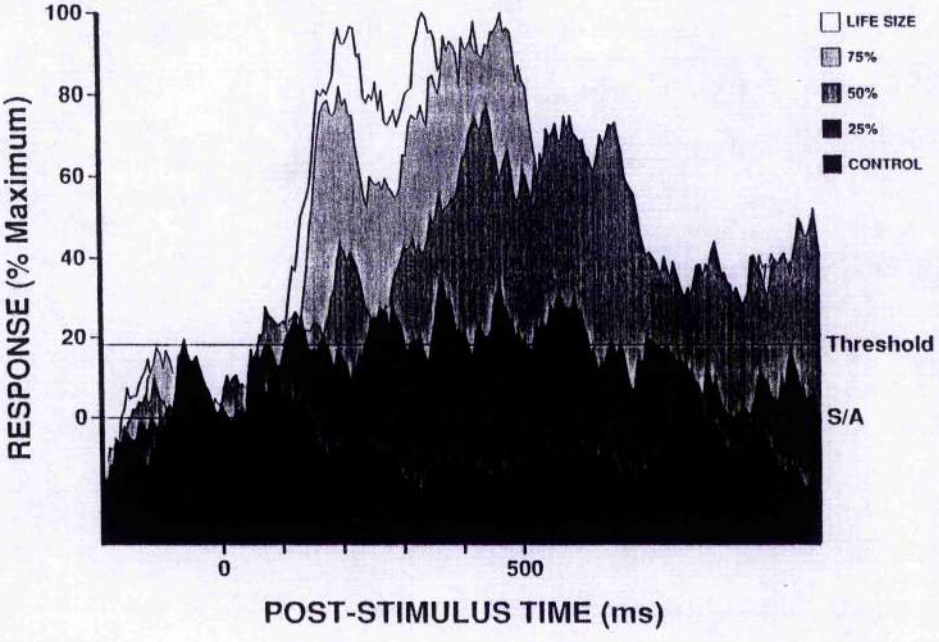


Figure 8.7. Cell population response to four different stimulus sizes and control objects. a) Combined responses of 12 cells to the five different stimuli. The clear area displays the population response to the largest (100%; 1.73 m, head to toe height, subtending 24.4°) whole body stimulus size, light grey area to the 75% whole body stimulus size, medium grey area to the 50% whole body stimulus size, dark grey area to the 25% whole body stimulus size and black area population response to various control objects. Firing rate is expressed as a percentage of the peak response to the best (in this case 75%) stimulus. Stimulus presentation occurs at point 0 ms. b) Cumulative Response curve of size population estimate. The colour coding for different stimuli responses is identical to (a). Firing rate is expressed as a percentage of the maximum response to the best (in this case 75%) stimulus. Stimulus presentation occurs at point 0 ms. The average population response rate was calculated after normalising each cell's activity (S/A=0, Max Response=100). The cumulative population response is defined as the running total of this average population response from 200 ms pre-stimulus onset.



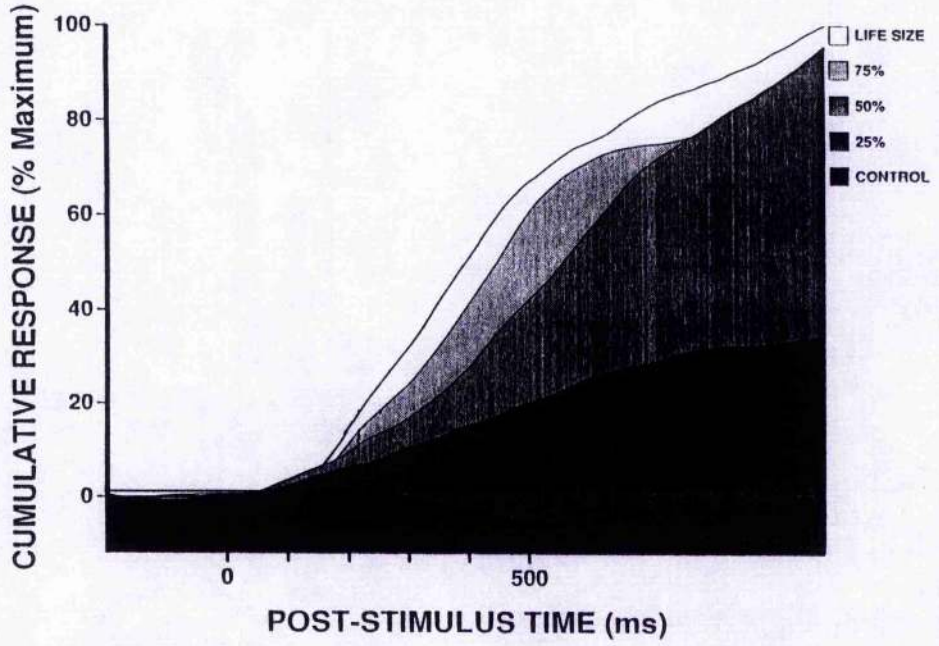
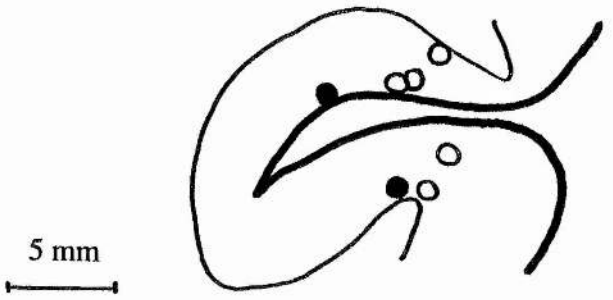


Figure 8.8. Histological reconstruction. A frontal section of subject E taken at 15 mm anterior to the interaural plane showing the location of cells tuned to the largest (100%, life-size) image size tested (open circles), cells tuned to size 75% (filled circles), and cells generalising across all sizes of the head/body tested (open triangles). The thick lines indicate the brain surface, the thin lines show the boundary between white and grey matter. Cells were located in both the upper bank and fundus of the superior temporal sulcus (see Appendix 5).



Therefore at, for example, 300 ms after stimulus presentation, the cell population responded most when a life-sized stimulus was presented. Also, if a response level is chosen (to trigger for example behavioural response times) at say 40% of the maximum response, then this level is reached much faster when life sized images are presented (approximately 350 ms after stimulus presentation), rather than smaller image sizes (e.g. 50%, response level reached only after 480 ms after stimulus presentation). Furthermore, it is interesting to note that the neuronal population response to the 50% sized stimuli is very late. This is reflected in the cumulative response curve where at a high post-stimulus time the cumulative response shadows the response to the 75% stimulus size. However, this response pattern is mainly due to one cell which had a very late response on-set, and because the number of cells in the population was so small, the late response of this cell was highly influential.

Histological reconstruction

The hisological reconstruction of two monkeys indicated that cells tested for size specificity of the head/body were mainly located in the upper and to some extent in the lower bank of the anterior part of the STS (see Fig. 8.8 for example and Appendix 5 for more detailed representation for one subject E). For subject E, the cells tested were situated at a quite anterior position of the STS, close to the temporal pole. There was no clustering of the different cell types observed.

DISCUSSION

What is size generalisation and specificity important for?

For effective, successful and fast object identification, generalising across object attributes, such as changes in image size, is very important. Coding objects in a (real life) size specific manner, on the other hand, is important for interaction with objects, e.g. adjusting the distance between fingers when picking up an object. It has been suggested that cells in parietal cortex (area LIP) code such size specific information (Taira, Mine et al., 1990; 1994; Sakata and Taira, 1994). This coding of size specific information is very important, since the parietal cortex is believed to be involved in guiding motor movements of interaction with objects. In addition, information about an object's size can carry important social information and hence

play a role in social interaction. If, for instance, a monkey or human face is visible but very small, it either does not appear to be a threat to the observer because of its small absolute size, or alternatively, it is still at a great distance from the viewer and therefore does not endanger the observer immediately.

What exactly is generalisation and specificity?

No fixed definitions of the terms 'generalisation' (analogous to object-centred coding) or 'specificity' (analogous to viewer-centred coding) are given by authors, resulting in these terms to be rather arbitrary. When comparing the findings of different studies, one has to understand the definition of generalisation and specificity given by each author. Often these definitions vary and to be able to compare across studies, these definitions have to be adjusted to each other. Before discussing some findings of different neurophysiological studies on size sensitivity, it is important to note the different terms used for defining tolerance levels by different authors (see table 1).

Table 1. Different size tolerance measurements commonly used

% of size change (from small to large)	12.5%	25%	50%	75%	100%	200%
% of size change (from large to extra large)	100%	200%	400%	600%	800%	1600%
Times / fold change	1x	2x	4x	6x	8x	10x
Doubling / octaves	0	1	2	2.6	3	4

Previous and empirical findings

Findings of previous and the current study are compared. To do so, the studies under investigation will be outlined and their findings discussed.

Experiments investigating subjects (human and monkey) with brain damage suggest that ventral stream lesions disrupt size generalisation processes. Lesion studies of monkey's IT (Humphrey and Weiskrantz, 1969; Ungerleider et al., 1977) support the proposal that the IT plays an important role in size constancy (see introduction).

Different results and conclusions were formed on the basis of single cell studies. Studies of stimulus size affects in the striate cortex suggested that these cells are 'finely' tuned to stimulus size (Kulikowski and Bishop, 1981; De Valois, Albrecth et al., 1982). That is, 68.5% (245/358 cells) of all the cells tested responded (greater

than half of the maximum response) to different stimuli sizes with a change which was more than 1.4 octaves (i.e. a 2.8 times change in stimulus size). The remaining 31.5% of the cells were more finely tuned to size, tolerating a size change of less than 1.4 octaves. The average tolerance of these cells was 1.07 octaves. The stimulus size preference of most V1 cells was approximately between 1° and 2°.

This size specific coding is also found further along the ventral stream in areas V4, PIT and AIT. Ito et al. (1995) studied size selectivity of cells in PIT and AIT and found that 43% (12/28) of the cells tested, which were selectively responsive to features of complex objects, were sensitive to stimulus size. These cells had a stimulus size change tolerance smaller than 2 octaves. The remaining 57% (16/28) of cells responded (greater than half the maximum response) to a stimulus size change of more than two octaves (size change of up to 4 octaves tested). Of these cells, 62% (10/16 cells; or 36%, 10/28) responded to a size change of more than 2 octaves but less than 4 octaves, whereas 38% (6/16; or 21%, 6/28) responded with size tolerance of at least a 4 octaves change. Rolls and Baylis (1986) recorded from single cells selectively responsive to faces in area AIT (named by them areas TEm and TEa) in the inferotemporal cortex and area TPO of the STS. They found that again 43% (14/33) of the cells were relatively specific in their responses to different stimuli sizes (gave a response within half of the maximum response). The remaining 57% (19/33) of the cells were described as more tolerant to a "wide range" of stimulus sizes (one of their examples tolerates a 4 octave size change). However, the size tolerance of these cells types are not specified in this study. Rolls and Baylis (1986) only mention the median tolerance of all cells tested which is a "12 fold size change", i.e. a tolerance across 3.5 octaves.

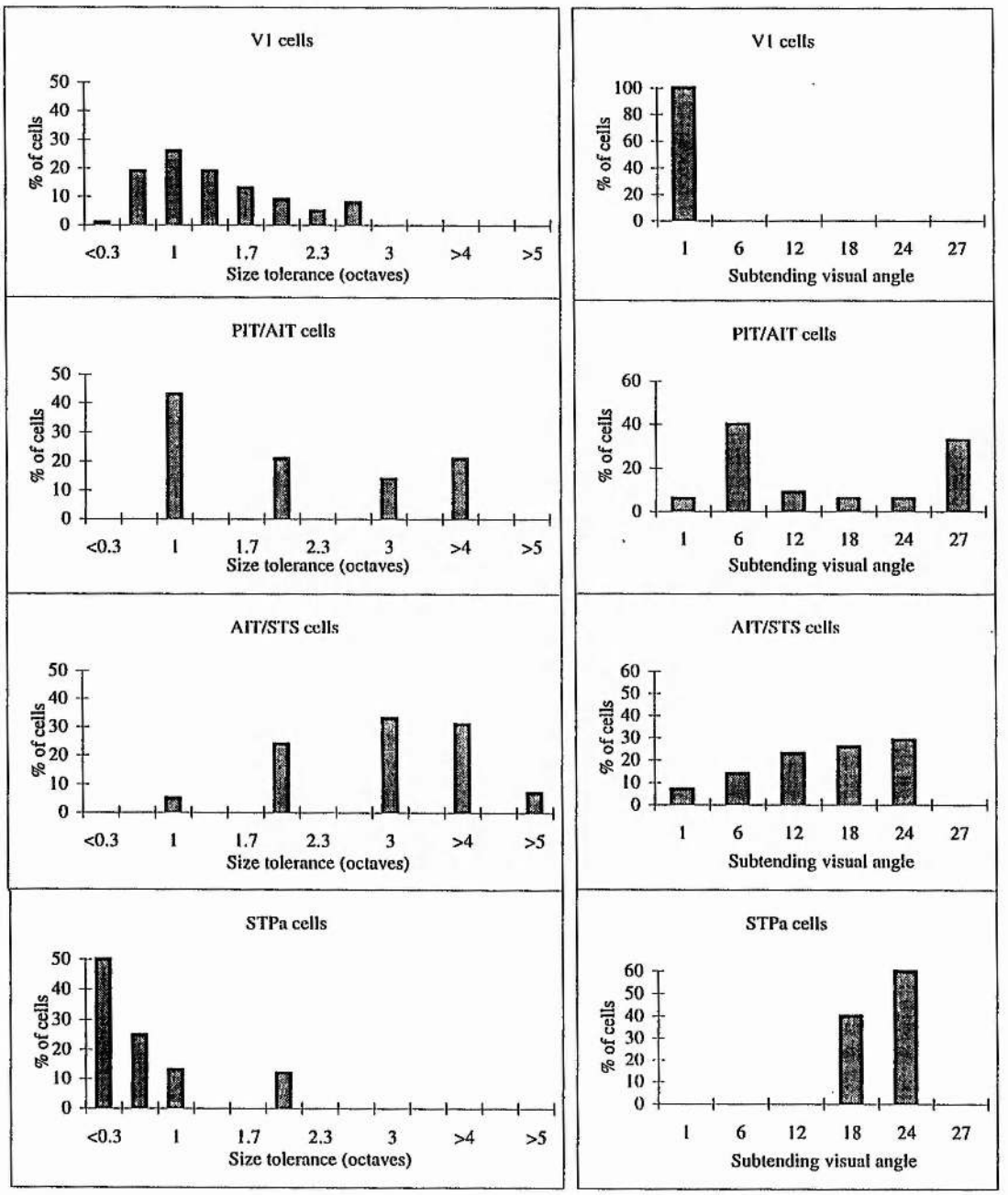
The current study was, unlike the studies reported above, classified on the basis of statistical difference in neuronal responses to different stimuli sizes using 1-way ANOVAs and post-hoc (protected least significant difference, PLSD) tests. However, for direct comparison with other studies, re-classification was carried out with an effective response being within half of the maximal response for that cell. This showed that the majority of cells (75%, 12/16) which were selectively responsive to the body (or a body part) tolerated a size change of less than one octave (8 cells responded to only one particular size, whereas 4 cells tolerated a stimulus size change from 75% to

100%). This tightly tuned neuronal response pattern to different stimulus sizes is analogous to a viewer-centred coding. The remaining 4 cells exhibit a more tolerant response pattern to stimulus size change. Two cells tolerated a size change of one octave (from e.g. 50%-100%), whereas the remaining 2 cells tolerated a stimulus size change of at least 2 octaves (25%-100%). Unfortunately, it was not possible to test even larger stimuli in this study, since at the fixed distance of 4 m the largest whole body size to fit onto the screen was already tested. If larger stimuli were tested, then only e.g. a very large head or even only two extremely large eyes would be visible on the screen. Previous findings (not included in this study; Perrett et al., 1991) do, however, indicate that many cells responsive to faces respond to very large faces of approximately 12.5° and real life sized faces (4°), but are less likely to respond to very small faces (2°) presented at 4 m. Thus, the empirical study reported here might have missed the size tolerance cells when extra large faces were presented. Nonetheless, the current study shows the size tolerance of cells when stimuli were tested which were decreased in sizes in comparison with real life sized images at an appropriate and constant distance.

Figure 8.9 gives an overview of the size tolerance of the cells in different ventral visual areas. The figure indicates a hint of greater size tolerance of cells tested at more anterior locations. The empirical study possibly distorting the size tolerance representation (see above). Nonetheless, no study reported only cells which generalised across size differences and hence size specific coding is highly important. This size specific coding be parallel to the increased complexity of the stimuli. Biological objects such as heads and bodies have a real life size at particular distances, whereas more abstract features found to be represented in PIT and AIT are less size specific in real life, though they do change retinal image size with varrying distances. This might be reflected in that the more anterior cells are located in the processing ventral stream the larger the optimal size of the stimulus (see Fig. 8.9).

It should be noted that cells in different visual areas have different sized receptive fields (see chapter II). That is, early visual areas such as V1 have very small receptive field and as cell's receptive field sizes are tested further along the ventral visual pathway (in areas V4, PIT, AIT, STPa), they become larger and larger. This will influence the preference of absolute stimulus size and the amount of size change which

Figure 8.9. Size tolerance. A cross study comparison of size tolerance of cells in different ventral areas (see text). V1 cells tested and reported by DeValois et al., 1982; PIT/AIT tested and reported by Ito et al., 1995; AIT/STS tested and reported by Rolls and Baylis, 1986; and STPa cells taken from the empirical study reported here.



can be tested and tolerated by the cells. If a visual field is e.g. less than 1 degrees^2 as for V1 cells, then only a small size change can be tested. If, however, the visual field is very large (for example, 150 degrees^2 as for AIT cells) a much greater range of different sizes can be examined. Therefore, if a STS cell is selectively responsive to a large (life-sized) stimulus then this stimulus might not cover the entire receptive field of that cell.

It is also suggested, as for orientation and view coding, that size specific (viewer-centred) cells may be pooled together, i.e. feed into cells with size generalisation (object-centred) properties.

Familiarity and Experience

It was observed that cells in the ventral pathway which coded size specific information were tuned to the real, or close to real life size of the stimulus. In Ito's et al. (1995) study, 33% (5/12) of the size specific PIT and AIT cells were found to selectively respond to only the large stimulus, i.e. with a visual angle greater than 25°. These stimuli were, however, not biological stimuli such as faces or bodies and therefore one cannot say that these stimuli were tuned to life size (i.e. there is no defined life size for a triangle with a bar across). Studies which tested cells which were selectively responsive to faces or bodies showed that the majority of size specific cells were tuned to real life size images (faces/bodies). Rolls and Baylis (1996) reported that 79% of a size specific cells in AIT and STPa were tuned to large faces (close to real life size). The empirical study described here reported that all STPa cells tested which showed some size tuning were either maximally activated by the 75% or the 100% stimuli (life size). It is therefore suggested that experience and familiarity of particular object size increases neuronal representations of that object in that size within the STPa.

Coding different image sizes

The size discrimination index indicates that the majority of STPa cells tested are relatively efficient in discriminating between the largest (100%) and the smallest (25%) body size. The closer the index value is to 0, the more the cell generalises across different sizes. Only very few cells are able to generalise across all image sizes tested

(that is from 25%-100%, which is two octaves, or two doublings). Note that there are no cells which are optimally tuned to the smallest image size, i.e. cells with a negative index. This lack of coding extreme small faces/bodies might be due to the familiarity effect. Monkeys are not familiar with seeing such small human heads/bodies at a distance of 4 m (testing distance used in empirical study).

Population response pattern to different stimulus sizes

The PSTH and the cumulative response curves described earlier reflects the overall neuronal response of cells selectively responsive to the face/body to different image sizes presented at a constant distance. The PSTH indicates that there is a slight response on-set (latency) difference between responses to different stimuli sizes. The smaller the stimulus the longer the latency. Looking at the cumulative response curves, however, it is clear that life-sized bodies (100% sized images) always evoke a stronger response at any response level (or time), than responses to smaller stimuli. This could explain behavioural findings which suggest that closer to life-sized images (or the image size the subject has been trained on, or is familiar with) will be recognised faster than non-life-sized images of objects (Besner, 1983; Jolicoeur and Besner, 1987). Furthermore, the physiological data described here supports the psychological finding that the greater the stimuli size change, the greater the recognition impairment (Jolicoeur, 1987).

Chapter IX

GENERAL DISCUSSION

In this final chapter of the thesis, issues on object recognition and effects of part occlusion, view, orientation and image size will be put forward and discussed. Particular cells which have been tested with different types of image transformation will be described and their function in visual object recognition debated. Examples of cells rather than population responses are noted, since only a few cells were tested with all transformations under investigation (part occlusion, perspective view, orientation and size change). This was often due to limited time during a recording session (that is the monkey was only allowed to be restrained up to 4 hours), and the decreasing amplitude of the response signal of a cell when recording over a long period of time. Furthermore, neuronal signals to 2D slide image are usually weaker than responses to 3D objects and therefore one is more prone to losing the cell's signal.

Cells which code information by inhibition

Some cells were observed to decrease their neuronal firing rate from spontaneous activity (S/A) when a head/body was viewed. There was no change in neuronal activity compared to S/A when any other (control) objects was presented. For example, one cell (E21_3190R) discriminated between different views of the head/body, i.e. the cell response was only inhibited when the head/body stimulus faced the monkey directly (i.e. tuned to view 0°), but neuronal response activity was as high as S/A when other views of the body were presented. When the cell response was tested for its sensitivity to the optimal stimulus (front view of the head/body) under different transformations (part occlusion, different orientations in the picture plane and different image sizes), then the cell response continued to display dramatic inhibition. Inhibitory behaviour to the optimal stimulus is quite rarely observed (4/77 cells), representing an unusual way of coding information. Such inhibitory coding is expensive in metabolic energy since the cell will fire at a constantly high rate when no head/body (view 0°) is in view. However, considering the experimental monkeys

environment (living in the same room as a large breeding colony and its frequent encounter with humans) such inhibitory neuronal coding maybe explainable. At most times, the experimental monkey is likely to view another facing monkey or human, which would result in the decreased response of such a cell. Previous studies have also reported inhibitory responses of visually responsive cells. It is interesting to note that cells in early visual areas such as V1 show frequent inhibitory responses (e.g. simple, complex and end-stop cells; see Hubel and Wiesel, 1962, 1968), whereas studies of cells in temporal cortical areas report a low percentage of inhibitory responses (Gross et al. 1972; Logothetis et al., 1995).

What kind of information do single cells code?

Some cells (9/20) which have been tested with all transformational attributes show selectivity to a particular view, orientation, and size of the image. That is if a cell is sensitive to view then this cell is likely to be sensitive to orientation and size. Other cells (8/20) are sensitive to most but one transformation of the image (e.g. sensitive to view and orientation but not to image size). Only very few cells (3/20) show generalisation across most transformations. Examples of different types of cells are given below and their function in coding information about the head/body are discussed.

As stated in the various empirical chapters, the majority of cells reflect a reference frame based on the viewer when coding information about object transformations. Such viewer-centred descriptions have been strongly reflected in coding object parts, perspective view, orientations in the picture plane and transformations of the object size. For instance, one cell (E24_33.44R) which was examined with all transformation tests described in this thesis was found to code specific part information (head alone), preferred a particular perspective view (front view), orientation (upright) and size (life-size) of the head/body. Such a cell would signal when a life-sized upright person is facing or coming towards the observer. The cell would continue to signal even when the body (torso and limbs) is occluded from sight, which is often the case when viewing heads/bodies in the real world.

Nonetheless, often information about one of the transformational attributes tested (part occlusion, view, orientation or size) was coded on the basis of a different reference frame than other attributes. Different combinations of coding specific and general information about an object may contribute to representing particular information about that object. For instance, to recognise a full cup of coffee it would be important to identify the view of the cup and its upright orientation. If the cup was not upright then it could not contain coffee. The size of the cup is irrelevant at the stage of recognition, though it may become significant for visuomotor interaction. One cell (E83_38.31L), for example, was found to be selectively responsive to one body part (head alone), one perspective view (front view) and to the life-sized image, but generalised across different head/body orientations in the picture plane. This cell signals that a life-sized (or very close) person is facing them, independent of whether they hang upside down or not. This orientation invariance can be accounted for by the frequent change in head/body orientation observed in monkeys.

Another cell (E79_35.14L) generalised across all transformational attributes except orientation. For the cell to respond, the head/body image had to be viewed in its upright orientation, though it did not matter whether the image was small or large, which perspective view of the head/body was observed and whether parts of the body were occluded or not. Therefore, only information about the head/body orientation was coded in a viewer-centred manner where the reference frame is based on the observer or gravity rather than on the object itself.

On the basis of cells coding such specific transformational attributes of the object, it is suggested that some of the STPa cells reported here may code not only information about the structure of an object, but play an important role in coding social signals and the direction of attention of the person/monkey being viewed. For example, attention to the right of the observer may be reflected in neuronal coding of the right profile in its upright position or the left profile inverted. Furthermore, it is suggested that the head carries more social information than the body presented in isolation. It would therefore be expected to find more STPa cells coding information about the head than the body presented in isolation. This suggestion is supported by the greater number of cells coding information about the head. Nonetheless, some information can

be obtained from the body alone (e.g. a person walking away), so that these cells are not redundant in the processing of social signals.

Little object-centred coding at more anterior parts of the processing pathway

It is interesting to note that there was only very small amounts (10%) of object-centred coding observed at such anterior sites of the ventral visual processing pathway. There was little temporal cortex anterior to the site of recording, indicating that the cells described here were located close to the temporal pole at the most anterior parts of the STS. It is therefore suggested that if object descriptions which are based on the object itself and are able to tolerate all existing image transformations, then such descriptions would mainly be located at different brain sites to the STPa. These sites are expected to receive neuronal inputs from the temporal cortex and may not be predominantly visual areas (e.g. Amygdala, see Leonard, Rolls et al., 1985). Pure object-centred descriptions carry only information about the shape of the object and no additional information about, e.g. view or orientation, can be obtained. Therefore, such descriptions are rather limited in providing specific information to guide physical interaction or allow social interpretation. Object-centred representations might support specialisation of what an object is but will not aid in interaction with the object (Milner and Goodale, 1993).

An outline of a object processing scheme

Cells located in the earlier cortical areas of the ventral visual processing pathway (V4 and PIT) have been found to be selectively responsive to simple shapes presented in a particular orientation and size. Such cells may be the building blocks of elaborate cells (Tanaka et al., 1991, 1992) located mainly in AIT which are selectively responsive to a combination of simple features, while maintaining selectivity for orientation and size. The output of these elaborate cells can be combined to create cells which are highly selective for particular shapes and may even be capable of representing complex 3D objects such as heads/bodies.

The neuronal representations of complex objects may, as reported above, tolerate transformational changes in one domain (e.g. size), but not in others (e.g. view,

orientation). Such coding maybe explained by either a functional advantage (see 'upright coffee cup' or social attention example given above) or by experience of associating certain transformations with certain actions. For example, if a person is walking toward the observer, the (front) view and (upright) orientation of that person is held constant, thought the image size changes. Experience and familiarity with that combination of image transformations may be reflected in neuronal coding of a body moving towards the observer. Cells may therefore code specific view and orientation information of the body but generalise across changes in image size.

The empirical studies reported here indicate extensive selectivity for view, orientation and image size. It is suggested that experience appears to effect neuronal coding in two ways: a) Cells become selective for multiple object components due to spatial and temporal association between parts; and b) more cells become tuned to views, orientations and image sizes that are commonly experienced by the observer. When learning about and becoming familiar with an object, one learns about the structural and spatial relationship between the object's parts. For example, the head is always associated with the torso and limbs of the body when viewing a moving whole body image in a crowded scene. Furthermore, the head is always attached at the shoulders enforcing the learning of that structural arrangement. Such association is reflected in the neuronal responses of the cells reported in this thesis. This association is a form of familiarity. Other familiarity concerning frequent encounter with an object in a particular viewing conditions is also suggested to be highly influential on the neuronal coding of the object. This notion is supported by the great number of cells which are selectively responsive to commonly experienced viewing conditions such as the body in its upright position and/or presented at life size for a given distance. Furthermore, the population analysis carried out in this thesis shows that such preferential coding of familiar views, orientations and image sizes leads to more efficient and faster object recognition described in psychological and behavioural experiments.

The concluding remarks of the statement given by W. A. H. Rushton (1963) quoted at the beginning of this thesis were: "Never in history has the view been more exciting than it is today when new techniques on every side (of the eye) invites us to explore gardens that for centuries have been locked to all but speculations." This was 32 years ago and we have since slowly made our way to the understanding of the underlying mechanisms of vision. Nonetheless, the statement is today more true than ever before. The development of new techniques in neuroscience have not only opened the gate to explore the 'gardens of vision', but to an entire new world of understanding the workings of the mind.

References

Albright, T. D. (1984). 'Direction and orientation selectivity of neurons in the visual area MT of the macaque.' Journal of Neurophysiology 52: 1106-1130.

Albright, T. D. (1992). 'Form-cue invarient motion processing in primate visual cortex.' Science 255: 1141-1143.

Albright, T. D., Desimone, R. and Gross, C. G. (1984). 'Columnar organization of directionally selective cells in visual area MT of the macaque.' Journal of Neurophysiology 51: 16-31.

Allison, T., Ginter, H., McCarthy, G., Nobre, A. C., Puce, A., Luby, M. and Spencer, D. D. (1994). 'Face recognition in human extrastriate cortex.' *Journal of Neurophysiology 71:*(2) 821-835.

Allison, T., McCarthy, G., Nobre, A., Puce, A. and Belger, A. (1994). 'Human extrastriate visual cortex and the perception of faces, words, numbers, and colors.' *Cerebral Cortex 5:*(4) 544-554.

Baizer, J. S., Ungerleider, L. G. and Desimone, R. (1991). 'Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques.' *Journal of Neuroscience 11:* 168-190.

Baker-Cave, C. and Kosslyn, M. (1993). 'The role of parts and spatial relations in object identification.' *Perception 22:* 229-248.

Barlow, H. (1994). 'The neuron doctrine in perception.' Cognitive Neurosciences. New York, MIT Press. 415-435.

Barlow, H. B. (1972). 'Single units and sensation: A neural doctrine for perceptual psychology.' *Perception 1:* 371-394.

Barlow, H. B. (1985). 'The twelfth Bartlett Memorial Lecture: The role of single neurons in the psychology of perception.' *Quarterly Journal of Experimental Psychology 37A:* 121-145.

Barnes, C. L. and Pandya, D. N. (1992). 'Efferent cortical connections of multimodal cortex of the superior temporal sulcus in the rhesus monkey.' *Journal of Comparative Neurology 318:* 222-244.

Baylis, G. C., Rolls, E. T. and 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 macaque monkey.' *Brain Research 342:* 91-102.

Benson, D. F., Segarra, J. and Albert, M. L. (1974). 'Visual agnosia-prosopagnosia.' *Arch. Neurology 3:* 307-310.

Besner, D. (1983). 'Visual pattern recognition: size pre-processing re-examined.' *Quarterly Journal of Experimental Psychology 35:*(A) 209-216.

Biederman, I. (1987). 'Recognition by components: a theory of human image understanding.' *Psychological Review 94:* 115-145.

Biederman, I. and Cooper, E. (1992). 'Size invariance in visual object priming.' *Journal of Experimental Psychology: Human Perception and Performance 18:*(1) 121-133.

Biederman, I. and Cooper, E. E. (1991a). 'Object recognition and laterality: null effects.' *Neuropsychologia 29:*(7) 685-694.

Biederman, I. and Cooper, E. E. (1991b). 'Priming contour-deleted images: Evidence for intermediate representations in visual object recognition.' *Cognitive Psychology 23:* 393-419.

Biederman, I. and Ju, G. (1988). 'Surface versus edge-based determinants of visual recognition.' *Cognitive Psychology 20:* 38-64.

Binford, T. O. (1971). 'Visual perception by computer.' *Paper presented at the IEEE Conference on Systems and control, December 1971, Miami.*

Boetzel, K. and Grüsser, O.-J. (1989). 'Electric brain potentials evoked by pictures of faces and non-faces: a search for "face specific" EEG potentials.' *Experimental Brain Research 77:* 349-360.

Brothers, L. and Ring, B. (1993). 'Mesial temporal neurons in the macaque monkey with responses selective for aspects of social stimuli.' *Behavioural Brain Research 57:* 53-61.

Bruce, C. (1982). 'Face recognition by monkeys: absence of an inversion effect.' *Neuropsychologia 20:*(5) 515-521.

Bruce, C. J., Desimone, R. and Gross, C. G. (1981). 'Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque.' *Journal of Neurophysiology 46:* 369-384.

Bruce, V., Valentine, T. and Baddeley, A. (1987). 'The basis of 3/4 view advantage in face recognition.' *Applied Cognitive Psychology 1:* 109-120.

Bruce, V. and Young, A. (1986). 'Understanding face recognition.' *British Journal of Psychology 77:* 103-327.

Budesen, C. and Larsen, A. (1975). 'Visual transformation of size.' *Journal of Experimental Psychology: Human Perception and Performance 1:* 214-220.

Carey, S. and Diamond, R. (1977). 'From piecemeal to configurational representation for faces.' *Science 195:* 312-314.

Cohen, L., Grey, F., Meyrignac, C., Dehaene, S. and Degos, J.-D. (1994). 'Selective deficit of visual size perception: two cases of hemimicropsia.' *Journal of Neurology, Neurosurgery, and Psychiatry 57:* 73-78.

Cooper, E. and Biederman, I. (1991). 'Evidence for size-invariant representations in visual object recognition (Poster Abstract).' *Investigative Opthalmology and Visual Science 32:*(4) 1279.

Cooper, L. A. and Shepard, R. N. (1973). 'Chronometric studies of the rotation of mental images.' New York, Academic Press.

Corballis, M. (1988). 'Recognition of disoriented shapes.' *Psychological Review 951:* 115-123.

Corballis, M. C. and Nagourney, B. A. (1978). 'Decisions about identity and orientation of rotated letters and digits.' *Canadian Journal of Psychology 32:* 186-188.

Cowey, A. and Gross, C. (1970). 'Effects of foveal prestriate and inferotemporal lesions on visual discrimination by rhesus monkeys.' *Exp. Brain Research 11:* 128-144.

Cowey, A. and Weiskrantz, L. (1967). 'A comparison of the effects of inferior temporal and striate cortex lesions on the visual behaviour of rhesus monkeys.' *The Quarterly Journal of Experimental Psychology 15:* 91-115.

Cragg, B. G. (1969). 'The topography of the afferent projections in the circumstriate visual cortex (C.V.C.) of the monkey studied by the Nauta method.' *Vision Research 9:* 733-747.

Cutzu, F. and Edelman, S. (1992). 'View-point-dependence of response time in object recognition.' *Weizmann Institute of Science Technical report:*

De Renzi, E., Perani, D., Carlesimo, G. A., Silveri, M. C. and Fazio, F. (1994). 'Prosapognosia can be associated with damage confined to the right hemisphere - an MRI and PET study and a review of the literature.' *Neuropsychologia 32:*(8) 893-902.

de Schonen, S. (1992). 'Discussion after: Becoming a face expert; by S. Carey.' Processing the Facial Image. Philosophical Transactions of the Royal Society of London B. 102.

De Valois, R. L., Albrecth, D. G. and Thorell, L. G. (1982). 'Spatial frequency selectivity of cells in macaque visual cortex.' *Vision Research 22:* 545-559.

De Yoe, E. A. and Van essen, D. C. (1985). 'Segregation of efferent connections and receptive field properties in visual area V2 of the macaque.' *Nature 317:* 58-61.

De Yoe, F. A. and Van-Essen, D. C. (1988). 'Concurrent processing streams in monkey visual cortex.' *Trends in Neuroscience 11:* 219-226.

Dean, P. (1976). 'Effects of inferotemporal lesions on the behaviour of monkeys.' *Psychological Bulletin 83:* 41-71.

Dean, P. and Weiskrantz, L. (1974). 'Loss of preoperative habits in rhesus monkeys with inferotemporal lesions: recognition failure or relearning deficit?' *Neuropsychologia 12:* 299-311.

Desimone, R., Albright, T. D., Gross, C. G. and Bruce, C. (1984). 'Stimulus-selective properties of inferior temporal neurons in the macaque.' *Journal of Neuroscience 4:*(8) 2051-2062.

Desimone, R., Fleming, J. and Gross, C. G. (1980). 'Prestriate afferents to inferior temporal cortex: an HRP study.' *Brain 184:* 41-55.

Desimone, R. and Gross, C. G. (1979). 'Visual areas in the temporal cortex of the macaque.' *Journal of Neuroscience 4:* 2051-2062.

Desimone, R., Schein, S. J., Morgan, J. and Ungerleider, L. G. (1985). 'Contour, color and shape analysis beyond the striate cortex.' Vision Research 25: 441-452.

Desimone, R. and Schein, S. J. (1987). 'Visual properties of neurons in area V4 of the macaque - sensitivity to stimulus form.' *Journal of Neurophysiology 57:*(3) 835-868.

Desimone, R., Wessinger, M., Thomas, L. and Schneider, W. (1990). 'Attentional control of visual perception: cortical and subcortical mechanisms.' *Cold Spring Habor Symposia on Quantitative Biology LV:* 963-971.

Dickerson, J. A. (1991). 'The recognition of transformed objects (PhD Thesis).' Birkbeck College, University of London.

Dittrich, W. (1990). 'Representation of faces in longtailed macaques (macaca fascicularis).' *Ethology 85:* 265-278.

Eacott, M. J., Heywood, C. A., Gross, C. G. and Cowey, A. (1993). 'Visual discrimination impairments following lesions of the superior temporal sulcus are not specific for facial stimuli.' *Neuropsychologia 31:*(6) 609-619.

Edelman, S. and Bülthoff, H. H. (1990). 'View-point-specific representations in three-dimensional object recognition.' Massachusetts Institute of Technology.

Edelman, S. and Bülthoff, H. H. (1992). 'Orientation dependence in the recognition of familiar and novel views of three dimensional objects.' *Vision Research 32:*(12) 2385-2400.

Edelman, S. and Weinshall, D. (1991). 'A self-organising multiple-view representation of 3D objects.' *Biological Cybernetics 63:* 209-219.

Ellis, R., Allport, D. A., Humphreys, G. W. and Collis, J. (1989). 'Varieties of object constancy.' *Quarterly Journal of Experimental Psychology 41:*(A) 775-796.

Farah, M. J. and Hammond, K. M. (1988). 'Mental rotation and orientation-invariant object recognition: Dissociable processes.' *Elsevier Science Publishers B.V.*

Felleman, D. and Van Essen, D. (1991). 'Distributed hierarchical processing in the primate cerebral cortex.' *Cerebral Cortex 1:* 1-47.

Flechsig, P. (1896). 'Gehirn und Seele.' Leibzig: von Veit.

Flechsig, P. (1920). 'Anatomie des menschlichen Gehirns und Rückenmarks auf myelogentischer Grundlage.' Leibzig: Thieme.

Flin, R. and Dziurawiec, S. (1989). 'Development factors in face processing.' Handbook of research on face processing. Amsterdam, North Holland.

Foldiak, P. (1990). 'Forming sparse representation by local anti-Hebbian learning.' *Biological Cybernetics 64:* 165-170.

Foldiak, P. (1991). 'Learning invariance from transformation sequences.' *Neural Computation 3:* 194-200.

Foldiak, P. (1993). 'The 'ideal homunculus': Statistical inference from neural population responses.' Computation and Neural System. Norwell, MA, USA, Kluwer Academic Publishers. 55-60.

Freeman, W. J. (1975). 'Mass action in the nervous system.' New York, Academic Press.

Fujita, I., Tanaka, K., Ito, M. and Cheng, K. (1992). 'Columns for visual features of objects in monkey inferotemporal cortex.' *Nature 36:* 343-346.

Fukamachi, A., Ohye, C. and Narabayashi, H. (1973). 'Delineation of the thalamic nuclei with a microelectrode in stereotaxic surgery for parkinsonism and cerebral palsy.' Journal of Neurosurgery 39: 214-225.

Georgopoulos, A. P., Kalaska, J. F., Crutcher, M. D., Caminitty, R. and Massey, J. T. (1994). 'The representation of movement direction in the motor cortex: Single cell and population studies.' Dynamic aspects of neocortical function. New York, Wiley. 47-52.

Gerhardstein, P. C. and Biederman, I. (1992). 'Recognising depth-rotated objects: evidence for 3D view-point invariance.' Investigative Opthalmology and Visual Sciences,

Goodale, M. A. (1993). 'Visual pathways supporting perception and action in the primate cerebral cortex.' *Current Opinion in Neurobiology 3:* 578-585.

Goodale, M. A. and Milner, A. D. (1992). 'Separate visual pathways for perception and action.' *Trends in Neurosciences 15:*(1) 20-25.

Goodale, M. A., Milner, A. D., Jacobson, L. S. and Carey, D. P. (1991). 'A neurological dissociation between perceiving objects and grasping them.' *Nature 349:*(6305) 154-156.

Gross, C. G. (1973). 'Visual functions of inferotemporal cortex.' Handbook of sensory Physiology VII/3. Berlin: Springer.

Gross, C. G. (1992). 'Representation of visual stimuli in inferior temporal cortex.' *Philosophical Transactions of the Royal Society of London (Series B) 335:* 3-10.

Gross, C. G., Bender, D. B. and Mishkin, M. (1977). 'Contributions of the corpus callosum and the anterior commisure to visual activation of inferior temporal neurons.' *Brain Research 131:* 227-239.

Gross, C. G., Bender, D. B. and Rocha-Miranda, C. E. (1969). 'Visual receptive fields of neurons in inferotemporal cortex of the monkey.' *Science 166:* 1303-1306.

Gross, C. G., Rocha-Miranda, C. E. and Bender, D. B. (1972). 'Visual properties of neurons in inferotemporal cortex of the macaque.' *Journal of Neurophysiology 35:* 96-111.

Gross, C. G., Desimone, R., Albright, T. D. and Schwartz, E. L. (1985). 'Inferior temporal cortex and pattern recognition.' Pattern Recognition Mechanisms. New York, Springer-Verlag. 179-201.

Haenny, P. E. and Schiller, P. H. (1988). 'State dependent activity in monkey visual cortex: 1. single cell-activity in V1 and V4 on visual tasks.' *Experimental Brain Research 69:*(2) 225-244.

Hamilton, C. R. and Vermeire, B. A. (1988). 'Complementary hemispheric specialization in monkeys.' *Science 242:* 1691-1694.

Harries, M. H. and Perrett, D. I. (1991). 'Visual processing of faces in the temporal cortex: physiological evidence for a modular organization and possible anatomical correlates.' *Journal of Cognitive Neuroscience 3:*(1) 9-24.

Harris, C. S. (1980). 'Insight or out of sight? Two examples of perceptual placicity in the human adult.' Visual Coding and Adaptability. Hillsdale, Erlbaum. 95-149.

Hasselmo, M. E., Rolls, E. T. and Baylis, G. C. (1989a). 'The role of expression and identity in the face-selective responses 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. and Nalwa, V. (1989b). 'Object-centred encoding by face-selective neurons in the cortex of the superior temporal sulcus of the monkey.' *Exp. Brain Res. 75:* 417-429.

Henry, G. H., Dreher, B. and Bishop, P. O. (1974). 'Orientation specificy of cells in cat striate cortex.' *Journal of Neurophysiology 37:* 1394-1409.

Hietanen, J. K., Perrett, D. I., Oram, M. W., Benson, P. J. and Dittrich, W. H. (1992). 'The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex.' *Experimental Brain Research 89:* 157-171.

Hinton, G. (1981). 'A parallel computation that assigns canonical object-based frames of reference.' 7th International Joint Conference on Artificial Intelligence, Vancouver, Canada.

Hoffman, D. D. and Richards, W. A. (1984). 'Parts of recognition.' *Cognition 18:* 65-96.

Holmes, E. J. and Gross, C. G. (1984a). 'Effects of inferior temporal lesions on discrimination of stimuli differing in orientation.' *Journal of Neuroscience 4:*(12) 3063-3068.

Holmes, E. J. and Gross, C. G. (1984b). 'Stimulus equivalence after inferior temporal lesions in monkeys.' *Behavioural Neuroscience 98:* 898-901.

Hubel, D. H. and Wiesel, T. N. (1962). 'Receptive fields of single neurons in the cat's striate cortex.' *Journal of Physiology 148:* 574-591.

Hubel, D. H. and Wiesel, T. N. (1968). 'Receptive fields and functional architecture of monkey striate cortex.' *Journal of Physiology 195:* 215-243.

Hubel, D. H. and Wiesel, T. N. (1970). 'Stereoscopic vision in macaque monkey.' *Nature 225:* 41-42.

Hummel, J. E. and Biederman, I. (1992). 'Dynamic binding in a neural network for shape recognition.' *Psychological Review 99:*(3) 480-517.

Humphrey, G. K. (1989). 'Visual object identification: some effects of image foreshortening and monocular depth cues.' Computational Processes in Human Vision: An Interdisciplinary Perspective (in press).

Humphrey, N. K. and Weiskrantz, L. (1969). 'Size constancy in monkeys with inferotemporal lesions.' *Experimental Psycholgy 21:* 225-238.

Humphreys, G. (1983). 'Reference frames and shape perception.' *Cognitive Psychology 15:* 151-196 .

Humphreys, G. (1984). 'Shape constancy: The effects of changing shape orientation and the effects of changing focal features.' *Perception and Psychophysics 36:* 50-64.

Humphreys, G. W. and Quinlan, P. T. (1987). 'Normal and pathological processes in visual object constancy.' Visual object processing: A cognitive neurological approach. London: Lawrence Erlbaum Associates Ltd.

Intrator, N. and Gold, I. (1992). '3-D object recognition using unsupervised feature extraction.' *Neuronal Computation 4:*(1) 98-107 .

Ito, M., Tamura, H., Fujita, I. and Tanaka, K. (1995). 'Size and position invariance of neuronal responses in monkey inferotemporal cortex.' *Journal of Neurophysiology 73:*(1) 218-226.

Iwai, E. and Mishkin, M. (1968). 'Two visual foci in the temporal lobe of monkeys.' Neurophysiological Basis of Learning and Behaviour. Osaka University Press,

Iwai, E. and Mishkin, M. (1990). 'Vision, Memory and the Temporal Lobe.' Netherlands, Holland, Elsvier.

Jakobson, L. S., Archibald, Y. M., Carey, D. P. and Goodale, M. A. (1991). 'A kinematic analysis of reaching and grasping movements in a patient recovering from optic ataxia.' *Neuropsychologia 29:*(8) 803-809.

Jeannerod, M. (1981). Attention and Performance IX. Hillsdale, New Jersey, Erlbaum. 153-168.

Jeannerod, M., Decety, J. and Michel, F. (1994). 'Impairment of grasping movements following a bilateral posterior parietal lesion.' *Neuropsychologia 32:*(4) 369-380.

Jeffreys, D. A. (1989). 'A face-responsive potential recorded from the human scalp.' *Experimental Brain Research 78:* 193-202.

Jeffreys, D. A. (1993). 'The influence of stimulus orientation on the vertex-positive scalp potential evoked by faces.' *Experimental Brain Research 96:*(1) 163-172.

Jeffreys, D. A., Tukmachi, E. S. A. and Rockley, G. (1992). 'Evoked potential evidence for human brain mechanisms that respond to single, fixated faces.' *Brain Research 91:* 315-362.

Jeffreys, D. A. and Turkmachi, E. S. A. (1992). 'The vertex-positive scalp potential evoked by faces and objects.' *Experimental Brain Research 91:* 340-350.

Jolicoeur, P. (1985). 'The time to name disorientated natural objects.' *Memory and Cognition 13:*(4) 289-303.

Jolicoeur, P. (1987). 'A size-congurency effect in memory for visual shape.' *Memory and Cognition 15:* 531-543.

Jolicoeur, P. (1992). 'Identification of Disoriented Objects: A Dual-system Theory.' Blackwell, Cambridge, Mass.

Jolicoeur, P. and Besner, D. (1987). 'Additivity and interaction between size ratio and response category in the comparison of size-discrepant shapes.' *Journal of Experimental Psychology: Human Perception and Performance 13:*(478-487)

Jones, E. G. and Burton, H. (1976). 'Areal differences in the laminar distribution of thalmic afferents in cortical fields of the insular, parietal and temporal regions of primates.' *The Journal of Comparative Neurology 168:* 197-248.

Kelly, P. J., Ahlskog, J. E., Goerss, S. J., Daube, J. R., Duffy, J. R. and Kall, B. A. (1987). 'Computer-assisted stereotaxic ventralis lateralis thalomotomy with microelectrode recording control in patients with parkinson's disease.' Mayo Clinic Proceedings 62: 655-664.

Kendrick, K. M. and Baldwin, B. A. (1987). 'Cells in temporal cortex of conscious sheep can respond preferentially to the sight of faces.' *Science 236:* 448-450.

Kimura, D. (1963). 'Right temporal lobe damage.' *Arch. Neurology 8:* 264-271.

Kobatake, E. and Tanaka, K. (1993). 'Selectivity for features beyond orientation, color, size and simple texture in the prestirate areas V2 and V4.' *Society of Neuroscience Abstracts 18:* 146.

Kobatake, E. and Tanaka, K. (1994). 'Neural selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex.' *Journal of Neurophysiology 71:*(3) 856-867.

Koenderink, J. J. and van Doorn, A. J. (1976). 'The singularities of the visual mapping.' Biological Cypernetics.

Koenderink, J. J. and van Doorn, A. J. (1979). 'The internal representation of solid shape with respect to vision.' *Biological Cybernetics 32:* 211-216.

Kohoen, T., Oja, E. and Lehtio, P. (1981). 'Storage and processing of information in distributed associative memory systems.' Parallel models of associative memory. Hillsdale, N.J., Laurace Erlbaum Associates Inc.

Komatsu, H. and Ideura, Y. (1993). 'Relationships between color, shape and pattern selectivities of neurons in the inferior temporal cortex of the monkey.' *Journal of Neurophysiology 70:*(2) 677-694.

Konorski, J. (1967). 'Intergrative activity of the brain.' Chicago, Illinois, University of Chicago Press.

Kosslyn, S. M., Flynn, R. A., Amsterdam, J. B. and Wang, G. (1990). 'Components of high-level vision: A cognitive neuroscience analysis and account of neurological syndromes.' *Cognition 34:* 203-277.

Kulikowski, J. J. and Bishop, P. O. (1981). 'Linear analysis of the responses of simple cells in the cat visual cortex.' *Experimental Brain Research 44:* 386-400.

Kypers, H. G. J. M., Scwarcbart, M. K., Mishkin, M. and Rosvold, H. E. (1965). 'Occipitotemporal corticocortical connections in the rhesus monkey.' *Experimental Neurology 11:* 245-262.

Landsdell, H. (1968). 'Effect of extent of temporal lobe ablations on two lateralized deficits.' *Physiological Behaviour 3:* 271-273.

Larsen, A. (1985). 'Pattern matching: Effects of size ratio, angular difference in orientation, and familiarity.' *Perception and Psychophysics 38:* 63-68.

Larsen, A. and Bundesen, C. (1978). 'Size scaling in human pattern recognition.' *Journal of Experimental Psychology: Human Perception and Performance 4:* 1-20.

Leonard, C. M., Rolls, E. T., Wilson, F. A. W. and Bylis, G. C. (1985). 'Neurons in the amygdala of the monkey with responses selective for faces.' *Behavioural Brain Research 15:* 159-176.

Lindsay, P. H. and Norman, D. A. (1972). 'Human information processing New York: Academic Press.'

Livingstone, M. S. and Hubel, D. H. (1984). 'Anatomy and physiology of a colour system in primate visual cortex.' *Journal of Neuroscience 4:*(1) 309-356.

Livingstone, M. S. and Hubel, D. H. (1987). 'Psychophysical evidence for separate channels for the perception of form, colour, movement and depth.' *Journal of Neuroscience 7:* 3416-3468.

Logothetis, N. K., Pauls, J., Bulthoff, H. H. and Poggio, T. (1994). 'View-dependent object recognition by monkeys.' *Current Biology 4:*(5) 401-414.

Logothetis, N. K., Pauls, J. and Poggio, T. (1995). 'Shape representations in the inferior temporal cortex of monkeys.' *Current Biology 5:* 552-563.

Logothetis, N. K., Vetter, T., Hurlbert, A. and Poggio, T. (in press). 'Models of 3D object recognition: invariance to image transformations.' *Cerebral Cortex.*

Lowe, D. G. (1985). 'Perceptual Organisation and Visual Recognition.' Boston, MA. Kluwer-Nijhoff.

Lowe, D. G. (1987). 'Three-dimensional object recognition form single two-dimensional images.' *Artificial Intelligence 31:* 355-395.

Lueschow, A., Miller, E. K. and Desimone, R. (1994). 'Inferior temporal mechanisms for invariant object recognition.' *Cerebral Cortex 5:*(4) 523-531.

Marr, D. (1977). 'Analysis of occluding contour.' *Proceedings of the Royal Society of London 197:*(Series B) 441-475.

Marr, D. (1982). 'Vision: A computational investigation into the human representation and processing of visual information.' San Francisco, W. H. Freeman and Company.

Marr, D. (1987). 'Understanding vision from images to shapes.' Matters of Intelligence. D. Reidel Publishing Company. 7-58.

Marr, D. and Nishihara, H. K. (1978). 'Representation and recognition of the spatial organization of three dimensional shapes.' *Proceedings of the Royal Society of London 200:*(Series B) 269-294.

Marr, D. and Vaina, L. (1982). 'Representation and recognition of the movements of shapes.' *Proceedings of the Royal Society of London (Series B) 214:* 501-524.

Maunsell, J. H. R. (1992). 'Functional visual streams.' *Current Opinion in Neurobiology 2:* 506-510.

Maunsell, J. H. R. and Newsome, W. T. (1987). 'Visual processing in monkey extrastriate cortex.' *Annual Review of Neuroscience 10:* 363-401.

McFie, J., Piercy, M. F. and Zangwill, O. L. (1950). 'Visual-spatial agnosia associated with lesions of the right cerebral hemisphere.' *Brain 73:* 167-190.

McMullen, P. A. and Farah, M. J. (1991). 'Viewer-centred and object-centred representations in the recognition of naturalistic line drawings.' *American Psychological Society 2:*(4) 275-277.

McMullen, P. A. and Jolicoeur, P. (1990). 'The spatial frame of reference in object naming and discrimination of left-right reflections.' *Memory and Cognition 18:*(1) 99-115.

McMullen, P. A. and Jolicoeur, P. (1992). 'The reference frame and effects of orientation on finding the top of rotated objects.' *Journal of Experimental Psychology: Human Perception and Performance 18:* 807-820.

Meadows, J. C. (1974). 'The anatomical basis of prosopagnosia.' *Journal of Neurology, Neurosurgery and Psychiatry 37:* 489-501.

Merrill, E. G. and Ainsworth, A. (1972). 'Glass-coated platinum-plated tungsten microelectrodes.' *Med. Biol. Eng. 10:* 662-672.

Mikami, A., Nakamura, K. and Kubota, K. (1992). 'Comparison of single neuronal activities in the temporal cortex and the amygdala of the rhesus monkey during a visual discrimination and memory task.' Brain mechanisms of perception and memory. New York, Oxford, Oxford University Press. 405-412.

Miller, E. K. and Desimone, R. (1991). 'A neural mechanism for working memory and recognition memory in inferior temporal cortex.' *Science 254:* 1377-1379.

Miller, E. K. and Desimone, R. (1994). 'Parallel neuronal mechanisms for short-term memory.' *Science 263:* 520-522.

Miller, E. K., Gochin, P. M. and Gross, C. G. (1993). 'Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus.' *Brain Research 616:* 25-29.

Miller, E. K., Li, L. and Desimone, R. (1993). 'Activity of neurons in anterior inferior temporal cortex during a short-term memory task.' *Journal of Neuroscience 13:*(4) 1460-1478.

Milner, A. D. and Goodale, M. A. (1993). 'Visual pathways to perception and action.' *Progress in Brain Research 95:* 317-337.

Milner, A. D., Perrett, D. I., Johnston, R. S., Benson, P. J., Jordan, T. R., Heeley, T. R., Bettucci, D. W., Mortara, D., Mutani, R., Terazzi, E., et al. (1991). 'Perception and action in visual form perception.' *Brain 114:* 405-428.

Milner, B. (1958). 'Psychological defects produced by temporal-lobe excision.' *Res. Publ. Assoc. Res. Nerv. Ment. Dis. 36:* 244-257.

Milner, B. (1968). 'Visual recognition and recall after right temporal excision in man.' *Neuropsychologia 6:* 191-209.

Mishkin, M. (1972). 'Cortical visual areas and their interactions.' In: Brain and Human Behaviour. Berlin: Springer,

Mishkin, M., Ungerleider, L. G. and Macko, A. (1983). 'Object vision and spatial vision: Two cortical pathways.' *Trends in the Neurosciences 6:* 414-417.

Miyashita, Y. (1988). 'Neuronal correlate of visual associative long-term memory in the primate temporal cortex.' *Nature 335:* 817-820.

Miyashita, Y. (1990). 'Associative representation of visual long term memory in the neurons of the primate temporal cortex.' Vision, memory and the temporal lobe. New York, Elsevier Publishers. 75-87.

Miyashita, Y. and Chang, H. S. (1988). 'Neuronal correlate of pictorial short-term memory in the primate temporal cortex.' *Nature 331:* 68-70.

Miyashita, Y., Date, A. and H., O. (1993). 'Configurational encoding of complex visual forms by single neurons of monkey temporal cortex.' *Neuropsychologia 31:* 1119-1131.

Neisser, U. (1967). Cognitive Psychology. New York: Appleton-Century-Crofts,

Newcombe, F. and Russell, W. R. (1969). 'Dissociated visual perceptual and spatial deficits in focal lesions of the right hemisphere.' *Journal Neurological Neurosurg. Psychiat. 32:* 73-81.

Ockleford, E. M., Milner, A. D., Dewar, W. and Sneddon, I. A. (1977). 'Form perception in stumptail macaques following posterior parietal and lateral frontal lesions.' *Cortex 13:* 350-360.

Oram, M. W. and Perrett, D. I. (1992). 'Time course of neural responses discriminating different views of the face and head.' *Journal of Neurophysiology 68:* 70-84.

Oram, M. W. and Perrett, D. I. (1994a). 'Modeling visual recognition from neurobiological constraints.' *Neural Networks 7:*(6/7) 945-972.

Oram, M. W. and Perrett, D. I. (1994b). 'Neural processing of biological motion in the macaque temporal cortex.' Proceedings of SPIE. 155-165.

Oram, M. W. and Perrett, D. I. (1994c). 'Responses of anterior superior temporal polysensory (STPa) neurons to `biological motion' stimuli.' *Journal of Cognitive Neuroscience 6:* 99-116.

Oram, M. W., Perrett, D. I. and Hietanen, J. K. (1993). 'Directional tuning of motion-sensitive cells in the anterior superior temporal polysensory area of the macaque.' *Experimental Brain Research 97:* 274-294.

Overman, W. H. and Doty, R. W. (1982). 'Hemispheric-specialization displayed by man but not macaques for analysis of faces.' *Neuropsychologia 20:*(2) 113.

Palmer, S. E., Rosch, E., Chase, P. and Baddely, D. (1981). 'Canonical perspective and the perception of objects.' Attention and Performance. Hillsdale, Lawrence Erlbaum Associates Inc.

Pandya, D. N. and Yeterian, E. H. (1985). 'Architecture and connections of cortical association areas.' *Cerebral Cortex 4:* 3-61.

Perenin, M. and Vighetto, A. (1988). 'Optic ataxia: a specific disruption in visuomotor mechanisms.' *Brain 111:* 643-674.

Perrett, D. I., Benson, P. J., Hietanen, J. K., Oram, M. W. and Dittrich, W. H. (1994). 'When is a face not a face? Correlations between perception and single neurons.' The Artful Brain. Bristol,

Perrett, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., Chitty, A. J., Hietanen, J. and Ortega, J. E. (1989). 'Frameworks of analysis for the neural representation of animate objects and actions.' *Journal of Experimental Biology 146:* 87-114.

Perrett, D. I. and Harris, M. H. (1988). 'Characteristic views and the visual inspection of simple faceted and smooth objects: 'tetrahedra and potatoes'.' *Perception 17:* 703-720.

Perrett, D. I., Hietanen, J. K., Oram, M. W. and Benson, P. J. (1992). 'Organization and functions of cells responsive to faces in the temporal cortex.' *Philosophical Transactions of the Royal Society of London B 335:* 23-30.

Perrett, D. I., Mistlin, A. J., Chitty, A. J., Smith, P. A. J., Potter, D. D., Broennimann, R. and Harries, M. H. (1988). 'Specialised face processing and hemispheric asymmetry in man and monkey: Evidence from single unit and reaction time studies.' *Behavioural Brain Research 29:* 245-258.

Perrett, D. I. and Oram, M. W. (1993). 'The neurophysiology of shape processing.' *Image and Visual Computing 11:*(6) 317-333.

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J. and 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., Oram, M. W. and Wachsmuth, E. (1993). 'Understanding minds and expression from facial signals: studies at the brain cell level.' *IEEE Conference Proceedings* 3-12.

Perrett, D. I., Rolls, E. T. and Caan, W. (1982). 'Visual neurons responsive to faces in the monkey temporal cortex.' *Experimental Brain Research 47:* 329-342.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D. and Jeeves, M. A. (1984). 'Neurons responsive to faces in the temporal cortex: Studies of functional organization, sensitivity to identity and relation to perception.' *Human Neurobiology 3:* 197-208.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D. and Jeeves, M. A. (1985). 'Visual cells in the temporal cortex sensitive to face view and gaze direction.' *Proceedings of the Royal Society London B223:* 293-317.

Phelps, M. T. and Roberts, W. A. (1994). 'Memory for pictures of upright and inverted primate faces in humans(*homo sapiens*), squirrel monkeys (*saimiri sciureus*), and pigeons (*columba livia*).' *Journal of Comparative Psychology 108:*(2) 114-125.

Pinker, S. (1984). 'Visual cognition: an introduction.' *Cognition 18:* 1-63.

Plaut, D. C. and Farah, M. J. (1990). 'Visual object recognition: Interpreting neurophysiological data within a computational framework.' *Journal of Cognitive Neuroscience 2:*(4) 320-343.

Poggio, T. and Edelman, S. (1990). 'A network that learns to recognise three-dimensional objects.' *Nature 343:* 263-266.

Pohl, W. (1973). 'Dissociation of spatial discrimination deficits following frontal and parietal lesions in monkeys.' *Journal of Compmaritve Physiolsiology and Psychology 82:* 227-239.

Porrill, J., Pollard, S. B., Pridmore, T. P., Bowen, J. B., Mayhew, J. E. W. and Frisby, J. P. (1988). 'TINA: a 3D vision system for pick and place.' *Image and Vision Computing 6:*(2) 91-99.

Rapcsak, S. Z., Polster, M. R., Comer, J. F. and Rubens, A. B. (1994). 'False recognition and misidentification of faces following right hemisphere dameage.' *Cortex 30:* 565-583.

Ratcliff, G. and Newcombe, F. (1973). 'Spatial orientation in man: effects of the left, right, and bilateral posterior cerebral lesions.' *Journal of Neurol. Neurosurg. Psychiat. 36:* 448-454.

Rock, I. (1973). 'Orientation and Forms.' New York, Academic Press.

Rockland, K. S. and Pandya, D. N. (1981). 'Cortical connections of the occipital lobe in the rhesus monkey: interconnections between areas 17, 18, 19 and the superior temproal sulcus.' *Brain Research 212:* 249-270.

Rolls, E. T. (1994). 'Brain mechanisms of invariant visual recognition and learning.' *Behavioural Processes 33:* 113-138.

Rolls, E. T. and 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., Baylis, G. C. and Hasselmo, M. E. (1987). 'The responses of neurons in the cortex in the superior temporal sulcus of the monkey to band-pass spatial frequency filtered faces.' *Vision Research 27:*(3) 311-326.

Rolls, E. T., Baylis, G. C. and Leonard, C. M. (1985). 'Role of low and high spatial frequency in the face-selective responses of neurons in the cortex in the superior temporal sulcus in the monkey.' *Vision Research 25:* 1021-1035.

Rolls, E. T., Perrett, D. I., Caan, W. and Wilson, F. (1982). 'Neuronal responses related to visual recognition.' *Brain 105:* 611-646.

Rosene, D. L., Roy, N. J. and Davis, B. J. (1986). 'A cryoprotection method that facilitates cutting frozen sections of whole monkey brains for histological and

histochemical processing without freezing artefact.' *Journal of Histochem. Cytochem 34:* 1301-1315.

Rosenfeld, S. A. and van Hoesen, G. W. (1979). 'Face recognition in the rhesus monkey.' *Neuropsychologia 17:* 503-509.

Rumelhart, D. E. and Zipser, D. (1985). 'Feature discovery by competitive learning.' *Cognitive Sciences 9:* 75-112.

Sakata, H., Shibutani, H., Ito, Y., Tsurugai, K., Mine, S. and Kusunoki, M. (1994). 'Functional-properties of rotation-sensitive neurons in the posterior parietal association cortex of the monkey.' *Experimental Brain Research 101:*(2) 183-202.

Sakata, H. and Taira, M. (1994). 'Parietal control of hand action.' *Current Opinion in Neurobiology 4:* 847-856.

Schwartz, E. L., Desimone, R., Albright, T. D. and Gross, C. G. (1983). 'Shape recognition and inferior temporal neurons.' *Proceedings National Academic Science US 80:* 5776-5778.

Seibert, M. and Waxman, A. (1992a). 'Adaptive 3-D object recognition from multiple views.' *IEEE Transactions on Pattern Analysis and Machine Intelligence 14:*(2) 107-124.

Seibert, M. and Waxman, A. (1992b). 'Learning and recognizing 3D objects from multiple views in a neural system.' Neural Networks for Perception. USA, Academic Press. 426-444.

Seibert, M. and Waxman, A. M. (1991). 'Learning aspect graph representations from view sequences.' Advances in Neural Network Information Processing Systems. San Mateo, California, USA, Morgan Kaufman 258-265.

Selfridge, O. G. (1959). 'Pandemonium: A paragraph for learning.' The mechanisation of thought processes. London: H.M.S.O.

Seltzer, B. and 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.

Seltzer, B. and Pandya, D. N. (1984). 'Further observations on parieto-temporal connections in the rhesus monkey.' *Experimental Brain Research 55:* 301-312.

Seltzer, B. and Pandya, D. N. (1989). 'Frontal lobe connections of the superior temporal sulcus in the rhesus monkey.' *Journal of Comparative Neurology 281:* 97-113.

Seltzer, B. and Pandya, D. N. (1989). 'Intrinsic connections and architectonics of the superior temporal sulcus in the rhesus monkey.' *Journal of Comparative Neurology 290:* 451-471.

Seltzer, B. and Pandya, D. N. (1991). 'Post-Rolandic cortical projections of the superior temporal sulcus in the rhesus monkey.' *Journal of Comparative Neurology 312:* 625-640.

Seltzer, B. and Pandya, D. N. (1994). 'Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study.' *Journal of Comparative Neurology 343:* 445-463.

Shepard, R. N. and Cooper, L. A. (1982). 'Mental images and their transformations.' Cambridge, MA, MIT Press.

Shepard, R. N. and Metzler, J. (1971). 'Mental rotation of three-dimensional objects.' *Science 171:* 701-703.

Shipp, S. and Zeki, S. (1989). 'The organisation of connections between areas V5 and V1 in the macaque monkey visual cortex.' *European Journal of Neuroscience 1:* 309.

Shipp, S. and Zeki, S. M. (1985). 'Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex.' *Nature 315:* 322.

Snedecor, G. W. and Cochran, W. G. (1980). Statistical methods (7th edn). Iowa State University Press, Ames Iowa, 215-237.

Snodgrass, J. G. and Feenan, K. (1990). 'Priming effects in picture fragment completion: Support for the perceptual closure hypothesis.' *Journal of Experimental Psychology General 119:*(3) 276-296.

Sobotka, S. and Ringo, J. L. (1993). 'Investigation of long term recognition and association memory in unit responses from inferotemporal cortex.' *Experimental Brain Research 96:* 28-38.

Sutherland, N. S. (1968). 'Outlines of a theory of pattern recognition in animals and man.' *Proceedings of the Royal Society London B 171:* 297-317.

Taira, M., Mine, S., Georgopoulos, A. P., Murata, A. and Sakata, H. (1990). 'Parietal cortex neurons of the monkey related to the visual guidance of hand movements.' *Experimental Brain Research 83:* 29-36.

Tanaka, J. W. and Farah, M. J. (1993). 'Parts and wholes in face recognition.' *The Experimental Psychology Society* 225-245.

Tanaka, K. (1992). 'Inferotemporal cortex and higher visual functions.' *Current Opinion in Neurobiology 2:* 502-505.

Tanaka, K. (1993). 'Neuronal mechanisms of object recognition.' *Science 262:* 685-688.

Tanaka, K. and Fujita, I. (1991). 'Coding visual images of objects in the inferotemporal cortex of the macaque monkey.' *Journal of Neurophysiology 66:* 170-189.

Tanaka, K., Fujita, I., Kobatake, E., Cheng, K. and Ito, M. (1993). 'Serial processing of visual object-features in the posterior and anterior parts of the inferotemporal cortex.' Brain Mechanisms of Perception and Memory: From Neuron to Behaviour. Oxford, England, Oxford University Press. 34-46.

Tarr, M. J. and Pinker, S. (1989). 'Mental rotation and orientation dependence in shape recognition.' *Cognitive Psychology 5:* 233-282.

Thompson, P. (1980). 'Margaret Thatcher: a new illusion.' Perception 9: 483-484.

Tootell, R. B. H., Silverman, M. S., De Valois, R. L. and Jacobs, G. H. (1983). 'Functional organization of the second cortical visual area in primates.' *Science 220:* 737.

Trotter, Y., Thorpe, S. J., Pouget, A., S., C. and Imbert, M. (1989). 'Processing of orientation in V1 of the awake monkey.' *Society of Neuroscience Abstracts 15:* 1056.

Turnbull, O. H., Laws, K. R. and McCarthy, R. A. (in press). 'Object recognition without knowladge of object orientation.' *Cortex.*

Ullman, S. (1989). 'Aligning pictorial descriptions: an approach to object recognition.' *Cognition 32:* 193-254.

Ungerleider, L. G. and Brody, B. A. (1977). 'Extrapersonal spatial orientation: the role of posterior parietal, anterior frontal, and inferotemporal cortex.' *Experimental Neurology 56:* 265-280.

Ungerleider, L. G., Ganz, L. and Pribram, K. H. (1977). 'Size constancy in rhesus monkeys: effects of pulvinar, prestriate, and inferotemporal lesions.' *Experimental Brain Research 27:* 251-269.

Ungerleider, L. G. and Mishkin, M. (1982). 'Two cortical visual systems.' In: Analysis of visual behaviour. Cambridge Massachusetts: MIT Press, 549-585.

Van Essen, D. C. and Zeki, S. (1978). 'The topographic organization of rhesus monkey presriate cortex.' *Journal of Physiology (London) 277:* 193-226.

Verfaillie, K. (1992). 'Variant points of view on view-point invariance.' *Canadian Journal of Psychology 46:*(2) 215-236.

Vogels, R. and Orban, G. A. (1994). 'Activity of inferior temporal neurons during orientation discrimination with successively presented gratings.' *Journal of Neurophysiology 71:*(4) 1428-1451.

von Bonin, G., Garol, H. W. and McCulloch, W. S. (1942). 'The functional organization of the occipital lobe.' *Biological Symposium 7:* 165-192.

Wachsmuth, E., Oram, M. W. and Perrett, D. I. (1993). 'Single unit responses in the anterior superior temporal polysensory area (STPa) of the the monkey to objects and their components.' *Perception 22 (Suppl):* 134-139.

Wachsmuth, E., Oram, M. W. and Perrett, D. I. (1994). 'Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque.' *Cerebral Cortex 4:*(5) 509-522.

Walsh, V. and Butler, S. R. (in press). 'The perception of rotated shapes: lessons fron lesions.' *Behavioural Brain Research.*

Walsh, V., Butler, S. R., Carden, D. and Kulikowski, J. J. (1992). 'The effects of V4 lesions on the visual abilities of macaques: shape discrimination.' *Behavioural Brain Research 50:* 115-126.

Walsh, V. and Perrett, D. I. (1994). 'Visual attention in the occipitotemporal processing stream of the macaque.' *Cognitive Neuropsychology 11:*(2) 243-263.

Warrington, E. K. and James, M. (1986). 'Visual object recognition in patients with right hemisphere lesions: axes or features?' *Perception 15:* 355-366.

Warrington, E. K. and Taylor, A. M. (1973). 'The contribution of the right parietal lobe to object recognition.' *Cortex 9:* 152-164.

Weiskrantz, L. and Saunders, R. C. (1984). 'Impairments of visual object transforms in monkeys.' *Brain 107:* 1033-1072.

Wilson, M. (1957). 'Visual system: Pulvinar-extrastriate cortex.' In: Handbook of Behavioral Neurobiology. New York: Plenum.

Yamashiro, K., Tasker, R. R., Iwayama, K., Mori, K., Albe-Fessard, D., Dostrovsky, J. O. and Chodakiewitz, J. W. (1989). 'Evoked potentials from the human thalamus: correlation with microstimulation and single unit recordings.' Microelectrode Meeting, Evian-Les-Bains, France.

Yeterian, E. H. and van Hosen, G. W. (1978). 'Cortico-striate projections in the rhesus monkey: The organization of certain cortico-caudate connections.' *Brain Research 139:* 43-63.

Yin, R. K. (1969). 'Looking at upside-down faces.' *Journal of Experimental Psychology 81:*(1) 141-145.

Yin, R. K. (1970). 'Face recognition by brain-injured patients: a dissociable ability?' *Neuropsychologia 8:* 395-402.

Young, M. P. (1992). 'Objective analysis of the topological organization of the primate cortical visual system.' *Nature 358:* 152-155.

Young, M. P. (1993). 'The organization of neural systems in the primate cerebral cortex.' *Proceedings of the Royal Society of London B 252:* 13-18.

Young, M. P. and Yamane, S. (1992). 'Sparse population coding of faces in the inferotemporal cortex.' *Science 256:* 1327-1331.

Young, M. P. and Yamane, S. (1993). 'An analysis at the population level of the processing of faces in the inferotemporal cortex.' Brain Mechanisms of Perception and Memory: From Neuron to Behaviour. Oxford, England, Oxford University Press. 47-70.

Zeki, S. (1976). 'The functional organisation of projections from striate to prestriate visual cortex in the rhesus monkey.' *Cold Spring Habor Symposium in Quantitative Biology 40:* 591.

Zeki, S. (1989). 'Modular connections between areas V2 and V4 of the macaque monkey visual cortex.' *European Journal of Neuroscience 1:* 494.

Zeki, S. (1990a). 'Colour vision and functional specialisation in the visual cortex.' *Discussions in Neurosciences VI:*(2)

Zeki, S. (1990b). 'Parallelism and functional specialization in human visual cortex.' *Cold Spring Harbour Symposia on Quantitative Biology LV:* 651-661.

Zeki, S. and Shipp, S. (1988). 'The functional logic of cortical connections.' *Nature 335:* 311-317.

Zeki, S., Watson, J. D. G., Lueck, C. J., Friston, K. J. and Kennard, C. F. (1991). 'A direct demonstration of functional specialization in human visual cortex.' *Journal of Neuroscience 11:* 641-649.

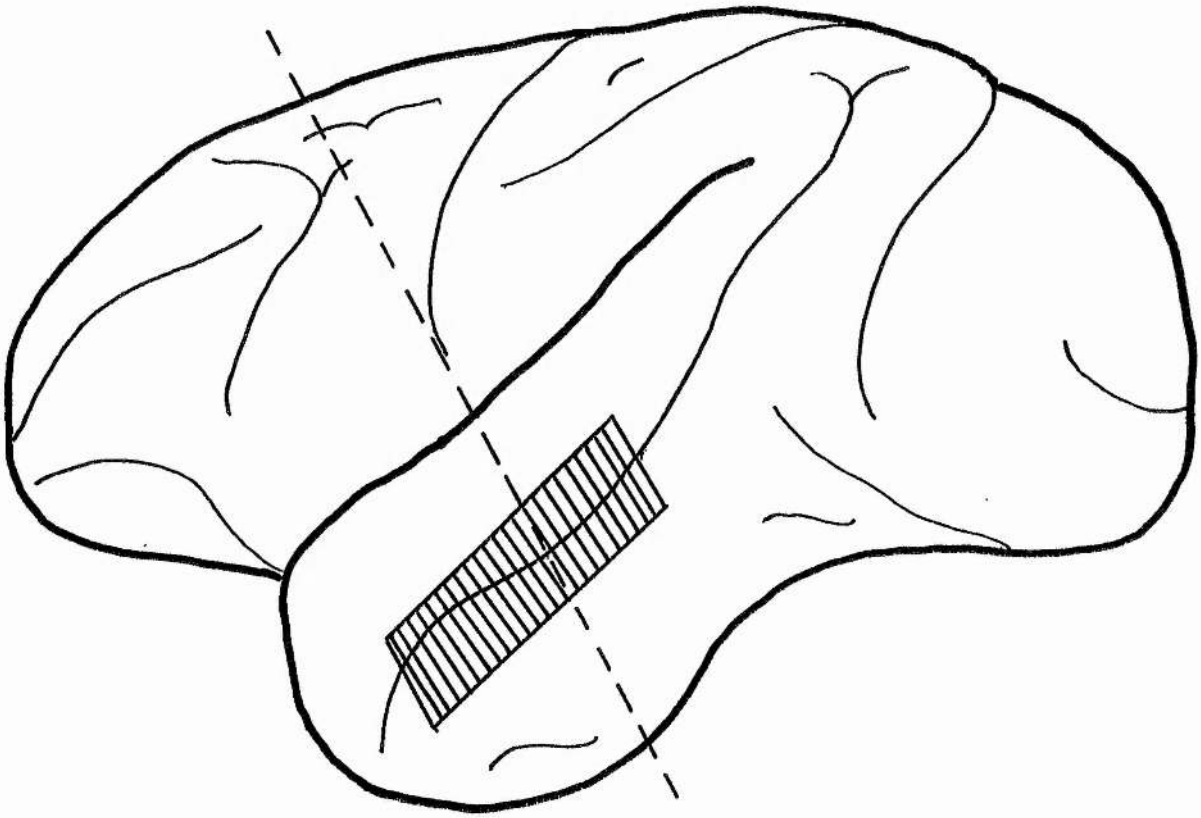
Zeki, S. M. (1978a). 'The third visual complex of rhesus monkey prestriate cortex.' *Journal of Physiology (London) 277:* 245-272.

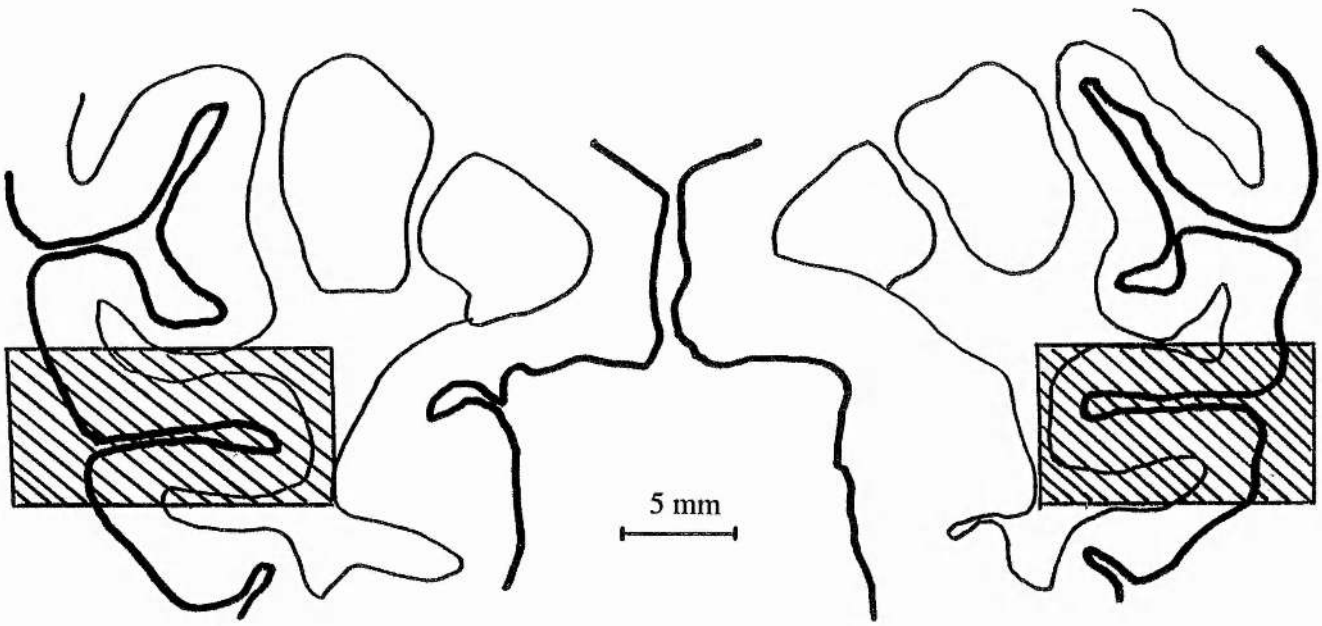
Zeki, S. M. (1978b). 'Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex.' *Journal of Physiology (London) 277:* 273-290.

Zihl, J. D., Von Cramon, D. and Mai, N. (1983). 'Selective disturbance of vision after bilateral brain damage.' *Brain 106:* 313-340.

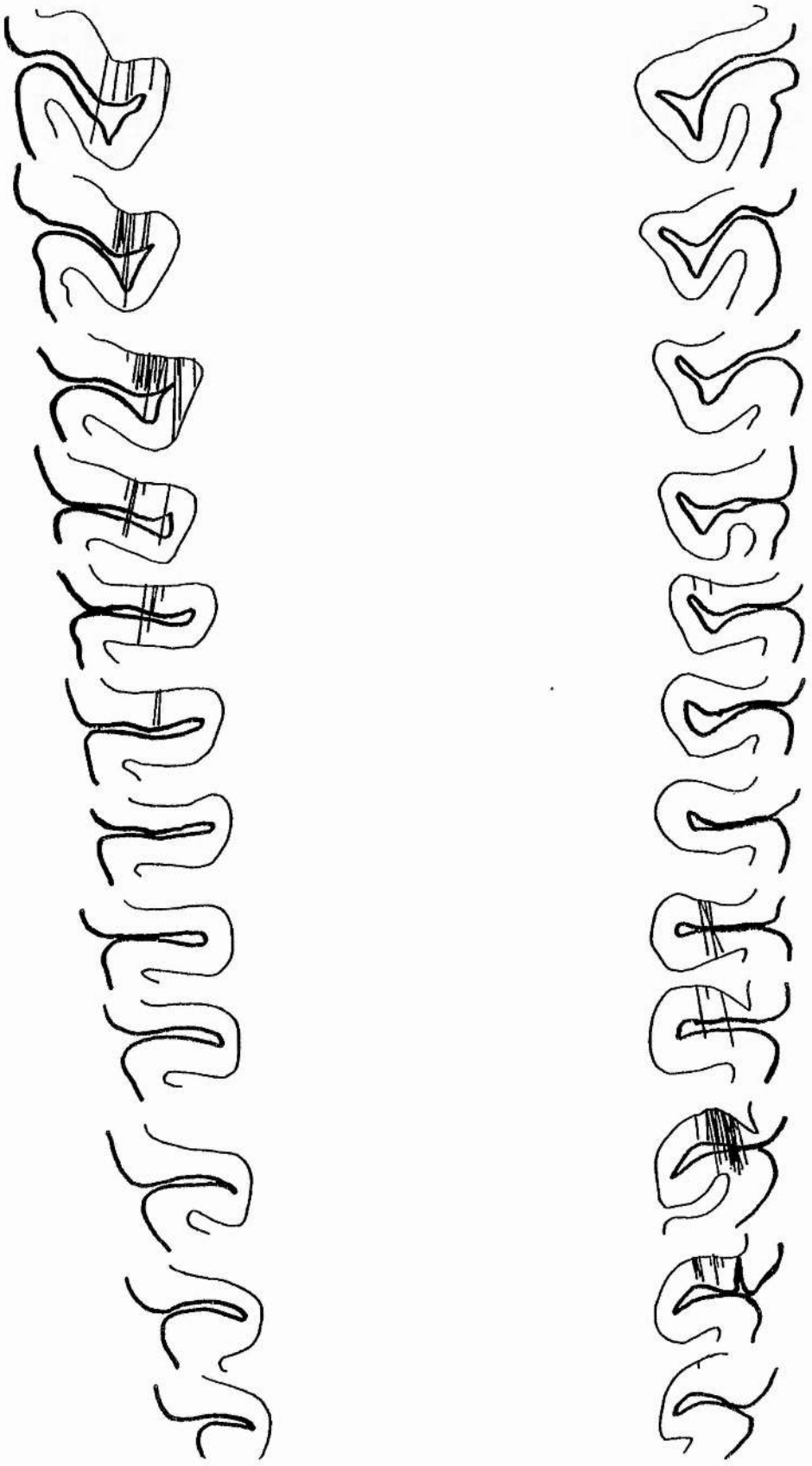
Appendix

Appendix 1. a) Schematic drawing of the left hemisphere of the macaque brain. Shaded region indicates the location area STPa. **b)** Frontal section of the brain of one monkey (E) at 13.5 mm anterior to the interaural plane. The superior temporal sulcus of the left and right hemispheres indicated by the shaded areas.

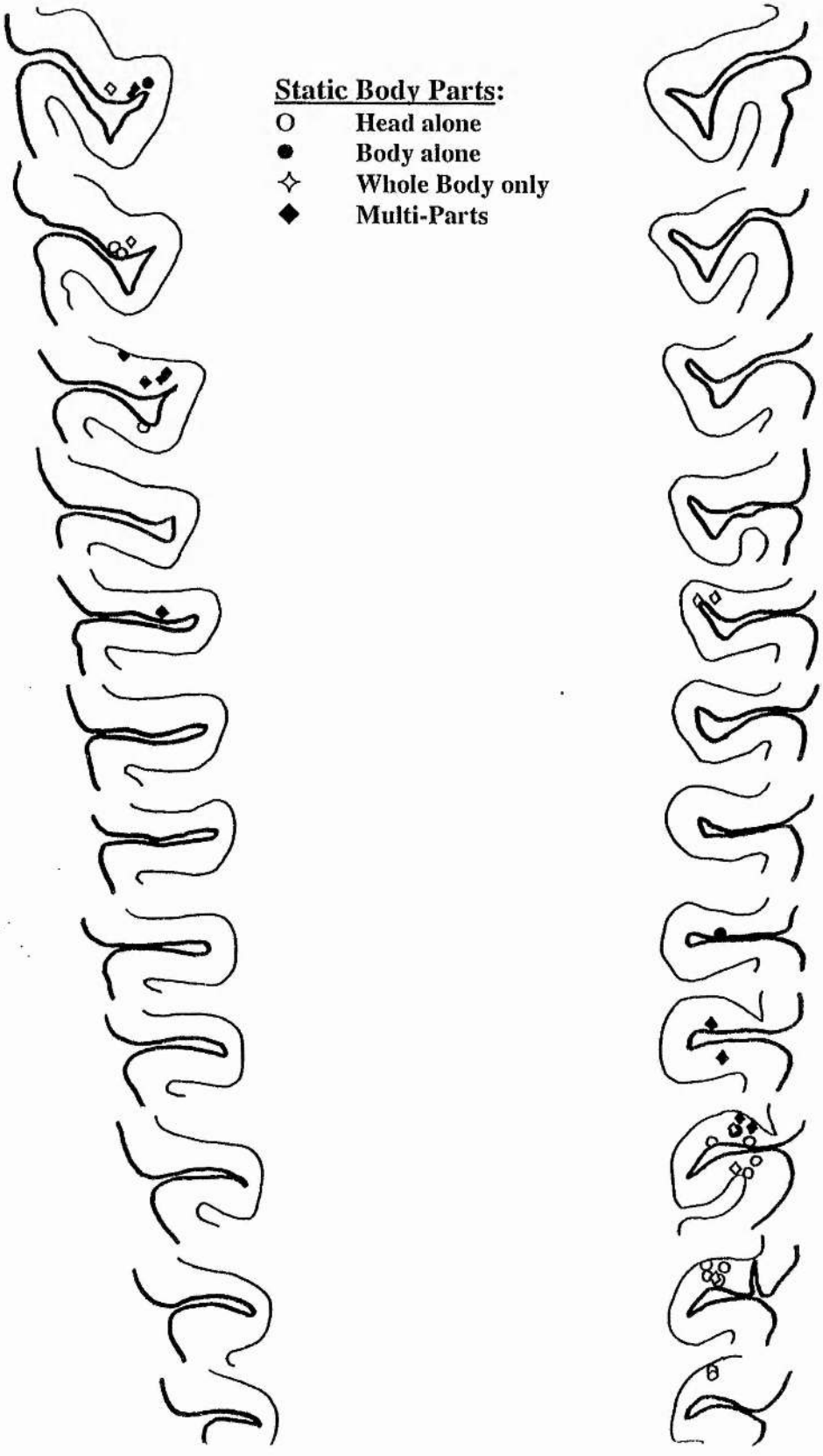




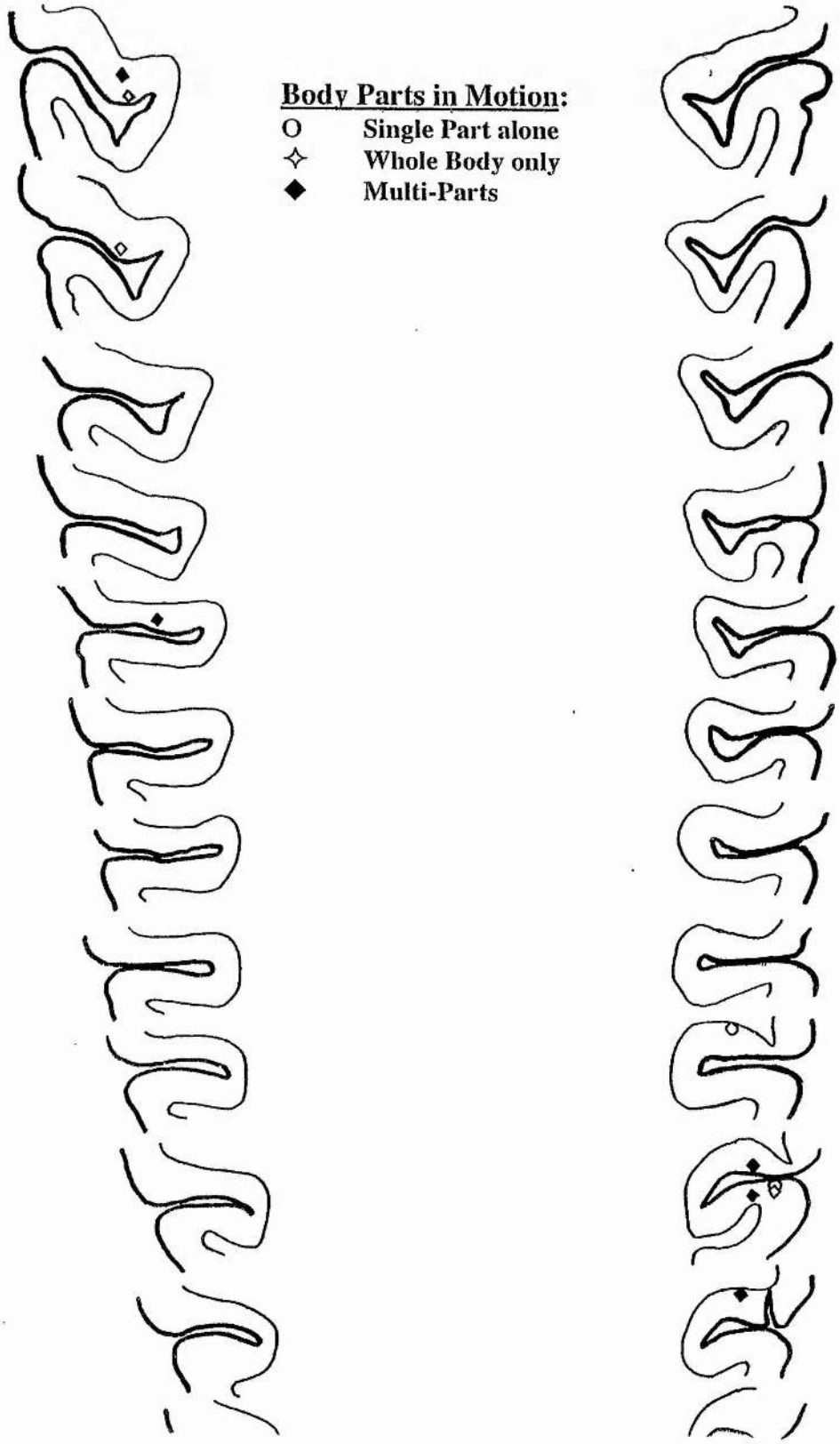
Appendix 2. Histological reconstruction: Series of frontal sections of the STS at every 1 mm (6.0 to 17.0 mm anterior to the interaural plane) reconstructed from one monkey (E) showing the location of recording tracks. The thick lines indicate the brain surface, the thin lines show the boundary between white and grey matter. Cells were located in both the upper bank and fundus of the superior temporal sulcus.



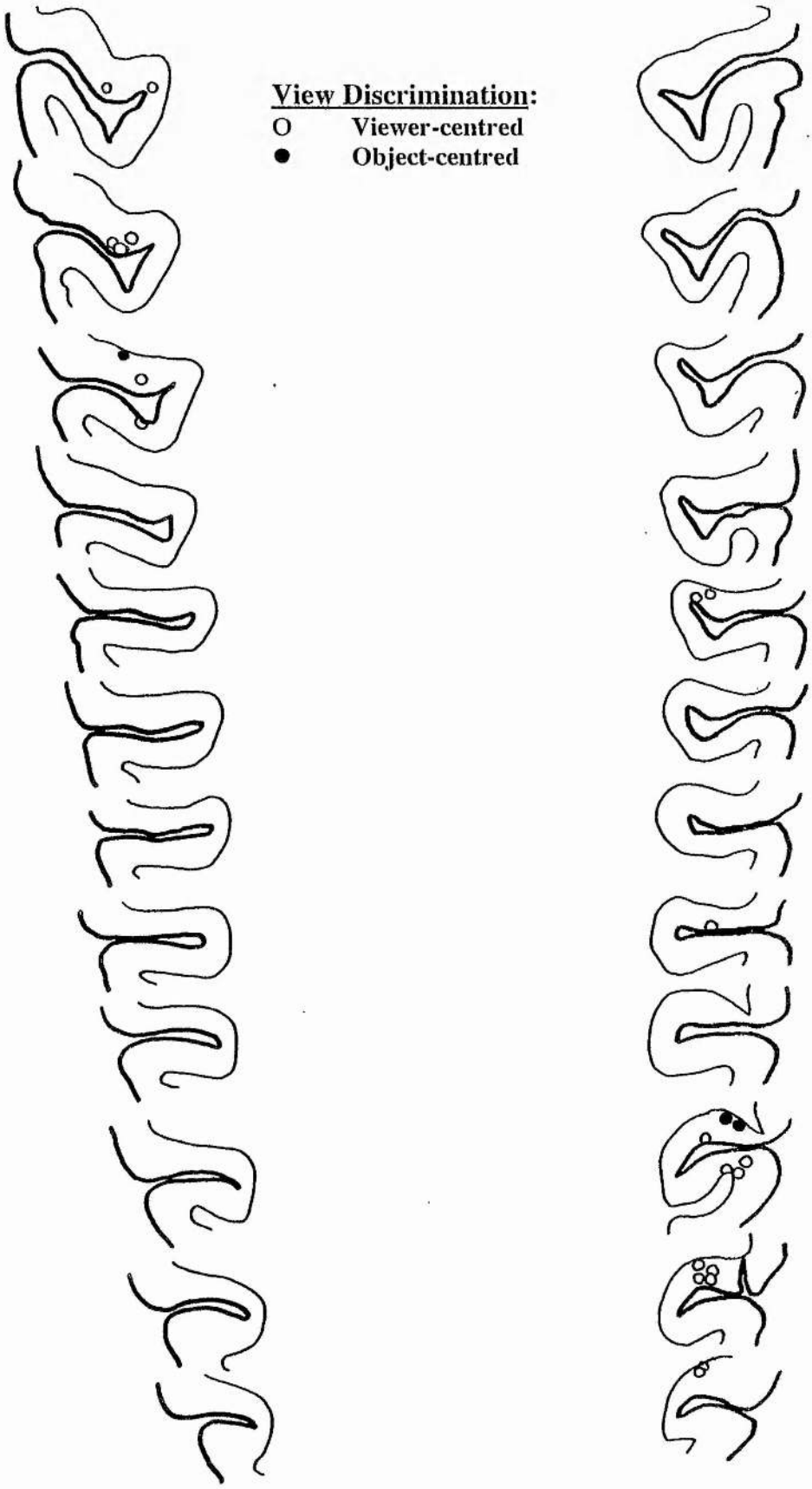
Appendix 2a. Histological reconstruction: Coding body parts. Series of frontal sections of the STS at every 1 mm (6.0 to 17.0 mm anterior to the interaural plane) reconstructed from one monkey (E) showing the location of cells responsive to the head alone (open circles), body alone (filled circles), whole body only (open triangles) and responsive to all parts (filled triangles). The thick lines indicate the brain surface, the thin lines show the boundary between white and grey matter. Cells were located in both the upper bank and fundus of the superior temporal sulcus.



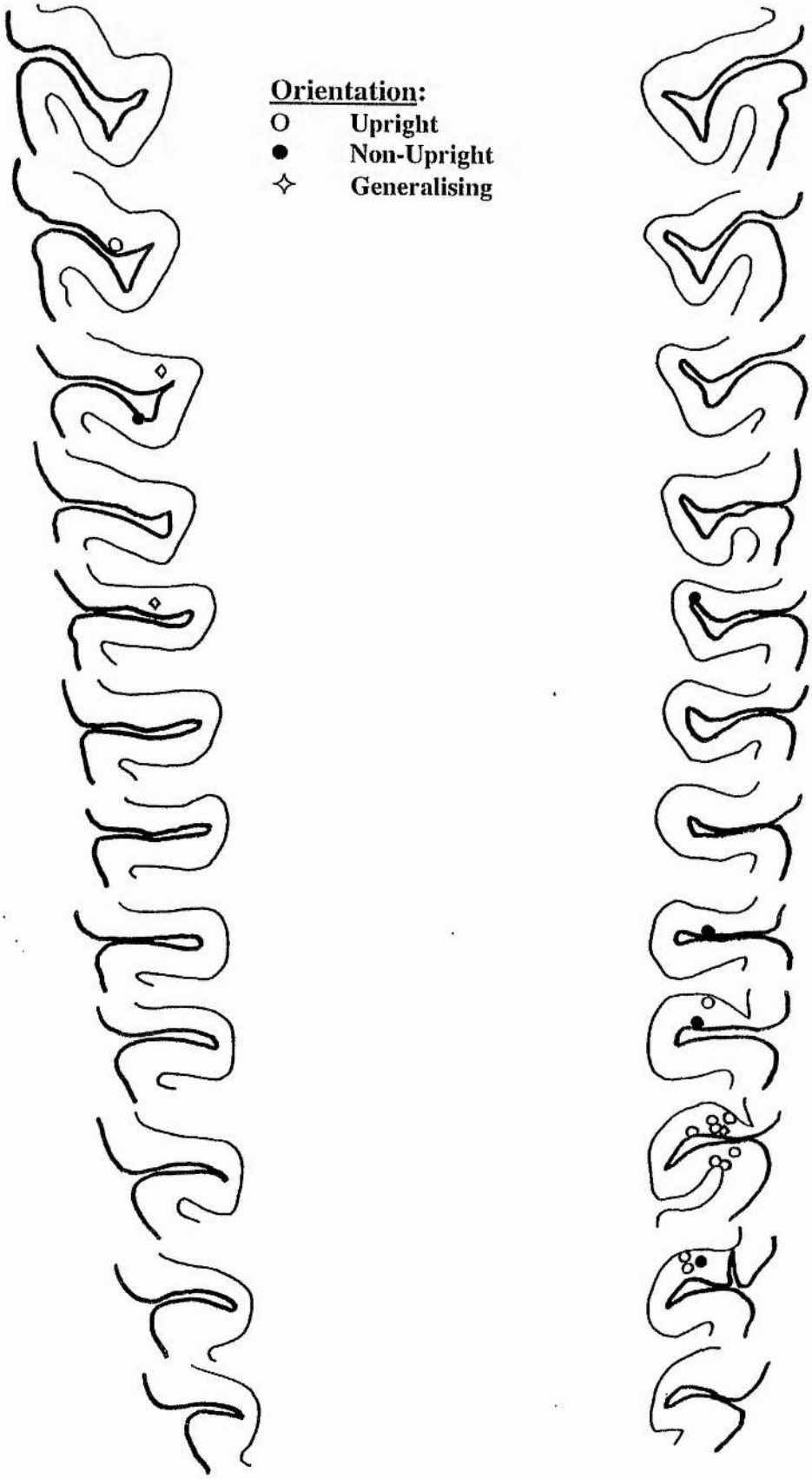
Appendix 2b. Histological reconstruction: Coding Body parts in motion. Series of frontal sections of the STS at every 1 mm (6.0 to 17.0 mm anterior to the interaural plane) reconstructed from one monkey (E) showing the location of cells responsive to the head alone moving in the optimal direction (open circles), body alone moving (filled circles), whole body only moving (open triangles) and responsive to all parts in motion (filled triangles). The thick lines indicate the brain surface, the thin lines show the boundary between white and grey matter. Cells were located in both the upper bank and fundus of the superior temporal sulcus.



Appendix 3. Histological reconstruction: View coding. Series of frontal sections of the STS at every 1 mm (6.0 to 17.0 mm anterior to the interaural plane) reconstructed from one monkey (E) showing the location of cells coding head/body information in a viewer-centred manner (open circles) and cells coding head/body information in an object-centred manner (filled circles).



Appendix 4. Histological reconstruction: Orientation coding. Series of frontal sections of the STS at every 1 mm (6.0 to 17.0 mm anterior to the interaural plane) reconstructed from one monkey (E) showing the location of cells tuned to upright orientation of the head/body (open circles), cells tuned to non-upright orientations of the head/body (filled circles), and cells generalising across all orientations of the head/body tested (open triangles).



Appendix 5. Histological reconstruction: Size coding. Series of frontal sections of the STS at every 1 mm (6.0 to 17.0 mm anterior to the interaural plane) reconstructed from one monkey (E) showing the location of cells tuned to the largest (100%, life-size) image size tested (open circles), cells tuned to size 75% (filled circles), and cells generalising across all sizes of the head/body tested (open triangles).